

A Quantitative Model of Banking Industry Dynamics

Dean Corbae

University of Wisconsin–Madison, NBER, and Visiting Scholar, Federal Reserve Bank of Philadelphia Research Department

Pablo D’Erasmus

Federal Reserve Bank of Philadelphia

WP 26-45

PUBLISHED

October 2026

ISSN: 1962-5361

Disclaimer: This Philadelphia Fed working paper represents preliminary research that is being circulated for discussion purposes. The views expressed in these papers are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. Philadelphia Fed working papers are free to download at: <https://www.philadelphiafed.org/search-results/all-work?searchtype=working-papers>.

DOI: <https://doi.org/10.21799/frbp.wp.2026.45>

A Quantitative Model of Banking Industry Dynamics*

Dean Corbae

University of Wisconsin - Madison and NBER

Pablo D'Erasmo[†]

Federal Reserve Bank of Philadelphia

September 15, 2026

Abstract

We develop a simple model of banking industry dynamics to study the relation between market structure and financial stability. Motivated by data, we analyze an environment where a small number of dominant (non-atomistic) banks strategically interact with a large number of small (atomistic) banks. A nontrivial endogenous bank size distribution arises out of entry and exit in response to aggregate and regional shocks to borrowers' production technologies and banks' idiosyncratic deposit inflows. Since the model has non-atomistic banks that generate granular spillovers from idiosyncratic shocks and pose systemic risk, we use the model to study the implications of too-big-to-fail policies.

*The authors wish to thank Manuel Amador, Andy Atkeson, Javier Bianchi, John Boyd, Jeff Campbell, Gianni DeNicolò, Luca Guerrieri, Hugo Hopenhayn, Nobu Kiyotaki, Jim MacGee, Fabrizio Perri, 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 Macroeconomic Policy Perspectives Conference held at the Minneapolis Fed, 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. Edited by Manuel Amador and Javier Bianchi.

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

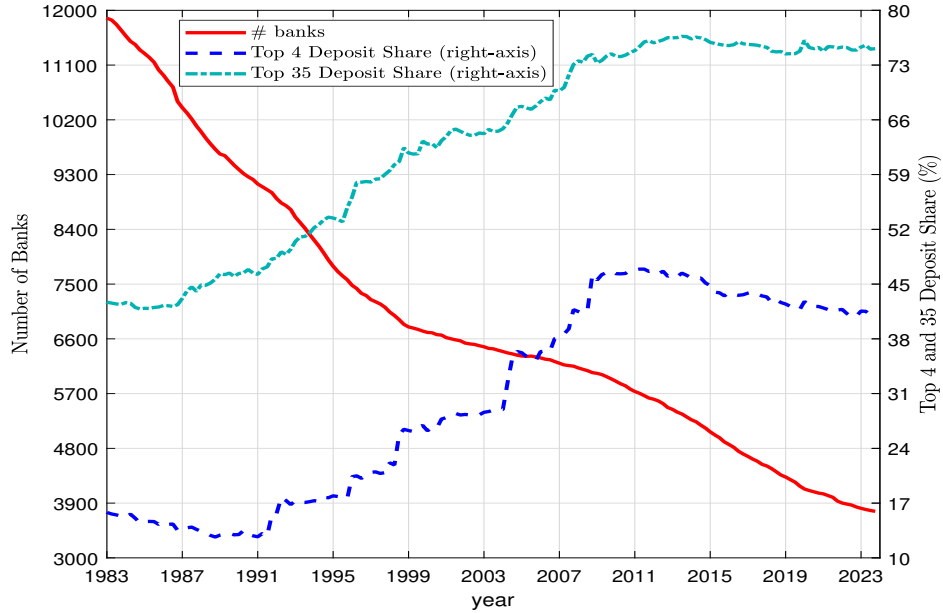
“I want to be very, very clear: too-big-to-fail is one of the biggest problems we face in this country, and we must take action to eliminate too-big-to-fail.”

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 those institutions commonly referred as too-big-to-fail (TBTF)) 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 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 are associated with regional banks (which make up another 30% of deposit market share), and the remaining roughly 4,000 banks are associated with the fringe banks (which make up the remainder of deposit market share). We think of the top 35 banks as being non-atomistic exerting some degree of market power while the fringe are atomistic (of measure zero).¹

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 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 would predict countercyclical policies to bail out failing banks. However, there is the moral hazard concern echoed by Ben Bernanke that TBTF policies may induce more risk taking.

Our model with heterogeneous and nonatomistic 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 big bank failure in the presence of moral hazard arising from deposit insurance?

The idea that big banks pose a systemic risk to the economy echoed in Bernanke’s statement is related to the notion of granularity in Gabaix (2011). While Gabaix focused on how idiosyncratic shocks to the growth rate of the largest 100 firms can explain 30% of aggregate output fluctuations, idiosyncratic funding shocks to our dominant banks can spill over to the aggregate loan market which funds economic activity unlike idiosyncratic shocks to atomistic (measure zero) fringe banks.² In this way, dominant banks can pose systemic risk to the economy through strategic spillovers posing a policy challenge for regulators.

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).³ Instead of assuming perfect competition among atomistic (measure zero banks), we propose a Stackelberg model where non-atomistic 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). Our framework actually nests the competitive model under certain assumptions on entry costs. To account for the rapid rise of the non-bank (or shadow sector), we include a simple competitive non-bank financial intermediary. In our baseline, we assume that there exists the

¹While the number of banks that are stress tested varies by year, the top 35 were stress-tested in 2017.

²See also Galaasen et al. (2020), who study how shocks to large firms spill over to banks and the rest of the economy.

³There is a large literature following this tradition. See, for example, Guerrieri et al. (2019), who assesses the performance of five different models when faced with bank capital shortfalls.

possibility of big bank bailouts. We model these bailouts as probabilistic equity injections in negative bank profit states that cover only a fraction of the losses.

Idiosyncratic shocks to deposit funding consistent with [Aguirregabiria, Clark, and Wang \(2016\)](#) along with endogenous exit and entry generate 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 like TBTF or idiosyncratic shocks to big, non-atomistic banks spill 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 that 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 TBTF policy that 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 inducing excess failure of smaller banks that are not covered by TBTF as we show in [Table 10](#). Further, the presence of non-atomistic banks in our model implies the existence of a pecuniary externality; any given big bank internalizes its effect on its own profitability and the reaction of others but not its negative consequences on other banks.⁴ Here, unlike models with atomistic banks, we can quantitatively assess the impact of one dominant bank’s effect on equilibrium prices and outcomes in an environment with moral hazard associated with deposit insurance and costly exit.

Here we can use our model to assess the influence of granular shocks to big bank funding. For instance, in the regional model we study, even a shock to a regional bank (e.g. SVB in the West) can spill over to the national economy. Specifically, a decrease in deposits at a regional bank *ceteris paribus* lowers loan supply in that region and puts upward pressure on interest rates there. Higher profitability in that region incentivizes the national bank and fringe banks in that region to increase their loan supply. The ensuing difference in profitability across regions induces the national bank to transfer funds to the more profitable region (e.g. the West), decreasing loans in the other region (e.g. the East). Thus a shock to a regional bank can spill over to create an aggregate disturbance (i.e. in both regions).

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

⁴This pecuniary externality (or business stealing) exists in virtually all industrial organization models with strategic interactions.

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.⁵ In particular, we compare our benchmark economy where the government is committed to covering a fraction of negative profits with some positive probability thereby reducing the likelihood of failure compared to one where those bailouts are absent. In the benchmark case, the possible bailout induces the national bank to take on more risk and augment its loan supply as the bailout reduces the need for costly equity injections in negative profit cases. In the counterfactual without bailouts, the loss of charter value and costs of equity issuance are 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. Regional and fringe banks increase their loan supply to make use of the profit opportunity but the decrease in loans by national banks dominates, generating higher interest rates. We also observe that a portion of the decline in loans by national banks is met by increased nonbank lending. Due to the risk shifting effect, borrowers take on more risk and default frequencies increase. The higher default frequency does not lead to an increase in failure by smaller banks because the rise in interest rates more than compensates this increase and results in higher markups and a lower failure rate. The decline in failures together with the decrease in taxes due to elimination of national bank bailouts results in an overall decrease in the tax-to-output ratio.

Literature: Our paper is most closely related to the following literature on the industrial organization of banking.⁶ 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.⁷ 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, as we did in [Corbae and D’Erasmus \(2020\)](#) and [Corbae and D’Erasmus \(2021\)](#). To keep the model simple, we focus only on loan market competition. There is an important IO literature on deposit market competition (see for example [Aguirregabiria, Clark, and Wang \(2025\)](#)) and both markets (see for example [Wang et al. \(2022\)](#) and [Oberfield et al. \(2024\)](#)).

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 spirit of work by [Buchak et al. \(2018\)](#) or the competitive models of [Begenau and Landvoigt \(2022\)](#) and [Dempsey \(2026\)](#), our model includes competition between banks and nonbanks.

Outline: The remainder of the paper is organized as follows. Section 2 documents a select

⁵See [Meehl \(2024\)](#) for a related paper examining the role of different bank resolution policies on financial stability and real outcomes. Earlier contributions in this literature include [Ennis and Malek \(2005\)](#) and [Stern and Feldman \(2004\)](#). More recently, [Dávila and Walther \(2020\)](#) explore how size heterogeneity affects leverage decisions and risk in a model with bailouts and [Berndt, Duffie, and Zhu \(2025\)](#) estimate the market-implied probabilities of government bailouts.

⁶There are several related papers that make a competitive banking industry assumption: [Bianchi and Bigio \(2022\)](#), [Jamilov and Monacelli \(2026\)](#), and [Rios-Rull, Takamura, and Terajima \(2023\)](#).

⁷A dynamic version of the model is considered in [Martinez-Miera and Repullo \(2010\)](#).

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 untargted 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 studies how idiosyncratic shocks to bank deposit flows spill over to the aggregate economy via changes in credit. Section 8 conducts our main counterfactual assessing the costs and benefits of a too-big-to-fail policy. Section 9 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.⁸ 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 Appendix A-1, 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, 1,383 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.⁹ We also use the Summary of Deposits (SOD), the annual survey of branch office deposits as of June 30 for all FDIC-insured institutions.¹⁰

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 4,000 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.¹¹ 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 Riegle-Neal, there was significant rise in market share of top 4 and top 35 banks until the

⁸The public version of balance sheet and income statements items can be found at <https://cdr.ffiec.gov/public/> (link last accessed 07/21/2025).

⁹We 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.

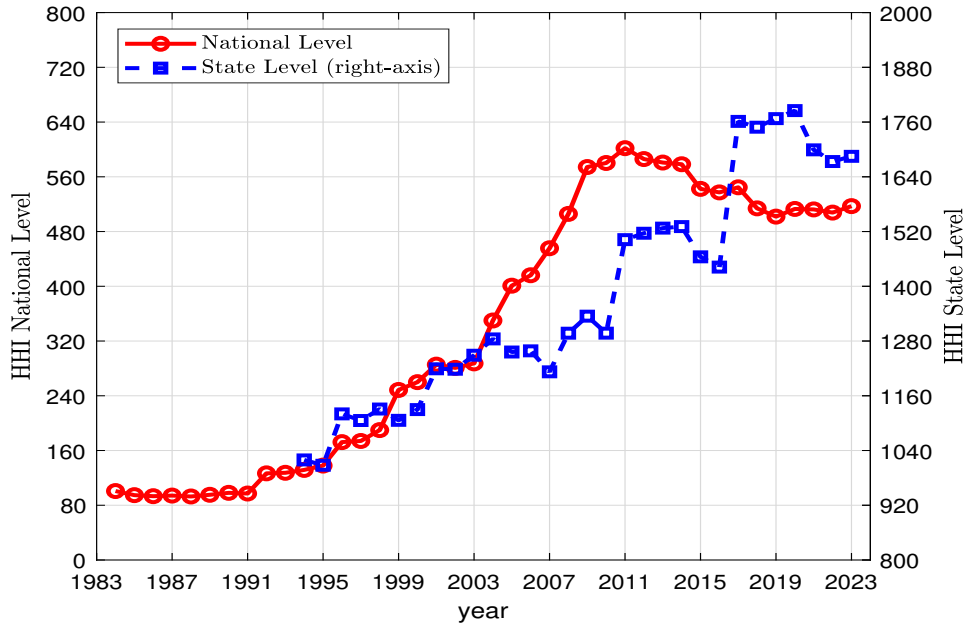
¹⁰Data can be found here: <https://banks.data.fdic.gov/bankfind-suite/SOD>.

¹¹While our paper equates size with systemic risk and interconnections through strategic interactions, there may be other (G-SIB) criteria like complexity or other forms of interconnections that also pose a systemic risk.

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 4,000 in 2023.

Figure 2 graphs deposit Herfindahl-Hirschman indices (HHI) at the national and state levels.¹² High levels of HHI are considered to be 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% since 1994.

Figure 2: Deposit Herfindahl–Hirschman Index (HHI)



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

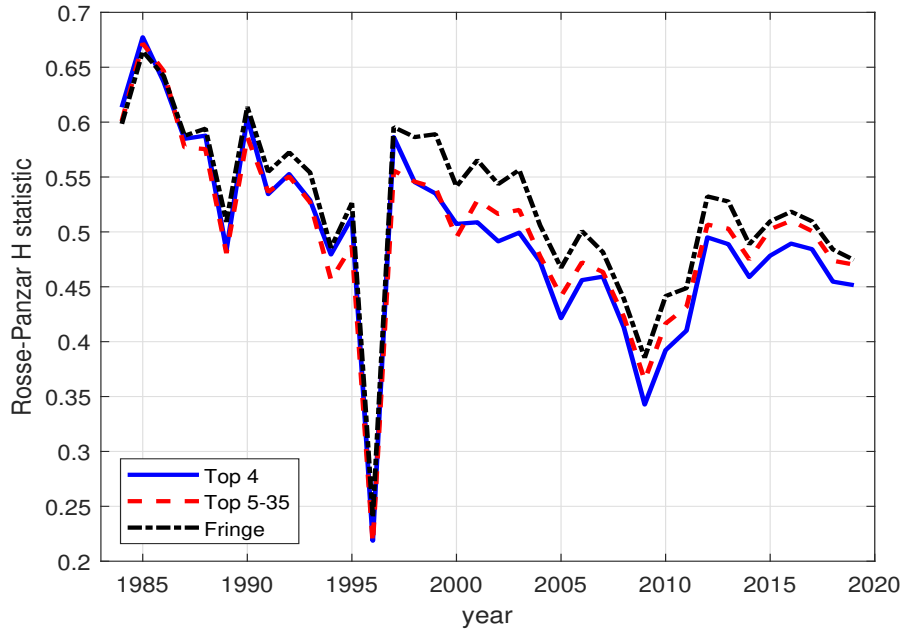
To obtain an additional measure of competition, we follow [Shaffer \(2004\)](#) and estimate the elasticity of marginal revenue with respect to factor prices (i.e. the Panzar-Rosse H statistic) by a log-linear regression in which the dependent variable is the natural logarithm of total revenue (or income)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

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

¹²The 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 1,500 and 2,500 points to a market moderately concentrated and one with an HHI of 2,500 is considered highly concentrated by the anti-trust division of the Justice Department.

The Panzar-Rosse H statistic equals the simple sum of coefficients on the respective log input price terms, $\beta_1 + \beta_2 + \beta_3$.¹³ This equation is estimated year by year with robust standard errors. Figure 3 presents the estimates by bank-size over time. The average Panzar-Rosse 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 40.1% and Figure 3 makes evident that this measure of pass-through is lower for larger banks.

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



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

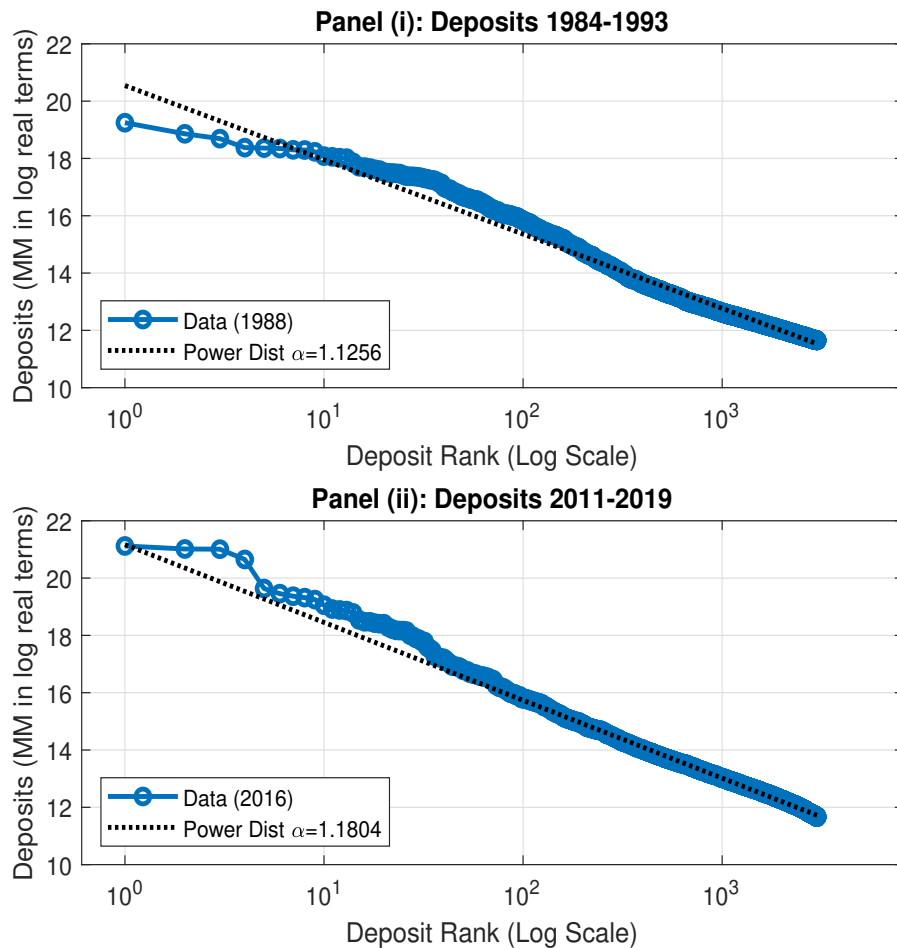
Some firm dynamics studies have found that the firm size distribution is well approximated by a Power law distribution due to a thicker right tail than the lognormal.¹⁴ Let $d(r) = b \cdot r^{-\alpha}$, where r is the rank and $d(r)$ is the deposits of bank with rank r . Zipf's law predicts that $\alpha = 1$. We estimate α as in Gabaix and Ibragimov (2011) year by year between 1984 and 2019. Figure 4 presents the average estimate for two different periods in our sample: 1984-1993 (Panel (i)) and 2011-2019 (Panel (ii)). We plot the actual distribution in the data for the midpoints of each sample to illustrate the growth of the right tail against the

¹³The log-linear form typically improves the regression's goodness of fit and may reduce simultaneity bias.

¹⁴See Janicki and Prescott (2006) for a study with a focus on the U.S. bank distribution between 1960 and 2005.

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, from 1.118 in 1984 to 1.282 in 2019 implying that the right tail is getting thicker over time.¹⁵

Figure 4: 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 3,000 banks (like [Janicki and Prescott \(2006\)](#)). The reported coefficient (and corresponding black dotted line) in each panel corresponds to the average during the period. The data distribution presented for 1988 (Panel i) and 2016 (Panel ii) correspond to the year that is the middle point in the time series. Source: Reports of Condition and Income (Call Reports)

Next we examine the geographic expansion that followed the Riegle-Neal Act in 1994 in Figure 5. We provide two measures of geographic expansion. First, the top panel graphs the number of states in which a bank in each size category has an active branch for each

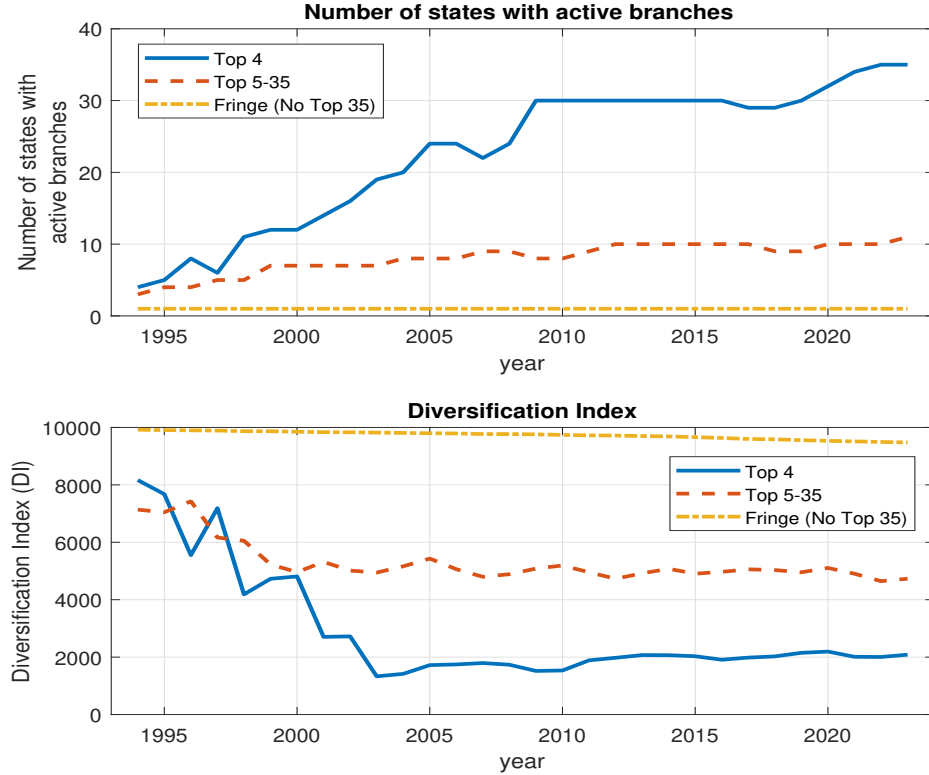
¹⁵See Table A.2 in Appendix A-2 for estimates and standard errors.

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:

$$DI_{i,t} = \sum_{m \in M_{i,t}} s_{i,m,t}^2. \quad (1)$$

This index ranges between 0 and 10,000 and a smaller value indicates a more diversified funding base. The bottom panel in Figure 5 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 5: 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 quarter t :

$$\log(d_{\theta,t}^i) = (1 - \rho_{\theta}^d) v_{\theta,t}^0 + \rho_{\theta}^d \log(d_{\theta,t-4}^i) + u_{\theta,t}^i, \quad (2)$$

where $\delta_{\theta,t}^i$ is the sum of deposits and other borrowings in quarter t for bank i , and $u_{\theta,t}^i$ is iid and distributed $N(0, \sigma_{\theta,u}^2)$. The parameter ρ_{θ}^d captures the one year autocorrelation consistent with the time period in the model.¹⁶ From these estimates, we can construct the variance of deposits by bank size (i.e. $\sigma_{\theta,d} = \frac{\sigma_{\theta,u}}{(1 - (\rho_{\theta}^d)^2)^{1/2}}$) that we present in the last column of Table 1. Thus, consistent with big banks having a geographically diversified pool of funding, we find big banks have less volatile funding inflows.¹⁷

Table 1: Deposit Process Parameters

	ρ_{θ}^d	$\sigma_{\theta,u}$	$\sigma_{\theta,d}$
Top 4 Banks	0.700	0.092	0.129
Top 5-35 Banks	0.825	0.149	0.264
Fringe Banks	0.803	0.206	0.345
All Banks	0.806	0.207	0.351

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)

Another key element of our analysis is the presence of increasing returns in the banking data we study. We estimate the marginal cost of producing a loan and fixed costs following the empirical literature on banking (see, for example, [Berger, Klapper, and Turk-Ariss \(2009\)](#)) and our previous work [Corbae and D’Erasmus \(2021\)](#).¹⁸ The estimated (asset-weighted) average of marginal non-interest expenses is reported in the first column of Table 2, while the estimated (asset-weighted) average of marginal non-interest income is reported in the second column of Table 2. Net marginal expenses (Mg Net Exp.) are computed as the difference between marginal non-interest expenses and marginal non-interest income and its estimate is reported in the third column of Table 2. The fact that marginal net non-interest costs are decreasing in size is consistent with economies of scope. The fixed cost is estimated as the total cost on expenses of premises and fixed assets. The estimated (asset-weighted) average fixed cost, taken as costs on premises and fixed assets, scaled by loans is reported in

¹⁶Estimates are robust to employing yearly data and using the method proposed by [Arellano and Bond \(1991\)](#) to estimate the autocorrelated process (see [Corbae and D’Erasmus \(2021\)](#)).

¹⁷These findings are consistent with [Liang and Rhoades \(1988\)](#) and [Aguirregabiria, Clark, and Wang \(2016\)](#).

¹⁸Details of the equations we use to estimate these costs can be found in Appendix A-2.

the fourth column of Table 2. The final column of Table 2 presents our estimate of average costs (average marginal net expense plus average fixed cost) for big and small banks. We find a statistically significant lower average cost 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 2: Banks' Cost Structure

	Mg Non-Int Exp.	Mg Non-Int Inc.	Mg Net Exp.	Fixed Cost	Avg.
Top 4 Banks	5.48 ^{†,‡}	4.74 ^{†,‡}	0.74 ^{†,‡}	0.73 ^{†,‡}	1.47
Top 5 - 35 Banks	4.32 [‡]	3.25 [‡]	1.07 [‡]	0.70 [‡]	1.77
Fringe Banks	3.55	1.69	1.86	0.75	2.62
All Banks	4.41	3.16	1.25	0.73	1.98

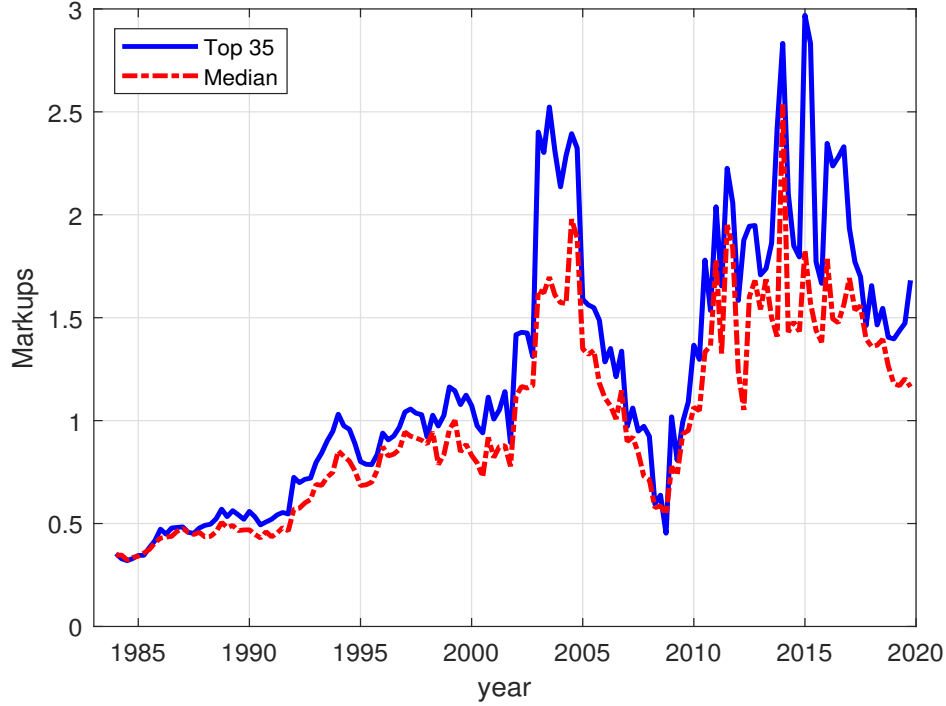
Note: Top 4 Banks and Top 5-35 Banks refers to the top 4 banks and top 5-35 banks when sorted by assets, respectively. Fringe Banks refers to all banks outside the top 35. [†] denotes statistically significant difference between the top 4 and top 5-35. [‡] denotes statistically significant difference 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 from Table 2 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 2). 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 6 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 6: Rise in Bank Loan Markups



Note: We estimate markups using the estimates of marginal costs in equation (A.2.2). We compute markups as price over marginal cost where price corresponds to interest income from loans over loans plus marginal non-interest income and cost corresponds to interest expense on deposits over deposits plus marginal non-interest expense. Source: Reports of Condition and Income (Call Reports).

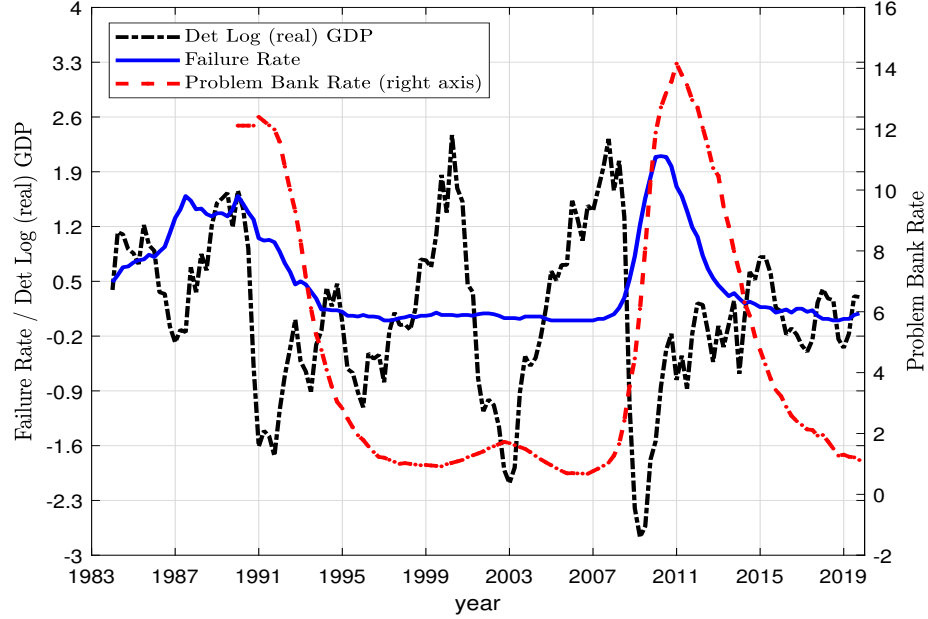
As we showed in Figure 1 the total number of commercial banks has declined consistently in the last 4 decades. In our previous work (Corbae and D’Erasmus (2020)) we showed that changes in regulations resulted in a wave of bank exits due to mergers accounting for the trend. We focus here on the business cycle properties of bank failure. Figure 7 presents the dynamics of the bank 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.¹⁹ The correlation between the failure rate and detrended log real GDP is -0.237 while the correlation between the problem bank rate and log real GDP is -0.72 (which is available only since mid 1990).²⁰ That is, there is evidence that bank failure is countercyclical. Table A.3 in Appendix A-2 documents that the bulk of entry and failures correspond to banks that

¹⁹Banks 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>).

²⁰If 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 -0.469. That is, the failure rate is significantly more countercyclical during this period.

are in the bottom of the distribution (i.e., outside the top 35).

Figure 7: 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. The failure rate corresponds to a four quarter moving average.

Source: Consolidated Report of Condition and Income (Call Reports) and FDIC.

Table A.4 in Appendix A-2 describes the structure of asset returns and funding costs across the bank size distribution. 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, and loan interest spreads decline with bank size. We do not find statistically significant differences in means 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 6, we find that average markups and the Lerner index are increasing in bank size (and statistically significantly different from one another).

The dynamics of entry and exit discussed above as well as differences in intensive margin responses lead to interesting business cycle dynamics. Table A.5 in Appendix A-2 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 detrended 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 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 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 banks in Figure 6).
2. Large right tail deviations from Zipf’s law in Figures A.1 and 4 and Table A.2.
3. Geographic diversification by the top 4 and 35 banks in Figure 5.
4. Diversification of idiosyncratic deposit shocks resulting in lower variance of funding inflows in Table 1.
5. Increasing returns to scale as in Table 2.
6. Countercyclical failure rate in Figure 7 as well as the predominance of bank entry and exit by small banks in Table A.3.
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, and loan interest spreads, decline with bank size in Table A.4.
8. While we do not find statistically significant differences in means across banks of different sizes in the cost of deposits, loan interest rates, or the return of (net) securities, we do find statistically significant differences in markups and the Lerner index that are increasing with bank size in Table A.4.²¹
9. Countercyclical charge-offs, net interest margins, and markups in Table A.5.

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. Motivated by high concentration of the top 35 national and regional banks versus the other 4,000 fringe banks, we consider a dominant-fringe framework along the lines of the Gowrisankaran and Holmes (2004) with Cournot competition in the loan market. The fact that there has been rising loan market competition from non-banks motivates us to add a non-bank (or shadow) sector along the lines of Buchak et al. (2018) to capture regulatory arbitrage. The fact that charge-off rates and default are countercyclical as is bank failure motivates us to allow for exit (and entry) generating an endogenous size distribution of banks.

²¹We note that d’Avernas et al. (2023), using Call Report data, find that small banks offer saving deposit rates about 24 basis points larger than those offered by big banks (14 large institutions) but no differences for time and transaction deposits between 2001 and 2020. Using the same dataset but different size groupings and sample period, we do not find evidence for statistically different deposit rates across our size groups. See Appendix A-2 for more details.

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 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. In all that follows, to keep notation manageable, we let any beginning of period t variable be denoted x and any end of period variable be denoted x' .

To simplify the model, we organize bank balance sheets by asset size along the lines of Table 3. Assets are grouped into securities, loans, and fixed and other assets. Liabilities are grouped into deposits, fed funds borrowed, and equity. Securities include cash plus federal funds sold plus securities. Loans include all loans on the balance sheet and fixed and intangible assets includes fixed assets and other assets not classified into securities or loans (except trading assets). Deposits include total deposits and other borrowings (which include home loan advances). Fed Funds refer to federal funds purchased. By working with this categorization we account for approximately 96% of total assets and about 95% of total liabilities from the Call Report data. We normalize the categories so the data totals add up to 1 in Table 3.

Table 3: Balance Sheet by Bank Size

	Bank Size			
	All Banks	Top 4	Top 5-35	Fringe
Assets (% of Total Assets)				
Securities (a^+)	35.68	30.43	30.81	35.70
Loans (ℓ)	53.83	61.30	59.96	53.80
Fixed & Intangible Assets (Υ)	10.49	8.27	9.22	10.50
Liabilities (% of Total Assets)				
Deposits (d)	88.38	83.80	83.25	88.41
Fed Funds (a^-)	1.14	7.93	7.53	1.11
Equity (e)	10.49	8.27	9.22	10.50

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

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 that may fail

or a household loan like a mortgage that may be foreclosed.²² The project requires one unit of investment (i.e. a loan from either 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:

$$\begin{cases} 1 + z'_j Z' R_{k,j} & \text{with prob } p_j(R_{k,j}, z'_j, Z') \\ 1 - \lambda & \text{with prob } [1 - p_j(R_{k,j}, z'_j, Z')] \end{cases} \quad (3)$$

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 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 $R_{k,j}$ 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 $k \in \{\mathcal{B}, \mathcal{N}\}$. Bank and non-bank interest rates on their loans to the borrower can differ. Taking the vector of interest rates $\mathbf{r}_j = \{r_{\mathcal{B},j}, r_{\mathcal{N},j}\}$ 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 $\epsilon = \{\epsilon_{\mathcal{B}}, \epsilon_{\mathcal{N}}\}$ affecting the value of taking a loan from each type of lender additively. We assume that ϵ_k are i.i.d. 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 $\omega \in [0, \bar{\omega}]$ 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

²²Long term mortgages can be thought of as a sequence of one period loans as in [Jeske, Krueger, and Mitman \(2013\)](#).

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 4 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 4: Borrower's Problem (conditional on investing)

Borrower Chooses $R_{k,j}$	Receive	Pay	Probability
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.²³ 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

²³While 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'Erasmus \(2021\)](#).

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 as well taxes to cover bank bailouts. 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.

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 $\theta \in \Theta = \{n, r, f\}$ for national, regional, and fringe respectively, with size ranking $f < r < n$. 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.²⁴ 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 Section 2. 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 $\ell_{\theta,j}^i$. As in [Corbae and D'Erasmus \(2021\)](#), bank type θ determines the mean and variance of a bank's deposits $d_\theta^i \in D_\theta$. 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 1, the mean of the deposit process satisfies $\bar{d}_n > \bar{d}_r > \bar{d}_f$ so that higher types have a bigger funding base. Furthermore, also consistent with the data presented in Table 1, the variance of deposits satisfy $\sigma_n \leq \sigma_r \leq \sigma_f$ so that bigger banks have lower variance consistent with diversification. We discretize the continuous deposit process d_θ^i 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 (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 (1) and regions are symmetric with respect to state variables, to simplify on notation we abstract from denoting the regional origin of d_θ^i .

A given incumbent bank i is randomly matched with a set of potential household depositors d_θ^i who receive deposit interest rate $r_{D,j}$ and then decide how many loans $\ell_{\theta,j}^i$ to

²⁴As in many industrial organization models, one could make a differentiated loan market assumption to yield variation in loan rates across banks.

extend. The flow balance sheet constraint for regional or fringe banks is $\ell_{\theta,j}^i + a_\theta^i = d_\theta^i$. National banks are geographically diversified in the sense that they extend loans and receive deposits in both regions ($\sum_j \ell_{n,j}^i + a_n^i = d_n^i$). Bank i can fund an amount of loans larger than its deposits by using external borrowing $a_\theta^i < 0$ at rate $r_A^-(a_\theta^i) > r_{D,j}$ where $r_A^-(a_\theta^i)$ is an increasing function of the amount borrowed $|a_\theta^i| < 0$. If a bank chooses an amount of loans lower than its capacity constraint d_θ^i , the leftover deposits $a_\theta^i > 0$ can be invested at rate $r_A^+ \leq r_A^-(a_\theta^i)$.²⁵ This assumption, which rules out arbitrage opportunities (i.e., borrowing securities at a rate lower than the saving rate), can be justified by making the borrowing rate $r_A^-(a_\theta^i)$ an increasing function of the amount borrowed a_θ^i reflecting the riskiness of excessive borrowing.²⁶ Note that since the outside option for a household matched with a bank is to store at rate $\bar{r}$, we know that $r_{D,j} \geq \bar{r}$.

In Table 2 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 $c_\theta(\ell_{\theta,j}, Z)$ (i.e., the difference between non-interest expenses and non-interest income on loans) per unit of loans extended. To capture heterogeneity among fringe banks, we assume c_ℓ^1 are drawn from a distribution with cdf $\Xi(c_\ell^1)$. The linear terms for banks of type $\theta \in \{n, r\}$ 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 $\ell_{\theta,j}^i$ 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.²⁷ 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 $\ell_{\theta,j}^i$ in region j for an incumbent bank i of type θ in industry state μ (to be described below) depends on its individual state s_θ with $s_n = d_n$, $s_r = (j, d_r)$, and $s_\ell = (j, d_\ell, c_\ell^1)$, and its end-of-period aggregate state $S' = (\mu, Z, z'_j, z'_{-j}, Z')$

$$\pi_{\theta,j}^\ell(s_\theta, S') = [p_j(R_{\mathcal{B},j}, z'_j, Z') r_{\mathcal{B},j}(\mu, Z) - (1 - p_j(R_{\mathcal{B},j}, z'_j, Z')) \lambda] \ell_{\theta,j}^i - c_\theta(\ell_{\theta,j}^i, Z), \quad (4)$$

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

²⁵Such an assumption is consistent with the endogenous results in [Rios-Rull, Takamura, and Terajima \(2023\)](#).

²⁶A parameterization where $r_A^+ > r_A^-(0)$ would lead to qualitatively similar results. Specifically, banks would borrow in the security market until reaching the no-arbitrage opportunity condition where $r_A^+ = r_A^-(b_\theta)$ for some $b_\theta < 0$. Then any $a_\theta^i < 0$ choice incurs cost $r_A^-(b_\theta + a_\theta^i)$.

²⁷Since projects are indivisible and the borrower outside option ω is unobservable, there is not heterogeneity in rates $r_{\mathcal{B},j}$ in any given region j .

Cournot equilibrium interest rate on bank loans in region j , and $c_\theta(\ell_{\theta,j}^i, Z)$ is the marginal cost of extending $\ell_{\theta,j}^i$ loans. Then, end of period profits associated with beginning-of-period deposits d_θ^i for an incumbent bank i of type $\theta \in \{r, \ell\}$ equals

$$\pi_\theta(s_\theta, S') = \pi_{\theta,j}^\ell(s_\theta, S') + [r_A^+ \mathbf{1}_{\{a_\theta^i \geq 0\}} + r_A^- \mathbf{1}_{\{a_\theta^i < 0\}}] a_\theta^i - r_D d_\theta^i - \kappa_\theta, \quad (5)$$

where κ_θ is the fixed operating cost in Table 2. Profits for an incumbent bank of type $\theta = n$ can be defined similarly by summing over loan revenues across regions (i.e., $\sum_j \pi_{n,j}^\ell(s_n, S')$).

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 but can issue equity in order to fund negative profits. 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.

We consider a baseline with the possibility of a government bailout when banks experience negative earnings and may fail. Berndt, Duffie, and Zhu (2025) estimate that the probability of a large bank bailout is not 1. Hence we assume that with probability $p^b \in [0, 1]$, the government extends bank bailouts in states where profits are negative. We assume bailouts are iid over time. If the government bails out, it covers a fraction $\varphi_\theta \in (0, 1]$ of negative profits (i.e., when $\varphi_\theta = 1$ ($\varphi_\theta = 0$) there is a full (no) bailout for bank type θ).²⁸ These losses are paid for by taxes as in the case of deposit insurance. We define $\iota^b = 1$ if the government extends bailouts and $\iota^b = 0$ if it does not. Dividends net of equity injections for bank i of type $\theta \in \{n, r, \ell\}$ in state (s_θ, S', ι^b) are given by

$$\mathcal{D}_\theta(s_\theta, S', \iota^b) = \pi_\theta(s_\theta, S')(1 - \varphi_\theta \iota^b) - \mathbf{1}_{\{\pi_\theta(s_\theta, S')(1 - \varphi_\theta \iota^b) < 0\}} \xi_\theta(|\pi_\theta(s_\theta, S')(1 - \varphi_\theta \iota^b)|). \quad (6)$$

A bank that has negative expected continuation value can exit, in which case it receives value zero. We denote the exit decision $x_\theta(s_\theta^i, S', \iota^b)$. If a national bank exits, it must 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 $\Upsilon_n \geq \Upsilon_r \geq \Upsilon_\ell = 0$.²⁹ 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_ℓ^1 exceeds the highest cost of fringe incumbents

²⁸In this specification, we can control what type of bank receives bailouts through φ_θ . For example, the case where only national banks receive bailouts is one where $\{\varphi_n, \varphi_r, \varphi_\ell\} = \{1, 0, 0\}$. The alternative where p^b is also a function of bank type θ would require us to add another state variable (ι_θ^b) as we would need to track which bank receives the bailout.

²⁹As 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.

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.³⁰ The entry decision is taken after the realization of the aggregate shock (and c_ℓ^1 in the case of fringe banks) but before learning their deposit funding base d_θ . Initial deposits are drawn from the stationary distribution $G_\theta^*(d_\theta)$ associated with $G_\theta(d'_\theta, d_\theta)$.

We will assume that all banks in a given state $(\theta, d_\theta) \in \{\{\boldsymbol{n}, \boldsymbol{r}\} \times D_\theta\}$ and $(\ell, d_\ell, c_\ell^1) \in \{D_\ell \times C_\ell^1\}$ are treated symmetrically, so the cross-sectional distribution $\boldsymbol{\mu}$ in any period 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

$$\boldsymbol{\mu}' = \mathcal{T}^*(\boldsymbol{\mu}, M^e; S', \iota^b), \quad (7)$$

where $M^e = \{M_\theta^e\}$ 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 (30).

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_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_N . Like banks, the representative non-bank can diversify entrepreneurs' idiosyncratic risk but it is subject to regional and aggregate fluctuations.

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

$$\pi_N(S') = \sum_j \{ [p_j(R_N, z'_j, Z') r_{N,j}(\boldsymbol{\mu}, Z) - (1 - p_j(R_N, z'_j, Z'))\lambda] - c_N \} \ell_{N,j} \quad (8)$$

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

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.

³⁰In any given period, there are M_ℓ 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'Erasmus \(2021\)](#).

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 observable only after incurring the monitoring cost in $c_\theta(\ell_{\theta,j}^i)$. 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 $\theta \in \{r, \ell\}$ in a given state (s_θ, S') symmetrically, post-liquidation cost to the deposit insurance fund from those banks in region j is given by

$$\begin{aligned} \Delta_\theta(s_\theta, S') &= (1 + r_D)d_\theta + (1 + r_A^-) \max\{(\ell_{\theta,j} - d_\theta), 0\} - \zeta_\theta(1 + r_A^+) \max\{(d_\theta - \ell_{\theta,j}), 0\} \\ &\quad - \zeta_\theta[p \cdot (1 + r_{\mathcal{B},j}(Z, \mu)) + (1 - p)(1 - \lambda)]\ell_{\theta,j} - c_\theta(\ell_{\theta,j}, Z) - \kappa_\theta, \end{aligned} \quad (9)$$

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 (9) by summing over the recovery of the loan portfolio in both regions. The government also collects taxes (or pays transfers) to cover the cost (or receive the proceeds) from the security market. Then, aggregate taxes are given by

$$\tau(S') \cdot H = \int_{\theta, s_\theta} x_\theta(s_\theta, S') \max\{0, \Delta_\theta(s_\theta, S')\} d\mu_\theta(s_\theta) + A^+ r_A^+ - A^- r_A^-, \quad (10)$$

where $A^+ = \int_{\theta, s_\theta} \max\{0, a_\theta\} d\mu_\theta(s_\theta)$ and $A^- = \int_{\theta, s_\theta} \max\{0, -a_\theta\} d\mu_\theta(s_\theta)$.

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 μ).
 - (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 ϵ 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' , z'_j , and ι^b 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 $\pi_\theta(s_\theta, S', \iota^b)$ 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 and bank failures as well as government bailouts, 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

$$\max_{\{\iota\}} (1 - \iota) \cdot \omega + \iota \cdot E_{\boldsymbol{\epsilon}}[\Pi_E(Z, \mathbf{r}_j, \boldsymbol{\epsilon})], \quad (11)$$

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

$$\begin{aligned} \Pi_E(Z, \mathbf{r}_j, \boldsymbol{\epsilon}) = \max_{\{k, R_{k,j}\}} & \left\{ \mathbf{1}_{\{k=\mathcal{B}\}} E_{z'_j, Z'|Z} [\pi_E(r_{\mathcal{B},j}, R_{\mathcal{B},j}, z'_j, Z') + \epsilon_{\mathcal{B}}] \right. \\ & \left. + \mathbf{1}_{\{k=\mathcal{N}\}} E_{z'_j, Z'|Z} [\pi_E(r_{\mathcal{N},j}, R_{\mathcal{N},j}, z'_j, Z') + \epsilon_{\mathcal{N}}] \right\}, \end{aligned} \quad (12)$$

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 (12) implies that the share of borrowers choosing a loan from a lender of type k in region j is

$$\psi_{k,j}(Z, \mathbf{r}_j) = \frac{\exp\left(\alpha E_{z'_j, Z'|Z} [\pi_E(r_{k,j}, R_{k,j}, z'_j, Z')]\right)}{\sum_{\hat{k} \in \{\mathcal{B}, \mathcal{N}\}} \exp\left(\alpha E_{z'_j, Z'|Z} [\pi_E(r_{\hat{k},j}, R_{\hat{k},j}, z'_j, Z')]\right)}. \quad (13)$$

The expected value of taking out a loan in region j is³¹

$$V_{E,j}(Z, \mathbf{r}_j) = \int \Pi_E(Z, \mathbf{r}_j, \epsilon) dF_\epsilon(\epsilon; \alpha). \quad (14)$$

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

$$\frac{\partial E_{z'_j, Z'|Z} [\pi_E(r_{k,j}, R_{k,j}, z'_j, Z')]}{\partial r_{k,j}} = -E_{z'_j, Z'|Z} [p_j(R_{k,j}, Z', z'_j)] < 0. \quad (15)$$

Thus, participating borrowers (i.e. those who choose to run a project rather than take the outside option) are worse off the higher the interest rate is 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

$$L_j^d(Z, \mathbf{r}_j) = \int_0^{\bar{\omega}} \mathbf{1}_{\{\omega \leq V_{E,j}(Z, \mathbf{r}_j)\}} d\Omega(\omega). \quad (16)$$

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

$$L_{\mathcal{B},j}^d(Z, \mathbf{r}_j) = \psi_{\mathcal{B},j}(Z, \mathbf{r}_j) L_j^d(Z, \mathbf{r}_j). \quad (17)$$

In that case, everything else equal, (15) implies $\frac{\partial L_{\mathcal{B},j}^d(\mathbf{r}_j, Z)}{\partial r_{\mathcal{B},j}} < 0$. 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. $\frac{\partial L_j^d(\mathbf{r}_j, Z)}{\partial r_{\mathcal{B},j}} \leq 0$).

4.2 Saver Problem

The problem of a representative household in any given region j is

$$V_H(\{\Phi_{s_\theta}\}_{\forall \theta}, \Phi_N) = \max_{\{A^h, D, \{\Phi'_{s_\theta}\}_{\forall \theta}, \Phi'_N\}} E [C' + \beta V_H(\{\Phi'_{s_\theta}\}_{\forall \theta}, \Phi'_N)] \quad (18)$$

subject to

$$\begin{aligned} C' + A^h + D + \int_{\theta, s_\theta} [P_{s_\theta} + \mathbf{1}_{\{e_\theta=1\}} \cdot \Upsilon_\theta] \Phi'_{s_\theta} d\mu_\theta(s_\theta) + \Phi'_N P_N \\ = y + (1 + \bar{r})A^h + (1 + r_D)D + \int_{\theta, s_\theta} (\mathcal{D}_{s_\theta} + P_{s_\theta}) \Phi_{s_\theta} d\mu_\theta(s_\theta) + (\mathcal{D}_N + P_N) \Phi_N - \tau(S'), \end{aligned} \quad (19)$$

and $C' \geq 0$ where P_{s_θ} and Φ'_{s_θ} are the post-dividend stock price and stock holding of bank of type θ in region j , respectively, and P_N and Φ'_N are the price of a claim to non-bank

³¹The expected value of taking out a loan has a convenient closed form: $V_{E,j}(Z, \mathbf{r}_j) = \frac{\gamma_E}{\alpha} + \frac{1}{\alpha} \ln \left(\sum_k \exp \left(\alpha E_{z'_j, Z'|Z} [\pi_E(r_{k,j}, R_{k,j}, z'_j, Z')] \right) \right)$, where γ_E is Euler's constant.

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 (19), 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.³² 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, $r_D = \bar{r}$. Differentiating π_θ with respect to $\ell_{\theta,j}$ (for expositional purposes for cases when $\ell_{\theta,j} \leq d_\theta$)

$$\frac{d\pi_\theta}{d\ell_{\theta,j}} = \underbrace{\left[p_j r_{\mathfrak{B},j} - (1 - p_j)\lambda - \frac{dc_\theta(\cdot)}{d\ell_{\theta,j}} - \bar{r} \right]}_{(+)\text{ or }(-)} + \ell_{\theta,j} \left[\underbrace{\left[\underbrace{p_j}_{(+)} + \underbrace{\frac{\partial p_j}{\partial R_j} \frac{\partial R_j}{\partial r_{\mathfrak{B},j}} (r_{\mathfrak{B},j} + \lambda)}_{(-)} \right]}_{(-)\text{ for dominant banks}} \right] \underbrace{\frac{dr_{\mathfrak{B},j}}{d\ell_{\theta,j}}}_{(-)\text{ for dominant banks}}. \quad (20)$$

The first bracket represents the marginal change in profits from extending an extra unit of loans.³³ 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 $\frac{dr_{\mathfrak{B},j}}{d\ell_{\theta,j}} < 0$ while for fringe banks $\frac{dr_{\mathfrak{B},j}}{d\ell_{\theta,j}} = 0$. Note that a change in interest rates also endogenously affects the fraction of delinquent loans faced by banks (the term $\frac{\partial p_j}{\partial R_j} \frac{\partial R_j}{\partial r_{\mathfrak{B},j}} < 0$). 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 by³⁴

$$V_\theta(s_\theta, Z, \boldsymbol{\mu}) = \max_{\{\ell_{\theta,j} \geq 0, a_\theta\}} \beta E_{S', \ell^b | Z} [W_\theta(s_\theta, S', \ell^b)] \quad (21)$$

³²In [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 $R_{\mathfrak{B}}$, 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.

³³In cases where $\ell_{\theta,j} > d_\theta$ (i.e., $a_\theta < 0$) the marginal cost of funds $\bar{r}$ changes to r_A^- (which is the cost of external funding for the bank at the lending stage when loans exceed deposits).

³⁴While each bank takes all other banks' strategies as given, for notational simplicity we abstract from writing that in the problem.

subject to bank balance sheet constraints

$$\begin{aligned} \ell_{\theta,j} + a_{\theta} &= d_{\theta} & \text{if } \theta \in \{\mathbf{r}, \mathbf{\ell}\}, \\ \sum_j \ell_{\theta,j} + a_{\theta} &= d_{\theta} & \text{if } \theta = \mathbf{n} \end{aligned} \quad (22)$$

and market clearing in each region³⁵

$$L_{\mathcal{B},j}^s(Z, \boldsymbol{\mu}) \equiv \int_{\theta, s_{\theta}} \ell_{\theta,j}(s_{\theta}, Z, \boldsymbol{\mu}) d\mu_{\theta}(s_{\theta}) - L_{\mathcal{B},j}^d(Z, \mathbf{r}_j) = 0, \forall j, \quad (23)$$

where $L_{\mathcal{B},j}^d(Z, \mathbf{r}_j)$ is given in (17) and the exit decision $x_{\theta}(s_{\theta}, S')$ derives from

$$W_{\theta}(s_{\theta}, S', \iota^b) = \max_{\{x_{\theta} \in \{0,1\}\}} \{W_{\theta}^{x=0}(s_{\theta}, S', \iota^b), W_{\theta}^{x=1}(s_{\theta}, S')\} \quad (24)$$

with

$$W_{\theta}^{x=0}(s_{\theta}, S', \iota^b) = \mathcal{D}(d_{\theta}, s', \iota^b) + E_{s'_{\theta}, \boldsymbol{\mu}'}[V_{\theta}(s'_{\theta}, Z', \boldsymbol{\mu}')], \quad (25)$$

$$\mathcal{D}(s_{\theta}, S', \iota^b) = \begin{cases} \pi_{\theta}(s_{\theta}, S') & \text{if } \pi_{\theta}(s_{\theta}, S') \geq 0 \\ \pi_{\theta}(s_{\theta}, S')(1 - \iota^b \varphi_{\theta}) - \xi_{\theta}(\pi_{\theta}(s_{\theta}, S')(1 - \iota^b \varphi_{\theta})) & \text{if } \pi_{\theta}(s_{\theta}, S') < 0 \end{cases}, \quad (26)$$

$$\begin{aligned} W_{\theta}^{x=1}(s_{\theta}, S') = \max \Big\{ & 0, \zeta_{\theta} \pi_{\theta}(s_{\theta}, S') + (1 - \zeta_{\theta})[(1 + r_A^-) \max\{\ell_{\theta} - d_{\theta}, 0\} \\ & - (1 + r_D) d_{\theta} - c_{\theta}(\ell_{\theta}) - \kappa_{\theta}] \Big\}. \end{aligned} \quad (27)$$

Equations (25)-(27) make clear that government interventions are triggered when a bank is in trouble in a static sense as the bailout decision only takes into account whether the bank currently has negative profits but does not incorporate the charter value of the bank. Of course if equity issuance costs are sufficiently high in negative earnings states, then an equity injection from the government is necessary to avoid bank failure. Our assumption is consistent with the determinants of the “FDIC Problem Bank List”, which includes historical book measures such as earnings and liquidity indicators.³⁶

4.3.1 Bank Entry

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

$$V_{\theta}^e(S') = -\Upsilon_{\theta} + E_{s'_{\theta}}[V_{\theta}(s'_{\theta}, Z', \boldsymbol{\mu}')]. \quad (28)$$

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

$$V_{\theta}^e(S') \times M_{\theta}^e = 0. \quad (29)$$

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

³⁵The national bank faces the clearing condition in each region.

³⁶Incorporating a dynamic evaluation (or market based accounting) into the decision of the government to extend the bailout or not complicates the solution of the model significantly as the bailout would be extended conditional on $W_{\theta}^{x=1}(s_{\theta}, S') > W_{\theta}^{x=0}(s_{\theta}, S', \iota^b = 0)$ (i.e., the bank would have chosen to exit without the bailout), so the bailout policy creates an additional fixed point problem. See [Begenau et al. \(2026\)](#) for a model where policy also considers market value not just historical book values.

4.3.2 Industry Law of Motion

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

$$\mu'_\theta(s'_\theta) = \int_{s_\theta} (1 - x_\theta(s_\theta, S', \iota^b)) G_\theta(d'_\theta, d_\theta) d\mu_\theta(s_\theta) + M_\theta^e G_\theta^*(d'_\theta). \quad (30)$$

Equation (30) makes clear how the law of motion for the distribution of banks is affected by entry and exit decisions as well as the exogenous shocks to the economy. The loan and security decision rules together with the realization of $\{z'_j, Z', \iota^b\}$ determine bank profitability and $x_\theta(s_\theta, S', \iota^b)$ which then together with the realization of deposits (for incumbents and entrants) determine the updated distribution. In the case of banks of type $\theta = \ell$, the second term in the right hand side of (30) incorporates the draws from the distribution of c_ℓ^1 and the last term becomes $M_\ell^e G_\ell^*(d'_\ell) \Xi(c_\ell^1)$.

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 $\beta(1 + \bar{r}) = 1$, the first order condition of the non-bank with respect to $\ell_{\mathcal{N},j}$ is given by

$$\bar{r} = E_{Z', z'_j, z'_{-j} | Z} [p(R_{\mathcal{N}}(r_{\mathcal{N},j}, Z), z'_j, Z') - (1 - p(R_{\mathcal{N}}(r_{\mathcal{N},j}, Z), z'_j, Z'))\lambda] - c_{\mathcal{N}}, \quad (31)$$

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 (31) 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 (31) 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 (11)-(12) inducing an aggregate loan demand function $L_j^d(Z, \mathbf{r}_j)$ in (16) and bank/nonbank market shares in (13).
2. $\{A', D', \Phi'_{s_\theta}, \Phi'_{\mathcal{N}}\}$ are consistent with household optimization in (18) 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 (21)-(24) conditional on the government bailout policy ι^b inducing an aggregate bank loan supply function $L_{\mathcal{B},j}^s$ defined in (23).
4. e_θ is such that free entry (29) is satisfied.
5. The law of motion for the industry state $\boldsymbol{\mu}' = \mathcal{T}^*(\boldsymbol{\mu}, M^e; S', \iota^b)$ in equation (30) 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 (31).
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 $\tau_D'(S')$ in (10) cover the cost of failing banks.

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”).³⁷ 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), Total Factor Productivity (TFP) series for the U.S. Business Sector, produced by Fernald (2014), 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 $\theta \in \{\mathbf{n}, \mathbf{r}, \ell\}$. Consistent with the data size differences we described in Section 2, we identify National banks $\theta = \mathbf{n}$ with those in the top 4 (when banks are sorted by assets), Regional banks $\theta = \mathbf{r}$ with those in the top 5-35, and the rest (i.e., the competitive fringe $\theta = \ell$) with those outside the top 35.

We parameterize the stochastic process for the borrower’s project as follows. For each borrower, let $y_j^e = aZ' - bR + \varepsilon_e$, where ε_e is iid (across agents and time) and drawn from

³⁷We obtain the data from the internal Federal Reserve server. The public version of the data can be found here: <https://cdr.ffiec.gov/public/PWS/DownloadBulkData.aspx>. See Appendix A-1 for a description of the data.

$N(z'_j, \sigma_\varepsilon^2)$. We define success to be the event that $y^e > 0$, so in states with higher ε_e success is more likely. Then

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

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

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 \(1986\)](#) 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 1. We calibrate the cost of deposits using the ratio of interest expenses on deposits to total deposits. We let $r_A^-(a_\theta) = r_{A,1}^-|a_\theta| + r_{A,2}^-|a_\theta|^2$ for $a_\theta < 0$. We set $r_A^+ = r_{A,1}^+$ and calibrate r_A^+ to match the average cost of net securities which equals the sum of cash, federal funds purchased and securities minus federal funds bought. We calibrate $r_{A,2}^-$ below to match the deposit to asset ratio for regional banks (as it captures the amount of borrowing relative to assets). We assume $c_\theta(\ell_\theta, Z)$ is a quadratic function of ℓ_θ with linear term $c_\theta^1 Z$ and quadratic term $c_\theta^2 Z$.

We let the external seasoned equity financing cost take the following form $\xi_\theta(x) = \xi_\theta^1 x$ for $\theta \in \{n, r\}$ and assume that external financing is prohibitively costly for fringe banks $\theta = f$. 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

$$\hat{\pi}_f(s_f, Z, \mu) \equiv \max_{\{\ell_f \geq 0, a_f\}} E_{Z', s'_j | Z} [\max\{\pi_f(s_f, Z, \mu), W_f^{x=1}(s_f, S')\}] = 0. \quad (32)$$

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. We set $\varphi_n = 0.90$ to make it consistent with the evidence presented in [Atkeson et al. \(2019\)](#) that study the ratio of the value of government guarantees to book equity for the period 1996-2007 and set $\varphi_r = \varphi_f = 0$ as only national banks receive

bailouts in our baseline. We incorporate p^b (the probability that the government extends the bailout) as one of the parameters to be calibrated via the simulated method of moments.

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. Table A.6 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 29 parameters and 29 targeted moments as part of the simulated method of moments estimation. Table A.6 in the Appendix presents all the parameters of the model. Tables A.7 and A.8 also in the Appendix present a set of data moments (targeted and not targeted) together with their model generated counterparts.

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 5 provides the correlation between key aggregate variables with GDP.³⁸ The data estimates presented are derived from a regression of the detrended log variable on detrended log real GDP. We observe that, as in the data, the model generates countercyclical failure rates, default frequencies, charge-off rates, interest rates, interest spread and procyclical loan supply. The calibration of the model implies a procyclical markup in contrast with the countercyclical markup we observe in the data. The mismatch arises because the countercyclicity of interest rates is not strong enough to offset the strong procyclicity of project success (i.e., the countercyclicity of default rates) in the model.

³⁸When calculating business cycle correlations in the model, we correlate variables that depend on $\{Z, \mu\}$ and S' (which includes $\{Z, \mu\}$) 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 $r_{\mathcal{B}}(Z, \mu)$ and $GDP(S')$.

Table 5: Business Cycle Statistics

Variable correlated with GDP	Data	Model
Charge-off Rate	-0.230***	-1.704***
Default freq.	-0.521***	-0.563***
Loan return	0.052	1.469***
Loan Interest Rate	-0.213 **	-2.232***
Loan Interest Spread	-0.116 ***	-2.229***
Markup	-2.508*	0.332***
Failure Rate	-0.149***	-1.558***
Loan Supply	0.945**	0.461***

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

6.2 Measures of Bank Competition

Table 6 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 and Lerner index that are in line with the data. The markup implied by the model is somewhat higher than the data as the model estimated average costs for national and regional banks are lower than in the data (see Table A.8).

Table 6: Measures of Bank Competition

Moment (%)	Data	Model
Net Interest Margin	4.65	4.08
Lerner	46.23	63.80
Markup	102.26	176.24

6.3 Empirical Studies of Banking Crises, Default and Concentration

Many authors have empirically estimated the relation between bank concentration (as a measure of lack of 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 using model simulated data. 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 using model simulated data. 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 aggregate loan supply.³⁹ Table 7 displays the estimated coefficients and their standard errors.

Table 7: Banking Crises, Default Frequencies and Concentration

Dependent Variable	Crisis _t	Default Freq _t
Concentration _t	-1.717	0.029
s.e.	(0.0747)**	(0.001)***
Δ GDP _t	0.078	0.0136
s.e.	(2.089)	(0.030)
Δ Loan Supply _t	-1.152	-0.012
se	(2.006)	(0.029)***
R^2	0.565	0.086
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 crisis. 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 \(2004\)](#). 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 owing to selection effects arising from heightened loan interest rates.

³⁹[Beck, Demirgüç-Kunt, and Levine \(2006\)](#) also include other controls like “economic freedom,” which are outside of our model.

7 Bank Run Granularity

We study how idiosyncratic bank shocks to bank deposit flows spill over to the aggregate economy via changes in credit. The idea we pursue is what happens if dominant banks receive the worst deposit flow; that is, face a run. Since our model features granular banks, in the language of [Gabaix \(2011\)](#), we can use the simulation of the model to decompose how much of the changes in total credit, interest rates, and output can be explained by deposit shocks (i.e. runs) to national and regional banks.

[Gabaix \(2011\)](#) focused on how idiosyncratic shocks to the growth rate of the largest 100 firms can explain 30% of aggregate output fluctuations. Here, we focus on idiosyncratic shocks to the largest 35 banks which were the focus of stress tests in 2017 represented by the national and regional banks in our model. In this granular view, shocks to large banks have the potential to generate nontrivial aggregate fluctuations.

We start with an experiment to compare aggregate dynamics in a simulation of our baseline model (i.e. where all shocks are present) to a simulation where aggregate and regional shocks are kept constant at their mean level (i.e. to an economy driven only by shocks to granular dominant banks). The differential response captures how much shocks to the top 35 banks matter for aggregate fluctuations.

Table 8 presents the results. We find that dominant bank level shocks explain all variation in the bank loan supply ($L_{\mathcal{B}}^s$) and about 77% of the variation in bank intermediated output ($\text{gdp}_{\mathcal{B}}$). It is interesting to see that the bank shocks spill over to the nonbank sector. As the bank interest rate becomes less volatile (implied by the reduction in the volatility of the bank loan supply), the volatility of the loan supply of the nonbank sector also declines. As the table shows, dominant bank shocks explain 75% of the variation of the nonbank loan supply ($L_{\mathcal{N}}^s$) and this leads to dominant bank shocks explaining about 55% of the variation of nonbank intermediated output ($\text{gdp}_{\mathcal{N}}$). Interestingly, we observe that dominant bank shocks explain only 14% of the variation in the total loan supply and 2.2% of the variation total gdp. The effect on each of the components of the loan supply and output does not translate into a larger aggregate effect as the covariance between the loan supply of the bank sector and the nonbank sector remains almost unchanged and the covariance between output in the bank sector and output originated with loans from nonbanks increases (in absolute terms) significantly.

Table 8: Experiment No Aggregate or Regional Shocks (i.e., only bank shocks)

Variable (%)	Baseline All Shocks	Only Dominant Bank Shock	% Change	Explained by Bank Shocks %
Var(gdp)	0.0125	0.0003	-97.81	2.19
Var(gdp _ℬ)	0.0107	0.0082	-23.06	76.94
Var(gdp _{ℕℬ})	0.0100	0.0055	-45.02	54.98
Cov(gdp _ℬ , gdp _{ℕℬ})	-0.0041	-0.0067	64.50	164.50
Var(L ^s)	0.0015	0.0002	-86.08	13.92
Var(L _ℬ ^s)	0.0064	0.0064	-0.05	99.95
Var(L _ℕ ^s)	0.0057	0.0043	-24.74	75.26
Cov(L _ℬ ^s , L _ℕ ^s)	-0.0053	-0.0053	-1.03	98.97

Note: “Baseline All Shocks” corresponds to the simulation of the baseline model where all shocks (aggregate, regional, and bank level shocks) are present. “Only Dominant Bank Shock” refers to the simulation of the model keeping aggregate shocks and regional shocks at their mean (i.e., the only shocks driving the time series variation are dominant bank level shocks where dominant banks refers to national and regional banks). Var refers to variance and Cov to covariance. “gdp” refers to total output, “gdp_ℬ” bank intermediated output, and “gdp_ℕ” nonbank intermediated output.

Next we map our model directly to the approach in [Gabaix \(2011\)](#) to measure the contribution of dominant bank shocks to aggregate fluctuations. We define two measures of idiosyncratic fluctuations, one based on deposits (truly exogenous variation in the model) and another based on loans (endogenous in the model). In particular, let $\epsilon_{i,t}^d = d_{\theta,t} - d_{i,t-1}$ and $\epsilon_{i,t}^\ell = \ell_{i,t} - \ell_{\theta,t-1}$ be the unexplained change in deposits or loans for bank i in period t . As in [Gabaix \(2011\)](#), we construct the granular residual as follows:

$$\Gamma_t^j \equiv \sum_{i=1}^K \frac{\ell_{it}}{L_{\mathcal{B}}^s} \epsilon_{i,t}^j \quad (33)$$

for $j = \{d, \ell\}$ and $K = 35$ (the number of dominant banks in our model). The goal is to investigate how much of the growth in output can be explained by Γ_t^j for $j = \{d, \ell\}$. As in [Gabaix \(2011\)](#) we regress gdp, gdp_ℬ and gdp_ℕ on Γ_t^j and Γ_{t-1}^j to find support for the granular hypothesis. The key statistic is the R^2 of these regressions.

Table 9 presents the key results.⁴⁰ We find that idiosyncratic shocks explain between 34% and 45% of fluctuations in bank intermediated output, between 22% and 29% of fluctuations of nonbank intermediated output and approximately 1% of aggregate output. These results are consistent with those presented in Table 8 since we find that shocks to the dominant bank in both tables provide a consistent ordering of variation; larger fraction of the variation in bank intermediated output than nonbank intermediated output as well as variation in total output. In addition, while values in Table 8 and 9 derive from different measures, the final outcomes (i.e., the fraction of variation explained by shocks to dominant banks) are also on the same order of magnitude.

⁴⁰Table A.9 in the Appendix presents all the estimated coefficients.

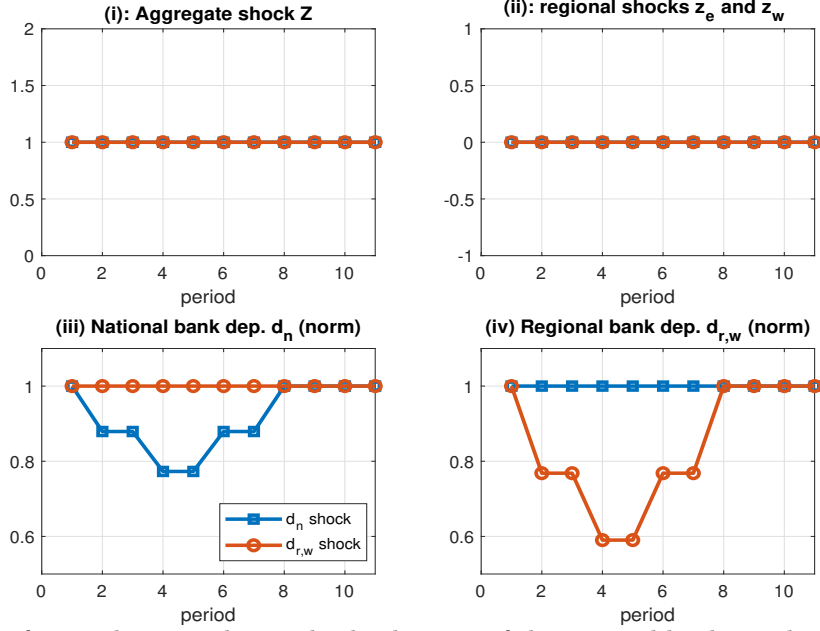
Table 9: Explanatory Power of Granular Residuals of Dominant Banks (R^2)

Independent Variable	Dependent Variable		
	$\Delta \text{ gdp}$	$\Delta \text{ gdp}_{\mathcal{B}}$	$\Delta \text{ gdp}_{\mathcal{N}}$
Deposits Γ_t^d	0.009	0.338	0.217
Loans Γ_t^ℓ	0.011	0.454	0.289

Note: Table presents the R^2 from a regression of the corresponding dependent variable on Γ_t^j and Γ_{t-1}^j . “ $\Delta \text{ gdp}$ ” refers to growth rate of total output, “ $\Delta \text{ gdp}_{\mathcal{B}}$ ” to the growth rate of bank intermediated output, and “ $\Delta \text{ gdp}_{\mathcal{N}}$ ” to the growth rate of nonbank intermediated output.

In order to better understand the economics behind the previous results, we next study separately how a decline in deposits (say due to an exogenous run) at national banks versus regional banks affects aggregate dynamics. More specifically, we perform two experiments. In our first experiment, we consider the case where all aggregate shocks, regional shocks, and regional bank deposits remain constant at their mean and let the deposit shocks of the national bank d_n gradually decline to the lowest value of the finite support of deposit shocks for one period and recover (i.e. a run on all national banks). In our second experiment, we set aggregate and regional shocks as in the first experiment but this time we set the deposits of the national bank to their mean and let the deposit shocks of the regional bank in the west $d_{r,w}$ gradually decline to the lowest value of the finite support of deposit shocks for one period and recover (i.e. a run on all regional banks). These two experiments are meant to resemble, in the same way that [Gabaix \(2011\)](#) considered subsets of K banks, bank run episodes for national and regional banks. The regional bank shock is much like the recent collapse of SVB and Signature Bank that quickly evolved into a panic at banks that shared their asset structure and size ([Acharya, Cecchetti, and Schoenholtz \(2023\)](#) document the large deposit outflows from midsize banks). [Figure 8](#) presents the evolution of the shocks in these two experiments. Deposits of national and regional banks are normalized to their mean value. We note that as the shocks of the regional banks are more volatile than those of their national bank counterparts, the lowest value of the regional shock is significantly lower (when normalized).

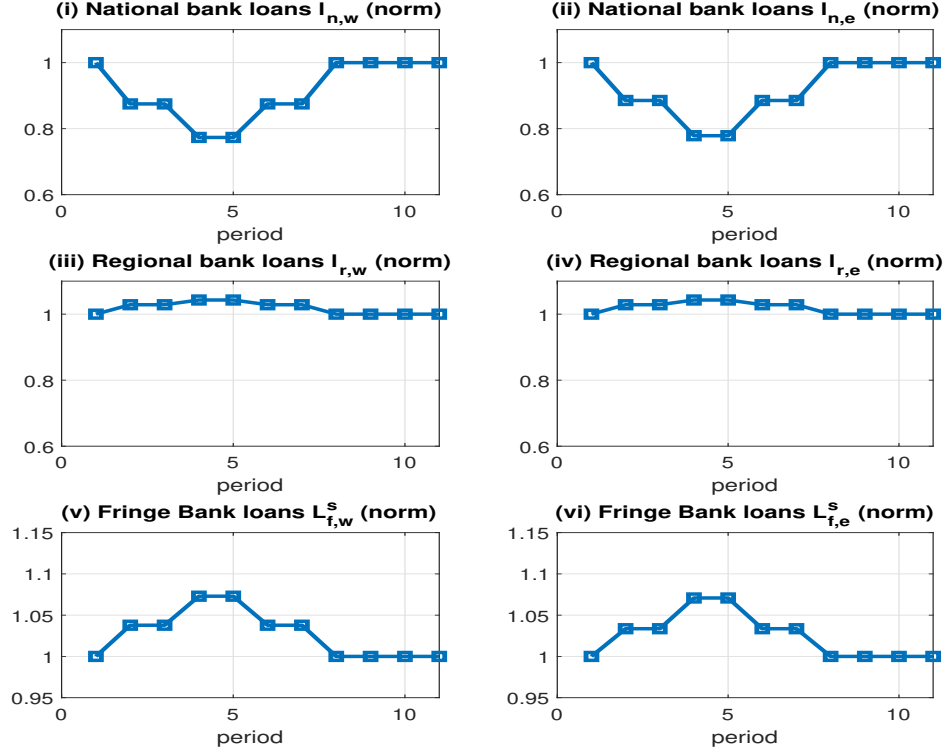
Figure 8: Aggregate Spillovers: National and Regional Bank Shocks



Note: “ d_n shock” refers to the case where only the deposits of the national bank are shocked. “ $d_{r,w}$ shock” refers to the case with a transitory decline only in regional bank deposits in the west. “(norm)” denotes normalized and correspond to a panel where the variables shown is normalized to make the starting value equal to 1.

Figures 9 and 10 present lending by the national bank in each region (Panels (i) and (ii)), total lending by regional banks in the west and the east (Panels (iii) and (iv)), and total lending by fringe banks in the west and the east (Panels (v) and (vi)) in the case of national bank shocks and in the case of regional bank shocks, respectively. In the case of national bank shocks, Figure 9 shows that as deposits in the national bank decline, total national bank lending also declines almost one to one (as national banks are reluctant to borrow at the more costly rate $r_A^- > r_D$ in the fed fund market to keep lending at the same level). This decline in national bank lending is compensated partially by an increase in lending by regional banks and fringe banks. The increase in fringe banks derives in some part from the extensive margin as more fringe banks enter the industry.

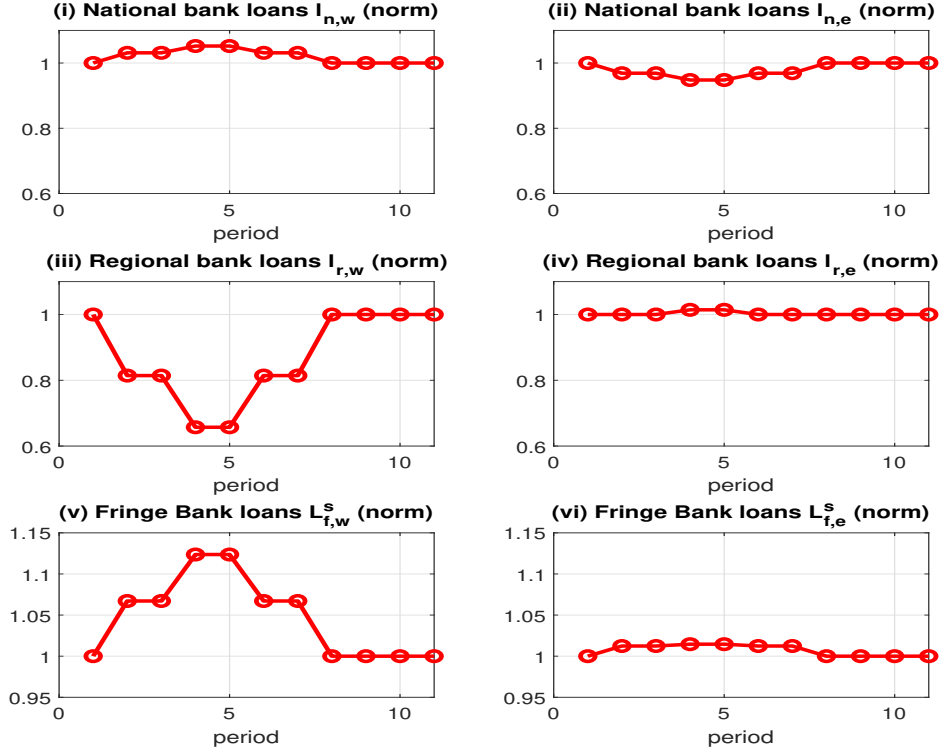
Figure 9: National Bank Shocks - Lending across bank sizes



Note: “(norm)” denotes normalized and corresponds to a panel where the variables shown are normalized to make the starting value equal to 1.

In Figure 10 we observe that as deposits in the regional bank in the west decline, lending by regional banks in the west region declines (similar to the national bank, regional banks are not willing to borrow a rate $r_A^- > r_D$ to keep lending at the level observed prior to the shock). This decline in regional bank lending in the west is compensated partially by an increase in lending by the national bank in that region as it shifts lending from the east to the west since the west now offers more profitable lending opportunities (Panels (i) and (ii)). Fringe banks also increase their lending in the west as a result of the decline in regional bank lending. In addition, the reduction in lending by the national bank in the east induces a mild increase in lending by regional and fringe banks in that region (Panels (iv) and (vi)).

Figure 10: Regional Bank Shocks - Lending across bank sizes



Note: “(norm)” denotes normalized and correspond to a panel where the variables shown is normalized to make the starting value equal to 1.

Figure 11 compares the evolution of aggregate variables under the two experiments. In particular, this figure shows the evolution of bank interest rates (Panel (i) and (ii)), the bank loan supply in each region (Panels (iii) and (iv)), the default probability in each region (Panels (v) and (vi)), the total bank loan supply (Panel (vii)) and GDP (Panel (viii)) for the case of a temporary national bank shock (denoted d_n shock) and for the case of a temporary regional shock (denoted d_r shock). Figure 11 shows that the increase of lending by competitors not affected by the shocks is not sufficient to compensate for the decline in lending driven by the decline in deposits as the total loan supply in both regions declines under both experiments (Panels (iii) and (iv)). The total decline in lending in the west is proportionally larger in the case of regional bank shocks as the more volatile deposits makes for a relatively larger deposit outflow when taking deposits to the lowest value in their support. The decline in total bank lending in the east is larger in the case of the national bank shock as the national bank reduces lending proportionally in each region when faced with a decline in deposits. The observed decline in lending in each region results in an increase in interest rates and default frequencies induced by higher borrower risk-taking (Panels (i)-(ii) and (iv)-(v)).

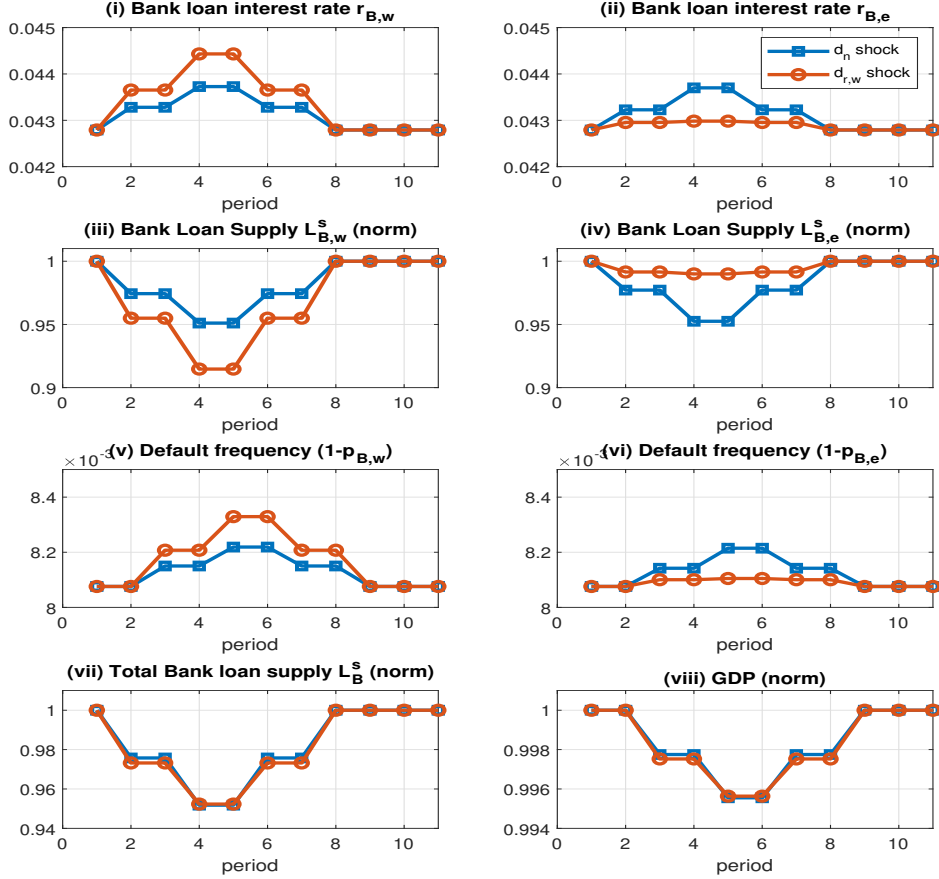
Since the loan market shares of national and regional banks are similar (37% in our

baseline), the overall response in total bank lending across the two experiments does not differ much (there is a slightly larger decline (0.1%) at the bottom of the cycle in the case of national banks (panel (vii))) in Figure 11. This similar behavior of total bank lending leads to an almost indistinguishable decline in GDP (of about one half percent) across the two experiments (panel (viii)). It is not surprising that a shock to *all* national banks has the same effect as a shock to *all* regional banks given the similarity of market shares between the two types.

An alternative to shocking a set of banks (e.g. the top 100 banks) as in Gabaix (2011) would be to run an experiment shocking only *one* national bank or *one* regional bank in each region. Such an experiment shocking deposits to a particular national bank (1 out of the 4 that are included in this group) or a particular regional bank (1 out of the 15 or 16 that are included in this group in each region) would potentially yield more variation in the aggregate effects of panels (vii) and (viii) since the market share of one national bank is substantially larger than the market share of one regional bank.⁴¹

⁴¹Implementing an experiment where only one bank receives the shocks requires imposing some assumptions about bank behavior. More specifically, take the case of a shock to a particular national bank. The response of the representative national bank and the representative regional bank can be scaled by one fourth of the observed response (working with a linear approximation). This calculation is missing the response of other national banks to the one affected. That is, it assumes that other national banks do not change their lending when confronted with a decline in lending by a bank of similar size and scope. This approach would be similar to the solution method proposed in Weintraub, Benkard, and Van Roy (2008) along with assuming that changes in the national bank deposits have no effect on the industry state (i.e., the case of many firms).

Figure 11: Aggregate Spillovers: National and Regional Bank Shocks - Aggregates



Note: “ d_n shock” refers to the case where only the deposits of the national bank are shocked. “ $d_{r,w}$ shock” refers to the case with a transitory decline only in regional bank deposits in the west. “(norm)” denotes normalized and correspond to a panel where the variables shown is normalized to make the starting value equal to 1.

8 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 counterfactuals like that described in the Introduction by Ben Bernanke.

In our baseline economy, we consider the possibility of national bank bailouts in states where national banks experience negative profits which, without a sufficiently high seasoned equity injection, would lead to bank failure. A policy of big bank bailouts or “too-big-to-fail” (TBTF) implies that with some probability (95% in our calibrated economy) 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 baseline economy with one where there are no government bailouts to national banks (i.e., we set $p^b = 0$ and $\varphi_n = 0$). As before, the bank will optimally decide to stay in operation or to exit after observing the realization of profits. Since the charter value of a bank is always non-negative due to limited liability, as φ_n goes to 0 (i.e. the government covers a smaller fraction of the loss), the probability of the bank continuing in operation decreases. The problem of a national bank ($\theta = n$) solves (21)-(27) with (26) now evaluated at $\iota^b = 0$.

In Table 10, we present the comparison of our calibrated baseline where $p^b = 0.95$ and $\varphi_n = 0.90$ to the case with no bailouts where $p^b = 0$ and $\varphi_n = 0$.

Table 10: Baseline vs. No Bailouts

	Baseline National Bank Bailout	No Bailouts Change (%)
	(1)	(2)
Bank Loan Supply	0.24	-0.75
Total Loan Supply	0.48	-0.07
Share Bank Loans (%)	49.41	-0.68
Markup (%)	176.24	0.98
Loan Market Share Fringe (%)	25.36	1.63
Loan Market Share Regional (%)	37.18	1.46
Loan Market Share National (%)	37.46	-2.55
Borrower Risk Taking	13.20	0.003
Default Frequency (%)	2.23	0.19
Net Interest Margin (%)	4.08	0.33
Bank Loan Interest Rate (%)	4.32	0.32
Failure Rate (%)	0.95	-0.35
Prob. $\pi_r < 0$	7.36	-0.04
Prob. $\pi_n < 0$	5.69	-2.59
Output	0.53	-0.07
Taxes Bank Failures	0.0002	-0.43
Taxes	0.0003	-27.10
Taxes/Output Bank Failures (%)	0.044	1.72
Taxes/Output (%)	0.062	-28.22

Note: Change (%) corresponds to the percentage change between the corresponding model and the Baseline with National Bank Bailouts.

In our benchmark equilibrium with national bank bailouts there was no national bank

failure on-the-equilibrium path but we observe a 5.7% probability of reaching states with negative profits for the national bank. We find that after removing national bank bailouts, national banks reduce their loan exposure in order to maintain their charter value, which leads to a decline of 2.6% in the probability of a national bank reaching a state with negative profits. There is a significant decline in taxes to output (-28%) as the government eliminates the bailout policy.

As evident in Table 10, eliminating TBTF induces the national bank to reduce its loan supply and a portion of the market is taken over by regional and fringe banks (their market share increases by 1.5% and 1.6%, respectively). The increase in the loan supply of smaller banks is not enough to compensate for the decline in lending by the national bank resulting in a decline in total bank lending (-0.75%) which in turn leads to higher interest rates (+0.32%) and default frequency (+0.19%). The higher default frequency is compensated by higher interest rates resulting in higher profits along the equilibrium path (markups increase by 0.98%) which leads to a decline in failure rates (-0.35%). The decline in lending induces a reduction in output by 0.07% but the reduction in failures does not translate into a reduction in the ratio of taxes for failure to output.

To assess the welfare consequences of changes in bailout policies, we ask the question, “What would households and entrepreneurs be willing to pay (or be paid) to remove national bank subsidies?” To answer this question, we calculate consumption equivalents for each type of representative agent household (H) and entrepreneur (E). Specifically, let $C_{h,t}^{pre}$ denote type $h \in \{H, E\}$ equilibrium aggregate consumption in the baseline economy with TBTF and $C_{h,t}^{post}$ denote type h equilibrium aggregate consumption post-reform without TBTF. Then, assuming the consumption flow of the entrepreneurs is discounted at the same rate as the households, the consumption equivalent α_h for a representative type h agent as a result of the reforms is defined as the constant percentage increase in consumption in the pre-reform case that gives the agent the same expected utility in the post-reform economy. Specifically, α_h solves:

$$E_0 \left[\sum_{t=0}^{\infty} \beta_h^t C_{h,t}^{post} \right] = E_0 \left[\sum_{t=0}^{\infty} \beta_h^t (1 + \alpha_h) C_{h,t}^{pre} \right]. \quad (34)$$

We let the ex-ante value (i.e., before being born as $h \in \{H, E\}$) for an agent in the pre-reform economy be

$$W_{pre} = \frac{\mathcal{H}}{1 + \mathcal{H}} E_0 \left[\sum_{t=0}^{\infty} \beta_H^t C_{H,t}^{pre} \right] + \frac{1}{1 + \mathcal{H}} E_0 \left[\sum_{t=0}^{\infty} \beta_E^t C_{E,t}^{pre} \right], \quad (35)$$

where $\mathcal{H}$ denotes the mass of households in the economy.⁴² We define the ex-ante value post-reform similarly. Then, we can write the ex-ante value of the reform in consumption equivalent terms $\bar{\alpha}$ as $W_{post} = (1 + \bar{\alpha})W_{pre}$. Equations (34) and (35) make clear that, since preferences are linear, computing ex-ante consumption equivalents in the long-run amounts to computing the difference between expected consumption pre- and post-reform.

Column (2) in Table 11 shows that both households and entrepreneurs would be willing to pay in order to remain in the economy with national bank bailouts as $\alpha_E = -0.045\%$

⁴²Since there is a mass $\mathcal{H}$ of households and a unit mass of entrepreneurs the probability of being born as a household is $\frac{\mathcal{H}}{1+\mathcal{H}}$.

and $\alpha_H = -0.002\%$. In the case of entrepreneurs, the decline in welfare (measured in consumption equivalents) derives from the decline in lending. In the case of the households, while in the baseline the households face taxes to cover deposit insurance, after removing the bailout policy, as the owners of the banks, households are required to still pay for those negative profits plus the cost of issuing equity so the change in taxes almost cancels out. We calibrate $\mathcal{H} = 4.88$ as in Corbae and D’Erasmus (2021) so the fraction of entrepreneurs equals 17%. This leads to an aggregate welfare loss (in consumption equivalents) of 0.006%.

Since households and entrepreneurs are risk neutral, this normative result abstracts from changes in aggregate second moments. Table 11 also includes a calculation of the variance of GDP and consumption in the absence of TBTF. Without TBTF, banks internalize that risky lending raises the likelihood of failure and loss of charter value. Concern about the loss of charter value induces banks to take less risk which leads to a 3% lower variance of lending in Table 12. A lower variance of lending induces a drop in the variance of GDP without TBTF evident in Table 11. While the variance of GDP drops, the variance of aggregate consumption actually rises. As expected, the drop in variance of lending leads to a drop in the variance of the entrepreneur’s consumption.

However, household consumption variability actually rises because once bailouts are gone, households need to inject equity to cover negative profits and keep banks afloat (i.e., equity issuance costs add variability to household cash flows).⁴³ These costly equity injections (absent when TBTF is in place) increase consumption variability. On the other hand, banks take less risk without TBTF, making states with negative profits less likely reducing (but not eliminating) the need of equity issuance. Hence which effect dominates is a quantitative question. In our parameterization, we find an increase in the variability of household consumption results in an increase in the variability of aggregate consumption. Thus, our estimate yields a lower bound on welfare losses because TBTF policies actually generate lower variance of aggregate consumption.

⁴³This can be seen by looking at the second line in equation (26). When $\iota^b = 1$ (i.e., TBTF is in place) households pay for taxes to cover negative profits in the first term and the second term is reduced significantly. When $\iota^b = 0$ (no TBTF) households cover negative profits in the first term *and* the equity issuance costs in the second term.

Table 11: Large Bank Bailouts: Normative Results

	Baseline National Bank Bailout	No Bailouts Change (%)
	(1)	(2)
Consumption Entrep	1.028	-0.045
Consumption Households	2.006	-0.002
Agg Consumption	1.839	-0.006
Var(Cons Entrep) %	0.0472	-0.251
Var(Cons Households) %	0.0042	1.274
Var(Agg Consumption) %	0.0080	0.666
Var(gdp) %	0.0125	-0.563

Note: Change (%) corresponds to the percentage change between the corresponding model and the Baseline with National Bank Bailouts. “Agg Consumption” refers to aggregate consumption, Consumption Entrep to consumption entrepreneurs.

In Appendix A-6, we also consider another experiment where we evaluate the consequences of providing bailouts to both national and regional banks. As expected, adding bailouts to regional banks as well as national banks exacerbates the moral hazard problem increasing variability of gdp while consumption variability falls as in the earlier experiment owing to the lack of need to inject equity in failing states which exacerbates a consumption decline.

8.1 Too-Big-To-Fail and Bank Granularity

In this Section, we evaluate how spillovers are affected by the presence of TBTF policies (see Table 12). When comparing the economy without TBTF versus our baseline, we find a decline in the variance of GDP and a reduction in the fraction of output variation that can be explained by dominant bank shocks which is consistent with the drop in variance of gdp in Table 11. Without TBTF policies in place, funding shocks to large banks induce smaller aggregate effects through lower risk taking (i.e., banks become less systemically important).

We evaluate how spillovers are affected by the presence of TBTF policies. We perform an experiment similar to that described in Section 7 where we compare the aggregate dynamics of the model (with and without TBTF) when all shocks are present to a simulation where the aggregate productivity shock and regional shocks are kept constant at their mean (i.e., the economy is driven by shocks to dominant banks). Table 12 presents the results. We observe that the model without TBTF displays lower standard deviation of output and loan supply. Without TBTF, banks are more prudent (i.e., make less risky loans) and this change in behavior induces smaller fluctuations in aggregate output. This table also shows that this reduction in volatility implies that idiosyncratic shocks in the model without TBTF subsidies

explain a smaller fraction of aggregate fluctuations (2.2% versus 1.5% of aggregate output, for example).

Table 12: Too-Big-to-Fail Policy and Spillovers

	Baseline: National Bank Bailouts		No Bank Bailouts	
Variable	All shocks	Explained by bank shocks %	All shocks	Explained by bank shocks %
$\text{Var}(\text{gdp})$	0.0125	2.19	0.0124	1.48
$\text{Var}(\text{gdp}_{\mathcal{R}})$	0.0107	76.94	0.0080	70.29
$\text{Var}(\text{gdp}_{\mathcal{N}\mathcal{R}})$	0.0100	54.98	0.0082	46.18
$\text{Cov}(\text{gdp}_{\mathcal{R}}, \text{gdp}_{\mathcal{N}\mathcal{R}})$	-0.0041	164.50	-0.0019	243.49
$\text{Var}(L^s)$	0.00153	13.92	0.0015	9.72
$\text{Var}(L_{\mathcal{R}}^s)$	0.0064	99.95	0.0044	101.11
$\text{Var}(L_{\mathcal{N}}^s)$	0.0057	75.26	0.0043	69.37
$\text{Cov}(L_{\mathcal{R}}^s, L_{\mathcal{N}}^s)$	-0.0053	98.97	-0.0036	100.93

Note: “Baseline: National Bank Bailouts” corresponds to the baseline model with National Bank Bailouts. “No Bank Bailouts” corresponds to the model without national bank bailouts. The columns “All shocks” refer to the simulation of the corresponding model where all shocks (aggregate, regional, and bank level shocks) are present. The columns “Explained by bank shocks %” correspond to the fraction of those standard deviation that is explained by shocks to dominant banks. The calculation is similar to that presented in Table 8 and derives from comparing the simulation with all shocks to one where aggregate shocks and regional shocks are kept constant at their mean (i.e., the only shocks driving the time series variation are dominant bank level shocks).

9 Concluding Remarks

We document the evolution of the US banking industry since the early 1980s and provide a model where “big” geographically diversified banks coexist in equilibrium with regional and fringe banks that are restricted to a geographical area. Since we allow for regional specific shocks to the success of borrower projects, smaller banks (both regional and fringe) may not be well diversified. This assumption generates ex-post differences between big and smaller banks.

Idiosyncratic deposit shocks to both national and regional banks spill over via the loan market to aggregate economic activity as in the granular hypothesis of [Gabaix \(2011\)](#) posing a systemic challenge for regulators unlike idiosyncratic shocks to atomistic (measure zero) fringe banks. As echoed in Bernanke’s statement, this highlights why studying the positive and normative effects of TBTF policies that target dominant, systemically important banks, is a critical question for financial stability.

Here we quantify the effects of subsidies to big banks in distressed states, or what it is commonly referred to as “too-big-to-fail” (TBTF) policies. Consistent with a moral hazard concern, TBTF induces dominant banks to extend more loans in risky states (i.e., increase

their exposure), and this spills over to smaller banks inducing them to lower their loan supply exposure. Since there are costs and benefits of the TBTF policy, it becomes a quantitative question. We find that the net effect of TBTF can actually lead to a fall in interest rates on loans, lower default frequencies owing to a selection effect, and higher intermediated output. The policy, however, comes at the expense of the regional and fringe banks shrinking their market shares and higher taxes. The welfare effects of eliminating too-big-to-fail as suggested by the quote from Bernanke are nuanced; aggregate consumption can fall due to less lending and the variance of consumption may rise for some segments of the population but fall for others.

There is much work left to do as part of this research agenda. One important agenda item is to understand how regulatory changes may have contributed to the endogenous transition of rising market shares in Figure 1. In Corbae and D’Erasmus (2020) we interpret the relaxation of branching restrictions in the Riegle-Neal Act of 1994 as a decrease in the cost of climbing a bank funding ladder (along the lines of the industrial organization model of Besanko and Doraszelski (2004)). Rising concentration in our paper occurs along an endogenous transition path between two steady states as banks invest to acquire a larger, more diversified deposit base and lower average costs. Corbae, D’Erasmus, and Smith (2025) explicitly model the merger process extending the Gowrisankaran and Holmes (2004) framework to account for data features of the U.S. banking industry. As in the data where big, dominant banks overwhelmingly acquire smaller targets subject to anti-trust policy, endogenous mergers can account for the transition path evident in Figure 1.

References

- Acharya, Viral V, Stephen G Cecchetti, and Kermit L Schoenholtz. 2023. “Overview of recent banking stress.” *SVB and Beyond: The Banking Stress of 2023* 2023:14–32.
- 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.
- . 2025. “The geographic flow of bank funding and access to credit: Branch networks, synergies, and local competition.” *American Economic Review* 115 (6):1818–1856.
- Allen, Franklin and Douglas Gale. 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.
- Atkeson, Andrew G, Adrien d’Avernas, Andrea L Eisfeldt, and Pierre-Olivier Weill. 2019. “Government guarantees and the valuation of American banks.” *NBER Macroeconomics Annual* 33 (1):81–145.
- Bassett, William F, Mary Beth Chosak, John C Driscoll, and Egon Zakrajšek. 2014. “Changes in bank lending standards and the macroeconomy.” *Journal of Monetary Economics* 62:23–40.
- 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, Saki Bigio, Jeremy Majerovitz, and Matias Vieyra. 2026. “A q-theory of banks.” *Review of Economic Studies* 93 (1):106–143.
- 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.
- Berndt, Antje, Darrell Duffie, and Yichao Zhu. 2025. “The decline of too big to fail.” *American Economic Review* 115 (3):945–974.

- Besanko, David and Ulrich Doraszelski. 2004. "Capacity dynamics and endogenous asymmetries in firm size." *RAND Journal of Economics* :23–49.
- Bianchi, Javier and Saki Bigio. 2022. "Banks, liquidity management, and monetary policy." *Econometrica* 90 (1):391–454.
- 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’Erasmus. 2020. "Rising bank concentration." *Journal of Economic Dynamics and Control* 115.
- . 2021. "Capital buffers in a quantitative model of banking industry dynamics." *Econometrica* 89:2975–3023.
- Corbae, Dean, Pablo D’Erasmus, and Charles Smith. 2025. "A quantitative model of bank merger dynamics." Mimeo.
- Corbae, Dean and Ross Levine. 2024. "Competition, stability, and efficiency in the banking industry." Mimeo.
- d’Avernas, Adrien, Andrea L Eisfeldt, Can Huang, Richard Stanton, and Nancy Wallace. 2023. "The deposit business at large vs. small banks." Tech. rep., National Bureau of Economic Research.
- 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.
- Dempsey, Kyle. 2026. "Capital requirements with non-bank finance." *Review of Economic Studies* 93 (3):1635–1670.
- Den Haan, Wouter J, Steven W Sumner, and Guy M Yamashiro. 2007. "Bank loan portfolios and the monetary transmission mechanism." *Journal of Monetary Economics* 54 (3):904–924.

- Diamond, Douglas W. 1984. “Financial intermediation and delegated monitoring.” *The Review of Economic Studies* 51 (3):393–414.
- Diebold, Francis X and Kamil Yilmaz. 2009. “Measuring financial asset return and volatility spillovers, with application to global equity markets.” *The Economic Journal* 119 (534):158–171.
- Dávila, Eduardo and Ansgar Walther. 2020. “Does size matter? Bailouts with large and small banks.” *Journal of Financial Economics* 136 (1):1–22.
- Ennis, Huberto M and H Sam Malek. 2005. “Bank risk of failure and the too-big-to-fail policy.” *FRB Richmond Economic Quarterly* 91 (2):21–44.
- 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.
- Fernald, John. 2014. “A quarterly, utilization-adjusted series on total factor productivity.” Federal Reserve Bank of San Francisco.
- Gabaix, Xavier. 2011. “The granular origins of aggregate fluctuations.” *Econometrica* 79 (3):733–772.
- 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.
- Galaasen, Sigurd, Rustam Jamilov, Ragnar Juelsrud, and H  lene Rey. 2020. “Granular credit risk.” Tech. rep., National Bureau of Economic Research.
- Gallin, Joshua. 2013. “Shadow banking and the funding of the nonfinancial sector.” In *Measuring Wealth and Financial Intermediation and Their Links to the Real Economy*. University of Chicago Press, 89–123.
- 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.
- Guerrieri, Luca, Matteo Iacoviello, Francisco Covas, John C Driscoll, Mohammad Jahan-Parvar, Michael Kiley, Albert Queralto, and Jae Sim. 2019. “Macroeconomic effects of banking-sector losses across structural models.” *59th issue (September 2019) of the International Journal of Central Banking* .

- 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.
- Jamilov, Rustam and Tommaso Monacelli. 2026. “Bewley banks.” *Review of Economic Studies* 93 (3):1889–1925.
- 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.
- Meehl, April. 2024. “Bailouts, bail-ins, and banking industry dynamics.” Mimeo.
- 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*.
- Pozsar, Zoltan, Tobias Adrian, Adam Ashcraft, and Hayley Boesky. 2010. “Shadow banking.” *Federal Reserve Bank of New York Staff Report No 458*:3–9.
- Rios-Rull, José-Victor, Tamon Takamura, and Yaz Terajima. 2023. “Banking dynamics, market discipline and capital regulations.” *mimeo*.
- Shaffer, S. 2004. “Patterns of competition in banking.” *Journal of Economics and Business* 56:287–313.
- Stern, Gary H and Ron J Feldman. 2004. *Too big to fail: The hazards of bank bailouts*. Rowman & Littlefield.
- 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.
- Weintraub, Gabriel Y, C Lanier Benkard, and Benjamin Van Roy. 2008. “Markov perfect industry dynamics with many firms.” *Econometrica* 76 (6):1375–1411.
- 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>.

Online Appendix

A Quantitative Model of Banking Industry Dynamics

by Dean Corbae and Pablo D’Erasmus

A-1 Data Description

As in [Corbae and D’Erasmus \(2021\)](#), we compile a large panel of banks from 1984 to 2019 using quarterly data.⁴⁴ The source for the data is the Consolidated Report of Condition and Income (known as Call Reports) that banks submit to the Federal Reserve each quarter.⁴⁵ Report of Condition and Income data are available for all banks regulated by the Federal Reserve System, Federal Deposit Insurance Corporation, and the Comptroller of the Currency. All financial data are on an individual bank basis.

We consolidate individual commercial banks to the bank holding company level and retain those bank holding companies and commercial banks (if there is not top holder) for which the share of assets allocated to commercial banking (including depository trust companies, credit card companies with commercial bank charters, private banks, development banks, limited charter banks, and foreign banks) is higher than 25 %. We follow [Kashyap and Stein \(2000\)](#) and [Den Haan, Sumner, and Yamashiro \(2007\)](#) in constructing consistent time series for our variables of interest. Finally, we include only banks located within the 50 states and the District of Columbia. In addition to information from the Call Reports, we identify bank failures using public data from the Federal Deposit Insurance Corporation (FDIC).⁴⁶ We also identify mergers and acquisitions using the Transformation Table in <https://www.ffiec.gov/npw/FinancialReport/DataDownload>. We identify “events” where the acquired and acquiring firms are commonly owned in some form before the acquisition (i.e., the listed merger is only a corporate reorganization) and discard these events from the merger sample.

To deflate balance sheet and income statement variables, we use the CPI index. When we report weighted aggregate time series, we use the asset market share as the weight. To control for the effect of a small number of outliers when constructing the loan returns, cost of funds, charge offs rates and related series, we eliminate observations in the top and bottom 1% of the distribution of each variable. We also control for the effects of bank entry, exit and mergers by not considering the initial period, the final period or the merger period (if relevant) of any given bank.

⁴⁴There was a major overhaul to the Call Report format in 1984. Since 1984 banks are, in general, required to provide more detailed data concerning assets and liabilities. Due to changes in definitions and the creation of new variables after 1984, some of the variables are available only after this date.

⁴⁵Balance Sheet and Income Statements items can be found at <https://cdr.ffiec.gov/public/>.

⁴⁶Data is available at <https://www.fdic.gov/bank/individual/failed/banklist.html>.

Table A.1: Model Definitions

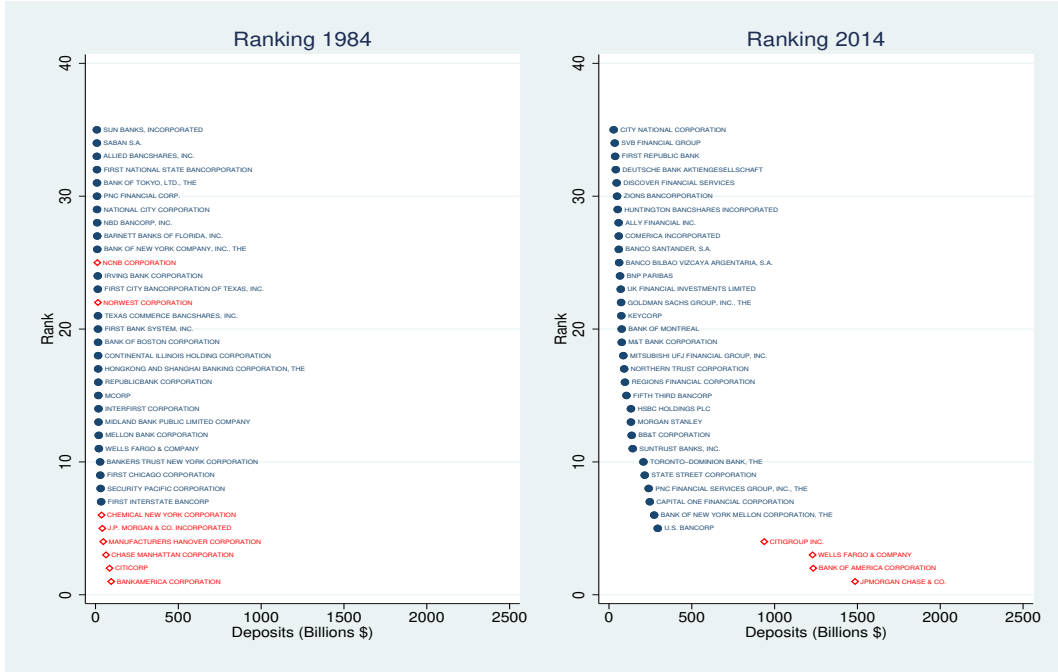
Assets (Assets_θ)	$\ell_\theta + \max\{a_\theta, 0\} + \Upsilon_\theta$
Liabilities (Liab_θ)	$d_\theta + \max\{-a_\theta, 0\}$
Loans / Assets	$\ell_\theta / \text{Assets}_\theta$
Securities / Assets	$\max\{a_\theta, 0\} / \text{Assets}_\theta$
Deposit / Assets	$d_\theta / \text{Assets}_\theta$
Default Frequency	$(1 - p_{k,j})$
Charge off Rate	$(1 - p_{k,j})\lambda$
Loan return	$p_{k,j} \cdot r_{k,j} - (1 - p_{k,j}) \cdot \lambda$
Loan interest margin	$p_{k,j} \cdot r_{k,j} - r_D$
Loan interest spread	$r_{k,j} - r_D$
Markup	$p_{\mathcal{B},j} \cdot r_{\mathcal{B},j} / (r_D + c_\theta^1 Z + 2c_\theta^2 Z \ell_\theta(s_\theta, Z, \mu))$
Total Output	$\sum_j L_{\mathcal{B},j}^s \left\{ p_{\mathcal{B},j} \cdot Z'(1 + z'_j) R_{\mathcal{B},j} + (1 - p_{\mathcal{B},j})(1 - \lambda) \right\}$ $+ \sum_j L_{\mathcal{N}\mathcal{B},j}^s \left\{ p_{\mathcal{N}\mathcal{B},j} \cdot Z'(1 + z'_j) R_{\mathcal{N}\mathcal{B}} + (1 - p_{\mathcal{N}\mathcal{B},j})(1 - \lambda) \right\}$
Failure Rate	$\int_{\theta, s_\theta} x_\theta(s_\theta, S') d\mu_\theta(s_\theta) / \int_{\theta, s_\theta} d\mu_\theta(s_\theta)$
Deposit Market Share θ	$\int_{s_\theta} d_\theta d\mu_\theta(s_\theta) / \int_{\theta, s_\theta} d_\theta d\mu_\theta(s_\theta)$
Loan Market Share θ	$\int_{s_\theta} \ell_\theta(s_\theta, z, \mu) d\mu_\theta(s_\theta) / \int_{\theta, s_\theta} \ell_\theta(s_\theta, z, \mu) d\mu_\theta(s_\theta)$
Borrower Risk Taking	$\sum_{k,j} L_{k,j}^s R_{k,j} / \sum_{k,j} L_{k,j}^s$
Borrower Return	$\sum_{k,j} L_{k,j}^s p_{k,j} Z'(1 + z'_j) R_{k,j} / \sum_{k,j} L_{k,j}^s$

A-2 Banking Data Facts

A-2.1 Fat Right Tail Bank Size Distribution

Figure A.1 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.

Figure A.1: Deposit Distribution 1984 & 2014



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

Table A.2 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. In Figure 4 in the main text, we present the average estimate for two different periods in our sample: 1984-1993 (Panel (i)) and 2011-2019 (Panel (ii)). The estimated averages for these periods are 1.1303 and 1.1701, respectively.

Table A.2: Zipf's Law Estimates Assets, Loans, and Deposits (across time)

	Power Exponent α				
	Year				
	1984	1994	2004	2014	2019
Assets	1.151	1.132	1.122	1.173	1.298
(se)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Loans	1.180	1.151	1.113	1.185	1.319
(se)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Deposits	1.118	1.094	1.081	1.152	1.282
(se)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: We estimate α and its standard error (se) as in [Gabaix and Ibragimov \(2011\)](#) and restrict the sample to the top 3,000 banks (as in [Janicki and Prescott \(2006\)](#)). All estimates are significantly different from 1 at the 1% level. Source: Reports of Condition and Income (Call Reports).

A-2.2 Cost Estimation

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:

$$\begin{aligned}
 \log(NIE_t^i) = & g_1 \log(W_t^i) + \varsigma_1 \log(\ell_{\theta,t}^i) + g_2 \log(q_t^i) + g_3 \log(W_t^i)^2 \\
 & + \varsigma_2 [\log(\ell_t^i)]^2 + g_4 \log(q_t^i)^2 + \varsigma_3 \log(\ell_t^i) \log(q_t^i) + \varsigma_4 \log(\ell_t^i) \log(W_t^i) \\
 & + g_5 \log(q_t^i) \log(W_t^i) + \sum_{j=1,2} g_6^j t^j + g_{8,t} + g_9^i + \epsilon_t^i,
 \end{aligned} \tag{A.2.1}$$

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.⁴⁷ Non-interest marginal expenses are then computed as:

$$\text{Mg Non-Int Exp.} \equiv \frac{\partial NIE_t^i}{\partial \ell_t^i} = \frac{NIE_t^i}{\ell_t^i} \left[\varsigma_1 + 2\varsigma_2 \log(\ell_t^i) + \varsigma_3 \log(q_{it}) + h_4 \log(W_t^i) \right]. \tag{A.2.2}$$

Marginal non-interest income (Mg Non-Int Inc.) is estimated using an equation similar to equation (A.2.1) (without input prices) where the left hand side corresponds to log total

⁴⁷We 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.

non-interest income. Net marginal expenses (Mg Net Exp.) are computed as the difference between marginal non-interest expenses and marginal non-interest income. The fixed cost κ_θ is estimated as the total cost on expenses of premises and fixed assets. The average cost is simply the sum of marginal net expenses and fixed costs.

A-2.3 Entry and Exit Statistics by Size

Table A.3 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.

Table A.3: 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	0.31	0.03
Fringe (Not Top 35)	99.69	99.97
Fraction of Loans of Banks in x accounted by:		
Top 4 Banks	0.60	0.00
Top 5 - 35 Banks	9.02	0.95
Fringe (Not Top 35)	90.37	99.05

Note: Top 4 Banks and Top 5-35 Banks refer 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

A-2.4 Returns and Funding Costs

Table A.4 describes the structure of asset returns and funding costs across the bank size distribution.⁴⁸

⁴⁸Following 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 $p \cdot r_{\mathcal{B},L} - (1 - p)\lambda$ where $r_{\mathcal{B},L}$ is the bank loan interest rate and $p \cdot r_{\mathcal{B},L}$ 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 $p \cdot r_{\mathcal{B},L}$ and our estimate of p . The net interest margin corresponds to $p \cdot r_{\mathcal{B},L} - r_D$ and the loan interest spread to $r_{\mathcal{B},L} - r_D$. The net security return is estimated as the ratio of interest income

Table A.4: 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 ^{†,‡}	0.30	0.80 [‡]	0.48	0.53	0.67
Default freq.	2.14	1.35	3.00 ^{†,‡}	0.76	1.96 [‡]	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 [‡]	0.62	4.39	0.84
Net Interest Margin	4.65	0.62	4.72 ^{†,‡}	0.44	4.31 [‡]	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 [†]	0.46	4.46 [‡]	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	102.26	52.80	143.73 ^{†,‡}	47.22	103.89 [‡]	41.39	78.38	30.98
Lerner	46.23	9.27	56.28 ^{†,‡}	6.63	46.17 [‡]	8.42	40.09	7.30
Net Mg Cost	1.25	0.73	0.74 ^{†,‡}	0.61	1.07 [‡]	0.64	1.86	0.57

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 refer 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 are 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

Using Call Report data, [d'Avernas et al. \(2023\)](#) find that small banks offer saving deposit rates about 24 basis points larger than those offered by big banks (14 large institutions) but no differences for time and transaction deposits between 2001 to 2020. Also using Call Reports but for our size groups and between 1984 and 2019, we do not find evidence for statistically different deposit rates between the Top 4, the Top 5-35, and fringe banks for the period 1984 - 2019. Focusing on 2001-2019, we obtain differences of 27 and 24 basis points between fringe banks and the Top 4 and the Top 5-35, respectively. However, we do not find these differences to be statistically significant.

A-2.5 Business Cycle Statistics

The dynamics of entry and exit discussed above as well as differences in intensive margin responses lead to interesting business cycle dynamics. Table A.5 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 detrended 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

on securities plus the net cost of federal funds divided by cash plus securities plus net federal funds lending.

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 A.5: Business Cycle Statistics by Bank Size (1984-2019)

	All Banks	Top 4	Top 5-35	Fringe
Charge-off Rate	-0.230***	-0.323***	-0.193***	-0.132 ***
Default freq.	-0.521***	-0.680***	-0.411***	-0.346 ***
Cost Deposits	-0.096	-0.044	-0.108	-0.153
Loan return	0.052	0.064	0.083	-0.055
Net Interest Margin	-0.085***	-0.107***	-0.041**	-0.041
Loan Interest Rate	-0.213**	-0.252**	-0.156	-0.209 **
Loan Interest Spread	-0.116***	-0.150***	-0.063 ***	-0.061**
Net securities return	-1.294	-1.487	-0.977	-0.913
Markup	-2.508 *	-4.980 *	-2.211	-2.404**
Lerner	-0.699	-1.239	-0.533	-0.710*
Net Marginal Cost	-0.107***	-0.136***	-0.130***	-0.052 ***

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 as interest expenses on deposits (and derived variables) are deflated using the CPI. Top 4 Banks and Top 5-35 Banks refer 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

A-3 Computational Algorithm

In this section, we describe the computational algorithm that allows us to compute bank strategies and the equilibrium of the model. We use an extension of the algorithm proposed by [Pakes and McGuire \(1992\)](#) that incorporates a competitive fringe. The description is based on the case where fringe banks do not have access to outside funding, entering fringe banks draw if c_f^1 exceeds the highest cost of fringe incumbents in the market that period, the regional shocks are iid over time, and the number of representative national banks and regional banks in each region is restricted to one.

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

1. Solve the entrepreneur problem. This amounts to define a grid over $r_{k,j}$ and for

each pair $\{r_{k,j}, Z\}$, conditional on the choice of financial institution $k \in \{\mathcal{B}, \mathcal{NB}\}$ obtain the type of project the entrepreneur operates $R_{k,j}(r_{k,j}, Z)$, the choice of financial institution k , and the decision on whether to operate the technology or not $\iota_j \in \{0, 1\}$. Using $\iota_j(\omega, r_{\mathcal{B},j}, r_{\mathcal{NB},j}, Z)$ we can derive the total loan demand in each region j : $L_j^d(r_{\mathcal{B},j}, r_{\mathcal{NB},j}, Z)$ and the bank loan demand $L_{\mathcal{B},j}^d(r_{\mathcal{B},j}, r_{\mathcal{NB},j}, Z)$.

2. Obtain the solution to the Non-Bank problem by solving for $r_{\mathcal{NB},j}$ using equation (31).
3. Solve for the reaction function of the fringe sector $L_\ell^s(r_{\mathcal{B},j}, Z)$ as a function of the bank interest rate and aggregate shock as well as the exit threshold for fringe banks. For each pair $\{r_{\mathcal{B},j}, Z\}$ and using the solution to the entrepreneurs problem $R_{\mathcal{B},j}(r_{\mathcal{B},j}, Z)$ to derive the failure probability, compute expected profits for fringe banks with cost c_ℓ^1 in $\Xi(c_\ell^1)$ and derive $\bar{c}_{\ell,j}^1(r_{\mathcal{B},j}, Z)$, i.e., the maximum cost such that a fringe bank would choose to operate and the amount of loans they offer. The loan supply from fringe banks is

$$L_\ell^s(r_{\mathcal{B},j}, Z) = \int_0^{\bar{c}_{\ell,j}^1} \ell_\ell(s_\ell, Z, r_{\mathcal{B},j}) d\Xi(c_{\ell,j}^1) dG^*(d_\ell).$$

The exit threshold $\bar{c}_{\ell,j}^{1,x}(s_\ell, Z, r_{\mathcal{B},j}, Z', z'_e, z'_w)$ is the one that solves $\pi_\ell(s_\ell, Z, r_{\mathcal{B},j}, Z', z'_e, z'_w) = W_\ell^{x=1}(s_\ell, Z, r_{\mathcal{B},j}, Z', z'_e, z'_w)$.

4. Using the bank loan demand and the loan supply by the fringe sector, define the residual loan demand $\tilde{L}_j^d(r_{\mathcal{B},j}, Z)$ as follows:

$$\tilde{L}_{\mathcal{B},j}^d(r_{\mathcal{B},j}, Z) = \max\{0, L_{\mathcal{B},j}^d(r_{\mathcal{B},j}, r_{\mathcal{NB},j}, Z)\}.$$

5. Define a grid over all possible combinations of deposits for active and non-active dominant banks: $\mu = \{I_n \times d_n, I_{r,e} \times d_{r,e}, I_{r,w} \times d_{r,w}\}$ where I_θ takes value 1 if bank θ is active and holds a particular value of deposits. If the number of deposit states is Nd for each bank type θ , the size of this grid is $(Nd + 1) \times (Nd + 1) \times (Nd + 1)$ (i.e., all possible states when active plus the empty set when non-active).
6. For each state $\{\mu, Z\}$, solve the problem of dominant banks (that also determines the solution of the fringe banks). We use value function iteration in this step:
 - (a) Guess a loan decision rule, exit decision rule, entry decision rule, and value function for each dominant bank: $\ell_\theta^0(s_\theta, Z, \mu)$, $V_\theta^0(s_\theta, Z, \mu)$, $e_\theta^0(s_\theta, Z, \mu)$ and $x_\theta^0(s_\theta, S')$ for $\theta = \{n, r\}$ and $j = e, w$. Using entry and exit decision rules, define $\mu' = \mathcal{F}^0(\mu, M^e, S')$.
 - (b) For each bank, solve the optimal loan and exit decision rule taking as given the strategy of other banks. The equilibrium in the loan market in region j (i.e., the equilibrium interest rate) is derived from the following equation:

$$\int_{\theta \in \{n, r\}, s_\theta} \ell_{\theta,j}^0(s_\theta, Z, \mu) d\mu_\theta(s_\theta) - \tilde{L}_{\mathcal{B},j}^d(r_{\mathcal{B},j}, Z) = 0, \quad (\text{A.3.3})$$

where $\tilde{L}_{\mathcal{B},j}^d(r_{\mathcal{B},j}, Z)$ is the residual loan demand defined above. The solution to this problem will provide a loan decision rule $\ell_{\theta,j}^1(s_\theta, Z, \mu)$, an exit decision rule $x_\theta^1(s_\theta, S')$, and a new value function $V_\theta^1(s_\theta, Z, \mu)$.

- (c) If $|x_\theta^1(s_\theta, S') - x_\theta^0(s_\theta, S')| < \epsilon^x$ and $|\ell_{\theta,j}^1(s_\theta, Z, \mu) - \ell_{\theta,j}^0(s_\theta, Z, \mu)| < \epsilon^\ell$ for ϵ^v , ϵ^x , and ϵ^ℓ small you have obtained the solution to the bank problem and can continue. If not, set $V_\theta^0(s_\theta, Z, \mu) = V_\theta^1(s_\theta, Z, \mu)$, $x_\theta^0(s_\theta, S') = x_\theta^1(s_\theta, S')$, $\ell_{\theta,j}^0(s_\theta, Z, \mu) = \ell_{\theta,j}^1(s_\theta, Z, \mu)$, update $\mu' = \mathcal{F}^0(\mu, M^e)$, and return to step 6b.
- (d) Solve the entry problem of dominant banks, that is select $e_\theta^1(s_\theta, Z, \mu) \in \{0, 1\}$. Each dominant bank will enter (i.e., $e_\theta^1 = 1$) if there is no other active dominant bank of the same type and

$$-\Upsilon_\theta + E_{s'_\theta} V_\theta(s'_\theta, Z', \hat{\mu}') \geq 0,$$

where $\hat{\mu}'$ is the distribution that would arise if the bank of type θ enters and takes into account the entry decision rule by other banks. If $|e_\theta^1(s_\theta, Z, \mu) - e_\theta^0(s_\theta, Z, \mu)| < \epsilon^e$ you have obtained the solution to the entry problem and can continue. If not, update $\mu' = \mathcal{F}^0(\mu, M^e, S')$ and return to step 6b.

7. The equilibrium dominant banks loan decision rules together with equation (A.3.3) determine the equilibrium loan interest rate $r_{\mathcal{B},j}(Z, \mu)$, the equilibrium loan supply by fringe banks $L_\ell^s(r_{\mathcal{B},j}, Z)$ as well as the thresholds $\bar{c}_{\ell,j}^1(r_{\mathcal{B},j}, Z)$ and $\bar{c}_{\ell,j}^{1,x}(s_\ell, Z, r_{\mathcal{B},j}, Z', z'_e, z'_w)$ for each region j .

A-4 Calibration

Table A.6 presents all the parameters of the model.

Table A.6: 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 (%)	$\bar{r} = r^D$	0.0014	Avg Interest Expense Deposits
Bank Discount Factor	β	0.9986	$1/(1 + \bar{r})$
Return on Securities	r_A^+	0.0180	Return on Net Securities
Cost of borrowing (linear term)	$r_{A,0}^-$	0.0180	Return on Net Securities
Deposit Process Parameters	$\{\rho_\theta^d, \sigma_{\theta,u}\}$		Deposit Process Estimates (eq 2)
Bailout parameter	φ_n	0.900	Atkeson et al. (2019)
Borrower Success Prob. Function	a	4.077	Avg. Borrower Return
Borrower Success Prob. Function	b	28.944	Avg. Default Frequency
Borrower Success Prob. Function	σ_ϵ	0.107	Avg. Loan Interest Rate
Outside Option	$\bar{\omega}$	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	$\bar{d}_\ell$	0.028	Relative Size Fringe to Top 5-35
Mean deposit process r	$\bar{d}_r$	1.000	Loans to Output Ratio
Mean deposit process n	$\bar{d}_n$	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	0.200	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	0.005	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	0.005	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	0.020	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
Cost of borrowing	$r_{A,2}^-$	0.100	Deposit to Asset ratio Regional
External finance param. r	ξ_r^1	0.200	Avg. equity issuance to loan ratio Top 5-35
External finance param. n	ξ_n^1	0.100	Avg. equity issuance to loan ratio top 4
Entry Cost r	Υ_r	0.202	Bank Loan Market Share Regional
Entry Cost n	Υ_n	0.606	Bank Loan Market Share National
Nonbank Marginal Cost	c_N	0.035	Bank Loan to Total Credit Ratio
Bailout probability	p^b	0.950	Failure Rate

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.

Tables A.7 and A.8 present a set of data moments together with their model generated counterparts. Table A.7 presents the targets of the calibration exercise and Table A.8 presents additional moments not targeted. The average borrower return refers to the average real equity return (12.94%) as reported by Diebold and Yilmaz (2009), added to help identify parameters associated with the borrower’s return. We employ the elasticity of loan demand (-1.10) estimated by Bassett et al. (2014). Determining the size of the non-bank sector is difficult and depends on the definition used, estimates of the ratio of bank loans to total credit (i.e., a measure that includes the non-bank sector) vary significantly from a lower bound of 35% (see Pozsar et al. (2010)) to an upper bound of 65% (see Gallin (2013)). We use the middle range of these estimates and target a bank loan to total credit ratio equal to 50%. In addition, as a way to determine the size of the commercial banking sector in the economy, we use the deposits in commercial banks to GDP ratio, and the loans and leases in commercial banks to GDP ratio, which average 57.78% and 60.34%, respectively, between 2011-2019.⁴⁹ All other moments listed in Tables A.7 and A.8 are derived from the Call Reports using a sample between 1984 and 2019.

⁴⁹To construct these ratios we use Gross Domestic Product (GDP), Deposits, All Commercial Banks (DPSACBM027SBOG), and Loans and Leases in Bank Credit, All Commercial Banks (“LOANS”) retrieved from FRED, Federal Reserve Bank of St. Louis.

Table A.7: Moments Data vs Model (Targets)

	Data	Model
Avg. Borrower Return	12.94	12.93
Avg. Default Frequency	2.14	2.22
Avg. Loan Interest Rate	4.85	4.32
Elasticity of Loan Demand	-1.10	-0.90
Std Dev Loan Returns	0.84	1.29
Std Dev Charge-offs Top 4	0.30	1.14
Avg Charge Off Rate	0.85	0.70
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 Fringe	1.86	3.41
Elasticity Net Marginal Expenses Fringe	0.74	1.00
Net Marginal Expenses Top 5-35	1.07	1.10
Elasticity Net Marginal Expenses Top 5-35	0.92	1.00
Net Marginal Expenses Top 4	0.74	1.19
Elasticity Net Marginal Expenses Top 4	1.02	1.90
Fixed cost over loans Fringe	0.75	2.00
Fixed cost over loans Top 5-35	0.70	0.57
Fixed cost over loans Top 4	0.73	0.28
Bank Loans to Output Ratio	43.75	44.04
Deposit To Output Ratio	57.82	57.73
Deposit Market Share Top 5-35	32.45	32.95
Deposit Market Share Top 4	29.60	28.87
Deposits to Assets Top 5-35	83.25	82.82
Avg. equity issuance to asset ratio Top 5-35	0.07	0.09
Avg. equity issuance to asset ratio Top 4	0.02	0.11
Bank Loan Market Share Top 5-35	33.97	37.17
Bank Loan Market Share Top 4	29.87	37.45
Bank Loan to Total Credit Ratio	50.00	49.41
Bank failure rate (all)	0.404	0.95

Table A.8: Moments Data vs Model (Untargeted Additional Moments)

	Data	Model
Bank failure rate Fringe	0.41	0.98
Bank Failure rate Regional	0.09	0.00
Bank Failure rate National	0.00	0.00
Loan to Assets Fringe	53.80	53.87
Loan to Assets Top 5-35	59.96	72.24
Loan to Assets Top 4	61.30	91.70
Deposits to Assets Fringe	88.41	100.00
Deposits to Assets Top 4	83.80	92.27
Securities to Assets Fringe	35.70	46.13
Securities to Assets Top 5-35	30.81	11.17
Securities to Assets Top 4	30.43	0.64
Fed Funds to Assets Fringe	1.11	0.00
Fed Funds to Assets Top 5-35	7.53	0.58
Fed Funds to Assets Top 4	7.93	0.08
Avg Cost Fringe	2.62	5.41
Avg Cost Top 5-35	1.77	1.66
Avg Cost Top 4	1.47	1.46
Dividends to Assets Fringe	0.10	0.11
Dividends to Assets Top 5-35	0.63	1.72
Dividends to Assets Top 4	0.63	2.20

A-5 Idiosyncratic Bank Shocks and Aggregate Fluctuations

Table A.9 presents the estimated coefficients for the regressions of Δgdp on Γ_t^j and Γ_{t-1}^j as in Gabaix (2011).

Table A.9: Explanatory power of Granular Residuals of Dominant Banks

	Dependent Variable					
	$\Delta \text{ gdp}$		$\Delta \text{ gdp}_{\mathcal{B}}$		$\Delta \text{ gdp}_{\mathcal{N}}$	
	Deposits	Loans	Deposits	Loans	Deposits	Loans
(intercept)	0.040 (0.061)	0.044 (0.061)	0.122 (0.072)*	0.146 (0.066)**	0.029 (0.074)	0.014 (0.071)
Γ_t	-0.0378 (174)	0.088 (0.251)	0.118 (0.206)	0.3014 (0.271)	-0.168 (0.211)	-0.072 (0.291)
Γ_{t-1}	0.686 (0.174)***	1.169 (0.251)***	6.33 (0.206)***	10.628 (0.271)***	-4.787 (0.211)***	-7.962 (0.291)***
N	1899	1899	1899	1899	1899	1899
R^2	0.009	0.011	0.338	0.454	0.217	0.289

Note: “ $\Delta \text{ gdp}$ ” refers to growth rate of total output, “ $\Delta \text{ gdp}_{\mathcal{B}}$ ” to the growth rate of bank intermediated output, and “ $\Delta \text{ gdp}_{\mathcal{N}}$ ” to the growth rate of nonbank intermediated output. “Deposits” columns refer to regressions where the dependent variable is constructed using $\epsilon_{i,t}^d$. “Loans” columns correspond to regressions where the dependent variable is constructed using $\epsilon_{i,t}^\ell$. Standard errors in parenthesis. *, **, and *** refer to variable significant at 10, 5, and 1 percent level, respectively.

A-6 National and Regional Bank Bailouts

In this section, we compare our baseline economy with one where there are bailouts for all dominant banks. More specifically, we keep $p^b = 0.95$ as in the baseline and set $\varphi_n = \varphi_r = 0.90$. Column (3) in Table A.10 presents the results.

Table A.10: Baseline vs National and Regional Bank Bailouts

	Baseline National Bank Bailout	National and Reg Bank Bailout Change (%)
	(1)	(2)
Bank Loan Supply	0.24	1.94
Total Loan Supply	0.48	0.18
Share Bank Loans (%)	49.41	1.76
Markup (%)	176.24	-3.61
Loan Market Share Fringe (%)	25.36	-4.32
Loan Market Share Regional (%)	37.18	5.07
Loan Market Share National (%)	37.46	-2.11
Borrower Risk Taking	13.20	-0.007
Default Frequency (%)	2.23	-0.49
Net Interest Margin (%)	4.08	-0.87
Bank Loan Interest Rate (%)	4.32	-0.85
Failure Rate (%)	0.95	-0.87
Prob. $\pi_r < 0$	7.36	7.94
Prob. $\pi_n < 0$	5.69	0.08
Output	0.53	0.18
Taxes Bank Failures	0.0002	-1.28
Taxes	0.0003	31.46
Taxes/Output Bank Failures (%)	0.044	-1.45
Taxes/Output (%)	0.062	32.12

Note: Change (%) corresponds to the percentage change between the corresponding model and the Baseline with National Bank Bailouts.

After incorporating regional bank bailouts, we find that the loan supply of the regional bank increases to take a larger portion of the market (+5%). The increase in lending by regional banks results in an increase in total bank supply (+1.94%), a decline in the interest rate (-0.85%) and a reduction in borrower risk taking and default frequencies (-0.49%). The change in policy induces regional banks to take on more risk, which results in a higher probability of reaching a state with negative profits. Interestingly, this also translates into an increase in the probability of reaching a state with negative profits for national banks as well.

The decline in the default frequency leads to a reduction in bank failures and a drop in the taxes due to failure to output ratio. Total taxes to output increases significantly as the government needs to collect taxes for regional banks as well. The increase in total lending results in an increase in total output (+0.18%).

Table A.11 shows that the increase in output and lending leads to an increase in welfare (measured in consumption equivalents) for the entrepreneurs of about 0.12% and small but

also an increase in welfare for households (+0.004%). The increase in aggregate consumption is accompanied by a reduction in the variance of consumption but an increase in the variance of output.

Table A.11: Baseline vs National and Regional Bank Bailouts: Normative Results

	Baseline National Bank Bailout	National and Reg Bank Bailout Change (%)
	(1)	(2)
Consumption Entrep	1.028	0.117
Consumption Households	2.006	0.004
Agg Consumption	1.839	0.015
Var(Cons Entrep)	0.0472	0.402
Var(Cons Households)	0.0042	-2.632
Var(Agg Consumption)	0.0080	-1.434
Var(gdp)	0.0125	0.903

Note: Change (%) corresponds to the percentage change between the corresponding model and the Baseline with National Bank Bailouts. Agg Consumption refers to aggregate consumption, Consumption Entrep to consumption entrepreneurs.