A Quantitative Model of Banking Industry Dynamics*

Dean Corbae
University of Wisconsin - Madison and NBER

Pablo D'Erasmo†
Federal Reserve Bank of Philadelphia

October 14, 2024

Abstract

We develop a simple 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. Given the model has a subset of non-atomistic banks, we study the costs and benefits of too-big-to-fail policies.

*PRELIMINARY. The authors wish to thank Andy Atkeson, John Boyd, Jeff Campbell, Gianni DeNicolo, Hugo Hopenhayn, 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, McMaster, Tilburg, UCLA, USC, University of Toronto, University of Wisconsin, the BU/Boston Fed Conference on Macro-Finance Linkages, and the Workshop on Financial Systems (CFSP) at MIT.

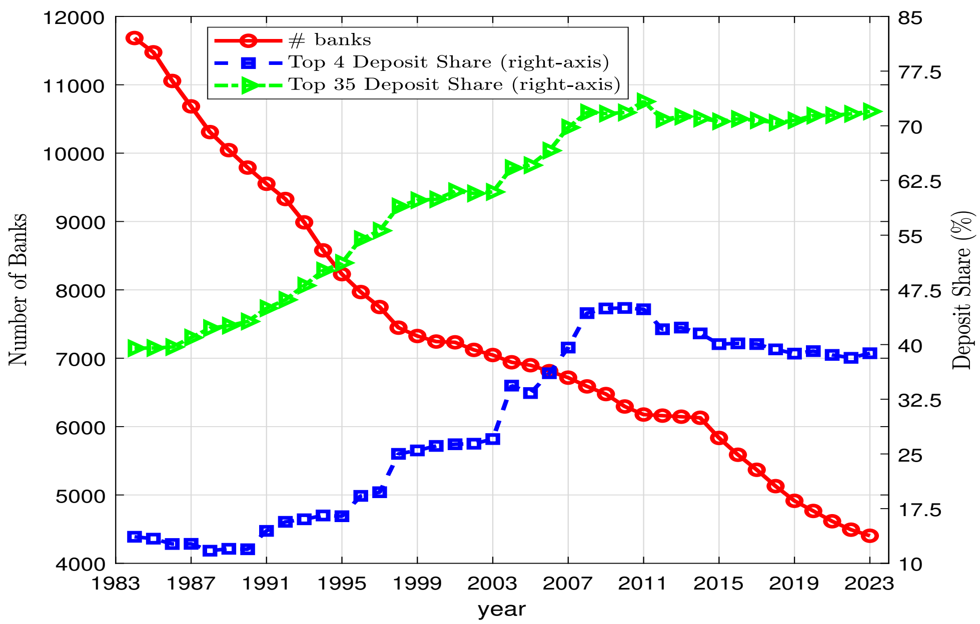
†The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

"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."
Ben Bernanke, Time, December 28, 2009/January 4, 2010, p. 78.

1 Introduction

The objective of this paper is to formulate a quantitative structural model of the banking industry consistent with data in order to assess the impact of too-big-to-fail policies on bank risk taking. 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 as in a Diamond (1984) delegated monitoring model. 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.

Figure 1: Concentration Measures and Number of Banks



Note: Top 4 Deposit Share corresponds to the share of total deposits accounted by the Top 4 banks (when sorted by deposits). Top 35 Deposit Share corresponds to the share of total deposits accounted by the Top 35 banks. Source: Reports of Condition and Income (Call Reports)

Our model assumes spatial heterogeneity between banks; there is a representative "national" geographically diversified bank (representing too-big-to-fail banks like Chase) that coexist in equilibrium with big "regional" (such as SVB) and "fringe" banks that are both restricted to a geographical area. This breakdown is roughly consistent with a market structure as in Figure 1 where the Top 4 banks are associated with national banks (which have approximately 40% of the deposit market share), banks 5 to 35 associated with regional banks (which make up another 30% of deposit market share), and the remaining roughly 4000 banks which are associated with the fringe banks (which make up the remainder of deposit market share).

Since we allow for regional specific shocks to the success of borrower projects, smaller banks' asset portfolios may not be well diversified. This assumption generates ex-post differences between big and smaller (regional and fringe) banks in a given spatial area. As documented in our data Section 2, 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. This would be the benefit but if it creates more risk taking maybe not.

Our model with heterogeneous and non-atomistic banks is well suited to address the following questions: 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 in the presence of moral hazard arising from deposit insurance?

As described above, we require our quantitative model to be consistent with U.S. data on market concentration. This motivates us to relax the assumption of a perfectly competitive banking sector with exogenous exit and a representative bank as assumed in the important central bank workhorse model of Gertler and Karadi (2011). 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. Specifically, we apply the Markov Perfect Industry equilibrium concept of Ericson and Pakes (1995) augmented with a competitive fringe along the lines of Gowrisankaran and Holmes (2004). The framework actually nests the competitive model under certain assumptions on entry costs.

Idiosyncratic shocks to loan payoffs and deposit funding along with endogenous exit and entry 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, non-atomistic banks (i.e. banks which are not of measure zero as in a competitive economy) 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. Further, the fact that some banks are non-atomistic implies there is a well-defined pecuniary externality; any given big bank internalizes its effect on its own profitability and the reaction of others but not their negative consequences. Here, unlike models with atomistic banks, we can quantitatively assess the impact of one dominant banks' effect on equilibrium prices and outcomes in an environment with moral hazard associated with deposit insurance and costly exit.

There is a vast empirical literature that takes up the "concentration-stability" versus "concentration-fragility" debate. For example, Beck, Demirgüç-Kunt, and Levine (2006) 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. and Corbae and Levine (2024) 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. As validation of our approach, we find that our model is consistent with results from the empirical literature.

Our main counterfactual studies the effects of bank bailout policies to mitigate exit. In particular, we compare our benchmark economy with one where the government is always 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 so they do not exit on-the-equilibrium path. In the counterfactual case, the representative national bank increases exposure to risk since its continuation (charter) value is guaranteed. Regional and fringe banks reduce their loan supply in order to avoid their own exit. The increase in loans by national banks dominates generating lower interest rates. Due to the risk shifting effect, borrowers take on less risk and default frequencies fall. A lower default frequency leads to less failure by smaller banks and a decline in the tax-to-output ratio necessary to fund failing banks.

Literature: 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 (2004) and Boyd and De Nicolo (2005) which feature an endogenous risk shifting effect.[1] Unlike the previous papers, we do not exogenously fix the number of banks but 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 (2016) and Oberfield et al. (2024)) and both markets in Wang et al. (2022).

While documented rising concentration in the banking industry motivates us to consider a model of imperfect competition, the financial intermediation industry has experienced a major change through the emergence of the less regulated, nonbank industry. Thus, in the

[1] A dynamic version of the model is considered in Martinez-Miera and Repullo (2010).

spirit of work by Buchak et al. (2018) and Begenau and Landvoigt (2022), our model includes competition between banks and nonbanks.

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 parameter estimation and Section 6 provides untargeted validation of the model against business cycle statistics and shows consistency with empirical studies of the role of industry concentration on banking stability. Section 7 conducts our main counterfactual assessing the costs and benefits of too-big-to-fail policy. Section 8 provides directions for future research.

2 Some Banking Data Facts

In this section, we document the evolution of the banking industry since the early 1980's. As in Kashyap and Stein (2000) our main data source is the Consolidated Report of Condition and Income (known as Call Reports) that banks submit to the Federal Reserve each quarter.[2] We compile quarterly data from 1984 to 2023 but for most of the analysis we focus on the period between 1984 and 2019.We follow Kashyap and Stein (2000) 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. 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. We aggregate bank level data to the bank holding company level using information on the regulatory high holder for each bank.[3] We also use the Summary of Deposits (SOD), the annual survey of branch office deposits as of June 30 for all FDIC-insured institutions.[4]

In the data work that follows in this section we organize banks into three different cohorts based on size: top 4, top 5-35, and the rest (approximately 4000 banks). We do so with our model in mind since we will have banks of three different types: national, regional, and fringe. We think of both national and regional banks as being individually systemically important and large enough to individually impact prices. We think of fringe banks as small enough that they do not have individual price impact. This organization of the data is consistent with a dominant-fringe model as in the industrial organization framework of Gowrisankaran and Holmes (2004).

Figure 1 graphs the deposit market share of the top 4 and top 35 banks across time. It shows that prior to the Riegle-Neal Act of 1994 which repealed state level branching restrictions at the national level, market shares were relatively constant. However, following

[2]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/.

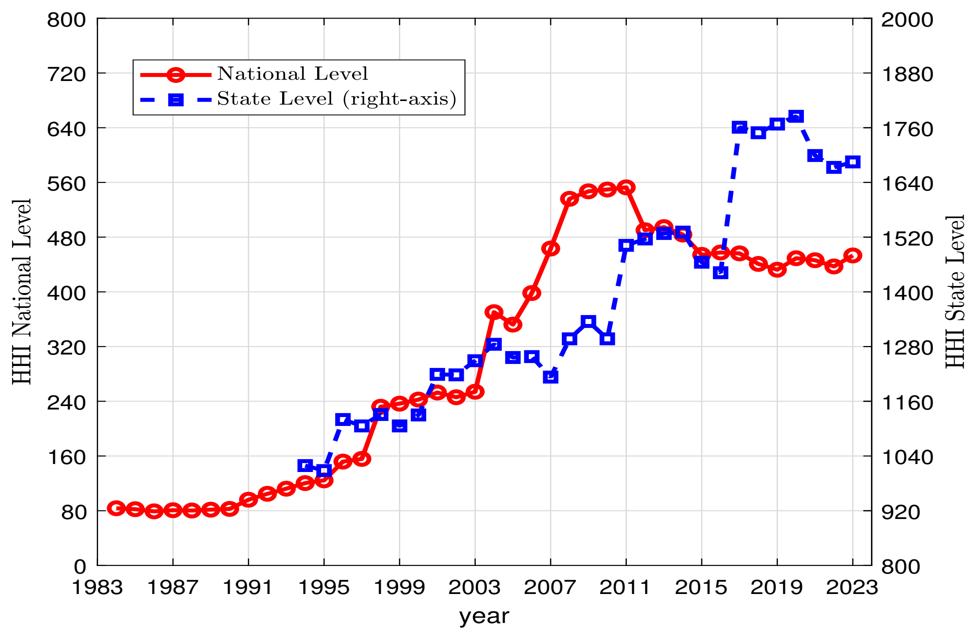
[3]We refer to bank holding companies and individual banks (that do not belong to a bank holding company) as banks. Kashyap and Stein (2000) focus on individual commercial banks in the U.S. As Kashyap and Stein (2000) argue, there are not significant differences in modeling each unit.

[4]Data can be found here https://banks.data.fdic.gov/bankfind-suite/SOD.

Riegle-Neal, there was significant rise in market share of top 4 and top 35 banks until the 2008 financial crisis, followed by another relatively constant share. The figure also documents the threefold drop in the number of banks in the U.S. from nearly 12,000 in 1983 to nearly 4000 in 2023.

Figure 2 graphs deposit Herfindahl Indices (HHI) at the national and state levels.[5] High levels of HHI are considered to anti-competitive. Figure 2 makes clear the large rise in HHI at both the national and state levels following Riegle-Neal until the financial crisis when it started to level off most notably at the national level. The current average of state-level Herfindahl indices falls into the "moderately concentrated" designation by the Justice Department's Antitrust Division. The Herfindahl has grown by 80 percent from 1994.

Figure 2: Deposit Herfindahl–Hirschman Index (HHI)



Source: Reports of Condition and Income (Call Reports) and Summary of Deposits (SOD)

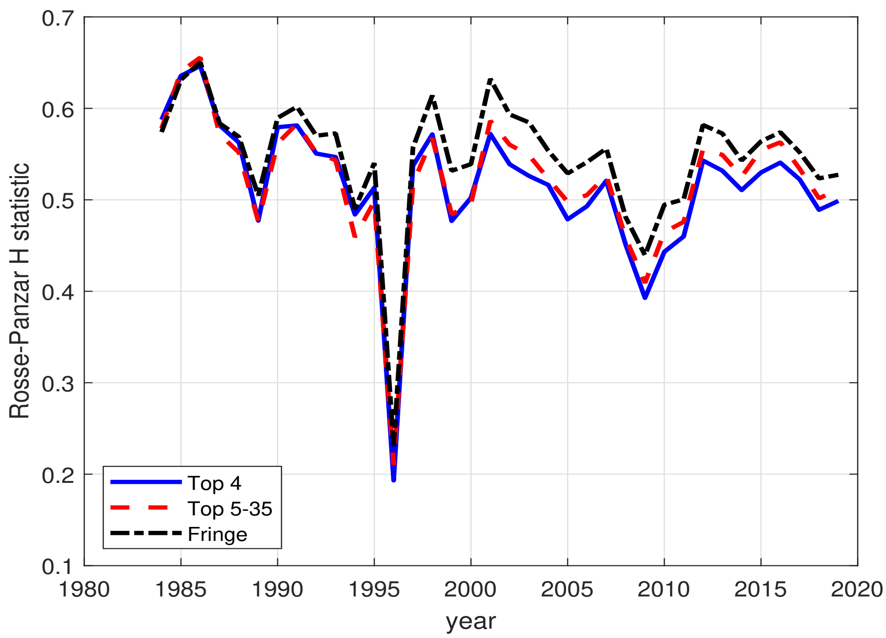
To obtain an additional measure of competition, we follow Shaffer (2004) and estimate the elasticity of marginal revenue with respect to factor prices (i.e. the Rosse-Panzar H statistic) by a log-linear regression in which the dependent variable is the natural logarithm of total revenue (or income) ($\ln(TR_{it})$ and the explanatory variables include the logarithms of input prices ($w_{1_{it}}$ funds (deposits and other borrowings), $w_{2_{it}}$ labor and $w_{3_{it}}$ fixed assets):

$$\ln(TR_{it}) = \alpha_t + \sum_{k=1}^{3} \beta_k \ln(w_{k_{it}}) + u_{it}.$$

[5]The HHI is given by $\sum_i s_i^2$ where s_i is bank i's market share in a given region (in percent). An HHI between 1500 and 2500 points to a market moderately concentrated and one with an HHI of 2500 is considered highly concentrated by the anti-trust division of the Justice Department.

The Rosse-Panzar H statistic equals the simple sum of coefficients on the respective log input price terms, $\beta_1 + \beta_2 + \beta_3$.[6] This equation is estimated year by year with robust standard errors. Figure 3 presents the estimates by bank-size over time. The average Rosse-Panzar H statistic is statistically different from 1 (the value that indicates the presence of perfect competition) with 99% confidence. Using this technique, Bikker and Haaf (2002) 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. We also find an average (across all bank sizes) of 47% and Figure 3 makes evident that this measure of pass-through is lower for larger banks.

Figure 3: Rosse Panzar H-statistic (by bank size)

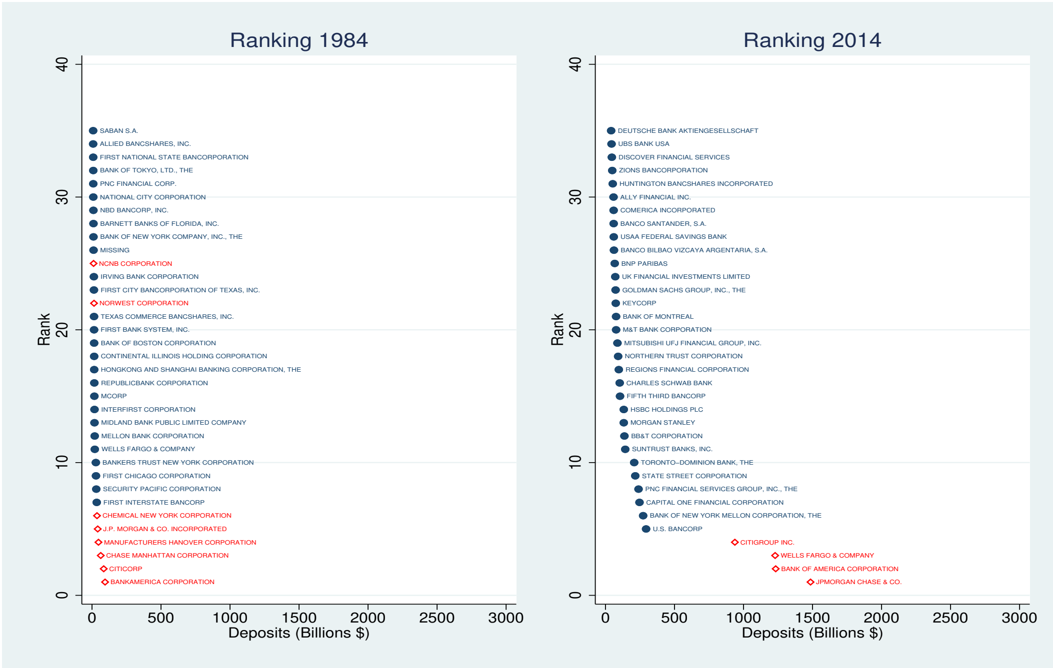


Note: Source: Reports of Condition and Income (Call Reports)

Figure 4 graphs how the top 35 individual banks ranked according to deposit size grew over time. Notable is the divergence of the top 4 even from the remaining top 5-35. This motivates our decision to categorize banks into 3 size bins (top 4, top 5-35, and the rest). We present two years in our sample and highlight in red the charters of banks that will eventually become part of the top 4 via mergers.

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

Figure 4: Deposit Distribution 1984 & 2014

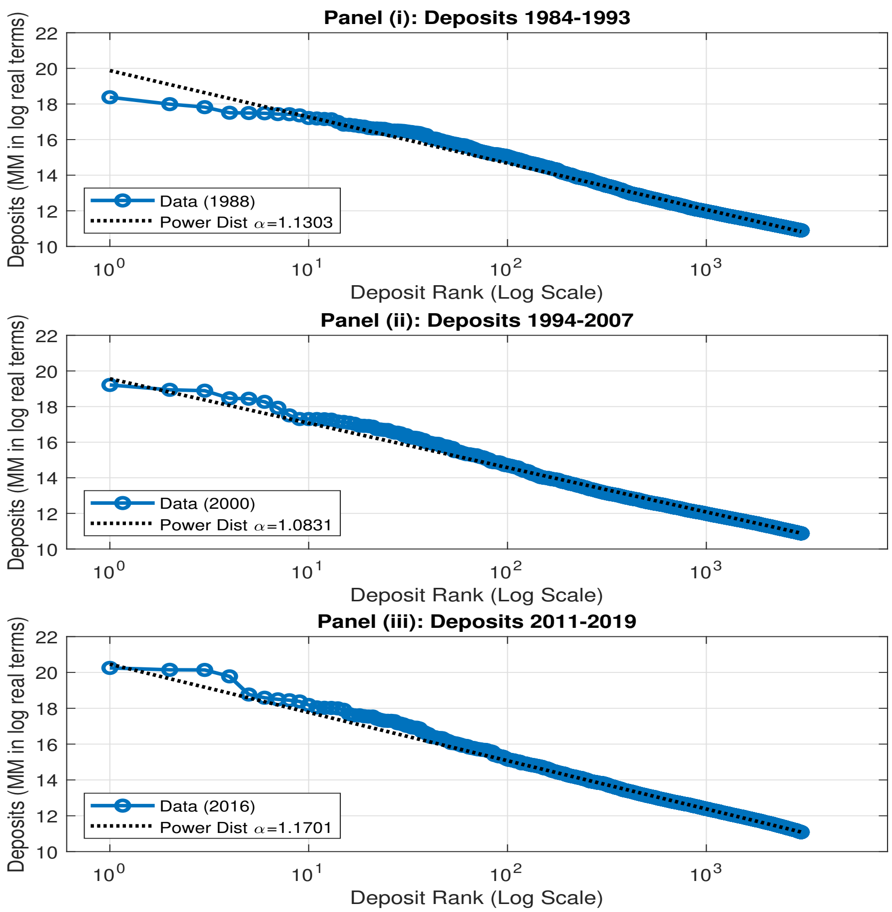


Note: Banks are ranked according to deposits. Deposits are in reported in real terms. Red bubbles identify banks that end up in the Top 4 of the distribution. Source: Reports of Condition and Income (Call Reports)

Some firm dynamics studies have found that the firm size distribution is well approximated by a Power law distribution due to a thicker right tail than the lognormal.[7] Let $d(r) = b \cdot r^{-\alpha}$ where r is the rank and $d(r)$ is the deposits of bank with rank r. Zipf's law predicts that $\alpha = 1$. We estimate α as in Gabaix and Ibragimov (2011) year by year between 1984 and 2019.Figure 5 presents the estimates for three different periods in our sample: 1984-1993 (Panel (i)), 1994-2007 (Panel (ii)), and 2011-2019 (Panel (iii)). We plot the actual distribution in the data for the midpoints of each sample to illustrate the growth of the right tail against the average estimated distribution (i.e., the distribution derived from the average estimated tail parameter during the sample period). We find a significant increase in the (absolute) value of α over time going from -1.043 in the period that goes between 1984 and 1993 to -1.17 in the period that goes between 2011 and 2019 implying that the right tail is getting thicker over time.

[7]See Janicki and Prescott (2006) for a study with a focus on the U.S. bank distribution between 1960 and 2005.

Figure 5: Deviations from Zipf's Law



Note: We estimate α (year by year) as in Gabaix and Ibragimov (2011) and restrict the sample to the top 3000 banks (like Janicki and Prescott (2006)). The reported coefficient in each panel corresponds to the average during the period. The data distribution to the middle point in the sample. Source: Reports of Condition and Income (Call Reports)

Table 1 reports the estimates for assets, loans, and deposits for different years in the sample. All coefficients are significantly different from the Zipf's law prediction of 1 at the 1% significance level.

Table 1: Zipfs Law estimates Assets, Loans, and Deposits (across time)

	Power Exponent α				
	Year				
	1984	1994	2004	2014	2019
Assets	1.160	1.133	1.120	1.171	1.272
(se)	(0.03)	(0.029)	(0.029)	(0.03)	(0.033)
Loans	1.191	1.151	1.108	1.175	1.280
(se)	(0.031)	(0.03)	(0.029)	(0.03)	(0.033)
Deposits	1.129	1.092	1.074	1.147	1.256
(se)	(0.029)	(0.028)	(0.028)	(0.03)	(0.032)

Note: We estimate α and its standard error (se) as in Gabaix and Ibragimov (2011) and restrict the sample to the top 3000 banks (as in Janicki and Prescott (2006)). All estimates significantly different from 1 at the 1% level. Source: Reports of Condition and Income (Call Reports).

We next discuss the balance sheet structure by bank size. In anticipation of the model we present below, we group the asset size into net securities, loans, and fixed and other assets, and the liabilities into deposits and equity. Net securities include cash plus federal funds sold plus securities minus federal funds purchased. Loans includes all loans on the balance sheet and fixed and other assets includes fixed assets and other assets not classified into net securities or loans (except trading assets). Deposits include total deposits and other borrowings (which include home loan advances). By working with this categorization we account for approximately 96% of total assets and about 95% of total liabilities. Table 2 presents the resulting balance sheet (as a ratio of total assets) by bank size. We normalize the categories so the totals add up to 1.

Table 2: Balance Sheet by Bank Size

	Bank Size			
	All Banks	Top 4	Top 5-35	Fringe
Assets (% Total Assets)				
Net Securities (a)	27.62	24.87	26.75	27.60
Loans (ℓ)	63.32	66.81	64.32	62.73
Fixed & Intangible Assets (κ)	9.05	8.32	8.93	9.67
Liabilities (% Total Assets)				
Deposits (d)	90.95	91.68	91.07	90.33
Equity (e)	9.05	8.32	8.93	9.67

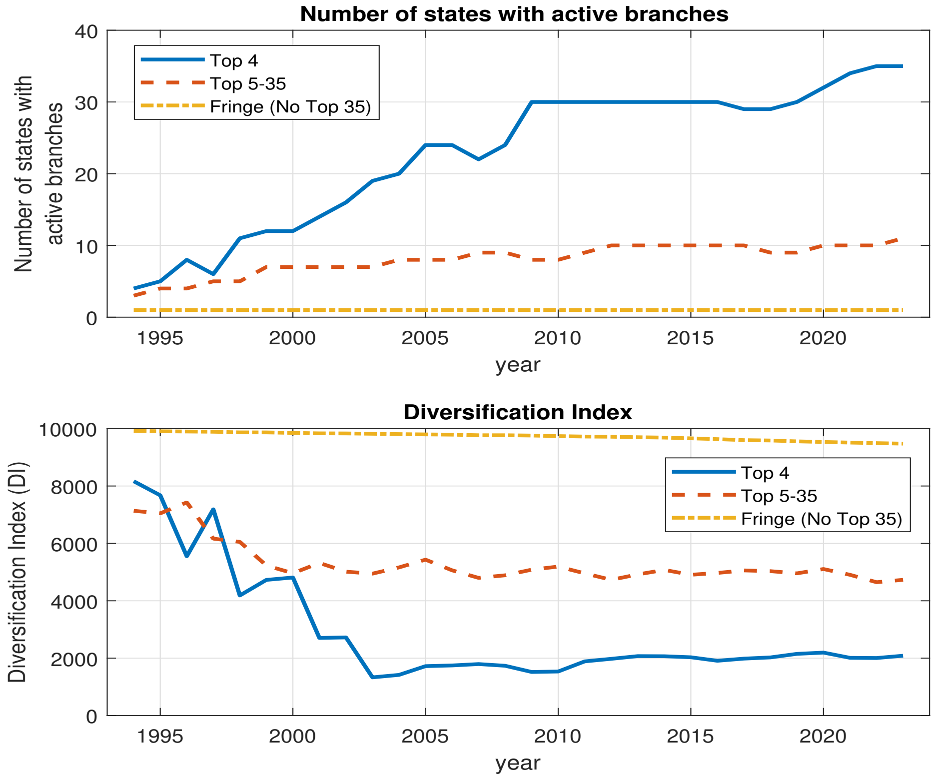
Note: We use the period that extends from 1984 to 2019. Bank Size defined using the asset distribution. Fringe includes all banks not in the Top 35 of the asset distribution. Net securities include cash plus federal funds sold plus securities minus federal funds purchased. Loans includes all loans on the balance sheet and fixed and other assets includes fixed assets and other assets not classified into net securities or loans (except trading assets). Deposits include total deposits and other borrowings (which include home loan advances). Source: Reports of Condition and Income (Call Reports).

Next we examine the geographic expansion that followed the Riegle-Neal Act in 1994 in Figure 6. We provide two measures of geographic expansion. First, the top panel graphs the number of states that a bank in each size category has an active branch for each of our 3 size categories: Top 4, Top 5-35, and the rest. Second, the bottom panel plots a diversification index. Let $d_{i,m,t}$ denote the amount of deposits received by bank i in market m in period t. Here we take m to be a state. The share of deposits of bank i in state m in period t is $s_{i,m,t} = \frac{d_{i,m,t}}{\sum_{m \in M_{i,t}} d_{i,m,t}} \times 100$ where $D_{i,t} = \sum_{m \in M_{i,t}} d_{i,m,t}$ is the total amount of deposits collected by bank i in period t and $M_{i,t}$ denotes the states in which lender i operates. We define a diversification index as follows

$$DI_{i,t} = \sum_{m \in M_{i,t}} s_{im,t}^2. \tag{1}$$

This index ranges between 0 and 10,000 and a smaller value indicates a more diversified lender. The bottom panel in Figure 6 shows the (deposit-weighted) average of this diversification index for our size categories. It is clear from both measures that there is a positive relationship between size and geographic diversification and that most of the expansion happened between the relaxation of branching restrictions and the early 2000's. Importantly, the bigger the bank, the more geographically diversified it is.

Figure 6: Deposit Space and Size over time



Note: Banks are ranked according to deposits. Source: Summary of Deposits (SOD)

In our model, the amount of depositors a bank is matched with is stochastic. Motivated by our findings of geographic diversification of big banks, we use information on deposits from our panel of commercial banks in the U.S. to estimate this process focusing on the variance of inflows by bank size. In particular, after controlling for firm and year fixed effects as well as a time trend, we estimate the following autoregressive process for log-deposits for bank i of type $\theta \in \{n, r, f\}$ (where n represents National banks (Top 4 in the asset distribution), r regional banks (Top 5-35 in the asset distribution), and f fringe (all banks not in the Top 35)) in period t:

$$\log(d^i_{\theta,t}) = (1-\rho^d_\theta)\upsilon^0_{\theta,t} + \rho^d_\theta \log(d^i_{\theta,t-1}) + u^i_{\theta,t}, \tag{2}$$

where $\delta^i_{\theta,t}$ is the sum of deposits and other borrowings in period t for bank i, and $u^i_{\theta,t}$ is iid and distributed $N(0, \sigma^2_{\theta,u})$. Since this is a dynamic model we use the method proposed by Arellano and Bond (1991). From these estimates, we can construct the variance of deposits by bank size (i.e. $\sigma_{\theta,d} = \frac{\sigma_{\theta,u}}{(1-(\rho^d_\theta)^2)^{1/2}}$) that we present in the last column of Table 3. Thus, consistent with big banks having a geographically diversified pool of funding, we find big banks have less volatile funding inflows.[8]

[8]These findings are consistent with Liang and Rhoades (1988) and Aguirregabiria, Clark, and Wang (2016).

Table 3: Deposit Process Parameters

	ρ_θ^d	$\sigma_{\theta,u}$	$\sigma_{\theta,d}$
Top 4 Banks	0.593	0.095	0.118
Top 5-35 Banks	0.839	0.134	0.247
Fringe Banks	0.889	0.164	0.358
All Banks	0.891	0.164	0.362

Note: Top 4 and Top 5-35 refers to Top 4 and Top 5-35 banks in the asset distribution when sorted by assets, respectively. Fringe Banks refers to all banks outside the top 35. Source: Consolidated Reports of Condition and Income (Call Reports)

A key aspect of the analysis is the presence of increasing returns in the banking data we study. We estimate the marginal cost of producing a loan and the fixed cost following the empirical literature on banking (see, for example, Berger, Klapper, and Turk-Ariss (2009)) and our previous work Corbae and D'Erasmo (2021).[9] The marginal cost is derived from an estimate of marginal net expenses that is defined to be marginal non-interest expenses net of marginal non-interest income. Marginal non-interest expenses are derived from the following trans-log cost function:

$$\begin{aligned}\log(NIE_t^i) &= g_1 \log(W_t^i) + \varsigma_1 \log(\ell_{\theta,t}^i) + g_2 \log(q_t^i) + g_3 \log(W_t^i)^2 \\ &+\varsigma_2[\log(\ell_t^i)]^2 + g_4 \log(q_t^i)^2 + \varsigma_3 \log(\ell_t^i)\log(q_t^i) + \varsigma_4 \log(\ell_t^i)\log(W_t^i) \\ &+g_5 \log(q_t^i)\log(W_t^i) + \sum_{j=1,2} g_6^j t^j + g_{8,t} + g_9^i + \epsilon_t^i,\end{aligned} \tag{3}$$

where $NIE_{\theta,t}^i$ is non-interest expenses (calculated as total expenses minus the interest expense on deposits, the interest expense on federal funds purchased, and expenses on premises and fixed assets), g_9^i is a bank fixed effect, W_t^i corresponds to input prices (labor expenses over assets), ℓ_t^i corresponds to real loans (one of the two bank i's outputs), q_t^i represents safe securities (the second bank output), the t regressor refers to a time trend, and $g_{8,t}$ refers to time fixed effects. We estimate this equation by panel fixed effects with robust standard errors clustered by bank.[10] Non-interest marginal expenses are then computed as:

$$\text{Mg Non-Int Exp.} \equiv \frac{\partial NIE_t^i}{\partial \ell_t^i} = \frac{NIE_t^i}{\ell_t^i}\Big[\varsigma_1 + 2\varsigma_2 \log(\ell_t^i) + \varsigma_3 \log(q_{it}) + h_4 \log(W_t^i)\Big]. \tag{4}$$

The estimated (asset-weighted) average of marginal non-interest expenses is reported in the second column of Table 4. Marginal non-interest income (Mg Non-Int Inc.) is estimated using an equation similar to equation (3) (without input prices) where the left hand

[9] The marginal cost estimated is also used to compute our measure of Markups and the Lerner Index.

[10] We eliminate bank-year observations in which the bank organization is involved in a merger or the bank is flagged as being an entrant or a failing bank. We only use banks with three or more observations in the sample.

side corresponds to log total non-interest income. The estimated (asset-weighted) average of marginal non-interest income is reported in the first column of Table 4. Net marginal expenses (Mg Net Exp.) are computed as the difference between marginal non-interest expenses and marginal non-interest income. The estimated (asset-weighted) average of net marginal non-interest expenses is reported in the third column of Table 4. The fixed cost κ_θ is estimated as the total cost on expenses of premises and fixed assets. The estimated (asset-weighted) average fixed cost (scaled by loans) is reported in the fourth column of Table 4. The final column of Table 4 presents our estimate of average costs for big and small banks. We find a statistically significant lower average cost for across the bank size distribution. This is consistent with increasing returns as in the delegated monitoring model of Diamond (1984) and with empirical evidence on increasing returns as in, among others, Berger and Mester (1997).

Table 4: Banks' Cost Structure

	Mg Non-Int Exp.	Mg Non-Int Inc.	Mg Net Exp.	Fixed Cost	Avg.
Top 4 Banks	$5.19^{\dagger,\ddagger}$	$4.05^{\dagger,\ddagger}$	$1.14^{\dagger,\ddagger}$	$0.78^{\dagger,\ddagger}$	1.92
Top 5 - 35 Banks	$4.18^{\ddagger}$	$2.77^{\ddagger}$	$1.42^{\ddagger}$	$0.69^{\ddagger}$	2.11
Fringe Banks	3.37	1.58	1.79	0.71	2.50
All Banks	4.29	2.77	1.52	0.73	2.25

Note: Top 4 Banks and Top 5-35 Banks refers to the Top 4 banks and Top 5-35 banks when sorted by assets, respectively. Fringe Banks refers to all banks outside the top 35. † denotes statistically significant difference between the Top 4 and Top 5-35. ‡ denotes statistically significant difference between the Top 4 or Top 5-35 with Fringe. Mg Non-Int Inc. refers to marginal non-interest income, Mg Non-Int exp. to marginal non-interest expenses. Mg Net Exp. corresponds to net marginal expense and it is calculated as marginal non-interest expense minus marginal non-interest income. The fixed cost is scaled by loans. Data correspond to commercial banks in the U.S. Source: Consolidated Reports of Condition and Income

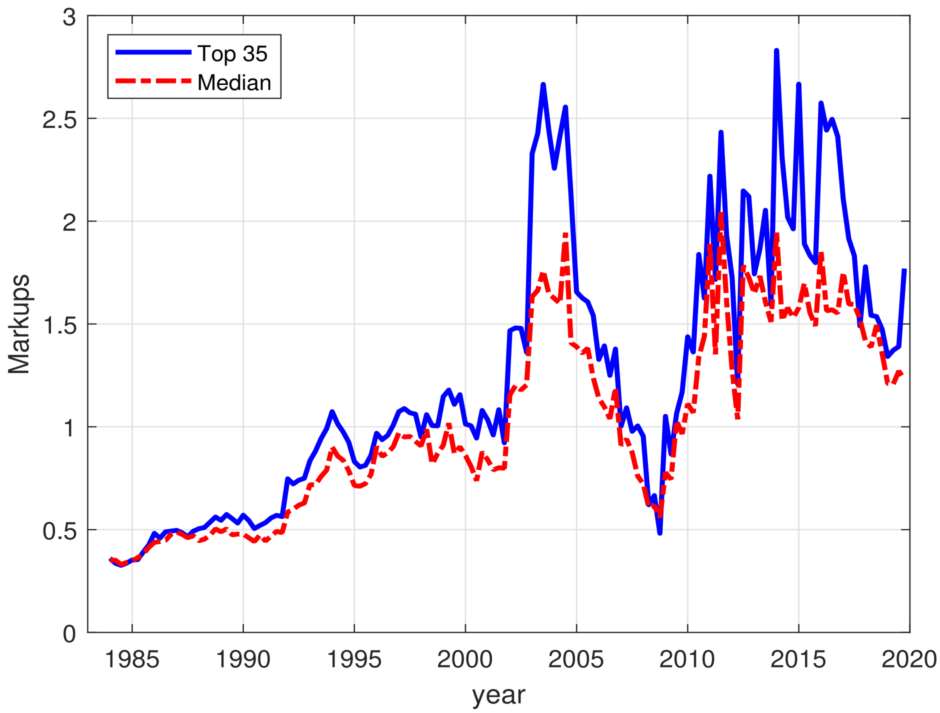
Rising bank concentration over time motivates us to consider the evolution of bank markups over time. We Use our estimates of marginal costs to compute a measure of markups across banks. The markup m_{it} for bank i in quarter t is defined as

$$m_{it} = \frac{\text{price}_{it}}{\text{marginal-cost}_{it}} - 1$$

where price_{it} denotes a measure of loan prices and $\text{marginal-cost}_{it}$ a measure the loan marginal cost. We estimate price_{it} as the ratio of interest income from loans to total loans and $\text{marginal-cost}_{it}$ as the ratio of interest expenses from deposits over deposits (r_D) plus net marginal non-interest expenses (as reported in column 3 of Table 4). Consistent with our previous paper (Corbae and D'Erasmo (2021)) and the evidence presented in De Loecker, Eeckhout, and Unger (2020) for non-financial firms, we find that average markups have been rising over time and the rise has been fueled by the upper tail of the distribution. Figure

7 presents the evolution of the median (weighted by assets) as well as the (asset-weighted) average across banks in the Top 35 and shows that markups have increased consistently during this period and that markups for large dominant banks are well above the median in most of the sample.

Figure 7: Rise in Bank Loan Markups



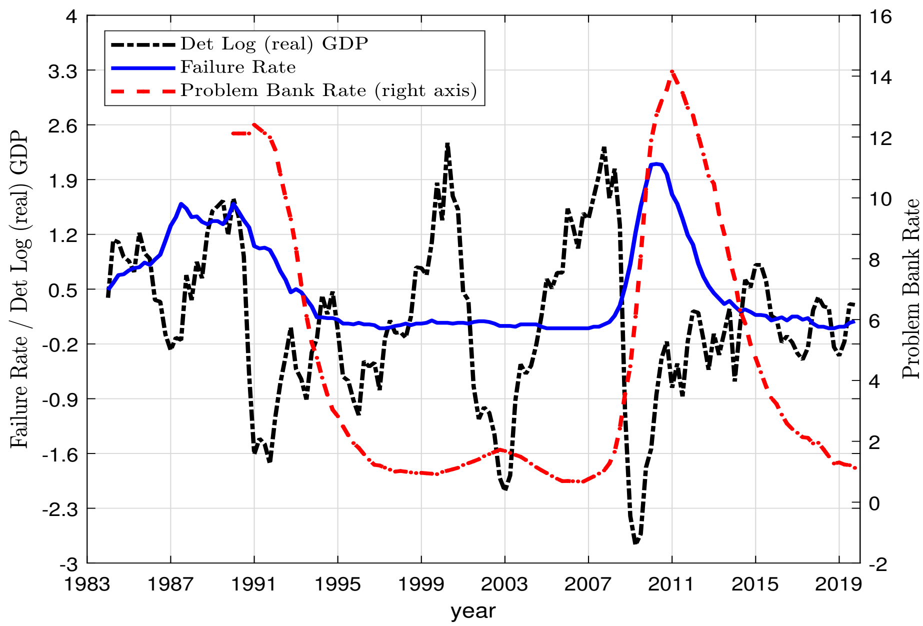
Source: We estimate markups using the estimates of marginal costs in equation (4). We compute markups as price over marginal cost where price corresponds to interest income from loans over loans plus marginal no-interest income and cost corresponds to interest expense on deposits over deposits plus marginal non-interest expense. Reports of Condition and Income (Call Reports)

As we showed in Figure 1 the total number of commercial banks has declined consistently in the last 4 decades. This trend was the consequence of regulatory changes but also industry dynamics at the business cycle frequency. In our previous work (Corbae and D'Erasmo (2021)) we showed that changes in regulations resulted in a wave of bank exits due to mergers. We focus here on the business cycle properties of bank failure. Figure 8 presents the dynamics of the banks failure rate (i.e., the ratio of the number of banks that exit due to failure to the total number of banks) and detrended log real GDP since 1984. We also include the FDIC's "Problem bank rate" which fluctuates considerably with the business cycle.[11] The correlation between the failure rate and detrended log real GDP is -0.162 while

[11]Banks on this list have a CAMELS composite rating of "4" or "5" due to financial, operational, or managerial weaknesses, or a combination of such issues (see for example https://www.fdic.gov/news/speeches/2024/fdic-quarterly-banking-profile-first-quarter-2024).

the correlation between the problem bank rate and log real GDP is -0.72 (which is available only since mid 1990).[12] That is, there is evidence that bank failure is countercyclical.

Figure 8: Bank Industry Dynamics and Business Cycles



Note: Data corresponds to commercial banks in the US. (grouped to the bank holding company level) and Det. Log (real) GDP refers to detrended log real GDP (quarterly frequency). The trend is extracted using the H-P filter with parameter 1600. The problem bank rate is derived from FDIC data at the bank level. Source: Consolidated Report of Condition and Income (call reports) and FDIC.

Table 5 documents that the bulk of entry and failures correspond to banks that are in the bottom of the distribution (i.e., outside the Top 35). The time series average accounted for by those not in the Top 35 is close to 99% across both categories. The pattern is similar when we measure the fraction of loans in each category accounted by banks of different sizes. In particular, 90% of the loans of entrants and 99% of the loans of banks that exit correspond to banks outside the Top 35 of the asset distribution.

[12] If we restrict the sample period to 1990 (Q3) to 2019 (Q4) (the same period used for the problem bank rate) the correlation between the failure rate and detrended log real GDP increases (in absolute terms) to -0.461. That is the failure rate is significanlty more countercyclical during this period.

Table 5: Entry and Exit Statistics by Bank Size (1984-2019)

Fraction of Total x accounted by:	x Entry	Failure
Top 4 Banks	0.01	0.00
Top 5 - 35 Banks	0.29	0.02
Fringe (Not Top 35)	99.70	99.98
Fraction of Loans of Banks in x accounted by:		
Top 4 Banks	0.06	0.00
Top 5 - 35 Banks	9.09	0.88
Fringe (Not Top 35)	90.36	99.12

Note: Top 4 Banks and Top 5-35 Banks refers to the Top 4 banks and Top 5-35 banks when sorted by assets, respectively. Fringe Banks refers to all banks outside the top 35. Data correspond to commercial banks in the U.S. between 1984 and 2019. Source: Consolidated Reports of Condition and Income

Table 6 describes the structure of asset returns and funding costs across the bank size distribution.[13] Consistent with the evidence on deposits, we find that the standard deviation of charge-off rates, default frequencies, loan returns, net interest margins, loan interest rates, loan interest spreads, and the net marginal cost decline with bank size. We do not find statistically significant differences across banks of different sizes in the cost of deposits, loan interest rates, or the return of (net) securities. Consistent with the evidence presented in Figure 7, we find that markups and the lerner index are increasing in bank size (and statistically significantly different from one another).

[13]Following the notation of the model we present in the next section, the charge off rate corresponds to $(1-p)\lambda$ where $(1-p)$ is the default probability and λ the recovery in default as it is estimated as the charge-off on loans net of recoveries over total loans. Our estimate of the default probability $(1-p)$ is found by taking the ratio of loans past due 90 days plus non-accrual loans over total loans. The cost of deposits (denoted as r_D in the model) is estimated as the ratio of interest expenses on deposits over total deposits. The loan return is $p \cdot r_{\mathscr{B},L} - (1-p)\lambda$ where $r_{\mathscr{B},L}$ is the bank loan interest rate and $p \cdot r_{\mathscr{B},L}$ is estimated from the ratio of interest income from loans over total loans. The loan interest rate is computed by taking the ratio of our estimate of $p \cdot r_{\mathscr{B},L}$ and our estimate of p. The net interest margin corresponds to $p \cdot r_{\mathscr{B},L} - r_D$ and the loan interest spread to $r_{\mathscr{B},L} - r_D$. The net security return is estimated as the ratio of interest income on securities plus the net cost of federal funds divided by cash plus securities plus net federal funds lending.

Table 6: Asset Returns, Liability Costs, and Markup by bank size (1984-2019)

	All Banks		Top 4		Top 5-35		Fringe	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Charge-off Rate	0.85	0.69	$1.13^{\dagger,\ddagger}$	0.30	$0.80^{\ddagger}$	0.48	0.53	0.67
Default freq.	2.14	1.35	$3.00^{\dagger,\ddagger}$	0.76	$1.96^{\ddagger}$	0.91	1.63	1.36
Cost Deposits	0.03	0.39	0.12	0.18	-0.01	0.29	0.11	0.39
Loan return	3.86	0.84	3.85	0.46	$3.54^{\ddagger}$	0.62	4.39	0.84
Net Interest Margin	4.65	0.62	$4.72^{\dagger,\ddagger}$	0.44	$4.31^{\ddagger}$	0.51	4.80	0.62
Loan Interest Rate	4.85	0.65	5.12	0.46	4.46	0.55	5.04	0.63
Loan Interest Spread	4.82	0.65	$4.98^{\dagger}$	0.46	$4.46^{\ddagger}$	0.52	4.94	0.65
Net securities return	1.80	13.48	1.31	12.86	1.43	6.42	2.78	13.38
Markup	110.44	51.72	$132.10^{\dagger,\ddagger}$	41.68	$114.62^{\ddagger}$	42.26	91.31	31.80
Lerner	47.73	8.91	$52.54^{\dagger,\ddagger}$	5.80	$48.34^{\ddagger}$	8.32	43.65	7.24
Net Mg Cost	1.52	0.50	$1.14^{\dagger,\ddagger}$	0.37	$1.42^{\ddagger}$	0.44	1.79	0.46

Note: Loan and net securities interest income (and derived variables) as well interest expenses on deposits (and derived variables) are deflated using the CPI. All estimates correspond to annualized values (computed using quarterly data). Top 4 Banks and Top 5-35 Banks refers to the Top 4 banks and Top 5-35 banks when sorted by assets, respectively. Fringe Banks refers to all banks outside the top 35. † denotes statistically significant difference (at 1% level) between the average for Top 4 banks and the average of Top 5-35 banks. ‡ denotes statistically significant difference (at 1% level) between the average of Top 4 banks and fringe banks as well as the average of Top 5-35 banks and fringe banks. Standard deviations computed after removing time and bank fixed effects. Data correspond to commercial banks in the U.S. between 1984 and 2019. Source: Consolidated Reports of Condition and Income

The dynamics of entry and exit discussed above as well as differences in intensive margin responses lead to interesting business cycle dynamics. Table 7 presents the relationship between key asset returns, funding costs, and markups with real gdp. We estimate this relationship via a regression of the detrended log real variable and detrendedn log real GDP. We find that the charge-off rate, the default frequency, the net interest margin, the loan interest rate, the loan interest spreads, and markups are countercyclical. These cyclical properties hold also across banks of different sizes. We do not find evidence of statistically significant cyclical movements in the cost of deposits, loan returns, or net securities return. In the case of loan returns, the result derives from the countercyclical loan interest rate together with a strongly countercyclical charge-off rate and default frequency.

Table 7: Business Cycle Statistics by Bank Size (1984-2019)

	All Banks	Top 4	Top 5-35	Fringe
Charge-off Rate	-0.223***	-0.318***	-0.191***	-0.127***
Default freq.	-0.455***	-0.580***	-0.377***	-0.308 ***
Cost Deposits	-0.098	-0.037	-0.108	-0.157
Loan return	0.030	0.045	0.062	-0.067
Net Interest Margin	-0.096***	-0.121***	-0.059**	-0.047*
Loan Interest Rate	-0.221**	-0.256**	-0.169*	-0.215**
Loan Interest Spread	-0.123***	-0.160***	-0.076***	-0.065***
Net securities return	-1.528	-1.678	-1.223	-1.461
Markup	-2.359*	-5.065*	-2.341*	-2.113***
Lerner	-0.663	-1.222	-0.664	-0.672*
Net Marginal Cost	-0.020	0.032	-0.036	-0.048***

Note: Values presented in this table are derived from a linear regression (at quarterly frequency) of the detrended log variable on detrended log real GDP. *** denotes significant at 1%, ** denotes significant at 5%, * denotes significant at 10%. Loan and net securities interest income (and derived variables) as well interest expenses on deposits (and derived variables) are deflated using the CPI. Top 4 Banks and Top 5-35 Banks refers to the Top 4 banks and Top 5-35 banks when sorted by assets, respectively. Fringe Banks refers to all banks outside the top 35. Data correspond to commercial banks in the U.S. between 1984 and 2019. Source: Consolidated Reports of Condition and Income

In summary, we find evidence for:

1. Imperfect competition (Top 4 and 35 banking industry concentration in Figure 1, elevated Herfindahl indices in Figure 2, incomplete passthrough as measured by the Rosse-Panzer H-statistic in Figure 3, rising markups as concentration rises and higher markups for top 35 banks than fringe in Figure 7).

2. Large right tail deviations from Zipf's law in Figures 4 and 5 and Table 1.

3. Geographic diversification by the top 4 and 35 banks in Figure 6.

4. Diversification of idiosyncratic deposit shocks resulting in lower variance of funding inflows in Table 3.

5. Increasing returns to scale as in Table 4.

6. Countercyclical failure rate in Figure 8 as well as the predominance of bank entry and exit by small banks in Table 5.

7. Consistent with a diversification argument for big banks, the standard deviation of charge-off rates, default frequencies, loan returns, net interest margins, loan interest rates, loan interest spreads, and the net marginal cost decline with bank size in Table 6

8. We do not find statistically significant differences across banks of different sizes in the cost of deposits, loan interest rates, or the return of (net) securities but do find statistically significant differences in markups and the Lerner in bank size in Table 6.

9. Countercylical chargeoffs, net interest margins, and markups in Table 7.

3 Model Environment

Our empirical findings in Section 2 motivate us to consider a Diamond (1984) delegated monitoring model (capturing increasing returns and geographic diversification) with Cournot competition among banks along the lines of the Gowrisankaran and Holmes (2004) dominant-fringe framework (motivated by high concentration of the top 35 banks versus the other 4000 fringe banks). To simplify the analysis, we organize bank balance sheets along the lines of Table 2 with size dependent stochastic insured deposits on the liabilities side versus an asset portfolio of risky loans and safe net assets (securities plus cash - short term borrowing).

Time is discrete and there is an infinite horizon. There are two regions $j \in \{e, w\}$, for instance "east" and "west". Each period and in each region, there is a mass B of ex-ante identical borrowers who have a profitable project which needs to be funded. There is also a mass $H > B$ of identical savers in each region that deposit their savings in banks and inject equity finance to banks and non-banks $k \in \{\mathcal{B}, \mathcal{N}\}$ where $\mathcal{B}$ denotes traditional banks and $\mathcal{N}$ non-banks. Financial intermediaries (banks and non-banks) intermediate between potential borrowers and savers. Except for regional specific productivity shocks, we treat regions symmetrically.

To keep notation manageable, we let any beginning of period t variable be denoted x and any end of period variable be denoted x'.

3.1 Borrowers

Ex-ante identical borrowers in region j demand one period loans in order to fund a large risky project. We think of the risky project as being either a commercial loan which may fail or a household loan like a mortgage which may be foreclosed.[14] The project requires one unit of investment (i.e. a loan either from a bank $k = \mathcal{B}$ or non-bank $k = \mathcal{N}$) at the beginning of period t. The borrower chooses the scale $R_{k,j}$ of the risky project in which she is investing those funds, which can be indexed on the lender type. The project returns $R_{k,j}$ at the end of the period according to:

$$\begin{cases} 1 + z'_j Z' R_{k,j} & \text{with prob } p_j(R_{k,j}, z'_j, Z') \\ 1 - \lambda & \text{with prob } [1 - p_j(R_{k,j}, z'_j, Z')] \end{cases} \tag{5}$$

in the successful and unsuccessful states respectively. That is, borrower gross returns are by $1 + z'_j Z' R_{k,j}$ in the successful state and by $1 - \lambda$ in the unsuccessful state where z'_j is a regional specific shock, Z' is an aggregate shock, and λ is the fraction lost in default. The

[14]Long term mortgages can be thought of as a sequence of one period loans as in Jeske, Krueger, and Mitman (2013).

aggregate technology shock Z' follows a persistent process with transition matrix $F_Z(Z', Z)$. The regional shocks z'_j are assumed to be independent over time and drawn from a bivariate normal distribution $F_z(\mu_z, \sigma_z, \rho_z)$ where μ_z denotes the mean, σ_z the standard deviation, and ρ_z covariance between regions. The success of a borrower's project, which occurs with probability $p_j(R_{k,j}, z'_j, Z')$, is independent across borrowers and time *conditional* on the borrower's endogenous choice of technology $R_k \geq 0$, the exogenous regional shock z'_j, and the exogenous aggregate shock Z'.

As for the likelihood of success or failure, a borrower who chooses to run a project with a higher return $R_{k,j}$ has more risk of failure. Specifically, $p_j(R_{k,j}, z'_j, Z')$ is assumed to be decreasing in $R_{,j}$ and increasing in z'_j, and Z'. Thus, the technology exhibits a risk-return trade-off. Further, since $R_{k,j}$ is a choice variable, project returns and failure rates are endogenously determined. While borrowers are ex-ante identical, they are ex-post heterogeneous owing to the realizations of the shocks to the return on their project.

The borrower makes a discrete choice over which type of financial institution to borrow from $k \in \{\mathscr{B}, \mathscr{N}\}$. Bank and non-bank interest rates on their loans to the borrower can differ. Taking the vector of interest rates $\mathbf{r_j} = \{r_{\mathscr{B},j}, r_{\mathscr{N},j}\}$ on loans as given, borrowers decide whether they want to fund a project given their outside option and then make a discrete choice over whether to borrow from a bank or non-bank in their region. Once with a lender type k offering a loan at interest rate $r_{k,j}$, the borrower chooses the risk-return tradeoff of their project $R_{k,j}$. This explains why we allow project choice to depend on k; since the borrower potentially faces different rates from different lenders, they may make different risk-return project choices. Following Buchak et al. (2018), we assume that the value associated with financing the project with each type of lender in any region j is subject to an unobservable idiosyncratic shock $\boldsymbol{\epsilon} = \{\epsilon_{\mathscr{B}}, \epsilon_{\mathscr{N}}\}$ affecting the value of taking a loan from each type of lender additively. We assume that ϵ_k are iid shocks drawn from a type one extreme value distribution $F_\epsilon(\epsilon; \alpha)$ with scale parameter $1/\alpha$.

Borrowers have an outside option. At the beginning of period t, they receive a realization of their reservation utility of consumption $\omega \in [0, \overline{\omega}]$ if they decide not to run the project. These draws from distribution function $\Omega(\omega)$ are i.i.d. over time and across regions. The outside option choice $\iota \in \{0, 1\}$ leads to a downward sloping aggregate demand for loans while conditional on choosing to borrow $\iota = 1$ the extreme value shocks determine loan demand across financial institution types.

There is limited liability on the part of the borrower at the project level so that the project return net of interest payments is bounded below at zero. If $r_{k,j}$ is the interest rate on a loan that the borrower faces, the borrower receives $\max\{z'_j Z' R_{k,j} - r_{k,j}, 0\}$ in the successful state and 0 in the failure state. Specifically, in the unsuccessful state she receives $1 - \lambda$ which must be relinquished to the lender. Table 8 summarizes the risk-return tradeoff that the borrower faces. Since the choice of $R_{k,j}$ is endogenous, changes in borrowing costs $r_{k,j}$ can affect the default frequencies on loans through a risk shifting motive. This also captures possible negative selection effects as loan interest rates rise.

Table 8: Borrower's Problem (conditional on investing)

Borrower Chooses $R_{k,j}$	Receive	Pay	Probability
			$\quad - \quad + \quad +$
Success	$(1 + z_j' Z' R_{k,j})$	$(1 + r_{k,j})$	$p(\ R_{k,j},\ z_j'\ Z')$
Failure	$(1 - \lambda)$	$\min\{(1-\lambda), (1 + r_{k,j})\}$	$1 - p(\ R_{k,j},\ z_j'\ Z')$

Both $R_{k,j}$ and ω are private information to the borrower. As in Bernanke and Gertler (1989), success or failure is also private information to the borrower unless the loan is monitored by the lender.[15] With one period loans, since reporting failure (and hence repayment of $1 - \lambda < 1 + r_{k,j}$) is a dominant strategy in the absence of monitoring, loans must be monitored. Monitoring is costly as in Diamond (1984).

3.2 Savers

In each region j, infinitely lived, risk neutral households with discount factor β are endowed with one unit of the good each period. We assume households are sufficiently patient such that they choose to exercise their savings opportunities. In particular, households have access to an exogenous risk-free storage technology yielding $1 + \bar{r}$ between any two periods with $\bar{r} \geq 0$ and $\beta(1 + \bar{r}) = 1$. They can also choose to supply their endowment to a bank, a non-bank, or to an individual borrower. We assume that after observing the deposit interest rate $r_{D,j}$, households who choose to deposit their earnings are randomly matched with a bank in their region at the beginning of any period t. Given deposit insurance, even if the bank fails, they receive their deposit with interest at the end of the period. Households can hold a portfolio of bank stocks yielding dividends (claims to bank cash flows) and can inject equity to banks. They can also invest in shares of the representative non-bank, which gives a claim to non-bank cash-flows. They pay lump-sum taxes/transfers τ' at the end of any period t which include a lump-sum tax τ_D' used to cover deposit insurance for failing banks. Finally, if a household wants to match directly with a borrower (e.g. directly fund an entrepreneur's project), it must compete with bank loans. Hence, the household could not expect to receive more than the bank lending rate $r_{k,j}$ in successful states and must pay a monitoring cost. Since households can purchase claims to bank cash flows, and banks can more efficiently minimize costly monitoring along the lines of Diamond (1984) and Williamson (1986), there is no benefit to matching directly with borrowers.

[15] While one interpretation of our borrowers is that they are effectively one period lived (born at the beginning of the period and dead at the end as in the OG model of Bernanke and Gertler (1989)), we could have effectively modeled borrowers as long lived and added enough inter-period anonymity so that financial contracts are one period lived as in Carlstrom and Fuerst (1997) and borrowers are sufficiently impatient to not want to augment net worth. We follow this approach in our previous paper Corbae and D'Erasmo (2021).

3.3 Banks

As in Diamond (1984), banks exist in our environment to pool risk and economize on monitoring costs. We assume there are three classes of banks $\theta \in \Theta = \{\mathpzc{n}, \mathpzc{r}, \mathpzc{f}\}$ for national, regional, and fringe respectively, with size ranking $\mathpzc{f} < \mathpzc{r} < \mathpzc{n}$. We assume a dominant-fringe market structure as in Gowrisankaran and Holmes (2004). Specifically, national and regional banks Cournot compete in a homogeneous loan market as in Ericson and Pakes (1995) by strategically choosing the quantity of loans they supply before fringe banks choose their loan supply taking prices as given. National banks are geographically diversified in the sense that they extend loans and receive deposits in both $\{e, w\}$ regions. Regional and fringe banks are restricted to make loans and receive deposits in 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. This assumption can, in principle, generate ex-post differences in loan returns documented in the data section. The model allows for endogenous entry and exit in response to bank and regional specific shocks.

We denote loans made by bank i of type θ in region j in period t by $\ell^i_{\theta,j}$. As in Corbae and D'Erasmo (2021), bank type θ determines the mean and variance of a bank's deposits $d^i_\theta \in D_\theta$. In particular, banks in the model face the deposit process we estimated in equation (2). To make our definition of type consistent with the data presented in Table 3, the mean of the deposit process satisfies $\overline{d}_{\mathpzc{n}} > \overline{d}_{\mathpzc{r}} > \overline{d}_{\mathpzc{f}}$ so that higher types have a bigger funding base. Furthermore, also consistent with the data presented in Table 3, the variance of deposits satisfy $\sigma_{\mathpzc{n}} \leq \sigma_{\mathpzc{r}} \leq \sigma_{\mathpzc{f}}$ so that bigger banks have lower variance consistent with diversification. We discretize the continuous deposit process d^i_θ in equation (2) into a finite support and denote its transition matrix by $G_\theta(d'_\theta, d_\theta)$. Deposits are collected at the regional level but we assume that national banks ($\mathpzc{n}$) can move deposits freely across regions. Since we do not take a stand on what a "region" is both in the model and in the estimation in equation (3) and regions are symmetric with respect to state variables, to simplify on notation we abstract from denoting the regional origin of d^i_θ.

A given incumbent bank i is randomly matched with a set of potential household depositors d^i_θ who receive deposit interest rate $r_{D,j}$ and then decide how many loans to extend. Bank i can fund an amount of loans larger than its deposits by using external borrowing $a^i_\theta < 0$ at rate $r^-_A > \overline{r}$. If a bank chooses an amount of loans lower than its capacity constraint, the leftover deposits $a^i_\theta > 0$ can be invested at rate $r^-_A > r^+_A > \overline{r}$. The flow balance sheet constraint for regional or fringe banks is $\ell^i_{\theta,j} + a^i_\theta = d^i_\theta$. National banks are geographically diversified in the sense that they extend loans and receive deposits in both regions ($\sum_j \ell^i_{\mathpzc{n},j} + a^i_{\mathpzc{n}} = d^i_{\mathpzc{n}}$). Note that since the outside option for a household matched with a bank is to store at rate $\overline{r}$, we know that $r_D \geq \overline{r}$.

In Table 4 of 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 pay net non-interest expenses $c_\theta(\ell_{\theta,j})$ (i.e., the difference between non-interest expenses and non-interest income on loans) per unit of loans extended. We assume $c_\theta(\cdot)$ is a quadratic function with linear term c^1_θ and quadratic term c^2_θ. To capture heterogeneity among fringe banks, we assume $c^1_{\mathpzc{f}}$ are drawn from a distribution with cdf $\Xi(c^1_{\mathpzc{f}})$. The linear terms for banks of type $\theta \in \{\mathpzc{n}, \mathpzc{r}\}$ are identical across all banks of the given type. Net non-interest expenses are constant over the lifespan of each bank. 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.

The timing in the loan stage follows the standard treatment of a dominant firm model as in Gowrisankaran and Holmes (2004). The dominant firms, our national and regional banks, move first. They compete in a Cournot fashion and choose quantities $\ell^i_{\theta,j}$ 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 bank loan interest rate $r_{\mathcal{B},j}$) in region j and simultaneously decide on the amount of loans to extend.[16] We assume that $H > B$ so there are sufficient funds to cover all possible loans.

End-of-period loan revenues associated with beginning-of-period lending $\ell^i_{\theta,j}$ in region j for an incumbent bank i of type θ in industry state $\boldsymbol{\mu}$ (to be described below) depends on its individual state s_θ with $s_n = d_n$, $s_r = (j, d_r)$, and $s_\ell = (j, d_\ell, c^1_\ell)$, and its end-of-period aggregate state $S' = (\boldsymbol{\mu}, Z, z'_j, z'_{-j}, Z')$

$$\pi^\ell_{\theta,j}(s_\theta, S') \;=\; \left[p_j(R_{\mathcal{B},j}, z'_j, Z')r_{\mathcal{B},j}(\boldsymbol{\mu}, Z) - (1 - p_j(R_{\mathcal{B},j}, z'_j, Z'))\lambda\right] \ell^i_{\theta,j} - c_\theta(\ell^i_{\theta,j}), \quad (6)$$

where $p_j(\cdot)$ denotes the fraction of bank loans that are repaid at the end of the period in region j (an endogenous object that is consistent with the borrower's problem), $r_{\mathcal{B},j}(\cdot)$ is the Cournot equilibrium interest rate on bank loans in region j, $c_\theta(\ell^i_{\theta,j})$ is the marginal cost of extending $\ell^i_{\theta,j}$ loans. Then, end of period profits associated with beginning-of-period deposits d^i_θ for an incumbent bank i of type $\theta \in \{r, \ell\}$ equals

$$\pi_\theta(s_\theta, S') \;=\; \pi^\ell_{\theta,j}(s_\theta, S') + [r^+_A \mathbf{1}_{\{a^i_\theta \geq 0\}} + r^-_A \mathbf{1}_{\{a^i_\theta < 0\}}]a^i_\theta - r_D d^i_\theta - \kappa_\theta \quad (7)$$

where κ_θ is the fixed operating cost in Table 4. Profits for an incumbent bank of type $\theta = n$ can be defined similarly by summing over loan revenues across regions (i.e. $\sum_j \pi^\ell_{n,j}(s_n, S')$).

There is limited liability on the part of banks. As in Cooley and Quadrini (2001) and Hennessy and Whited (2007), 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. For tractability, unlike in Corbae and D'Erasmo (2021), we assume that banks cannot retain earnings. Dividends net of equity injections for bank i of type $\theta \in \{n, r, \ell\}$ in state (s_θ, S') are given by

$$\mathcal{D}_\theta(s_\theta, S') = \pi_\theta(s_\theta, S') - \mathbf{1}_{\{\pi_\theta(s_\theta, S') < 0\}}\xi_\theta(|\pi_\theta(s_\theta, S')|). \quad (8)$$

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 to incorporate an extra state variable. A bank that has negative expected continuation value can exit, in which case it receives value zero. We denote the exit decision $x_\theta(s^i_\theta, S')$. If a national bank exits, it must

[16] Since projects are indivisible and borrower outside option ω is unobservable, there is not heterogeneity in rates $r_{\mathcal{B},j}$ in any given region j.

exit both regions. The objective function of the bank is to maximize the expected present discounted value of future dividends net of equity injections with discount factor β.

Entry costs for the creation of national and regional banks are denoted by $\Upsilon_{n} \geq \Upsilon_{r} \geq \Upsilon_{f} \geq 0$.[17] Entry costs correspond to the initial injection of equity into the bank by its owners. Every period a countable mass M_θ of potential banks make the decision to enter the market or not. We assume that each entrant satisfies a zero expected discounted profits condition net of this entry cost. We denote the entry decision by e_θ. To simplify the analysis, we assume that entering fringe banks' draw of c_{f}^{1} 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.[18] The entry decision is taken after the realization of the aggregate shock (and c_{f}^{1} in the case of fringe banks) but before learning their deposit funding base d_θ. Initial deposits are drawn from $G^*_\theta(d_\theta)$ the stationary distribution associated with $G_\theta(d'_\theta, d_\theta)$.

We will assume that all banks in a given state $(\theta, d_\theta) \in \{\{n, r\} \times D_\theta\}$ and $(f, d_{f}, c_{f}^{1}) \in \{D_{f} \times C_{f}^{1}\}$ are treated symmetrically, so the cross-sectional distribution $\boldsymbol{\mu}$ in any period t specifies the number (or countable measure) of banks across states and regions. More specifically, the cross-sectional distribution $\boldsymbol{\mu}$ is given by $\{\mu_\theta(s_\theta)\}$ where each element of $\boldsymbol{\mu}$ is a measure μ_θ corresponding to *active* banks of type θ in state s_θ. Further, the law of motion for the industry state is denoted

$$\boldsymbol{\mu}' = \mathcal{T}^*(\boldsymbol{\mu}, M^e) \tag{9}$$

where $M^e = \{M^e_\theta\}$ denotes the set with the mass of entrants of each bank type θ and the transition function $\mathcal{T}^*$ is defined explicitly below in equation (32).

3.4 Non-Bank Lenders

A representative national non-bank that discounts the future at rate β specializes in extending loans to entrepreneurs (in both regions) in a perfectly competitive market. To keep the analysis simple, the non-bank is financed with equity $e_{\mathcal{N}}$ raised from the household sector and is not subject to limited liability. When lending to entrepreneurs non-banks face a marginal monitoring cost $c_{\mathcal{N}}$. Like banks, the representative non-bank can diversify entrepreneurs' idiosyncratic risk but it is subject to regional and aggregate fluctuations.

Let $\pi_{\mathcal{N}}(s')$ denote the profits of the non-bank after the realization of next period's aggregate and regional shocks associated with its current lending $\ell_{\mathcal{N},j}$ given by

$$\pi_{\mathcal{N}}(s') = \sum_j \left\{ \left[p_j(R_{\mathcal{N}}, z'_j, Z') r_{\mathcal{N},j}(\boldsymbol{\mu}, Z) - (1 - p_j(R_{\mathcal{N}}, z'_j, Z'))\lambda\right] - c_{\mathcal{N}} \right\} \ell_{\mathcal{N},j}. \tag{10}$$

[17] As in Pakes and McGuire (1992) we will assume that these costs become infinite after a certain number of firms of the given type are in the market.

[18] In any given period, there are M_{f} 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 banks 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, Saure, and Weintraub (2012) as we do in Corbae and D'Erasmo (2021).

subject to flow constraint $S_{\mathcal{N}} \cdot e_{\mathcal{N}} = \sum_j \ell_{\mathcal{N},j}$ where $S_{\mathcal{N}}$ are household shareholdings of the non-bank. Since the non-bank operates in a perfectly competitive market it takes the regional interest rate $r_{\mathcal{N},j}$ as given. The non-bank issues dividends according to $\mathcal{D}_{\mathcal{N}} = \pi_{\mathcal{N}}(s')$.

The objective function of the non-bank is to maximize the expected present discounted value of future cash-flows to households with discount factor β. We assume that there is free entry into the non-bank sector, and to simplify the analysis we set the entry cost to zero.

3.5 Information

There is asymmetric information on the part of borrowers and lenders (banks and households). Only borrowers know their outside option (ω) and success or failure of their project is only observable after incurring the monitoring cost in $c_\theta(\ell^i_{\theta,j})$. Other information is observable.

3.6 Government Budget Constraint

The government collects lump-sum taxes to cover the cost of deposit insurance and unpaid net securities. Treating banks of type $\theta \in \{\mathscr{r}, \mathscr{f}\}$ in a given state (s_θ, S') symmetrically, post-liquidation cost to the deposit insurance fund from those banks in region j are given by

$$\begin{aligned}\Delta'_\theta(s_\theta, S') &= (1+r_D)d_\theta + (1+r_A^-)\max\{(\ell_{\theta,j} - d_\theta), 0\} - \zeta_\theta[p \cdot (1 + r_{\mathcal{B},j}(Z,\mu)) + (1-p)(1-\lambda)]\ell_{\theta,j} - c_\theta(\\ &- \zeta_\theta(1+r_A^+)\max\{(d_\theta - \ell_{\theta,j}), 0\} - \kappa_\theta\end{aligned}$$

where $\zeta_\theta \leq 1$ is the post-liquidation value of the bank's asset portfolio. The post liquidation cost for a national bank can be defined similar to equation (11) by summing over the recovery of the loan portfolio in both regions. Then, aggregate taxes are given by

$$\tau'_D(S') \cdot H = \int_{\theta, s_\theta} x_\theta(s_\theta, S') \max\{0, \Delta'_\theta(s_\theta, S')\} d\mu_\theta(s_\theta) \tag{12}$$

3.7 Timing

In any period, the timing of events is as follows:

1. At the beginning of the period, given Z,
   (a) The mass of depositors that the bank is matched with d_θ is realized given household asset decisions. That determines the industry state (i.e. cross-sectional distribution $\boldsymbol{\mu}$).
   (b) After observing ω, borrowers choose whether to invest in the risky technology $\iota = 1$ or to choose their outside option $\iota = 0$.
   (c) If $\iota = 1$, those borrowers choose the type of lender $k \in \{\mathcal{B}, \mathcal{N}\}$ given a draw $\boldsymbol{\epsilon}$ from an extreme value distribution. The borrower also chooses $R_{k,j}$ (i.e. the risk-return tradeoff of their project).

(d) Banks in individual state s_θ choose how many deposits to accept and their portfolio of loans and net securities. National and regional banks extend loans before fringe banks (i.e. as a Stackelberg game). Non-banks receive their equity injections from households and make loans.

(e) The loan market is cleared determining $\mathbf{r}_j = \{r_{\mathcal{B},j}, r_{\mathcal{N},j}\}$

2. At the end of the period, Z' and z'_j are realized

(a) Project returns for borrowers are determined according to $p_j(R_{k,j}, z'_j, Z')$. This determines the portfolio of performing and non-performing loans resulting in a realization of $\pi_\theta(s_\theta, S')$ and $\pi_{\mathcal{N}}(S')$.

(b) Bank exit x_θ and entry e_θ choices are made.

(c) Dividend payments and equity injections are undertaken by banks and nonbanks.

(d) Households pay taxes $\tau'(S')$ to fund deposit insurance, bank stock choices, equity injections for banks and nonbanks, and consume.

4 Equilibrium

As described above, to save on notation, we assume all banks in a given individual state s_θ are treated symmetrically (and hence drop the notation i.

4.1 Borrower Problem

Every period, given $\mathbf{r}_j = \{r_{\mathcal{B},j}, r_{\mathcal{N},j}\}$, Z, and ω, borrowers located in region j choose whether ($\iota = 1$) or not ($\iota = 0$) to undertake their project. Conditional on choosing $\iota = 1$, entrepreneurs observe $\boldsymbol{\epsilon} = \{\epsilon_{\mathcal{B}}, \epsilon_{\mathcal{N}}\}$ and then choose which type of lender $k \in \{\mathcal{B}, \mathcal{N}\}$ to borrow from and the scale of the technology to operate $R_{k,j}$ to solve

$$\max_{\{\iota\}}(1-\iota)\cdot\omega + \iota \cdot E_{\boldsymbol{\epsilon}}[\Pi_E(Z, \boldsymbol{r_j}, \boldsymbol{\epsilon})] \tag{13}$$

where the value of investing (conditional on $\boldsymbol{\epsilon}$) $\Pi_E(\boldsymbol{r_j}, \boldsymbol{\epsilon}, z'_j, Z')$ is

$$\begin{aligned}\Pi_E(Z, \boldsymbol{r_j}, \boldsymbol{\epsilon}) = \max_{\{k, R_{k,j}\}} \Big\{ & \mathbf{1}_{\{k=\mathcal{B}\}} E_{z'_j, Z'|Z}\left[\pi_E(r_{\mathcal{B},j}, R_{\mathcal{B},j}, z'_j, Z') + \epsilon_{\mathcal{B}}\right] \\ & + \mathbf{1}_{\{k=\mathcal{N}\}} E_{z'_j, Z'|Z}\left[\pi_E(r_{\mathcal{N},j}, R_{\mathcal{N},j}, z'_j, Z') + \epsilon_{\mathcal{N}}\right] \Big\},\end{aligned} \tag{14}$$

where $\mathbf{1}_{\{\cdot\}}$ is an indicator function that takes the value one if the argument $\{\cdot\}$ is true and zero otherwise, and

$$\pi_E(r_{k,j}, R_{k,j}, z'_j, Z') = \begin{cases} \max\{0, z'_j Z' R_{k,j} - r_{k,j}\} & \text{with prob } p_j(R_{k,j}, z'_j, Z') \\ \max\{0, -(\lambda + r_{k,j})\} & \text{with prob } 1 - p_j(R_{k,j}, z'_j, Z') \end{cases}$$

The solution to (14) implies that the share of borrowers choosing a loan from a lender of type k in region j is

$$\psi_{k,j}(Z,\mathbf{r}_j)=\frac{\exp\left(\alpha E_{z'_j,Z'|Z}\left[\pi_E(r_{k,j},R_{k,j},z'_j,Z')\right]\right)}{\sum_{\hat{k}\in\{\mathcal{B},\mathcal{N}\}}\exp\left(\alpha E_{z'_j,Z'|Z}\left[\pi_E(r_{\hat{k},j},R_{\hat{k},j},z'_j,Z')\right]\right)}. \tag{15}$$

The expected value of taking out a loan in region j is[19]

$$V_{E,j}(Z,\mathbf{r}_j) \;=\; \int \Pi_E(Z,\boldsymbol{r_j},\boldsymbol{\epsilon})dF_\epsilon(\epsilon;\alpha). \tag{16}$$

If the borrower undertakes the project financed by lender type k, then an application of the envelope theorem implies

$$\frac{\partial E_{z'_j,Z'|Z}\left[\pi_E(r_{k,j},R_{k,j},z'_j,Z')\right]}{\partial r_{k,j}}=-E_{z'_j,Z'|Z}[p_j(R_{k,j},Z',z'_j)]<0. \tag{17}$$

Thus, participating borrowers (i.e. those who choose to run a project rather than take the outside option) are worse off the higher is the interest rate on loans. This has implications for the aggregate demand for loans determined by the participation decision (i.e. $\omega \leq V_{E,j}(Z,\mathbf{r}_j)$). In particular, the total demand for loans in region j is given by

$$L^d_j(Z,\mathbf{r}_j)=\int_0^{\overline{\omega}}\mathbf{1}_{\{\omega\leq V_{E,j}(Z,\mathbf{r}_j)\}}d\Omega(\omega). \tag{18}$$

Then loan demand for commercial banks in region j is given by

$$L^d_{\mathcal{B},j}(Z,\mathbf{r}_j)=\psi_{\mathcal{B},j}(Z,\mathbf{r}_j)L^d_j(Z,\mathbf{r}_j). \tag{19}$$

In that case, everything else equal, (17) implies $\frac{\partial L^d_{\mathcal{B},j}(\mathbf{r}_j,Z)}{\partial r_{\mathcal{B},j}}<0$. That is, the bank loan demand curve is downward sloping. Furthermore, bank market shares are decreasing in bank lending rates (i.e. $\frac{\partial s_{\mathcal{B},j}(\mathbf{r}_j,Z)}{\partial r_{\mathcal{B},j}}<0$) and aggregate loan demand decreases with an increase in bank lending rates (i.e. $\frac{\partial L^d_j(\mathbf{r}_j,Z)}{\partial r_{\mathcal{B},j}}\leq 0$).

4.2 Saver Problem

The problem of a representative household in region j is

$$V_H(A,D,\{\Phi_{s_\theta}\}_{\forall\{\theta\}},\Phi_{\mathcal{N}})=\max_{\{A',D',\{\Phi'_{s_\theta}\}_{\forall\theta},\Phi'_{\mathcal{N}}\}}E\left[C'+\beta V_H(A',D',\{\Phi'_{s_\theta}\}_{\forall\{\theta\}},\Phi'_{\mathcal{N}})\right] \tag{20}$$

[19] The expected value of taking out a loan has a convenient closed form: $V_{E,j}(Z,\mathbf{r}_j)=\frac{\gamma_E}{\alpha}+\frac{1}{\alpha}\ln\left(\sum_k\exp\left(\alpha E_{z'_j,Z'|Z}\left[\pi_E(r_{k,j},R_{k,j},z'_j,Z')\right]\right)\right)$ where γ_E is Euler's constant.

subject to

$$C' + A' + D' + \int_{\theta,s_\theta} [P_{s_\theta} + \mathbf{1}_{\{e_\theta=1\}} \cdot \Upsilon_\theta]\Phi'_{s_\theta} d\mu_\theta(s_\theta) + \Phi'_{\mathcal{N}} P_{\mathcal{N}} \tag{21}$$
$$= y + (1 + r_A^+)A + (1 + r_D)D + \int_{\theta,s_\theta} (\mathcal{D}_{s_\theta} + P_{s_\theta})\Phi_{s_\theta} d\mu_\theta(s_\theta) + P_{\mathcal{N}})\Phi_{\mathcal{N}} - \tau'_D(S'),$$

and $C' \geq 0$ where $P_{\theta,j}$ and $\Phi'_{\theta,j}$ are the post-dividend stock price and stock holding of bank of type θ in region j, respectively, and $P_{\mathcal{N}}$ and $\Phi'_{\mathcal{N}}$ are the price of a claim to non-bank dividends cum equity and stock holdings of the non-bank, respectively, given zero initial holdings. Given exit and entry decision rules, in cases in which a bank has exited, $P_{s_\theta} = 0$ on the right-hand side of the budget constraint (21), and, in cases in which a bank has entered, $P_{s_\theta} > 0$ on its left-hand side.

4.3 Bank Problem

As in Ericson and Pakes (1995), we consider symmetric equilibrium in the sense that all banks of a given type θ in the same region and individual state (d_θ, c_θ^1) are treated identically. Hence, an incumbent bank i of type θ in state (d_θ, c_θ^1) in region j chooses loans $\ell_{\theta,j}$ and net securities a_θ understanding that these decisions interact with the exit decision $x_{\theta,j}$ that is taken after the realization of the aggregate and regional shocks Z' and z'_j for $j = e, w$, respectively.[20] To save on notation, we suppress the explicit dependence of each bank function on its competitor's actions. As standard in the IO literature, we take the expectation over the end-of-period payoffs and its continuation value as each dominant bank forms expectations of the likely future states of its competitors conditional on the different possible outcomes of the aggregate and regional shocks.

In equilibrium, $r_D = \overline{r}$. Differentiating π_θ with respect to $\ell_{\theta,j}$ (for expositional purposes for cases when $\ell_{\theta,j} \leq d_\theta$)

$$\frac{d\pi_\theta}{d\ell_{\theta,j}} = \underbrace{\left[p_j r_{\mathcal{B},j} - (1-p_j)\lambda - \frac{dc_\theta(\cdot)}{d\ell_{\theta,j}} - \overline{r}\right]}_{(+)\text{ or }(-)} + \ell_{\theta,j}\Big[\underbrace{p_j}_{(+)} + \underbrace{\frac{\partial p_j}{\partial R_j}\frac{\partial R_j}{\partial r_{\mathcal{B},j}}(r_{\mathcal{B},j} + \lambda)}_{(-)}\Big] \underbrace{\frac{dr_{\mathcal{B},j}}{d\ell_{\theta,j}}}_{(-)\text{ for dominant banks}}. \tag{22}$$

The first bracket represents the marginal change in profits from extending an extra unit of loans.[21] 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 $\frac{dr_{\mathcal{B},j}}{d\ell_{\theta,j}} < 0$ while for fringe banks $\frac{dr_{\mathcal{B},j}}{d\ell_{\theta,j}} = 0$. Note that a change in interest rates also endogenously affects the fraction of delinquent loans faced by banks (the term

[20] 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 $R_{\mathcal{B}}$ 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.

[21] In cases where $\ell_{\theta,j} > d_\theta$ (i.e., $a_\theta < 0$) the marginal cost of funds $\overline{r}$ changes to r_A^- (which is the cost of external funding for the bank at the lending stage when loans exceed deposits).

$\frac{\partial p_j}{\partial R_j}\frac{\partial R_j}{\partial r_{\mathcal{B},j}} < 0$). That is, given limited liability entrepreneurs take on more risk when their financing costs rise.

The value of an incumbent bank of any type θ at the beginning of the period is given by[22]

$$V_\theta(s_\theta, Z, \boldsymbol{\mu}) = \max_{\{\ell_{\theta,j}\geq 0, a_\theta\}} \quad \beta E_{S'|Z}\left[W_\theta(s_\theta, S')\right] \tag{23}$$

subject to bank balance sheet constraints

$$\begin{array}{cc} \ell_{\theta,j} + a_\theta = d_\theta & \text{if } \theta \in \{\boldsymbol{r}, \boldsymbol{f}\}, \\ \sum_j \ell_{\theta,j} + a_\theta = d_\theta & \text{if } \theta = \boldsymbol{n} \end{array} \tag{24}$$

and market clearing in each region[23]

$$L^s_{\mathcal{B},j}(Z, \boldsymbol{\mu}) \equiv \int_{\theta, s_\theta} \ell_{\theta,j}(s_\theta, Z, \boldsymbol{\mu}) d\mu_\theta(s_\theta) - L^d_{\mathcal{B},j}(Z, \boldsymbol{r}_j) = 0, \forall j \tag{25}$$

where $L^d_{\mathcal{B},j}(Z, \boldsymbol{r}_j)$ is given in (19) and the exit decision $x_\theta(s_\theta, S')$ derives from

$$W_\theta(s_\theta, S') = \max_{\{x_\theta \in \{0,1\}\}} \left\{W^{x=0}_\theta(s_\theta, S'), W^{x=1}_\theta(s_\theta, S')\right\} \tag{26}$$

with

$$W^{x=0}_\theta(s_\theta, S') = \mathcal{D}(d_\theta, s') + E_{s'_\theta, \boldsymbol{\mu}'}[V_\theta(s'_\theta, Z', \boldsymbol{\mu}')], \tag{27}$$

$$\mathcal{D}(s_\theta, S') = \left\{ \begin{array}{cc} \pi_\theta(s_\theta, S') & \text{if } \pi_\theta(s_\theta, S') \geq 0 \\ \pi_\theta(s_\theta, S') - \xi_\theta(\pi_\theta(s_\theta, S')) & \text{if } \pi_\theta(s_\theta, S') < 0 \end{array} \right., \tag{28}$$

$$W^{x=1}_\theta(s_\theta, S') = \max\left\{0, -\Delta'(s_\theta, S') - (1 + r^-_A)\max\{\ell_\theta - d_\theta, 0\} - \kappa_\theta\right\}. \tag{29}$$

4.3.1 Bank Entry

The value of an entrant of type θ net of entry costs when the aggregate state is S' is

$$V^e_\theta(S') \;\; = \;\; -\Upsilon_\theta + E_{s'_\theta}\left[V_\theta(s'_\theta, Z', \boldsymbol{\mu}')\right]. \tag{30}$$

It should be understood that for banks of type $\theta \in \{\boldsymbol{r}, \boldsymbol{f}\}$ the entry value also depends on the region where the bank is entering. Potential entrants will decide to enter if $V^e_\theta(S') \geq 0$. We denote the entry decision by $e_\theta(S') \in \{0, 1\}$ The mass of entrants M^e_θ is determined endogenously in equilibrium. Free entry implies that

$$V^e_\theta(S') \times M^e_\theta = 0. \tag{31}$$

That is, in equilibrium, either the value of entry is zero, the mass of entrants is zero, or both.

[22]While each bank takes all other banks' strategies as given, for notational simplicity we abstract from writing that in the problem.

[23]The national bank faces the clearing condition in each region.

4.3.2 Industry Law of Motion

The distribution of banks evolves according to $\boldsymbol{\mu}' = \mathcal{T}^*(\boldsymbol{\mu}, M^e)$ where each component is given by:

$$\mu'_\theta(s'_\theta) = \int_{s_\theta} (1 - x_\theta(s_\theta, S'))G_\theta(d'_\theta, d_\theta)d\mu_\theta(s'_\theta) + M^e_\theta G^*_\theta(d'_\theta). \tag{32}$$

In the case of banks of type $\theta = \not{f}$, the second term in the right hand side of (32) incorporates the draws from the distribution of $c^1_{\not{f}}$. That is, the last term becomes $M^e_{\not{f}} G^*_{\not{f}}(d'_{\not{f}})\Xi(c^1_{\not{f}})$. Equation (32) makes clear how the law of motion for the distribution of banks is affected by entry and exit decisions.

4.4 Non-Bank Problem

The representative non-bank operates in a competitive industry, so when making lending decisions it takes the loan interest rate $r_{\mathcal{N},j}$ as given. In any state Z, and taking into account that $\beta(1 + \overline{r}) = 1$, the first order condition of the non-bank with respect to $\ell_{\mathcal{N},j}$ is given by

$$\overline{r} = E_{Z',z'_j,z'_{-j}|Z} \left[p(R_{\mathcal{N}}(r_{\mathcal{N},j}, Z), z'_j, Z') - (1 - p(R_{\mathcal{N}}(r_{\mathcal{N},j}, Z), z'_j, Z'))\lambda \right] - c_{\mathcal{N}}, \tag{33}$$

where $R_{\mathcal{N},j}(r_{\mathcal{N}}, Z)$ is the optimal choice of technology by the entrepreneur in region j when taking a loan from a non-bank facing interest rate $r_{\mathcal{N},j}$. Equation (33) is one equation in one unknown which pins down the interest rate $r_{\mathcal{N},j}$ of the non-bank sector as a function of Z. The fact that $r_{\mathcal{N},j}$ is independent of the entire distribution of banks is a form of block recursivity as in Menzio and Shi (2010). Evaluating the non-bank loan demand at this price we can determine the level of lending of the non-bank. Equation (33) also makes clear that the expected net return between a bank deposit and non-bank investment is equalized, with the spread depending on $c_{\mathcal{N}}$. However, while bank deposits guarantee a risk-free return (since there is deposit insurance), equity injections in a non-bank are subject to aggregate and regional risk.

4.5 Definition of Equilibrium

A Markov Perfect Equilibrium is:

1. $\{\iota_j, k_j, R_{k,j}\}$ are consistent with borrower optimization in (13)-(14) inducing an aggregate loan demand function $L^d_j(Z, \mathbf{r}_j)$ in (18) and bank/nonbank market shares in (15).

2. $\{A', D', \Phi'_{s_\theta}, \Phi'_{\mathcal{N}}\}$ are consistent with household optimization in (20) and the deposit matching process $G_\theta(d'_\theta, d_\theta)$.

3. $\{\ell_\theta, a_\theta, x_\theta, V_\theta\}$ are consistent with bank optimization in (23)-(26) inducing an aggregate bank loan supply function $L^s_{\mathcal{B},j}$ defined in (25).

4. e_θ is such that free entry (31) is satisfied.

5. The law of motion for the industry state $\boldsymbol{\mu'} = \mathcal{T}^*(\boldsymbol{\mu}, M^e)$ in equation (32) induces a sequence of cross-sectional distributions over time that are consistent with entry and exit.

6. Non-bank loans $\ell_{\mathcal{N},j}$ are consistent with optimization given by (33).

7. The vector of interest rate $\mathbf{r}_j(Z, \boldsymbol{\mu})$ is such that the loan market clears.

8. Bank stock prices P_{s_θ} are consistent with bank valuation V_θ, as well as nonbank stock prices $P_{\mathcal{N}}$.

9. Taxes $\tau'_D(S')$ cover the cost of deposit insurance.

5 Calibration

We calibrate the model parameters via Simulated Method of Moments to match the key statistics of the U.S. banking industry described in Section 2. Our main source for bank level variables (and aggregates derived from them) is the Consolidated Report of Condition and Income for Commercial Banks (regularly called "call reports").[24] We aggregate commercial bank level information to the Bank Holding Company level. As discussed above, moments from the Call Report data are computed beginning in 1984 (due to an overhaul of the data in that year). We also use data from the Summary of Deposits (SOD), and FRED, federal reserve economic data for aggregate economic series (GDP, CPI, etc).

A model period is set to be one year. Before moving into the details of the calibration, we provide functional forms for the stochastic process of the borrower idiosyncratic shock, the distribution of borrower's outside option, the aggregate shock, the regional shock, the distribution of net expenses for fringe banks and banks' external financing cost function.

In our model, banks of three types operate $\theta \in \{\mathcal{n}, \mathcal{r}, \mathcal{f}\}$. Consistent with the data size differences we described in the data section, we identify National banks $\theta = \mathcal{n}$ with those in the Top 4 (when banks are sorted by assets), Regional banks $\theta = \mathcal{r}$ with those in the Top 5-35, and the rest (i.e., the competitive fringe $\theta = \mathcal{f}$) with those outside the Top 35.

We parameterize the stochastic process for the borrower's project as follows. For each borrower, let $y^E_j = aZ' - bR + \varepsilon_e$, where ε_e is iid (across agents and time) and drawn from $N(z'_j, \sigma^2_\varepsilon)$. We define success to be the event that $y^E > 0$, so in states with higher ε_e success is more likely. Then

$$\begin{aligned} p(R, Z', z'_j) &= 1 - \Pr(y^e \leq 0 | R, Z', z'_j) \\ &= 1 - \Pr\left(\varepsilon_e \leq -aZ' + bR\right) \\ &= \Phi_{z'_j}\left(aZ' - bR\right), \end{aligned}$$

where $\Phi_{z'_j}(x)$ is a normal cumulative distribution function with mean z'_j and variance σ^2_ε. The stochastic process for the borrower outside option, $\Omega(\omega)$, is taken to be the uniform distribution $[0, \overline{\omega}]$.

[24]Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement (http://www2.fdic.gov/hsob/SelectRpt.asp?EntryTyp=10). See Appendix **??** for a description of the data.

To calibrate the stochastic process for aggregate technology shocks $F(Z', Z)$, we detrend the sequence of TFP using the H-P filter and estimate the following equation:

$$\log(Z_t) = \rho_Z \log(Z_{t-1}) + u_t^Z,$$

with $u_t^Z \sim N(0, \sigma_{u^Z})$. Once parameters ρ_Z and σ_{u^Z} are estimated, we discretize the process using the Tauchen Tauchen (1986) method. We assume that the vector of regional shocks $\{z_j\}$ is drawn from a bivariate normal distribution with zero mean, standard deviation σ_z and covariance between the two shocks equal to ρ_z.

We calibrate the parameters of the deposit process d_θ using the estimates presented in Table 3. We calibrate the cost of deposits using the ratio of interest expenses on deposits to total deposits. We calibrate the cost of external borrowing r_A^- using interest expenses on federal funds over federal funds and the return on securities r_A^+ using interest income from cash, securities and federal funds sold over the equivalent asset categories. We assume $c_\theta(\ell_\theta)$ is a quadratic function with linear term c_θ^1 and quadratic term c_θ^2.

We let the external financing cost take the following form $\xi_\theta(x) = \xi_\theta^1 x$ for $\theta \in \{\boldsymbol{n}, \boldsymbol{r}\}$ and assume that external financing is prohibitively costly for fringe banks $\theta = \boldsymbol{f}$.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 there will be a threshold $\bar{c}_f^1(s_f, Z, \mu)$ that solves

$$\widehat{\pi}_f(s_f, Z, \mu) \equiv \max_{\{\ell_f \geq 0, a_f\}} E_{Z', s'_j|Z} \left[\max\{\pi_f(s_f, Z, \mu), 0\}\right] = 0. \tag{34}$$

This threshold $\bar{c}_f^1(s_f, Z, \mu)$ determines the mass of fringe banks that are active each period. Finally, we assume that c^f is distributed exponentially with location parameter equal to μ_{c^1}.

We do not have enough information on the liquidation value of the assets of large banks (since we do not observe liquidations in the largest category) we set $\zeta_\theta = \zeta$ and calibrate ζ using data from the FDIC.

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 solving the model). After those are set, a second group is estimated using the simulated method of moments. In what follows, we describe both groups of parameters as well as our targeted moments. Table 9 presents the parameters of the model and the targets that were used. Entries above the line correspond to parameters chosen outside the model while entries below the line correspond to parameters chosen within the model by simulated method of moments. In all we have 27 parameters and 27 targeted moments as part of the simulated method of moments estimation.

Table 9: Parameters and Targets

Parameter		Value	Target
Mass of borrowers	B	1	Normalization
Autocorrel Agg Productivity	ρ_Z	0.2990	TFP US (Fernald)
Std. Dev Agg Productivity	σ_{u^Z}	0.0100	TFP US (Fernald)
Failure Value Recovery	ζ	0.804	Recovery Value Bank Failures (FDIC)
Deposit Interest Rate (%)	$\overline{r} = r^D$	0.0014	Avg Interest Expense Deposits
Bank Discount Factor	β	0.9986	$1/(1+\overline{r})$
Return on Securities	r_A^+	0.0233	Return Cash, Securities, and Fed Fund sold
Cost of External Borrowing	r_A^-	0.0147	Cost Fed Funds Purchased
Deposit Process Parameters	$\{\rho_\theta^d, \sigma_{\theta,u}\}$		Deposit Process Estimates (eq 2)
Borrower Success Prob. Function	a	4.291	Avg. Borrower Return
Borrower Success Prob. Function	b	28.940	Avg. Default Frequency
Borrower Success Prob. Function	σ_ϵ	0.107	Avg. Loan Interest Rate
Outside Option	$\overline{\omega}$	0.420	Elasticity of Loan Demand
Std. Dev Reg Shocks	σ_z	0.050	Std Dev Loan Returns
Correlation Regional Shocks	ρ_z	0.002	Std Dev Charge-offs Top 4
Mean deposit process ℓ	$\overline{d}_\ell$	0.028	Relative Size Fringe to Top 5-35
Mean deposit process r	$\overline{d}_r$	1.000	Loans to Output Ratio
Mean deposit process n	$\overline{d}_n$	7.370	Relative Size Top 4 to Top 5-35
Loan Loss Rate	λ	0.314	Avg Charge Off Rate
Mean Dist Cost Loans ℓ	c_ℓ^1	0.035	Net Marginal Expenses No Top 35
Quadratic Cost Loans ℓ	c_ℓ^2	0.085	Elasticity Net Marginal Expenses No Top 35
Linear Cost Loans r	c_r^1	0.002	Net Marginal Expenses Top 5-35
Quadratic Cost Loans r	c_r^2	0.003	Elasticity Net Marginal Expenses Top 5-35
Linear Cost Loans n	c_n^1	0.005	Net Marginal Expenses Top 4
Quadratic Cost Loans n	c_n^2	0.001	Elasticity Net Marginal Expenses Top 4
Fixed cost ℓ	κ_ℓ	0.001	Fixed cost over loans No Top 35
Fixed cost r	κ_r	0.005	Fixed cost over loans Top 5-35
Fixed cost n	κ_n	0.020	Fixed cost over loans Top 4
Mass Potential ℓ	M_ℓ	2.000	Deposit To Output Ratio
Mass Potential r	M_r	0.049	Deposit Market Share Regional
Mass Potential n	M_n	0.012	Deposit Market Share National
External finance param. r	ξ_r^1	0.025	Avg. equity issuance to loan ratio Top 5-35
External finance param. n	ξ_n^1	0.025	Avg. equity issuance to loan ratio top 4
Entry Cost r	Υ_r	2.429	Bank Loan Market Share Regional
Entry Cost n	Υ_n	12.172	Bank Loan Market Share National
Nonbank Marginal Cost	$c_{\mathcal{N}}$	0.045	Bank Loan to Total Credit Ratio

Note: The entry cost is set as part of the equilibrium selection. In particular, in the baseline case, the entry for national and regional is such that there is entry whenever there is failure by national or regional banks.

Table 10 presents a set of data moments together with their model generated counterparts.

Table 10: Moments Data vs Model (Targets and Additional Moments)

	Data	Model
Avg. Borrower Return	12.94	13.66
Avg. Default Frequency	2.14	1.96
Avg. Loan Interest Rate	4.85	4.08
Elasticity of Loan Demand	-1.10	-1.15
Std Dev Loan Returns	0.84	1.19
Std Dev Charge-offs Top 4	0.30	1.05
Avg Charge Off Rate	0.85	0.62
Relative Dep Size Fringe to Top 5-35	0.03	0.03
Relative Dep Size Top 4 to Top 5-35	7.37	7.37
Net Marginal Expenses No Top 35	1.79	2.92
Elasticity Net Marginal Expenses No Top 35	0.74	1.00
Net Marginal Expenses Top 5-35	1.42	1.07
Elasticity Net Marginal Expenses Top 5-35	0.90	1.00
Net Marginal Expenses Top 4	1.14	0.84
Elasticity Net Marginal Expenses Top 4	1.00	1.37
Fixed cost over loans No Top 35	0.71	0.92
Fixed cost over loans Top 5-35	0.69	0.27
Fixed cost over loans Top 4	0.78	0.38
Bank Loans to Output Ratio	60.34	68.57
Deposit To Output Ratio	57.78	47.30
Deposit Market Share Regional	32.45	28.68
Deposit Market Share National	29.60	37.84
Avg. equity issuance to loan ratio Regional	0.04	0.06
Avg. equity issuance to loan ratio National	0.04	0.02
Bank Loan Market Share Regional	33.97	47.74
Bank Loan Market Share National	29.87	18.20
Bank Loan to Total Credit Ratio	50.00	77.46
Bank failure rate (all)	0.490	11.91
Bank failure rate Fringe	0.493	12.52
Bank Failure rate Regional	0.09	0.00
Bank Failure rate National	0.00	0.00
Loan to Assets Fringe	62.73	100.0
Loan to Assets Regional	64.32	89.25
Loan to Assets National	67.20	65.47
Deposits to Assets Fringe	90.33	36.74
Deposits to Assets Regional	91.07	49.15
Deposits to Assets National	92.44	82.28
Avg Cost Fringe	2.50	3.84
Avg Cost Regional	2.11	1.34
Avg Cost National	1.92	1.23

Note: Moments below the line are not targets of the calibration exercise.

6 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.

6.1 Business Cycle Statistics

Table 11 provides the correlation between key aggregate variables with GDP.[25] We observe that, as in the data, the model generates countercylical failure rates, default frequencies, charge-off rates, interest rates, interest spread and procyclical loan supply. At the current calibration the model implies a procyclical interest margin in contrast with the countercyclical interest margin we observe in the data. This mismatch leads to the misalignment of other correlations in the model relative to the data.

Table 11: Business Cycle Statistics

Variable correlated with GDP	Data	Model
Charge-off Rate	-0.223***	-0.823***
Default freq.	-0.455***	-0.823***
Loan return	0.030	0.803***
Net Interest Margin	-0.096***	0.538***
Loan Interest Rate	-0.221**	-0.287**
Markup	-2.359*	0.341**
Failure Rate	-0.162*	-0.721***
Loan Supply	0.945**	0.380***

Note: Data values presented in this table are derived from a regression (at quarterly frequency) of the detrended log variable on detrended log real GDP. Data correspond to commercial banks in the U.S. between 1984 and 2019. Model values derive from a regression of the log variable on log GDP. *** denotes significant at 1%, ** denotes significant at 5%, * denotes significant at 10%. Source: Consolidated Reports of Condition and Income and Model Outcome

Table 12 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.

[25] When calculating business cycle correlations in the model, we correlate variables that depend on $\{Z, \boldsymbol{\mu}\}$ and S' (which includes $\{Z, \boldsymbol{\mu}\}$) to be consistent with the definition of a period in the model. For example, the correlation of the bank loan interest rate corresponds to the correlation between $r_{\mathscr{B}}(Z, \boldsymbol{\mu})$ and $GDP(S')$.

Table 12: Measures of Bank Competition

Moment (%)	Data	Model
Net Interest Margin	4.18	3.86
Lerner	52.48	54.44
Markup	110.44	119.47

6.2 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, Demirgüç-Kunt, and Levine (2006), 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, Klapper, and Turk-Ariss (2009), 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 bank 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 the aggregate loan supply.[26] Table 13 displays the estimated coefficients and their standard errors.

[26]Beck, Demirgüç-Kunt, and Levine (2006) also include other controls like "economic freedom" which are outside of our model.

Table 13: Banking Crises, Default Frequencies and Concentration

Dependent Variable	Crisis_t	Default Freq_t
Concentration_t	-3.935	0.128
s.e.	$(0.569)^{***}$	$(0.061)^{**}$
ΔGDP_t	-85.389	-0.8282
s.e.	$(7.059)^{***}$	$(0.032)^{***}$
Loan Supply_t	46.209	-0.059
se	$(12.184)^{***}$	$(0.052)^{***}$
R^2	0.692	0.421
Model	Logit	Linear

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, Demirgüç-Kunt, and Levine (2006), 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 (2000). Consistent with the evidence in Berger, Klapper, and Turk-Ariss (2009) 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 (2005), who showed that higher concentration can be associated with riskier loan portfolios.

7 Too-Big-To-Fail

The top 4 commercial bank loan market share was 40% in the last quarter of 2007. Since our paper admits a nontrivial endogenous size distribution of banks and includes non-atomistic banks, it is well suited to analyze size dependent policy changes like that described above by Ben Bernanke.

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 loan exposure in order to maintain their charter value. However, a policy of big bank bailouts or "too-big-to-fail" (TBTF) 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 ($\theta = \boldsymbol{n}$) solves (23)-(29) with (28) now given by:

$$\mathscr{D}(s_\theta, S') = \begin{cases} \pi_\theta(s_\theta, S') & \text{if } \pi_\theta(s_\theta, S') \geq 0 \\ \pi_\theta(s_\theta, S')(1-\varphi) - \xi_\theta(\pi_\theta(s_\theta, S')(1-\varphi)) & \text{if } \pi_\theta(s_\theta, S') < 0 \end{cases} . \tag{35}$$

Note that there is full certainty about the bailout and that the bank receives funds from the government only when realized profits are negative.[27] These losses are paid for by taxes as in the case of the deposit insurance. In Table 14, we present the results of the case when $\varphi = 1$.

Table 14: Benchmark vs Model with National Banks Bailouts

	Benchmark	National Bank Bailout Change (%)
Bank Loan Supply	0.39	3.27
Total Loan Supply	0.50	0.86
Share Bank Loans (%)	77.46	2.39
Markup (%)	119.47	-5.46
Loan Market Share Fringe (%)	34.06	-13.44
Loan Market Share Regional (%)	47.74	-8.16
Loan Market Share National (%)	18.20	46.55
Borrower Risk Taking	13.92	-0.01
Default Frequency (%)	1.96	-1.41
Net Interest Margin (%)	3.86	-2.93
Bank Loan Interest Rate (%)	4.08	-2.85
Failure Rate (%)	11.91	-4.03
Output	0.57	0.85
Taxes	0.00	-7.26
Taxes/Output (%)	0.87	-7.90

Unlike our benchmark equilibrium where there was no national bank failure on-the-equilibrium path, with TBTF national banks make negative profits on-the-equilibrium path when the economy heads into a recession. The unconditional probability of a government bailout equals 6.54% and it can cost up to 0.024% of output.

However, as evident in Table 14, the introduction of TBTF induces the national bank to make more loans (i.e. total bank loan supply increases 3.27%) since there is no cost to the increased exposure. The increased loan supply by national banks raises its market share

[27]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.

at the expense of regional and fringe banks. Increased bank loan supply lowers bank loan interest rates by 2.85% which induces borrower default frequencies to fall. Lower equilibrium default induces a lower failure rate by fringe banks which ultimately reduces the level of taxes over GDP necessary to cover bank losses (a 7.90% decline) despite experiencing infrequent national bank bailouts over the long run. The reduction in the interest rate results in a reduction in bank profitability generating a drop in markups (-5.46%) which induces a lower entry rate (-4.03%, similar to the reduction in failure rates).

8 Concluding Remarks

To be added.

References

Aguirregabiria, Victor, Robert Clark, and Hui Wang. 2016. "Diversification of geographic risk in retail bank networks: evidence from bank expansion after the Riegle-Neal Act." *The RAND Journal of Economics* 47 (3):529–572.

Allen, Franklin and Douglas Gale. 2000. *Comparing financial systems.* MIT press.

———. 2004. "Competition and financial stability." *Journal of money, credit and banking* :453–480.

Arellano, Manuel and Stephen Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The review of economic studies* 58 (2):277–297.

Beck, Thorsten, Asli Demirgüç-Kunt, and Ross Levine. 2006. "Bank concentration, competition, and crises: First results." *Journal of banking & finance* 30 (5):1581–1603.

Begenau, Juliane and Tim Landvoigt. 2022. "Financial regulation in a quantitative model of the modern banking system." *The Review of Economic Studies* 89 (4):1748–1784.

Berger, Allen N, Leora F Klapper, and Rima Turk-Ariss. 2009. "Bank competition and financial stability." *Journal of Financial Services Research* 35 (2):99–118.

Berger, Allen N and Loretta J Mester. 1997. "Inside the black box: What explains differences in the efficiencies of financial institutions?" *Journal of banking & finance* 21 (7):895–947.

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

Bikker, Jacob and Katharina Haaf. 2002. "Competition, concentration and their relationship: An empirical analysis of the banking industry." *Journal of Banking and Finance* 26 (11):2191–2214.

Boyd, John H and Gianni De Nicolo. 2005. "The theory of bank risk taking and competition revisited." *The Journal of finance* 60 (3):1329–1343.

Buchak, Greg, Gregor Matvos, Tomas Piskorski, and Amit Seru. 2018. "Fintech, regulatory arbitrage, and the rise of shadow banks." *Journal of Financial Economics* 130:453–483.

Carlstrom, Charles and Timothy Fuerst. 1997. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis." *American Economic Review* 87 (5):893–910.

Cooley, Thomas F and Vincenzo Quadrini. 2001. "Financial markets and firm dynamics." *American economic review* 91 (5):1286–1310.

Corbae, Dean and Pablo D'Erasmo. 2021. "Capital Buffers in a Quantitative Model of Banking Industry Dynamics." *Econometrica* 89:2975–3023.

Corbae, Dean and Ross Levine. 2024. “Competition, Stability, and Efficiency in the Banking Industry.” *mimeo* .

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. “The rise of market power and the macroeconomic implications.” *The Quarterly Journal of Economics* 135 (2):561–644.

Diamond, Douglas W. 1984. “Financial intermediation and delegated monitoring.” *The review of economic studies* 51 (3):393–414.

Ericson, Richard and Ariel Pakes. 1995. “Markov-perfect industry dynamics: A framework for empirical work.” *The Review of economic studies* 62 (1):53–82.

Farias, Vivek, Denis Saure, and Gabriel Y Weintraub. 2012. “An approximate dynamic programming approach to solving dynamic oligopoly models.” *The RAND Journal of Economics* 43 (2):253–282.

Gabaix, Xavier and Rustam Ibragimov. 2011. “Rank- 1/2: a simple way to improve the OLS estimation of tail exponents.” *Journal of Business & Economic Statistics* 29 (1):24–39.

Gertler, Mark and Peter Karadi. 2011. “A model of unconventional monetary policy.” *Journal of monetary Economics* 58 (1):17–34.

Gowrisankaran, Gautam and Thomas J Holmes. 2004. “Mergers and the evolution of industry concentration: results from the dominant-firm model.” *RAND Journal of Economics* :561–582.

Hennessy, Christopher A and Toni M Whited. 2007. “How costly is external financing? Evidence from a structural estimation.” *The Journal of Finance* 62 (4):1705–1745.

Janicki, Hubert and Edward S Prescott. 2006. “Changes in the size distribution of US banks: 1960-2005.” *FRB Richmond Economic Quarterly* 92 (4):291–316.

Jeske, Karsen, Dirk Krueger, and Kurt Mitman. 2013. “Housing and the Macroeconomy: The Role of BAilout Guarantees for Government Sponsored Enterprises.” *Journal of Monetary Economics* 60:917–935.

Kashyap, Anil K and Jeremy C Stein. 2000. “What do a million observations on banks say about the transmission of monetary policy?” *American Economic Review* 90 (3):407–428.

Liang, Nellie and Stephen A Rhoades. 1988. “Geographic diversification and risk in banking.” *Journal of Economics and Business* 40 (4):271–284.

Martinez-Miera, David and Rafael Repullo. 2010. “Does competition reduce the risk of bank failure?” *The Review of Financial Studies* 23 (10):3638–3664.

Menzio, Guido and Shouyong Shi. 2010. “Block recursive equilibria for stochastic models of search on the job.” *Journal of Economic Theory* 145 (4):1453–1494.

Oberfield, Ezra, Esteban Rossi-Hansberg, Nicholas Trachter, and Derek Wenning. 2024. "Banks in Space. NBER Working Paper No. 32256." *National Bureau of Economic Research* URL `https://www.nber.org/papers/w32256`.

Pakes, Ariel and Paul McGuire. 1992. "Computing markov perfect nash equilibria: Numerical implications of a dynamic differentiated product model. NBER Technical Working Paper No. 0119." *National Bureau of Economic Research Cambridge, Mass., USA* .

Shaffer, S. 2004. "Patterns of Competition in Banking." *Journal of Economics and Business* 56:287–313.

Tauchen, George. 1986. "Finite state markov-chain approximations to univariate and vector autoregressions." *Economics letters* 20 (2):177–181.

Wang, Yifei, Toni M Whited, Yufeng Wu, and Kairong Xiao. 2022. "Bank market power and monetary policy transmission: Evidence from a structural estimation." *The Journal of Finance* 77 (4):2093–2141.

Williamson, Stephen D. 1986. "Costly monitoring, financial intermediation, and equilibrium credit rationing." *Journal of Monetary Economics* 18 (2):159–179. URL `https://ideas.repec.org/a/eee/moneco/v18y1986i2p159-179.html`.