REVISED

MACROECON & INT'L FINANCE WORKSHOP
presented by ***Dean Corbae***
FRIDAY, March 1, 2013
3:30 pm – 5:00 pm, **Room: HOH-706**

A Quantitative Model of Banking Industry Dynamics*

Dean Corbae
University of Wisconsin - Madison

Pablo D'Erasmo
University of Maryland at College Park

March 1, 2013

Abstract

We develop a model of banking industry dynamics to study the relation between commercial bank market structure, entry and exit along the business cycle, and the riskiness of commercial bank loans as measured by default frequencies. We analyze a Stackelberg environment where a small number of dominant banks choose their loan supply strategically before a large number of small banks (the competitive fringe) make their loan choices. A nontrivial endogenous bank size distribution arises out of entry and exit in response to aggregate and regional shocks to borrowers' production technologies. The model is estimated using first moments of aggregate and cross-sectional statistics for a panel of the entire U.S. commercial banking industry. The model is qualitatively consistent with many non-targeted moments; for instance, the model generates countercyclical loan interest rates, bank failure rates, default frequencies, and markups as well as procyclical loan supply and entry rates. The model is used to study the effects of increased bank competition and the benefits/costs of a set of policies: (i) branching restrictions; (ii) lower costs of loanable funds; and (iii) big bank bailouts.

1 Introduction

The objective of this paper is to formulate a quantitative structural model of the banking industry consistent with data in order to understand the relation between market structure and risk taking by financial intermediaries. Once the underlying technological parameters are consistently estimated, we can also use the model to address important banking policy

*INCOMPLETE. The authors wish to thank John Boyd, Jeff Campbell, Gianni DeNicolo, Nobu Kiyotaki, Jim MacGee, Raphael Repullo, John Rust, Juan Sanchez, Kenechi Ueda, and Steve Williamson for comments on an earlier version of this paper as well as seminar participants at the Bank of Canada, Bank of England, IMF, Federal Reserve Banks of Atlanta, Chicago, Dallas, Philadelphia and St. Louis, SED, University of Toronto, University of Wisconsin, the BU/Boston Fed Conference on Macro-Finance Linkages, the Workshop on Financial Systems (CFSP) at MIT, McMaster University and Tilburg University.

questions. We want the model to be rich enough to address issues like those posed by Ben Bernanke "I want to be very, very clear: too big to fail is one of the biggest problems we face in this country, and we must take action to eliminate too big to fail."[1]

Banks in our environment intermediate between a large number of households who supply funds and a large number of borrowers who demand funds to undertake risky projects. By lending to a large number of borrowers, a given bank diversifies risk that any particular household cannot accomplish individually. Since we estimate banking sector technology parameters, we require our model to be parsimonious. When mapping the model to data, we attempt to match long run average and cross-sectional statistics for the U.S. banking industry.

Our model assumes spatial heterogeneity between banks; there is a representative "national" geographically diversified bank that may coexist in equilibrium with "regional" and "fringe" banks that are both restricted to a geographical area. This breakdown is roughly consistent with a market structure where the Top 10 banks are associated with national banks (which have 51% of the loan market share), Top 1% (minus Top 10) associated with regional banks (which make up 25% of loan market share), and the bottom 99% which are associated with the fringe banks (which make up the remaining 24% of loan market share). Since we allow for regional specific shocks to the success of borrower projects, smaller banks may not be well diversified. This assumption generates ex-post differences between big and smaller (regional and fringe) banks. As documented in the data section, the model generates not only procyclical loan supply but also countercyclical interest rates on loans, borrower default frequencies, and bank failure rates. Since bank failure is paid for by lump sum taxes to fund deposit insurance, the model predicts countercylical policies to bail out failing banks.

Some of the questions to be addressed in this paper are: Are crises (defined as bank exit rates or default frequencies higher than some threshold) less likely in more concentrated banking industries? What are the costs of policies to mitigate bank failure? Besides our quantitative approach, the benefit of our model relative to the existing literature is that the size distribution of banks is derived endogenously and varies over the business cycle - a fact which is evident in the data.

Our paper is most closely related to the following literature on the industrial organization of banking. Our underlying model of banking is based on the static models of Allen and Gale [4] (hereafter A-G) and Boyd and De Nicolo [11] (hereafter B-D). A dynamic version of the model is considered in Martinez-Miera and Repullo [34].[2] In all of those models, the authors study the implications of exogenously varying the number of banks on loan supply and borrower risk taking. A-G make model assumptions such that more concentration leads to more stability while B-D show that in other cases, more concentration leads to more fragility.[3] Unlike the previous papers, we do not exogenously fix the number of banks but

[1] Time, December 28, 2009/January 4, 2010, p. 78.

[2] Another strand of literature uses the costly state verification approach of Townsend [39] either ex-ante as in Diamond [17] or ex-post as in Williamson [41] to rationalize the existence of banks. These papers all study a competitive market structure however.

[3] In particular, an exogenous increase in the number of banks in both A-G and B-D raises interest rates that banks must pay their depositors. However, in A-G those costs are passed on to borrowers since there is not loan market competition; this results in higher borrower default probabilities and ultimately lower

instead solve for an equilibrium where banks enter and exit so that the number of banks is endogenously determined. To keep the model simple, here we focus only on loan market competition while there is an important IO literature on deposit market competition (see for example Aguirregabiria, Clark, and Wang [2]).

As described above, we require our quantitative model to be consistent with U.S. data on market concentration. In particular, measures of imperfect competition we compute from the U.S. data suggest less than perfect competition. Hence, instead of assuming perfect competition, we propose a Stackelberg model where national and regional banks strategically choose the quantity of loans they supply before fringe banks choose their loan supply taking prices as given. We apply the Markov Perfect Industry equilibrium concept of Ericson and Pakes [22] augmented with a competitive fringe along the lines of Gowrisankaran and Holmes [24]. In this way, we depart from quantitative competitive models of banking such as Bernanke, et. al. [9], Carlstrom and Fuerst [12], or Diaz-Gimenez, et. al. [19], thereby allowing big banks to act strategically in the loan market. Further, dropping the competitive assumption along with our spatial restrictions generates a nontrivial size distribution of banks where both intensive and extensive margins can vary over the business cycle which is broadly consistent with data. The Stackelberg game allows us to examine how changes in the environment which affect big banks spills over to the rest of the industry. While this is not systemic risk in the sense of interconnected balance sheets (whereby a failure in a big bank worsens the balance sheet of other banks who hold its interbank market liabilities), it does allow us to consider how a big bank's loan behavior can worsen the balance sheet position of other banks. For instance, a too-big-to-fail policy which leads the national bank to make more loans in risky states of the world lowers the interest rate on loans which in turn lowers the profitability of other banks. When mapping the model to data, we use the same dataset as Kashyap and Stein [30].

There is a vast empirical literature that takes up the "concentration-stability" versus "concentration-fragility" debate. For example, Beck, Demirguc-Kunt, and Levine [6] run probit regressions where the probability of a crisis depends on banking industry concentration as well as a set of controls. In their regressions a "crisis" is defined to be a significant fraction of insolvent banks (or a fraction of nonperforming loans exceeding 10%). While Beck, et. al. find evidence in favor of the concentration-stability view, in general there are mixed results from this empirical work. We address this debate using our quantitative structural model. After estimating the model by Simulated Method of Moments (SMM) using aggregate and cross-sectional statistics for the U.S. banking industry, we simulate banking industry market structure in response to the aggregate and regional shocks in our model and then run the same types of regressions with "crisis" dependent variables as in Beck, et. al. that the empirical literature studies across the endogenously determined differences in market structure. We find that our model is consistent with results from the empirical literature.

We then take up a series of policy counterfactuals. First, we study the effects of branching restrictions. In particular, we increase the cost of entry for national banks to a prohibitively

realized profits. In B-D, on the other hand, increased loan market competition lowers interest rates on loans as well and this lowers borrower default probabilities which may ultimately raise realized profits. Which effect dominates in the B-D case is a quantitative matter.

large value. Since regional banks become the only dominant players in the market, we find that this policy reduces competition, increasing interest rates and default frequencies but reducing exit rates. The effect on interest rates is smaller with regional monopolies than in a counterfactual with national monopolies because regional banks are less diversified.

Next, we study the effects of policies to mitigate bank exit by lowering the cost of loanable funds. In particular, we compare our benchmark economy to one where the cost of loanable funds is zero (as opposed to 0.72%). The lower cost leads to more lending and lower interest rates. The ensuing drop in loan returns is actually large enough that profits and markups are reduced, thereby leading to a drop in entry. Interestingly, the lower costs of loanable funds also changes the cyclical properties of loan rates. In particular at zero cost, loan returns are procyclical as opposed to countercyclical in the benchmark. This provides an example where government policy can lead to different cyclical properties for loan interest rates.

Finally, we study the effects of bank bailout policies to mitigate exit. In particular, we compare our benchmark economy with one where the government is committed to cover negative profits of national banks preventing them from exit. In the benchmark case, the possible loss of charter value and costs of equity issuance is enough to induce national banks to lower loan supply in order to reduce exposure to risk. In the counterfactual case, the representative national bank increases exposure to the region with high downside risk since its continuation (charter) value is guaranteed. Regional and fringe banks reduce their loan supply in order to avoid their own exit. The decrease in loans by smaller banks dominates generating higher interest rates. Due to the risk shifting effect, borrowers take on more risk and default frequencies rise by 16%. Lower profitability of smaller banks lowers entry. Exit by regional banks means that the tax to output ratio rises by 4%.

The remainder of the paper is organized as follows. Section 2 documents a select set of banking data facts. Section 3 lays out a simple model environment to study bank risk taking and loan market competition. Section 4 describes a markov perfect equilibrium of that environment. Section 5 discusses how the preference and technology parameters are chosen and section 6 provides results for the simple model. Section 7 conducts four counterfactuals: (i) one experiment assesses the effects of bank competition on business failures and banking stability; (ii) another experiment assesses the effects of regulation which restricts banks to a geographical region; (iii) another experiment assesses the consequences of a "too big to fail" policy; and (iv) a final experiment assesses the effects of a policy that reduces the cost of funds that banks use to make loans on risk taking and exit. Section 8 concludes and lists a set of extensions to the simple model which we are currently pursuing.

2 Some Banking Data Facts

In this section, we document the cyclical behavior of entry and exit rates, bank lending, measures of loan returns and the level of concentration in the U.S. As in Kashyap and Stein [30] we focus on individual commercial banks in the U.S.[4] The source for the data

[4]In some cases, commercial banks are part of a larger bank holding company. For example, in 2008, 1383 commercial banks (20% of the total) were part of a bank holding company. As Kashyap and Stein [30] argue, there are not significant differences in modeling each unit. The holding company is subject to limited

is the Consolidated Report of Condition and Income (known as Call Reports) that insured banks submit to the Federal Reserve each quarter.[5] We compile a data set from 1976 to 2010 using data for the last quarter of each year.[6] We follow Kashyap and Stein [30] in constructing consistent time series for our variables of interest. In the Data Appendix, we provide a detailed description of variable definitions and how all the statistics reported are constructed.

One clear trend of the commercial banking industry during the last three decades is the continuous drop in the number of banking institutions. In 1980, there were approximately 14,000 institutions and this number has declined at an average of 360 per year, bringing the total number of commercial banks in 2010 to less than 6,600. This trend was a consequence of important changes in regulation that were introduced during the 1980's and 1990's (deposit deregulation in the early 1980's and the relaxation of branching restrictions later).

This decline in the number of active banks is evidenced by flow measures of exit and entry. The number of exits (including mergers and failures) and entrants expressed as a fraction of the banking population in the previous year are displayed in Figure 1. We also incorporate detrended real log-GDP to understand how entry and exit rates move along the cycle.[7]

---

liability protection rules with respect to the losses in any individual bank.

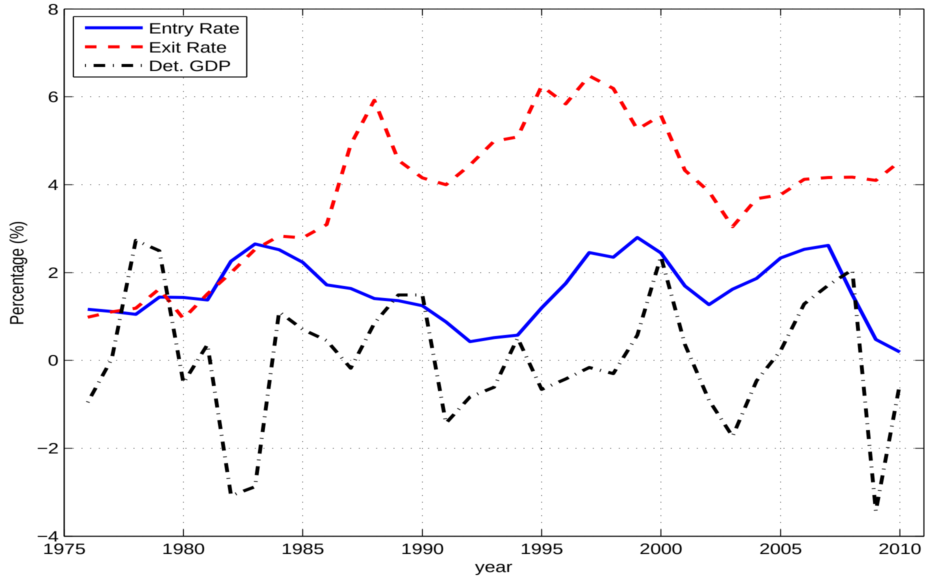
[5]The number of institutions and its evolution over time can be found at http://www2.fdic.gov/hsob/SelectRpt.asp?EntryTyp=10.

Balance Sheet and Income Statements items can be found at https://cdr.ffiec.gov/public/.

[6]We extend the data to 2010 (the last year available at the moment of writing the paper) but all our results are robust to the inclusion of years 2008/09/10 (i.e the financial crisis period).

[7]The H-P filter with parameter equal to 6.25 is used to extract the trend from log real GDP data.

Figure 1: Bank Industry Dynamics and Business Cycles

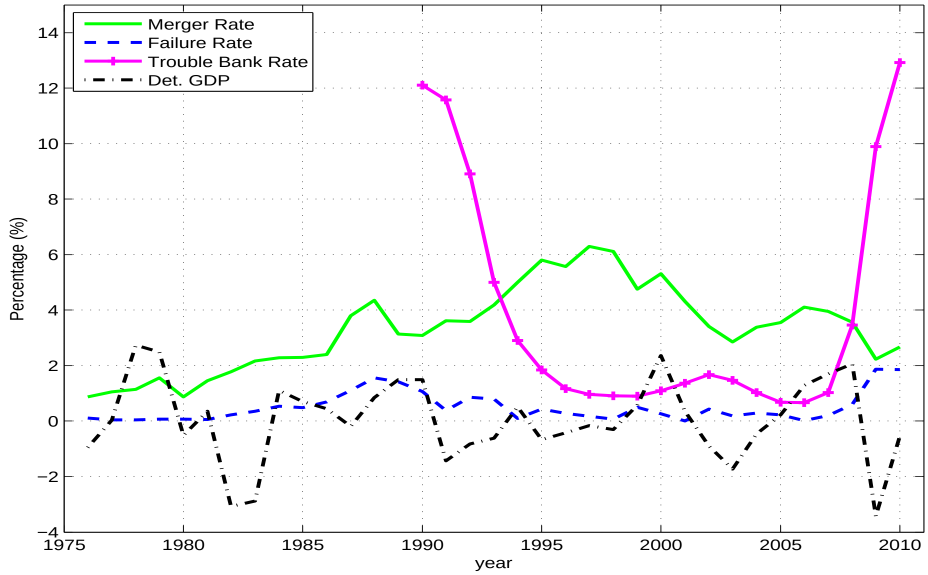


Note: Data corresponds to commercial banks in the US. Source: Consolidated Report of Condition and Income. Entry corresponds to new charters and conversions. Exit correspond to unassisted mergers and failures. Det. GDP refers to detrended log real GDP. The trend is extracted using the H-P filter with parameter 6.25.

Figure 1 shows that there was an important increase in the fraction of banks that exited starting in the 1980's and the high level continued through the late 1990's due to the aforementioned regulatory changes. The figure also shows that there has been a consistent flow of entry of new banks, cycling around 2%. Since 1995 the net decline in the number of institutions has trended consistently lower (except for the most recent period 2008-2010) so that the downward trend is leveling off. This recent leveling off of the trend is also documented in Table 6 of Janicki and Prescott [29]. Figure 2 decomposes the exit rate into mergers and failures as well as the fraction of 'troubled' banks for a subset of the period.[8] The bulk of the decline was due to mergers and acquisitions. However, from 1985 through 1992, failures also contributed significantly to the decline in the number of banks.

[8] A troubled bank, as defined by the FDIC, is a commercial bank with CAMEL rating equal to 4 or 5. CAMEL is an acronym for the six components of the regulatory rating system: **C**apital adequacy, **A**sset quality, **M**anagement, **E**arnings, **L**iquidity and market **S**ensitivity (since 1998). Banks are rated from 1 (best) to 5 (worst), and banks with rating 4 or 5 are considered 'troubled' banks (see FDIC Banking Review 2006 Vol 1). This variable is only available since 1990.

Figure 2: Exit Rate Decomposed



Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement. Det. GDP refers to detrended log real GDP. The trend is extracted using the H-P filter with parameter 6.25.

Figures 1 and 2 also make clear that there was a significant amount of cyclical variation in entry and exit. The correlation of the entry and exit rates with detrended GDP is 0.25 and 0.22 respectively.[9] If we restrict to the post-reform years - after 1990 - the correlations with detrended GDP are 0.60 and 0.38 for entry and exit rates respectively.

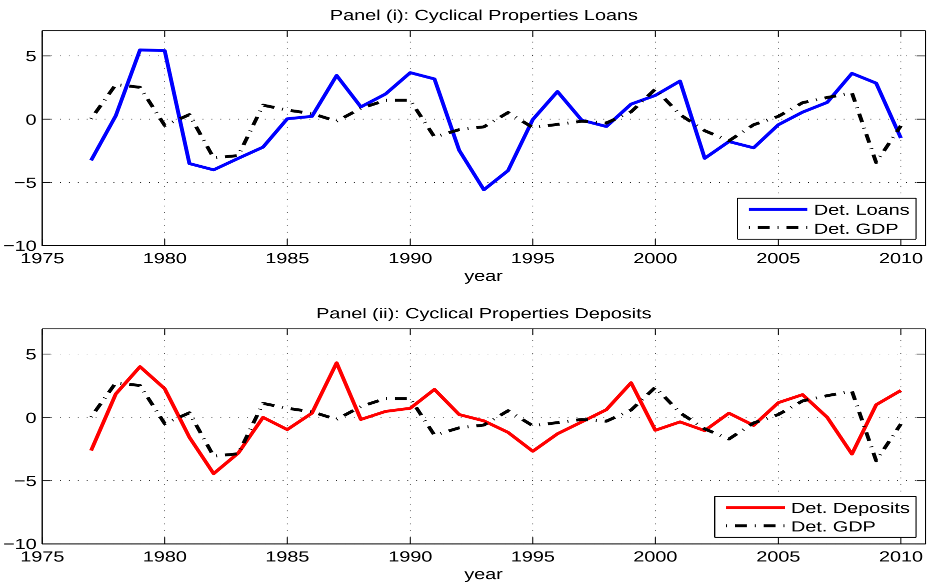
Since exits can occur as the result of a merger, these correlations hide what we usually think of as the cyclical component of exits; failures and "troubled" banks have a more important cyclical component than mergers. We find that the failure and "troubled" bank rates are countercyclical while the merger rate displays a procyclical behavior. Specifically, the correlation with real detrended GDP of the failure and "troubled" bank rates (after 1990) are -0.47 and -0.72 respectively. On the other hand, the correlation of the merger rate for the same period equals 0.58.

Figure 3 displays how bank lending and deposits move along the cycle. The series for loans and deposits are constructed by aggregating the individual commercial bank level data.[10]

[9] All cyclical correlations are computed using H-P filetered series.

[10] The CPI index is used to convert the nominal loan and deposit variables into real. The H-P filter with parameter equal to 6.25 is used to extract the trend from the data.

Figure 3: Loans, Deposits and Business Cycles



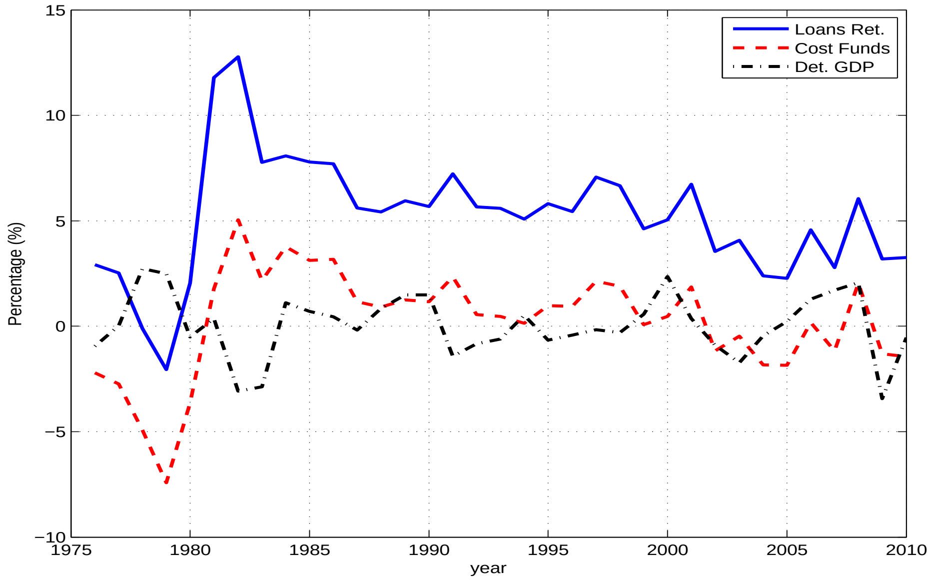
Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement. Det. Loans and Det. Deposits refer to detrended real stock of loans and deposits respectively. Det. GDP refers to detrended real log-GDP. The trend is extracted using the H-P filter with parameter 6.25.

This figure shows that the stock of loans and deposits have an important cyclical component. We find that both measures of bank activity are procyclical where correlations with detrended GDP equal 0.72 and 0.22 for loans and deposits respectively.

Figure 4 presents the behavior of the rate of return on loans and the cost of funds.[11] Rates of return on loans and the cost of funds display a countercyclical behavior. Their correlation with detrended log real GDP equals -0.26 and -0.23 respectively.

[11]The rate of return on loans is defined as interest income from loans divided by total loans. The cost of funds is constructed as interest expense on deposits and federal funds divided by the sum of deposits and federal funds. Variables reported are loan-weighted averages. See the Data Appendix for a detailed definition. Nominal returns are converted into real returns using the CPI index.

Figure 4: Loan Returns, Costs of Funds and Business Cycles

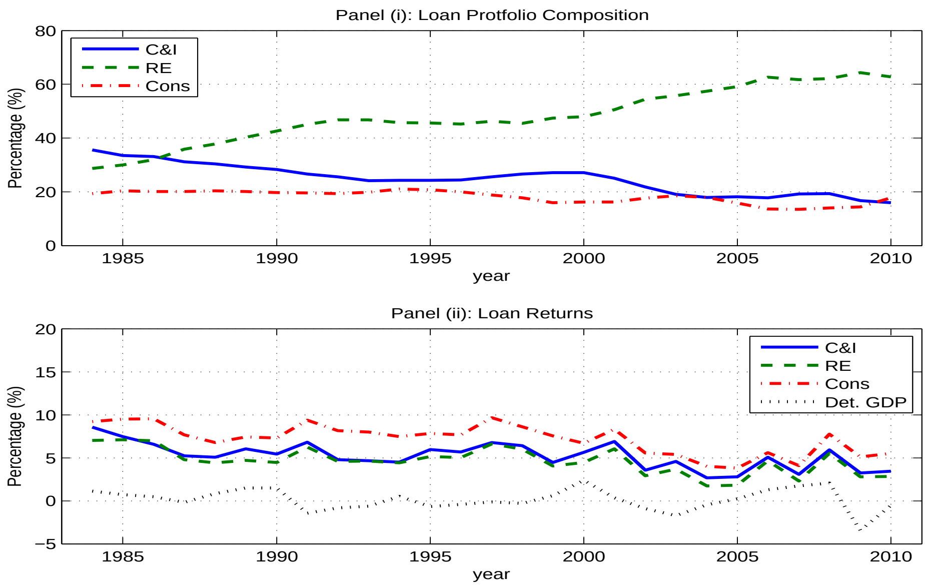


Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement. See footnote 11 for loan return and deposit cost definition. Det. GDP refers to detrended real GDP. The trend is extracted using the H-P filter with parameter 6.25.

An important trend observed in the loan portfolio composition of commercial banks is the increase in the fraction of loans secured by real estate and the decrease in the amount of commercial and industrial lending. In Panel (i) of Figure 5, we present the fraction of total loans accounted by Industrial and Commercial loans $(C\&I)$, loans Secured by Real Estate (RE) and Consumer loans $(Cons)$. Panel (ii) in Figure 5 shows the loan return by loan type.[12]

[12]Variables reported are loan-weighted averages. Consistent data dissagregated by loan type is available from 1984 - 2010.

Figure 5: Loan Portfolio Composition and Loan Returns



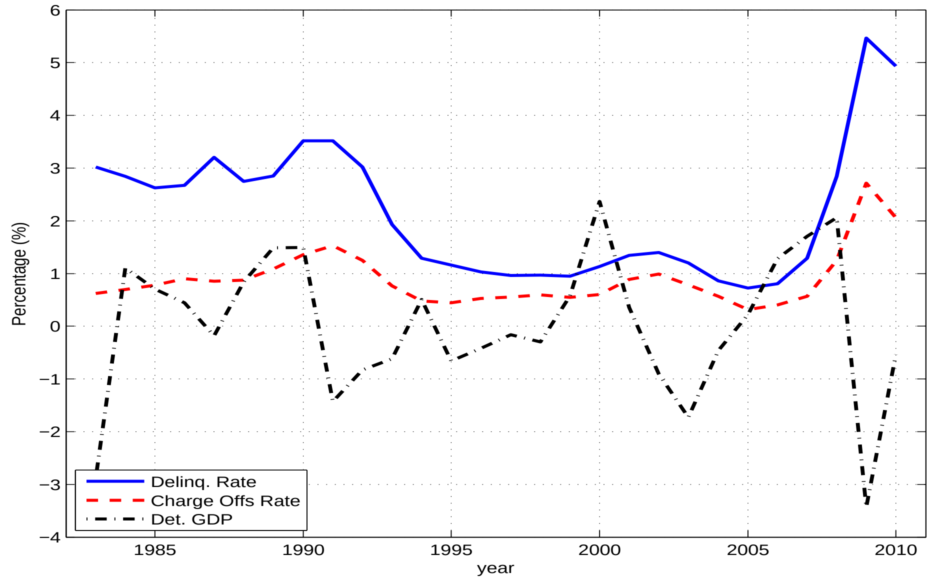
Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement. See footnote 11 for loan return definition. Det. GDP refers to detrended log real GDP. The trend is extracted using the H-P filter with parameter 6.25.

The fraction of loans secured by real estate loans more than doubled during this period and that most of the reduction came from commercial and industrial loans. We observe that loan returns display a milder but still countercyclical behavior when dissagregated by loan type (commercial and industrial loans, real estate loans and consumer loans). The correlation with detrended GDP equals -0.01, -0.13 and -0.26 for commercial and industrial, secured by real estate and consumer loans respectively.

In Figure 6, we present the evolution of loan delinquency rates and charge off rates.[13]

[13] The delinquency rate is the ratio of loans past due 90 days or more plus non accrual loans divided by total loans. Consistent data for delinquency rates is only available since 1983. Charge-off rates are defined as the flow of a bank's net charge-offs (charge-offs minus recoveries) divided by total loans. We report loan-weighted averages.

Figure 6: Loan Delinquency Rates, Charge Off Rates and Business Cycles

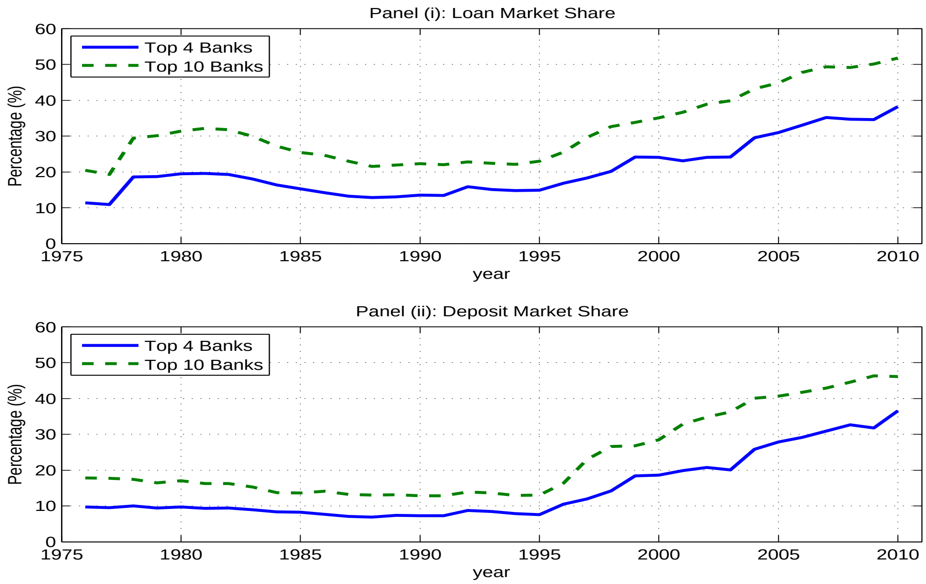


Note: Data corresponds to commercial banks in the US.Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement. See footnote 13 for delinquency rates and charge offs rate definition. Det. GDP refers to detrended real GDP. The trend is extracted using the H-P filter with parameter 6.25.

Delinquency rates and charge off rates are countercyclical. Their correlation with detrended GDP is -0.61 and -0.56 respectively. Considering only the period until 2007, provides correlations with detrended GDP equal to -0.46 and -0.44 respectively.

The size distribution of banks has always been skewed but the large number of bank exits (mergers and failures) that we documented above resulted in an unparalleled increase in loan and deposit concentration during the last 35 years. Figure 7 displays the trend in the share of loans and deposits in the hands of the four and ten largest banks (when sorted by loans) since 1976.

Figure 7: Increase in Concentration: Loan and Deposit Market

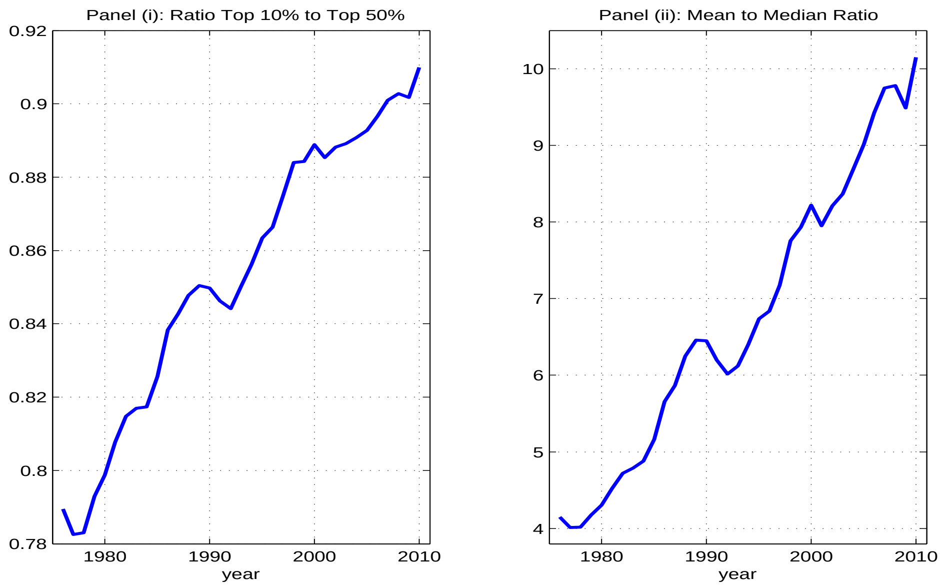


Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement.

Figure 7 shows that for example, in 1976 the four largest banks (when sorted by loans) held 11 and 10 percent of the banking industry's loans and deposits respectively while by 2010 these shares had grown to 38 and 37 percent.

The increase in the degree of concentration is also evident in the evolution of the mean-to-median ratio and the ratio of total loans in hands of the top 10 percent banks to the total loans in hands of top 50 percent of the loan distribution. Figure 8 displays the trend in these two measures of concentration in the loan market since the year 1976.

Figure 8: Increase in Concentration: Loan Market



Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement.

The increase in concentration is also the result of considerable exit (merger and failure) and entry by banks of small size. Table 1 shows entry and exit statistics by bank size (when sorted by loans).

Table 1: Entry and Exit Statistics by Bank Size (sorted by loans)

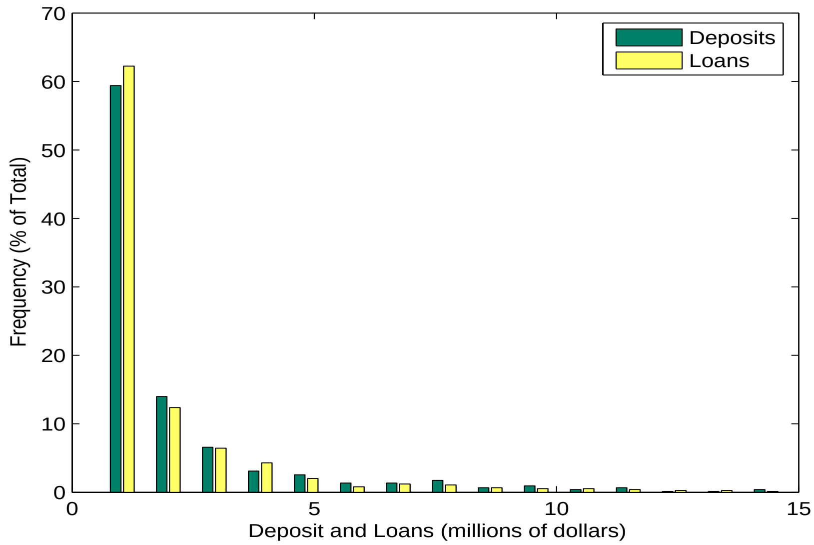
Fraction of Total x, accounted by:	Entry	Exit	Exit by Merger	Exit by Failure
Top 4 Banks	0.00	0.01	0.02	0.00
Top 10 Banks	0.00	0.09	0.16	0.00
Top 1% Banks	0.33	1.07	1.61	1.97
Top 10% Banks	4.91	14.26	16.17	15.76
Bottom 99% Banks	99.67	98.93	98.39	98.03
Fraction of Loans of Banks in x, accounted by:	**Entry**	**Exit**	**Exit by Merger**	**Exit by Failure**
Top 4 Banks	0.00	2.32	2.44	0.00
Top 10 Banks	0.00	9.23	9.47	0.00
Top 1% Banks	21.09	35.98	28.97	15.83
Top 10% Banks	66.38	73.72	47.04	59.54
Bottom 99% Banks	75.88	60.99	25.57	81.14

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Entry and Exit period extends from 1976 to 2010. Merger and Failure period is 1990 - 2007. Let $y \in \{\text{Top 4}, \text{Top 1\%}, \text{Top 10\%}, \text{Bottom 99\%}\}$ and let $x \in \{\text{Enter}, \text{Exit}, \text{Exit by Merger}, \text{Exit by Failure}\}$. Each value in the table is constructed as the time average of "y banks that x in period t" over "total number of banks that x in period t".

We note that the bulk of entry, exit, mergers and failures correspond to banks that are in the bottom 99% of the distribution. The time series average accounted for by the bottom 99% is close to 99% across all categories. The pattern is similar when we measure the fraction of loans in each cathegory accounted for banks of different sizes. In particular, 75% of the loans of entrants and 60.99% of the loans of banks that exit correspond to banks in the bottom 99% of the loan distribution.

The high degree of concentration in the banking industry is the reflection of the presence of a large number of small banks and only a few large banks. In Figure 9, we provide the distribution of deposits and loans for the year 2010. Given the large number of banks at the bottom of the distribution we plot only banking institutions with less than 15 million dollars in deposits (93% of the total). Banks with 1 million dollars of deposits and loans account for approximately sixty percent of the total number of banks. However, total deposits and loans in these banks make up only twenty percent of the total loans and deposits in the industry.

Figure 9: Distribution of Bank Deposits and Loans in 2010



Note: Data corresponds to commercial banks in the US.
Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement.

Table 2 shows measures of deposit and loan concentration for commercial banks in the U.S. for the year 2010.[14] The table shows the high degree of concentration in deposits and loans. It is striking that the the four largest banks (measured by the C_4) hold approximately forty percent of deposits and loans and that the top 1 percent hold 71 and 76 percent of total deposits and loans respectively. We also observe a ratio of mean-to-median of around 10 suggesting sizeable skewness of the distribution. This high degree of inequality is also evident in the Gini coefficient of around 0.9 (recall that perfect inequality corresponds to a measure of 1).

[14] C_4 refers to the top 4 banks concentration index. The Herfindahl Index (HHI) is a measure of the size of firms in relation to the industry and often indicates the amount of competition among them. It is computed as $\sum_{n=1}^{N} s_n^2$ where s_n is market share of bank n. The Herfindahl Index ranges from $1/N$ to one, where N is the total number of firms in the industry.

Table 2: Bank Deposit and Loan Concentration (in 2010)

Measure	Deposits	Loans
Percentage of Total in top 4 Banks (C_4)	36.6	38.2
Percentage of Total in top 10 Banks	46.1	51.7
Percentage of Total in top 1% Banks	71.4	76.1
Percentage of Total in top 10% Banks	87.1	89.6
Ratio Mean to Median	11.1	10.2
Ratio Total Top 10% to Top 50%	91.8	91.0
Gini Coefficient	.91	.90
HHI : Herfindahl Index (National) (%)	5.6	4.3

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Total Number of Banks 6,544. Top 1% banks corresponds to 65 banks. Top 10% banks corresponds to 654 banks.

The national Herfindahl Index is between 4-5%. This is because the largest four banks have the same market share (around 10% each) and there is a large number of firms that have a very small market share (more than 95% of banking institutions have deposit and loan market shares below 1%). The national values are much higher than the values that one would obtain when all firms have equal market shares (i.e. with 1/N, $HHI = 0.13\%$). The national Herfindahl values are a lower bound since they do not consider regional market shares. Bergstresser [8] documents (see his Table 1) that when computed for Metropolitan Statistical Areas (MSA) the Herfindahl Index is much higher (around 20%). Those numbers are typically associated with a highly concentrated industry (values between 10-20%).

If we follow the traditional approach to competition that associates more firms with more price competition and fewer firms with less-competitive behavior, these numbers can be understood as evidence in favor of an imperfectly competitive banking industry. However, an alternative view is one where firms that have higher productive efficiency have lower costs and therefore higher profits. These firms tend to do better and so naturally gain market share, which can lead to concentration. Therefore, by this logic, concentration reflects more efficient banks, not necessarily an increase in market power. For this reason, different approaches have been suggested to attempt to measure the competitive conduct of banks without explicitly using information on the number of firms in the market.

We present several proposed measures computed from our sample of U.S. commercial banks. First, we use the simplest approach and present data on the difference between the realized return on loans and the cost of funds. As it is standard in the banking literature (see for example Boyd and Gertler (1994)) we call this measure the "net interest margin".[15] Second, we present data on markups for the commercial bank industry, the most standard

[15] Note that we partially control for the risk premium charged in interest rates by using realized income from loans instead of ex-ante interest rates. As long as the average observed default frequency is consistent with the default frequency priced into loan contracts, realized interest income provides a good measure of risk-adjusted interest income from loans.

measure of competition. The markup is defined as the ratio of price over marginal cost. Third, following Berger et. al. [7], we present data on the Lerner index, another proxy of the degree of market power. The Lerner index is defined as the difference between price and marginal cost over the price. Fourth, we follow an approach known as contestability that estimates deviations from competitive pricing (i.e. the difference between marginal revenue and marginal cost). One of the most widely used contestability tests is proposed by Panzar and Rosse [36] which essentially tests if the elasticity of marginal revenue with respect to factor prices (marginal cost) is sufficiently below 1 (which is the perfect competition prediction).

We start by formally defining each measure of competition and later we present the results. The markup is defined as

$$\text{Markup}_{it} = \frac{p_{\ell_{it}}}{mc_{\ell_{it}}} - 1 \tag{1}$$

where $p_{\ell_{it}}$ is the price of loans or marginal revenue for bank i in period t and $mc_{\ell_{it}}$ is the marginal cost of loans for bank j in period t.[16] Following Berger et. al. [7], the Lerner index for bank i in period t is defined as follows:

$$\text{Lerner}_{it} = \frac{p_{\ell_{it}} - mc_{\ell_{it}}}{p_{\ell_{it}}} \tag{2}$$

To obtain both, the Markup and the Lerner index, from the data we need to estimate a measure of marginal revenue and marginal cost. Marginal revenue is defined as the sum of the real return on loans and average total non-interest income from loans (computed as total non-interest income from loans divided by loans).[17] The total variable cost for bank i in period t is defined as the sum of the real cost of funds (deposits and securities) plus real non-interest expenses.[18] The total marginal cost is the sum of marginal real cost of funds plus marginal non-interest expenses. The marginal real cost of funds corresponds to the expenses on deposits and securities divided by total deposits and securities (adjusted for inflation, as presented in Figure 4). To derive marginal real non-interest expenses, we

[16] To be consistent with the model we present below, we split commercial banks' output into two categories: loans and other assets (net of premises, fixed assets and cash). Results are similar if we use total assets as the proxy for commercial banks' output (i.e without separating bank's output between loans and securities) as done in other empirical papers.

[17] We deviate from the empirical literature that generally defines marginal revenue as total interest income plus non-interest income divided by loans (or assets) since we deflate the interest return on loans. Our data expands from 1976 to 2010, so adjusting the return on loans for inflation is important because the early 80's was a period of relatively high inflation in the U.S. economy. More specifically, let $\hat{p}_t$ denote the price index in period t, ι_{it} be the nominal interest rate (set in period $t-1$), ℓ the real value of loans and φ be non-interest income. Then, total income from loans divided by loans equals $\frac{\hat{p}_t \ell(\iota_{it}+\varphi)}{\hat{p}_t \ell} = \iota_{it} + \varphi$. Thus, if we let $r_{it} = (1+\iota_{it})/(1+\pi_{it}) - 1$ be the real return on loans, the real marginal revenue equals $r_{it} + \varphi$, i.e. the real return on loans plus the average non-interest income from loans.

[18] Again, as with the return on loans we adjust the cost of funds using CPI inflation. The average fraction of interest expenses as total expenses is approximately 60%.

assume that real Non-interest Expenses take the standard trans-log cost form:

$$
\begin{aligned}
\log(T_{it}) \;=\;& a_i + k_1 \log(w^1_{it}) + h_1 \log(\ell_{it}) + k_2 \log(y_{it}) + k_3 \log(w^1_{it})^2 \\
&+h_2[\log(\ell_{it})]^2 + k_4[\log(y_{it})]^2 + h_3 \log(\ell_{it}) \log(y_{it}) + h_4 \log(\ell_{it}) \log(w^1_{it}) \\
&+k_5 \log(y_{it}) \log(w^1_{it}) + k_6 \log(x_{it}) + \sum_{j=1,2} k_{7,j} t^j + k_{8,t} + \epsilon_{it}
\end{aligned} \tag{3}
$$

where T_{it} is total non-interest expenses minus expenses on premises and fixed assets, w^1_{it} corresponds to input prices (labor), ℓ_{it} corresponds to real loans (one of the two bank j's output), y_{it} represents securities and other assets (the second bank output measured by real assets minus loans minus fixed assets minus cash), x_{it} is equity (a fixed netput), the t regressor refers to a time trend and $k_{8,t}$ refer to time fixed effects. We estimate this equation by panel fixed effects with robust standard errors clustered by bank. The marginal non-interest expense is then computed as:

$$
\frac{\partial T_{it}}{\partial \ell_{it}} = \frac{T_{it}}{\ell_{it}} \Big[h_1 + 2h_2 \log(\ell_{it}) + h_3 \log(y_{it}) + h_4 \log(w^1_{it}) \Big]. \tag{4}
$$

Once we have the estimates of marginal revenue and marginal cost, we can compute the markup and the lerner index for each bank-period as presented in equations (1) and (2).

To obtain our final measure of competition, we follow Shaffer [38] and estimate the elasticity of marginal revenue with respect to factor prices (i.e. the Rose-Panzar H index) by a log-linear regression in which the dependent variable is the natural logarithm of total revenue ($\ln(TR_{it})$ measured as interest income and non-interest income from loans) and the explanatory variables include the logarithms of input prices ($w_{1_{it}}$ funds, $w_{2_{it}}$ labor and $w_{3_{it}}$ fixed assets) and other bank specific factors:

$$
\ln(TR_{it}) = \alpha + \sum_{k=1}^{3} \beta_k \ln(w_{k_{it}}) + \text{Bank Specific Factors}_{it} + u_{it}.
$$

The Rosse-Panzar H equals the simple sum of coefficients on the respective log input price terms, $\beta_1 + \beta_2 + \beta_3$.[19] Bank Specific Factors are additional explanatory variables which reflect differences in risk, cost, size structures of banks and include the value of loans, cash, equity and securities scaled by assets. This equation is estimated by pooled OLS with time fixed effects and robust standard errors.

Table 3 reports the estimated values for the different measures of competition in the U.S. banking industry.

[19]The log-linear form typically improves the regression's goodness of fit and may reduce simultaneity bias.

Table 3: Measures of Competition in Banking

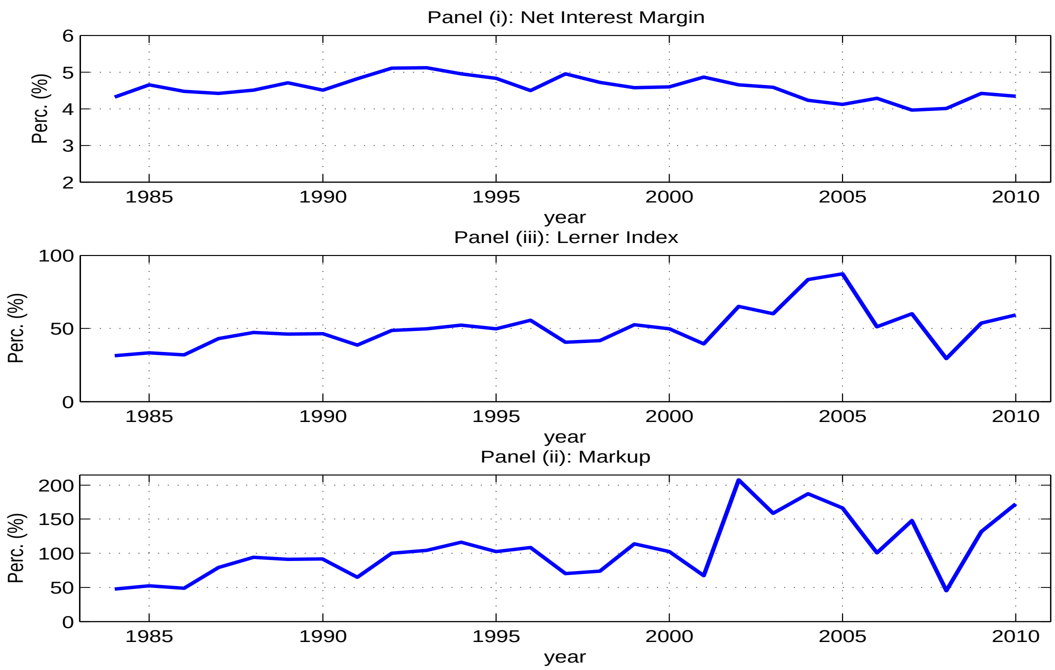
Moment	Value (%)	Std. Error (%)	Corr w/ GDP
Net interest margin	4.56	0.01	-0.31
Lerner Index	43.11	0.38	-0.21
Markup	90.13	1.42	-0.27
Rosse-Panzar H	50.15	0.87	-

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Values correspond to the time series average of the loan weighted cross-sectional averages. Consistent data available from 1984 to 2010.

All the measures presented in Table 3 provide evidence of an imperfectly competitive industry. More specifically, we observe that net interest margins and markups are well above zero. The Lerner index is approximately 49%. This value is similar to the estimated by Berger et.al (2008) for a different U.S. sample. Recall that under perfect competition marginal revenue equals marginal cost so both the markup and Lerner index equal zero in that case. Finally, the Rose-Panzar H measure is statistically different from 100 (the value that indicates the presence of perfect competition) with 99% confidence. Using this technique, Bikker and Haaf [10] estimate the degree of competition in the banking industry for a panel of 23 (mostly developed) countries. They find that for all slices of the sample, perfect competition can be rejected convincingly, i.e. at the 99% level of confidence. The value estimated for the U.S. banking industry in their paper ranges from 54% to 56%, very close to our estimate reported in Table 3. In summary, taken together these measures suggest the banking industry is less than perfectly competitive.

We are also interested in the cyclical properties of net interest margin, markups, and the Lerner index. Table 3 shows that interest margins, markups and the Lerner index are countercyclical with correlation with detrended GDP equal to -0.31, -0.21 and -0.27 respectively. These values are consistent with evidence presented in Aliaga Diaz and Olivero [1]. Figure 10 shows the evolution of these measures and detrended GDP.

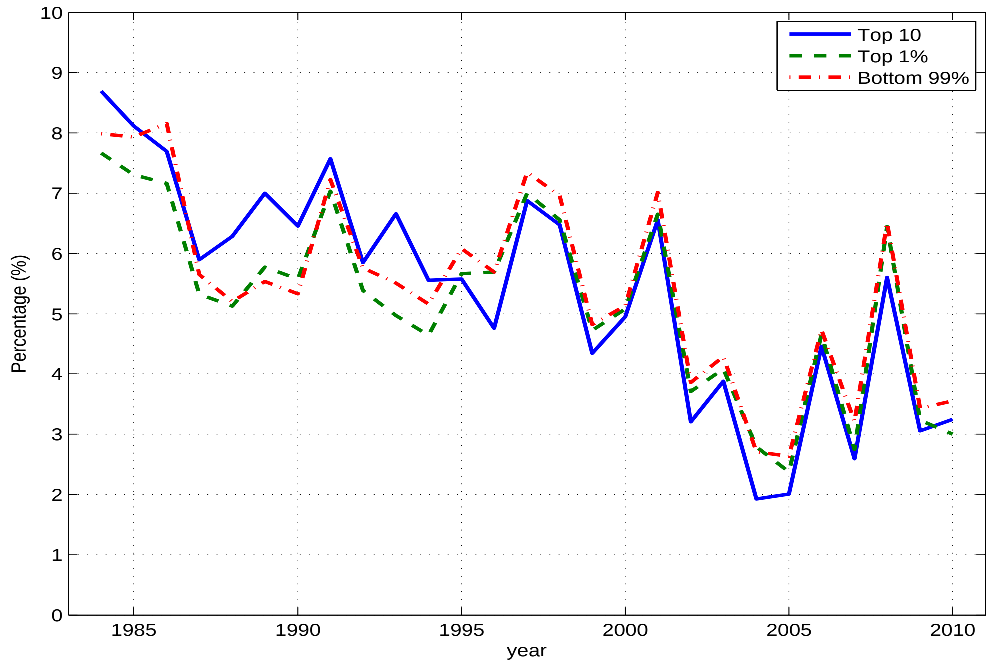
Figure 10: Evolution of Net Interest Margin, Markups and Lerner Index



Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Net interest margin is the difference between the return on loans minus the cost of deposits. Markups are defined as $\frac{p_{\ell_{ti}}}{mc_{\ell_{ti}}} - 1$ where $p_{\ell_{ti}}$ refers to marginal revenue and $mc_{\ell_{ti}}$ to marginal cost. The Lerner index is $\frac{p_{\ell_{ti}} - mc_{\ell_{ti}}}{p_{\ell_{ti}}}$.

In Figure 11, we analyze the evolution of loan returns by bank size (when sorted by loans). For most periods in the sample and in particular after the period of banking deregulation, we note that small banks have higher returns than big banks.

Figure 11: Loan Returns by Bank Size



Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Each year observation corresponds to the weighted cross-sectional average of loan returns for the particular bank group in the given year. Bank size corresponds to the position of the bank in the loan distribution. Top 1% Banks do not include the Top 10 Banks.

We use our rich panel data set to conduct a deeper analysis on loan returns and its standard deviation (a measure of how diversified banks are).[20] We estimate loan returns and standard deviation of loan returns for bank i in period t as a function of bank size. Banks are grouped into "Top 10", "Top 1%" and "Bottom 99%" using the distribution of loans. Size dummies are created using these categories and used as regressors leaving out the smallest class. Panel (a) of Table 4 presents the average return, its standard deviation and the correlation with GDP by bank size. Panel (b) of Table 4 presents the statistical tests of significant differences.

[20] The standard deviation for bank i in period t is constructed using a 5 period rolling window using the observed loan returns in periods $t-4$ to t. Results are robust to using 3 and 7 periods.

Table 4: Loan Return and Volatility by Bank Size

Panel (a): Size Coefficients

Loan Returns	Avg.(%)	Std. Dev. (%)	Corr. with GDP
Top 10 Banks	$5.24^{*,\dagger}$	$1.14^{*,\dagger}$	-0.24
Top 1% Banks	$5.46^{\dagger}$	$1.20^{\dagger}$	-0.29
Bottom 99% Banks	6.05	1.22	-0.29

Note: * Denotes statistically significant difference with Top 1% value.
† Denotes statistically significant difference with Bottom 99% value.

Panel (b): Tests of Size Effect

Loan Returns	Avg.	Std. Dev.
H_0 : Top 10 equals Top 1% coeff.		
$F-$stat	3.01	3.41
$p-$value	0.08	0.06
H_0 : Top 1% equals Bottom 99% coeff.		
$t-$stat	-11.99	-1.67
$p-$value	0.00	0.09
H_0 : Top 10 equals Bottom 99% coeff.		
$t-$stat	-5.45	-2.68
$p-$value	0.00	0.00

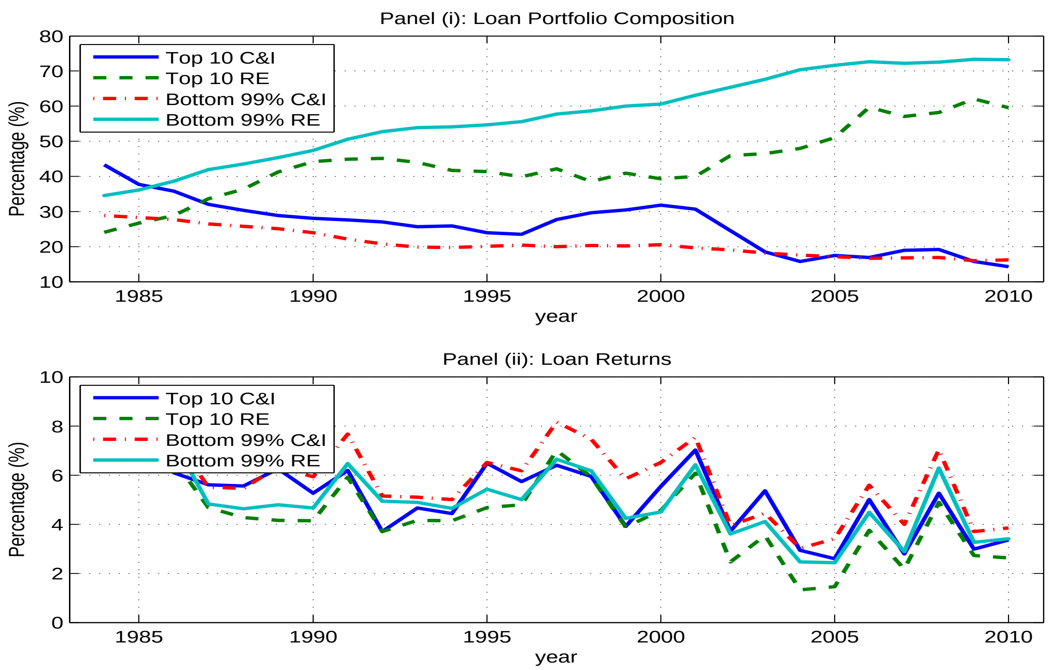
Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Bank size corresponds to the position of the bank in the loan distribution. Consistent data available from 1984 to 2010.

Panel (a) shows that smaller banks have higher returns and higher volatility of returns but larger banks have a stronger negative correlation with detrended GDP. The results of the tests in Panel (b) imply that loan returns and its standard deviation are significantly different (at the 10% significance level) between Top 10 banks and banks in the Top 1% and Bottom 99%. The same is true between banks in the Top 1% and banks in the Bottom 99% group. An important component of loan returns is the fraction of loans that are not repaid. For this reason, in Figures 13 and 14 we present charge-off rates and delinquency rates by bank size over time.

Independent evidence for the benefits of geographic diversification associated with bigger banks is given in Liang and Rhoades [33]. They test the hypothesis that geographic diversification lowers bank risk by regressing alternative measures of risk like the probability of bank insolvency, probability of failure, and the standard deviation of net income-to-assets on, among other controls, geographic diversification proxied by the inverse of the sum of squares of the percentage of a bank's deposits in each of the markets in which it operates. They find (see their Table 1) that the standard deviation of net-income-to-assets is significantly (both statistically and quantitatively) lower for firms that operate in a greater number of geographic markets. This will be consistent with our model.

Changes in the aggregate composition of the loan portfolio that we described before are also present for banks of different sizes (when sorted by loans). In Panel (i) of Figure 12, we document this trend for the largest 10 banks and the bottom 99% "small" banks when sorted by loans. We compute the share of total loans that corresponds to commercial and industrial loans ($C\&I$) and real estate loans (RE) for each bank and plot the weighted average of these shares for each group and year.[21] Panel (ii) of Figure 12 shows the loan return by bank size and loan type.

Figure 12: Loan Portfolio Composition and Loan Returns by Loan Type and Bank Size



Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement. Each year observation corresponds to the weighted cross-sectional average of loan returns for the particular group and loan class in the given year. Bank size corresponds to the position of the bank in the loan distribution.

We observe that for both small and large banks, loans secured by real estate have become much more important. In the case of small banks, the share of loans secured by real estate more than doubled during this period (the share went from approximately 35 percent to more than 70 percent). A similar trend is observed for the largest banks. The share of real estate loans in their portfolio increased from 33 percent to 60 percent. For this group of banks

[21] A consistent series for $C\&I$ loans at the individual bank level is only available since 1984.

we also note a faster increase in this share during the last decade. The counterpart of the increase in real estate loans is the decrease in the share of commercial and industrial loans. We note one of the differences between small and big banks is the portfolio composition. For most of the period (since 1990), loans secured by real estate constitute a more important component for small banks than for big banks. The opposite is true for commercial and industrial loans. Finally, Panel (*ii*) in Figure 11, shows that small banks have higher returns for both real estate and commercial and industrial loans than big banks (top 4).

Both charge-off rates and delinquency rates have an important negative cyclical component. Figure 13 presents the evolution over time. For most periods in the sample, charge-off rates and delinquency rates are higher for big banks than for small banks.

Figure 13: Charge-Off Rates by Bank Size (when sorted by loans)

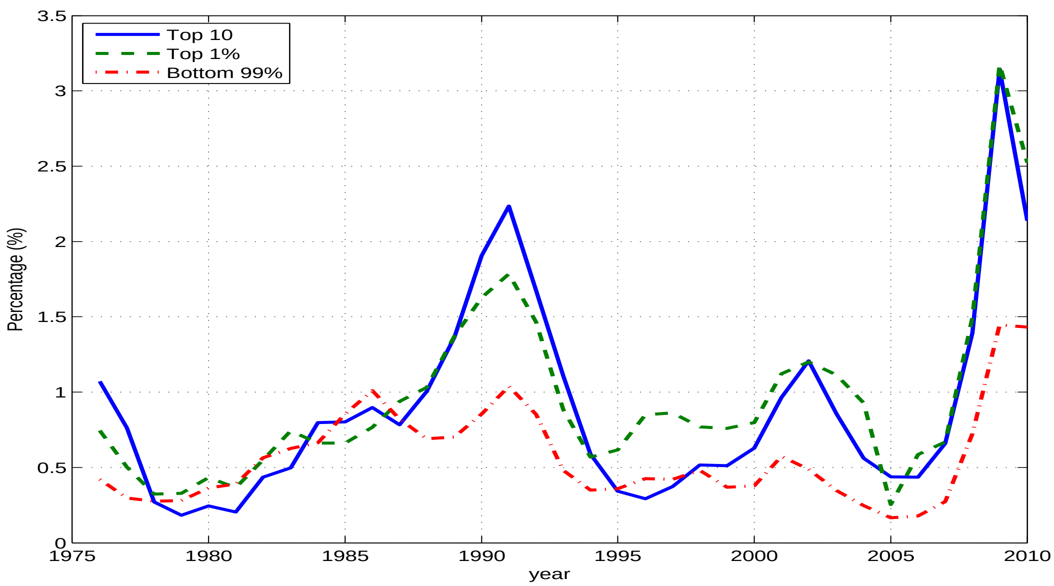


Figure 14: Delinquency Rates by Bank Size (when sorted by loans)

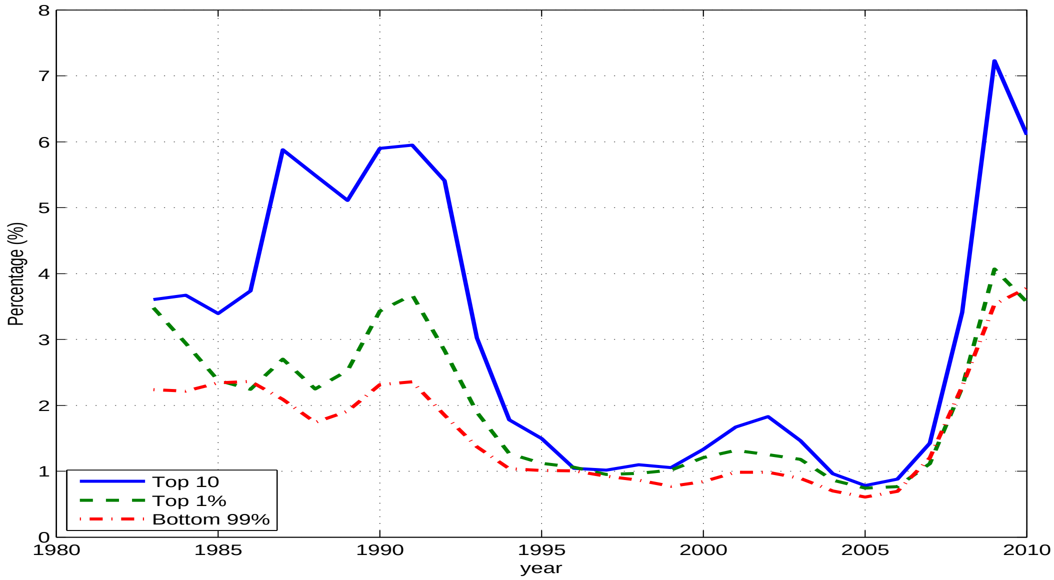


Table 5 presents a formal test of differences in the level of charge off rates and delinquency rates by bank size (as well as their standard deviation). We use our panel to estimate charge off rates and delinquency rates for bank i in period t as a function of the size dummies. The standard deviation in period t is computed using charge off rates and delinquency rates for years $t-4$ to t.

Table 5: Charge-Offs and Delinquency Rates by Bank Size (sorted by loans)

Panel (a): Size Coefficients

Moment	Avg. (%)	Std. Dev. (%)	Corr. with GDP
Charge Off Rate Top 10 Banks	$1.06^{\dagger}$	0.41	-0.51
Charge Off Rate Top 1% Banks	$1.00^{\dagger}$	$0.40^{\dagger}$	-0.57
Charge Off Rate Bottom 99% Banks	0.57	0.42	-0.54
Del. Rate Top 10 Banks	$2.82^{*,\dagger}$	0.85^{*}	-0.59
Del. Rate Top 1% Banks	$1.93^{\dagger}$	$0.66^{\dagger}$	-0.56
Del. Rate Bottom 99% Banks	1.64	0.81	-0.49

Note: * Denotes statistically significant difference with Top 1% value.
† Denotes statistically significant difference with Bottom 99% value.

Panel (b): Tests of Size Effect

	Charge Off	Del. Rate
H_0 : Top 10 equals Top 1% coeff.		
$F-$stat	0.69	38.99
$p-$value	0.41	0.00
H_0 : Top 1% equals Bottom 99% coeff.		
$t-$stat	15.16	7.81
$p-$value	0.00	0.00
H_0 : Top 10 equals Bottom 99% coeff.		
$t-$stat	6.29	8.64
$p-$value	0.00	0.00

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Bank size corresponds to the position of the bank in the loan distribution. Consistent data available from 1984 to 2010.

Table 5 results are consistent with the pattern documented in Figures 13 and 14. Charge off rates and delinquency rates for Top 10 and Top 1% banks are statistically higher than those computed for Bottom 99% banks. There are two factors to consider when reading the results in Table 5. First, delinquency rates are computed as the ratio of the value of loans that are delinquent (90 days or more plus those in nonaccrual status) to total loans. Thus, to be more precise, the statistics we display correspond to the fraction of delinquent loans. The default frequency (i.e. the ratio of the number of delinquent loans to the total number of loans) and the fraction of delinquent loans coincide only when all loans are of the same size. Second, there is a selection effect present in the data. We observe only active banks and, as we showed above, most exit happens for banks in the Bottom 99% group. Thus, we observe only those banks in the Bottom 99% group with low default rates.

Table 3 presented evidence on the level of competition at the aggregate level. In Table 6 we disaggregate net interest margins, markups and the lerner index by banks size.

Table 6: Net Interest Margins, Markups and Lerner Index by Bank Size

Panel (a): Size Coefficients

Moment (in %)	Net Int.	Markups	Lerner Index
Top 10 Banks (%)	$4.19^{*,\dagger}$	$70.38^{*,\dagger}$	$42.94^{*,\dagger}$
Top 1% Banks (%)	$4.51^{\dagger}$	$96.72^{\dagger}$	$45.11^{\dagger}$
Bottom 99% Banks (%)	5.17	146.06	54.70

Note: * Denotes statistically significant difference with Top 1% value.
† Denotes statistically significant difference with Bottom 99% value.

Panel (b): Tests of Size Effect

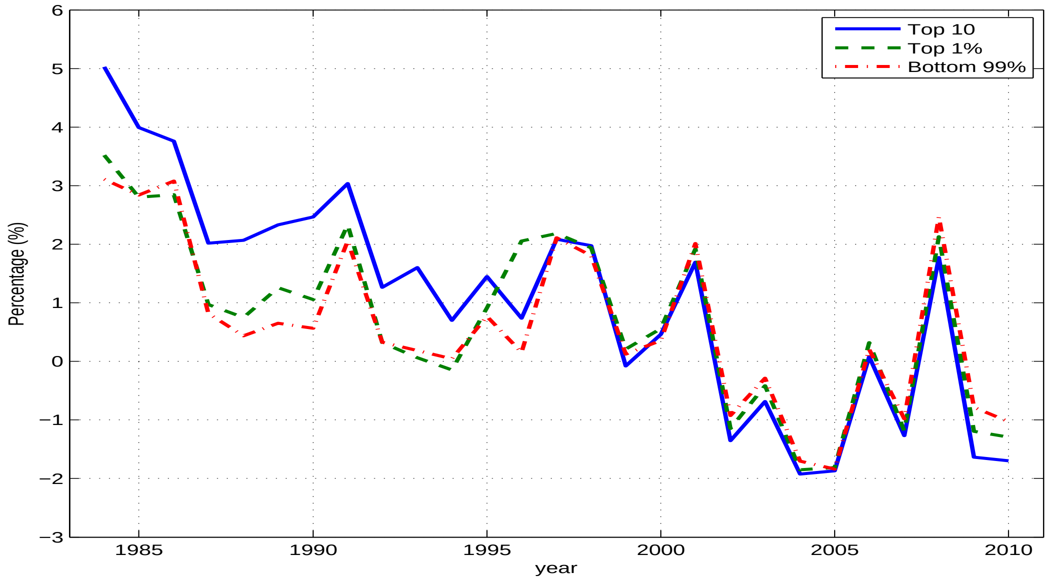
	Net Int.	Markups	Lerner Index
H_0 : Top 10 equals Top 1% coeff.			
$F-$stat	6.23	6.03	2.76
$p-$value	0.01	0.01	0.09
H_0 : Top 1% equals Bottom 99% coeff.			
$t-$stat	-6.90	-8.09	-26.57
$p-$value	0.00	0.00	0.00
H_0 : Top 10 equals Bottom 99% coeff.			
$t-$stat	-12.03	-8.53	-9.22
$p-$value	0.00	0.00	0.00

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Bank size corresponds to the position of the bank in the loan distribution. Consistent data available from 1984 to 2010.

Bigger banks have lower net interest margins, markups and lerner index. This is consistent with bigger banks lower loan returns (see Table 4) and with a selection effect. We only observe those banks that remain active and most exit happens for small banks, so we the small banks that remain active have a low default frequency and high margins.

One important component of the cost structure of banks is the cost of funds (deposits and securities). Figure 15 shows the evolution of the cost of funds across bank sizes. This figure shows that they are very similar (specially since 1995). After conducting a formal test, we find that there are no statistical differences across size classes.

Figure 15: Cost of Funds by Bank Size (when sorted by loans)



Banks of different sizes differ in their non-interest income and expenses. It is important to consider these differences since their relevance as a fraction of total profits has been rising during the past three decades. We present the estimates of average non interest income from loans and the marginal non-interest expenses that we constructed to obtain measures of marginal cost and marginal revenue. We define net expenses as marginal non interest expenses minus non interest income from loans. We use our panel and size dummies to obtain the values presented in Table 7.

Table 7: Marginal Non-Interest Income, Expense and Net Expense

	Non-Int Inc.	Non-Int Exp.	Net Exp.
Top 10 Banks (%)	$2.41^{\dagger}$	$4.19^{*,\dagger}$	$1.78^{*,\dagger}$
Top 1 % Banks (%)	$2.30^{\dagger}$	$3.91^{\dagger}$	1.61
Bottom 99 % Banks (%)	0.89	2.48	1.59

Note: * Denotes statistically significant difference with Top 1% value. † Denotes statistically significant difference with Bottom 99% value.

Panel (b): Tests of Size Effect

	Non Int. Inc.	Non Int. Exp.	Net Exp.
H_0 : Top 10 equals Top 1% coeff.			
$F-$stat	0.16	5.41	7.71
$p-$value	0.68	0.02	0.01
H_0 : Top 1% equals Bottom 99% coeff.			
$t-$stat	24.58	27.57	0.39
$p-$value	0.00	0.00	0.69
H_0 : Top 10 equals Bottom 99% coeff.			
$t-$stat	25.89	15.71	3.10
$p-$value	0.00	0.00	0.00

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Net expense is calculated from our measure of marginal cost as marginal cost - cost of funds - non-interest income from loans. Consistent data available from 1984 to 2010.

It is evident from Table 7 that net expenses are increasing in size. While large banks perceive higher Non-Interest income than small banks, their non-Interest expenses is also higher resulting in the observed pattern of net-expenses.

The final piece of the cost structure of commercial banks in the U.S. that remains to be estimated corresponds to fixed costs. We defined fixed costs as total expenses on premises and fixed assets. Table 8 presents the estimates scaled by total assets at the bank level.

Table 8: Fixed Costs (as a fraction of loans)

	Fixed Cost (%)	Std. Error (%)
Top 10 Banks (%)	$0.485^{*,\dagger}$	0.02
Top 1 % Banks (%)	$0.427^{\dagger}$	0.01
Bottom 99 % Banks (%)	0.459	0.01

Note: * Denotes statistically significant difference with Top 1% value. † Denotes statistically significant difference with Bottom 99% value. Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports. Consistent data available from 1984 to 2010.

3 Model Environment

Time is infinite. There are two regions $j \in \{e, w\}$, for instance "east" and "west". Each period, a mass B of one period lived ex-ante identical borrowers and a mass $H = 2B$ of one period lived ex-ante identical households (who are potential depositors) are born in each region.[22]

3.1 Borrowers

Borrowers in region j demand bank loans in order to fund a project. The project requires one unit of investment at the beginning of period t and returns at the end of the period:

$$\begin{cases} 1 + z_{t+1}R_t^j & \text{with prob } p^j(R_t^j, z_{t+1}, s_{t+1}) \\ 1 - \lambda & \text{with prob } [1 - p^j(R_t^j, z_{t+1}, s_{t+1})] \end{cases} \tag{5}$$

in the successful and unsuccessful states respectively. Borrower gross returns are given by $1 + z_{t+1}R_t^j$ in the successful state and by $1 - \lambda$ in the unsuccessful state. The success of a borrower's project in region j, which occurs with probability $p^j(R_t^j, z_{t+1}, s_{t+1})$, is independent across borrowers but depends on several things: the borrower's choice of technology $R_t^j \geq 0$, an aggregate technology shock at the end of the period z_{t+1}, and a regional shock s_{t+1} (the dating convention we use is that a variable which is chosen/realized at the end of the period is dated $t + 1$).

The aggregate technology shock is denoted $z_t \in \{z_b, z_g\}$ with $z_b < z_g$ (i.e. good and bad shocks). This shock evolves as a Markov process $F(z', z) = prob(z_{t+1} = z'|z_t = z)$. The regional specific shock $s_{t+1} \in \{e, w\}$ also evolves as a Markov process $G(s', s) = prob(s_{t+1} = s'|s_t = s)$ which is independent of z_{t+1}.

At the beginning of the period when the borrower makes his choice of R_t both z_{t+1} and s_{t+1} have not been realized. As for the likelihood of success or failure, a borrower

[22] The assumption $H = 2B$ is a normalization that simplifies the analysis below. Furthermore, the assumption that borrowers and depositors are one period lived is simply to restrict attention to one period loan and deposit contracts rather than to resort to anonymity as in, for instance, Carlstrom and Fuerst [12].

who chooses to run a project with higher returns has more risk of failure and there is less failure in good times. Specifically, $p^j(R_t^j, z_{t+1}, s_{t+1})$ is assumed to be decreasing in R_t^j and $p^j(R_t^j, z_g, s_{t+1}) > p^j(R_t^j, z_b, s_{t+1})$. Moreover, we assume that the borrower success probability depends positively on which region $s_{t+1} \in \{e, w\}$ receives a favorable shock. Specifically, $p^{j=s_{t+1}}(R_t^j, z_{t+1}, s_{t+1}) > p^{j \neq s_{t+1}}(R_t^j, z_{t+1}, s_{t+1})$. That is, in any period, one region has a higher likelihood of success than the other. While the aggregate shock cannot be diversified away, the negative correlation between regional shocks provides a role for cross-region diversification.

While borrowers in a given region are ex-ante identical, they are ex-post heterogeneous owing to the realizations of the shocks to the return on their project. We envision borrowers either as firms choosing a technology which might not succeed or households choosing a house that might appreciate or depreciate.

There is limited liability on the part of the borrower. If $r_t^{L,j}$ is the interest rate on bank loans that borrowers face in region j, the borrower receives $\max\{z_{t+1}R_t^j - r_t^{L,j}, 0\}$ in the successful state and 0 in the failure state. Specifically, in the unsuccessful state he receives $1-\lambda$ which must be relinquished to the lender. Table 9 summarizes the risk-return trade off that the borrower faces.

Table 9: Borrower's Problem

Borrower chooses R^j	Receive	Pay	Probability			
				$-$	$+$	$+$
Success	$1+z'R^j$	$1+r^{L,j}(\mu,z,s)$	p^j	$(R^j,$	$z',$	$s')$
Failure	$1-\lambda$	$1-\lambda$	$1-p^j$	$(R^j,$	$z',$	$s')$

Borrowers have an outside option (reservation utility) $\omega_t \in [\underline{\omega}, \overline{\omega}]$ drawn at the beginning of the period from distribution function $\Omega(\omega_t)$.

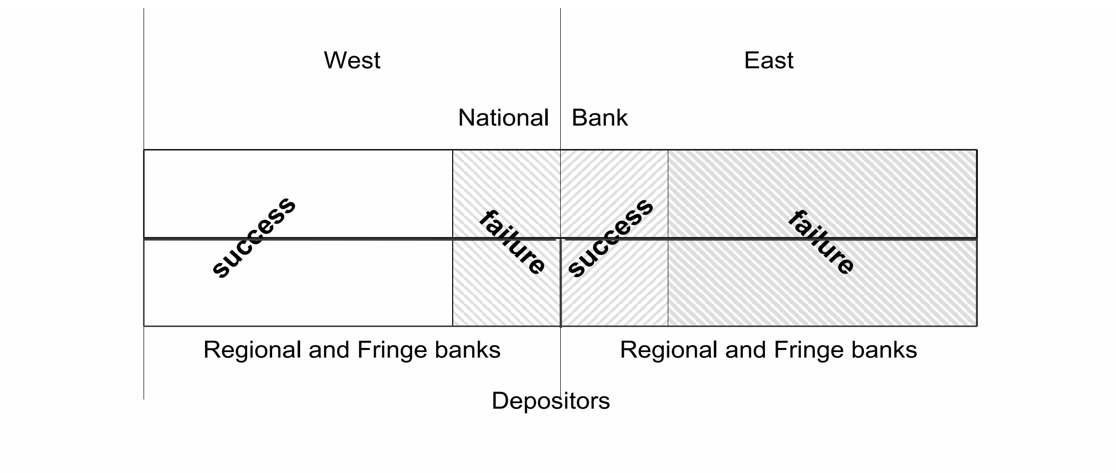
3.2 Depositors

All households are endowed with 1 unit of the good and have preferences denoted $u(C_t)$. All households have access to a risk free storage technology yielding $1+\overline{r}$ with $\overline{r} \geq 0$ at the end of the period. They can also choose to supply their endowment to a bank in their region or to an individual borrower. If the household deposits its endowment with a bank, they receive $r_t^{D,j}$ whether the bank succeeds or fails since we assume deposit insurance. If they match with a borrower, they are subject to the random process in (5). At the end of the period they pay lump sum taxes τ_{t+1} which are used to cover deposit insurance for failing banks.

3.3 Banks

Motivated by the data described in Section 2, we assume there are three classes of banks $\theta \in \{n, r, f\}$ for national, regional, and fringe respectively. National banks are geographically diversified in the sense that they extend loans and receive deposits in both $\{e, w\}$ regions. Regional banks are restricted to make loans and receive deposits in one geographical area (i.e. either e or w). Fringe banks are also restricted on one geographical area (i.e. either e or w). Since we allow regional specific shocks to the success of borrower projects, regional and fringe banks may not be well diversified.[23] This assumption can, in principle, generate ex-post differences in loan returns documented in the data section. A bank's type is represented by the two-tuple $(\theta, \{e, w\})$ where, for instance (r, e) denotes an eastern regional bank. See Figure 16 for a graphical description of the regional segmentation in the model.

Figure 16: Regional Segmentation



We denote loans made by bank i of type (θ, j) to borrowers at the beginning of period t by $\ell_{i,t}(\theta, j)$ and accepted units of deposits by $d_{i,t}(\theta, j)$. All banks have the same linear technology for producing loans. Without an interbank market, if i is a regional or fringe bank then $\ell_{i,t}(\theta, j) \leq d_{i,t}(\theta, j)$. If i is a national bank then $\ell_{i,t}(n, e) + \ell_{i,t}(n, w) \leq d_{i,t}(n, e) + d_{i,t}(n, w)$. We assume that national and regional banks do not face any restriction on the number of deposits they can accept in their region. On the other hand, fringe banks face a capacity constraint $\bar{d}$ of available deposits. Since fringe banks take prices as given, their expected profit function is linear in the amount of loans they extend, so we need to impose this capacity constraint in order to prevent the amount of loans of a fringe bank from exceeding the total amount of deposits in the region.

The timing in the loan stage follows the standard treatment of the dominant firm model (see for example Gowrisankaran and Holmes [24]). The dominant firms, our national and regional banks, move first. They compete in a Cournot fashion and choose quantities $\ell_{i,t}(\theta, j)$ taking as given not only the reaction function of other dominant banks but also the loan supply of the competitive fringe. Each fringe bank observes the total loan supply of dominant banks and all other fringe banks (that jointly determine the loan interest rate $r_t^{L,j}$) in region

[23]In an interesting paper, Koeppl and MacGee [31] consider whether a model with regional banks which operate within a region with access to interbank markets can achieve the same allocation under uncertainty as a model with national banks which operate across regions.

j and simultaneously decide on the amount of loans to extend. Since, at a given interest rate, the production technology is linear in loans supplied, the fringe banks decision reduces simply to whether to bring all their available funds to the market or not, i.e. $\ell_{i,t}(f,j) \in \{0, \overline{d}\}$.

In principle one could also have banks be Cournot competitors in the deposit market as in Boyd and DeNicolo [11]. However, since we assume that $H > 2B$ there are sufficient funds to cover all possible loans if banks offer the lowest possible deposit rate $r_t^{D,j} \geq \overline{r}$.

In Section 2, we documented important differences in non-interest income and non-interest expenses across banks of different sizes. Based on this evidence, we assume that banks receive proportional non-interest income $c^{\theta,inc}$ and pay proportional non-interest expenses $c^{\theta,exp}$ that differ across bank type. We denote net-expenses by $c^\theta = c^{\theta,exp} - c^{\theta,inc}$ and assume that all national and regional banks face the same net expenses c^n and c^r, respectively.[24] Net-expenses for fringe bank i is denoted c_i^f which is drawn from a distribution with cdf $\Xi(c^f)$. For simplicity, we assume that costs are constant over the lifespan of the bank and they are identical across regions. Differences in fixed costs are also evident in the data presented in Section 2. Consistent with this, we assume that banks pay fixed costs κ^θ every period.

Entry costs for the creation of national and regional banks are denoted by $\Upsilon^n \geq \Upsilon^r \geq 0$.[25] Entry costs correspond to the initial injection of equity into the bank by its owner. We normalize the cost of creating a fringe bank to zero. Every period a large number of potential banks M make the decision to enter the market or not. We assume that each entrant satisfies a zero expected discounted profits condition. To simplify the analysis, we assume that entering fringe banks draw of c^f exceeds the highest cost of fringe incumbents in the market that period. This assumption makes the computation much easier since the only relevant variable to predict the number of active fringe banks is the threshold of the active bank with the highest cost. In any given period, there are M fringe banks potentially ready to extend loans. This allows us to track the entire distribution of banks by simply keeping track of the distribution of dominant firms and a moment (that is a sufficient statistic) of the distribution of fringe banks. Without such an assumption, we would have to generalize an algorithm proposed by Farias, Ifrach, and Weintraub [23] as we do in Corbae and D'Erasmo (2012).

There is limited liability on the part of banks. As in Cooley and Quadrini [13] and Hennesy and Whited [25], we assume that banks with negative profits have access to outside funding or equity financing at cost $\xi^\theta(x)$ per x units of funds raised, where $\xi^\theta(x)$ is an increasing function. The benefit of introducing external financing of this form is that it allows us to consider a problem where banks face a dynamic exit decision (i.e. one where a non-negative future value of the bank plays a role in the exit decision) without the need of incorporating an extra state variable. A bank that has negative expected continuation value can exit, in which case it receives value zero. We assume that if a national bank exits, it must exit both regions.

[24] As it will be clear below, net-expenses c^θ is the only relevant variable to solve the bank problem. However, non-interest income $c^{\theta,inc}$ and non-interest expense $c^{\theta,exp}$ become important to bring the model to the data and we present results on markups and the Lerner index.

[25] As in Pakes and McGuire [35] we will assume that these costs become infinite after a certain number of firms of the given type are in the market.

An injection of equity would never be optimal for our one period lived consumers because they do not benefit from the charter value of the bank. For this reason, and to keep the model tractable, we assume that banks are owned by infinitely lived risk-neutral investors who discount the future at rate β and have a large endowment (i.e. they have "deep pockets" and do not face any borrowing constraint). Investors choose the number of shares to buy in each bank (incumbent and newly created) to maximize their expected sum of present discounted value of current and future cash flows. We abstract from agency problems, so the objective of the individual bank is aligned with that of investors, i.e. they maximize the expected discounted sum of dividends (net of equity financing costs in cases where profits are negative but charter value is non-negative) and discount the future at rate β. In Appendix A-2, we present a full description of the problem of the investor.

We denote the industry state by

$$\mu_t = \{\mu_t^{(n,\cdot)}, \mu_t^{(r,e)}, \mu_t^{(r,w)}, \mu_t^{(f,e)}, \mu_t^{(f,w)}\}, \tag{6}$$

where each element of μ_t is a measure $\mu_t^{(\theta,j)}$ corresponding to *active* banks of type θ in region j. It should be understood that $\mu_t(\theta, j)$ is a counting measure for $\theta \in n, r$.

3.4 Information

There is asymmetric information on the part of borrowers and lenders. Only borrowers know the riskiness of the project they choose (R) and their outside option (ω). All other information is observable.

3.5 Timing

At the beginning of period t,

1. Starting from beginning of period state (μ_t, z_t, s_t), borrowers draw ω_t.
2. National and regional banks choose how many loans $\ell_{i,t}(\theta, j)$ to extend and how many deposits $d_{i,t}(\theta, j)$ to accept.
3. Fringe banks in each region choose loan supply and how many deposits to accept. Borrowers in region j choose whether or not to undertake a project of technology R_t^j. Depositors in each region decide where to deposit.
4. Return shocks z_{t+1} and s_{t+1} are realized, as well as idiosyncratic borrower shocks.
5. Equity finance and exit decisions are made in that order.[26]
6. Entry occurs sequentially (national and regional then fringe).[27]

[26]Consistent with Appendix A-2 investors receive dividends net of equity financing costs in cases where profits are negative but charter value is non-negative for banks that survive.

[27]Consistent with Appendix A-2, in the entry period the dividend to investors takes the form of financing entry costs -Υ_i^θ in exchange for a unit share of the firm.

7. Households pay taxes τ_{t+1} and consume.[28]

4 Industry Equilibrium

Since we will use recursive methods to define an equilibrium, let any variable a_t be denoted a and a_{t+1} be denoted a'.

4.1 Borrower Decision Making

Starting in state z, borrowers take the loan interest rate $r^{L,j}$ as given and choose whether to demand a loan and if so, what technology R^j to operate. Specifically, if a borrower in region j chooses to participate, then given limited liability his problem is to solve:

$$v(r^{L,j}, z, s) = \max_{R^j} E_{z',s'|z,s}\Big[p^j(R^j, z', s')\left(z'R^j - r^{L,j}\right)\Big]. \tag{7}$$

Let $R(r^{L,j}, z, s)$ denote the borrower's decision rule that solves (7). We assume that the necessary and sufficient conditions for this problem to be well behaved are satisfied. The borrower chooses to demand a loan if

$$v(r^{L,j}, z, s) \geq \omega. \tag{8}$$

In an interior solution, the first order condition is given by

$$E_{z',s'|z,s}\Big\{\underbrace{p^j(R^j, z', s')z'}_{(+)} + \underbrace{\frac{\partial p^j(R^j, z', s')}{\partial R^j}}_{(-)}\left[z'R^j - r^{L,j}\right]\Big\} = 0 \tag{9}$$

The first term is the benefit of choosing a higher return project while the second term is the cost associated with the increased risk of failure.

To understand how bank lending rates influence the borrower's choice of risky projects, one can totally differentiate (9) with respect to $r^{L,j}$

$$\begin{aligned} 0 \;=\; & E_{z',s'|z,s}\left\{\frac{\partial p^j(R^{j*}, z', s')}{\partial R^{j*}}\frac{dR^{j*}}{dr^{L,j}}z' + \frac{\partial^2 p^j(R^{j*}, z', s')}{\left(\partial R^{j*}\right)^2}\left[z'R^{j*} - r^{L,j}\right]\frac{dR^{j*}}{dr^{L,j}}\right. \\ & \left. + \frac{\partial p^j(R^{j*}, z', s')}{\partial R^{j*}}\left[z'\frac{dR^{j*}}{dr^{L,j}} - 1\right]\right\} \end{aligned}$$

where $R^{j*} = R^j(r^{L,j}, z)$. But then

$$\frac{dR^{j*}}{dr^{L,j}} = \frac{E_{z',s'|z,s}\left[\frac{\partial p^j(R^{j*},z',s')}{\partial R^{j*}}\right]}{E_{z',s'|z,s}\left\{\frac{\partial^2 p^j(R^{j*},z',s')}{\left(\partial R^{j*}\right)^2}\left[z'R^{j*} - r^{L,j}\right] + 2\frac{\partial p^j(R^{j*},z',s')}{\partial R^{j*}}z'\right\}} > 0 \tag{10}$$

[28] Consistent with Appendix A-2, the investor re-balances its portfolio of bank shares and consumes.

since both the numerator and the denominator are strictly negative (the denominator is negative by virtue of the sufficient conditions). Thus a higher borrowing rate implies the borrower takes on more risk. Boyd and De Nicolo [11] call $\frac{dR^{j*}}{dr^{L,j}} > 0$ in (10) the "risk shifting effect". Risk neutrality and limited liability are important for this result.

An application of the envelope theorem implies

$$\frac{\partial v(r^{L,j}, z, s)}{\partial r^{L,j}} = -E_{z',s'|z,s}[p^j(R^j, z', s')] < 0. \tag{11}$$

Thus, participating borrowers are worse off the higher are borrowing rates. This has implications for the demand for loans determined by the participation constraint. In particular, since the demand for loans is given by

$$L^{d,j}(r^{L,j}, z, s) = B \cdot \int_{\underline{\omega}}^{\overline{\omega}} 1_{\{\omega \leq v(r^{L,j},z,s)\}} d\Omega(\omega), \tag{12}$$

then (11) implies $\frac{\partial L^{d,j}(r^{L,j},z,s)}{\partial r^{L,j}} < 0$.

4.2 Depositor Decision Making

If $r^{D,j} = \overline{r}$, then a household would be indifferent between matching with a bank and using the autarkic storage technology so we can assign such households to a bank. If it is to match directly with a borrower, the depositor must compete with banks for the borrower. Hence, in successful states, the household cannot expect to receive more than the bank lending rate $r^{L,j}$ but of course could choose to make a take-it-or-leave-it offer of their unit of a good for a return $\widehat{r} < r^{L,j}$ and hence entice a borrower to match with them rather than a bank. Given state contingent taxes $\tau(\mu, z, s, z', s')$, the household matches with a bank if possible and strictly decides to remain in autarky otherwise provided

$$\begin{aligned} U \equiv \quad & E_{z',s'|z,s}\left[u(1+\overline{r}-\tau(\mu, z, z', s'))\right] > \\ & \max_{\widehat{r}<r^{L,j}} E_{z',s'|z,s}\Big[p^j(\widehat{R}^j, z', s')u(1+\widehat{r}-\tau(\mu, z, s, z', s')) \\ & +(1-p^j(\widehat{R}^j, z', s'))u\left(1-\lambda-\tau(\mu, z, s, z', s')\right)\Big] \equiv U^E. \end{aligned} \tag{13}$$

If this condition is satisfied, then the total supply of deposits in region j is given by

$$D^{s,j} = \sum_{\theta} \int d_i(\theta, j)\mu^{(\theta,j)}(di) \leq H \tag{14}$$

Condition (13) makes clear the reason for a bank in our environment. By matching with a large number of borrowers, the bank can diversify the risk of project failure and this is valuable to risk averse households. It is the loan side uncertainty counterpart of a bank in Diamond and Dybvig [18].

4.3 Incumbent Bank Decision Making

An incumbent bank i of type (θ, j) chooses loans $\ell_i(\theta, j)$ in order to maximize profits and chooses whether to exit $x_i(\theta, j)$ after the realization of the aggregate shock z' and the regional shock s'.[29]

It is simple to see that no bank would ever accept more total deposits than it makes total loans.[30] Further, the deposit rate $r^{D,j} = \overline{r}$. [31] Simply put, a bank would not pay interest on deposits that it doesn't lend out and with excess supply of funds, households are forced to their reservation value associated with storage.

Let $\sigma_{-i} = (\ell_{-i}, x_{-i}, e)$ denote the industry state dependent lending, exit, and entry strategies of all other banks. The end-of-period realized profits in state (z', s') for bank i of type (θ, j) with cost c^θ extending loans ℓ_i starting in state (μ, z, s) is given by:

$$\pi_{\ell_i(\theta,j)}(\theta, j, c^\theta, \mu, z, s, z', s'; \sigma_{-i}) \equiv \Big\{ p^j(R^j, z', s')(1 + r^{L,j}) + (1 - p^j(R^j, z', s'))(1 - \lambda) \\ -(1 + \overline{r}) - c^\theta \Big\} \ell_i(\theta, j).$$

The first two terms represent the net return the bank receives from successful and unsuccessful projects respectively and the last terms correspond to its costs.

Differentiating with respect to ℓ_i we obtain

$$\frac{d\pi}{d\ell_i} = \Big[\underbrace{p^j r^{L,j} - (1 - p^j)\lambda - \overline{r} - c^\theta}_{(+)\text{ or }(-)}\Big] + \ell_i \Big[\underbrace{p^j}_{(+)} + \underbrace{\frac{\partial p^j}{\partial R^j}\frac{\partial R^j}{\partial r^{L,j}}(r^{L,j} + \lambda)}_{(-)}\Big] \underbrace{\frac{dr^{L,j}}{d\ell_i}}_{(-)}. \tag{15}$$

The first bracket represents the marginal change in profits from extending an extra unit of loans. The the second bracket corresponds to the marginal change in profits due to a bank's influence on the interest rate it faces. This term will reflect the bank's market power: for dominant banks $\frac{dr^{L,j}}{d\ell_i} < 0$ while for fringe banks $\frac{dr^{L,j}}{d\ell_i} = 0$.

The value function of a "national" incumbent bank i at the beginning of the period is given by

$$V_i(n, \mu, z, s; \sigma_{-i}) = \max_{\{\ell_i(n,j)\}_{j=e,w}} \beta E_{z',s'|z,s} \left[W_i(n, \mu, z, s, z', s'; \sigma_{-i}) \right] \tag{16}$$

subject to

$$\sum_\theta \int \ell_i(\theta, j, \mu, s, z; \sigma_{-i}) \mu^{(\theta,j)}(di) - L^{d,j}(r^{L,j}, z, s) = 0, \tag{17}$$

[29]In Allen and Gale (2004), banks compete Cournot in the deposit market and offer borrowers an incentive compatible loan contract that induces them to choose the project R which maximizes the bank's objective. As in Boyd and De Nicolo (2005), we assume that banks compete Cournot in the loan market and offer borrowers an incentive compatible loan contract which is consistent with the borrower's optimal decision rule.

[30]Suppose not and $d_i > \ell_i$. The net cost of doing so is $r^{D,j} \geq 0$ while the net gain on $d_i - \ell_i$ is zero, so it is weakly optimal not to do so.

[31]Suppose not and some bank is paying $r^{D,j} > \overline{r}$. Then the bank can lower $r^{D,j}$, still attract deposits since H=2B and make strictly higher profits, so it is strictly optimal not to do so.

where $L^{d,j}(r^{L,j}, z, s)$ is given in (12) and

$$W_i(n, \mu, z, s, z', s'; \sigma_{-i}) = \max_{\{x \in \{0,1\}\}} \left\{ W_i^{x=0}(n, \mu, z, s, z', s'; \sigma_{-i}), W_i^{x=1}(n, \mu, z, s, z', s'; \sigma_{-i}) \right\} \tag{18}$$

$$W_i^{x=0}(n, \mu, z, s, z', s'; \sigma_{-i}) = \mathcal{D}_i + V_i(n, \cdot, \mu', z', s'; \sigma_{-i})$$

$$\mathcal{D}_i = \begin{cases} \sum_j \pi_{\ell_i(n,j)}(n, j, \cdot) - \kappa^n & \text{if} \sum_j \pi_{\ell_i(n,j)}(n, j, \cdot) - \kappa^n \geq 0 \\ \sum_j \pi_{\ell_i(n,j)}(n, j, \cdot) - \kappa^n - \xi^n(\sum_j \pi_{\ell_i(n,j)}(n, j, \cdot) - \kappa^n) & \text{if } \sum_j \pi_{\ell_i(n,j)}(n, j, \cdot) - \kappa^n < 0 \end{cases}.$$

$$W_i^{x=1}(n, \mu, z, s, z', s'; \sigma_{-i}) = \max \left\{ 0, \sum_j \pi_{\ell_i(n,j)}(n, j, \cdot) - \kappa^n \right\}.$$

Constraint (17), which is simply the loan market clearing condition, is imposed as a consistency condition due to the Cournot assumption whereby a national bank realizes its loan supply will influence the interest rate $r^{L,j}$. The exit decision rule is the solution to problem (18) reflects the choice between continuing (and possibly obtaining outside funding in case of negative profits) or exiting.[32] The value of exit is bounded below by zero due to limited liability.

The value function of a "regional" incumbent bank i in region j at the beginning of the period is given by

$$V_i(r, j, \mu, z, s; \sigma_{-i}) = \max_{\ell_i(r,j)} \quad \beta E_{z',s'|z,s} W_i(r, j, \mu, z, s, z', s'; \sigma_{-i}) \tag{19}$$

subject to (17) and where

$$W_i(r, j, \mu, z, s, z', s'; \sigma_{-i}) = \max_{\{x \in \{0,1\}\}} \left\{ W_i^{x=0}(r, j, \mu, z, s, z', s'; \sigma_{-i}), W_i^{x=1}(r, j, \mu, z, s, z', s'; \sigma_{-i}) \right\} \tag{20}$$

$$W_i^{x=0}(r, j, \mu, z, s, z', s'; \sigma_{-i}) = \mathcal{D}_i + V_i(r, j, \mu', z', s'; \sigma_{-i})$$

$$\mathcal{D}_i = \begin{cases} \pi_{\ell_i(r,j)}(r, j, \cdot) - \kappa^r & \text{if } \pi_{\ell_i(r,j)}(r, j, \cdot) - \kappa^r \geq 0 \\ \pi_{\ell_i(r,j)}(r, j, \cdot) - \kappa^r - \xi^r(\pi_{\ell_i(r,j)}(r, j, \cdot) - \kappa^r) & \text{if } \pi_{\ell_i(r,j)}(r, j, \cdot) - \kappa^r < 0 \end{cases}.$$

$$W_i^{x=1}(r, j, \mu, z, s, z', s'; \sigma_{-i}) = \max \left\{ 0, \pi_{\ell_i(r,j)}(r, j, \cdot) - \kappa^r \right\}.$$

The problem of fringe bank i in region j is different from that of a dominant national or regional bank. When fringe banks make their loan supply decision, dominant banks have

[32] To save on notation, we denote $\pi_{\ell_i(n,j)}(n, j, c^n, \mu, z, s, z', s'; \sigma_{-i})$ by $\pi_{\ell_i(n,j)}(n, j, \cdot)$. We use similar notation in what follows.

already made their move and since fringe banks are sufficiently small they take $r^{L,j}$ as given. As discussed following equation (15), in this case the profit function is linear in $\ell_i(f,j)$ so the quantity constraint $\ell_i(f,j) \leq \bar{d}$ will in general bind the loan decision. In particular, the value function of an incumbent fringe bank which drew cost c_i^f at entry and takes the $r^{L,j}$ which solves (17) is given by

$$V_i(f,j,c_i^f,\mu,z,s;\sigma_{-i}) = \max_{\ell_i(f,j)\leq \bar{d}} \quad \beta E_{z',s'|z,s}\left[V_i(f,j,c_i^f,\mu,z,s,z',s';\sigma_{-i})\right] \tag{21}$$

$$W_i(f,j,c_i^f,\mu,z,s,z',s';\sigma_{-i}) = \max_{\{x\in\{0,1\}\}} \left\{W_i^{x=0}(f,j,c_i^f,\mu,z,s,z',s';\sigma_{-i}), W_i^{x=1}(f,j,c_i^f,\mu,z,s,z',s';\sigma_{-i})\right\} \tag{22}$$

$$W_i^{x=0}(f,j,c_i^f,\mu,z,s,z',s';\sigma_{-i}) = \mathcal{D}_i + V_i(f,j,c_i^f,\mu',z',s';\sigma_{-i})$$

$$\mathcal{D}_i = \begin{cases} \pi_{\ell_i(f,j)}(f,j,c_i^f,\cdot) - \kappa^f & \text{if } \pi_{\ell_i(f,j)}(f,j,c_i^f,\cdot) - \kappa^f \geq 0 \\ \pi_{\ell_i(f,j)}(f,j,c_i^f,\cdot) - \kappa^f - \xi^f(\pi_{\ell_i(f,j)}(f,j,c_i^f,\cdot) - \kappa^f) & \text{if } \pi_{\ell_i(f,j)}(f,j,c_i^f,\cdot) - \kappa^f < 0 \end{cases}.$$

$$W_i^{x=1}(f,j,c_i^f,\mu,z,s,z',s';\sigma_{-i}) = \max\left\{0, \pi_{\ell_i(f,j)}(f,j,c_i^f,\cdot) - \kappa^f\right\}.$$

Since the loan interest rate is taken as given and the technology is linear in loans made, the fringe bank's decision is simply whether to bring all their available funds to the market or not, i.e. $\ell_i(f,j) \in \{0,\overline{d}\}$. Total loan supply by fringe banks in region j will be

$$L^s(f,j,\mu,z,s;\sigma_{-i}) = M\Xi\left(\overline{c}^j(\mu,z,s;\sigma_{-i})\right)\bar{d}.$$

where the cutoff $\overline{c}^j(\mu,z,s;\sigma_{-i})$ denotes the highest cost such that a fringe bank will choose to offer loans in region j.

The new distribution of banks after entry and exit μ' is determined by the measure of active banks of type (θ,j) that remain active after the exit stage $\mu^{x,(\theta,j)}$ and the mass of entrants $\mu^{e,(\theta,j)}$ of type (θ,j) as follows:

$$\begin{aligned} \mu' \;&=\; \{\mu^{x,(n,\cdot)} + \mu^{e,(n,\cdot)}, \mu^{x,(r,e)} + \mu^{e,(r,e)}, \\ &\quad\; \mu^{x,(r,w)} + \mu^{e,(r,w)}, \mu^{x,(f,e)} + \mu^{e,(f,e)}, \mu^{x,(f,w)} + \mu^{e,(f,w)}\}. \end{aligned} \tag{23}$$

The mass of banks of type (θ,j) in the industry after exit is given by

$$\mu^{x,(\theta,j)} = \mu^{(\theta,j)} - \int_i x_i(\theta,j,\mu,z,s,z',s';\sigma_{-i})\mu^{(\theta,j)}(di). \tag{24}$$

The mass of fringe banks in region j after exit can be defined as a function of the bank with the highest cost among the survivors. Denote this value by $c^{x,j}$. Then, the mass of fringe banks in region j after exit is:

$$\mu^{x,(f,j)} = M \cdot \min\left\{\Xi\left(c^{x,j}(f,j,\mu,s,z,z',s';\sigma_{-i})\right), \Xi\left(\overline{c}^j(f,j,\mu,s,z,z',s';\sigma_{-i})\right)\right\}.$$

The mass of fringe banks that exits in region j is $\mu^{(f,j)} - \mu^{x,(f,j)}$.

4.4 Entrant Bank Decision Making

In each period, new banks of type θ can enter the industry by paying the setup cost Υ^θ. They will enter the industry if the net present value of entry is nonnegative. For example, taking the entry and exit decisions by other banks as given, a potential regional entrant in the west region will choose $e_i\left(r, w, \{\cdots, \mu^{x,(r,w)} + \mu^{e,(r,w)}, \cdots\}, z', s'\right) = 1$ if

$$V_i(r, w, \{\cdots, \mu^{x,(r,w)} + \mu^{e,(r,w)}, \cdots\}, z', s'; \sigma_{-i}) - \Upsilon^r > 0. \tag{25}$$

4.5 Definition of Equilibrium

A pure strategy Markov Perfect Equilibrium (MPE) is a set of functions $\{v(r^{L,j}, z, s)$ and $R(r^{L,j}, z, s)\}$ describing borrower behavior, a set of functions $\{V_i(\theta, j, \mu, z, s; \sigma_{-i})$, $\ell_i(\theta, j, \mu, z, s; \sigma_{-i})$, $x_i(\theta, j, \mu, z, s, z', s'; \sigma_{-i})$, and $e_i\left(\theta, j, \mu, z', s'; \sigma_{-i}\right)\}$ describing bank behavior, a loan interest rate $r^{L,j}(\mu, z, s)$ for each region, a deposit interest rate $r^D = \overline{r}$, an industry state μ, a function describing the number of entrants $N^e(\theta, j, \mu, z')$, and a tax function $\tau(\mu, z, s, z', s')$ such that:

1. Given a loan interest rate $r^{L,j}$, $v(r^{L,j}, z, s)$ and $R(r^{L,j}, z, s)$ are consistent with borrower's optimization in (7) and (8).

2. For any given interest rate $r^{L,j}$, loan demand $L^{d,j}(r^{L,j}, z, s)$ is given by (12).

3. At $r^D = \overline{r}$, the household deposit participation constraint (13) is satisfied.

4. Given the loan demand function, the value of the bank $V_i(\theta, j, \mu, z, s; \sigma_{-i})$, the loan decision rules $\ell_i(\theta, j, \mu, z, s; \sigma_{-i})$, and exit rules $x_i(\theta, j, \mu, z, s, z', s'; \sigma_{-i})$, are consistent with bank optimization in (16), (18), (19), (20), (21) and (22).

5. The entry decision rules $e_i\left(\theta, j, \mu, z', s'; \sigma_{-i}\right)$ are consistent with bank optimization in (25).

6. The law of motion for the industry state (23) is consistent with entry and exit decision rules.

7. The interest rate $r^{L,j}(\mu, z, s)$ is such that the loan market (17) clears. That is,

$$L^{d,j}(r^{L,j}, z, s) = B \cdot \int_{\underline{\omega}}^{\overline{\omega}} 1_{\{\omega \leq v(r^{L,j}, z, s)\}} d\Upsilon(\omega) = \sum_{\theta} \int \ell_i(\theta, j, \mu, s, z; \sigma_{-i}\mu^{(\theta,j)}(di) = L^{s,j}(\mu, s, z; \sigma_{-i}).$$

8. Across all states (μ, z, s, z', s'), taxes cover deposit insurance:

$$\tau(\mu, z, s, z', s') = \sum_{\theta, j} \int x_i(\theta, j, \mu, z, s, z', s'; \sigma_{-i}) \pi_{\ell_i(\theta,j)}(\theta, j, \mu, z, s, z', s'; \sigma_{-i}) \mu^{(\theta,j)}(di).$$

5 SMM Estimation

We estimate the model parameters by Simulated Method of Moments (SMM) to match the key statistics of the U.S. banking industry described in Section 2. A model period is set to be one year.

Before moving into the details of the estimation, we provide functional forms for the stochastic process of the borrower idiosyncratic shock, the distribution of borrower's outside option, the utility function of the consumer, the distribution of net expenses for fringe banks and banks' external financing cost function.

The stochastic process for the borrower's project is parameterized as follows. For each borrower in region j, let $y^j = \alpha z' + (1-\alpha)\varepsilon_e - bR^{\psi}$ where ε_e is drawn from $N(\phi(s'), \sigma_{\varepsilon}^2)$. The regional shock affects the mean of the idiosyncratic shock through $\phi(s') \in \{-\overline{\phi}, \overline{\phi}\}$. We assume that if $s' = j$, $\phi(s') = \overline{\phi}$ and $\phi(s') = -\overline{\phi}$ otherwise. The borrower's idiosyncratic project uncertainty is iid across agents. We define success to be the event that $y > 0$, so in states with higher z or higher ε_e success is more likely. Then

$$\begin{aligned} p^j(R, z', s') &= 1 - prob(y \leq 0 | R, z', s') \\ &= 1 - prob\left(\varepsilon_e \leq \frac{-\alpha z' + bR^{\psi}}{(1-\alpha)}\right) \\ &= \Phi\left(\frac{\alpha z' - bR^{\psi}}{(1-\alpha)}\right) \end{aligned} \tag{26}$$

where $\Phi(x)$ is a normal cumulative distribution function with mean $\phi(s')$ and variance σ_{ε}^2. We assume that s follows a Markov process and that the transition matrix has diagonal values equal to $\overline{G}$.

We let the distribution of borrower's outside option $\Upsilon(\omega)$ to be a uniform distribution with support defined by $[\underline{\omega}, \overline{\omega}]$ and let consumer's preferences be given by $u(x) = \frac{x^{1-\sigma}}{1-\sigma}$.

We let the external financing cost for national and regional banks take the following form $\xi^{\theta}(x) = \xi_0^{\theta} + \xi_1 x$ for regional and national banks and assume that external financing is prohibitively costly for fringe banks.[33]

Given our assumption about external financing for fringe banks, this implies that fringe banks with negative profits exit. More specifically, for a given loan interest rate, fringe banks will choose to offer loans whenever expected profits net of fixed costs are greater than or equal to zero. This implies that $\overline{c}^j(\mu, z, s; \sigma_{-i})$ solves

$$\widehat{\pi}_{\bar{d}}(f, j, \mu, z, s; \sigma_{-i}) \equiv E_{z', s'|z,s} \max\left\{\pi_{\bar{d}}(f, j, \overline{c}^j, \mu, z, s, z', s'; \sigma_{-i}) - \kappa^f, 0\right\} = 0. \tag{27}$$

Finally, we assume that c^f is distributed exponentially with location parameter equal to μ_c.

The full set of parameters of the model are divided into two groups. The first group of parameters can be estimated directly from the data (i.e. they can be pinned down without

[33] We also estimated the model allowing for full functional differences across dominant banks; however, we find that by setting $\xi_1^n = \xi_1^r = \xi_1$ we reduce the number of estimable parameters without affecting the goodness of fit of the model.

solving the model). After those are set, a second group is estimated via SMM. In what follows, we describe both groups of parameters as well as our targeted moments.

To calibrate the stochastic process for aggregate technology shocks $F(z', z)$, we use the NBER recession dates and create a recession indicator. More specifically, for a given year, the recession indicator takes a value equal to one if two or more quarters in that year were dated as part of a recession. The correlation of this indicator with HP filtered GDP equals -0.87. Then, we identify years where the indicator equals one with our periods of $z = z_b$ and construct a transition matrix. In particular, the maximum likelihood estimate of F_{kj}, the $(j, k)th$ element of the aggregate state transition matrix, is the ratio of the number of times the economy switched from state j to state k to the number of times the economy was observed to be in state j. We normalize the value of $z_g = 1$. We include z_b in the set of the parameters to estimate via SMM and select the variance of detrended US GDP as one of the targeted moments.

We calibrate $\bar{r} = r^D$ using data from the banks' balance sheet. We target the average cost of funds computed as the ratio of interest expense on funds (deposits and federal funds) over total deposits and federal funds for commercial banks in the US from 1976 to 2008.[34] The discount factor β is set to $1/(1 + r^D)$. Using information on the average charge off rate (0.80%) and average default frequency (2.15%) observed in our data, the parameter λ can be set to 0.3721 since the model counterpart of the charge off rate is equal to $(1 - p)\lambda$.

The mass of borrowers is normalized to 1. We set $\underline{\omega} = 0$ and assume that $\sigma = 2$, a standard value in the macro literature. At this level of risk aversion the consumer participation constraint is satisfied.

We identify "national" banks with the top 10 banks (when sorted by loans), "regional" banks with the top 1% banks (also when sorted by loans and excluding the top 10 banks) and the fringe banks with the bottom 99% of the bank loan distribution. The model, on the other hand, has only one "representative" national bank, one "representative" regional bank in each region, and a mass M of potential fringe banks. We link loan supply from the model to data in the following way. The model delivers aggregate loan supply $\ell(\theta, j, \mu, z, s)$ for each bank type given by

$$\ell(\theta, j, \mu, z, s) = \int \ell_i(\theta, j, \mu, z, s)\mu^{(\theta,j)}(di) \equiv w(\theta, j)\bar{\ell}(\theta, j, \mu, z, s) \tag{28}$$

where $w(\theta, j)$ is the relative fraction of banks of type θ in region j and the definition amounts to a representative bank assumption so $\bar{\ell}(\theta, j, \mu, z, s)$ is the "representative" or "average" loan supply by banks of type θ in region j. Since we work under the assumption of a representative bank for each dominant bank type, the relative mass $w(\theta, j)$ is not relevant for the determination of the equilibrium. However, it is important when taking the model to the data and reporting parameters or functions that are expressed in levels (e.g. fixed costs, entry costs, individual loan decision rules and value functions). For example, the average loan supply by a national bank in region j is $\ell(n, j, \mu, z, s)/w(n, \cdot)$.

[34]Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement (http://www2.fdic.gov/hsob/SelectRpt.asp?EntryTyp=10). The nominal interest rate is converted to a real interest rate by using the CPI.

We set $w(\theta, j)$ using data from the distribution of banks. In particular, we associate the national banks with the top 10 banks, the regional banks with the top 1% banks (not including the top 10) and the fringe banks with the bottom 99%. The number of banks in 2010 was 6544 so $w(n, \cdot) = \frac{10}{6544} = 0.153\%$, $w(r, j) = \frac{(654-10)/2}{6544} = 0.4235\%$, and $w(f, j) = 0.99/2 = 44.5\%$.

Dominant banks' net non-interest expenses are calibrated using the information in Table 7. The value of $c^n = 0.0178$ and $c^r = 0.0161$. We set M to one and estimate the value of $\bar{d}$. As long as the mass of fringe banks does not bind, the value of M does not affect the qualitative and quantitative properties of the model as long as d is adjusted accordingly since all relevant equilibrium conditions depend on $M \times \bar{d}$.

We are left with eighteen parameters to estimate via the Simulated Method of Moments (SMM). The full set is given by

$$\Theta = \{\alpha, b, \sigma_\varepsilon, \psi, \overline{G}, \overline{\phi}, \overline{\omega}, z_b, \bar{d}, \mu_c, \kappa^n, \kappa^r, \kappa^f, \xi_0^n, \xi_0^r, \xi_1, \Upsilon^n, \Upsilon^r\}.$$

The SMM procedure consists of minimizing the distance between data moments and moments extracted from the simulated model. That is, the parameters are chosen to minimize

$$J(\Theta) = [\mu^d - \mu^s(\Theta)]W^*[\mu^d - \mu^s(\Theta)]' \tag{29}$$

with respect to parameters Θ, where μ^d are the moments from the data, $\mu^s(\Theta)$ are the moments from the simulated model at parameters Θ.[35] The estimator is consistent for any positive definite weighting matrix. However, there is exists a matrix W^* that gives also efficient estimators. Thus, we say that W^* is an optimally derived weighting matrix. The weighting matrix is calculated via the bootstrap method. More specifically, after an initial set of consistent estimates are obtained using the identity matrix as an initial weight matrix, the model is simulated 4000 times using different draws for the innovations to the exogenous processes to create panels with the same length that our data has (36 years). After each simulation, the auxiliary moments are calculated and stored. The weighting matrix is then calculated as the inverse of the covariance matrix of these 4000 sets of auxiliary moments. Efficient estimates are then obtained using this weighting matrix.

The estimator $\widehat{\Theta}$ is asymptotically normal for fixed number of simulations S:

$$\sqrt{T}(\widehat{\Theta} - \Theta_0) \to N(0, Q_S(W^*))$$

where

$$Q_S(W^*) = \left(1 + \frac{1}{S}\right)\left[E_0 \frac{\partial J'}{\partial \Theta} W^* \frac{\partial J}{\partial \Theta}\right]^{-1}. \tag{30}$$

is the asymptotic variance covariance matrix. The covariance matrix of the estimator converges to the covariance matrix of the standard GMM estimator when $S \to \infty$.

Obtaining the standard errors of the parameters, requires a number of moments larger than the number of parameters (i.e. the model needs to be over-identified). We estimate

[35]For every set of parameters, we simulate twenty eight panels of banks. Each panel consists of up to 10,000 banks (this number never binds along the simulation path) and 7,000 periods. To compute the moments, we discard the initial 2,000 periods and average over all the panels created.

the model using seventeen data moments. Except for three data moments, we use the data for commercial banks described in Section 2 to compute the necessary set of moments. Two of these extra moments are computed using data from the Federal Reserve Bank of St Louis (FRED). They are the standard deviation of detrended log real GDP (1.48%) and the standard deviation of the detrended ratio of loans from commercial banks to GDP (0.82%). The third extra moment corresponds to the average real equity return (12.94%) as reported by Diebold and Yilmaz [20]. This last moment is added to shed light on the borrower's return R^*. The set of targets from commercial bank data includes the average default frequency (2.15%), the average loan return (5.17%), the average entry rate (1.60%), the average ratio of new equity issuances to loans for top 10 and top 1% banks with negative profits (0.01% and 0.42%), the maximum ratio of new equity issuances to loans ratio for dominant banks (5.85%), the ratio of loan returns of Top 10 banks to Top 1% banks (95.97%), the market share of Top 1% banks (30.36%) and Bottom 99% banks (37.91%), the average fixed cost over loans for top 10, top 1% and bottom 99% banks (0.82%, 0.70%, 0.99% respectively), the average net expense of the Bottom 99% banks (1.59%) and the standard deviation of the price cost margin (0.37%).

Before presenting the estimation outcome, we discuss the selection of these moments. Since every moment that results from the model is a function of all parameters, there is no one-to-one link between parameters and moments. However, we can point to moments that are more informative to pin down a given parameter or set of parameters than others. As explained above, the level of the aggregate shock in the bad state is selected to match the standard deviation of aggregate output. The average borrower return, default frequency, loan return and the relative loan returns across banks of different sizes are informative about the parameters of the the stochastic process for the borrower's project $\{\alpha, b, \sigma_\varepsilon, \psi, \overline{G}, \overline{\phi}\}$. These parameters control the underlying relation between loan risk and return and the extent to which banks of different sizes are willing to extend more or less loans affecting the equilibrium interest rate, the borrower project, default probability and the probability of bank exit. The parameter $\overline{\varpi}$ controls the slope of the loan demand function (see equation 12), so the volatility of the ratio of aggregate loans to output is useful to pin down this parameter. The estimated average net-expenses for the bottom 99% banks presented in Table 7 (estimated net-expenses by bank size) provides valuable information to set μ_c (the average net-expense for fringe banks). The average size of fringe banks $\bar{d}$ can be identified using the loan market share of the bottom 99% banks. The market share of the top 1% banks, the average net equity issuance to loans ratio when profits are negative and its maximum value are informative to estimate the parameters of external financing $\{\xi_0^n, \xi_0^r, \xi_1\}$. These parameters also affect the loan supply of dominant banks (national and regional) by allowing them to continue even when they face negative profits, so are an important determinant of the equilibrium interest rate and average default frequency. The average fixed cost over loans for top 10, top 1% and bottom 99% banks are informative about κ^θ. Since loans are endogenous, and we target the fixed cost to loan ratio, it is necessary to incorporate κ^θ to the SMM procedure. The entry rate and the fraction of entry accounted by top 1% banks are informative to estimate Υ^n and Υ^r.

One important aspect when estimating the model by SMM is that a moment is informative if it is sensitive to changes in parameters. The SMM standard errors are a function of the inverse of the partial derivative of the objective function with respect to the parameters of the model(see equation 30). For this reason, the size of the standard errors will provide additional and very valuable information on the identification of model. As it turns, standard errors are small reflecting that moments are very informative.

We use the following definitions to connect the model to the variables we presented in the data section.

Definition Model Moments

Aggregate loan supply	$L^s(\mu, z, s) = \sum_j L^{s,j}(\mu, z, s)$
Aggregate Output	$\sum_j L^{s,j}(\mu, z, s)\Big\{p^j(R^*(\mu, z, s), z', s')(1 + z'R)$ $+(1 - p^j(R^*(\mu, z, s), z', s'))(1 - \lambda)\Big\}$
Entry Rate	$\sum_j \sum_\theta \mu_t^{e,(\theta,j)}/\mu_{t-1}^{(\theta,j)}$
Default frequency	$1 - p^j(R^*(\mu, z, s), z', s')$
Borrower return	$p^j(R^*(\mu, z, s), z', s')(z'R^*(\mu, z, s))$
Loan (interest) return	$p^j(R^*(\mu, z, s), z', s')r^{L,j}(\mu, z, s)$
Loan Charge-off rate	$(1 - p^j(R^*(\mu, z, s), z', s'))\lambda$
Profit Rate	$\pi_{\ell_i(\theta,j)}(\theta, j, \mu, z, s, z', s')/\ell_i(\theta, j, \mu, z, s)$
Equity Issuance	$\max\{-(\pi_{\ell_i(\theta,j)}(\theta, j, \mu, z, s, z', s') - \kappa^\theta), 0\}$
Net Interest Margin	$p^j(R^*(\mu, z, s), z', s')r^{L,j}(\mu, z, s) - r^d$
Lerner Index	$1 - \left[r^d + c^{\theta,exp}\right] / \left[p^j(R^*(\mu, z, s), z', s')r^{L,j}(\mu, z, s) + c^{\theta,inc}\right]$
Markup reg. j	$\left[p^j(R^*(\mu, z, s), z', s')r^{L,j}(\mu, z, s) + c^{\theta,inc}\right] / \left[r^d + c^{\theta,exp}\right] - 1$

As in the data, aggregates are computed as loan-weighted averages. For example, the weighted-average loan return in any given period is

$$\sum_{\theta,j} \int p^j(R^*(\mu, z, s), z', s')r^{L,j}(\mu, z, s)\frac{\ell_i(\theta, j, \mu, z, s)}{L^s(\mu, z, s)}d\mu^{(\theta,j)}(di).$$

Table 10 shows the estimated parameters.[36]

[36]We are currently working on computing the standard errors of estimated parameters.

Table 10: Model Parameters

Parameter		Value	s.e.	Target
Mass of borrowers	B	1	-	Normalization
Mass of households	H	$2B$	-	Assumption
Dep. preferences	σ	2	-	Part. constraint
Agg. shock in good state	z_g	1	-	Normalization
Transition probability	$F(z_g, z_g)$	0.86	-	NBER data
Transition probability	$F(z_b, z_b)$	0.43	-	NBER data
Deposit interest rate (%)	$\bar{r}$	0.72	-	Int. expense
Discount Factor	β	0.99	-	Int. expense
Net. non-int. exp. n bank	c^n	1.78	-	Net non-int exp. Top 10
Net. non-int. exp. r bank	c^r	1.61	-	Net non-int exp. Top 1%
Charge-off rate	λ	0.21	-	Charge off rate
Weight agg. shock	α	0.883	-	Default freq.
Success prob. param.	b	3.773	-	Loan return
Volatility borrower's dist.	σ_ϵ	0.059	-	Borrower Return
Success prob. param.	ψ	0.784	-	Loan ret. top 10 to top 1%
Regional shock	$\overline{\phi}$	0.095	-	Std. dev. net-int. margin
Persistence reg. shock	$\overline{G}$	0.850	-	Std. dev. charge-off rate
Max. reservation value	$\overline{\omega}$	0.227	-	Std. dev. L^s/Output
Agg. shock in bad state	z_b	0.969	-	Std. dev. Output
Dist. net-non int. exp f bank	μ_c	0.014	-	Net non-int exp. bottom 99%
Deposit f banks	$\bar{d}$	0.16	-	Loan mkt share bottom 99%
Fixed cost n bank	κ^n	0.556	-	Fixed cost over loans top 10
Fixed cost r bank	κ^r	0.083	-	Fixed cost over loans top 1%
Fixed cost f bank	κ^f	0.001	-	Fixed cost over loans bottom 99%
External finance param.	ζ_0^n	137.3	-	Avg. equity issuance to loan ratio top 10
External finance param.	ζ_0^r	23.0	-	Avg. equity issuance to loan ratio top 1%
External finance param.	ζ_1	0.250	-	Max. equity issuance/loan
Entry Cost National*	Υ^n	274.5	-	Bank entry rate
Entry Cost Regional*	Υ^r	24.7	-	Entry accounted by top 1%
				Loan mkt share top 1%

Note: * Value reported corresponds to the upper bound of the set consistent with an equilibrium.

Table 11 displays the targeted moments of the model and a comparison with the data.

Table 11: Model and Targeted Moments

Moment (%)	Model	Data
Default freq.	1.22	2.15
Loan return	5.62	5.17
Borrower Return	14.45	12.94
Loan ret. top 10 to top 1%	90.57	95.97
Std. dev. net-int. margin	0.47	0.37
Std. dev. charge-off rate	0.61	0.22
Std. dev. L^s/Output	1.33	0.82
Std. dev. Output	1.85	1.48
Net non-int exp. bottom 99%	1.23	1.59
Loan mkt share bottom 99%	39.64	37.91
Fixed cost over loans top 10	0.55	0.49
Fixed cost over loans top 1%	0.46	0.43
Fixed cost over loans bottom 99%	0.46	0.46
Avg. equity issuance to loan ratio top 10	0.06	0.01
Avg. equity issuance to loan ratio top 1%	1.27	0.41
Max. equity issuance/loan	7.04	5.85
Bank entry rate	2.78	1.60
Entry accounted by top 1%	0.81	0.46
Loan mkt share top 1%	39.37	30.36

In general, the model does a good job in matching the targeted moments. However, the standard deviation of the loan supply to output ratio and the standard deviation of aggregate output are higher than in the data. Also, the default frequency, the net non-interest expense and the fixed cost for fringe banks is lower than observed in the data.

6 Results

We can summarize the equilibrium that corresponds to the parameter values displayed in Table 10 by describing the entry and exit decisions of dominant and fringe banks. In this equilibrium, national banks do not exit while regional banks and fringe banks exit in bad times (though fringe market share takes up some of the slack of the regional bank market share in bad times). In particular, we find: (i) if there is no “regional” bank in one of the regions ($N(r, j) = 0$ for $j = e$ and/or $j = w$), there is entry by a “regional” bank in the region with $s = j$ for $z = z_b$ (this is on-the-equilibrium path); (ii) if $N(n, \cdot) = 0$, there is entry by a “national” bank (this is off-the-equilibrium path); (iii) a “regional” bank in region j exits when $s' \neq j$ and $z' = z_b$ for all $z = z_g$.

To understand the equilibrium, we first describe borrower decisions. Figure 17 shows the borrower’s optimal choice of project riskiness $R^*(r^{L,j}, z, s)$ and the inverse demand function associated with $L^d(r^{L,j}, z, s)$ for region 1 (those corresponding to region 2 are similar).

Figure 17: Borrower's Risk Taking R and Loan Inverse Demand

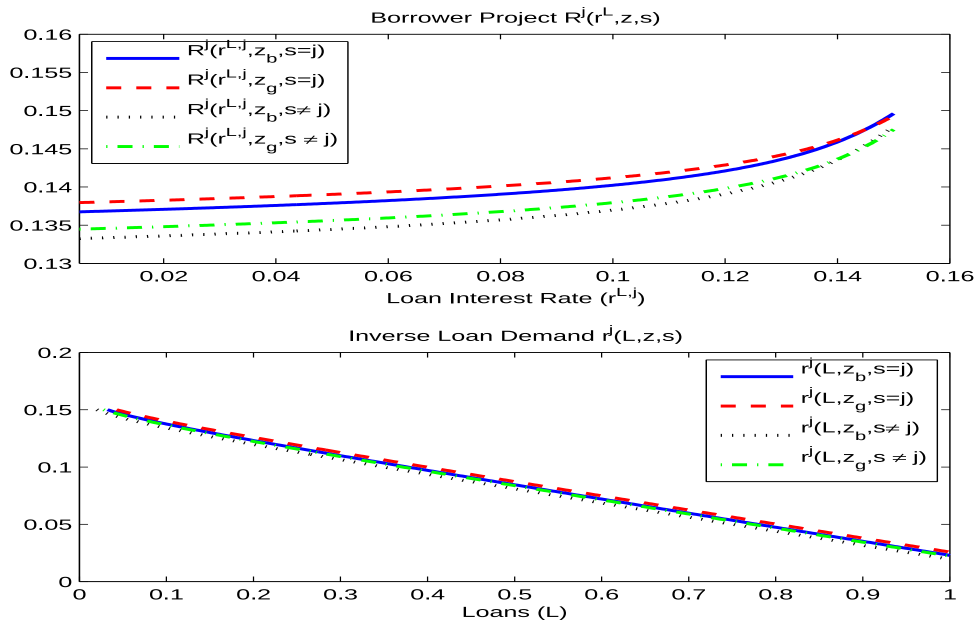


Figure 17 shows that the borrower's optimal project R is an increasing function of the loan interest rate $r^{L,j}$. An equilibrium with lower competition and higher prices results in more risky projects being undertaken by borrowers. Moreover, given that the value of the borrower is decreasing in $r^{L,j}$, loan demand is a decreasing function of $r^{L,j}$.

In Figure 18 and Table (12), we provide a description of bank decision rules. Note that while these are equilibrium functions not every state is necessarily on-the-equilibrium path (starting with Table 15 we evaluate the behavior of the model on-the-equilibrium path).

Figure 18: Competitive Fringe Thresholds $\overline{c}^j$ and $c^{x,j}$

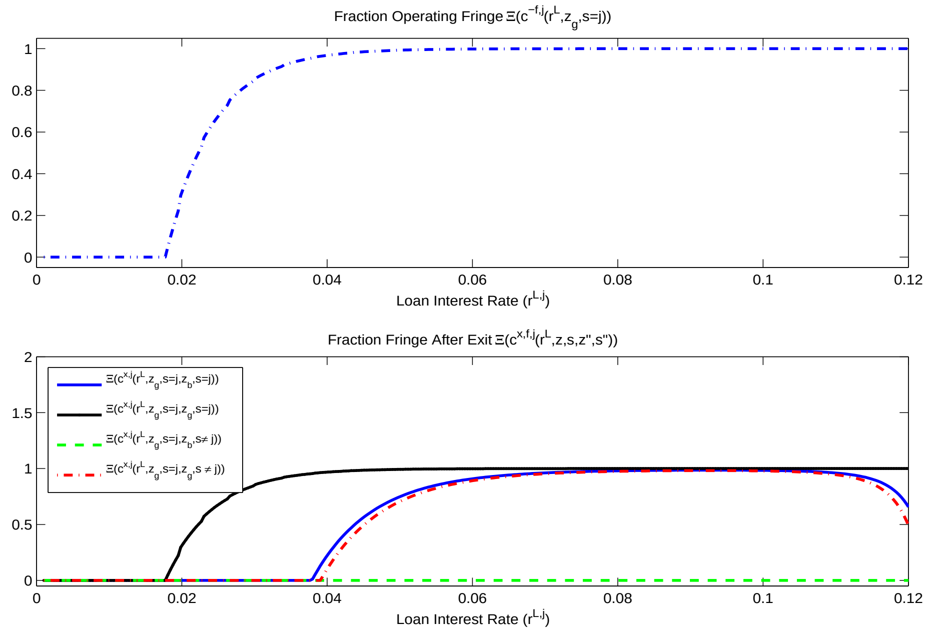


Figure 18 shows how the the fraction of active fringe banks changes with the loan interest rate for the case of $z = z_g$ and $s = j$ (a similar figure results for other combinations of z and s). In the top panel, we observe the fraction of fringe banks that decide to extend loans $\Xi(\overline{c}^j)$. This fraction is increasing in the loan interest rate since expected profits are increasing in $r^{L,j}$. Limited liability imposes a lower bound of bank losses and allows the direct positive effect of $r^{L,j}$ to exceed the indirect negative effect of p^j on expected profits. In the bottom panel, we observe the fraction of fringe banks that survive after the exit stage $\Xi(c^{x,j})$. If the aggregate shock stays in z_g (i.e. $z' = z_g$), all the fringe banks that extended loans will remain active, $c^{x,j} = \overline{c}^j$. On the other hand, a fraction of the fringe banks will exit when $z' = z_b$ and $s' = j$ and the the full fringe sector in region j disappears when $z' = z_b$ and $s' \neq j$. This latter case is very unlikely (probability 0.004) since aggregate and regional shocks are highly persistent.

Table 12 provides the loan decision rules for national and regional banks. The first four columns correspond to the loans made by a national bank in region e and w respectively. The next two columns correspond to the regional bank in region e and the last two columns to the regional bank in region w.

Table 12: Dominant Bank Loan Decision Rules $\bar{\ell}(\theta, j, \mu, z, s; \sigma_{-n})$

	$\bar{\ell}(n,e)$		$\bar{\ell}(n,w)$		$\bar{\ell}(r,e)$		$\bar{\ell}(r,w)$	
μ	(z_b,e)	(z_g,e)	(z_b,e)	(z_g,e)	(z_b,e)	(z_g,e)	(z_b,e)	(z_g,e)
$\{0,1,0,\cdot\}$	-	-	-	-	45.5	46.6	-	-
$\{0,1,1,\cdot\}$	-	-	-	-	45.5	46.6	44.0	45.5
$\{1,0,0,\cdot\}$	122.5	117.9	121.2	123.2	-	-	-	-
$\{1,1,0,\cdot\}^*$	81.2	20.3	121.2	123.2	31.0	43.6	-	-
$\{1,1,1,\cdot\}^*$	79.9	7.2	81.9	82.6	31.5	45.5	31.7	31.5
$\{0,1,0,\cdot\}$	-	-	-	-	44.0	45.5	-	-
$\{0,1,1,\cdot\}$	-	-	-	-	44.0	45.5	45.5	46.6
$\{1,0,0,\cdot\}$	121.2	123.2	122.5	117.9	-	-	-	-
$\{1,1,0,\cdot\}^*$	81.9	82.6	6.6	117.9	31.2	31.5	-	-
$\{1,1,1,\cdot\}^*$	81.9	82.6	79.9	7.2	31.7	31.5	31.5	45.5

Note: * denotes a distribution that arises on the equilibrium path. Loan values reported correspond to $\bar{\ell}(\theta, j, \mu, z, s)$ in equation (28).

We observe that, in general, banks offer more loans when $z = z_g$ than when $z = z_b$. On the equilibrium path, when facing a competitor in region j, the national bank offer a small amount of loans (see for example $\ell(n, e)$ when $\mu = \{1, 1, 1, \cdot\}$ and the state is $(z = z_g, s = e)$). National banks reduce the amount of loans in region j since the region has a positive probability of entering a deeper recession (being hit by a regional shock from $s = e$ to $s' = w$ and from $z = z_g$ to $z' = z_b$) to induce regional banks to exit. In fact, this is the state when regional banks exit on the equilibrium path. Independent of the aggregate shock, national banks offer more loans in regions where they have more market power (i.e. when they do not face competition by a regional bank). They also reduce the amount of loans in region j when $s \neq j$ and they are competing with a regional bank. They increase the level of markups with no risk of exit since they are diversified and a regional shock that would cause an increase in loan failure in region j generates a similar reduction in failures in the other region.

The following equation presents the exit decision rules for regional banks.

$$x_i(r, j, \mu, z, s, z', s') = \begin{cases} 1 & \text{if } z = z_g,\ s = j,\ z' = z_b, \text{ and } s' \neq j \\ 0 & \text{otherwise} \end{cases} . \quad (31)$$

Exit occurs for a regional bank when its regional shock is negative during a recession (on-the-equilibrium path). This happens because borrowers take on more risk in good times and project failure is more likely in bad states. The national bank loan decision lowers realized profits of regional banks enough to induce them to exit in order to become the only dominant bank in that particular region the following period. To see this dynamic aspect of strategic behavior, we compare decision rules on an equilibrium path of the benchmark dynamic model versus a static economy evaluated at $\mu = \{1, 1, 1, \cdot\}$, $z = z_g$, $s = e$ in the following table.

Table 13: Dynamic vs Static Model

	Loan Decision Rules $\ell(\theta, j, \mu, z, s)$ ($\mu = \{1, 1, 1, \cdot\}, z = z_g, s = e$)			
Model	$\bar{\ell}(n, e, \cdot)$	$\bar{\ell}(n, w, \cdot)$	$\bar{\ell}(r, e, \cdot)$	$\bar{\ell}(r, w, \cdot)$
Dynamic (benchmark)	7.2	82.6	45.4	31.5
Static	53.1	51.8	20.4	19.9
	Exit Rule $x(\theta, j, \mu, z, s, z', s')$ ($\mu = \{1, 1, 1, \cdot\}, z = z_g, s = e, z' = z_b, s' = w$)			
Model	$x(n, \cdot)$		$x(r, e, \cdot)$	$x(r, w, \cdot)$
Dynamic (benchmark)	0		1	0
Static	1		1	0

Note: Loan values reported correspond to $\bar{\ell}(\theta, j, \mu, z, s)$ in equation (28).

As evident in Table 13, in general, dominant banks offer more loans in the dynamic case than in the static case. A larger loan supply induces a reduction in $r^{L,j}$ leading the borrower to reduce R^j which in turn increases the success probability p^j. The national bank offers less loans in the dynamic cases than in the static case only in states where changes in the loan interest rate and default frequencies can induce regional bank exit. Expected profits are lower but since the continuation value is still positive, dominant banks adjust their strategies to prevent exit (recall that in order to continue when profits are negative they need to raise enough equity). In the dynamic case, National banks do not exit and regional banks only exit when $\{z = z_g, s = j, z' = z_b, s' \neq j\}$. On the other hand, in the static case, since limited liability imposes a lower bound on bank losses, banks reduce the amount of loans to increase static expected profits as much as possible.

Table 14 shows the relation between the degree of bank competition measured by the number of dominant banks and important first moments for the economy. More competition (i.e. more active banks) implies a higher loan supply, a lower interest rate on loans, lower bank profit rates, and higher borrower returns. Despite lower interest rates on loans with more competition, exit rates, entry rates and default frequency display a non-linear relation with the number of dominant banks in the market. This table is constructed using bank decision rules. Not all distributions are observed on the equilibrium path.

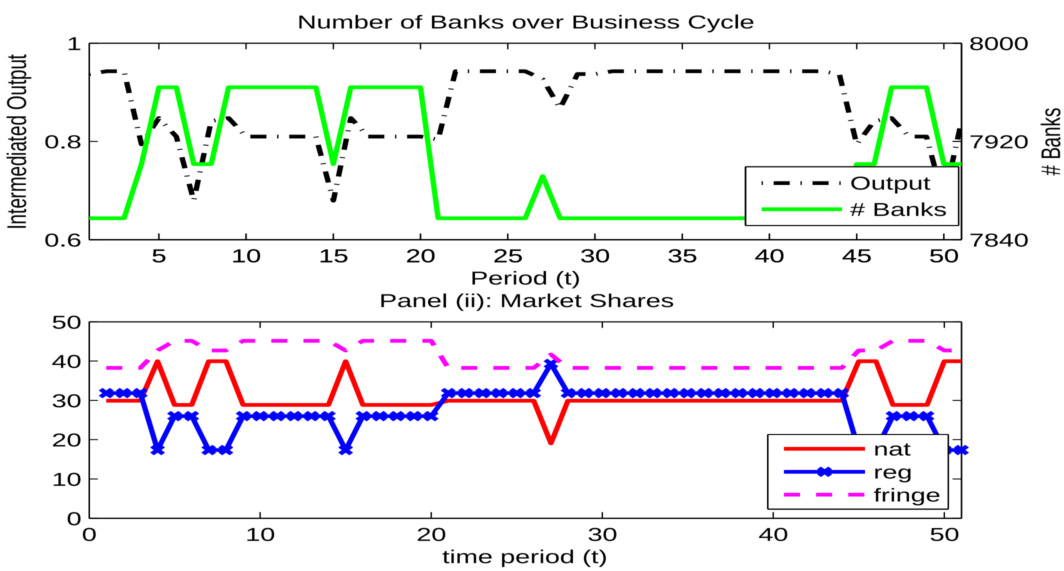
Table 14: Model Moments and Market Concentration

	$N(n,\cdot)+\sum_j N(r,j)=$		
Moment Average	1	2	3
Default frequency	3.83	3.37	3.06
Borrower return	14.37	14.40	14.18
Loan return rate	6.67	6.24	5.70
Loan interest rate (r^L)	7.33	6.61	5.98
Charge-off rate	1.23	1.17	1.08
Entry Rate	0.81	1.32	1.89
Bank profit rate	3.34	2.74	2.16

6.1 Industry Dynamics

We simulate a pseudo-panel of firms for 5000 model periods (years) and cut the last 50 periods. In Figure 19, we plot the number of firms and market shares across time noting periods of state z_b.

Figure 19: Sample Path of Industry Dynamics



As evident in Figure 19, the cyclical properties of the number of banks are similar to that of total output. This is the reflect of exit being countercyclical and entry being procyclical. Also, we see that the fringe accounts for a larger market share when regional banks market share increase. The sample exhibits periods of high concentration following recessions. Large swins correspond to entry or exit by regional banks following a switch in regional shock. The drop around period 20 corresponds to the case where the regional shock changes while $z = z' = z_g$. This generates exit only by fringe banks, with the corresponding reduction in their market share, that is not accompanied by exit by dominant banks. The national bank increases the amount of loans and takes over that portion of the market increasing the loan supply and output. Periods of high concentration following recessions raise interest rates and amplify downturns.

6.2 Tests of the Model

We now move on to moments that the model was not calibrated to match, so that these tables can be considered simple tests of the model. Table 15 provides the correlation between key aggregate variables with GDP.[37] We observe that, as in the data, the model generates countercylical loan interest rates, exit rates, default frequencies, loan returns, charge-off rates, price-cost margins and markups. Moreover, the model generates procyclical aggregate loan supply, entry rates and profit rates.

Table 15: Business Cycle Correlations

Variable Correlated with GDP Shock	Model	Data
Loan Interest Rate r^L	-0.14	-0.18
Exit Rate	-0.46	-0.47
Entry Rate	0.02	0.25
Loan Supply	0.34	0.72
Default Frequency	-0.55	-0.61
Loan Return	-0.03	-0.26
Charge-off rate	-0.55	-0.56
Profit Rate	0.32	0.36
New Equity/Loans	-0.07	-0.39
Price Cost Margin	-0.08	-0.31
Markup	-0.17	-0.27

Table 16 displays a comparison of the measures of the degree of competition in the banking industry between the model and the data. This table shows that the model generates a price cost margin, markup, and Lerner index that are in line with the data .

[37] We use the following dating convention in calculating correlations. Since most variables depend on z, s, z', and s' (e.g. default frequency $1 - p(R(r^L(\mu, z, s)), z', s')$), we display $corr(GDP(\mu, z, s, z', s'), k(z, z'))$.

Table 16: Measures of Bank Competition

Moment (%)	Model	Data
Net Interest Margin	4.9	4.56
Lerner Index	48.32	43.11
Markup	108.44	90.13

To further analyze the differences between the model and the data, in Table 17, we present moments across banks of different size, "national" (Top 10 banks), "regional" (Top 1% banks) and "fringe" (Bottom 99%).

Table 17: Model Moments by Bank Size

	National		Regional		Fringe	
Moment Average	Model	Data	Model	Data	Model	Data
Loan returns*	5.22	5.24	5.76	5.46	5.99	6.05
Variance Return	0.24	1.14	0.74	1.20	0.96	1.22
Default Freq.	1.03	2.82	3.11	1.93	1.68	1.64
Charge-off Rate	0.21	1.06	0.65	1.00	0.35	0.57
Loan Interest Rate	5.27	5.39	5.94	5.57	6.09	6.15
Net Interest Margin	4.50	4.19	5.04	4.51	5.27	5.17
Lerner	35.58	42.94	40.00	45.11	60.98	54.70
Markup	55.39	70.38	72.94	96.72	168.98	146.06

Note: * Denotes calibration Target.

As evident in the table, the model is consistent in generating the pattern of loan return across different size banks as in the data, but this is not surprising since they were targeted in Table 11. The model is also consistent in generating the relation between variance of returns across bank size the relation between size and interest margins, markups and Lerner indexes. The model is unable to match the monotonically decreasing delinquency and chargeoff rates in the data.

6.3 Empirical Studies of Banking Crises, Default and Concentration

Many authors have tried to empirically estimate the relation between bank concentration, bank competition and banking system fragility and default frequency using a reduced form approach. In this section, we follow this approach using simulated data from our model to show that the model is consistent with the empirical findings. As in Beck et. al. [6], we estimate a logit model of the probability of a crisis as a function of the degree of banking industry concentration and other relevant aggregate variables. Moreover, as in Berger et. al. [7], we estimate a linear model of the aggregate default frequency as a function of banking industry concentration and other relevant controls. The banking crisis indicator takes value equal to one in periods whenever: (i) the loan default frequency is higher than 10%; (ii) deposit insurance outlays as a fraction of GDP are higher than 2%; (iii) large dominant banks are liquidated; or (iv) the exit rate is higher than two standard deviations from its mean. The concentration index corresponds to the loan market share of the national and regional banks. We use as extra regressors the growth rate of GDP and lagged growth rate of loan supply.[38] Table 18 displays the estimated coefficients and their standard errors.

Table 18: Banking Crises, Default Frequencies and Concentration

Model Dependent Variable	Logit Crisis_t	Linear Default Freq._t
Concentration_t	-3.77	0.0294
	(0.86)***	(0.001)***
GDP growth in t	0.81	-1.423
	(0.09)***	(0.021)***
$\text{Loan Supply Growth}_t$	-3.38	1.398
	(1.39)**	(0.0289)***
R^2	0.76	0.53

Note: Standard Errors in parenthesis. R^2 refers to Pseudo R^2 in the logit model.
*** Statistically significant at 1%, ** at 5% and * at 10%.

Consistent with the empirical evidence in Beck, et. al. [6], we find that banking system concentration is highly significant and negatively related to the probability of a banking crises. The results suggest that concentrated banking systems are less vulnerable to banking crises. Higher monopoly power induces periods of higher profits that prevent bank exit. This is in line with the findings of Allen and Gale [4]. Consistent with the evidence in Berger et. al. [7] we find that the relationship between concentration and loan portfolio risk is positive. This is in line with the view of Boyd and De Nicolo [11], who showed that higher concentration is associated with riskier loan portfolios.

7 Counterfactuals

7.1 On the effects of Bank Competition

Given entry costs, aggregate and regional shocks determine equilibrium entry/exit and hence the degree of industry concentration. To disentangle the effect of bank competition on risk taking and the probability of crises we run a counterfactual where entry costs Υ^r are

[38] Beck et. al. [6] also include other controls like "economic freedom" which are outside of our model.

raised by 12% in which case “regional” banks choose not to enter the market, thus endogenously generating a more concentrated industry (inducing a market structure identical to the Gowrisankaran and Holmes [24]).

Table 19: Effects of Lower Competition

Moment	Benchmark	$\uparrow \Upsilon^r$ change (%)
Loan Supply	0.78	-11.84
Loan Interest Rate (%)	5.69	19.90
Markup	108.44	27.73
Market Share bottom 99%	39.64	16.02
Market Share top 1%	39.37	-100.00
Market Share top 10	20.99	157.31
Borrower Risk Taking R (%)	14.78	0.17
Default Frequency (%)	1.22	49.52
Entry/Exit Rate (%)	2.78	-21.85
Int. Output	0.89	-11.88
Taxes/Output (%)	17.84	-46.06

As the level of competition decreases, the aggregate loan supply is also reduced.[39] We observe that in the less competitive environment default frequencies, loan interest rates and borrower risk taking are higher, while the entry/exit rate, and taxes over intermediated output are lower. In line with the predictions of A-G, a reduction in the level of competition reduces the entry rate (and by construction the exit rate). In this counterfactual, national banks are the only dominant bank and operate in a region of interest rates and default frequency that avoids failure. This also increases expected profits for fringe banks reducing exit for them as well. Note that the market share of the national bank increases but the market share of fringe banks is also higher than in the benchmark. The reduction in the loan supply by 11.84% increases the loan interest rate as well as markups (they increase by 19.9% and 27.73% respectively). Consistent with the predictions of B-D, a reduction in the level of competition increases the equilibrium interest rate on loans which induces borrowers to take on slightly more risk (i.e. R is 0.17% higher). This in turn, leads to some of the increase in default frequency by 49.52%. The increase in the default frequency is due to the fact that fringe banks take some of the market left unattended by regional banks. Since higher expected profits imply that banks with higher costs can survive a negative shock, the loan weighted average default frequency increases. The increase in default frequency reduces output by 11.88%. In this equilibrium, there is no exit by national banks. The reduction in exit rates by dominant banks reduces the amount of taxes that need to be collected to pay for deposit insurance by 46.06%.

[39]Note that loan supply denotes the fraction of borrower (unit) projects which are financed by banks.

7.2 On the effects of Branching Restrictions

Important regulatory changes took place during the late eighties and early nineties in the U.S. banking industry. In 1994, the Riegle-Neal Interstate Banking and Branching Efficiency Act was passed.[40] The act allows banks to freely establish branches across state lines opening the door to the possibility of substantial geographical consolidation in the banking industry. To study the implications of branching restrictions, we study a counterfactual where we increase Υ^n by 20%, a value that prevents entry of national banks resulting in only regional and fringe banks in equilibrium. By contrasting this case with our benchmark model, we can study the benefits and costs of removing branching restrictions.

Table 20: Counterfactual: Effects of Branching Restrictions

Moment	Benchmark	$\uparrow \Upsilon^n$ Change (%)
Loan Supply	0.78	-1.64
Loan Interest Rate (%)	5.69	18.42
Markup	108.44	29.56
Market Share bottom 99%	39.64	14.71
Market Share top 1%	39.37	38.51
Market Share top 10	20.99	-100.00
Borrower Risk Taking R (%)	14.78	0.16
Default Frequency (%)	1.22	48.14
Entry/Exit Rate (%)	2.78	-21.77
Int. Output	0.89	-10.93
Taxes/Output (%)	17.84	-46.80

Increasing Υ^n such that no national banks enter, each of the regions becomes a more concentrated market since there is at most one incumbent dominant bank each period. In fact, given that the regional banks become regional monopolies, there is no exit by regional banks in this counterfactual. Thus, the qualitative properties of this counterfactual are close to those of the previous counterfactual where there is a national monopoly. There is a reduction in the aggregate loan supply of 1.64%. The loan interest rate and markup increase by 18.42% and 29.56% respectively. This results in an increase in the riskiness of the borrower's project choice (+0.16%), the default frequency (+48.14%). The increase in margins overturns the effects of a higher default frequency and this results in a lower exit rate (-21.77%). Changes in the default frequency and exit rate balance to generate no changes in the tax to GDP ratio that needs to be collected. Finally, the increase in loan interest rates reduces the number of borrowers that choose to operate the technology resulting in a lower level of output (10.93% lower). Thus, the effect on the level of output clearly dominates the reduction in exit rates observed. The increase in market power results in a reduction of exit

[40]The act removed the final restrictions that were in place in 1994, but the consolidation of the banking industry was a process that started during the eighties.

rates by regional banks. The reduction in exit rates by dominant banks reduces the amount of taxes that need to be collected to pay for deposit insurance.

7.3 Policies to Reduce the Cost of Loanable Funds

In response to the crisis, the Fed has lowered the cost of loanable funds. In this counterfactual we compare the benchmark model where $\overline{r} = 0.72\%$ with one where $\overline{r} = 0$. Table 21 presents the results.

Table 21: Counterfactual: Effects of Lower $\overline{r}$

Moment	Benchmark	$\bar{r} = 0$ Change (%)
Loan Supply	0.78	4.68
Loan Interest Rate (%)	5.69	-7.28
Markup	108.44	-12.39
Market Share bottom 99%	39.64	-3.93
Market Share top 1%	39.37	2.93
Market Share top 10	20.99	1.92
Borrower Risk Taking R (%)	14.78	-0.07
Default Frequency (%)	1.22	25.73
Entry/Exit Rate (%)	2.78	-11.60
Int. Output	0.89	4.70
Taxes/Output (%)	17.84	-57.74

Ceteris paribus, a decrease in the cost of funds increases banks' profits. Higher profitability alters the nature of the counterfactual equilibrium relative to the benchmark; now there is no regional bank exit at all. The lower cost of funds opens the possibility for dominant banks to make more loans without increasing their probability of exit. The loan supply increases by 4.68% and the market share of dominant banks increase (+1.92% for national banks and +2.93% for regional banks). The increase in the loan supply reduces the loan interest rate (-7.28%) that in place results in lower markups and lower entry by fringe banks (-11.60%). The loan market share of the fringe sector decreases by -3.93%. As a consequence of lower loan rates, borrowers take on less risk (i.e. the choice of R falls slightly). The increase in loan market share by regional banks (who have the highest average default frequency across banks of different size) results in an increase in the overall default frequency. Finally, the increase in total loan supply implies that more projects are funded resulting in an increase in total Output of 4.70%.

It is interesting to look at the effect of this policy change on the cyclical properties of the equilibrium aggregates. Table 22 presents these correlations.

Table 22: Counterfactual: Effects of Lower $\overline{r}$

Correlation with output	Benchmark	$\overline{r} = 0$
Loan Interest Rate r^L	-0.14	0.11
Exit Rate	-0.46	-0.49
Entry Rate	0.02	-0.01
Loan Supply	0.34	-0.11
Default Frequency	-0.55	-0.63
Loan Return	-0.03	-0.05
Charge-off rate	-0.55	-0.63
Profit Rate	0.32	0.46
New Equity/Loans	-0.07	-0.59
Price Cost Margin	-0.08	-0.05
Markup	-0.17	-0.13

We observe that a reduction in the cost of loanable funds generates a fundamental change in the cyclicality of the loan interest rate, from countercyclical to procyclical. All entry and exit is by fringe banks in this equilibrium. The first thing to note is there is now countercyclical entry. This arises because in downturns arising from regional shocks there is less exit in that region and actually more entry in the region which is doing well. Hence in a downturn, more loans are made, thereby decreasing interest rates (generating procyclicality). The other fundamental difference is that price-cost margins become procyclical as well. This arises from the changes in the cyclical properties of loan interest rates. When the cost of loanable funds is zero, the marginal cost of producing a loan is only c^θ. In downturns, we have more exit by fringe banks (only those with low cost survive). This induces markups to be countercyclical. In the benchmark case, this correlation is reinforced by the fact that loan returns (the marginal price) are also countercyclical. However, in the case with $\overline{r} = 0$, loan returns are highly procyclical making markups less countercyclical and price-cost margins procyclical.

7.4 Too Big To Fail

We documented that the top 10 commercial banks control more than 51% of total loans. As far as we know, ours is the first structural quantitative model of banking which admits a nontrivial endogenous size distribution of banks. This makes the model suitable for analyzing changes in policies that affect banks of particular sizes.

In our benchmark economy, there is no failure by national banks on-the-equilibrium path but it could happen off-the-equilibrium path. Failure doesn't arise on-the-equilibrium path because national banks reduce their exposure in the region with higher risk in order to maintain their charter value. However, a policy of big bank bailouts or "too big to fail" guarantees that the the government will bail out national banks in the event of realized losses. Such a policy changes the ex-ante incentives of national banks since they can take on more risk guaranteed that they receive ex-post bailouts.

In this section, we compare our benchmark economy with one where there are government bailouts to national banks with negative profits. More specifically, we consider the case where if realized profits for a national bank are negative the government will cover a fraction φ of the losses. The bank will optimally decide to stay in operation or to exit after receiving the funds. Since the value of the bank is positive, as φ goes to 1 (i.e. the government covers a larger fraction of the loss), the probability of the bank continuing in operation increases. The problem of a national bank becomes

$$V_i(n,\cdot,\mu,z,s;\sigma_{-i}) = \max_{\{\ell_i(n,j)\}_{j=e,w}} E_{z',s'|z,s}\left[W_i(n,\cdot,\mu,z,s,z',s';\sigma_{-i})\right]$$

subject to the loan market clearing condition (17) and

$$W_i(n,\cdot,\mu,z,s,z',s';\sigma_{-i}) = \max_{\{x\in\{0,1\}\}} \left\{W_i^{x=0}(n,\cdot,\mu,z,s,z',s';\sigma_{-i}), W_i^{x=1}(n,\cdot,\mu,z,s,z',s';\sigma_{-i})\right\}$$

$$W_i^{x=0}(n,\cdot,\mu,z,s,z',s';\sigma_{-i}) = \mathcal{D}_i + \beta V_i(n,\cdot,\mu',z',s';\sigma_{-i})$$

$$\mathcal{D}_i = \begin{cases} \sum_j \pi_{\ell_i(n,j)}(\cdot) & \text{if } \sum_{j=e,w}\pi_{\ell_i(n,j)}(\cdot) \geq 0 \\ \sum_j \pi_{\ell_i(n,j)}(\cdot)(1-\varphi) - \xi^b(\sum_{j=e,w}\pi_{\ell_i(n,j)}(\cdot)(1-\varphi)) & \text{if } \sum_{j=e,w}\pi_{\ell_i(n,j)}(\cdot) < 0 \end{cases}.$$

$$W_i^{x=1}(n,\cdot,\mu,z,s,z',s';\sigma_{-i}) = \max\left\{0, \sum_j \pi_{\ell_i(n,j)}(n,j,c^n,\mu,z,s,z',s';\sigma_{-i})\right\}.$$

Note that there is full certainty about the bailout and that the bank receives funds from the government only when realized profits are negative.[41] These losses are paid for by taxes as in the case of the deposit insurance. In Table 23, we present the results of the case when $\varphi = 1$.

[41]More generally, one might think that the probability of a bailout is in $[0,1]$ not $\{0,1\}$, but this induces a much more complicated computational algorithm where the evolution of the banking industry depends on the realization of government bailouts.

Table 23: Benchmark vs Model with National Banks Bailouts

Moment	Benchmark	Nat. Bank Bailout Change (%)
Loan Supply	0.78	6.13
Loan Interest Rate (%)	5.69	-8.85
Markup	108.44	-15.04
Market Share bottom 99%	39.64	-7.06
Market Share Top 10 / Top 1%	20.97 / 39.38	52.02 / -20.57
Prob. Exit Top 10 / Top 1%	0 / 1.67	- / 65.87
Borrower Risk Taking R (%)	14.78	-0.02
Default Frequency (%)	1.22	-2.13
Entry/Exit Rate (%)	2.78	-0.11
Int. Output	0.89	6.15
Taxes/Output (%)	17.84	9.79

Along the equilibrium path national banks make negative profits which introduces government bailouts when the economy heads into a recession. The unconditional probability of a government bailout equals 2.72% and it can cost up to 5.29% of output. The introduction of "big" banks bail outs increases the level of taxes over GDP necessary to cover the losses (a 9.79% increase).

The introduction of government bailouts induces the national bank to increase its exposure to the region with the highest risk. This "excessive" risk taking behavior is what concerns policymakers. The highest fraction of defaults happens when the regional shock changes (i.e. $s = j$ and $s' \neq j$) during a recession $z = z_b$. Thus, provided that regional shocks are persistent, borrowers in different regions have a different risk profile. A national bank will increase its exposure to risk if it increases the amount of loans to the region with the good regional shock during good times. This is precisely what happens under the too big to fail policy. In Table 24, we compare the loan decision rules for dominant banks when $\mu\{1, 1, 1, \cdot\}$, $z = z_b$ and $s = e$ in the benchmark model versus the National Bank Bailouts policy.

Table 24: Benchmark vs Too Big to Fail

	Loan Decision Rules $\ell(\theta, j, \mu, z_g, e)$ ($\mu = \{1, 1, 1, \cdot\}, z = z_b, s = e$)			
Model	$\ell(n, e, \cdot)$	$\ell(n, w, \cdot)$	$\ell(r, e, \cdot)$	$\ell(r, w, \cdot)$
Dynamic (benchmark)	7.209	82.562	45.450	31.483
National Bank Bailouts	85.837	82.562	32.668	31.483

Under the National Bank Bailouts policy, the national bank extends 10 times more loans in region e than in our benchmark economy. If $s' = w$ is realized, the government will effectively need to bailout the national bank (this is an on-the equilibrium-path action). Induced by the actions of national banks, regional banks, however, extend less loans than before. The net effect results in an increase of the aggregate loan supply (+6.13%). This change in aggregate loans made results in a lower interest rate (-8.85%). Since more projects are financed output increases (+6.15%).

The reduction in the interest rate induces a reduction the in default frequency and bank profitability generating a drop in markups (-15.04%) than induces a lower entry rate (-0.11%).

8 Concluding Remarks

We organize Call Report data from commercial banks in the U.S. from 1976 to 2008 (the same data employed by Kashyap and Stein [30]) in order to be consistent with a structural model of the banking industry. In particular, we document that exit by failure is countercyclical, total loans and deposits are procyclical, loan returns and markups are countercyclical and delinquency rates are countercyclical. We also organize the data by bank size: Top 10, Top 1%, and bottom 99%. We show there are important differences between small and large banks using these categories. For example, we find that smaller banks have higher returns and higher volatility of returns than large banks, which is consistent with a model of geographic diversification. We also document that many of the measures of the degree of imperfect competition in a market used in industrial organization suggest the market structure is not perfectly competitive.

We provide a model where "big" geographically diversified banks coexist in equilibrium with "smaller" regional and fringe banks that are restricted to a geographical area. Since we allow for regional specific shocks to the success of borrower projects, small banks (both regional and fringe) may not be well diversified. This assumption generates ex-post differences between big and small banks. As documented in the data section, the model generates not only procyclical loan supply but also countercyclical interest rates and returns on loans, bank failure rates, default frequencies, charge-off rates. Since bank failure is paid for by lump sum taxes to fund deposit insurance, the model predicts countercylical taxes. Also, the model generates differences in loan interest rates, loan returns, profit rates and default frequencies between banks of different sizes (national, regional and fringe) since large banks are able to diversify across both regions. The variance of returns is also a decreasing function of bank size but it is smaller than in the data.

The benefit of our model relative to the existing theoretical literature is that the number of banks is derived endogenously and varies over the business cycle. One novel part of our environment is that the Stackelberg game between banks provides an framework whereby "big banks" actions in the loan market spread to affect the profitability of smaller banks. Thus it provides a complement to the systemic effects which act through interconnected balance sheets. Another novel aspect of the model is that countercyclical markups in the financial industry can provide an amplification mechanism for the business cycle. For instance, more bank exit in a downturn reduces competition which raises interest rates, thereby increasing risk taking behavior by borrowers (i.e. default frequencies rise via borrowers' choice of R).

Given our structural model, we run a series of counterfactuals. First, to disentangle the effects of bank competition on default frequencies, borrower returns, bank exit rates and intermediated output we run a counterfactual where we increase entry costs into the banking sector to endogenously generate a more concentrated industry. As in Allen and Gale [4], we find that a reduction in the level of competition reduces bank exit. Moreover, in line with the predictions of Boyd and De Nicolo [11], less competition increases interest rates and induces borrowers to take on more risk resulting in higher default frequencies. We also quantify one aspect of the effects of a "too-big-to-fail" policy. As expected, "too-big-to-fail" induces big banks to extend more loans in risky states (i.e. increase their exposure), but this spills over to smaller banks inducing them to lower their loan supply exposure. The net effect is actually a fall in interest rates on loans, lower default frequencies, and higher intermediated output. This comes at the expense of the regional and fringe banks however; their market shares shrink.

There is much work left to do. First, in a companion paper Corbae and D'Erasmo [14], we expand the bank balance sheet by including net assets (securities minus borrowings). While our current paper captures the largest part of assets on the bank balance sheet (about 67%), security holdings are the next highest (about 22%). While this adds another state variable to our analysis, it allows us to consider interesting policy experiments like the effects of raising capital requirements.[42] Since smaller banks have higher volatility of deposits than larger banks (a fact we find in the data), this implies smaller banks should have higher capital ratios via a precautionary motive like a buffer stock model of saving. Again, this prediction is consistent with data. We then undertake a policy experiment; what are the effects of raising capital requirements on banks of different sizes (an industry equilibrium version of Van Den Heuvel [40])? Further, as in Kashyap and Stein [30] we study whether the impact of Fed policy on lending behavior is stronger for banks with less liquid balance sheets.

Second, we remove the segmented markets assumption in this paper and embed the model into a general equilibrium environment with infinitely lived risk averse consumers who can hold a portfolio of bank stocks in Corbae and D'Erasmo [15]. The paper can be thought of as Gowrishankaran and Holmes [24] meets Hopenhayn and Rogerson [27]. While this paper assumed a risk neutral deep pockets investor who held a portfolio of bank stocks and made "seasoned" equity injections, the new paper allows us to endogenize the bank discount factor which may have important implications for risk taking.

[42] The environment requires us to keep track of the distribution of bank assets with aggregate shocks so we use an approximation technique as in Krusell and Smith [32].

References

[1] Aliaga Diaz, R. and M., Olivero (2011) "THe cyclicality of price-cost margins in banking: An empirical analysis of its determinants," *Economic Inquiry*, 49, 26-46.

[2] Aguirregabiria, V., R. Clark, and H. Wang. (2010) "Bank Expansion after the Riegle-Neal Act: The Role of Diversification of Geographic Risk", unpublished manuscript.

[3] Allen, F., E. Carletti and R. Marquez (2008) "Credit Market Competition and Capital Regulation", unpublished manuscript.

[4] Allen, F. and D. Gale (2000) *Comparing Financial Systems,* Cambridge, MA: The MIT Press.

[5] Allen, F. and D. Gale (2000) "Financial Contagion", *Journal of Political Economy,* Vol. 108, p. 1-33.

[6] Beck, T., A. Demirguc-Kunt, and R. Levine (2003) "Bank concentration and crises", NBER Working Paper 9921.

[7] Berger, A. L. Klapper and R. Turk-Ariss (2008) "Banking competition and financial stability," *Journal of Financial Services Research*, pp. 99-118.

[8] Bergstresser, D. (2004) "Market concentration and commercial bank loan portfolios ", Harvard Business School, manuscript.

[9] Bernanke, B., M. Gertler, and S. Gilchrist (1999) "The Financial Accelerator in a Quantitative Business Cycle Framework", in John Taylor and Michael Woodford (eds.), *The Handbook of Macroeconomics*, Amsterdam: North-Holland.

[10] Bikker, J.A. and K. Haaf (2002) "Competition, concentration and their relationship: An empirical analysis of the banking industry," *Journal of Banking & Finance*, Vol. 26(11), p. 2191-2214.

[11] Boyd, J. and G. De Nicolo (2005) "The Theory of Bank Risk-Taking and Competition Revisited", *Journal of Finance*, 60, p. 1329-43.

[12] Carlstrom, C. and T. Fuerst (1997) "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis", *American Economic Review*, 87, p. 893-910.

[13] Cooley, T. and V. Quadrini (2001) "Financial Markets and Firm Dynamics", *American Economic Review*, 91(5), p. 1286-1310.

[14] Corbae, D. and P. D'Erasmo (2012) "Capital Requirements in a Quantitative Model of Banking Industry Dynamics", mimeo.

[15] Corbae, D. and P. D'Erasmo (2013) "A Simple Quantitative General Equilibrium Model of Banking Industry Dynamics", mimeo.

[16] den Haan, W., S. Sumner and G. Yamashiro (2007) "Bank loan portfolios and the monetary transmission mechanism," *Journal of Monetary Economics*, 54(3), 904-924.

[17] Diamond, D. (1984) "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies*, 51, p. 393-414.

[18] Diamond, D. and P. Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity", *Journal of Political Economy*, 91, p. 401–19.

[19] Diaz-Gimenez, J., E. Prescott, F. Alvarez, and T. Fitzgerald (1992) "Banking in Computable General Equilibrium Models", *Journal of Economic Dynamics and Control*, 16, p.533-60.

[20] Diebold F. and K. Yilmaz (2009) "Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets," *Economic Journal*, 119(534), 158-171.

[21] Ennis, H. (2001) "On the Size Distribution of Banks", Federal Reserve Bank of Richmond *Quarterly Review*, 87, p. 1-25.

[22] Ericson, R. and A. Pakes (1995) "Markov-Perfect Industry Dynamics: A Framework for Empirical Work", *Review of Economic Studies*, 62, p. 53-82.

[23] Farias, V., B. Ifrach and G. Weintraub (2011) "A Framework for Dynamic Oligopoly in Concentrated Industries", *manuscript*.

[24] Gowrisankaran, G. and T. Holmes (2004) "Mergers and the Evolution of Industry Concentration: Results from the Dominant-Firm Model," *RAND Journal of Economics*, 35(3), p. 561-582.

[25] Hennessy, C. and T. Whited (2007) "How Costly is External Financing? Evidence from a Structural Estimation", *Journal of Finance*, LXII(4), p. 1705-45.

[26] Hopenhayn, H. (1992) "Entry, Exit, and Firm Dynamics in Long-Run Equilibrium", *Econometrica*, 60, p. 1127-50.

[27] Hopenhayn, H. and R. Rogerson (1993) "Job Turnover and Policy Evaluation: A General Equilibrium Analysis", *Journal of Political Economy*, 101, p. 915-38.

[28] Huggett, M. (1993) "The risk-free rate in heterogeneous-agent incomplete-insurance economies", *Journal of Economic Dynamics and Control*, 17, p. 953-69.

[29] Janicki, H. and E. Prescott (2006) "Changes in the Size Distribution of U.S. Banks: 1960-2005", Federal Reserve Bank of Richmond *Quarterly Review*, 92, p. 291-316.

[30] Kashyap A. and J. Stein (2000) "What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?" *American Economic Review*, vol. 90(3), p. 407-428.

[31] Koeppl, T. and J. MacGee (2009) “What broad banks do, and markets don’t: Cross-subsidization” *European Economic Review*, vol. 53, p. 222-236.

[32] Krusell P. and Anthony Smith (1998) “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy*, 106(5), 867-896.

[33] Liang, N. and S. Rhoades (1988) “Geographic Diversification and Risk in Banking” *Journal of Economics and Business*, vol. 40, p. 271-284.

[34] Martinez-Miera, D. and R. Repullo (2008) “Does Competition Reduce the Risk of Bank Failure?,” *Review of Financial Studies*, 23, p. 3638-3664.

[35] Pakes, A. and P. McGuire (1994) “Computing Markov Perfect Nash Equilibrium: Numerical Implications of a Dynamic Differentiated Goods Model”, *RAND Journal of Economics*, 25, p. 555-589.

[36] Panzar, J.C. and J.N. Rosse (1987) “Testing for monopoly equilibrium,” *Journal of Industrial Economics*, 35, 443456.

[37] Rhoades, S. (2000) “Bank Mergers and Banking Structure in the United States, 1980-98” *Board of Governors of the Federal Reserve System*, Staff Report 174.

[38] Shaffer, S. (2004) “Patterns of Competition in Banking”, *Journal of Economics and Business*, Vol. 56, p. 287-313.

[39] Townsend, R. (1979) “Optimal contracts and competitive markets with costly state verification” *Journal of Economic Theory*, 22, p. 265293.

[40] Van Den Heuvel, S. (2008) “The Welfare Cost of Bank Capital Requirements” *Journal of Monetary Economics*, 55, p. 298-320.

[41] Williamson, S. (1986) “Costly Monitoring, Financial Intermediation and Equilibrium Credit Rationing”, *Journal of Monetary Economics*, 18, p. 159-179.

Appendix

A-1 Data Appendix

We focus on individual commercial banks in the U.S. We compile a large panel of banks from 1976 to 2010 using data for the last quarter of each year. The source for the data is the Consolidated Report of Condition and Income (known as Call Reports) that banks submit to the Federal Reserve each quarter.[43] Report of Condition and Income data are available for all banks regulated by the Federal Reserve System, Federal Deposit Insurance Corporation, and the Comptroller of the Currency. All financial data are on an individual bank basis.

We follow Kashyap and Stein [30] and den Haan, Summer and Yamashiro [16] in constructing consistent time series for our variables of interest. There was a major overhaul to the Call Report format in 1984. Since 1984 banks are, in general, required to provide more detailed data concerning assets and liabilities. Thus, the complexity of the panel requires careful work. Due to changes in definitions and the creation of new variables after 1984, some of the variables are only available after this date. To avoid miss-reporting errors, we perform the following filter to the original data: Year-bank observations for which total assets (RCFD2170) or total loans (RCFD1400) have a non positive entry are deleted from our sample. We also restrict the bank universe to insured banks that are chartered as commercial banks (including depository trust companies, credit card companies with commercial bank charters, private banks, development banks, limited charter banks, and foreign banks). Finally, we only include banks located within the fifty states and the District of Columbia. (0 $<$RSSD9210$<$ 57).

To deflate balance sheet and income statement variables we use the CPI index. To compute business cycle correlations, variables are detrended using the HP filter with parameter 6.25. When we report weighted aggregate time series we use the loan market share as weight. To control for the effect of a small number of outliers, when constructing the loan returns, cost of funds, charge offs rates and related series we restricted our sample to the interval defined by $\pm$ 5 standard deviations from the mean. The estimated markups and the Lerner index are restricted to the $[2, 99]$ percentiles interval. We also control for the effects of bank entry, exit and mergers by not considering the initial period, the final period or the merger period (if relevant) of any given bank. We use data provided by the Federal Reserve of New York and data available in the $RSSD$ series to reconstruct merger activity at the bank level.

Tables 25, 26 and 27 present the balance sheet variables, the income statement variables and derived variables used respectively.

[43] Balance Sheet and Income Statements items can be found at https://cdr.ffiec.gov/public/.

Table 25: Balance Sheet Variables

Name	Series	Dates
Assets	rcfd2170	1976 - 2010
Loans	rcfd1400	1976 - 1983
	rcfd1400-rcfd2165	1984 - 2010
C&I Loans (dom. Offices)	rcon1766	1984 - 2010
Loans secured by real estate (dom. offices)	rcon1410	1976 - 2010
Consumer Loans (dom. offices)	rcon1975	1976 - 2010
Agricultural loans (dom. offices)	rcon1590	1976 - 2010
Non-Accrual Loans	rcfd1403	1983 - 2010
Loans Past Due 90 days or more	rcfd1407	1983 - 2010
Loans Past Due 30-89 days or more	rcfd1406	1983 - 2010
Cash	rcfd0010	1976 - 2010
Securities	rcfd0400+rcfd0600	1976 - 1983
	+rcfd0900+rcfd0380	
	rcfd0390+rcfd2146	1984 - 1993
	rcfd1754+rcfd1773	1994 - 2010
Total Liabilities	rcfd2948	1976 - 2010
Total Deposits	rcfd2200	1976 - 2010
Total Liabilities Net of Sub. Debt	rcfd2950	1976 - 2010
Demand Deposits	rcfd2210	1976 - 2010
Bank Equity/Capital	rcfd3210	1976 - 2010

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports.

Table 26: Income Statement Variables

Name	Series	Dates
Interest Income From Loans	riad4010	1984 - 2010
Interest Income From Loans Secured by RE (dom. offices)	riad4011	1984 - 2010
Interest Income From C&I Loans (dom. offices)	riad4012	1984 - 2010
Interest Income From Consumer Loans (dom. offices)	riad4013	1984 - 2010
Interest Income all sources	riad4107	1984 - 2010
Total Operating Income	riad4000	1976 - 2010
Interest Expense on Deposits	riad4170	1976 - 2010
Interest Expense all sources	riad4073	1984 - 2010
Interest Expense on Fed Funds	riad4180	1976 - 2010
Total Operating Expense	riad4130	1976 - 2010
Net Interest Income	riad4074	1984 - 2010
Total Non Interest Income	riad4079	1984 - 2010
Total Non Interest Expense	riad4093	1984 - 2010
Salaries	riad4135	1976 - 2010
Total Net Income	riad4301	1976 - 2010
Total Net Income Net of Taxes	riad4300	1976 - 2010
Charge offs all loans	riad4635	1976 - 2010
Recoveries all loans	riad4605	1976 - 2010
Charge offs Loans Secured by RE	riad4613	1984 - 2010
Recoveries Loans Secured by RE	riad4616	1984 - 2010
Charge offs C&I	riad4638	1984 - 2010
Recoveries C&I loans	riad4608	1984 - 2010
Charge offs consumer loans	riad4639	1984 - 2010
Recoveries consumer loans	riad4609	1984 - 2010
Loan Loss provision	riad4230	1976 - 2010
Transfer Risk	riad4243	1976 - 2010
Expenditures on fixed assets	riad4217	1976 - 2010
Sale, conversion, acquisition of capital stock	riadb509	2001 - 2010

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports.

Table 27: Derived Variables

Entry Rate	# banks enter in t / total banks in t
Exit Rate	# banks exit in t / total banks in $t-1$
Return on Loans	(1+Int. Income From Loans / Loans)/(1+Inf. Rate) - 1
Cost of Deposits	(1+ Int. Expense From Dep. / Deposits)/(1+Inf. Rate)-1
Cost of Funds	[1+ Int. Exp From Dep. and Fed Funds / (Deposits+Fed Funds)]/(1+Inf. Rate)-1
Net Interest Margin	Return on Loans - Cost of Deposits
Net Charge off Rate	(Charge offs - Recoveries) / Loans
Delinquency Rate	(Loans Past Due 90 days or more + Non Accrual Loans) / Loans
Input Price w_1	Salaries / Employees
Output y_1	Loans
Output y_2	Assets-Loans-Cash-Fixed Assets
Netput z	Equity
Non-Int Exp	Salaries + Loan Loss Provision
Fixed Costs	Expenditures on Fixed Assets

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports.

In Table 28 we present the summary statistics of selected variables.

Table 28: Summary Statistics Selected Variables

Variable	# Obs.	Mean	Std. Dev.
Assets	366,732	413,522.00	9,400,472.00
Loans	366,732	237,761.20	4,702,208.00
C&I Loans	251,661	65,547.41	1,036,726.00
Loans Secured by Real Estate	366,732	107,257.50	2,272,504.00
Consumer Loans	366,732	38,204.73	707,009.80
Cash	366,732	32,449.85	705,714.70
Securities	366,732	73,961.38	1,374,771.00
Deposits	366,732	291,045.10	5,934,225.00
Liabilities	251,699	498,926.00	10,400,000.00
Equity	366,732	34,530.02	789,795.20
Interest Income From Loans	251,662	24,941.76	387,469.90
Interest Income	251,663	33,799.40	550,091.60
Interest Expense Deposits	366,710	10,095.15	161,442.30
Interest Expense	251,663	16,460.20	285,421.10
Non Interest Income	251,663	10,751.36	239,071.80
Non Interest Expense	251,663	17,466.09	314,241.00
Net Income	251,663	7,682.32	162,953.70
Charge offs	366,710	2,247.21	58,674.62

Note: Data corresponds to commercial banks in the US. Source: FDIC, Call and Thrift Financial Reports.

A-2 Investor's Problem

Let Y represent the endowment of investors and $\overrightarrow{S} = \{S_i(\theta)\}_{i=1,\theta=n,r,f}^{N(\theta)}$ denote the vector of share holdings in each bank i. Investors choose the number of shares in each bank to maximize their present discounted value of current and future cash flows. The recursive representation of their problem is:

$$J(z, s, \mu, \overrightarrow{S}) = \max_{\overrightarrow{S}'} \beta E_{z',s'|z,s} \left[\mathcal{C} + J(z', s', \mu', \overrightarrow{S}') \right] \tag{A.2.1}$$

subject to

$$\mathcal{C} + \sum_{i=1}^{N(\theta)} \sum_{\theta} P_i(\theta, z', s', \mu') S_i'(\theta) = \sum_{i=1}^{N(\theta)} \sum_{\theta} S_i(\theta) \left[\mathcal{D}_i(\theta, z, s, \mu, z', s') + P_i(\theta, z', s', \mu') \right] + Y. \tag{A.2.2}$$

where $P_i(\theta, z', s', \mu')$ is the price of a share of bank i after the realization of z' and s'.

The first order condition is:

$$S_i'(\theta) : P_i(\theta, z, s, \mu) = \beta E_{z',s'|z,s} \left[\mathcal{D}_i(\theta, z, s, \mu', z', s') + P_i(\theta, z', s', \mu') \right] \tag{A.2.3}$$

Normalizing the number of shares of each bank to 1 and letting the price of such share of bank i be $P_i(\theta, z, s, \mu) = V_i(\theta, z, s, \mu) - \beta E_{z',s'|z,s}[\mathcal{D}_i(\theta, z, s, \mu', z', s')]$ where $V_i(\theta, z, s, \mu)$ is the present discounted value of bank i dividends, the (A.2.3) can be written:

$$V_i(\theta, z, s, \mu) - \beta E_{z',s'|z,s}[\mathcal{D}_i(\theta, z, s, \mu', z', s')] = \beta E_{z',s'|z,s} \left[V_i(\theta, z', s', \mu') \right] \iff$$
$$V_i(\theta, z, s, \mu) = \beta E_{z',s'|z,s} \left[\mathcal{D}_i(\theta, z, s, \mu', z', s') + V_i(\theta', z', s', \mu') \right]$$

which is the dynamic programming problem of each bank i we are solving.

A-3 Computational Algorithm

In this section, we describe the computational algorithm that allows as to compute bank strategies and the equilibrium of the model. We use an extension of the algorithm proposed by Pakes and MacGuirre (1994 and 2000) that incorporates a competitive fringe. The description is based on the case where fringe banks do not have access to outside funding, entering fringe banks draw of c^f exceeds the highest cost of fringe incumbents in the market that period and the number of representative national banks and regional banks in each region is restricted to one.

After giving a functional form to the utility function, the failure probability, the equity issuance cost and the distribution of net costs for fringe banks as well as defining a grid and a transition probability for the aggregate shock the steps to compute the algorithm are the following:

1. Solve the entrepreneur problem. This amounts to define a grid over r^L and for each 4-tuple $\{r^{L,j}, z, s, j\}$ obtain the type of project the entrepreneur operates $R^j(r^{L,j}, z, s)$ and the decision on whether to operate the technology or not $\iota^j(\omega, r^L, z, s)$. Using $\iota^j(\omega, r^L, z, s)$ we can derive the loan demand in each region j: $L^{d,j}(r^{L,j}, z, s)$ (equation (12)).

2. Solve for the reaction function of the fringe sector $L^s(f, j, r^{L,j}, z, s)$ as a function of the interest rate. For each 4-tuple $\{r^{L,j}, z, s, j\}$ and using the solution to the entrepreneurs problem $R^j(r^{L,j}, z, s)$ to derive the failure probability, compute expected profits for fringe banks with cost c_f in $\Xi(c_f)$ and derive $\overline{c}^j(r^{L,j}, z, s)$, i.e. the maximum cost such that a fringe bank would choose to operate and offer loans. The loans supply from fringe banks is
$$L^s(f, j, r^{L,j}, z, s) = \left[M\Xi(\overline{c}^j(r^{L,j}, z, s))\right] \bar{d}.$$

3. Using the aggregate loan demand and the loan supply by the fringe sector, define the residual loan demand $\tilde{L}^{d,j}(r^{L,j}, z, s)$ as follows:
$$\tilde{L}^{d,j}(r^{L,j}, z, s) = \max\{0, L^{d,j}(r^{L,j}, z, s) - L^s(f, j, r^{L,j}, z, s)\}$$

4. Define a grid over the number of possible active dominant banks $\{\{0,0,1\},\{0,1,0\}, \{0,1,1\},\{1,0,0\},\{1,0,1\},\{1,1,0\},\{1,1,1\}\}$ where the first element in each set corresponds to an indicator of whether the representative national bank is active or not, the second element corresponds to an indicator for the regional bank in region e and the third element the indicator for regional bank in region w.

5. For each state $\{\mu, z, s\}$, solve the problem of dominant banks (that also determines the solution of the fringe banks). We use value function iteration in this step:

   (a) Guess a loan decision rule and value function for each dominant banks: $\ell^0(\theta, j, \mu, z, s)$ and $V^0(\theta, \mu, z, s)$ for $\theta = n, r$ and $j = e, w$.

(b) For each bank, solve the optimal loan and exit decision rule taking as given the strategy of other banks. Note that the equilibrium in the loan market in region j (i.e. the equilibrium interest rate) is derived from the following equation:

$$\ell^0(n, j, \mu, z, s) + \ell^0(r, j, \mu, z, s) = \tilde{L}^{d,j}(r^{L,j}, z, s). \tag{A.3.4}$$

where $\tilde{L}^{d,j}(r^{L,j}, z, s)$ is the residual loan demand defined above. The solution to this problem will provide a loan decision rule $\ell^1(\theta, j, mu, z, s)$, an exit decision rule and a new value function $V^1(\theta, \mu, z, s)$.

(c) If $|V^1(\theta, \mu, z, s) - V^0(\theta, \mu, z, s)| < \epsilon^v$ and $|\ell^1(\theta, j, \mu, z, s) - \ell^0(\theta, j, \mu, z, s)| < \epsilon^\ell$ for ϵ^v and ϵ^ℓ small you have obtained the solution to the bank problem and can continue. If not, set $V^0(\theta, \mu, z, s) = V^1(\theta, \mu, z, s)$, $\ell^0(\theta, j, \mu, z, s) = \ell^1(\theta, j, \mu, z, s)$ and return to previous step.

(d) The equilibrium dominant banks loan decision rules together with equation (A.3.4) determine the equilibrium loan interest rate $r^{L,j}(\mu, z, z)$, the equilibrium loan supply by fringe banks $L^s(f, j, r^{L,j}, z, s)$ as well as the thresholds $\overline{c}^j(r^{L,j}, z, s)$ and $c^x(r^{L,j}, z, s, z', s')$ for each region j.

6. For states where $\mu = \{0, 0, 0\}$, the equilibrium interest rate can be derived from $L^{d,j}(r^{L,j}, z, s) = L^s(f, j, r^{L,j}, z, s)$.

7. Solve the entry problem of dominant banks, that is select $e(\theta) \in \{0, 1\}$. Each dominant bank will enter (i.e. $e(\theta) = 1$) if there is other dominant bank of the same type and

$$-\Upsilon^\theta + V(\theta, \hat{\mu}, z, s) \geq 0.$$

where $\hat{\mu}$ is the distribution that would arise if the bank of type θ enters and takes into account the entry decision rule by other banks.

8. Using the exit and entry decision rules of dominant banks we can define the evolution of the distribution $\mu' = H(\mu, z, s, z', s')$.

9. Verify that given the level of risk aversion and the equilibrium interest rate, the consumer prefers to bring its deposit to the bank (or use the storage technology) rather than lend it directly to the entrepreneur.