Bank Net Worth and Frustrated Monetary Policy

Alexander K. Zentefis*

Yale School of Management

November 19, 2017

Abstract

This paper presents a model in which the effect of monetary policy depends on the state of bank net worth. When banks are flush with equity, changes in the central bank's policy interest rate pass through fully to bank lending rates. When banks have low equity, there is no such pass-through. Banks in the model are local monopolists for borrowers near them. When they have lots of equity, they compete for customers at the edges of their markets. When they have little equity, they retreat and exploit the monopoly power over their local customers. With very low equity, banks may even *raise* lending rates after a drop in the policy rate. The model posits novel connections between aggregate bank net worth, bank competition, and the effectiveness of monetary policy.

JEL classification: E32, E44, E52, G21, L13

Keywords: banking, monetary policy, business fluctuations, industrial organization, monopolistic competition

*I am grateful to Doug Diamond, Stavros Panageas, John Cochrane, and Pietro Veronesi for many fruitful conversations about this topic and for their invaluable assistance. I also thank Amy Boonstra, Will Cong, Jason Donaldson, Radha Gopalan, João Granja, Veronica Guerrieri, Zhiguo He, Yunzhi Hu, Anjini Kochar, Tara Levens, Adam Jørring, Asaf Manela, Gregor Matvos, Michael Minnis, Stefan Nagel, Paymon Khorrami, Aaron Pancost, Jacopo Ponticelli, Raghu Rajan, Uday Rajan, Jung Sakong, Amit Seru, Amir Sufi, Chad Syverson, Fabrice Tourre, Margarita Tsoutsoura, Luigi Zingales, Eric Zwick, the Chicago Booth Corporate Finance Reading Group, the Chicago Booth Finance Brownbag, the Chicago Booth Finance Lunch Seminar, and the 13th Annual Conference on Corporate Finance for very helpful comments.
The latest version of this paper is posted here: https://phd.chicagobooth.edu/alexander.zentefis/.

1 Introduction

Interest rate pass-through *is* the interest rate channel of monetary policy. Changes to a central bank's policy interest rate are meant to pass through to retail borrowing rates of consumers and firms in order to influence aggregate spending and fixed investment. One important class of borrowing rates for firms are interest rates on commercial and industrial (C&I) bank loans.

After the financial crisis, the pass-through to these loans appeared to be impaired. Despite the Federal Reserve holding the federal funds rate near zero starting in late 2008, interest rates on C&I loans were slow to respond. As a consequence, credit spreads on these loans remained elevated. Meanwhile, delinquency rates, the fraction of nonperforming loans, and net loan charge-offs for new C&I credit returned to pre-crisis levels, suggesting more than just a default-based explanation for the prolonged high spreads. See Figures 1(a)-1(b). Figure (12) in Appendix (A.1) plots both the C&I loan spread and a corporate bond spread measured in Gilchrist and Zakrajšek (2012), which is commonly referred to as the GZ spread. The figure shows that the C&I loan spread was slower to decline than corporate bond spreads following the crisis.

During this period, bank concentration in the C&I loan market also rose strongly, with a notably sharp decline between 2008 and 2010 in the number of banks that extended this kind of firm credit. The spike in concentration coincided with the extraordinary wave of bank consolidation that took place following the financial crisis. From 2007 to 2013, 492 commercial and savings banks were put into FDIC receivership and sold at auction to acquiring banks (Granja et al. (2015)). See Figures 2(a)-2(b).

Is there a connection between the rise in bank concentration and the lack of pass-through? Indeed, one of the most robust empirical findings on impediments to interest rate pass-through is increased bank concentration (Cottarelli and Kourelis (1994); Borio and Fritz (1995); Mojon (2000); Sørensen and Werner (2006); van Leuvensteijn et al. (2008); Gigineishvili (2011)). In light of this fact, the contribution of this paper is to jointly explain an obstructed interest rate channel and a rise in bank concentration following a financial crisis.

I present a dynamic stochastic general equilibrium model of credit and the real economy to understand why bank loan spreads can remain high and fixed investment low for a protracted period after a financial crisis despite accommodative monetary policy. I posit that aggregate equity capital in the banking sector provides an explanation. My central thesis is that bank oligopoly power can persistently impede the interest rate channel as long as bank net worth remains low.

I argue that when banks are flush with equity, their required cost of equity is low, so they have incentive to compete across different parts of the loan market, because doing so is relatively cheap. In this case, competition compels banks to pass through changes in the central bank's policy interest rate to their lending rates.

However, a severe contraction in total bank equity sharply raises the cost of equity across banks, which forces them to consolidate for survival. Competing across each other's territories is no longer profitable, so instead banks become local monopolists in separate parts of the loan market.

No longer facing the same competitive pressure, banks do not pass through changes in the policy rate to their lending rates. In some cases, the central bank may lower the interest rate, but banks perversely *raise* their lending rates in response. Over time, as long as bank equity remains strained, there is no transmission via the interest rate channel. High bank loan rates consequently lower investment and output.

In the model, banks lend money to entrepreneurs who run industrial projects that produce physical capital. These projects are located around a circle, and locations on the circle represent industries or geographic areas. Banks are local monopolists for borrowers near them, but they can compete for customers at the edges of their markets.

A bank has reason to diversify its loan portfolio around the circle to increase its leverage, which is constrained because of an agency friction with depositors. Diversification is increasingly expensive, though, in that the cost of liquidating projects in default grows in the size of the bank's loan portfolio. Banks that focus on narrow parts of the loan market are more efficient at recovering value from distressed assets than are banks that operate over broad stretches. Banks trade off the costs and benefits of diversification, and in equilibrium, they specialize in classes of industries or areas.

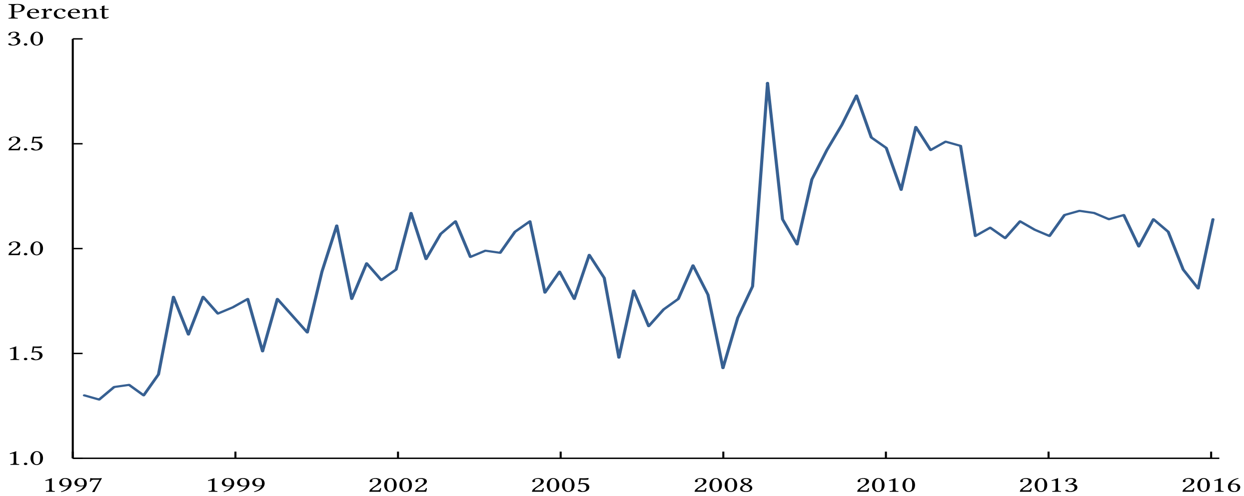
Bank market power arises out of an entrepreneur's preference to contract with a bank that is more specialized in that entrepreneur's project (i.e., closer to his or her location on the circle). Each bank can carve out a local monopoly market by offering credit at a price that entices entrepreneurs to borrow and pursue their projects. With more aggressive pricing, the bank can try to lure borrowers away from a neighboring bank, igniting competition. Hence the local monopoly power of a bank can always be softened by another bank's competitive entry.

When a bank lowers its lending rate to exactly match its neighbor's, the local monopoly markets of the two banks just touch. At this price of credit, the bank observes a kink in its residual demand curve for loans. Charging any higher lending rate shrinks the bank's local market, which the neighbor pays no attention to. Charging any lower lending rate expands the bank's market into the neighbor's territory. The amount the bank must offer as a price concession to get new customers doubles when it switches from a local monopolist to a competitor, which generates the kink. The kink plays a key role in the analysis, because a bank that operates there does not adjust its lending rate to small changes in its marginal cost of financing.

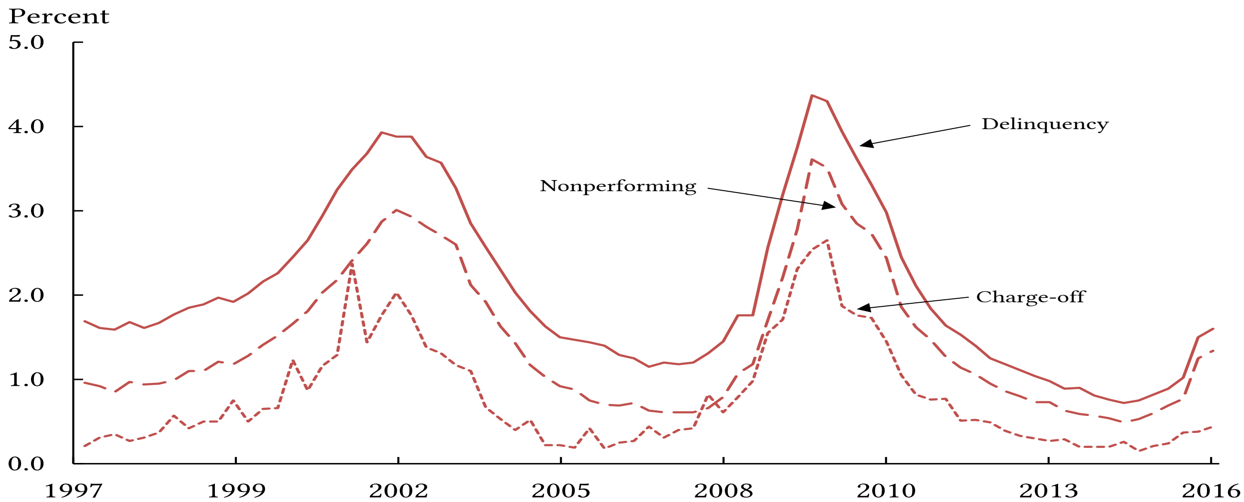
This price rigidity at the kink is the reason for an obstructed interest rate pass-through. A central implication of the model is that all banks in the credit market collectively settle at the kink

Figure 1: Commercial and Industrial Loans since the Crisis, All Commercial Banks

(a) C&I Loan Spread



(b) Delinquency, Nonperforming, Charge-off Rates

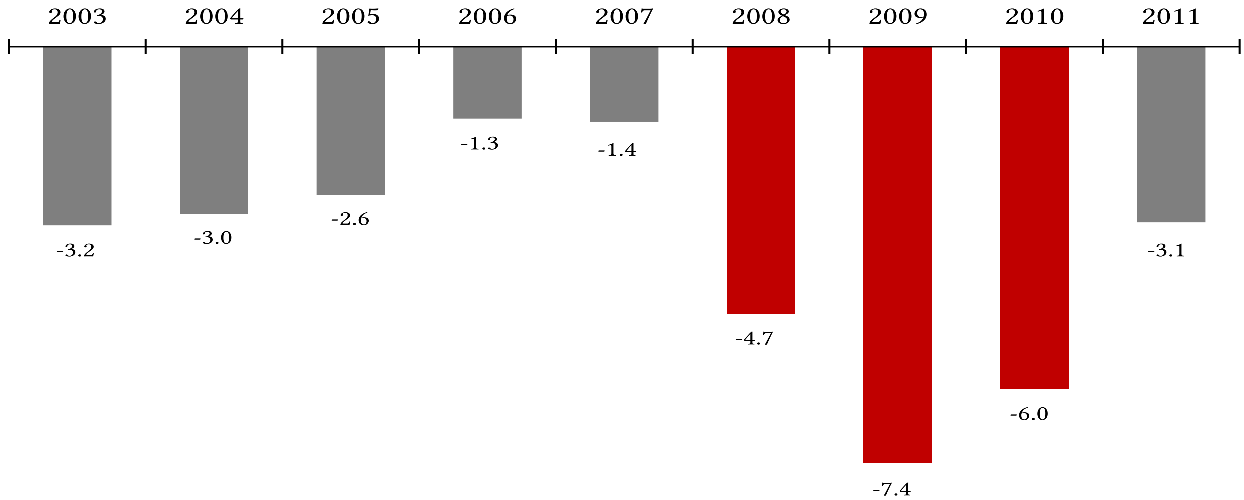


Notes: The *C & I loan spread* is the difference between the weighted-average effective annual loan rate on all commercial and industrial loans and the effective federal funds rate. Weights are by loan amount. The *delinquency rate* is the fraction of total C & I loans that are delinquent. Delinquent loans are those past due 30 days or more and still accruing interest, as well as those in nonaccrual status. The *nonperforming rate* is the fraction of total C & I loans that are nonperforming. Nonperforming loans are those that bank managers classify as 90 days or more past due or nonaccrual. The *charge-off rate* is the value of C & I loans removed from the books and charged against loss reserves divided by the total value of C & I loans. Charge-off rates are annualized, net of recoveries. Data are quarterly.

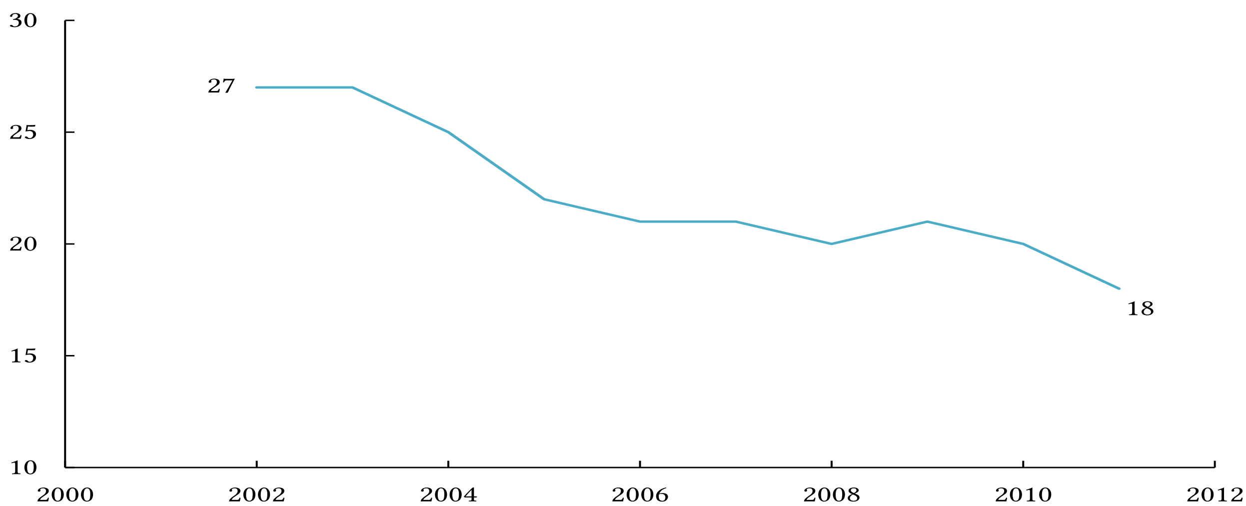
Sources: Board of Governors of the Federal Reserve System (C & I loan spread, delinquency rate, charge-off rate). Federal Financial Institutions Examinations Council (nonperforming rate). Data retrieved from FRED, Federal Reserve Bank of St. Louis.

Figure 2: Concentration in the C&I Loan Market since the Crisis

(a) Annual Percent Change in the Number of Banks with at Least 10% C&I Lending



(b) Number of Banks Constituting 50% of C&I Loan Market



Notes: Panel (a) depicts the annual percent change in the number of banks with at least 10% of their loan portfolio consisting of C&I lending. Panel (b) depicts the minimum number of banks required each year to amass 50% of the total market for C&I loans.

Source: Berger et al. (2017)

if aggregate bank equity capital is sufficiently depleted. A severe drop in net worth raises the cost of equity, forces bank consolidation, and transitions the loan market to the kink in equilibrium, where each surviving bank maintains a local monopoly over a distinct segment of the loan market. As long as equity capital positions stay impaired, no bank finds it optimal to deviate its price of credit away from the kink and trigger competition, because doing so would further damage profits.

Instead, banks tacitly collude to keep their lending rates fixed. Efforts by the central bank to get banks to pass through a low policy rate fail.

More accommodative monetary policy can even *raise* lending rates to firms because it increases bank market power. At the kink, a drop in the interest rate lowers bank funding costs, which increases rents. Higher potential rents encourage other banks to enter the lending market, which makes each one more specialized in a narrower class of industries or areas. The average entrepreneur can then contract with a more preferred bank closer to his or her location. Each bank exploits its position by charging a higher price of credit. An accommodative monetary policy thus inadvertently increases the local monopoly power of banks.

Because bank net worth in the economy is procyclical, the central bank's attempts to lower bank lending rates will be persistently thwarted in downturns. A financial accelerator effect emerges in which an initial shock to bank equity propagates through time because banks affect physical capital production. They do so in two ways. First, their loan rates influence aggregate investment. Second, their scale determines their specialization and hence their efficiency at recovering physical capital from failed projects.

When banks turn into local monopolists following an aggregate shock that pushes them to the kink, their higher price of credit curtails aggregate investment. In addition, after banks have consolidated from the drop in their net worth, their larger scale and weaker specialization hurt their ability to retrieve capital. Both effects deplete the physical capital stock and lower output, which reduce bank net worth the following period. The banking sector persists at the kink where pass-through is obstructed, and the initial shock to bank equity permeates the real economy through time.

In summary, this paper provides a framework for studying how aggregate bank equity and the industrial organization of the banking sector affect the real economy and monetary policy. A novelty of the analysis is that bank net worth becomes an indicator for the degree of competition in the bank credit market and the effectiveness of the interest rate channel. Poor health of bank balance sheets leads the economy to persist in an equilibrium where efforts by a central bank to lower the cost of firm credit are repeatedly frustrated, while output and investment remain low.

Literature

This paper combines the insights of several strands of literature to uniquely tie aggregate bank net worth to bank competition, monetary policy, and the real economy.

The first strand is the broad body of work exploring the effects of financial frictions on the macroeconomy. The papers most closely related to mine are Bernanke and Gertler (1989), Holmstrom and Tirole (1997), and Gertler and Kiyotaki (2010).

As in Bernanke and Gertler (1989), frictions in the financial market waste productive physical capital resources and affect aggregate output. A key distinction here is banks. Their efficiency at recovering value from defaulted projects influences the size of the physical capital stock.

Net worth in the banking sector plays a major role in the model, as in Gertler and Kiyotaki (2010). Here, the novelty is that net worth influences the degree of competition among banks and the effectiveness of monetary policy.

Holmstrom and Tirole (1997) analyze the effect of changes to the supply of intermediated financial capital on investment and credit spreads. A difference here is that the market for intermediated financial capital (bank loans) is imperfectly competitive.

A second strand is on bank diversification and specialization. Diamond (1984) stresses the benefits of bank diversification in lowering the uncertainty in a bank loan portfolio and shrinking costs of delegated monitoring. I adopt that insight. Greater diversified banks in the model can operate with more leverage. Liang and Rhoades (1991), McAllister and McManus (1993) and Demsetz and Strahan (1997) give empirical evidence of a positive correlation between bank diversification and leverage.

Winton (1999) explores some hazards of bank diversification. Expansion into new industries may reduce the effectiveness of loan monitoring and increase the chance of bank failure. Acharya et al. (2006) find that a less diversified bank loan portfolio is associated with higher return on assets and a lower fraction of nonperforming loans among Italian banks. Paravisini et al. (2015) find evidence of bank specialization in lending to exporting firms in Peru. Berger et al. (2017) document industry concentration in commercial and industrial loans among U.S. banks.

Banks in the model specialize in different segments of the credit market rather than fully diversify because of the increasing costs of liquidating failed projects. The convex cost of diversification I use is similar to that in Gârleanu et al. (2015), who work in a setting of investor portfolios and asset pricing, rather than one with banks and competition.

A third strand is the industrial organization of banking. Berger et al. (2004) provides a survey. Matutes and Vives (1996) present a model in which banks rival each other in deposits. Whether a bank is a local monopolist or a competitor depends on the perceptions of that bank's likelihood of failure. I fix depositor beliefs (banks cannot fail) to emphasize how bank net worth alters bank lending competition.

Matutes and Vives (2000) analyze how imperfect competition affects bank portfolio choice and whether deposit regulation intensifies or weakens risk-taking. Loans in my model all carry the same risk, so I can focus on the choice of pass-through rather than the choice of bank portfolio.

This paper owes a large debt to Salop (1979), whose structure of monopolistic competition on a circle I adopt. Other papers have also used the Salop framework to explore a variety of issues in banking:

Besanko and Thakor (1992) present a spatial model in which banks differentiate in loans and deposits and study the welfare implications of relaxed barriers to entry. To focus on bank lending, I have banks competing only in the credit market.

Chiappori et al. (1995) study the effects of deposit regulation on bank lending rates. In their model, the interest rate channel can also be hampered, but only when deposit rates are capped and deposits are bundled with credit services; otherwise, full transmission occurs. The deposit rate in my model is unregulated to put attention on bank lending.

Sussman and Zeira (1995) look at financial development across U.S. states and present a macroeconomic model in which costs of intermediation increase with the distance between the borrower and the bank. As in my model, greater bank specialization enhances physical capital production. Their economy displays no persistence.

Hauswald and Marquez (2006) feature bank-screening technology whose signal quality declines with the borrower's distance from the bank. Their focus is on banks strategically screening borrowers to carve out different segments of the loan market and soften competition. In my model, all borrowers are identical prior to obtaining a loan.

The final strand is the empirical and theoretical work on interest rate pass-through and its relation to banks. The papers mentioned in the introduction document the positive connection between impaired pass-through to lending rates and bank concentration. Early work by Hannan and Berger (1991) and Neumark and Sharpe (1992) find a similar relation in deposit rates, as does Drechsler et al. (2017) in more recent work. Aristei and Gallo (2014) and Hristov et al. (2014) provide evidence that pass-through deteriorated in the Euro area during the financial crisis. Recently, Scharfstein and Sunderam (2016) find that higher mortgage lender concentration reduces the pass-through of declines in RMBS yields to mortgage rates.

Models of interest rate pass-through in the banking sector typically assign market power to banks, but they treat incomplete pass-through using either sticky prices (Hülsewig et al. (2009)) or menu costs (Hannan and Berger (1991); Cottarelli and Kourelis (1994); Scharler (2008); Hülsewig et al. (2009); Gerali et al. (2010); Günter (2011)). I micro-found the pass-through impairment from the kink in the demand curve for bank credit. The kink arises endogenously from the competitive market structure among banks.

2 Economy

The model setting is an overlapping generations economy advancing through discrete time on a circle. Lining the circle each period are industrial projects that produce physical capital. Banks finance these projects with single-period loan contracts, and retain market power over borrowers, but engage in monopolistic competition à la Salop (1979). Banks will specialize in different segments of the loan market and finance themselves with deposits and equity.

The baseline model I present in this section features a single-unit scale for investment and fully depreciating physical capital. Cyclical persistence will originate purely from the banking sector rather than from the productive sector, because aggregate investment will be fixed through time. In section 7, I allow the scale of investment to vary, thus letting aggregate shocks to propagate through time from investment fluctuations as well.

2.1 Production

Distinct production technologies exist for output and physical capital. Output produced in period t can be consumed in the period or invested in the production of capital that becomes available for use in $t+1$. Capital cannot be consumed but only used in production of output.

Output

At time t, a final consumption good, denoted Y_t, is produced by a perfectly competitive, representative firm. The final good is produced using capital and labor in a Cobb-Douglas production technology:

$$Y_t = \tilde{A}_t K_t^{\alpha} L_t^{1-\alpha},$$

with $\alpha \in (0,1)$. The random productivity shock $\tilde{A}_t$ is continuously distributed over a finite positive support $[\underline{A}, \overline{A}]$, has mean A, and is i.i.d. over time. It is the only source of aggregate uncertainty in the economy.

Physical capital

Firms use projects to produce physical capital. A continuum of projects are uniformly distributed around the circle each period. I normalize the circumference of the circle to one, so projects take up a unit measure. A project is identified by its unique location $j \in [0,1)$ on the circle. I interpret projects on different parts of the circle as belonging to different industries or geographic areas.

Projects (and firms) are owned and managed by entrepreneurs, who operate a single project (and firm). Entrepreneurs borrow from banks to finance projects, posting the project as collateral. I elaborate on entrepreneurs in the section on entrepreneurs below.

A project is indivisible, lasts a single period, and is risky. A project takes the output good as investment and can produce one of two possible quantities of physical capital: high and low. Let H denote the high state and L the low state. All projects require a single unit of output to initiate.

Denote the high return on the project $\overline{\kappa}$ and the low return $\underline{\kappa}$. The returns are arranged

$$0 < \underline{\kappa} < 1 < \overline{\kappa},$$

making the low return a strict loss on investment (a failure), and the high return a strict gain (a success).

Project uncertainty

The probability that a project produces the high return takes a special form. This form allows all projects to bear the same expected probability of success prior to financing, but different probabilities after initiation. At the beginning of period t, the probability that project j reaches the high state at the end of that period is random. This random probability takes the following form:

$$\widetilde{\Pr}_{t^+}\left(H|j,\tilde{u}_{t^+}\right)=\frac{1}{2}\left(1+\cos\left(2\pi\left(j+\tilde{u}_{t^+}\right)\right)\right), \tag{1}$$

where $\tilde{u}_{t^+}\sim U\left[0,1\right]$. The object in (1) is a random measure that maps a realization of the uniform shock $\tilde{u}_{t^+}$ to a probability distribution over the two states at a location j. I call (1) the *success probability* of a project.

The periodicity of the cosine function guarantees $\widetilde{\Pr}_{t^+}:[0,1]\mapsto[0,1]$. The shock $\tilde{u}_{t^+}$ is realized in the middle of the period, which I represent using the t^+ notation, making the success probability measurable as of time t^+.

Properties of a project's success probability are presented in Lemma (1).

Lemma 1. *The* success probability *in (1) satisfies the following properties:*

1. *(Distributional symmetry) The probability density function of a project's success probability is the same at all locations.*
2. *(Mean and variance) Each project is expected to succeed half the time, with variance $\frac{1}{8}$.*
3. *(Distance-dependent covariance) The covariance between projects j and k in their success probabilities is $\frac{1}{8}\cos\left(2\pi\left(j-k\right)\right)$.*

Proof. See Appendix (A.2). □

The form in (1) is a way to make a project's probability of reaching the high state invariant to its location. Prior to the realization of $\tilde{u}_{t^+}$, a project's outcome distribution cannot be distinguished from its neighbors', because all projects bear the same uncertainty of success. As a result, projects at every location share the same expected probability of generating the high return $\overline{\kappa}$–namely, $\frac{1}{2}$.

Another important feature of physical capital production is that the covariance of success probabilities between projects depends exclusively on the distance between those projects rather than on their locations. From the lemma, the correlation between the success probabilities of projects located at positions j and k on the circle is

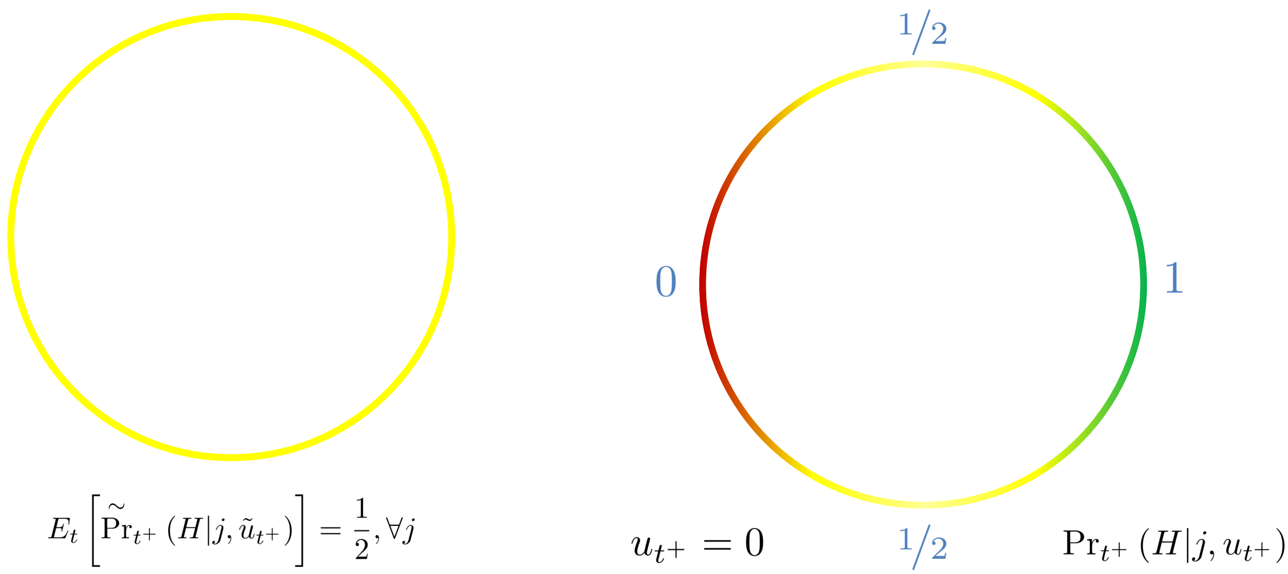
$$\text{corr}\left(\widetilde{\Pr}_{t^+}\left(H|j,\tilde{u}_{t^+}\right),\widetilde{\Pr}_{t^+}\left(H|k,\tilde{u}_{t^+}\right)\right)=\cos\left(2\pi\left(j-k\right)\right).$$

The above expression implies projects located near one another on the circle have more positively correlated probabilities of success than those located farther apart. Projects positioned opposite one another on the circle have the lowest correlated probability. This correlation structure is meant to capture the notion of integrated industries (e.g., metals and automobiles) or nearby geographic areas (e.g., neighboring cities) sharing more correlated production outcomes than more "distant" ones.

Figures 3(a) - 3(b) present an illustration of project uncertainty. At the start of the period, each project around the circle bears the same uncertainty of project success, having one-half chance of yielding the high return. Once the shock $\tilde{u}_{t+}$ is drawn in the middle of the period, projects bear different probabilities of success according to their locations, with those close to one another on the circle sharing similar likelihoods of yielding the high return.

Figure 3: Example of Success Probabilities across Projects

(a) Before realization of $\tilde{u}_{t+}$ (b) After realization of $\tilde{u}_{t+}$



Notes: At the beginning of each period, all projects share the same expected success probability of one-half. This common probability of success in expectation is represented in Figure 3(a) by the color yellow along the entire circle. In the middle of the period, the shock $\tilde{u}_{t+}$ is realized. The example in the figure has a realized value of $u_{t+}=0$. At that moment, projects differ in their success probabilities according to (1). In Figure 3(b), arcs of the circle with projects having high success probability are colored green. Arcs with projects having low success probability are colored red. The four numbers positioned around the circle are the success probabilities of the projects located at those positions.

A project's life follows this sequence: at the beginning of the period, the project is financed and the investment is made. In the middle of the period, $\tilde{u}_{t+}$ is realized, which determines the project's actual probability of the high return, denoted $\Pr_{t+}\left(H|j,u_{t+}\right)$. No action related to project financing or the project itself can be made at that time. Finally, at the end of the period, the project produces either the high or low amount of physical capital.

Physical capital formation

Total physical capital in the economy aggregates the high and low production from all projects. A successful project contributes its full quantity produced $\overline{\kappa}$ to the physical capital stock.

If a project produces a low return $\underline{\kappa}$, however, the project is in default, control rights over the project transfer to the bank, and the entrepreneur receives nothing. The productivity of the project after production is zero, so the bank's only recourse is to liquidate the project and recover as much as possible.[1]

I assume the bank is imperfect in liquidating a project, in that it retrieves only a fraction of the low physical capital amount. The difference between $\underline{\kappa}$ and the quantity the bank can retrieve from the distressed project is the *liquidation cost.* I go into more detail on liquidation costs in the section on banks below. For now, let $G(1, N_t)$ denote the aggregate liquidation costs of all banks in the economy, which is a function of the aggregate investment of 1 and the number of banks N_t.

Physical capital depreciates fully in the period. Because of how I defined the probability of project success in (1), and the presence of a continuum of projects, project outcomes in the aggregate feature no uncertainty. At the end of each period, exactly half the projects will succeed and half will fail.

Therefore, the aggregate quantity of physical capital produced and available for the next period will be known. For any period t, the next period physical capital stock K_{t+1} in equilibrium is given by

$$K_{t+1} = \underbrace{\frac{1}{2}\left(\overline{\kappa} + \underline{\kappa}\right)}_{\text{real side}} - \underbrace{G\left(1, N_t\right)}_{\text{banking side}}. \tag{2}$$

The equation for physical capital in (2) consists of two components. The first component originates purely from the real side of the economy, capturing the total quantity of physical capital produced by entrepreneurial investment. The second component captures the impact the banking sector has on the production of physical capital. That piece is a consequence of frictions in the banking sector in the liquidation of failed projects.

Banks will play a central role in the quantity of physical capital produced. The number of banks N_t will greatly influence the banking sector's efficiency in recovering physical capital from defaulted projects.

2.2 Agents

Three types of agents populate the economy: depositors, bankers, and entrepreneurs. Depositors and bankers live for two periods, whereas entrepreneurs live for one. Periods should be interpreted

[1]The entrepreneur alone observes the project return. Control rights transfer to the bank to compel the entrepreneur to report the outcome of the project truthfully. Such a setting justifies the use of a debt contract and is similar to the costly state verification assumptions of Townsend (1979).

as long enough to allow for the entry and exit (consolidation) of banks in the commercial loan market in response to aggregate shocks. Each agent knows his or her type at the beginning of life. Every generation produces a continuum of depositors, bankers, and entrepreneurs, with each continuum having unit measure.

Depositors

Bank deposits are the only financial asset depositors have access to. And in equilibrium, depositors will be the exclusive investors in bank deposits. Their location on the circle is immaterial, so I put them at the center. Each depositor is endowed with L^d units of labor while young and old. They have no disutility of labor, so supply it inelastically both periods in a competitive labor market. Individual preferences are identical over lifetime consumption:

$$U^d = \log c_t + \beta E_t \left(\log c_{t+1}\right).$$

The budget constraints in the two periods of life are

$$\begin{aligned} s_t + c_t &\leq L^d w_t + M_t, \\ c_{t+1} &\leq L^d w_{t+1} + R_{f,t} s_t \left(1 - \tau\right). \end{aligned}$$

Here, w_t is the real wage, $R_{f,t}$ is the gross real interest rate on deposits, τ is the marginal tax rate on deposit savings, and M_t is a lump-sum tax rebate or levy. Bank deposits will be free of default risk, so $R_{f,t}$ is the risk-free interest rate in the economy. The lump-sum amount $M_t = \tau s_t$, so that tax policy is government-budget neutral.

This economy is real and there is no money. I abstract from a nominal economy in order to put focus singularly on (monopolistic) competition in the banking sector as an impediment to the interest rate channel. The tax rate τ is set to directly alter the real interest rate $R_{f,t}$, and it is the single monetary policy tool in the economy. Hence monetary policy is inextricably linked to fiscal policy. Monetary policy amounts to a redistribution of wealth between young and old depositors that adjusts the supply of aggregate deposits. To simplify things, I restrict the tax rate to be constant, making τ a parameter of the model.

Bankers

Bankers are risk-neutral and have utility only over expected consumption when old. They have access to both bank deposits and bank equity for their financial investments. In equilibrium, they will exclusively invest in bank equity, because the return on equity will be at least as high as the return on deposits. While young, they are endowed with L^b units of labor and the exclusive ability to evaluate industrial projects. They supply labor inelastically and thus save the amount $w_t L^b$.

Like depositors, bankers are located at the center of the circle.

The ability to evaluate projects involves expertise in writing financial contracts with entrepreneurs, collecting payments, and liquidating projects in the event of non-payment. No single banker has enough personal wealth to finance a project on his own. So groups of bankers pool their savings and incorporate N_t institutions called banks at different locations on the circle[2]. Let $\mathcal{N}_t$ denote the set of bank locations on the circle at time t, so that $|\mathcal{N}_t| = N_t$ and $i \in [0, 1)$ for all $i \in \mathcal{N}_t$. I exclude explicit bank collusion in setting loan prices.

Banks finance projects with loans. Bankers are the equity investors in banks and can perfectly observe the loan portfolio of their banks. I denote the expected return on bank equity capital at time t by $\hat{R}_{E,t+1}$. A banker consumes the realized return on his investment in the second period.

Aggregate bank equity in the economy will turn out to be the savings of bankers $w_t L^b$, and will play a critical role. When aggregate equity is low, the cost of equity will be high, and vice versa. Bank equity is the source of a financial-accelerator effect similar to Bernanke and Gertler (1989), but here the effect is coming from the banking sector. Bank equity capital positions become a channel through which an i.i.d. aggregate productivity shock propagates through time despite fully depreciating physical capital. The tight link between banker savings and bank net worth is one way to capture the notion that bank cost of equity is low in good times and high in bad times.

Entrepreneurs

Entrepreneurs are endowed with both L^e units of labor and a potential project. They are risk neutral with no disutility of labor. An entrepreneur decides either to work for the representative output firm and consume income $w_t L^e$, or start a firm, pursue the project, and consume the net return. Projects are non-tradeable and non-transferable. Entrepreneurs lack any wealth, so their projects must rely entirely on bank financing.

Entrepreneurs are located at different positions on the circle and their individual locations are important. Entrepreneurs are identified by the locations of their projects $j \in [0, 1)$. Consider an entrepreneur positioned at location j who thinks about a loan from bank $i \in \mathcal{N}_t$. Let $R^i_{L,t}$ be the gross lending rate bank i charges on the loan. The expected utility of entrepreneur j who undertakes the project and borrows from bank i at time t is

$$U^e_{j,i} = \frac{1}{2}\left[\hat{q}_{t+1}\overline{\kappa} - R^i_{L,t}\right] - c|i - j|. \tag{3}$$

Prior to financing, the entrepreneur expects the project to succeed half the time, and expects to default and receive nothing in the other half. The term $\hat{q}_{t+1}$ is the expected relative price of physical capital as of time t. If the project is successful, the entrepreneur will receive the expected

[2]Agents who collaborate to form financial intermediaries is also featured in Ramakrishnan and Thakor (1984).

high return $\hat{q}_{t+1}\overline{\kappa}$ in the physical capital market. He must also repay the bank loan with interest: $R^i_{L,t}$.[3]

Entrepreneurs have preference to borrow from a bank "nearby." This preference for proximity is represented by the term $-c|i-j|$, where c captures the strength of the preference and $|i-j|$ is the shortest arc length between i and j.

There are many interpretations of the proximity preference. An entrepreneur could prefer a closer bank because the bankers there speak the same language, or demand less paperwork, or because the bank has a reputation for specifically lending to the industry or area the entrepreneur's project is in.

In fact, banks end up specializing in different segments of the industrial loan market. The preference for minimizing the distance will incline an entrepreneur to borrow from a bank with expertise in that entrepreneur's industry or geographic area. Paravisini et al. (2015) provide empirical evidence that firms with exports have a greater likelihood of borrowing from a bank that specializes in that firm's exporting country.

2.3 Banks

After incorporation, banks extend credit in the form of one-period bank loans to entrepreneurs. After the project returns, each bank collects the promised repayments from entrepreneurs with successful projects. They also liquidate failed projects, which involves selling the distressed assets in the physical capital market. After all payments are collected and sales are made, each bank repays its depositors and issues a liquidating dividend to its equity holders. Bank equity is therefore all external equity with no retained earnings. In the next period, a fresh crop of banks are incorporated under the new generation of bankers.

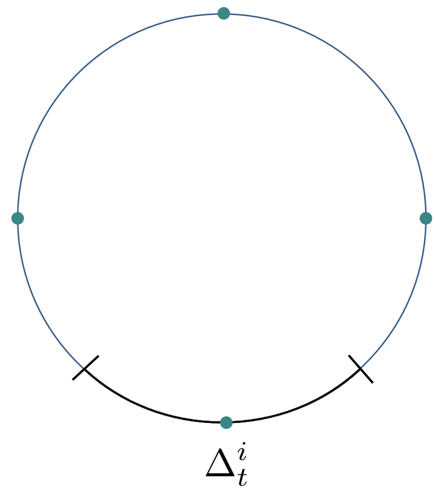
Banks confront two key frictions: deadweight costs from liquidating assets and a constraint on leverage. The first friction encourages them to be small, specialized, and "close" to their average borrower, whereas the second friction encourages them to be big, broad, and "far" from their average borrower. Banks trade off these two frictions when choosing their loan portfolios.

Loan portfolio

I restrict a bank to only finance projects that are positioned along arcs centered at the bank's home location (headquarters). Let $\Delta^i_t \in [0,1]$ denote the arc length of the projects financed by a bank headquartered at position i on the circle. With each project requiring a single unit of financing, the size of the total loan portfolio of bank i is then Δ^i_t. This length can also be considered the "size" of the bank. A visual depiction of a bank $i's$ portfolio is given in Figure 4.

[3]Physical capital is sold to the representative firm at the end of the period. However, the productivity of that capital is not known until the start of the following period once the shock $\tilde{A}_{t+1}$ realizes. Therefore, physical capital transactions are based on the expected relative price.

Figure 4: Bank Loan Portfolio Representation



Notes: Projects are uniformly distributed around the circle. Bank i is headquartered at the bottom dot. The bank's loan portfolio size Δ_t^i is the length of the arc centered at bank $i's$ headquarters. The three remaining dots represent other banks in the loan market.

Banks issue take-it-or-leave-it loan offers to entrepreneurs. Banks know entrepreneurs have a preference for proximity, and they are aware of the preference structure $(-c|i-j|)$. However, the location of an entrepreneur is unobservable to any bank at the time the loan is contracted. Also at that time, all projects share identical potential returns per unit of investment and bear the same risk. For these reasons, a bank does not engage in any degree of price discrimination, but instead posts a single lending rate $R_{L,t}^i$, while taking into account the rates of all other banks on the circle. Entrepreneurs will self-select into banks according to their preferences and the posted lending rates. In Appendix A.8, I show that the results of the model go through even if the bank can price discriminate.

Diversification

In financing multiple projects, a bank can reduce the uncertainty of its loan portfolio through diversification. For a given Δ_t^i, the bank's average probability per project of receiving payment $R_{L,t}^i$, prior to the realization of $\tilde{u}_{t^+}$, is given by

$$\begin{aligned} \widetilde{\Pr}_{t^+}\left(R_{L,t}^i|\Delta_t^i,\tilde{u}_{t^+}\right) &= \frac{1}{\Delta_t^i}\int_{-\left(\Delta_t^i\right)/2}^{\left(\Delta_t^i\right)/2}\frac{1}{2}\left(1+\cos\left(2\pi\left(i+j+\tilde{u}_{t^+}\right)\right)\right)dj \\ &= \frac{1}{2}+\underbrace{\frac{\sin\left(\pi\Delta_t^i\right)}{\Delta_t^i}}_{\text{diversification}}\underbrace{\frac{\cos\left(2\pi\left(i+\tilde{u}_{t^+}\right)\right)}{2\pi}}_{\substack{\text{residual}\\ \text{uncertainty}}}. \end{aligned} \tag{4}$$

I call (4) the *repayment rate* of bank $i's$ loan portfolio, as it is the fraction of projects whose owners can repay the bank.

Two components comprise the repayment rate of a portfolio: diversification and residual uncertainty. The diversification component captures the reduction in the uncertainty of a bank loan portfolio's payoff from choosing a larger arc length around the circle. The residual uncertainty component reflects the risk that remains in a loan portfolio that is imperfectly diversified.

Important properties of the repayment rate are presented in Lemma (2).

Lemma 2. *The* repayment rate *of bank $i's$ portfolio satisfies the following properties:*

1. *(Common mean) The expected repayment rate is always $\frac{1}{2}$, no matter the choice of Δ_t^i.*
2. *(No diversification) As $\Delta_t^i \downarrow 0$, the bank's repayment rate approaches the same probability that a single project succeeds, given in (1).*
3. *(Declining variance) As Δ_t^i increases, the variance of the repayment rate declines.*
4. *(Perfect diversification) When $\Delta_t^i \uparrow 1$, the repayment rate approaches $\frac{1}{2}$, no matter the realization of $\tilde{u}_{t^+}$.*

Proof. See Appendix (A.3). □

The repayment rate of an imperfectly diversified bank is a random variable prior to the realization of $\tilde{u}_{t^+}$. Though, no matter the bank's arc length, the *expected* repayment rate on its portfolio is always $\frac{1}{2}$: the bank expects half its loan portfolio to repay and half to default.

As a bank lends to more and more entrepreneurs around the circle, it reduces the variability of its repayment rate by diversifying its loan portfolio. This reduction in a bank's loan portfolio risk is similar in spirit to the benefits of bank diversification identified in Diamond (1984). Eventually, if a bank lends the circumference of the circle, its portfolio becomes risk-free, being immune to the random realization of $\tilde{u}_{t^+}$. In this case, half the portfolio *will* succeed and half *will* fail; the payoff of the portfolio is no longer uncertain.

Debt constraint

Diversification affects a bank's financial capital structure. Bank capital structure consists of deposits from depositors and equity capital from bankers. The bank chooses the composition of debt and equity to finance its operations. The first friction a bank encounters affects its capital structure.

Both bankers and depositors know their bank's loan portfolio Δ_t^i. Whereas bankers also perfectly observe the uniform shock $\tilde{u}_{t^+}$, and thus the realized profits of their bank's loan portfolio, depositors do not. However, depositors do know the loan repayment rate function in (4), and hence are certain of their bank's *minimum* possible repayment rate. The minimum profit on the loan portfolio is the maximum amount depositors can prove and recover from the bank in bankruptcy court. Depositors are only willing to finance their bank up to this amount.[4] The minimum of (4) over the shock $\tilde{u}_{t+}$ is

$$\Pr_{\min}\left(R_{L,t}^{i}|\Delta_{t}^{i}\right)=\frac{1}{2}\left(1-\frac{\sin\left(\pi\Delta_{t}^{i}\right)}{\pi\Delta_{t}^{i}}\right).$$

Denote the minimum loan profits for bank i at time t as $\Pi_{\min}^{i,t}$. Depositors are willing to lend an amount up to the discounted face value of $\Pi_{\min}^{i,t}$. Because deposits are safe, the deposit rate is the risk-free interest rate $R_{f,t}$. Let D_t^i be the amount of deposits a bank chooses. The constraint on D_t^i is

$$D_{t}^{i}\leq\frac{\Pi_{\min}^{i,t}}{R_{f,t}}. \tag{5}$$

I call the maximum amount a bank can raise in deposits the *debt capacity* of the bank. Whatever additional outside financial capital a bank requires to finance its operations, it obtains from the equity market at the required expected equity return $\hat{R}_{E,t+1}$. In equilibrium, equity will be at least as expensive as debt $\left(\hat{R}_{E,t+1}\geq R_{f,t}\right)$.

The minimum repayment rate of a bank determines its debt capacity, which influences its leverage. As a bank increases Δ_t^i, it expands its lending operations to more and more industries or areas across the circle, and diversifies its portfolio. Depositors, in turn, are then willing to lend more to the bank. The minimum repayment rate of the bank is increasing in Δ_t^i, and so too will its debt capacity and leverage. [5]

Liquidation costs

A bank has reason to extend its financing around the circle in order to diversify, loosen its debt constraint, and obtain cheaper deposit financing. Because project success probabilities near the bank are more highly correlated with each other than with those farther away, the bank achieves a greater diversification benefit per loan the larger its loan portfolio.[6] The debt constraint encourages banks to be diversified and big.

A second friction encourages them to be specialized and small. The friction limits how much a bank can recover from loans in default. The entrepreneur cannot make payment after a project

[4]Bank i suffers its minimum repayment rate if the shock lands at location $|\frac{1}{2}-i|$ if $i\in\left[0,\frac{1}{2}\right]$ or location $1-|\frac{1}{2}-i|$ if $i\in\left(\frac{1}{2},1\right)$. If the shock lands at some other location for all $i\in\mathcal{N}_t$, every bank's realized repayment rate exceeds the minimum.

[5]The debt constraint in (5) can equivalently be considered a minimum equity capital requirement. Denote the total assets of the bank by A_t^i. Substituting the balance sheet identity $A_t^i\equiv D_t^i+E_t^i$, the constraint can be written as $E_t^i\geq A_t^i-\frac{\Pi_{\min}^{i,t}}{R_{f,t}}$. So rather than choosing an amount D_t^i in deposits, the bank instead can choose an amount E_t^i in equity, provided its choice satisfies a minimum amount. A greater diversified bank loosens this equity capital requirement, allowing the bank to issue less equity and more debt.

[6]Formally, for small Δ_t^i, $-\frac{d^2\left(\sin\left(\pi\Delta_t^i\right)/\Delta_t^i\right)}{d\left(\Delta_t^i\right)^2}>0$. The sign is negative because the diversification component in (4) is decreasing in Δ_t^i, as a larger loan breadth reduces the exposure of the bank's repayment rate to the residual uncertainty.

failure. With the loan in default, the bank transfers control rights of the project to itself, seizes the collateralized assets, and liquidates those assets in the physical capital market. However, the bank recovers less than the full amount of low physical capital return $\underline{\kappa}$, which means liquidation is costly.

I assume the total liquidation costs of a bank, denoted $g\left(\Delta_t^i\right)$, are increasing in a bank's loan portfolio Δ_t^i at an increasing rate. I take the function g to be

$$g\left(\Delta_t^i\right) = \gamma\left(\Delta_t^i\right)^2, \tag{6}$$

with $\gamma > 0$. The bank loses the amount g in physical capital from every loan it liquidates. Because liquidation costs are convex in Δ_t^i, the bank suffers a diseconomy of scale, making diversification costly.[7]

A bank's loan portfolio Δ_t^i signifies the class of industries or geographic areas to which it lends. Successfully extending credit to a specific set of industries or locations requires expertise in those markets. I interpret the arc length Δ_t^i to be an indicator of a bank's expertise or specialization. It also represents the size of the bank. A greater arc length means a greater "distance" between the bankers who oversee individual loans and bank headquarters.

Part of a bank's expertise is intimate knowledge of the "soft" information about the industries or locations it lends to that is not easily communicated to others. This information includes the organizational structures of the firms, common production processes, the local customer markets, and importantly, the second best use of the physical capital assets. Convex liquidation costs is a reduced-form representation of the economic reasoning in Stein (2002): bank lending that relies on soft information weakens the research incentives of the line managers in a large, hierarchical bank. These weakened incentives make bankers less capable of liquidating distressed assets at their full value.

Bank decision

A typical bank i chooses a lending rate $R_{L,t}^i$ and quantity of deposits D_t^i to maximize expected profits over a single period, perfectly knowing and taking as given (1) the demand curve for bank credit (described below), (2) the lending rates of other banks, (3) the number of banks N_t on the circle, and (4) the factor prices w_t and $\hat{q}_{t+1}$, and the costs of debt and equity capital, $R_{f,t}$ and $\hat{R}_{E,t+1}$, respectively. Let FC_t^i denote the financing cost function for bank i.

[7] Having the liquidation costs be a function of the distance between the failed project and the bank headquarters greatly complicates the aggregation of physical capital. Instead, I make the liquidation costs a function of bank size, which is related to distance, because a larger bank will be farther from its average borrower.

Expected profits of the typical bank at time t are

$$\Pi_t^i = \frac{1}{2} R_{L,t}^i \Delta_t^i + \frac{1}{2} \hat{q}_{t+1} \left(\underline{\kappa} \Delta_t^i - g\left(\Delta_t^i\right)\right) - FC_t^i. \tag{7}$$

Prior to the realization of $\tilde{u}_{t^+}$, the bank expects half the projects in its loan portfolio to repay and half to default. The first term in (7) represents expected payments received from the fraction of projects that succeed. The second term is the expected proceeds from the fraction that fail. In this case, the bank recovers the low returns on physical capital net of the liquidation costs.

The financing cost function FC_t^i consists of the payments to depositors and equity holders. The function is

$$FC_t^i = R_{f,t} D_t^i + \hat{R}_{E,t+1} \left(\Delta_t^i + f - D_t^i \right), \tag{8}$$

where f is a fixed cost of entry into the commercial loan market (i.e., costs of chartering, complying with regulations, building the organizational form, etc.). Market power without a fixed cost would encourage an unlimited number of banks to enter the lending market. The bank requires an amount $\Delta_t^i + f$ in financing. The minimum loan profits that determines the debt constraint of (5) are

$$\begin{aligned} \Pi_{\text{min}}^{i,t} &= \Pr\nolimits_{\text{min}} \left(R_{L,t}^i | \Delta_t^i \right) R_{L,t}^i \Delta_t^i \\ &\quad + \left(1 - \Pr\nolimits_{\text{min}} \left(R_{L,t}^i | \Delta_t^i \right)\right) \hat{q}_{t+1} \left(\underline{\kappa} \Delta_t^i - g\left(\Delta_t^i\right)\right). \end{aligned}$$

The bank maximizes (7) subject to (5).

3 Demand Curve for Bank Credit

The purpose of this section is to construct the demand curve for bank credit from the perspective of a typical bank. In doing so, I describe the industrial organization of the banking sector.

Entrepreneurs choose which bank to finance a project in order to maximize utility presented in (3). For an entrepreneur to borrow from a bank and undertake the project at all, the expected return from the project must exceed the outside option: the income from working, $w_t L^e$. Because of a preference to be "close," as measured by the distance cost c, an entrepreneur will always choose between three alternatives: the outside option, a loan from the bank to his or her "left" on the circle, and a loan from the bank to his or her "right."

Because I will focus on symmetric equilibria, I derive the demand curve for a typical bank i assuming (1) N_t banks operate on the circle, located a distance $\frac{1}{N_t}$ from each other, and (2) all other banks on the circle charge the same fixed lending rate $R_{L,t}$, whereas bank i decides on its lending rate $R_{L,t}^i$. The demand curve for credit from the typical bank will consist of a monopoly, competitive, and kinked component.

3.1 Monopoly

The monopoly portion of bank $i's$ demand curve consists of the set of lending rates the bank can charge and face no competition from its neighboring banks.

To begin, if bank i sets $R^i_{L,t} > \hat{q}_{t+1}\overline{\kappa} - 2w_t L^e$, no entrepreneur on the circle would find it worthwhile to borrow from the bank. The price of the bank's loan would be so high that even the entrepreneur located at the bank's headquarters would rather work or borrow from a neighboring bank.

As bank i lowers $R^i_{L,t}$, however, it will start attracting entrepreneurs whose surplus from the project exceeds the outside option value. Denote by x the distance from the bank's headquarters such that the entrepreneur located at that distance has a surplus from the project equaling the value from working. The entrepreneur's surplus consists of the expected net return on the project $\frac{1}{2}\left(\hat{q}_{t+1}\overline{\kappa} - R^i_{L,t}\right)$ less the distance cost cx. Formally, x satisfies

$$\frac{1}{2}\left(\hat{q}_{t+1}\overline{\kappa} - R^i_{L,t}\right) - cx = w_t L^e.$$

The entrepreneur at distance x is indifferent between managing the project and working. Solving for x yields

$$x = \frac{\frac{1}{2}\left(\hat{q}_{t+1}\overline{\kappa} - R^i_{L,t}\right) - w_t L^e}{c}.$$

A typical bank will fund projects on either side of it, so the monopoly demand function for bank i, denoted $\Delta^{i,M}_t$, is

$$\Delta^{i,M}_t = \frac{\frac{1}{2}\left(\hat{q}_{t+1}\overline{\kappa} - R^i_{L,t}\right) - w_t L^e}{c/2}. \tag{9}$$

This quantity defines the potential local monopoly market of the typical bank. The monopoly demand function is increasing in the high physical capital return $\hat{q}_{t+1}\overline{\kappa}$. It is declining in the lending rate $R^i_{L,t}$, the outside option value $w_t L^e$, and distance cost c. Although a bank faces no competition from other banks in its local monopoly market, the bank implicitly competes with the outside option of entrepreneurs.

3.2 Competitive

The competitive part of bank $i's$ demand curve consists of the set of lending rates that would expand bank $i's$ loan portfolio into the lending market of a neighboring bank, igniting competition between the two.

If an entrepreneur is choosing between two banks, it must mean the expected return on the project exceeds the outside option value. The entrepreneur will borrow from the bank offering the lower financing and distance cost. Because the neighbor is located $\frac{1}{N_t}$ distance away on the circle, a typical bank will capture all projects within a distance x satisfying

$$\frac{1}{2}R_{L,t}^i + cx \leq \frac{1}{2}R_{L,t} + c\left(\frac{1}{N_t} - x\right).$$

The entrepreneur who is indifferent between the two competing banks is at a distance x' that satisfies the above relation with equality. Solving for x' gives

$$x' = \frac{1/2\left(R_{L,t} - R_{L,t}^i\right) + c/N_t}{2c}.$$

Because the typical bank competes against the two neighbors on either side, its competitive demand function $\Delta_t^{i,C} = 2x'$, making

$$\Delta_t^{i,C} = \frac{1/2\left(R_{L,t} - R_{L,t}^i\right) + c/N_t}{c}. \tag{10}$$

Bank $i's$ competitive credit market shrinks the more its lending rate exceeds the rates of the neighbors. Additionally, the more banks on the circle, the closer every entrepreneur is to a potential bank, which narrows the competitive market of any one bank.[8]

The size of bank $i's$ competitive lending market is determined by the marginal entrepreneur who is just indifferent between borrowing from bank i and borrowing from the bank's neighbors. Entrepreneurs located closer to bank i will strictly prefer borrowing from it. Conversely, entrepreneurs located outside bank $i's$ competitive market strictly prefer borrowing from the competing neighbor. For these reasons, each entrepreneur will prefer funding the project using a single bank. Petersen and Rajan (1994) document that small U.S. firms tend to concentrate their bank borrowing from one source. I assume the marginal entrepreneur flips a fair coin, and picks the bank according to the result.

3.3 Kinked

When bank i reduces its lending rate to exactly match the neighboring rate $R_{L,t}$, its local monopoly market will just touch the monopoly markets of its two neighbors, and a kinked market arises. I denote this kinked lending rate $R_{L,t}^{i,K}$.

The kinked market gets its name from the kink in the demand curve at the lending rate $R_{L,t}^{i,K}$. If bank i set its lending rate just above $R_{L,t}^{i,K}$, its local monopoly market would be segregated from

[8] When competing with neighboring banks, bank i could reduce $R_{L,t}^i$ enough to capture even those entrepreneurs *residing* at the neighbors' locations. Such a pricing strategy would drive the neighboring banks out of the market, and create a jump discontinuity in the demand curve for bank $i's$ credit. This *predatory* pricing would require posting a lending rate below marginal cost, which would necessarily lose money, so can be ruled out in equilibrium.

that of its neighbor. The bank would lend according to the monopoly demand function in (9). The slope of the corresponding monopoly demand curve is $\frac{dR^{i,M}_{L,t}}{d\Delta^i_t} = -c$. Alternatively, if the bank set a lending rate just below $R^{i,K}_{L,t}$, its local monopoly market would cross the markets of the two neighboring banks, which sets off competition. Bank i would collect demand according to the competitive demand function in (10). The slope of the corresponding competitive demand curve is $\frac{dR^{i,C}_{L,t}}{d\Delta^i_t} = -2c$. The difference in the slopes of the monopoly and competitive portions generates the kink in the demand for bank loans.

The slope of the competitive portion is twice that of the monopoly portion for the following reason. When the typical bank is a local monopolist that seeks to expand its market Δ^i_t by an increment, it must offer a price concession in the amount c in order to entice the marginal entrepreneur to borrow from a more distant, less specialized bank.

But when the typical bank tries to expand in a competitive market, it must offer the same price concession as before plus an additional amount c because the marginal entrepreneur is now closer to a neighboring bank that is more specialized in the entrepreneur's industry. The extra concession is meant to lure the entrepreneur away from the competition.

The kink in the demand curve for bank credit is a key feature of the lending market and critical for the main results. In the theory of kinked demand curves,[9] prices under oligopoly may "stick" around a focal price. That price is sustainable in equilibrium out of each firm's belief that undercutting will trigger a price war, but charging more leads no other firm to follow. The demand curve an individual firm faces will have a kink at the focal price.

The same economic reasoning applies here. If a bank reduced its lending rate below $R^{i,K}_{L,t}$, it would expand its segment of the loan market into the territories of the neighboring banks, sparking competition and hurting profits. Alternatively, raising the lending rate simply reduces the breadth of that bank's local monopoly market, which neighboring banks can safely ignore. The novelty here is that aggregate net worth in the banking sector will determine whether banks settle at that focal price.

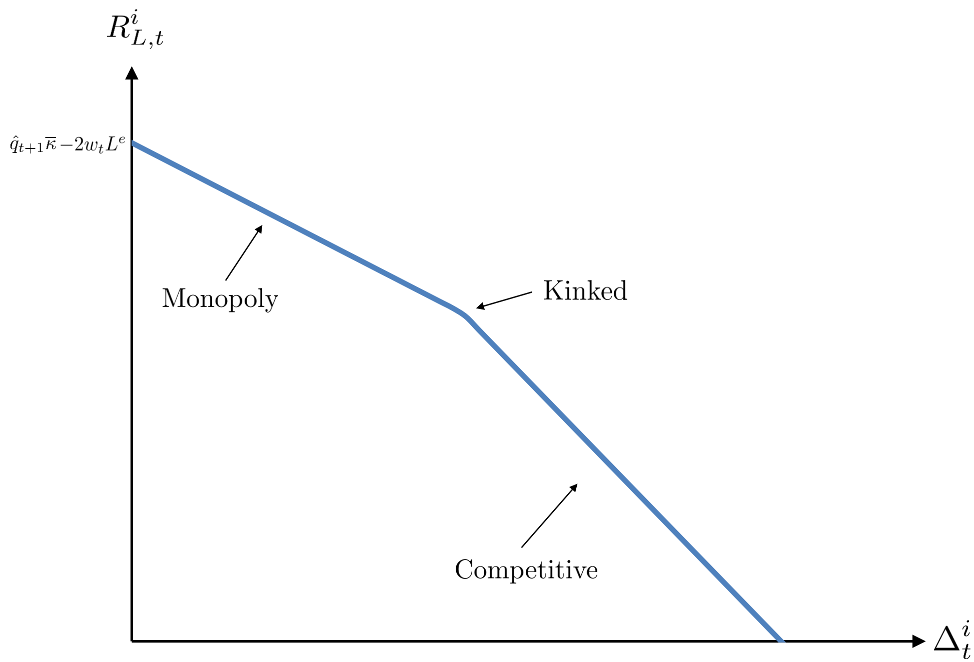
Figure (5) illustrates bank $i's$ demand curve for loans, stitching together the monopoly, kinked, and competitive lending markets.

4 Equilibrium

I study dynamic, symmetric, pure-strategy, zero-profit, Nash equilibria. The dynamic equilibrium is defined as a sequence of static Nash equilibria that will be linked together through aggregate savings. A static equilibrium is characterized by the tuple $\mathcal{E}_t \equiv \left\{R_{L,t}, N_t, \hat{q}_{t+1}, w_t, R_{f,t}, \hat{R}_{E,t+1}\right\}$,

[9] Hall and Hitch (1939) and Sweezy (1939) introduced the concept. Maskin and Tirole (1988), Bhaskar (1988), Rothschild (1992), and Sen (2004) give rigorous foundations. The literature on kinked demand curves is large. Reid (1981) provides a survey.

Figure 5: Demand Curve for Bank Credit from a Typical Bank i.



where $R_{L,t}$ is the single lending rate charged by all banks, N_t is the positive-integer number of equally spaced banks on the circle, $\hat{q}_{t+1}$ is the expected relative price of physical capital, w_t is the real wage, $R_{f,t}$ is the risk-free interest rate on deposits, and $\hat{R}_{E,t+1}$ is the required expected return on equity. The tuple is determined so that (1) every bank's choice of lending rate $R_{L,t}$ is profit-maximizing, (2) this choice of lending rate earns zero expected profits[10], (3) the circle contains no gaps $\left(\Delta^i_t = \frac{1}{N_t},\ \forall i\right)$, and (4) markets clear.

The market for bank credit will be characterized by monopolistic competition, as in Chamberlin (1933), Robinson (1969), and Salop (1979). Entrepreneur preference for proximity will be a source of differentiation among banks that gives them market power in loan pricing, even when competing with one another to fund projects. Banks perfectly compete for deposits and bank equity.

4.1 Competitive and kinked equilibria

Three types of equilibria are possible in the economy: monopoly, kinked, and competitive. These types correspond to the three parts of the demand curve for bank credit.

A convenient way to visualize the equilibrium of the economy is to plot the average revenue and average cost curves of banks, given a set of prices $\left(w_t, \hat{q}_{t+1}, R_{f,t}, \hat{R}_{E,t+1}\right)$ that clear the other

[10]In equilibrium, banks will enter or exit the lending market each period until no more profits can be made. A non-integer number of banks η_t might be required for zero profits. In this situation, the equilibrium number N_t will be the largest previous integer to η_t, and banks on the circle earn positive expected profits in equilibrium. The next bank to enter, however, would earn negative expected profits, which prevents entry.

markets. The point of tangency between the two curves indicates the equilibrium. Tangency ensures all N_t banks in the loan market jointly earn zero expected profits at the profit-maximizing lending rate $R_{L,t}$.

The average revenue curve is a simple affine transformation of the demand curve for bank loans. The part of the average revenue curve at which the average cost curve lies tangent indicates the equilibrium as monopoly, competitive, or kinked. If the average cost curve touches the kink in the average revenue curve, the first-order condition of the bank's problem will hold as a strict inequality.

I focus on kinked and competitive equilibria rather than monopoly. The monopoly equilibrium does not add much to the main results, and the economy can be in a monopoly equilibrium at only one point on the average revenue curve.[11] For these reasons, I ignore it.

Dividing the bank profit function in (7) by Δ_t^i gives the average revenue and average cost functions. Denote them by $AR\left(\Delta_t^i\right)$ and $AC\left(\Delta_t^i\right)$, respectively:

$$AR\left(\Delta_t^i\right) = \frac{1}{2}\left(R_{L,t}^i + \hat{q}_{t+1}\underline{\kappa}\right), \tag{11}$$

$$AC\left(\Delta_t^i\right) = \frac{1}{2}\hat{q}_{t+1}\gamma\Delta_t^i - \left(\frac{\hat{R}_{E,t+1}}{R_{f,t}} - 1\right)\frac{\Pi_{\min}^{i,t}}{\Delta_t^i} + \hat{R}_{E,t+1}\left(1 + \frac{f}{\Delta_t^i}\right). \tag{12}$$

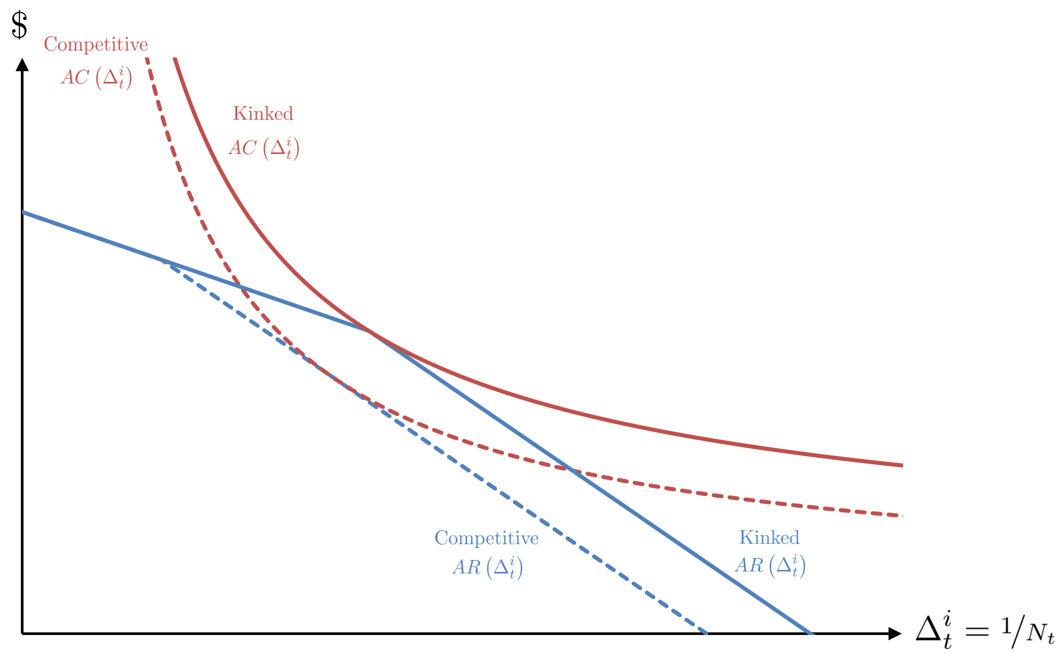
The second term in the average cost $AC\left(\Delta_t^i\right)$ is negative because it represents the cost savings from cheaper debt financing. Here, the lending rate $R_{L,t}^i$ is the inverse demand function for loans and a function of Δ_t^i. An illustration of a kinked and competitive equilibrium is presented in Figure 6.

The average cost curve is downward sloping and convex because of the fixed cost of bank entry f. In both equilibria, banks specialize their lending over non-overlapping segments of the credit market. Local bank markets cannot overlap since entrepreneurs will only select a single bank in equilibrium given their preferences. In the kinked case, *monopoly* markets just touch and competition is threatened, whereas in the competitive case, *competitive* markets just touch and competition is active.

A simple way to distinguish the two types of equilibria is to consider a deviation by a bank thinking to raise its lending rate. In a kinked equilibrium, if a bank were to raise the rate, its customers would elect to work rather than borrow. In a competitive equilibrium, its customers would still borrow, but from the neighboring bank. A positive deviation in the kinked equilibrium kicks entrepreneurs out of the credit market; in the competitive equilibrium, it relinquishes them

[11] The equilibrium requirement that $\Delta_t^i = \frac{1}{N_t}$ for all i implies local monopoly markets will just touch in a monopoly equilibrium. However, a kinked equilibrium also features local monopoly markets just touching. The only point of tangency in a monopoly equilibrium, therefore, is located immediately before the kink (approaching from the left) in the average revenue curve.

Figure 6: Kinked and Competitive Equilibria



Notes: The equilibrium lending rate $R_{L,t}$ and loan portfolio arc length Δ_t are determined at the point where the average revenue curve and average cost curve are tangent (competitive) or just touch (kinked). Kinked average revenue and average cost curves are solid; competitive are dashed.

to a competitor.

4.2 Market clearing

Five markets must clear in equilibrium: bank credit, labor, physical capital, deposit, and equity. I briefly describe each of the markets next.

The condition of no gaps on the circle ensures the market for bank credit clears: the aggregate demand for project financial capital (1) will match the aggregate supply of bank loans $\left(N_t \times \frac{1}{N_t}\right)$ at the equilibrium lending rate $R_{L,t}$.

I normalize the aggregate supply of labor to one, which means the labor endowments of depositors and bankers are arranged so that $2L^d + L^b = 1$. In equilibrium, all entrepreneurs will undertake their project rather than work, so they do not contribute to the labor supply. The wage is determined by the marginal productivity of labor:

$$w_t = (1 - \alpha)\, A_t K_t^{\alpha}. \tag{13}$$

The supply of physical capital is given by the capital-formation equation of (2). The capital demand curve is set so that the expected relative price of physical capital equates to the expected marginal product:

$$\hat{q}_{t+1} = \alpha A K_{t+1}^{\alpha-1}, \tag{14}$$

where A again is the mean of $\tilde{A}_t$.

The supply of bank deposits is the aggregate private savings of depositors s_t. From depositor preferences, their savings are

$$s_t = \frac{L^d}{(1+\beta(1-\tau))}\left(\beta w_t - \frac{\hat{w}_{t+1}}{R_{f,t}(1-\tau)}\right), \tag{15}$$

where $\hat{w}_{t+1}$ is the expected period $t+1$ wage. Banks will maximize their deposit financing, as deposits must be no more expensive than equity $\left(R_{f,t} \leq \hat{R}_{E,t+1}\right)$ for the equity market to clear. The deposit constraint (5) will bind. The demand curve for bank deposits, denoted D_t, is

$$D_t = N_t \left(\frac{\Pi_{\min}^t}{R_{f,t}}\right). \tag{16}$$

Finally, the supply of bank equity capital is the aggregate private savings of bankers: $w_t L^b$. Equity demand, denoted E_t, is the aggregate funding shortfall after deposit financing:

$$E_t = N_t \left(\frac{1}{N_t} + f - \frac{D_t}{N_t}\right). \tag{17}$$

Aggregate net worth in the banking sector is directly tied to banker savings, making it procyclical. In a downturn, the equity position of banks will deteriorate, putting upward pressure on the cost of equity capital $\hat{R}_{E,t+1}$. Bank equity will play an important role in whether the interest rate channel is effective.

4.3 Dynamics

In this section, I derive the dynamical system representing the equilibrium of the economy. The state variable of the economy turns out to be the lagged number of banks N_{t-1}, or equivalently, the current period physical capital stock K_t.

Physical capital stock

The economy displays non-trivial dynamics despite i.i.d. aggregate shocks and fully depreciating physical capital. From the formation equation in (2), the supply of physical capital in period t that is available for use in period $t+1$ is

$$K_{t+1} = \frac{1}{2}\left(\overline{\kappa} + \underline{\kappa}\right) - \frac{\gamma/2}{N_t^2}. \tag{18}$$

The first term of (18) comes from the real side of the economy. Half the projects will generate the high physical capital return $\overline{\kappa}$ and half will generate the low return $\underline{\kappa}$. The second term comes from the banking sector. It reflects the loss in the quantity of physical capital produced from the liquidation of failed projects.

For every project that fails, a bank will lose an amount γ/N_t^2 from liquidation costs. In aggregate, the total fraction of failed projects is $\frac{1}{2}$. So the amount $G(1, N_t) = \frac{\gamma/2}{N_t^2}$ *will* be lost from the physical capital stock across the banks in the credit market.

A greater number of banks in the economy means each will lend to a more specialized set of industries, making the banking sector as a whole more efficient at recovering the low physical capital return from projects in default. Aggregate physical capital production will be higher, as well as output.

Propagation

The link between aggregate savings, bank equity capital, and the number of banks in the lending market allows i.i.d. aggregate productivity shocks to propagate through time despite a fully depreciating physical capital stock. Substituting (18) into the factor market-clearing condition for the wage (13) gives

$$w_t = (1-\alpha) A_t \left(\frac{1}{2} (\overline{\kappa} + \underline{\kappa}) - \frac{\gamma/2}{N_{t-1}^2} \right)^{\alpha}, \tag{19}$$

The equilibrium evolution equation for physical capital in (18) and the wage equation in (13) is key to understand the source of the propagation. Shocks will affect the aggregate supply of deposits and bank equity capital through the wage, which will influence the cost of bank funding $R_{f,t}$ and $\hat{R}_{E,t+1}$. Costs of bank funding will partly determine the number of banks N_t sustainable in the lending market through the average revenue and average cost functions.[12] A high cost of funding, for instance, will lower potential profits in the credit market and restrict the number of banks.

In turn, the number of banks affects the efficiency of bank specialization in the production of the physical capital stock that is usable the following period. A shock A_t will thus influence production in period $t+1$, and so on, because the wage the next period is affected, as are the cost of bank funding, the number of banks N_{t+1}, and the physical capital stock the following period. A shock propagates through aggregate savings, the cost of bank funding, and bank specialization.

[12] The asset prices $R_{f,t}$ and $\hat{R}_{E,t+1}$ are available in closed form in both the competitive and kinked equilibria. I provide them in Appendix (A.5).

State variable

The state variable of the economy is the lagged number of banks N_{t-1}, as information about that variable is enough to determine the future evolution of the economy in the absence of any further shock. [13]

For a fixed level of productivity A_t, an evolution equation for N_t characterizes the equilibrium dynamical system. The evolution equation is an implicit non-linear function $F(N_{t-1}, N_t)$. In the kinked case, that function is the equity-market-clearing condition (20). In the competitive case, that function is the zero-profit condition (21). The zero-profit condition equates the average revenue and average cost curves of (11) and (12) and sets $\Delta_t^i = \frac{1}{N_t}$:

$$
\begin{aligned}
w_t L^b &= 1 + fN_t - N_t \left(\frac{\Pi_{\min}^t}{R_{f,t}} \right), && (20)\\
\frac{1}{2}\left(R_{L,t} + \hat{q}_{t+1}\underline{\kappa}\right) &= \frac{1}{2}\frac{\hat{q}_{t+1}\gamma}{N_t} - \left(\frac{\hat{R}_{E,t+1}}{R_{f,t}} - 1 \right) N_t \Pi_{\min}^t + \hat{R}_{E,t+1}\left(1 + fN_t\right). && (21)
\end{aligned}
$$

The lending rate $R_{L,t}$ in (21) is the kinked lending rate, whereas the lending rate that is implicit in $\Pi_{\min}^t$ in (20) is the competitive lending rate. We shall see next that the degree of pass-through of the interest rate to bank lending rates depends on the type of equilibrium.

5 Frustrated Monetary Policy

This section presents the main results on interest rate pass-through. In Proposition 1, I present the bank lending rates in both the competitive and kinked equilibria. I then discuss how the pass-through depends on the amount of bank equity and the type of equilibrium. The superscripts in the proposition signify different values of the endogenous objects across the two equilibria.

Proposition 1. *(Lending rates) The bank lending rate in a competitive equilibrium is*

$$
R_{L,t}^C = \frac{\hat{R}_{E,t+1}^C + \frac{1}{2}\left[\hat{q}_{t+1}^C g'\left(\frac{1}{N_t^C}\right) - \hat{q}_{t+1}^C \underline{\kappa} + \frac{2c}{N_t^C}\right] + \left(\frac{\hat{R}_{E,t+1}^C}{R_{f,t}^C} - 1\right)\phi\left(\frac{1}{N_t^C}\right)}{\frac{1}{2} + \left(\frac{\hat{R}_{E,t+1}^C}{R_{f,t}^C} - 1\right)\psi\left(\frac{1}{N_t^C}\right)}, \tag{22}
$$

where the functions ϕ and ψ are defined in Appendix A.4 by equations (35) and (36), respectively.

[13] From the physical capital equation in (2), a bijection exists between the physical capital stock K_t and the lagged number of banks N_{t-1}. Therefore, the state variable N_{t-1} can be replaced by K_t. I use N_{t-1} in the unit investment case for ease of notation. When investment can vary in section (7), the bijection is lost, and the physical capital stock K_t becomes the state variable.

The kinked equilibrium lending rate is

$$R_{L,t}^{K} = \hat{q}_{t+1}^{K}\overline{\kappa} - 2w_t L^e - \frac{c}{N_t^K}. \tag{23}$$

Proof. See Appendix A.4. □

5.1 Perfect pass-through

Consider first the competitive lending rate. Suppose the supply of equity were so large the equity market cleared at the lower-bound price $\hat{R}_{E,t+1} = R_{f,t}$. This price is the lower bound because a banker would have no reason to invest equity into a bank unless he or she earned at least the return of a depositor.

In this situation, deposits and equity are perfect substitutes, so the bank faces a single cost of financial capital $R_{f,t}$. Because the bank could finance itself entirely with equity, the credit constraint (5) would be slack. The functions ϕ and ψ in (22) reflect a bank's debt capacity. They enter the lending rate if the constraint binds. Here, they are set to zero.

The lending rate in such a competitive equilibrium would be

$$R_{L,t}^{C} = 2\hat{R}_{f,t}^{C} + \hat{q}_{t+1}^{C} g'\left(\frac{1}{N_t^C}\right) - \hat{q}_{t+1}^{C}\underline{\kappa} + \frac{2c}{N_t^C}. \tag{24}$$

The competitive lending rate reflects the marginal cost of doing business in the loan market plus a mark-up from bank market power. The first term is the marginal cost of financing passed onto entrepreneurs. The second and third terms are the marginal liquidation costs net of the recovery value from a project in default. The more a bank can get from liquidation, the less it can charge the entrepreneur. These first three terms represent the cost of borrowing if entrepreneurs picked banks solely on price with no preference for those nearby $(c = 0)$. In that case, perfect competition would drive banks to charge exactly the marginal cost of extending credit.

When banks can differentiate themselves by distance $(c > 0)$, they charge a markup over marginal cost, which is the last term in (24). The slope of the competitive demand curve and the number of banks determine the size of this markup. The costlier it is for entrepreneurs to contract with banks at a distance from their industries or areas (large c), the faster their demand for loans from remote banks drops off. The imperfect competition allows banks to exploit this feature of the demand curve by charging a larger markup against those entrepreneurs who pick them. More banks competing in the market for loans (large N_t^C) lowers individual market power and shrinks the markup.

The interest rate enters the bank lending rate linearly, so perfect pass-through occurs. Competition is the reason: if a bank observes its cost of funding drop, and does not lower its lending rate, it would lose customers to a neighboring bank. To remain in business, the bank must pass through any changes in the interest rate to its lending rate.

5.2 Imperfect pass-through

Now suppose equity were scarce, so that the equity market cleared at price $\hat{R}_{E,t+1} > R_{f,t}$. A bank would have strict preference for cheaper deposit financing, so the credit constraint would bind. The debt capacity of the bank now becomes important when the bank chooses the competitive loan rate.

The loan rate is now (22). The rate reflects the blend of debt and equity in the bank's financial capital structure. The functions ϕ and ψ capture this blend. They adjust the marginal cost of financing from an incremental growth in the loan portfolio. This fact can be seen clearly from Lemma (3), which presents a typical bank's marginal financing cost function $FC'\left(\Delta_t^i\right)$ as it increases its portfolio size Δ_t^i:

Lemma 3. *The marginal financing cost function of a typical bank i is*

$$\frac{\partial FC_t^i}{\partial \Delta_t^i} = \hat{R}_{E,t+1} + \left(\frac{\hat{R}_{E,t+1}}{R_{f,t}} - 1\right)\left[\phi\left(\Delta_t^i\right) - \psi\left(\Delta_t^i\right) R_{L,t}^i\right]. \tag{25}$$

Proof. See Appendix A.4. □

A marginal increase in a bank's loan portfolio has two effects on its cost of funding. The first effect is a higher financing cost from the need for more equity to fund the portfolio at price $\hat{R}_{E,t+1}$. The second effect is a decrease to the cost of funding as the bank tilts its capital structure to cheaper debt financing because of greater diversification.

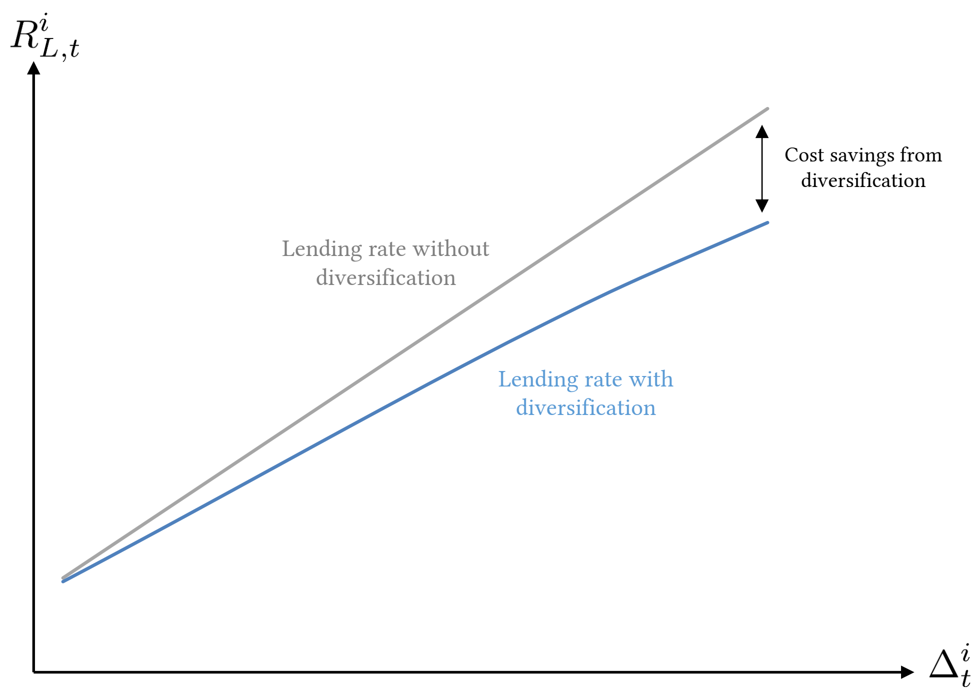
The second term in (25) are cost savings from diversification. They reflect changes to the minimum possible loan profit of the bank $\Pi_{\min}^{i,t}$, which determines its debt capacity. For a fixed lending rate $R_{L,t}^i$, an expansion in Δ_t^i increases the debt capacity of the bank at rate ψ. The higher debt capacity generates financing cost savings at rate $\left(\frac{\hat{R}_{E,t+1}}{R_{f,t}} - 1\right)$, which are passed onto entrepreneurs, in the form of a lower lending rate in equilibrium.

Greater diversification decreases the bank's minimum failure rate on its portfolio $(1 - \Pr_{\min})$. A lower failure rate means the bank will receive payment $R_{L,t}^i$ on more of its loans. It also means the bank will retrieve the low physical capital return $\hat{q}_{t+1}\underline{\kappa}$ net of the liquidation costs on less of its loans, as fewer will default. Less recovery values from fewer defaults reduces the minimum loan profit of the bank and its debt capacity. The function ϕ is the rate at which debt capacity decreases with Δ_t^i. Lower debt capacity increases marginal financing costs, raising the competitive lending rate.

Figure 7 depicts the benefits of diversification on the competitive lending rate. The figure plots the loan rate in (22) when the functions ϕ and ψ are zero (so there are no benefits of diversification) and when the functions are non-zero (so there are benefits). The larger the bank's Δ_t^i, the more diversified its loan portfolio, so the more deposits it can raise. Cheaper deposit financing lowers bank financing costs, which reduces the competitive lending rate. These cost savings decline as the bank becomes less diversified because the bank can raise fewer and fewer deposits when its loan portfolio becomes more and more specialized.

When the credit constraint on bank financing binds, the degree of pass-through now relies on bank financial capital structure. The relation is non-linear and imperfect. Interest rate pass-through depends on the functions $\phi\left(\Delta_t^i\right)$ and $\psi\left(\Delta_t^i\right)$, which reflect the bank's debt capacity and its diversification. The pass-through also depends on the cost of equity capital $\hat{R}_{E,t+1}$, which is a function of aggregate bank net worth.

Figure 7: Competitive Lending Rate With and Without Diversification Benefits



Notes: The gray curve is the competitive lending rate in (22) if the functions ϕ and ψ are set to zero so that the bank realizes no diversification benefits. The blue curve is the lending rate in (22) unaltered. Factor prices and costs of financial capital are held fixed in the illustration.

5.3 No pass-through

In a kinked equilibrium, banks operate off the kink in the demand curve for loans, which means their profit-optimality condition will hold as a strict inequality. It does not determine the equilibrium lending rate. The marginal cost of extending loans does not either. The lending rate instead is taken off the monopoly portion of the demand curve for bank credit, given by (23).

Banks in a kinked equilibrium are local monopolists in segmented industries of the credit market in which they specialize. They compete against the outside option of entrepreneurs rather than with each other. As seen in (23), a bank charges a higher lending rate if a project yields a higher successful return $\hat{q}_{t+1}\overline{\kappa}$. The bank exploits the attraction of pursuing a project. On the other hand, the bank cuts back on the lending rate if the wage w_t is high. A high wage implies working is a strong alternative to pursuing a project. To entice an entrepreneur to take out a loan, the bank has to lower the rate.

The kink in the demand curve generates a jump discontinuity in the marginal revenue curve of a bank. For this reason, if banks operate at the kink in equilibrium, small changes in the cost of *providing* a loan do not affect the cost of *obtaining* a loan. No component of the marginal financing cost function of the bank enters (23).

This result is critical to understanding why monetary policy is ineffective if the economy is in a kinked equilibrium. Any changes in the interest rate has no direct impact on the lending rate. Competition is missing to compel banks to adjust their loan rates after changes to their cost of funding. Each bank knows every other bank will not deviate from the kinked lending rate $R^K_{L,t}$. So no bank does. There is no pass-through.

The kink generates a sharp prediction of no pass-through. Generally, a region of a demand curve that features higher concavity leads a monopolist to pass through less of any changes to marginal cost. The kink creates a sharp concavity in the demand curve for loans. In Appendix A.9, I provide one way of "smoothing" the kink. I assume that banks are unsure of which part of the demand curve they are on when setting a price. Their uncertainty is a function of the size of a bank's loan portfolio Δ^i_t : the larger its portfolio, the more confidently it believes it competes with a neighbor. I provide conditions on their uncertainty that preserves the limited pass-through at the "smoothed" kink.

5.4 Negative pass-through

A central bank that wants to lower the lending rate for commercial credit would adjust the tax rate τ in order to influence the quantity of aggregate savings and reduce the interest rate $R_{f,t}$. In a kinked equilibrium, the only effect of a lower interest rate on bank lending rates will be indirect through the number of banks N^K_t.

A lower interest rate will reduce the average cost of operating a bank, increase profits, and encourage entry into the lending market. An important perverse feature of the kinked lending rate in (23), however, is that more banks in the credit market actually leads all of them to *raise* their loan rates.

In a kinked credit market, banks are local monopolists. More banks on the circle means that an entrepreneur can find one that specializes in an industry or area "closer" to the entrepreneur's. A bank takes advantage of its greater local monopoly power by charging a higher lending rate.

Conversely, fewer banks lead all of them to *reduce* their loan rates. When a bank exits the lending market, the average entrepreneur needs to "travel" a longer distance on the circle, contracting with a bank that is less specialized in his or her particular industry or location than before. Undertaking the project becomes less attractive to the entrepreneur relative to the outside option of working. Because the typical bank in a kinked market is competing against its borrowers' outside options, it needs to lower the lending rate to encourage the entrepreneur to borrow instead.

By encouraging bank entry, an accommodative monetary policy has the unintended effect of increasing the cost of bank credit to firms and worsening the commercial loan spread. I call a decrease to the interest rate that leads to an increase in the bank loan rate (and vice versa) "negative pass-through." This result is demonstrated formally in Proposition (2).

Proposition 2. *(Negative pass-through) In steady state, for a sufficiently low fixed cost of entry* f *and factor price-to-physical-capital ratios* ${}^{\hat{q}_{ss}}/_{K_{ss}}$ *and* ${}^{w_{ss}}/_{K_{ss}}$*, a change in the tax rate* τ *that lowers the interest rate* $R_{f,ss}$ *increases the number of banks* N_{ss} *and raises the kinked lending rate* $R^K_{L,ss}$*.*

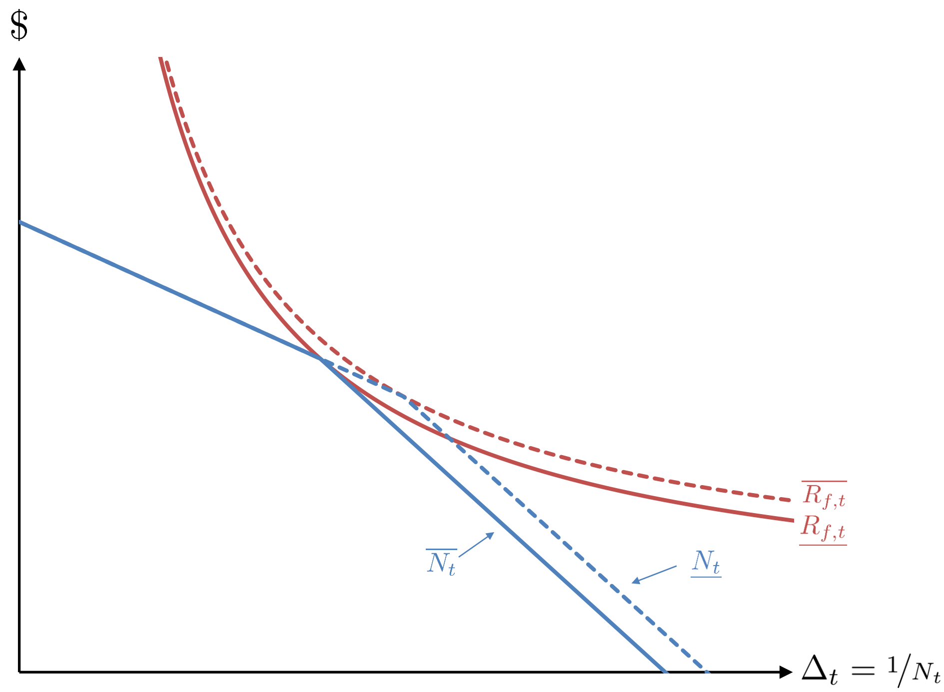
Proof. See Appendix A.6 □

Bank entry increases specialization, which raises the physical capital stock. A higher capital stock lowers the expected price of physical capital and increases the wage. The condition on the factor price-to-physical-capital ratios in the proposition ensures the decline in the expected price of physical capital and increase in the wage in steady state does not off set the positive effect a higher number of banks has on the lending rate. The condition on the fixed cost of entry ensures the increased demand for deposit financing by the entering banks does not raise the interest rate. The exact conditions are provided in Appendix A.6.

Figure (8) illustrates the effect of a lower interest rate on the average cost curve that encourages bank entry. The average cost curve shifts inward from the lower funding costs, increasing profits of existing banks. Positive profits give a reason for other banks to enter the lending market.

The negative pass-through requires the number of banks in the lending market to adjust. In reality, bank entry or consolidation takes time. The perversity is then a long-run rather than short-run effect. One can think of the short run as prior to any adjustment to N_t. Even in the short run, the interest rate channel is still closed, as any adjustment to the interest rate has no impact on the kinked lending rate.

Figure 8: Bank Entry from Lower Interest Rate, Kinked Equilibrium



Notes: The interest rate declines from $\overline{R_{f,t}}$ to $\underline{R_{f,t}}$, lowering the average cost curve from the dotted to solid line. This change leads the number of banks to increase from $\underline{N_t}$ to $\overline{N_t}$, which pushes the average revenue curve inward from the dotted to solid line. The wage w_t, expected wage $\hat{w}_{t+1}$, expected relative price of physical capital $\hat{q}_{t+1}$, and expected return on equity $\hat{R}_{E,t+1}$ are held fixed in the illustration.

6 Bank Net Worth

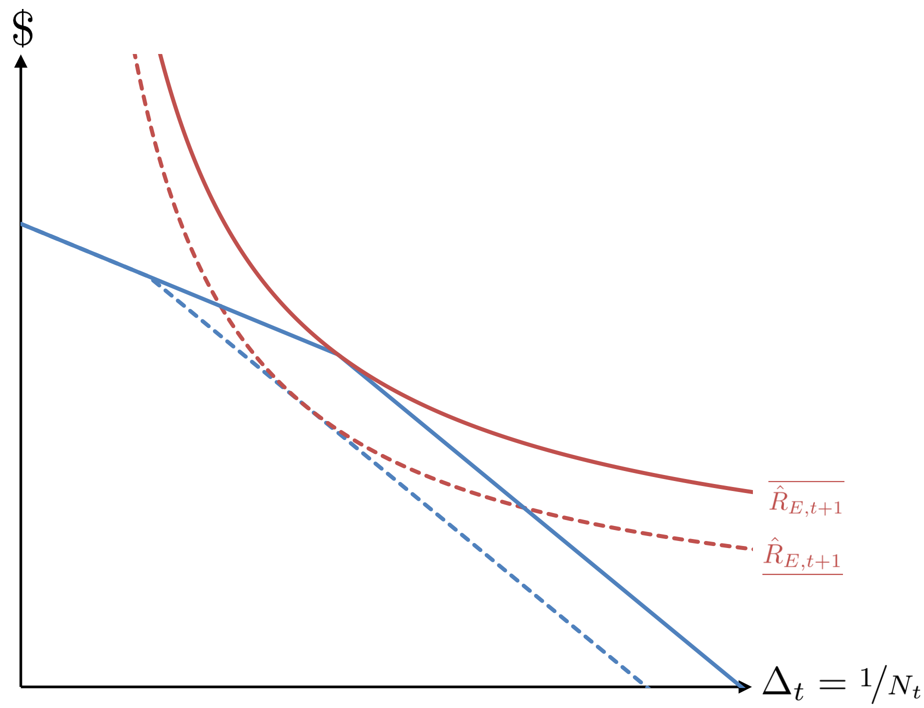
In this section, I discuss how bank equity determines whether the credit market is in a competitive or kinked equilibrium.

6.1 Role of Net Worth

Large enough changes in net worth transition the economy between equilibria. Indeed, a sudden drop in bank equity positions following a negative productivity shock can move the credit market to the kink. All banks in that situation retreat from competition. Each exploits its local monopoly power over entrepreneurs, and all settle at the focal kinked lending rate. No bank thinks any other will deviate in pricing, so all tacitly agree to refrain from competing with one another. The economy can get stuck at the kink for as long as net worth remains low.

Figure 9 presents an illustration of how a large enough decline in net worth raises the cost of equity $\hat{R}_{E,t+1}$ and the average cost curve of banks. The rise in cost transitions the credit market from a competitive to kinked equilibrium.

Figure 9: Transition to Kinked Equilibrium after Decline in Bank Net Worth



Notes: The economy starts in a competitive equilibrium with cost of equity $\underline{\hat{R}_{E,t+1}}$, represented by the dotted lines. An adverse productivity shock lowers aggregate bank net worth E_t and raises the cost of equity to $\overline{\hat{R}_{E,t+1}}$. The average cost curve increases, leads to bank consolidation, and pushes the economy to a kinked equilibrium. The kinked equilibrium is represented by the solid lines. The expected wage $\hat{w}_{t+1}$, expected relative price of physical capital $\hat{q}_{t+1}$, and interest rate $R_{f,t}$ are held fixed in the illustration.

6.2 Determining the Equilibrium

The equilibrium of an economy is determined by a set of conditions that I explain in detail in Appendix A.7. Broadly speaking, the economy is in a kinked equilibrium if the entrepreneur at the edge of a bank's market is indifferent between working and borrowing at the equilibrium lending rate. The economy is in a competitive equilibrium if the entrepreneur at the edge of the market strictly prefers borrowing to working.

Variation in either the productivity shock A_t or the physical capital stock K_t move an economy between equilibria. Aggregate bank net worth $E_t = w_t L^b$ is a composite of A_t and K_t, so it encodes information about both. Therefore, one can think of bank net worth as the state variable that determines the equilibrium.

The conditions that determine the equilibrium of an economy are expressible as bounds on the expected return on equity. The following condition uniquely determines when an economy is in a competitive equilibrium:

$$\hat{R}^{C}_{E,t+1} \leq h^{C}\left(E_{t}\right), \tag{26}$$

where h^C is a nonlinear function of net worth and defined in expression (46) in Appendix A.7. The intuition behind (26) is that the cost of equity cannot be so high that it restricts entry and competition among banks.

The expected return on equity in a competitive equilibrium $\hat{R}^{C}_{E,t+1}$ is equation (41) in Appendix A.5, and is also a nonlinear function of bank net worth. Values of E_t for which the two functions share the relation in (26) implies an economy featuring those levels of bank net worth is in a competitive equilibrium.

An economy is in a kinked equilibrium if and only if the following two conditions hold:

$$\hat{R}^{C}_{E,t+1} > h^{C}\left(E_{t}\right),$$
$$\hat{R}^{M}_{E,t+1} < h^{M}\left(E_{t}\right),$$

where h^M is defined in expression (47) in Appendix A.7, and $\hat{R}^{M}_{E,t+1}$ is the expected return on equity in a monopoly equilibrium, given in Appendix A.5. The intuition behind these two conditions is that the cost of equity must exceed the threshold that signifies a competitive equilibrium, but also be low enough to allow positive monopoly profits, which characterizes a kinked equilibrium.

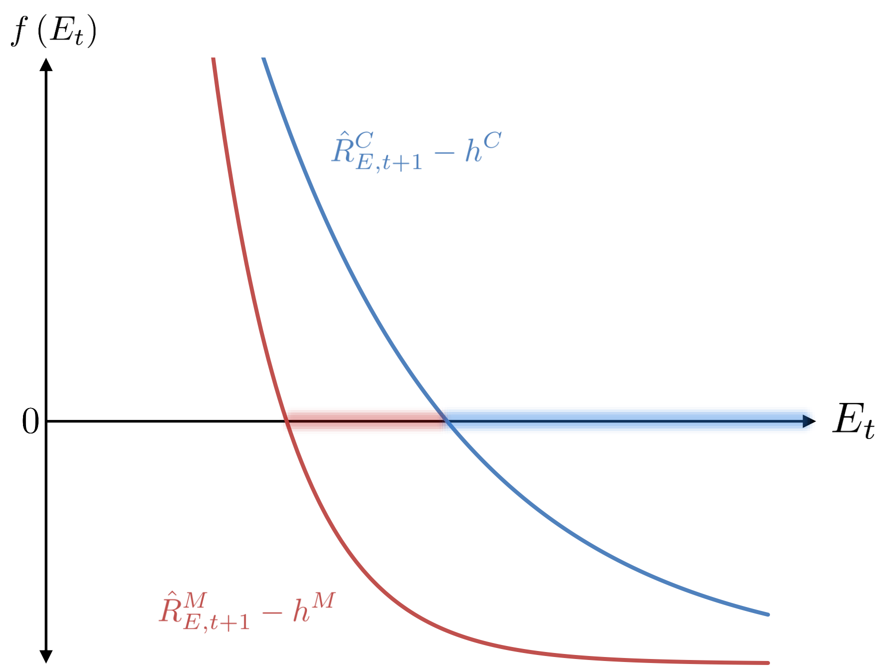
Figure 10 illustrates the "competitive" function $\hat{R}^{C}_{E,t+1} - h^C$ and the "monopoly" function $\hat{R}^{M}_{E,t+1} - h^M$. Values of bank net worth for which the competitive function is negative (shaded in blue) constitute a competitive equilibrium, whereas values of net worth for which the competitive function is positive and the monopoly function is negative (shaded in red) constitute a kinked equilibrium. The single value of net worth for which the monopoly function is exactly zero constitutes a monopoly equilibrium, because at that point, monopoly profits are zero.

The interval of bank net worth marking a kinked equilibrium must lie between those intervals marking a competitive and monopoly equilibrium. For this arrangement, both the monopoly and competitive functions must both be either monotonically increasing or monotonically decreasing.

The function h^C is monotonically decreasing in bank net worth. Therefore, if the competitive expected return on equity function $\hat{R}^{C}_{E,t+1}$ is decreasing in bank net worth, the competitive function $\hat{R}^{C}_{E,t+1} - h^C$ must also be decreasing. Hence, the monopoly function $\hat{R}^{M}_{E,t+1} - h^M$ is also decreasing, as illustrated in Figure 10.

When bank net worth is high, the cost of equity is low and banks actively compete in the lending market. Following a severe enough productivity shock, however, net worth in the banking sector becomes impaired and the economy shifts into the red kinked region where banks refrain from competition and the interest rate channel is closed. The economy persists in the kinked region until bank net worth recovers enough to reach a value in the shaded blue region. With

Figure 10: Bank Net Worth Determining the Equilibrium



Notes: The values of aggregate bank net worth E_t for which $\hat{R}^{C}_{E,t+1} - h^{C} < 0$ (represented by the blue shaded area) constitute a competitive equilibrium. The values of bank net worth for which both $\hat{R}^{C}_{E,t+1} - h^{c} > 0$ and $\hat{R}^{M}_{E,t+1} - h^{M} < 0$ (represented by the red shared area) constitute a kinked equilibrium.

equity capital positions improved, banks have incentive to break the tacit agreement in their pricing, enter the markets of other banks, and resume competition. The interest rate channel then opens.

6.3 Relation to the C&I Loan Market

Figure 10 can help us better understand the behavior of the C&I loan spread during and after the financial crisis. Severe losses impaired bank balance sheets and forced many banks to be absorbed by more solvent ones. The resulting sharp increase in bank concentration in the C&I loan market and reduction in aggregate bank net worth is in line with the bank credit market transitioning to the red region in the figure. At this stage, the central bank was limited in its ability to transmit a low policy rate to bank lending rates. Hence the persistently high C&I loan spread despite declining measures of C&I loan default risk. Over time, as banks shored up their equity capital positions, they resumed more aggressive competition, which would correspond to the credit market moving into the blue region. There the central bank would have more success in transmission. Hence the C&I loan spread gradually fell, moving more in tandem with the default-risk measures.

7 Variable Investment Scale

The baseline model featured a single unit of investment to initiate a project. Aggregate investment was fixed and the only dynamics originated from persistence in the banking sector's efficiency at liquidating failed projects. Here, the scale of project investment can vary. Importantly, this change allows a productivity shock to affect aggregate investment and to propagate through time from both the real side of the economy and the banking sector. Also, an impairment to interest rate pass-through now directly obstructs the central bank from affecting investment and influencing real output.

A simple way to allow variable investment is to insert concavity into the production function of a project. Investment thus depends on the cost of bank credit. Rather than a constant returns-to-scale technology, the production function is now

$$\kappa\left(\iota_t^j\right) = \tilde{\kappa}\left(1 - \exp\left(-\iota_t^j\right)\right), \tag{27}$$

where $\tilde{\kappa} = \overline{\kappa}$ in the high state and $\tilde{\kappa} = \underline{\kappa}$ in the low state. The probability of the high state is again given by (1).

The function in (27) is continuously differentiable, strictly increasing, strictly concave, satisfies $\kappa(0) = 0$, and has a first derivative that vanishes as $\iota_t^j \to \infty$. All these attributes are appropriate for a production function. This production function will generate results that are analogous to the unit investment case, which is why I use it.

Entrepreneurs choose an amount of output to invest into the project at time t. Investment decisions are independent of an entrepreneur's location. So I denote by ι_t^i an investment by an entrepreneur who considers a loan from bank i at rate $R_{L,t}^i$. That investment decision satisfies

$$V_{j,i}^e \equiv \max_{\iota_t^i}\ E_t\left[\pi\left(\iota_t^i\right)\right] - c|i-j|$$

with

$$E_t\left[\pi\left(\iota_t^i\right)\right] = \frac{1}{2}\left[\hat{q}_{t+1}\overline{\kappa}\left(1 - \exp\left(-\iota_t^i\right)\right) - R_{L,t}^i \iota_t^i\right].$$

The optimal investment is

$$\iota_t^i = \log\left(\frac{\hat{q}_{t+1}\overline{\kappa}}{R_{L,t}^i}\right). \tag{28}$$

Investment is now decreasing in the lending rate and increasing in the expected relative price of physical capital. To ensure positive investment, the high return of the project $\overline{\kappa}$ must be large enough so that

$$\hat{q}_{t+1}\overline{\kappa} > R_{L,t}^i.$$

So long as $\overline{\kappa}$ is sufficiently above 1, this relation will hold for all reasonable values of the gross lending rate $R^i_{L,t}$ and the expected relative price of physical capital $\hat{q}_{t+1}$.

7.1 Demand curve for loans

Exactly as in the unit investment case, the demand curve for bank credit will consist of monopoly, kinked, and competitive components.

The indifference condition of the marginal entrepreneur who defines the size of the local monopoly market is

$$E_t\left[\pi\left(\iota^i_t\right)\right] - cx - w_t L^e = 0.$$

Substituting the entrepreneur's expected profit at the optimal level of investment from (28), solving for x, and multiplying by 2 gives the size of the local monopoly market:

$$\Delta^{i,M}_t = \frac{\frac{1}{2}\left[\hat{q}_{t+1}\overline{\kappa} - R^i_{L,t}\left(1+\iota^i_t\right)\right] - w_t L^e}{c/2}. \tag{29}$$

The monopoly demand curve here is virtually the same as the one in (9), where projects required a single unit. The only difference is an adjustment for the optimal investment ι^i_t.

When banks compete and an entrepreneur has a choice between two banks for a loan, the entrepreneur will choose the lowest-cost financing. The indifference condition for the marginal entrepreneur is

$$\frac{1}{2}\left[R^i_{L,t}\left(1+\iota^i_t\right)\right] + cx = \frac{1}{2}\left[R_{L,t}\left(1+\iota_t\right)\right] + c\left(\frac{1}{N_t} - x\right),$$

making the competitive demand curve

$$\Delta^{i,C}_t = \frac{\frac{1}{2}\left[R_{L,t}\left(1+\iota_t\right) - R^i_{L,t}\left(1+\iota^i_t\right)\right] + \frac{c}{N_t}}{c}. \tag{30}$$

Here I denoted the investment of an entrepreneur who contracts with the competing bank by ι_t. As in the monopoly case, the only adjustment in the competitive demand curve compared to (10), when the scale of projects was fixed, is the inclusion of the investment scale.

The different slopes of the monopoly and competitive demand curves generate the kink. The slopes of the monopoly and competitive demand curves are:

$$\frac{\partial R^{i,M}_{L,t}}{\partial \Delta^i_t} = -\frac{c}{\iota^i_t},$$
$$\frac{\partial R^{i,C}_{L,t}}{\partial \Delta^i_t} = -\frac{2c}{\iota^i_t}.$$

Both slopes now reflect the investment scale, and again, the competitive demand curve slope is twice that of the monopoly.

7.2 Bank problem

Banks again are restricted to finance projects that are positioned along arcs Δ_t^i centered at the headquarters. But now, each loan has size ι_t^i, rather than a single unit. Because of these two dimensions to lending, I refer to Δ_t^i as the "breadth" of bank $i's$ loan portfolio at time t, and ι_t^i as the "depth" of the portfolio. The size of the total loan portfolio of bank i is then $\Delta_t^i \iota_t^i$.

The bank diversifies into new industries that require new expertise from expansions in Δ_t^i rather than ι_t^i. For that reason, I augment the total liquidation cost function g to be linear in ι_t^i, but again quadratic in Δ_t^i:

$$g\left(\Delta_t^i, \iota_t^i\right) = \gamma \iota_t^i \left(\Delta_t^i\right)^2.$$

Expected profit of the typical bank now includes the scale of project financing:

$$\Pi_t^i = \frac{1}{2} R_{L,t}^i \Delta_t^i \iota_t^i + \frac{1}{2} \hat{q}_{t+1} \left(\underline{\kappa}\left(\iota_t^i\right) \Delta_t^i - g\left(\Delta_t^i, \iota_t^i\right)\right) - FC_t^i.$$

So too does the financing cost function:

$$FC_t^i = R_{f,t} D_t^i + \hat{R}_{E,t+1}\left(\Delta_t^i \iota_t^i + f - D_t^i\right).$$

The bank requires a total of $\Delta_t^i \iota_t^i + f$ in financing. Depositors will supply a restricted quantity of deposits again, so the bank faces the constraint

$$D_t^i \leq \frac{\Pi_{\min}^{i,t}}{R_{f,t}},$$

with minimum loan profits now given by

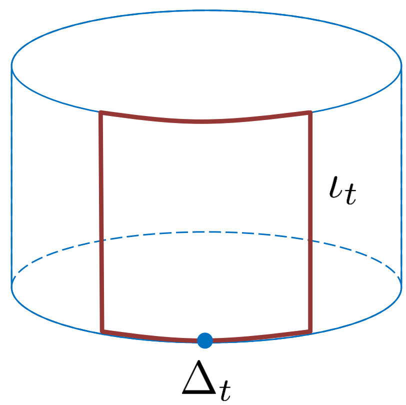
$$\begin{aligned} \Pi_{\min}^{i,t} \quad = \quad & \Pr_{\min}\left(R_{L,t}^i | \Delta_t^i\right) R_{L,t}^i \Delta_t^i \iota_t^i \\ & + \left(1 - \Pr_{\min}\left(R_{L,t}^i | \Delta_t^i\right)\right) \hat{q}_{t+1} \left(\underline{\kappa}\left(\iota_t^i\right) \Delta_t^i - g\left(\Delta_t^i, \iota_t^i\right)\right). \end{aligned}$$

The bank chooses $R_{L,t}^i$ and D_t^i to maximize expected profits subject to the deposit constraint, while taking factor prices, costs of financial capital, the number of banks, the loan-pricing strategies of those banks, and the demand curve for bank credit as given.

7.3 Equilibrium

I study the same types of Nash equilibria as in the main text. Because I look at symmetric equilibria, lending rates are the same across banks, so all entrepreneurs will make the same investment choices. Thus, $\iota_t^i = \iota_t$ for all i. Therefore, the equilibrium loan portfolio of a typical bank i can be represented graphically as the surface area of a sector of a cylinder. This representation is depicted in Figure 11.

Figure 11: Equilibrium Loan Market and Bank Loan Portfolio Representation



Notes: Projects are uniformly distributed around the circle. Bank i is located at the dot. The bank's loan portfolio in equilibrium has breadth $\Delta_t = \frac{1}{N_t}$ and depth $\iota_t = \log\left(\frac{\hat{q}_{t+1}\overline{\kappa}}{R_{L,t}}\right)$.

7.4 Bank lending rates

Given a set of factor prices w_t and $\hat{q}_{t+1}$ and costs of financial capital $R_{f,t}$ and $\hat{R}_{E,t+1}$, the equilibrium lending rate $R_{L,t}$ and number of banks N_t again are jointly determined by the point of tangency between the average revenue and average cost curves of a typical bank.

Dividing the expected bank profit function in (7) by Δ_t^i gives the average revenue and average cost functions:

$$AR\left(\Delta_t^i, \iota_t^i\right) = \frac{1}{2}\left(R_{L,t}^i\left(\iota_t^i - \underline{\kappa}/\overline{\kappa}\right) + \hat{q}_{t+1}\underline{\kappa}\right)$$
$$AC\left(\Delta_t^i, \iota_t^i\right) = \frac{1}{2}\hat{q}_{t+1}\gamma\Delta_t^i\iota_t^i - \left(\frac{\hat{R}_{E,t+1}}{R_{f,t}} - 1\right)\frac{\Pi_{\min}^{i,t}}{\Delta_t^i} + \hat{R}_{E,t+1}\left(\iota_t^i + \frac{f}{\Delta_t^i}\right).$$

The average revenue and average cost of a typical bank are functions of the bank's loan portfolio breadth Δ_t^i and depth ι_t^i. Tangency between the average revenue and average cost curves, along with all markets clearing, will pin down the number of banks and the lending rate in the economy.

In the kinked equilibrium, the lending rate is available in closed form by re-arranging (29) and setting $\Delta_t^i = \frac{1}{N_t}$. Doing so gives

$$R_{L,t}^K = \frac{\hat{q}_{t+1}\overline{\kappa} - 2w_t L^e - \frac{c}{N_t}}{1+\iota_t}. \tag{31}$$

The kinked lending rate is virtually the same as the case with unit investment, save for the adjustment for the investment level ι_t in the denominator. A greater level of project investment is associated with a lower lending rate. Because the investment ι_t is also a function of $R_{L,t}^K$, equation (31) expresses the lending rate implicitly. Note the cost of bank financing again does not enter the kinked lending rate.

7.5 Dynamics

The derivation of the equilibrium dynamical system proceeds exactly as in the unit investment case. The market-clearing equations are virtually identical, save for some minor adjustments to account for the investment scale ι_t. I include the full set of equations in Appendix A.10.

The state variable of the system is the physical capital stock K_t. Unlike the fixed-investment case, the lagged number of banks N_{t-1} cannot also serve as the state variable.

The most significant change to the dynamical system compared to the fixed-investment case is the formation equation of physical capital. When investment can vary, (2) becomes

$$K_{t+1} = \underbrace{\frac{1}{2}\left(\overline{\kappa} + \underline{\kappa}\right)}_{\substack{\text{unit}\\ \text{investment}}} - \underbrace{\frac{1}{2}\left(1 + {}^{\underline{\kappa}}\!/\!_{\overline{\kappa}}\right)\left(\frac{R_{L,t}}{\hat{q}_{t+1}}\right)}_{\substack{\text{scale}\\ \text{adjustment}}} - \underbrace{\frac{\gamma/2}{N_t^2}\log\left(\frac{\hat{q}_{t+1}\overline{\kappa}}{R_{L,t}}\right)}_{\substack{\text{liquidation}\\ \text{costs}}}. \tag{32}$$

The first term of (32) is the physical capital production from projects if they required a single unit of investment. Half will succeed and produce the high return $\overline{\kappa}$, and half will fail and produce the low return $\underline{\kappa}$. The second term is new. It reflects the adjustment to the physical capital stock from the scale of project investment ι_t. Investment is decreasing in the bank lending rate $R_{L,t}$ and increasing in the expected relative price of physical capital $\hat{q}_{t+1}$. The final term is the aggregate bank liquidation costs, which depress the physical capital stock. Because liquidation costs are proportional to investment, they now are a function of the lending rate and the expected relative price of physical capital.

7.6 Monetary policy and the real economy

When investment can vary, the industrial organization of the banking sector and impairment to the interest rate channel has a direct effect on the production of physical capital from the real side of the economy. The equilibrium lending rate $R_{L,t}$ will differ depending on whether banks actively compete across industries or act as local monopolists in segregated areas of the lending market.

If the economy is in a competitive equilibrium, an accommodative monetary policy through a decrease in the interest rate $R_{f,t}$ will pass through to bank lending rates and spur investment. In the kinked equilibrium, however, a lower interest rate will have neither effect on the cost of bank credit or investment.

Indeed, a lower interest rate would have the perverse effect of *decreasing* investment in the kinked equilibrium. A decline in the interest rate would lower the financing costs of banks and encourage entry. In the kinked equilibrium, entry increases the lending rate, which lowers investment. Again, the perverse effect is in the long run after time has elapsed for banks to enter the lending market.

A secondary effect of a lower interest rate is on the liquidation costs. A greater number of banks N_t and lower investment ι_t narrow bank specialization and lowers aggregate liquidation costs. So the total effect of an accommodative monetary policy in the kinked equilibrium is a reduction in aggregate investment but also a reduction in liquidation costs.

With investment dependent on the cost of bank credit and the relative price of physical capital, the economy now features cyclical persistence that originates both from investment fluctuations out of the productive sector as well as changes in the efficiency of projection liquidation out of the banking sector.

8 Conclusion

This paper presents a model in which the industrial organization of the bank credit market affects the real economy and monetary policy. A driving force of that competitive structure is the net worth of banks. A sufficient drop in aggregate bank equity transitions the economy to an equilibrium where banks consolidate for survival, retreat to local monopoly markets, and tacitly collude not to compete. A wide commercial loan spread and impaired interest rate pass-through can persist, lowering investment and output.

An important contribution of this paper is to tie bank net worth to the degree of competition and specialization in the banking sector, as well as to the functioning of the interest rate channel of monetary policy. Lower bank net worth raises the cost of equity and encourages local monopolies. This development obstructs the interest rate channel.

The negative interest rate pass-through over the long run in a kinked equilibrium presents a dilemma for the central bank. On the one hand, the government has reason to restrict entry (or encourage consolidation) among banks when aggregate equity capital is low to prevent bank loan rates from rising. On the other hand, fewer specialized banks reduce the physical capital stock and output. A central bank must trade-off these effects when choosing optimal policy.

The number of banks N_t in the commercial credit market plays a significant role in the economy. The number proxies for the efficiency of the banking sector at recovering distressed physical capital. This efficiency is one source of cyclicality.

One potential criticism is that the number of banks in the U.S. economy is not cyclical. Since the mid-1980s, the number of commercial banks has steadily fallen from 14,000 to 4,000.[14] This secular decline corresponds to the trend of bank consolidation that began with deregulation of interstate banking.

In the model, the number of banks is isomorphic to the specialization Δ_t of bank lending. Changes to N_t can equivalently be interpreted as changes to the degree of specialization in the banking sector. Acharya et al. (2006) and Berger et al. (2017) document significant variation in bank specialization across industries in the cross section and through time. Paravisini et al. (2015) does the same for specialization across exporting countries.

Regarding welfare, spatial models of firm competition with unit demand typically display excess entry in the decentralized equilibrium relative to the social optimum (see Vickrey (1964); Salop (1979); Anderson et al. (1992); Matsumura and Okamura (2006); Gu and Wenzel (2009)). Banks decide to enter based on the marginal borrower, while welfare is determined using the average one. Whether there is excess entry depends on the difference between the average borrower surplus and the surplus of the marginal borrower relative to the fixed cost of entry.

A novelty here is that a lot of banks has its benefits in physical capital production due to the evolution equation for physical capital in (20). More banks increases specialization in lending, and as a result, the physical capital stock, output, and consumption increase. Determining whether the equilibrium number of banks is too few, too many, or optimal will require a formal welfare analysis.

Regarding government policy, common tools to resolve a financial crisis are (1) reducing borrowing costs, (2) providing equity capital, and (3) purchasing distressed assets. The focus of the paper is the industrial organization of the banking sector. Therefore, the government could also consider (4) adjusting entry or exit to the lending market. These four tools are available to the government to try to restore transmission via the interest rate channel after a severe deterioration in aggregate bank equity.

In the model, the quality of assets does not enter the pricing decisions of banks, so purchasing assets will not resume transmission. Reducing borrowing costs is the same as lowering the interest rate. If the banking sector is in the kinked equilibrium, there will still be no pass-through despite accommodative monetary policy. In the long run, banks enter the lending market from a reduction in borrowing costs, which will increase the capital stock but also raise the cost of borrowing for firms due to the negative pass-through.

Providing equity will lower the cost of equity for banks and encourage the lending market to

[14] *Source*: Federal Financial Institutions Examination Council.
Data available: https://fred.stlouisfed.org/series/USNUM

transition from a kinked to a competitive equilibrium and open the interest rate channel. Therefore, this tool seems particularly effective at resuming transmission after a negative shock to the equity of the banking sector. The government in the model would finance an equity injection by taxing young depositors and becoming co-equity holders with bankers. The government would then rebate the equity earnings to the depositors when they are old.

Finally, it may be optimal for the government to restrict entry into the lending market if the economy is at the kink in order to prevent the negative pass-through. Such a policy will decrease bank specialization and lower the capital stock as a consequence, however.

Although the economic mechanism differs because of the element of imperfect competition, the model in this paper relates to the literature on intermediary asset pricing (e.g., He and Krishnamurthy (2013); Brunnermeier and Sannikov (2014)) in the following ways: (1) scarcity of bank equity makes the spread of risky assets (loans) over safe assets (deposits) rise after negative shocks to bank equity positions; (2) this elevated spread can persist as long as bank equity does not improve; and (3) equity capital injections are especially effective at reducing the risky asset spread. Key novelties here are that (1) the model can generate a stickiness in the spread despite improvements in equity positions, as long as banks continue to refrain from competing; and (2) the model features entry into the lending market, whereas the models in the intermediary asset pricing literature commonly do not.

In the model, banks enter or exit within the period. Typically bank exit (consolidation) occurs more rapidly than bank entry. If banks in the economy could only enter with delay–as in, after a certain number of periods–but could exit immediately, then existing banks in the lending market could earn positive profits until entry by competing banks. Allowing this asymmetry in entry and exit would allow banks to capture retained earnings and recapitalize their balance sheets. This kind of "self-healing" is missing from the model, but such a process would hasten a transition away from the kink that would re-open the interest rate channel.

The effects of the industrial organization of banks on the real economy is a topic ripe for future research. Recent empirical work has explored the issue. Paravisini et al. (2015) find credit-supply shocks to a specialized bank have a disproportionate effect on exports to that bank's country of expertise. Some unresolved questions are: What makes specialized bank debt difficult to replace? What distinguishes relationship skills from industry expertise? How does the failure of a specialized bank distort the allocation of credit? Much remains to be studied.

References

Acharya, Viral V., Iftekhar Hasan, and Anthony Saunders. Should banks be diversified? evidence from individual bank loan portfolios. *The Journal of Business*, 79(3):1355–1412, May 2006. [7], [45]

Anderson, Simon P., André Palmade , and Jacques-François Thisse. *Discrete Choice Theory of Product Differentiation*. The MIT Press, 1st edition, October 1992. [45]

Aristei, David and Manuela Gallo. Interest rate pass-through in the euro area during the financial crisis: A multivariate regime-switching approach. *Journal of Policy Modeling*, 36(2):273–295, March-April 2014. [8]

Berger, Allen N., Asli Demirgüç-Kunt, Ross Levine, and Joseph G. Haubrich. Bank concentration and competition: An evolution in the making. *Journal of Money, Credit, and Banking*, 36(3): 433–451, June 2004. [7]

Berger, Philip G., Michael Minnis, and Andrew Sutherland. Commercial lending concentration and bank expertise: Evidence from borrower financial statements. February 2017. [5], [7], [45]

Bernanke, Ben and Mark Gertler. Agency costs, net worth, and business fluctuations. *The American Economic Review*, 79(1):14–31, March 1989. [6], [14]

Besanko, David and Anjan V. Thakor. Banking deregulation: Allocational consequences of relaxing entry barriers. *Journal of Banking and Finance*, 16(5):909–932, September 1992. [7]

Bhaskar, Venkataraman. The kinked demand curve: A game-theoretic approach. *International Journal of Industrial Organization*, 6(3):373–384, 1988. [23]

Borio, Claudio E.V. and Wilhelm Fritz. The response of short-term bank lending rates to policy rates: A cross-country perspective. BIS Working Paper No. 27, May 1995. [2]

Brunnermeier, Markus K. and Yuliy Sannikov. A macroeconomic model with a financial sector. *The American Economic Review*, 104(2):379–421, February 2014. [46]

Chamberlin, Edward Hastings. *The Theory of Monopolistic Competition*, volume 6. Harvard University Press, Cambridge, MA, 1933. [24]

Chiappori, Pierre-André, David Perez-Castrillo, and Thierry Verdier. Spatial competition in the banking system: Localization, cross subsidies and the regulation of deposit rates. *European Economic Review*, 39(5):889–918, May 1995. [8]

Cottarelli, Carlo and Angeliki Kourelis. Financial structure, banking lending rates and the transmission mechanism of monetary policy. *IMF Staff Papers*, 41(4):587–623, December 1994. [2], [8]

Demsetz, Rebecca S. and Philip E. Strahan. Diversification, size, and risk at bank holding companies. *Journal of Money, Credit and Banking*, 29(3):300–313, August 1997. [7]

Diamond, Douglas W. Financial intermediation and delegated monitoring. *The Review of Economic Studies*, 51(3):393–414, July 1984. [7], [17]

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl. The deposits channel of monetary policy. *Quarterly Journal of Economics*, Forthcoming, March 2017. [8]

Gârleanu, Nicolae, Stavros Panageas, and Jianfeng Yu. Financial entanglement: A theory of incomplete integration, leverage, crashes, and contagion. *The American Economic Review*, 105 (7):1979–2010, July 2015. [7]

Gerali, Andrea, Stefano Neri, Luca Sessa, and Federico M. Signoretti. Credit and banking in a dsge model of the euro area. *Journal of Money, Credit and Banking*, 42(Supplement 1):107–141, September 2010. [8]

Gertler, Mark and Nobuhiro Kiyotaki. *Financial Intermediation and Credit Policy in Business Cycle Analysis*, volume 3 of *Handbook of Monetary Economics*, chapter 11, pages 547–599. North-Holland, 525 B Street, Suite 1800, San Diego, CA 92101-4495, 2010. [6], [7]

Gigineishvili, Nikoloz. Determinants of interest rate pass-through: Do macroeconomic conditions and financial market structure matter? IMF Working Paper, July 2011. [2]

Gilchrist, Simon and Egon Zakrajšek. Credit spreads and business cycle fluctuations. *The American Economic Review*, 102(4):1692–1720, June 2012. [2], [52]

Granja, Joao, Gregor Matvos, and Amit Seru. Selling failed banks. *Journal of Finance*, Forthcoming, April 2015. [2]

Gu, Yiquan and Tobias Wenzel. A note on the excess entry theorem in spatial models with elastic demand. *International Journal of Industrial Organization*, 27(5):567–571, September 2009. [45]

Günter, Jochen H.F. Competition among banks and the pass-through of monetary policy. *Economic Modelling*, 28(4):1891–1901, 2011. [8]

Hall, Robert L. and Charles Johnston Hitch. Price theory and business behaviour. *Oxford Economic Papers*, (2):12–45, May 1939. [23]

Hannan, Timothy H. and Allen N Berger. The rigidity of prices: Evidence from the banking industry. *The American Economic Review*, 81(4):938–945, September 1991. [8]

Hauswald, Robert and Robert Marquez. Competition and strategic information acquisition in credit markets. *The Review of Financial Studies*, 19(3):967–1000, February 2006. [8]

He, Zhiguo and Arvind Krishnamurthy. Intermediary asset pricing. *The American Economic Review*, 103(2):732–770, April 2013. [46]

Holmstrom, Bengt and Jean Tirole. Financial intermediation, loanable funds, and the real sector. *The Quarterly Journal of Economics*, 112(3):663–691, August 1997. [6], [7]

Hristov, Nikolay, Oliver Hülsewig, and Timo Wollmershäuser. The interest rate pass-through in the euro area during the global financial crisis. *Journal of Banking & Finance*, 48:104–119, November 2014. [8]

Hülsewig, Oliver, Eric Mayer, and Timo Wollmershäuser. Bank behavior, incomplete interest rate pass-through, and the cost channel of monetary policy transmission. *Economic Modelling*, 26(6): 1310–1327, November 2009. [8]

Liang, J. Nellie and Stephen A. Rhoades. Asset diversification, firm risk, and risk-based capital requirements in banking. *Review of Industrial Organization*, 6(1):49–59, January 1991. [7]

Maskin, Eric and Jean Tirole. A theory of dynamic oliopoly, II: Price competition, kinked demand curves, and edgeworth cycles. *Econometrica*, 56(3):571–599, May 1988. [23]

Matsumura, Toshihiro and Makoto Okamura. A note on the excess entry theorem in spatial markets. *International Journal of Industrial Organization*, 24(5):1071–1076, September 2006. [45]

Matutes, Carmen and Xavier Vives. Competition for deposits, fragility, and insurance. *Journal of Financial Intermediation*, 5(2):184–216, April 1996. [7]

Matutes, Carmen and Xavier Vives. Imperfect competition, risk taking, and regulation in banking. *European Economic Review*, 44(1):1–34, January 2000. [7]

McAllister, Patrick H. and Douglas McManus. Resolving the scale efficiency puzzle in banking. *Journal of Banking & Finance*, 17(2-3):389–405, April 1993. [7]

Mojon, Benoît. Financial structure and the interest rate channel of ecb monetary policy. ECB Working paper series, Working Paper No. 40, November 2000. [2]

Neumark, David and Steve A. Sharpe. Market structure and the nature of price rigidity: Evidence from the market for consumer deposits. *The Quarterly Journal of Economics*, 107(2):657–680, May 1992. [8]

Paravisini, Daniel, Veronica Rappoport, and Philipp Schnabl. Specialization in bank lending: Evidence from exporting firms. November 2015. [7], [15], [45], [46]

Petersen, Mitchell A. and Raghuram G. Rajan. The beneftis of lending relationships: Evidence from small business data. *The Journal of Finance*, 49(1):3–37, March 1994. [22]

Ramakrishnan, Ram T. S. and Anjan V. Thakor. Information reliability and a theory of financial intermediation. *The Review of Economic Studies*, 51(3):415–432, July 1984. [14]

Reid, Gavin C. *The Kinked Demand Curve Analysis of Oligopoly: Theory and Evidence*. Edingbugh University Press, 1981. [23]

Robinson, Joan. *The Economics of Imperfect Competition*. St. Martin's Press, Inc., 175 Fifth Avenue, New York, N.Y. 10010, 2nd edition, 1969. [24]

Rothschild, Robert. A simple proof of sweezy's 'kinked-demand' conjecture. *Scottish Journal of Political Economy*, 39(1):69–75, February 1992. [23]

Salop, Steven C. Monopolistic competition with outside goods. *The Bell Journal of Economics*, 10 (1):141–156, Spring 1979. [7], [8], [24], [45]

Scharfstein, David and Adi Sunderam. Market power in mortgage lending and the transmission of monetary policy. August 2016. [8]

Scharler, Johann. Do bank-based financial systems reduce macroeconomic volatility by smoothing interest rates? *Journal of Macroeconomics*, 30(3):1207–1221, September 2008. [8]

Sen, Debapriya. The kinked demand curve revisited. *Economics Letters*, 84(1):99–105, July 2004. [23]

Sørensen, Christoffer Kok and Thomas Werner. Bank interest rate pass-through in the euro area: A cross country comparison. European Central Bank Working Paper Series, January 2006. [2]

Stein, Jeremy C. Information production and capital allocation: Decentralized versus hierarchical firms. *Journal of Finance*, 57(5):1891–1921, October 2002. [19]

Sussman, Oren and Joseph Zeira. Banking and development. Centre for Economic Policy Research Discussion Paper 1127, February 1995. [8]

Sweezy, Paul M. Demand under conditions of oligopoly. *Journal of Political Economy*, 47(4): 568–573, August 1939. [23]

Townsend, Robert M. Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293, October 1979. [12]

Leuvensteijn, Michielvan , Christoffer Kok Sørensen, Jacob Antoon Bikker, and Adrian A.R.J.M. Rixtelvan . Impact of bank competition on the interest rate pass-through in the euro area. Tjalling C. Koopmans Research Institute, Discussion Paper Series 08-08, March 2008. [2]

Vickrey, William S. *Microstatics.* Harcourt, Brace & World, New York, 1st edition, 1964. [45]

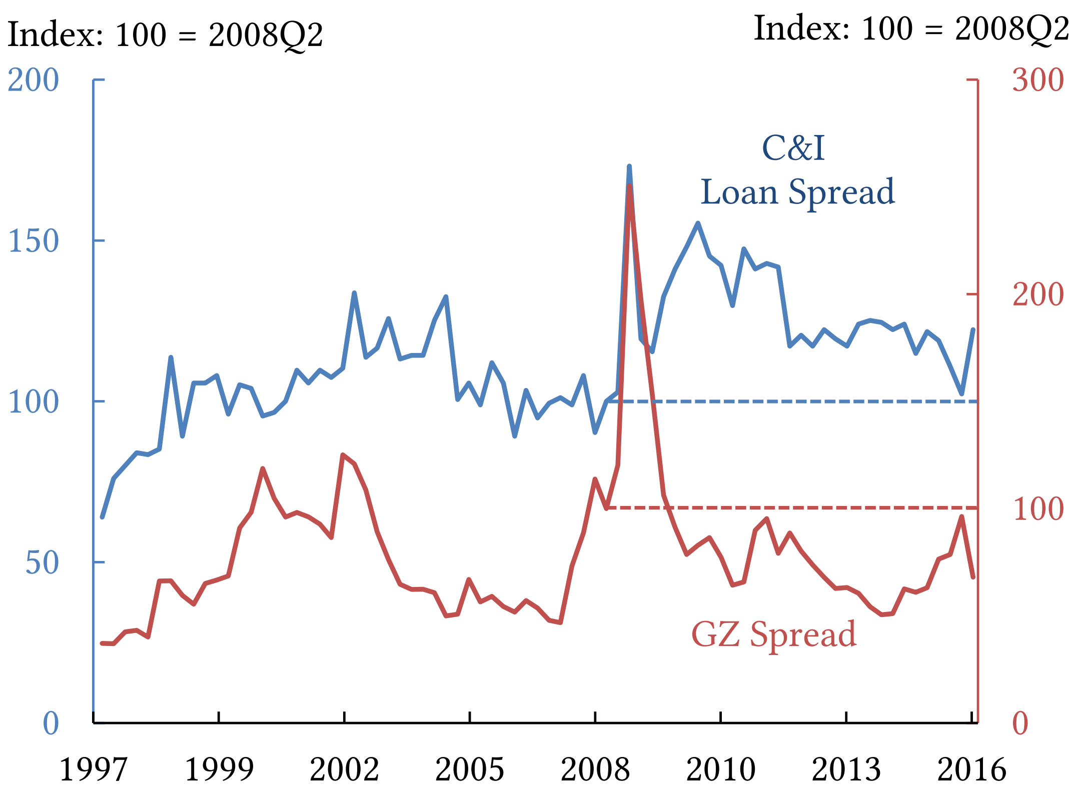
Winton, Andrew. Don't put all your eggs in one basket? diversification and specialization in lending. September 1999. [7]

A Appendix

A.1 Corporate Bond Spread

Figure 12 plots the C&I loan spread against the corporate bond spread calculated in Gilchrist and Zakrajšek (2012) in order to illustrate the sluggishness of the C&I loan spread to return to pre-crisis levels relative to the corporate bond spread.

Figure 12: C&I Loan and GZ Spreads



Notes: The C&I spread is the same as in Figure 1(a). The GZ spread from Gilchrist and Zakrajšek (2012) is an unweighted cross-sectional average of the spreads between the yields of senior unsecured bonds of a sample of U.S. non-financial firms and synthetic risk-free securities that match the cash flows of those bonds. Each time series is normalized to equal its value in 2008Q2. The dashed lines start from those values and extend to the end of the two series in the figure.

Sources: Board of Governors of the Federal Reserve System (C & I loan spread). Simon Gilchrist (GZ spread). C&I loan data retrieved from FRED, Federal Reserve Bank of St. Louis. GZ spread retrieved from Simon Gilchrist's website: `http://people.bu.edu/sgilchri/Data/data.htm`.

A.2 Proof of Lemma (1)

The success probability of a project at location j from (1) is

$$\widetilde{\Pr}_{t^+}\left(H|j,\tilde{u}_{t^+}\right)=\frac{1}{2}\Big(1+\cos\Big(2\pi\left(j+\tilde{u}_{t^+}\right)\Big)\Big)$$

The success probability can be treated as a transformation of a uniform random variable $x_j=2\pi\left(j+\tilde{u}_{t^+}\right)$, which has support $\left[2\pi j,2\pi\left(j+1\right)\right]$. For ease of notation, define

$$Y_j\equiv\frac{1}{2}\left(1+\cos\left(x_j\right)\right).$$

For $y\in\left[0,1\right]$, the equation

$$y=\frac{1}{2}\left(1+\cos\left(x_j\right)\right)$$

has two solutions in $\left[2\pi j,2\pi\left(j+1\right)\right]$. Therefore, the transformed density is

$$\begin{aligned}f_{Y_j}\left(y\right)&=2\times\frac{2}{\sqrt{1-y^2}}\times\frac{1}{2\pi}\\&=\frac{2}{\pi\sqrt{1-y^2}},\end{aligned}$$

for $y\in\left[0,1\right]$ and zero otherwise for all j. The leading factor of 2 accounts for the two solutions in the support. Thus, the density of the success probability is the same at all locations.

The expected probability of success for a single project is $\frac{1}{2}$. To see this, integrate the success probability over the unit interval since $\tilde{u}_{t^+}\sim U\left[0,1\right]$ to get

$$\begin{aligned}E_t\left[\widetilde{\Pr}_{t^+}\left(H|j,\tilde{u}_{t^+}\right)\right]&=\int_0^1\frac{1}{2}\Big(1+\cos\Big(2\pi\left(j+\tilde{u}_{t^+}\right)\Big)\Big)\,d\tilde{u}_{t^+}\\&=\frac{1}{2}+\frac{1}{2\pi}\left[\sin\left(2\pi\left(j+1\right)\right)-\sin\left(2\pi j\right)\right]\\&=\frac{1}{2}+\frac{1}{2\pi}\left[\sin\left(2\pi j\right)\cos\left(2\pi\right)+\cos\left(2\pi j\right)\sin\left(2\pi\right)-\sin\left(2\pi j\right)\right]\\&=\frac{1}{2}+\frac{1}{2\pi}\left[\sin\left(2\pi j\right)-\sin\left(2\pi j\right)\right]\\&=\frac{1}{2},\end{aligned}$$

where the third equality follows from the sum-difference formula for sine.

The variance of the success probability is

$$\begin{aligned}
\sigma^2\left[\widetilde{\Pr}_{t+}\left(H|j,\tilde{u}_{t+}\right)\right] &= E_t\left[\widetilde{\Pr}_{t+}\left(H|j,\tilde{u}_{t+}\right)^2\right] - E_t\left[\widetilde{\Pr}_{t+}\left(H|j,\tilde{u}_{t+}\right)\right]^2 \\
&= \frac{1}{4}E_t\left[\left(1+\cos\left(2\pi\left(j+\tilde{u}_{t+}\right)\right)\right)^2\right] - \frac{1}{4} \\
&= \frac{1}{4}\int_0^1 \cos^2\left(2\pi\left(j+\tilde{u}_{t+}\right)\right) d\tilde{u}_{t+}.
\end{aligned}$$

Using the half-angle trigonometric formula $\cos^2 u = \frac{1+\cos(2u)}{2}$, the variance can be written as

$$\begin{aligned}
\sigma^2\left[\widetilde{\Pr}_{t+}\left(H|j,\tilde{u}_{t+}\right)\right] &= \frac{1}{8}\int_0^1 \left[1+\cos 4\pi\left(j+\tilde{u}_{t+}\right)\right] d\tilde{u}_{t+} \\
&= \frac{1}{8}\left[1+\frac{1}{4\pi}\left[\sin\left(4\pi\left(j+1\right)\right) - \sin\left(4\pi j\right)\right]\right] \\
&= \frac{1}{8}.
\end{aligned}$$

Finally, the covariance of success probabilities between projects located at j and k on the circle is

$$\begin{aligned}
\mathrm{cov}\left(\widetilde{\Pr}_{t+}\left(H|j,\tilde{u}_{t+}\right), \widetilde{\Pr}_{t+}\left(H|k,\tilde{u}_{t+}\right)\right) &= E_t\left[\widetilde{\Pr}_{t+}\left(H|j,\tilde{u}_{t+}\right)\widetilde{\Pr}_{t+}\left(H|k,\tilde{u}_{t+}\right)\right] - \frac{1}{4} \\
&= \frac{1}{4}E_t\left[\cos\left(2\pi\left(j+\tilde{u}_{t+}\right)\cos\left(2\pi\left(k+\tilde{u}_{t+}\right)\right)\right]\right].
\end{aligned}$$

Using the cosine product formula $\cos(a)\cos(b) = \frac{1}{2}\left[\cos(a+b)+\cos(a-b)\right]$ gives

$$\begin{aligned}
\mathrm{cov}\left(\widetilde{\Pr}_{t+}\left(H|j,\tilde{u}_{t+}\right), \widetilde{\Pr}_{t+}\left(H|k,\tilde{u}_{t+}\right)\right) &= \frac{1}{8}\cos\left(2\pi\left(j-k\right)\right) + E_t\left[\cos\left(2\pi\left(j+k+2\tilde{u}_{t+}\right)\right)\right] \\
&= \frac{1}{8}\cos\left(2\pi\left(j-k\right)\right) + \frac{1}{8}\int_0^1 \cos\left(2\pi\left(j+k+2\tilde{u}_{t+}\right)\right) d\tilde{u}_{t+} \\
&= \frac{1}{8}\cos\left(2\pi\left(j-k\right)\right).
\end{aligned}$$

A.3 Proof of Lemma (2)

The repayment rate of bank i conditional on its chosen arc length Δ_t^i is

$$\widetilde{\Pr}_{t+}\left(R_{L,t}^i|\Delta_t^i,\tilde{u}_{t+}\right) = \frac{1}{2} + \frac{\sin\left(\pi\Delta_t^i\right)}{\Delta_t^i}\frac{\cos\left(2\pi\left(i+\tilde{u}_{t+}\right)\right)}{2\pi}.$$

As $\Delta_t^i \downarrow 0$, the bank's lending arc reduces to its home location alone. Apply L'Hôpital's rule to get

$$\begin{aligned}
\lim_{\Delta_i \downarrow 0} \widetilde{\Pr}_{t^+}\left(R_{L,t}^i | \Delta_t^i, \tilde{u}_{t^+}\right) &= \lim_{\Delta_t^i \downarrow 0} \left\{ \frac{1}{2} + \pi \cos\left(\pi \Delta_t^i\right) \frac{\cos\Big(2\pi\left(i + \tilde{u}_{t^+}\right)\Big)}{2\pi} \right\} \\
&= \frac{1}{2}\Big(1 + \cos\Big(2\pi\left(i + \tilde{u}_{t^+}\right)\Big)\Big) \\
&= \widetilde{\Pr}_{t^+}\left(H | j, \tilde{u}_{t^+}\right),
\end{aligned}$$

matching the probability of a single project generating the high physical capital return from (1).

Next, the expected repayment rate of a bank's portfolio is always $\frac{1}{2}$, no matter its arc length Δ_t^i. Integrate the repayment rate over the unit interval to get

$$\begin{aligned}
E_t\left[\widetilde{\Pr}_{t^+}\left(R_{L,t}^i | \Delta_t^i, \tilde{u}_{t^+}\right)\right] &= \int_0^1 \left[\frac{1}{2} + \frac{\sin\left(\pi \Delta_t^i\right)}{\Delta_t^i} \frac{\cos\Big(2\pi\left(i + \tilde{u}\right)\Big)}{2\pi}\right] d\tilde{u}_{t^+} \\
&= \frac{1}{2} + \frac{\sin\left(\pi \Delta_t^i\right)}{\left(2\pi\right)^2 \Delta_t^i}\left[\sin\left(2\pi\left(i+1\right)\right) - \sin\left(2\pi i\right)\right] \\
&= \frac{1}{2}.
\end{aligned}$$

The variance of the repayment rate is

$$\begin{aligned}
\sigma^2\left[\widetilde{\Pr}_{t^+}\left(R_{L,t}^i | \Delta_t^i, \tilde{u}_{t^+}\right)\right] &= \int_0^1 \left(\widetilde{\Pr}_{t^+}\left(R_{L,t}^i | \Delta_t^i, \tilde{u}_{t^+}\right) - E_t\left[\widetilde{\Pr}_{t^+}\left(R_{L,t}^i | \Delta_t^i, \tilde{u}_{t^+}\right)\right]\right)^2 d\tilde{u}_{t^+} \\
&= \int_0^1 \left(\widetilde{\Pr}_{t^+}\left(R_{L,t}^i | \Delta_t^i, \tilde{u}_{t^+}\right) - \frac{1}{2}\right)^2 d\tilde{u}_{t^+} \\
&= \int_0^1 \left(\frac{\sin\left(\pi \Delta_t^i\right)}{\Delta_t^i} \frac{\cos\Big(2\pi\left(i + d\tilde{u}_{t^+}\right)\Big)}{2\pi}\right)^2 d\tilde{u}_{t^+} \\
&= \left[\frac{\sin\left(\pi \Delta_t^i\right)}{2\pi \Delta_t^i}\right]^2 \int_0^1 \cos^2\Big(2\pi\left(i + \tilde{u}_{t^+}\right)\Big) d\tilde{u}_{t^+}
\end{aligned}$$

Using the half-angle formula, the variance can be written as

$$\begin{aligned}
\sigma^2\left[\widetilde{\Pr}_{t^+}\left(R_{L,t}^i | \Delta_t^i, \tilde{u}_{t^+}\right)\right] &= \frac{1}{2}\left[\frac{\sin\left(\pi \Delta_t^i\right)}{2\pi \Delta_t^i}\right]^2 \int_0^1 \left[1 + \cos\left(4\pi\left(i + \tilde{u}_{t^+}\right)\right)\right] d\tilde{u}_{t^+} \\
&= \frac{1}{2}\left[\frac{\sin\left(\pi \Delta_t^i\right)}{2\pi \Delta_t^i}\right]^2 + \frac{1}{8\pi}\left[\frac{\sin\left(\pi \Delta_t^i\right)}{2\pi \Delta_t^i}\right]^2 \left[\sin\left(4\pi\left(i+1\right)\right) - \sin\left(4\pi i\right)\right] \\
&= \frac{1}{2}\left[\frac{\sin\left(\pi \Delta_t^i\right)}{2\pi \Delta_t^i}\right]^2, \qquad (33)
\end{aligned}$$

where the third equality follows after using the sum-difference formula like before.

The variance of the repayment rate is strictly decreasing for $\Delta_t^i \in (0,1)$. Taking the first derivative of (33) with respect to Δ_t^i gives

$$\begin{aligned}
\frac{\partial\sigma^2\left[\widetilde{\Pr}_{t+}\left(R_{L,t}^i|\Delta_t^i,\tilde{u}_{t+}\right)\right]}{\partial\Delta_t^i} &= \left[\frac{\sin\left(\pi\Delta_t^i\right)}{2\pi\Delta_t^i}\right]\frac{\partial\frac{\sin\left(\pi\Delta_t^i\right)}{2\pi\Delta_t^i}}{\partial\Delta_t^i} \\
&= \left[\frac{\sin\left(\pi\Delta_t^i\right)}{2\pi\Delta_t^i}\right]\left[\frac{\pi\Delta_t^i\cos\left(\pi\Delta_t^i\right)-\sin\left(\pi\Delta_t^i\right)}{2\pi\left(\Delta_t^i\right)^2}\right] \\
&= \frac{\pi\Delta_t^i\cos\left(\pi\Delta_t^i\right)\sin\left(\pi\Delta_t^i\right)-\sin^2\left(\pi\Delta_t^i\right)}{4\pi^2\left(\Delta_t^i\right)^3}. \qquad (34)
\end{aligned}$$

The sign of $\frac{\partial\sigma^2\left[\widetilde{\Pr}_{t+}\left(H|\Delta_t^i\right)\right]}{\partial\Delta_t^i}$ is determined by the numerator of (34), as the denominator is always positive. The variance is non-increasing in Δ_t^i if

$$\pi\Delta_t^i\cos\left(\pi\Delta_t^i\right)\sin\left(\pi\Delta_t^i\right)-\sin^2\left(\pi\Delta_t^i\right)\leq 0.$$

For $\Delta_t^i=0$ or $\Delta_t^i=1$, the numerator is zero, so the above inequality holds. For $\Delta_t^i\in(0,1)$, $\sin\left(\pi\Delta_t^i\right)\neq 0$, so the expression can be written as

$$\pi\Delta_t^i\cos\left(\pi\Delta_t^i\right)\leq\sin\left(\pi\Delta_t^i\right).$$

Perform a change of variable $\theta=\pi\Delta_t^i$. The aim is to show

$$\theta\cos\theta<\sin\theta$$

over the domain $\theta\in(0,\pi)$. Over the upper-half of the interval $\theta\in\left[\frac{\pi}{2},\pi\right)$, since $\cos\theta\leq 0$ and $\sin\theta>0$, the relation holds. Now define the function

$$f(\theta)\equiv\sin\theta-\theta\cos\theta.$$

Note that $\lim_{\theta\downarrow 0} f(\theta)=0$ and $f'(\theta)=\theta\sin\theta>0$ for $\theta\in\left(0,\frac{\pi}{2}\right)$. Therefore, $f(\theta)>0$ over the lower half of the interval. This proves the numerator of (34) is negative for $\Delta_t^i\in(0,1)$ and that the variance of the repayment rate is non-increasing for $\Delta_i\in[0,1]$ and strictly decreasing over the open unit interval.

Finally, as $\Delta_t^i \uparrow 1$, the repayment rate has the following limit:

$$\begin{aligned}\lim_{\Delta_t^i \uparrow 1} \widetilde{\Pr}_{t+}\left(R_{L,t}^i | \Delta_t^i, \tilde{u}_{t+}\right) &= \frac{1}{2} + \sin(\pi) \frac{\cos\left(2\pi\left(i + \tilde{u}_{t+}\right)\right)}{2\pi} \\ &= \frac{1}{2}.\end{aligned}$$

Thus, the repayment rate becomes a constant $\frac{1}{2}$, no matter the realization of the random variable $\tilde{u}_{t+}$.

A.4 Proof of Proposition (1) and Lemma (3)

In a competitive equilibrium, the first order condition for optimality is

$$\frac{1}{2}\left[R_{L,t}^i + \hat{q}_{t+1}\underline{\kappa}\right] + \frac{1}{2}\Delta_t^i \left(\frac{dR_{L,t}^i}{d\Delta_t^i}\right) = \frac{1}{2}\hat{q}_{t+1} g'\left(\Delta_t^i\right) + \frac{\partial FC_t^i}{\partial \Delta_t^i}.$$

Using (8) and a binding deposit constraint (5), the marginal financing cost function,

$$\frac{\partial FC_t^i}{\partial \Delta_t^i} = \hat{R}_{E,t+1}^C + \left(1 - \frac{\hat{R}_{E,t+1}^C}{R_{f,t}}\right) \frac{d\Pi_{\min}^{i,t}}{d\Delta_t^i}.$$

Computing $\frac{d\Pi_{\min}^{i,t}}{d\Delta_t^i}$ gives for the marginal financing cost function

$$\begin{aligned}\frac{\partial FC_t^i}{\partial \Delta_t^i} = & \left(1 - \frac{\hat{R}_{E,t+1}^C}{R_{f,t}}\right)\left(\hat{q}_{t+1}\underline{\kappa} - \hat{q}_{t+1} g'\left(\Delta_t^i\right)\right) - \left(1 - \frac{\hat{R}_{E,t+1}^C}{R_{f,t}}\right) \mathrm{Pr}_{\min}\left(\Delta_t^i\right)\left(\hat{q}_{t+1}\underline{\kappa} - \hat{q}_{t+1} g'\left(\Delta_t^i\right)\right) \\ & + \left(1 - \frac{\hat{R}_{E,t+1}^C}{R_{f,t}}\right) \mathrm{Pr}_{\min}\left(\Delta_t^i\right)\left[R_{L,t}^i + \Delta_t^i \left(\frac{dR_{L,t}^i}{d\Delta_t^i}\right)\right] \\ & + \left(1 - \frac{\hat{R}_{E,t+1}^C}{R_{f,t}}\right) \frac{d\,\mathrm{Pr}_{\min}\left(\Delta_t^i\right)}{d\Delta_t^i}\left[\left(R_{L,t}^i - \hat{q}_{t+1}\underline{\kappa}\right)\Delta_t^i + \hat{q}_{t+1} g\left(\Delta_t^i\right)\right] \\ & + \hat{R}_{E,t+1}^C\end{aligned}$$

Substituting the slope of the competitive demand curve $\frac{dR_{L,t}^i}{d\Delta_t^i} = -2c$ and re-arranging terms gives

$$\begin{aligned}\frac{\partial FC_t^i}{\partial \Delta_t^i} = & \left(1 - \frac{\hat{R}_{E,t+1}^C}{R_{f,t}}\right)\left[\mathrm{Pr}_{\min}\left(\Delta_t^i\right)\left(R_{L,t}^i - 2c\Delta_t^i\right) + \left(1 - \mathrm{Pr}_{\min}\left(\Delta_t^i\right)\right)\left(\hat{q}_{t+1}\underline{\kappa} - \hat{q}_{t+1} g'\left(\Delta_t^i\right)\right)\right] \\ & + \left(1 - \frac{\hat{R}_{E,t+1}^C}{R_{f,t}}\right) \frac{d\,\mathrm{Pr}_{\min}\left(\Delta_t^i\right)}{d\Delta_t^i}\left[\left(R_{L,t}^i - \hat{q}_{t+1}\underline{\kappa}\right)\Delta_t^i + \hat{q}_{t+1} g\left(\Delta_t^i\right)\right] + \hat{R}_{E,t+1}^C\end{aligned}$$

The first line in the above expression is the marginal financing cost savings from greater diversification and larger debt capacity, holding the minimum probability fixed. The first term of the second line is the cost savings from a higher minimum repayment rate after a marginal increase in the loan portfolio breadth Δ_t^i. The bank gets repaid on more of its loans, recovers less of the low physical capital returns and saves on liquidation costs from fewer projects in default. The final term $\hat{R}^C_{E,t+1}$ of the second line is the marginal cost of equity financing.

The marginal financing cost function can be conveniently represented by separating terms involving $R^i_{L,t}$. Doing so gives

$$\begin{aligned}
\frac{\partial FC_t^i}{\partial \Delta_t^i} = & -\left(\frac{\hat{R}^C_{E,t+1}}{R_{f,t}} - 1\right)\left[\Pr_{\min}\left(\Delta_t^i\right) + \frac{d\Pr_{\min}\left(\Delta_t^i\right)}{d\Delta_t^i}\Delta_t^i\right] R^i_{L,t} \\
& + \left(\frac{\hat{R}^C_{E,t+1}}{R_{f,t}} - 1\right)\left[\Pr_{\min}\left(\Delta_t^i\right) 2c\Delta_t^i + \left(1 - \Pr_{\min}\left(\Delta_t^i\right)\right)\left(\hat{q}_{t+1} g'\left(\Delta_t^i\right) - \hat{q}_{t+1}\underline{\kappa}\right)\right] \\
& + \left(\frac{\hat{R}^C_{E,t+1}}{R_{f,t}} - 1\right)\frac{d\Pr_{\min}\left(\Delta_t^i\right)}{d\Delta_t^i}\left(\hat{q}_{t+1}\underline{\kappa}\Delta_t^i - \hat{q}_{t+1} g\left(\Delta_t^i\right)\right) \\
& + \hat{R}^C_{E,t+1}
\end{aligned}$$

Define the functions

$$\begin{aligned}
\phi\left(\Delta_t^i\right) \equiv\ & \Pr_{\min}\left(\Delta_t^i\right) 2c\Delta_t^i + \left(1 - \Pr_{\min}\left(\Delta_t^i\right)\right)\left(\hat{q}_{t+1} g'\left(\Delta_t^i\right) - \hat{q}_{t+1}\underline{\kappa}\right) \\
& + \frac{d\Pr_{\min}\left(\Delta_t^i\right)}{d\Delta_t^i}\left(\hat{q}_{t+1}\underline{\kappa}\Delta_t^i - \hat{q}_{t+1} g\left(\Delta_t^i\right)\right) \qquad (35)
\end{aligned}$$

$$\psi\left(\Delta_t^i\right) \equiv \Pr_{\min}\left(\Delta_t^i\right) + \frac{d\Pr_{\min}\left(\Delta_t^i\right)}{d\Delta_t^i}\Delta_t^i \qquad (36)$$

and re-write $\frac{\partial FC_t^i}{\partial \Delta_t^i}$ as

$$\frac{\partial FC_t^i}{\partial \Delta_t^i} = \hat{R}^C_{E,t+1} + \left(\frac{\hat{R}^C_{E,t+1}}{R_{f,t}} - 1\right)\left[\phi\left(\Delta_t^i\right) - \psi\left(\Delta_t^i\right) R^i_{L,t}\right]$$

The function $\psi > 0$. Provided $\hat{q}_{t+1}\underline{\kappa}$ is not too large, then also $\phi > 0$. The terms in ϕ are those that increase the marginal cost of financing for the bank, while those in ψ decrease the marginal cost of financing.

Using the representation of FC' in the last expression, the optimality condition becomes

$$\frac{1}{2}\left[R^i_{L,t} + \hat{q}_{t+1}\underline{\kappa}\right] - c\Delta^i_t = \frac{1}{2}\hat{q}_{t+1}g'\left(\Delta^i_t\right) + \hat{R}^C_{E,t+1} + \left(\frac{\hat{R}^C_{E,t+1}}{R_{f,t}} - 1\right)\left[\phi\left(\Delta^i_t\right) - \psi\left(\Delta^i_t\right)R^i_{L,t}\right]$$

Solving for $R^i_{L,t}$ and using the equilibrium condition $\Delta^i_t = \frac{1}{N^C_t}$ gives for the competitive equilibrium lending rate

$$R^C_{L,t} = \frac{\hat{R}^C_{E,t+1} + \frac{1}{2}\left[\hat{q}_{t+1}g'\left(\frac{1}{N^C_t}\right) - \hat{q}_{t+1}\underline{\kappa} + \frac{2c}{N^C_t}\right] + \left(\frac{\hat{R}^C_{E,t+1}}{R_{f,t}} - 1\right)\phi\left(\frac{1}{N^C_t}\right)}{\frac{1}{2} + \left(\frac{\hat{R}^C_{E,t+1}}{R_{f,t}} - 1\right)\psi\left(\frac{1}{N^C_t}\right)}.$$

In a kinked equilibrium, the first order condition for optimality does not hold with equality, so the lending rate is instead the monopoly demand curve. Solving (9) for the lending rate and setting $\Delta^i_t = \frac{1}{N^K_t}$ gives (23).

A.5 Derivation of Asset Prices $R_{f,t}$ and $\hat{R}_{E,t+1}$, unit investment scale

Here I derive the interest rate $R_{f,t}$ and expected equity return $\hat{R}_{E,t+1}$ across the kinked and competitive equilibria when the scale of investment is a single unit.

The derivation of the interest rate does not depend on the type of equilibrium. Combining deposit and equity market clearing delivers the interest rate. Equate deposit supply s_t from (15) with deposit demand from (16), and combine with equity market clearing using equity supply $w_t L^b$ and bank equity demand from (17) to get the real interest rate. The after-tax interest rate is

$$R_{f,t}\left(1-\tau\right) = \frac{\hat{w}_{t+1}}{\beta w_t - \frac{1+\beta(1-\tau)}{L_d}\left(1 + fN_t - w_t L^b\right)}. \tag{37}$$

If workers expected a downward sloping (super-martingale) income profile, whereby $\hat{w}_{t+1} < w_t$, they are encouraged to save, pushing down the interest rate. The effect of a change in the tax rate τ on the interest rate is ambiguous and depends on the response of entry N_t.

Bank demand for deposits is captured by the term $1 + fN_t - w_t L^b$, which is aggregate demand for financing $(1 + fN_t)$ in excess of the supply of equity capital $\left(w_t L^b\right)$. Bank debt and equity are substitutes in the economy, so any bank financing not fulfilled by equity increases the demand for deposits. Higher deposit demand puts upward pressure on $R_{f,t}$.

The derivation of the cost of equity does depend on the type of equilibrium. Depending on the equilibrium, the expected equity rate takes a different form. The derivation uses the equity market clearing and zero profit conditions, reprinted here:

$$
\begin{aligned}
w_t L^b &= 1 + fN_t - N_t \left(\frac{\Pi^{i,t}_{\min}}{R_{f,t}} \right), \\
\frac{1}{2}\left(R_{L,t} + \hat{q}_{t+1}\underline{\kappa}\right) &= \frac{1}{2}\frac{\hat{q}_{t+1}\gamma}{N_t} - \left(\frac{\hat{R}_{E,t+1}}{R_{f,t}} - 1 \right) N_t \Pi^{i,t}_{\min} + \hat{R}_{E,t+1}\left(1 + fN_t\right).
\end{aligned}
$$

A.5.1 Kinked

Substitute the equity market clearing condition into the zero profit condition:

$$
\frac{1}{2}\left(R_{L,t} + \hat{q}_{t+1}\underline{\kappa}\right) = \frac{1}{2}\frac{\hat{q}_{t+1}\gamma}{N_t} + \hat{R}_{E,t+1} w_t L^b + R_{f,t}\left(1 + fN_t - w_t L^b\right). \tag{38}
$$

Next, substitute the kinked lending rate from (23) into (38) and re-arrange to get

$$
\hat{R}_{E,t+1} = \frac{\frac{1}{2}\left(\hat{q}_{t+1}\left(\overline{\kappa} + \underline{\kappa}\right) - 2 w_t L^e - \frac{(c + \hat{q}_{t+1}\gamma)}{N_t} \right) - R_{f,t}\left(1 + fN_t - w_t L^b\right)}{w_t L^b}.
$$

The expression for the equity return in the kinked equilibrium is easily interpreted. The denominator is the supply of equity, which lowers the cost of bank equity capital after any expansion. The numerator captures the net demand for equity. The first term is the average operating profit (pre-financing costs) for banks and represents the “gross” demand for equity, holding fixed the demand for debt. An expansion in the operating profit of a bank, entices other banks to enter the lending market. More banks on the circle means the gross demand for equity capital increases, which raises $\hat{R}_{E,t+1}$. The second term of the numerator captures the aggregate demand for bank debt as a substitute for equity. A higher demand for debt relaxes the demand for equity, lowering the cost of equity capital.

A.5.2 Competitive

Only the equity market clearing condition is needed. Re-arrange the condition to get:

$$
R_{f,t}\left(1 + fN_t - w_t L^b\right) = N_t \Pi^{i,t}_{\min}. \tag{39}
$$

The right-hand-side of (39) is

$$
N_t \Pi^{i,t}_{\min} = \Pr_{\min}\left(\frac{1}{N_t}\right) R_{L,t} + \left(1 - \Pr_{\min}\left(\frac{1}{N_t}\right)\right)\left(\hat{q}_{t+1}\left(\underline{\kappa} - \frac{\gamma}{N_t}\right)\right).
$$

Substitute the above expression into (39) and isolate the lending rate to get

$$R_{L,t} = \frac{R_{f,t}\left(1 + fN_t - w_t L^b\right) - \left(1 - \Pr_{\min}\left(\frac{1}{N_t}\right)\right)\left(\hat{q}_{t+1}\left(\underline{\kappa} - \frac{\gamma}{N_t}\right)\right)}{\Pr_{\min}\left(\frac{1}{N_t}\right)}. \tag{40}$$

Let the right-hand-side of (40) be defined as $\mu\left(\frac{1}{N_t}\right)$. Substitute the competitive lending rate from (22) into (40), and solve for the expected equity rate to get

$$\hat{R}_{E,t+1} = \frac{\mu\left(\frac{1}{N_t}\right)\left(\frac{1}{2} - \psi\left(\frac{1}{N_t}\right)\right) + \phi\left(\frac{1}{N_t}\right) - \frac{1}{2}\left[\hat{q}_{t+1}g'\left(\frac{1}{N_t}\right) - \hat{q}_{t+1}\underline{\kappa} + \frac{2c}{N_t}\right]}{1 + \frac{\phi\left(\frac{1}{N_t}\right) - \psi\left(\frac{1}{N_t}\right)\mu\left(\frac{1}{N_t}\right)}{R_{f,t}}}. \tag{41}$$

A.5.3 Monopoly

I present the expected return on equity in the monopoly case because it is used in the conditions that determine which equilibrium an economy is in. The monopoly expected return on equity is the same as (41), except that the term $2c$ is replaced with c. This change also affects the function ϕ, given in (35), where $2c$ is also replaced with c.

A.6 Proof of Proposition (2)

Since the proposition applies in steady state, I drop all t notation for simpler notation. I first show how a change in the tax rate affects the interest rate.

The total derivative of the interest rate to the tax rate parameter consists of two components:

$$\frac{dR_f(\tau, N)}{d\tau} = \frac{\partial R_f}{\partial \tau} + \frac{\partial R_f}{\partial N}\frac{dN}{d\tau}.$$

The first component is the direct sensitivity of the interest rate to the tax rate, while the second component consists of the sensitivity of the interest rate to the number of banks and the sensitivity of bank entry to the tax rate.

I next establish a condition under which $\frac{\partial R_f}{\partial \tau} < 0$. Taking the partial derivative gives

$$\frac{\partial R_f}{\partial \tau} = \frac{-\left[\beta w - \frac{1+\beta(1-\tau)}{L_d}\left(1 + fN - wL^b\right)\right]\left(\frac{w}{(1-\tau)^2}\right) + \left(\frac{w}{1-\tau}\right)\left(\frac{\beta}{L_d}\right)\left(1 + fN - wL^b\right)}{\left[\beta w - \frac{1+\beta(1-\tau)}{L_d}\left(1 + fN - wL^b\right)\right]^2}.$$

The sign depends on the numerator only. Re-arranging reveals that the numerator is negative provided

$$\left(1 + fN - wL^b\right)\left(1 + 2\beta\left(1 - \tau\right)\right) > \beta w L^d.$$

This condition can hold for sufficiently low depositor labor endowment L^d.

Next I show a change to the tax rate affects the kinked lending rate. The kinked lending rate is a function of the tax rate only through the number of banks N. Therefore, the sensitivity of the lending rate to the tax rate can be decomposed as

$$\frac{dR_L}{d\tau} = \frac{\partial R_L}{\partial N}\frac{dN}{d\tau}.$$

The sign of the sensitivity of the number of banks to the tax rate $\operatorname{sgn}\left(\frac{dN}{d\tau}\right)$ is immaterial to the proof of the proposition. I need only show that *any* change to the tax rate τ that lowers the interest rate simultaneously encourages bank entry and increases the lending rate. Therefore, for the proof, it is sufficient to show that

$$\frac{\partial R_f}{\partial N} < 0,$$
$$\frac{\partial R_L}{\partial N} > 0.$$

The sensitivity of the interest rate to the number of banks is given by

$$\frac{\partial R_f}{\partial N} = \frac{w\left(1+\beta\left(1-\tau\right)\right)f - \frac{\partial w}{\partial N}\left(1+\beta\left(1-\tau\right)\right)\left[1+fN+w\left(1-L^b\right)\right]}{\left(1-\tau\right)\left[\beta w - \frac{1+\beta(1-\tau)}{L_d}\left(1+fN-wL^b\right)\right]^2}.$$

The sign of the derivative is determined by the sign of the numerator. The sensitivity of the wage to the number of banks is given by

$$\frac{\partial w}{\partial N} = \alpha\left(\frac{\gamma}{N^3}\right)\left(\frac{w}{K}\right),$$

where $K = \frac{1}{2}\left(\overline{\kappa}+\underline{\kappa}\right) - \frac{\gamma/2}{N^2}$ is the steady state capital stock. Substituting this quantity and re-arranging terms in the numerator reveals that $\frac{\partial R_f}{\partial N} < 0$ provided

$$fN\left(N^2K-\alpha\gamma\right) < \alpha\gamma\left(1+w\left(1-L^b\right)\right),$$

which holds for sufficiently small f.

Finally, the sensitivity of the lending rate to the number of banks is given by

$$\frac{\partial R_L}{\partial N} = \overline{\kappa}\frac{\partial q}{\partial N} - 2L^e\frac{\partial w}{\partial N} + \frac{c}{N^2}. \tag{42}$$

The sensitivity of the physical capital price to the number of banks is

$$\frac{\partial q}{\partial N} = -\left(1-\alpha\right)\left(\frac{\gamma}{N^3}\right)\left(\frac{q}{K}\right).$$

Substitute the two factor price sensitivities into (42) to get

$$\frac{\partial R_L}{\partial N} = -\left(\frac{\gamma}{N^3}\right)\left[\overline{\kappa}\left(1-\alpha\right)\left(\frac{q}{K}\right) + 2L^e\alpha\left(\frac{w}{K}\right)\right] + \frac{c}{N^2}.$$

The lending rate increases with the number of banks so long as

$$c > \frac{\gamma}{N}\left[\overline{\kappa}\left(1-\alpha\right)\left(\frac{q}{K}\right) + 2L^e\alpha\left(\frac{w}{K}\right)\right],$$

which holds if the factor-price-to-physical-capital ratios are not too high.

A.7 Determining the equilibrium

A.7.1 The conditions

I present the conditions that separate equilibria, then re-express those conditions in terms of the expected return on equity and aggregate bank net worth.

Given a vector of model parameters θ, productivity shock A_t, and state variable K_t, I determine the type of equilibrium of the economy in the unit scale investment case using a set of conditions. A necessary and sufficient condition for a monopoly equilibrium is that the optimal lending rate $R^M_{L,t}$ is on the monopoly demand curve. Mathematically that condition is

$$R^M_{L,t} = \hat{q}^M_{t+1}\overline{\kappa} - 2w_tL^e - \frac{c}{N^M_t}. \tag{43}$$

A necessary and sufficient condition for a competitive equilibrium is that the competitive lending rate is below the monopoly demand curve, so that

$$R^C_{L,t} \leq \hat{q}^C_{t+1}\overline{\kappa} - 2w_tL^e - \frac{c}{N^C_t}. \tag{44}$$

Finally, for a kinked equilibrium, it must be that expected monopoly profits, when the monopoly lending rate is replaced with the monopoly demand curve, as in (43), are positive. This condition distinguishes a kinked from monopoly equilibrium. Mathematically the condition is

$$\Pi|_{R^M_{L,t}=\hat{q}^M_{t+1}\overline{\kappa}-2w_tL^e-\frac{c}{N^M_t}} > 0, \tag{45}$$

where Π is the bank profit function in (7), and the vertical line represents “conditional on.” The inequality in (45), however, is necessary, but not sufficient, for a kinked equilibrium. The inequality also holds under a competitive equilibrium. Therefore, a necessary and sufficient condition for a kinked equilibrium is (45) and the failure of (44).

Finally, if the conditions fail for all equilibria, then no equilibrium exists under the parameter vector θ, productivity shock A_t, and state variable K_t.

A.7.2 Re-expressing the conditions

The conditions that define the equilibria can be expressed in terms of the expected return on equity $\hat{R}_{E,t+1}$. This way, bank net worth $E_t = w_t L^b$ can easily be used to demarcate equilibria.

I focus on the competitive and kinked equilibria, since I put attention on those two in the paper. The monopoly equilibrium condition in (43) can be easily expressed in terms of the expected return on equity as well.

Substituting (22) into (44) and re-arranging terms gives the inequality

$$\hat{R}^C_{E,t+1} \leq h^C\left(E_t\right),$$

with the function h^C given by

$$h^C\left(E_t\right) = \frac{\frac{1}{2}\left(\hat{q}^C_{t+1}\left(\overline{\kappa}+\underline{\kappa}\right) - \frac{\left(3c+2\hat{q}^C_{t+1}\gamma\right)}{N^C_t} - 2w_t L^e\right) + \phi\left(\frac{1}{N^C_t}\right) + \psi\left(\frac{1}{N^C_t}\right)\left(\hat{q}^C_{t+1}\overline{\kappa} - 2w_t L^e - \frac{c}{N^C_t}\right)}{1 + \frac{\phi\left(\frac{1}{N^C_t}\right)}{R^C_{f,t}} - \left(\hat{q}^C_{t+1}\overline{\kappa} - 2w_t L^e - \frac{c}{N^C_t}\right)\frac{\psi\left(\frac{1}{N^C_t}\right)}{R^C_{f,t}}}. \tag{46}$$

The expected return on equity in the competitive case $\hat{R}^C_{E,t+1}$ is given in (41).

A kinked equilibrium is defined by two conditions. The first is the failure of the above condition:

$$\hat{R}^C_{E,t+1} > h^C\left(E_t\right).$$

The second is (45), which can also be expressed as $N^M_t \Pi|_{R^M_{L,t}=\hat{q}^M_{t+1}\overline{\kappa}-2w_t L^e-\frac{c}{N^M_t}} > 0$. Substituting the bank profit function gives the inequality

$$\frac{1}{2}\left(R_{L,t} + \hat{q}^M_{t+1}\underline{\kappa}\right) - \frac{1}{2}\frac{\hat{q}^M_{t+1}\gamma}{N^M_t} - N^M_t \Pi^{t,M}_{\min} - \hat{R}^M_{E,t+1}\left(1 + fN^M_t - \frac{N^M_t \Pi^{t,M}_{\min}}{R^M_{f,t}}\right) > 0.$$

Substituting the monopoly demand curve and re-arranging produces the inequality

$$\hat{R}^M_{E,t+1} < h^M\left(E_t\right),$$

with the function h^M given by

$$h^M(E_t) = \frac{\frac{1}{2}\left(\hat{q}^M_{t+1}(\overline{\kappa}+\underline{\kappa}) - 2w_tL^e - \frac{c}{N^M_t}\right) - \frac{1}{2}\frac{\hat{q}^M_{t+1}\gamma}{N^M_t} - N^M_t\Pi^{t,M}_{\min}}{1+fN^M_t - \frac{N^M_t\Pi^{t,M}_{\min}}{R^M_{f,t}}}. \tag{47}$$

The expected return on equity in the monopoly case $\hat{R}^M_{E,t+1}$ is given in Appendix (A.5). The minimum profit function in the monopoly case when substituting the monopoly demand curve for the lending rate is

$$N^M_t\Pi^{t,M}_{\min} = \hat{q}^M_{t+1}\left(\underline{\kappa} - \frac{\gamma}{N^M_t}\right) + \Pr_{\min}\left(\frac{1}{N^M_t}\right)\left[\hat{q}^M_{t+1}(\overline{\kappa}-\underline{\kappa}) - 2w_tL^e + \frac{\hat{q}^M_{t+1}\gamma - c}{N^M_t}\right].$$

Summarizing, the economy is in a kinked equilibrium provided

$$\hat{R}^M_{E,t+1} < h^M(E_t),$$
$$\hat{R}^C_{E,t+1} > h^C(E_t).$$

A.8 Price discriminating banks

In this section, I assume that banks can identify the location of any prospective borrower. They are free to offer a loan rate that depends on the borrower's location. I do so to demonstrate the robustness of the interest rate pass-through results to price discrimination. I drop all time subscripts for simplicity.

I consider first-degree price discrimination in that a bank can capture the entire consumer surplus. A simple way to insert price discrimination is to allow banks to charge a personalized fixed premium to each entrepreneur for taking out a loan but keep all other ingredients of the model the same. The fixed premium could be a loan application or closing fee.

The premium would need to depend on the borrower's distance from the bank. It would be highest for those closest to the bank, because these borrowers would retain the largest surplus under uniform pricing, as in the baseline model.

The personalized fixed premium is a two-part tariff or affine pricing schedule. It is equivalent to a system of personalized prices with each borrower paying a sum equal to his willingness to pay.

Under both the kinked and competitive cases, let $S(x)$ be the net surplus for an entrepreneur located a distance x from bank i, which is charging lending rate R^i_L. The total amount of money $T^i(x)$ the entrepreneur pays for the loan from bank i would then be

$$T^i(x) = S(x) + R^i_L.$$

A.8.1 Kinked case

In the kinked case, the indifference condition for the entrepreneur located a distance x from bank i was

$$\frac{1}{2}\left(q\overline{\kappa} - R_L^i\right) - cx = wL^e.$$

Without price discrimination, the equilibrium kinked lending rate was $R_L = q\overline{\kappa} - 2wL^e - \frac{c}{N}$. Substituting the lending rate into the indifference condition gives the surplus for the borrower:

$$S(x) = c\left(\frac{1}{N} - x\right).$$

The surplus is positive for $x \leq \frac{1}{N}$. The upper bound is the edge of bank $i's$ potential local monopoly market before it reaches the headquarters of a neighboring bank. The personalized premium is decreasing in the borrower's distance from bank i. The bank has the most market power over those borrowers nearest to it and so it can charge the largest premium without them rejecting the loan.

Since the kinked interest rate with price discrimination does not depend on the interest rate (holding fixed the number of banks), the absence of interest rate pass-through in a kinked equilibrium holds again.

A.8.2 Competitive case

For the competitive case, it is easiest to assume a very simple expected profit function for the bank:

$$\Pi = \frac{1}{2}R_L^i\Delta^i - R\left(\Delta^i + f\right).$$

I have removed benefits from diversification, recovery value and liquidation costs. The equilibrium competitive lending rate in this case would be

$$R_L = 2R + \frac{2c}{N}.$$

This lending rate is very similar to the perfect pass-through lending rate in (24). The indifference condition for a borrower located a distance x from bank x is

$$\frac{1}{2}R_L^i + cx = \frac{1}{2}R_L + c\left(\frac{1}{N} - x\right).$$

In the competitive case, an entrepreneur minimizes costs between the two neighboring banks. The surplus from borrowing from bank i is the cost savings of doing so:

$$S(x) = \frac{1}{2}\left(R_L - R_L^i\right) + c\left(\frac{1}{N} - 2x\right).$$

The lower bank i sets the lending rate, the more surplus goes to the entrepreneur; the closer the entrepreneur is to bank i, the more surplus he or she receives.

In equilibrium, the lending rates match, so the surplus comes to

$$S(x) = c\left(\frac{1}{N} - 2x\right).$$

This surplus is positive for $x \leq \frac{1}{2N}$. An entrepreneur located that distance $x = \frac{1}{2N}$ is in between the two banks. That entrepreneur is marginal, so he or she will not be charged a personalized premium. Everyone else will be charged the fixed premium $S^i(x) = c\left(\frac{1}{N} - 2x\right)$ according to their distance from the bank. Compared to the kinked case, the surplus in the competitive case declines twice as fast due to the entrepreneur's credible alternative of contracting with a competitor.

A.9 Smoothing the kink

In this section I provide one way to "smooth" the kink in the demand curve for bank credit (make the region differentiable). I do so in order to demonstrate that the limited pass-through at the kink is robust even after smoothing it. I drop the time subscripts to simplify notation in this section.

A.9.1 General Case

Generally, the pass-through of marginal costs to prices is lower at points of a downward sloping demand curve that feature greater concavity. In the model, the kink is a sharp way of creating concavity in the demand curve for loans when consumer preferences would otherwise imply a linear demand curve (and perfect pass-through). As long as the smoothing procedure preserves the highest concavity at the smoothed kink, then the pass-through will be lowest there, just as when the kink is sharp.

One way to smooth the kink is to assume that banks are unsure about which demand curve they are on when setting a price. Since banks set prices at the margin, they may equivalently be unsure about the slope of the demand curve at a given price.

The two possible slopes a bank faces is either $-c$ or $-2c$. I assume the bank assigns a probability h that the slope is $-2c$, which is to say that it believes with probability h that it is competing with a neighboring bank. The probability with which the bank believes it is a local monopolist is then $(1-h)$.

I assume that h is an increasing function of the bank's market share Δ, is continuous, and is three times differentiable over the domain I specify below.This kind of uncertainty can be rationalized by a bank not knowing the precise boundary of its neighbors' market, but knowing that it has increasingly likely penetrated that boundary as it expands its market share. I assume the bank knows the number of banks operating in the lending market N.

For simplicity, I also assume the bank cannot recover any value from a loan in default ($\underline{\kappa} = 0$) and that it bears no liquidation costs ($g = 0$). I also assume the bank is flush with equity so that its cost of capital is the interest rate R, which I take as a parameter.

The expected profit function of a bank is then

$$\Pi = \frac{1}{2} R_L \Delta - R \left(\Delta + f \right).$$

The bank will chose a market share Δ that satisfies the first order condition

$$\Omega \equiv R_L \left(\Delta \right) + \Delta R_L' \left(\Delta \right) - 2R = 0.$$

By the implicit function theorem, the quantity pass-through pass-through $\frac{\partial \Delta}{\partial R}$ is

$$\begin{aligned} \frac{\partial \Delta}{\partial R} &= -\frac{\frac{\partial \Omega}{\partial R}}{\frac{\partial \Omega}{\partial \Delta}} \\ &= \frac{2}{R_L' + \Delta R_L'' \left(\Delta \right) + R_L'} \\ &= \frac{2}{2R_L' + \Delta R_L'' \left(\Delta \right)}. \end{aligned}$$

Now by the chain rule, the interest rate pass-through is

$$\frac{\partial R_L}{\partial R} = \frac{\partial R_L}{\partial \Delta} \frac{\partial \Delta}{\partial R}.$$

Substituting for the quantity pass-through gives

$$\frac{\partial R_L}{\partial R} = R_L' \left(\frac{2}{2R_L' + \Delta R_L''} \right).$$

Assuming symmetry in market shares and a completely served circle gives $\Delta = \frac{1}{N}$. Substituting and re-arranging terms gives

$$\begin{aligned}\frac{\partial R_L}{\partial R} &= \frac{2R_L'}{2R_L' + \frac{1}{N}R_L''} \\ &= \frac{2}{2 + \frac{1}{N}\frac{R_L''}{R_L'}}\end{aligned}$$

Define the *concavity* of the demand curve as

$$\omega\left(\Delta\right) \equiv \frac{R_L''\left(\Delta\right)}{R_L'\left(\Delta\right)}.$$

The interest rate pass-through is then

$$\frac{\partial R_L}{\partial R} = \frac{2}{2 + \omega\left(\Delta\right)/N}.$$

With a downward sloping demand curve, a larger concavity implies a lower interest rate pass-through. Also, as the number of banks tends to infinity $N \to \infty$, the market reaches perfect competition and features perfect pass-through.

Returning to a bank's uncertainty over the slope of its demand curve, we have

$$\begin{aligned}R_L'\left(\Delta\right) &= -\left(2c \times h\left(\Delta\right) + c \times \left(1 - h\left(\Delta\right)\right)\right) \\ &= -\left(c + ch\left(\Delta\right)\right) \\ &= -c\left(1 + h\left(\Delta\right)\right)\end{aligned}$$

This object is the slope of the bank's subjective demand curve given its beliefs about being a competitor or local monopolist. The second derivative gives

$$R_L''\left(\Delta\right) = -ch'\left(\Delta\right).$$

The concavity of the demand curve for loans is then

$$\omega\left(\Delta\right) = \frac{h'\left(\Delta\right)}{1 + h\left(\Delta\right)}.$$

Because h is a probability, I require $h\left(\Delta\right) \geq 0$ for all Δ. I also assume that $h' \geq 0$, making the bank increasingly believe it is competing as it expands. These assumptions make $\omega\left(\Delta\right) \geq 0$. Furthermore, a bank knows that its headquarters is located at $\Delta = 0$. It also knows that its loan portfolio reaches the neighboring bank's headquarters when $\Delta = \frac{2}{N}$. Therefore, it is reasonable to assume that $\lim_{\Delta\downarrow 0} h(\Delta) = 0$ and $\lim_{\Delta\uparrow\frac{2}{N}} h(\Delta) = 1$. I define the domain of h to be the closed interval $\left[0, \frac{2}{N}\right]$.

If the concavity is uniquely highest at the point midway between headquarters of banks, then the lowest pass-through would occur at the location of the kink in the baseline model. The function ω should be globally maximized at $\frac{1}{N}$ and hence satisfy

$$\omega'\left(\frac{1}{N}\right) = 0 \tag{48}$$

$$\omega''\left(\frac{1}{N}\right) < 0. \tag{49}$$

A simple way to ensure the global maximum is uniquely reached when $\Delta = \frac{1}{N}$ is to assume that the function $\omega(\Delta)$ is strictly concave along the entire interval. A sufficient but not necessary condition for the global maximum is

$$\omega''(\Delta) < 0, \qquad \forall\Delta \in \left[0, \frac{2}{N}\right]. \tag{50}$$

The restrictions on the belief function h imposed by the conditions (48)-(49) will smooth the kink in the demand curve in a way to minimize the interest rate pass-through at that point. If condition (50) is also imposed, then it is guaranteed that that point will uniquely minimize the pass-through.

A.9.2 Example for h

Given the assumptions for h, a cumulative distribution function with non-negative bounded support that satisfies (48)-(49) and where the maximum is unique would deliver an appropriate example.

I use the beta distribution $h(x) = \text{Beta}(x; \alpha, \beta)$, where I make the linear transformation $x = \frac{\Delta}{(2/N)}$ that adjusts the support to $[0, 1]$. Thus, a bank that has a portfolio size of $x = 1$ unit believes its loan portfolio has reached the headquarters of the two neighboring banks.

Choosing parameters α and β that deliver a global maximum for ω at $x = 1/2$ would give what is needed, which is the highest curvature at $\Delta = 1/N$, the location of the kink in the baseline model. I find those parameters computationally. I search for the parameter vector (α, β) that minimizes the function $|\text{argmax}_{x(\alpha,\beta)}\omega(x(\alpha,\beta)) - \frac{1}{2}|$ over the unit interval and confirm graphically that the solution associates with a global maximum of $\omega(x)$. I use 10,000 starting points for the optimization routine.

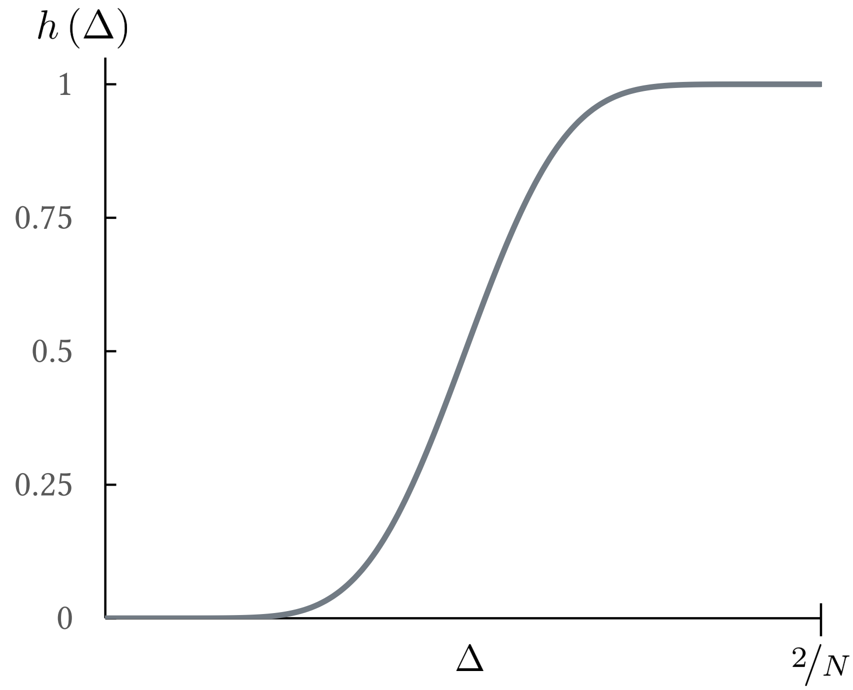
The optimal solution from the search is $\alpha = 1059$ and $\beta = 1046$. The differences between the smooth and original demand curves under these parameters, however, are difficult to see.

Therefore, in the figures below, I use the parameters $\alpha = 10.59$ and $\beta = 10.46$, which make the differences clearer and delivers a concavity function ω that approximately is maximized at $x = \frac{1}{2}$.

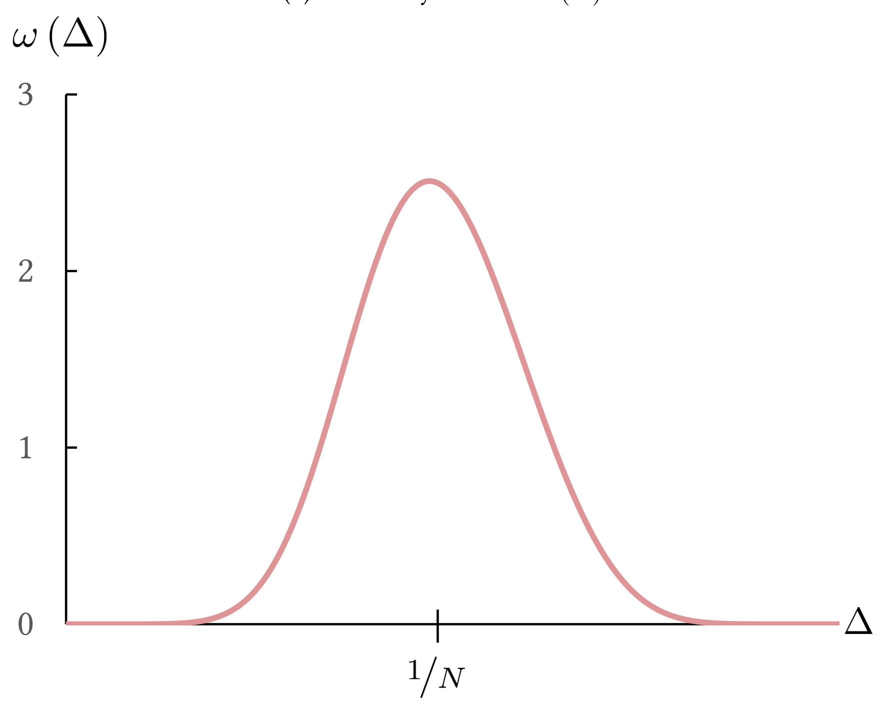
Figures 13(a)-13(b) plot the probability function $h\left(\Delta\right)$ and the concavity function $\omega\left(\Delta\right)$.

Figure 13: Uncertainty Over the Demand Curve for Loans

(a) Probability function $h(\Delta)$



(b) Concavity function $\omega(\Delta)$



Notes: The function $h(\Delta)$ represents the probability a bank believes it is competing with a neighboring bank. This uncertainty affects the weight a bank places on the two possible slopes in the demand curve for loans. The function $\omega(\Delta)$ is the concavity of the subjective demand curve implied by the bank's beliefs. The support of the beta distribution has been transformed from the unit interval to $[0, {}^2/_N]$ in the figures. The scale parameters for the beta distribution are $\alpha = 10.59$ and $\beta = 10.46$.

The figures reveal that all the uncertainty is concentrated around $\frac{1}{N}$, which is the location a bank would expect the boundary of a neighboring bank's market to be in equilibrium. This probability distribution leads the concavity of the subjective demand curve to be maximized around the location of the kink.

Figure 14 illustrates the original sharp kink featuring no uncertainty about the demand curve and the corresponding smoothed kink in which there is uncertainty. For $\Delta < \frac{1}{N}$, the smooth curve begins to deviate from the original curve once the bank starts assigning positive probability to competing with a neighbor. The bank reduces its lending rate faster in order to attract the marginal borrower because the bank believes it might now be competing for that customer. As Δ approaches $\frac{1}{N}$, the original kink is entirely "rounded out" and the smooth demand curve displays the most concavity. Interest rate pass-through will be lowest in that region, though not zero.

As the bank extends its loan portfolio further past $\frac{1}{N}$, it becomes more confident that it is competing with the neighbor, and so the bank puts more weight on the slope of the demand curve being $-2c$. The original and smooth demand curves converge.

A.10 Dynamical system, variable investment scale

In this section I present the equilibrium equations that define the dynamical system when the scale of project investment can vary. The state variable of the system is the current period physical capital stock K_t.

The scale of project investment is

$$\iota_t = \log\left(\frac{\hat{q}_{t+1}\overline{\kappa}}{R_{L,t}}\right). \tag{51}$$

From labor market clearing, the current wage and expected wage next period are given by

$$w_t = (1-\alpha)\, A_t K_t^{\alpha}, \tag{52}$$

$$E_t(w_{t+1}) = (1-\alpha)\, A K_{t+1}^{\alpha}. \tag{53}$$

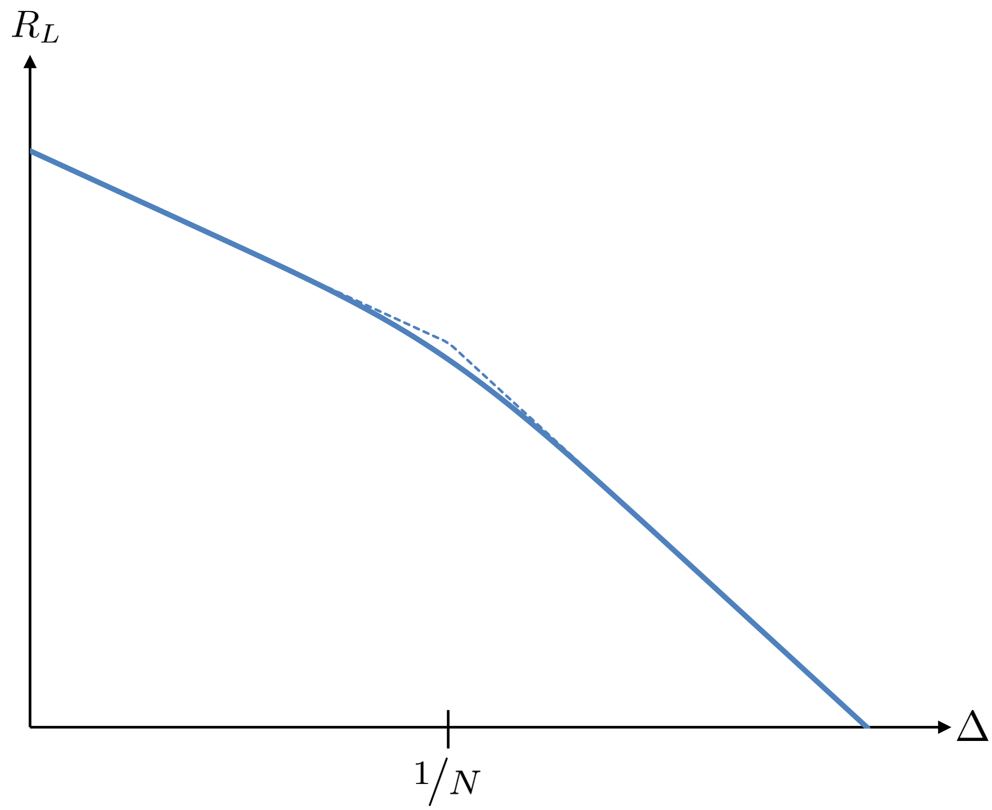
In the physical capital market, the supply of physical capital is from the physical-capital-formation equation:

$$K_{t+1} = \frac{1}{2}\left(\overline{\kappa} + \underline{\kappa}\right) - \frac{1}{2}\left(1 + {}^{\underline{\kappa}}/_{\overline{\kappa}}\right)\frac{R_{L,t}}{\hat{q}_{t+1}} - \frac{{}^{\gamma}/_{2}}{N_t^2}\log\left(\frac{\hat{q}_{t+1}\overline{\kappa}}{R^i_{L,t}}\right), \tag{54}$$

while the demand for physical capital is the expected marginal productivity:

$$\hat{q}_{t+1} = \alpha A K_{t+1}^{\alpha-1}. \tag{55}$$

Figure 14: Smooth and Original Demand Curves



Notes: The dashed curve is the original demand curve that features a kink at $\Delta = \frac{1}{N}$. The solid curve is the smooth demand curve that is derived from a bank being uncertain about the slope of the demand curve it faces. The uncertainty is captured by a beta distribution with scale parameters $\alpha = 10.59$ and $\beta = 10.46$.

The after-tax interest rate $R_{f,t}(1-\tau)$ is found by combining market clearing in the deposit and equity markets, and is given by:

$$R_{f,t}(1-\tau) = \frac{\hat{w}_{t+1}}{\beta w_t - \frac{1+\beta(1-\tau)}{L_d}\left(\iota_t + fN_t - w_t L^b\right)}. \tag{56}$$

The remaining endogenous objects are the expected return on equity $\hat{R}_{E,t+1}$ and number of banks N_t. Substituting the equity market clearing condition into the zero profit condition gives for the zero profit condition:

$$\frac{1}{2}\left(R_{L,t}\left(\iota_t - \underline{\kappa}/\overline{\kappa}\right) + \hat{q}_{t+1}\underline{\kappa}\right) = \frac{\hat{q}_{t+1}\gamma}{2}\frac{\iota_t}{N_t} - \left(\hat{R}_{E,t+1} - R_{f,t}\right)\left(\iota_t + fN_t - w_t L^b\right) + \hat{R}_{E,t+1}\left(\iota_t + fN_t\right). \tag{57}$$

Finally, the equity market clearing condition is

$$w_t L^b = \iota_t + f N_t - N_t \left(\frac{\Pi^{i,t}_{\min}}{R_{f,t}} \right). \tag{58}$$

Equations (51) - (58) define the dynamical system for a given lending rate $R_{L,t}$. In both the kinked and competitive equilibria, the lending rate is defined implicitly. In the kinked case, the lending rate is given by

$$R^K_{L,t} = \frac{\hat{q}_{t+1}\overline{\kappa} - 2w_t L^e - \frac{c}{N_t}}{1 + \iota_t}. \tag{59}$$

Because the kinked lending rate is not a function of the $\hat{R}_{E,t+1}$, the expected return on equity can be solved in closed form using (57):

$$\hat{R}_{E,t+1} = \frac{\frac{1}{2}\left(R_{L,t}\left(\iota_t - {\underline{\kappa}}/{\overline{\kappa}}\right) + \hat{q}_{t+1}\underline{\kappa}\right) - \frac{\hat{q}_{t+1}\gamma}{2}\frac{\iota_t}{N_t} - R_{f,t}\left(\iota_t + f N_t - w_t L^b\right)}{w_t L^b}. \tag{60}$$

The number of banks N_t and expected equity return $\hat{R}_{E,t+1}$ are determined jointly in the competitive case by (57) and (58).

Examining the system, one can see that the current period physical capital stock K_t is the state variable of the economy, just as it was in the unit investment case.