A Quantitative Model of Bank Mergers

Dean Corbae
University of Wisconsin - Madison and NBER

Pablo D'Erasmo*
Federal Reserve Bank of Philadelphia

Charles R. Smith†
University of Wisconsin - Madison

August 14, 2025

Abstract

We develop a simple model of the bank merger process to study the rise in bank concentration following the deregulation of bank branching in the Riegle-Neal Act of 1994. Motivated by the data where currently 10 (dominant) banks have over 55 percent of the U.S. deposit market share while the remaining over 4000 (fringe) banks cover the rest, we apply a dominant-fringe framework with a merger stage to model the rise in concentration following the change in regulation making interstate branching possible. We study the effect of the merger wave on competition, efficiency, and stability of the banking industry.

Keywords: Bank Mergers, Industry Dynamics, Imperfect Competition, Market Efficiency, Financial Stability.

JEL Classification Numbers: G21, E58.

---

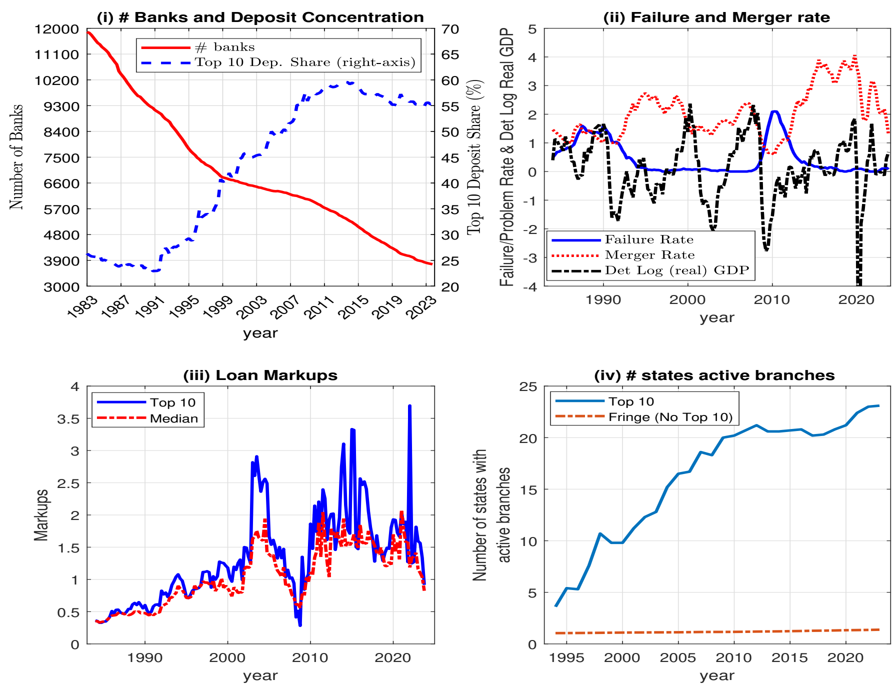
*The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

†Charles Smith gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship Program under Grant No. 2137424. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

1 Motivation

Geographic expansion of the banking industry followed from the elimination of cross-state branching restrictions that began in the McFadden Act of 1927 which gave states the ultimate authority. While some states permitted cross-state branching prior to 1994, the Riegle-Neal Act removed regulatory obstacles to banks opening branches in other states and provided a uniform set of rules regarding banking in each state. As evident in Figure 1 (Panel (i)), following a relatively stable bank concentration of the top 10 banks of under 25% prior to 1994, concentration rose dramatically to over 60% until stabilizing post-Great Recession. Also significant in (Panel (i)) of Figure 1 is the threefold drop in the number of banks (from over 12,000 in 1983 to under 4000 in 2023).

Figure 1: Concentration, Failure/Merger Rates, Markups, and Geographic Expansion (1984 - 2023)



Note: # of banks refers to number of commercial banks in the US (aggregated to the top holding company). Top 10 Dep Share refers to the share of total deposits accounted by the Top 10 banks (when sorted by deposits). Det. Log (real) GDP refers to detrended log real GDP (quarterly frequency). The trend is extracted using the H-P filter with parameter 1600. We estimate markups using the estimates of marginal costs as in Corbae and D'Erasmo (2021). Top 10 markups corresponds to the (asset-weighted) average within Top 10 banks when sorted by assets. Number of states refers to the number of states with active branches at the bank holding company level. Source: Call Reports and Summary of Deposits.

Panel (ii) in Figure 1 documents an elevated merger rate following Riegle-Neil as well as post-Great Recession. The figure also provides evidence for countercylical failures and procyclical mergers. Specifically, the correlation between the failure rate and detrended log-real GDP is

-0.149 while the correlation between the merger rate and detrended log-real GDP is 0.167. While the 1994 Riegle-Neal Act which facilitated unrestricted nationwide banking expansion can account for the first merger wave of 1994-2000, the Financial Crisis in 2008 resulted in a wave of bank failures that coupled with slower banking system growth and increased competition from other intermediation channels led to the second merger wave in the post crisis period.[1]

Regulators use measures based on concentration (e.g. Herfindahl indices) when evaluating whether to allow bank mergers.[2] Regulatory concern about mergers focuses on whether it erodes *competition* leading to market *inefficiency* as measured, for instance, by large markups. Panel (*iii*) in Figure 1 documents the increase in bank markups that may have occurred as a consequence of rising market power.

Regulators also face concerns about whether a banking industry characterized by a few large, systemically important banks can lead to *financial instability* when one big bank's troubles spill over to the rest of the economy.[3] While several papers have focused on balance sheet spillovers among financial institutions causing instability[4], granular spillovers along the lines of Gabaix (2011) can also yield instability. Spillovers from an idiosyncratic shock to a big bank to the broader financial system and economy generate challenges to policymakers; are big banks too-big-to-fail? If so, such policies may actually induce excessive risk taking amplifying financial instability that the policy is intended to reduce. On a technical level, an idiosyncratic shock to atomistic (measure zero) fringe banks has no aggregate consequences, while an idiosyncratic shock to a non-atomistic dominant bank can have aggregate consequences as in the granularity literature.

This paper studies the costs and benefits of bank merger activity that spread across states following the Riegle-Neil Act. While the previous paragraphs focused on some costs, there are potential benefits from such deregulation. For instance, geographic diversification as evidenced in Panel (*iv*) in Figure 1 can lead to financial stability in the presence of idiosyncratic regional shocks to banking portfolios. Specifically, the figure documents that the top 10 banks have diversified to 23 states on average while the remainder are isolated in 1.2 states on average.

To adequately address these costs and benefits given the market structure documented in Figure 1 with 10 dominant banks and 4000 fringe banks in the latest year, we apply and extend industrial organization's dominant-fringe framework of Gowrisankaran and Holmes (2004) with several important exceptions. First, consistent with other papers and our own work (e.g. Corbae and D'Erasmo (2021)) establishing increasing returns in the data, we assume fixed costs which break the constant returns to scale assumption in their paper. This generates fundamental nonlinearities in the pricing of mergers which necessitates a global solution method when solving

[1]The observed higher merger rate post-2010 than the period 1994-2000 is not due to a higher number of mergers (the numerator of the merger rate) but a lower number of banks documented in Panel (*i*) of Figure 1 (the denominator of the merger rate). This will also be evident in Figure 3 below.

[2]Among the techniques to assess the competitive effects of a proposed merger, the FDIC will consider the degree of concentration within the relevant geographic market(s) in order to approve a proposed merger. See Corbae and D'Erasmo (2025) for Herfindahl measures.

[3]Since enactment of the Dodd-Frank Act in 2010, the federal banking agencies have been required by statute to consider risks to financial stability when evaluating proposed bank mergers and acquisitions (12 U.S.C. §1842(c)(7)).

[4]See, for example, Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015).

for a Markov Perfect Equilibrium. Second, we allow for entry and exit which may vary across the cycle as established in Panel (*ii*) Figure 1. This is important with mergers since they can facilitate entry as in other settings; firms may enter to be acquired by a dominant firm. Third, mergers in our model generate a decline in *both* the number of banks and rise in dominant market share documented in Panel (*i*) of Figure 1 while the Gowrisankaran and Holmes (2004) can *only* account for the rise in market share, not the decline in numbers.

The paper is organized as follows. Section 2 provides data on bank mergers that informs our model. Section 3 runs "granular" regressions to study how shocks to large banks have translated to nontrivial aggregate fluctuations. Section 4 describes the model environment and equilibrium. Section 5 presents the calibration. Section 6 illustrates the model's merger dynamics. Section 7 runs a policy counterfactual to ask "What would have happened in the post-financial crisis period if Riegle-Neal was not enacted?" This requires a decomposition between effects associated Dodd-Frank regulation and non-bank technology improvements. Section 8 concludes.

2 Merger Data

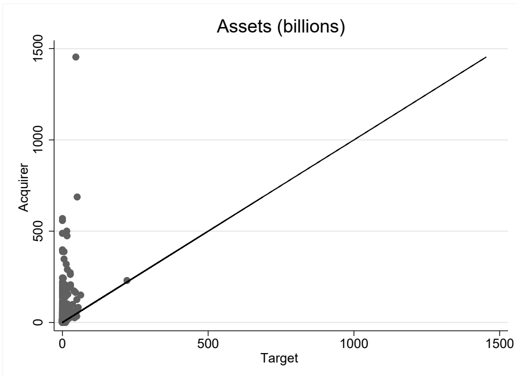
We present new empirical evidence establishing that the majority of mergers are by big (dominant) banks acquiring small (fringe) banks. This motivates the merger stage we introduce below in our model.

Below, we compare the size (book assets and market capitalization) of targets and acquirers between and 2023.[5] Panel (a) in Figure 2 shows that 96% of bank mergers involve an acquirer that is significantly larger than its target and the median acquirer has over seven times more assets than the target. Panel (b) in Figure 2 shows that 99% of bank mergers reported in Capital IQ involve in acquirer that has a larger market capitalization than its target and the median acquirer has a market capitalization that is seven times that of the target.

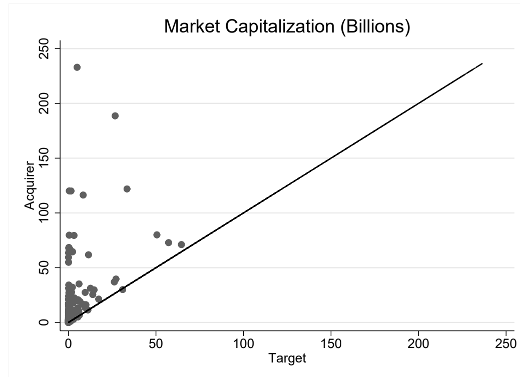
[5] We obtain information on merger activity from the "transformation" table and eliminate transactions between banks that belong to the same top holding company.

Figure 2: Bank Mergers 1990-2023

(a) Size of Acquirers vs Targets (Assets)

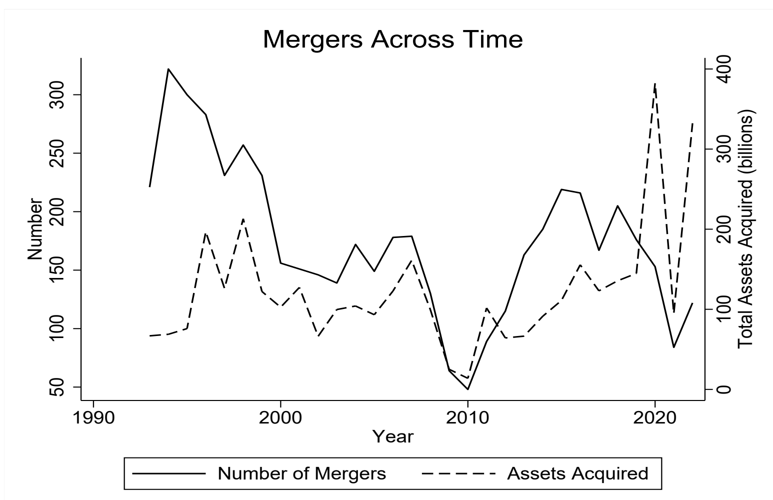


(b) Size of Acquirers vs Targets (Market Cap)



Note (Left Graph): We only include mergers where we can identify both the acquirer and target. We combine mergers that occur in the same quarter where there are multiple RSSDs with the same target holding company acquired by the same acquirer. Source: Statistics on Depository Institutions (SDI) and Federal Financial Institutions Examination Council (FFIEC). Note (Right Graph): We restrict to mergers from 1990 to 2023 where both the acquirer and target are banks as defined by SIC code. Source: Capital IQ

Figure 3: Number of Mergers over time



Note: We only include mergers where we can identify both the acquirer and target. We combine mergers that occur in the same quarter where there are multiple RSSDs with the same target holding company acquired by the same acquirer. Source: Statistics on Depository Institutions (SDI) and Federal Financial Institutions Examination Council (FFIEC).

Figure 3 makes evident that the number of mergers declined since the peak after the introduction of Riegle-Neal to the financial crisis. It rebounded slightly but has generally remained below pre-crisis levels. The total amount of assets acquired each year has remained relatively leveled over time as despite declining mergers, both acquirers and targets have grown over time.

3 Granularity and Financial Stability

Given the growth of the largest banks over the past 30 years, we study how idiosyncratic bank shocks to deposit flows or profitability can spill over to the aggregate economy via changes in credit. In the language of Gabaix (2011), and consistent with our model, we treat the largest bank in our model economy as a "granular" bank. In this granular view, an idiosyncratic shock to a dominant banks has the potential to generate nontrivial aggregate fluctuations.

Gabaix (2011) focused on how idiosyncratic shocks to the growth rate of the largest 100 firms can explain 30% of aggregate output fluctuations. We perform a similar empirical analysis by focusing on the largest commercial banks. More specifically, let x_{it} the log of the variable of interest for bank i in period t (for example, real deposits). The growth rate of x_{it} is $g^x_{it} = x_{it} - x_{it-1}$. As in Gabaix (2011), we assume that g_{it} evolves as

$$g^x_{it} = \beta' X_{it} + \epsilon^x_{it}, \tag{1}$$

where X_{it} is a vector of factors that may depend on bank characteristics at time $t-1$ and of factors at time (e.g., a time fixed effect or other aggregate variables). The goal is to investigate whether the purely idiosyncratic component ϵ^x_{it} of large banks can explain fluctuations in aggregate credit and aggregate output. As in Gabaix (2011), we construct the *granular residual* as follows

$$\Gamma^x_t \equiv \sum_{i=1}^{K} \frac{\ell_{it-1}}{L_{t-1}} \epsilon^x_{i,t}, \tag{2}$$

where K represents the K top banks when ranked by assets, ℓ_{it} is total lending by bank i in period t and $L_t = \sum_{i=1}^{N} \ell_{it}$ (N is the total number of banks). We estimate how much of aggregate fluctuations in credit and output can be explained by Γ^x_t.

To implement this, we estimate (1) and obtain $\hat{\epsilon}^x_{it} = g^x_{it} - \hat{\beta}' X_{it}$ to then construct

$$\hat{\Gamma}^x_t = \sum_{i=1}^{K} \frac{\ell_{it-1}}{L_{t-1}} \hat{\epsilon}^x_{it}. \tag{3}$$

We then regress aggregate loan growth and GDP growth on $\hat{\Gamma}^x_t$ and some lags and evaluate the value of the R^2. If only aggregate shocks were important (i.e., shocks to dominant banks would not influence aggregate fluctuations) the R^2 would be low. Table 1 presents the results when x_{it} corresponds to real deposits and aggregate fluctuations are measured using aggregate lending growth. We find that R^2 are relatively high with dominant bank idiosyncratic shocks explaining between 15% and 28% of aggregate lending growth. Changes in the predicted power of idiosyncratic shocks appear to increase at a smaller pace when we move from Top 35 to Top 100 than when we go from Top 10 to Top 35 banks.

Table 1: Explanatory power of granular residuals of dominant banks on aggregate lending (R^2)

	Dep. Var. $\Delta \log(L_t)$					
	Top 10 Banks		Top 35 Banks		Top 100 Banks	
(intercept)	0.00774***	0.00772***	0.00788***	0.00794***	0.00775***	0.00779***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Γ_t^x	0.220***	0.219***	0.328***	0.326***	0.381***	0.376***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Γ_{t-1}^x	-0.0221	-0.0225	-0.0363	-0.0365	-0.0335	-0.0314
	(0.567)	(0.562)	(0.420)	(0.419)	(0.466)	(0.495)
Γ_{t-2}^x		0.0118		0.0391		0.0464
		(0.761)		(0.388)		(0.315)
N	191	190	191	190	191	190
R^2	0.149	0.149	0.224	0.227	0.273	0.277
Adjusted R^2	0.140	0.136	0.215	0.214	0.265	0.265

Note: Table presents the R^2 from a regression of the corresponding dependent variable on Γ_t^x, Γ_{t-1}^x, and Γ_{t-2}^x when x is real deposits. "$\Delta \log(L_t)$" refers to growth rate of total real loans. Top 10 Banks, Top 35 Banks, and Top 100 Banks correspond to the cases when $K = 10$, $K = 35$, and $K = 100$, respectively. We let $X_{it} = \bar{g}_t = N^{-1} \sum_{i=1}^N g_{it}^x$, so $\hat{\Gamma}_t^x = \sum_{i=1}^K \frac{\ell_{it-1}}{L_{t-1}}(g_{it}^x - \bar{g}_t)$. Source: Call Reports

Table 2 presents the results when x_{it} corresponds to real deposits and aggregate fluctuations are measured using the growth of real total credit to the private non-financial sector.[6] We find that the R^2 declines compared to those presented in Table 1 but are still sizeble. In particular, we find that granular bank idiosyncratic shocks explaining up to 15.2% of the fluctuations in aggregate credit.[7] As before, we find that, while still increasing, the effect appears to flatten out when going beyond the Top 35 banks.

[6]We use data from the Federal Reserve Bank of St. Louis "Total Credit to Private Non-Financial Sector, Adjusted for Breaks, for United States (QUSPAMUSDA)" (see here.)

[7]Results are similar when using the total credit to the non-financial sector normalized by GDP.

Table 2: Explanatory power of granular residuals of dominant banks on aggregate credit (R^2)

	Dep. Var. $\Delta \log(totcredit_t)$					
	Top 10 banks		Top 35 banks		Top 100 banks	
(intercept)	-0.00589***	-0.00577***	-0.00579***	-0.00564***	-0.00583***	-0.00568***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Γ_t^x	0.101***	0.101***	0.167***	0.166***	0.212***	0.208***
	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Γ_{t-1}^x	0.0468	0.0471	0.0523	0.0528	0.0677*	0.0709*
	(0.145)	(0.144)	(0.172)	(0.169)	(0.086)	(0.073)
Γ_{t-2}^x		0.0222		0.0373		0.0545
		(0.490)		(0.333)		(0.169)
N	191	190	191	190	191	190
R^2	0.061	0.063	0.100	0.105	0.143	0.152
Adjusted R^2	0.051	0.048	0.091	0.091	0.134	0.138

Note: Table presents the R^2 from a regression of the corresponding dependent variable on Γ_t^x, Γ_{t-1}^x, and Γ_{t-2}^x when x is real deposits. "$\Delta \log(totcredit_t)$" refers to growth rate of real total credit to private non-financial sector. Top 10 Banks, Top 35 Banks, and Top 100 Banks correspond to the cases when $K = 10$, $K = 35$, and $K = 100$, respectively. We let $X_{it} = \overline{g}_t = N^{-1} \sum_{i=1}^N g_{it}^x$, so $\hat{\Gamma}_t^x = \sum_{i=1}^K \frac{\ell_{it-1}}{L_{t-1}} (g_{it}^x - \overline{g}_t)$. Source: FRED and Call Reports

Table 3 presents the results when x_{it} corresponds to real deposits and aggregate fluctuations are measured using the growth of real gdp. We find that R^2 are smaller but still significant with dominant bank idiosyncratic shocks explaining up to 4% of fluctuations in real gdp.[8]

[8]Table 3 in the Appendix shows that result also hold when using the growth rate of detrended GDP as the dependent variable.

Table 3: Explanatory power of granular residuals of dominant banks on aggregate output (R^2)

	Dep. Var. $\Delta\log(gdp_t)$					
	Top 10 banks		Top 35 banks		Top 100 banks	
(intercept)	0.00645***	0.00678***	0.00644***	0.00673***	0.00652***	0.00677***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Γ_t^x	-0.0201	-0.0224	-0.0278	-0.0323	-0.0345	-0.0423
	(0.522)	(0.472)	(0.469)	(0.397)	(0.395)	(0.292)
Γ_{t-1}^x	-0.0298	-0.0293	-0.0407	-0.0398	-0.0267	-0.0211
	(0.344)	(0.347)	(0.289)	(0.295)	(0.510)	(0.598)
Γ_{t-2}^x		0.0738**		0.0876**		0.102**
		(0.019)		(0.023)		(0.012)
N	191	190	191	190	191	190
R^2	0.007	0.036	0.009	0.036	0.006	0.039
Adjusted R^2	0.000	0.020	0.000	0.021	0.000	0.024

Note: Table presents the R^2 from a regression of the corresponding dependent variable on Γ_t^x, Γ_{t-1}^x, and Γ_{t-2}^x when x is real deposits. "$\Delta\log(gdp_t)$" refers to growth rate of real gdp. Top 10 Banks, Top 35 Banks, and Top 100 Banks correspond to the cases when $K = 10$, $K = 35$, and $K = 100$, respectively. We let $X_{it} = \overline{g}_t = N^{-1}\sum_{i=1}^{N} g_{it}^x$, so $\hat{\Gamma}_t^x = \sum_{i=1}^{K} \frac{\ell_{it-1}}{L_{t-1}}(g_{it}^x - \overline{g}_t)$. Source: Call Reports

In sum, we find evidence for a granular view (in the language of Gabaix (2011)) of how idiosyncratic shocks to bank deposit flows of the largest U.S. banks spill over to the aggregate economy via changes in credit. Specifically, in table 1 we document that idiosyncratic shocks to top 10 and top 35 bank deposits can explain 15% and 23% of aggregate bank lending growth (measured via R^2), respectively. In addition, in table 2, we document that granular shocks to systemically important banks (i.e. the top 35) explain up to 10% of the fluctuations in real total credit to the private non-financial sector. Finally, in table 3 we document that granular shocks to systemically important banks (i.e. the top 35) explain up to 4% of the growth of real GDP.

4 Model

Having documented that the largest banks buy smaller banks and that idiosyncratic shocks to the largest banks can have aggregate effects, here we present a version of a dominant-fringe model consistent with other data facts for the banking industry. We think of the dominant-fringe framework as a simple representation of a granular banking industry. Unlike papers that focus solely on the deposit side, we also study how bank mergers affect banking lending. As a result, our model must also account for the fact that growth of the nonbank sector has put further competitive pressure on the banking industry.

4.1 Environment

We consider an infinite horizon, discrete time model but to save on notation we use the nomenclature that variable x_t is denoted x and x_{t+1} is denoted x'. Each period begins with a measure γ of identical atomistic (i.e. measure zero) fringe banks each with D_f deposits as well as a non-atomistic (representative) dominant bank with D_d deposits. While we take D_f as parametrically given, D_d can expand endogenously through the acquisition of fringe banks. Specifically, to expand its deposit base, the dominant bank can make a take-it-or-leave-it (TOLI) offer to a measure $(\gamma - \Gamma)$ of fringe banks. Assuming all fringe banks accept, the measure of fringe banks falls from γ to Γ and the deposits of the dominant bank increases by $(\gamma - \Gamma)D_f$. The dominant bank chooses the size of the merger Γ to maximize its equity value. The cost of deposits, r_D, is exogenous and calibrated to the average deposit rate.

Following the merger stage, the dominant bank, the fringe banks, and the representative non-bank Cournot compete in the loan market. Banks choose between allocating their deposits to the loan market or securities, where the latter provides an exogenous risk free return of $r_A > r_D$. Ex-ante identical borrowers demand one-period loans in order to fund a risky project. The borrower decides whether they want to fund their project given an outside option, and then makes a discrete choice over whether to borrow from a bank B or a nonbank N. Using the discrete choice approach from IO, these borrowers create an aggregate demand for loans. Dominant banks move first followed by the fringe banks and the nonbank. Market clearing and the nonbank's first-order condition determine equilibrium loan rates (r_B^L, r_N^L).

Following the loan stage, loan defaults (θ') and chargeoffs (λ') are realized in the payoff stage. We assume that θ' follows a persistent Markov process with distribution $G(\theta', \theta)$, that $\lambda'_d = \lambda'_N = \overline{\lambda} \in [0, 1]$, and that $\lambda'_f \in [0, 1]$ is iid over time and across fringe banks distributed with a truncated normal $F(\lambda; \overline{\lambda}, \sigma_\lambda)$ where $\overline{\lambda}$ and σ_λ denote the mean and the standard deviation of the distribution, respectively.[9] We assume banks with non-negative profits pay them as dividends (there are no retained earnings). Banks with negative profits can choose to exit or inject costly equity to remain in the market. New banks can choose to enter the fringe at cost K_f. Endogenous exit and entry (γ_e and γ_x) imply that the measure of fringe banks can change through time. In sum, the size of the fringe next period is given by $\gamma' = \Gamma + \gamma_e - \gamma_x$.

We model merger regulatory policy as follows. First, in the pre-Riegle-Neal restrictions, banks were restricted to operating in one state. As such, we assume the cost of merging such that their deposit market share exceeds 25% (the pre-Riegle-Neal deposit market share of the top 10 banks) is infinite. Second, to capture costs associated with HHI constraints as shown in Nocke and Whinston (2022), we compute an implied HHI from the model and the cost associated from the merger is given by the following functional form:

$$H(\Gamma, s) = \begin{cases} \infty & \text{if pre-Riegle-Neal and } \frac{D'_d}{D_d + \gamma D_f} > 0.25 \\ 2h(\gamma - \Gamma)^2 & \text{if } HHI > 1800 \\ h(\gamma - \Gamma)^2 & \text{if } HHI < 1800 \text{ and } \Delta_{HHI} > 100 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

[9] This is consistent with the evidence presented in Table A.1 in the Appendix. Specifically, the standard deviation of loan loss rate is lower for top 10 banks than fringe banks in both pre- and post-reform periods.

Note that these costs change with the endogenous market structure of the banking sector as our dominant bank buys up a fraction $\gamma - \Gamma$ of fringe banks. In summary, the timing in any period is as follows:

M1. Merger Stage: The dominant bank makes a take-it-or-leave-it merger bid to a measure $(\gamma - \Gamma)$ of fringe banks subject to regulatory costs of doing so in order to grow its deposit base.

M2. Loan Stage: Banks allocate their deposits to the loan market or securities. Dominant banks move first in the loan market followed by the fringe banks and nonbanks.

M3. Payoff Stage: At the end of the period, loan defaults and charge-offs are realized. Exit and entry decisions are made and dividends paid.

4.2 Merger Equilibrium

Aggregate deposits in the banking sector are $D = D_d + \gamma D_f$. Starting in aggregate state $s = (\gamma, D_d, \theta)$, the dominant bank makes a TOLI offer to a measure $\gamma - \Gamma \geq 0$ of fringe banks to gain deposits. Assuming all fringe banks accept, the measure of the fringe banks after the merger adjusts to Γ and the deposits of the dominant bank is given by

$$D'_d = D_d + (\gamma - \Gamma)D_f \tag{5}$$

If the price of a unit of deposits in the merger stage is denoted $p(\Gamma, s)$, then the total cost of the merger deal to the dominant bank is given by $p(\Gamma, s)(\gamma - \Gamma)D_f$. Since mergers (weakly) increase the dominant bank's deposit market share, we can match the data fact of rising dominant bank market share and falling number of banks evident in figure 1. The latter result - the falling number of banks - arises as the endogenous measure of fringe banks evolves in our model which is absent in Gowrisankaran and Holmes (2004).

The dominant bank chooses a deposit market share to maximize its profitability recognizing that the fringe banks must be willing to accept their offer of size $p(\Gamma, s)(\gamma - \Gamma)D_f$. Specifically, we write the dominant bank's dynamic programming problem at the beginning of the merger stage as:

$$v_d(s) = \max_{\Gamma \in [0,\gamma]} -p(\Gamma, s)(\gamma - \Gamma)D_f - H(\Gamma, s) + w_d(\Gamma, D'_d, \theta) \tag{6}$$

subject to

$$p(\Gamma, s)D_f \geq w_f(S) \tag{7}$$

where $H(\Gamma, s)$ are any additional regulatory costs and $w_i(S)$ is the expected present discounted value of a type $i \in \{d, f\}$ bank in loan stage state $S = (\Gamma, D'_d, \theta)$. Constraint (7) incentivizes the fringe to accept the take-it-or-leave-it offer of the dominant bank.

Taking a first order condition with respect to Γ and assuming the constraint binds, we get:

$$\frac{\partial w_d(\Gamma)}{\partial \Gamma} = \frac{\partial w_f(\Gamma)}{\partial \Gamma}(\gamma - \Gamma) + \frac{\partial H(\Gamma)}{\partial \Gamma} - w_f(\Gamma, D'_d(\Gamma), \theta) \tag{8}$$

where we denote $\frac{\partial f(\Gamma, D'_d(\Gamma),\theta)}{\partial \Gamma} + \frac{\partial f(\Gamma, D'_d(\Gamma),\theta)}{\partial D'_d} D_f$ as $\frac{\partial f(\Gamma)}{\partial \Gamma}$. The left hand side is the marginal benefit for the dominant firm. It depends on how the acquisitions limits competition and improves the dominant bank profitability. The right hand side is the marginal cost of the merger, which incorporates both the marginal cost due to greater merger regulation (H) and due to rising fringe value functions (w_f).

Let $\Gamma^*(s)$ solve (6)-(7) noting that if the dominant bank chooses not to acquire any fringe banks, then $\Gamma^*(s) = \gamma$. Since, in the loan stage, the dominant bank will use its market power to withhold loans, fringe banks recognize that the more mergers occur, the more valuable it is to remain in the market. As a result, the dominant bank faces a trade-off between acquiring a few fringe banks at a low price or many fringe banks at a high price.

Following the merger stage, the dominant bank, fringe banks, and a representative nonbank compete in the loan market. Ex-ante identical borrowers demand one-period loans to fund a risky project. The project requires one unit of investment from either a bank B or nonbank N. Every period, given r^L_B, r^L_N, and shock ω, borrowers choose whether to invest ($\iota = 1$) or not ($\iota = 0$).

$$\max_{\{\iota\}} (1-\iota)\cdot\omega + \iota \cdot E_{\boldsymbol{\epsilon}}[\Pi_E(\theta, r^L_B, r^L_N, \boldsymbol{\epsilon})] \tag{9}$$

Conditional on choosing $\iota = 1$, entrepreneurs observe $\boldsymbol{\epsilon} = \{\epsilon_B, \epsilon_N\}$ and then choose which type of lender $k \in \{B, N\}$ to borrow from to solve:

$$\Pi_E(\theta, r^L_B, r^L_N, \epsilon) = \max_{k\in\{B,N\}} \alpha E_{\theta'|\theta}[\pi(r^L_k, \theta)] + \epsilon_k \tag{10}$$

where

$$\pi_E(r^L_k, \theta') = \begin{cases} \max\{0, R - r^L_k\} & \text{with prob } \theta' \\ \max\{0, -(\lambda' + r^L_k)\} & \text{with prob } 1-\theta' \end{cases}.$$

The solution to (10) implies that the share of borrowers choosing a loan from a lender of type k is

$$\Psi_k(\theta, r^L_B, r^L_N) = \frac{\exp\left(\alpha E_{\theta'|\theta}\left[\pi_E(r^L_k, \theta')\right]\right)}{\sum_{\hat{k}\in\{B,N\}} \exp\left(\alpha E_{\theta'|\theta}\left[\pi_E(r^L_{\hat{k}}, \theta')\right]\right)}. \tag{11}$$

The ex-ante value for the borrower is given by:

$$U_B(\theta, r^L_B, r^L_N) = \frac{1}{\alpha} E_\omega[\iota] E_\epsilon[\Pi_E(\theta, r^L_B, r^L_N, \epsilon)] \tag{12}$$

The borrowers problem creates a demand system that determines for a given level of loans by the banking system L_B and nonbank interest rate r^L_N, the requisite r^L_B that clears the market.

Following the loan stage, loan defaults (θ') and chargeoffs (λ) are realized. The profit of bank $i \in \{d, f\}$ is given by:

$$\pi_i(L_i; S, \theta', \lambda'_i) = [\theta' r^L_B - (1-\theta')\lambda'_i] L_i + r^A(D_i - L_i) - r^D D_i - C_i(L_i) - \kappa_i. \tag{13}$$

where $C_i(L_i)$ is the cost of loan monitoring. We assume $C_i(L_i)$ takes the following form:

$$C_i(L_i) = C_{1i}L_i + C_{2i}L_i^2 \tag{14}$$

Further, we denote dividends by:

$$\mathscr{D}_i(L_i; S, \theta', \lambda_i') = \pi_i(L_i; S, \theta', \lambda_i') - \mathbf{1}_{\{\pi_i(\cdot)<0\}}\psi_i(|\pi_i(L_i; S, \theta', \lambda_i')|). \tag{15}$$

where ψ_i captures the costs of equity issuance.

The dynamic programming problem of the dominant bank can then be written as:

$$w_d(S, r_N^L) = \max_{L_d \leq D_d'} \mathbb{E}_{\theta'|\theta}\Big[\max_{x_d' \in \{0,1\}} (1 - x_d')\big(\mathscr{D}_d(L_d; S, \theta', \lambda_d') + \beta v_d(s')\big)\Big], \tag{16}$$

subject to

$$\begin{aligned} L_d + \Gamma L_f^*(L_d, S, r_N^L) &= L_B(r_B^L, r_N^L), && (17)\\ \gamma' &= F(L_d, S, \theta'). && (18) \end{aligned}$$

where $s' = (\gamma', D_d', \theta')$ and the function $F(L_d, S, \theta')$ captures the transition of the measure of fringe banks. Equation (17) illustrates both loan market clearing and that the dominant bank takes into account both how their loan decision affects the fringe loan decision, L_f, and the bank interest rate, r_B^L. Equation (18) ensures that the dominant bank internalizes its impact on market structure in the future.

Given the stackelberg game in the loan market, taking the interest rates r_B^L and r_N^L as given, the problem of the fringe bank solves[10]

$$w_f(L_d, S, r_N^L) = \max_{\ell_f \leq D_f} \mathbb{E}_{\theta', \lambda_f'|\theta}\Big[\max_{\chi_f' \in \{0,1\}} (1 - \chi_f')\big(\mathscr{D}_f(\ell_f; L_d, S, \theta', \lambda_f') + \beta v_f(s')\big)\Big] \tag{19}$$

Turning to non-bank competition, the profits of nonbank N are given by:

$$\pi_N(\ell_N, S, \theta', \lambda') = [\theta' r_N^L - (1 - \theta')\lambda' - c_N]\ell_N \tag{20}$$

The first order condition of the non-bank with respect to ℓ_N is given by

$$r^D = E_{\theta'|\theta}\left[\theta' r_N^L - (1 - \theta')\lambda'\right] - c_N. \tag{21}$$

Banks with negative profits can choose to exit or inject equity to remain in the market. New fringe entrants choose whether to enter. The entry decision satisfies

$$v_i^e(L_d, S, r_N^L, \theta'; \gamma^e(x_d', e_d')) = \max_{e_i' \in \{0,1\}} (1 - e_i')\left\{-K_i + v_i(s')\right\}, \tag{22}$$

The evolution of the mass of fringe banks then is given by

$$\gamma' = F(L_d, S, \theta') = \Gamma - \gamma^x(L_d, S, \theta') + \gamma^e(L_d, S, \theta'). \tag{23}$$

[10] To save on notation, we neglect adding L_f and x_f as state variables in (19).

Definition 1. *Taking r^A and r^D as given, a **Markov Perfect Merger Equilibrium** is a set of value functions $\{v_i, w_i\}$ and policy functions $\{\Gamma, L_i, \ell_f, x'_i, \chi'_f, e'_i\}$ for $i \in \{d, f\}$, ℓ_N, prices $\{p, r^L_B, r^L_N\}$, and transition functions for $\{\gamma', D'_d\}$ such that:*

1. *The pre-merger value function v_d solves* (6)*. The merger quantity Γ maximizes* (6)*.*
2. *The merger pricing function p satisfies* (7)*.*
3. *The post-merger value functions w_d and w_f solve* (16) *and* (19)*. The loan supply policy functions (L_d, ℓ_f) and exit decision rules (x'_d,χ'_f) maximize* (16) *and* (19)*.*
4. *Consistency requires $\ell_f = L_f$ and $\chi'_f = x'_f$.*
5. *r^L_B clears the loan market* (17)*.*
6. *r^L_N satisfies the shadow bank first order condition (21)*
7. *The mass of entrants γ^e solves the entry problem (22)*
8. *Transition functions are consistent with mergers, entry, and exit (23)*

4.3 Multiple Steady States and Equilibria

As in Gowrisankaran and Holmes (2004), we have the possibility of multiple steady states. In Gowrisankaran and Holmes (2004), they find three steady states, one with no dominant firm, one with no fringe firms, and one with both a mass fringe firms and a dominant firm. Given that we have entry, as long as entry costs are sufficiently small, we always have a mass of fringe banks in any steady state. Given the large fixed costs for dominant banks, we find steady states with no dominant bank. We do not focus on these because banking is defined by a few large, dominant banks. Finally, we find a steady state with both a dominant bank and a mass of fringe banks as illustrated in table 6.

5 Calibration

Our calibration strategy is as follows. While all endogenous variables defined in our equilibrium are derived from an economy with aggregate uncertainty about borrower failure rates $1 - \theta$, we calibrate parameters for our Pre-Riegle-Neal long run equilibrium using a realization of aggregate shocks which do not include the small probability crisis event since our annual data span only 10 years prior to Riegle-Neal. The Pre-Riegle-Neal long run equilibrium assumes that $H(\Gamma, s) = \infty$ for all s which is how we enforce the state branching restrictions. In the transition between Pre-Riegle-Neal and Post-Dodd-Frank we simply relax the assumption $H(\Gamma, s) = \infty$ in equation (4) maintaining all other parameter values. We calibrate h to match the post-Dodd-Frank dominant bank market share. Thus, our transition path should be seen as a validation exercise. For the post-Dodd-Frank period, we maintain all of the previous parameters except for "regulatory" costs associated with the Dodd-Frank Act. Specifically, we choose the fixed costs to fringe and dominant banks, κ_f and κ_d respectively, to match rising fixed costs to loans. Further, we choose the marginal cost to extending loans by non-banks c_N to match the declining bank share of loans also in the post-Dodd-Frank data. All parametric changes are unexpected from the point of view of banks (i.e. implemented as MIT shocks).

We calibrate the model parameters via Simulated Method of Moments to match key statistics of the U.S. banking industry described in Corbae and D'Erasmo (2025). 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").[11] We aggregate commercial bank level information to the Bank Holding Company level. As discussed above, moments from the Call Report data are computed beginning in 1984 (due to an overhaul of the data in that year). We also use data from the Summary of Deposits (SOD), and FRED, federal reserve economic data for aggregate economic series (GDP, CPI, etc).

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

There are two types of banks $i \in \{d, f\}$ in our model. Consistent with the data size differences we described in the data section, we identify Dominant banks $i = d$ with those in the Top 10 (when banks are sorted by assets) and the rest (i.e., the competitive fringe $i = f$) with those outside the Top 10.

We parameterize the stochastic process for the aggregate likelihood of default with two states $\theta \in \{\theta_G, \theta_B\}$, where $\theta_G > \theta_B$. We calibrate θ to match the average default probability.

We calibrate $\overline{\lambda}$ to match the average charge-off rate (across banks and time) and set σ_λ to match the difference between the observed standard deviation of the charge-off rates of the fringe banks and that of dominant banks.[12] $\lambda'_d = \lambda'_N \equiv \overline{\lambda} \in [0, 1]$, $\lambda'_f \in [0, 1]$ is iid over time and across fringe banks distributed with a truncated normal $\mathcal{N}(\lambda; \overline{\lambda}, \sigma_\lambda)$ where $\overline{\lambda}$ and σ_λ denote the mean and the standard deviation of the distribution, respectively.

We assume $C(L_i)$ is a quadratic function of L_i with linear term C_i^1 and quadratic term C_i^2 and calibrate it to match the bank marginal expenses and the elasticity of net marginal expenses.

We let the discount factor be $\beta = 1/(1 + r^D)$ and calibrate the deposit interest rate r^D to match the ratio of interest expenses on deposits over total deposits. Similarly, we calibrate r^A to match the ratio of interest income on securities over total securities. We let the external seasoned equity financing cost take the following form $\psi_i(x) = \psi_i^1 x$ for $i \in \{d, f\}$ and calibrate it to the average equity issuance to asset ratio by banks of type $i \in \{d, f\}$.

We calibrate loan demand parameters α, R, N, and $\bar{\omega}$ to match the elasticity of loan demand, net interest margin, the deposit-to-output ratio, and average dividend issuance. We normalize $\underline{\omega}$ to be the value of never borrowing (i.e. the value of the borrower at the highest possible interest rates from both lender types $E_\epsilon[\Pi_E(\theta, r_B^L = R, r_N^L = R, \epsilon)]$).

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

[11]Source: FDIC, Call and Thrift Financial Reports, Balance Sheet and Income Statement, link here.

[12]See table A.1 in the Appendix. Specifically, the standard deviation of loan loss rate is lower for top 10 banks than fringe banks in both pre- and post-reform periods which is consistent with a diversification argument.

solving the model). After those are set, a second group is estimated using simulated method of moments. In what follows, we describe both groups of parameters as well as our targeted moments. Table 4 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 17 parameters and 23 targeted moments as part of the simulated method of moments estimation (i.e. an overidentified model).

Table 4 presents all the parameters of the model. Table 5 presents the parameters that change after Dodd-Frank's implementation. We increase fixed costs for both dominant and fringe banks to capture the rising fixed cost to loan ratio. We decrease the marginal cost of the nonbank to capture the declining share of bank loans to total loans. Since there were minimal changes to FDIC bank merger policy during this period, h is unchanged. Table 6 present a set of data moments (targeted and not targeted) together with their model generated counterparts.

Table 4: Parameters and Targets

Parameter		Value	Target
Deposit Interest Rate (%)	$\bar{r} = r^D$	0.0014	Avg Interest Expense Deposits
Bank Discount Factor	β	0.998	$(1 + r^D)^{-1}$
Return on Securities	r_A^+	0.0180	Return on Net Securities
Default Frequency (Good Times)	θ_G	0.969	Mean Default Frequency (Non-Crisis)
Default Frequence (Bad Times)	θ_B	0.80	Mean Default Frequency (Financial Crisis)
Loan Loss Rate	$\bar{\lambda}$	0.31	Average Charge-off rate
Standard Dev Loan Loss rate	σ_λ	0.15	Difference std dev charge off rates d vs f
Number of Borrowers	N	8.5	Deposit to Output
Return on investing	R	0.20	Net Interest Margin
Price coefficient	α	42.0	Elasticity of Loan Demand
Lower bound demand shock	$\underline{\omega}$	0.693	Normalization
Upper bound demand shock	$\bar{\omega}$	1.193	Dividend Issuance d and f
Mean size of Fringe Bank f	D_f	0.001	Relative Size Fringe to Top 10
Linear Cost Loans d	C_d^1	0.010	Net Marginal Expenses Top 10
Quadratic Cost Loans d	C_d^2	0.001	Elasticity Net Marginal Expenses Top 10
Fixed cost d	κ_d	0.0028	Fixed cost over loans Top 10
Mean Dist Cost Loans f	C_f^1	0.010	Net Marginal Expenses Fringe
Quadratic Cost Loans f	C_f^2	0.001	Elasticity Net Marginal Expenses Fringe
Fixed cost f	κ_f	$0.012 \times D_f$	Fixed cost over loans Fringe
External finance param. d	ψ_d^1	0.05	Avg. equity issuance to loan ratio Fringe
External finance param. f	ψ_f^1	0.50	Avg. equity issuance to loan ratio Fringe
Entry Cost f	K_f	$0.55 \times D_f$	Entry Rate
Regulatory Merger Cost	h	0.001	Post-Dodd-Frank Bank Market Share Top 10
Marginal Cost Nonbank	c_N	0.375	Bank Loan to Total Loans

Note: The entry cost is set as part of the equilibrium selection.

Table 5: Parameter Changes Post-Dodd-Frank

Parameter		Pre-Riegle-Neal	Post-Dodd-Frank
Fixed cost d	κ_d	0.0028	0.0035
Fixed cost f	κ_f	$0.012 \times D_f$	$0.028 \times D_f$
Marginal Cost nonbank	c_N	0.375	0.315

To illustrate the calibration, we present the moments in the data relative to our model output in both the pre (1984-1993) and post (2010-2019) periods.

Table 6: Moments Data vs Model (Targets)

Name	Calc	Data Pre	Data Post	Model Pre	Model Post
Average Charge-off rate	$E'_\theta[(1-\theta')\lambda']$	0.96%	0.94%	0.96 %	0.96 %
Elasticity of Loan Demand	$-\alpha r_B^L(1-s_B)$	-1.1	-1.1	-1.00	-1.39
Deposit Share Top 10 to Fringe	$\frac{D_d}{D_d+\gamma D_f}$	24.77%	57.79%	24.60%	55.26%
Loan Share Top 10 to Fringe	$\frac{L_d}{L_d+\gamma L_f}$	28.55%	52.86%	21.51%	40.68%
Bank Loans to Total Loans Ratio	$\frac{L_d+\gamma L_f}{L_n+L_d+\gamma L_f}$	44.54%	33.28%	51.51%	39.70%
Deposit to Output Ratio	$\frac{D_d+\gamma D_f}{\sum_{borrower} Output}$	39.01%	57.19%	46.09%	45.27%
Net Interest Margin	$E_\theta[\theta' r_B^L - r_D]$	4.94%	4.35%	4.63%	5.18%
Net Marginal Expenses Top 10	$\frac{c(L_D)}{L_D}$	1.15%	1.35%	1.03%	1.04%
Net Marginal Expenses Fringe	$\frac{c(L_F)}{L_F}$	2.00%	1.69%	1.00%	1.00%
Elasticity Net Mg Expenses Top 10	$\frac{dC_D(L_d)}{dL_d}$	0.95%	1.03%	1.05%	1.08%
Elasticity Net Mg Expenses Fringe	$\frac{dC_f(L_f)}{dL_f}$	0.78%	0.84%	1.00%	1.00%
Fixed cost over loans Top 10	$\frac{\kappa_D}{L_D}$	0.89%	0.78%	1.02%	0.88%
Fixed cost over loans Fringe	$\frac{\kappa_F}{L_F}$	0.99%	5.83%	1.20%	2.80%
Equity issuance to assets Top 10	$\frac{\mathbf{1}_{\{\pi_d<0\}}(-\pi_d)}{D_d}$	0.01%	0.04%	0.03%	0.01%
Equity issuance to assets Fringe	$\frac{\mathbf{1}_{\{\pi_f<0\}}(-\pi_f)}{D_f}$	0.07%	0.13%	0.06%	0.10%
Bank Failure Rate Top 10	x_d	0.00%	0.00%	0.00%	0.00%
Bank Failure Rate Fringe	$\frac{\gamma_x}{\Gamma}$	0.76%	0.44%	0.03%	0.44%
Bank Merger Rate	$\frac{(\gamma-\Gamma)}{\gamma}$	1.27%	2.69%	0.00%	0.34%
Relative Size Dominant to Fringe	$\frac{D_f}{D_d}$	324.79	688.47	325.00	717.36
Dividends/Assets Top 10	$\mathscr{D}_d/D_d$	0.36%	0.74%	1.65%	1.99%
Dividend/Assets Fringe	$\mathscr{D}_f/D_f$	0.39%	0.66%	1.50%	0.43%
Loan Markup Top 10	$\frac{\theta r^L}{r_D+c'(L_d)}$	56.29%	205.37%	218.16%	258.73%
Loan Markup Fringe	$\frac{\theta r^L}{r_D+c'(L_f)}$	46.64%	149.73%	204.25%	135.03%
Interest Rate	r_B^L	6.69%	3.18%	4.93%	5.49%
Ratio of Loans to Deposits Top 10	$\frac{L_d}{D_d}$	83.3%	64.1%	84.18%	55.51%
Ratio of Loans to Deposits Fringe	$\frac{L_f}{D_d}$	73.3%	79.7%	99.99%	99.99%

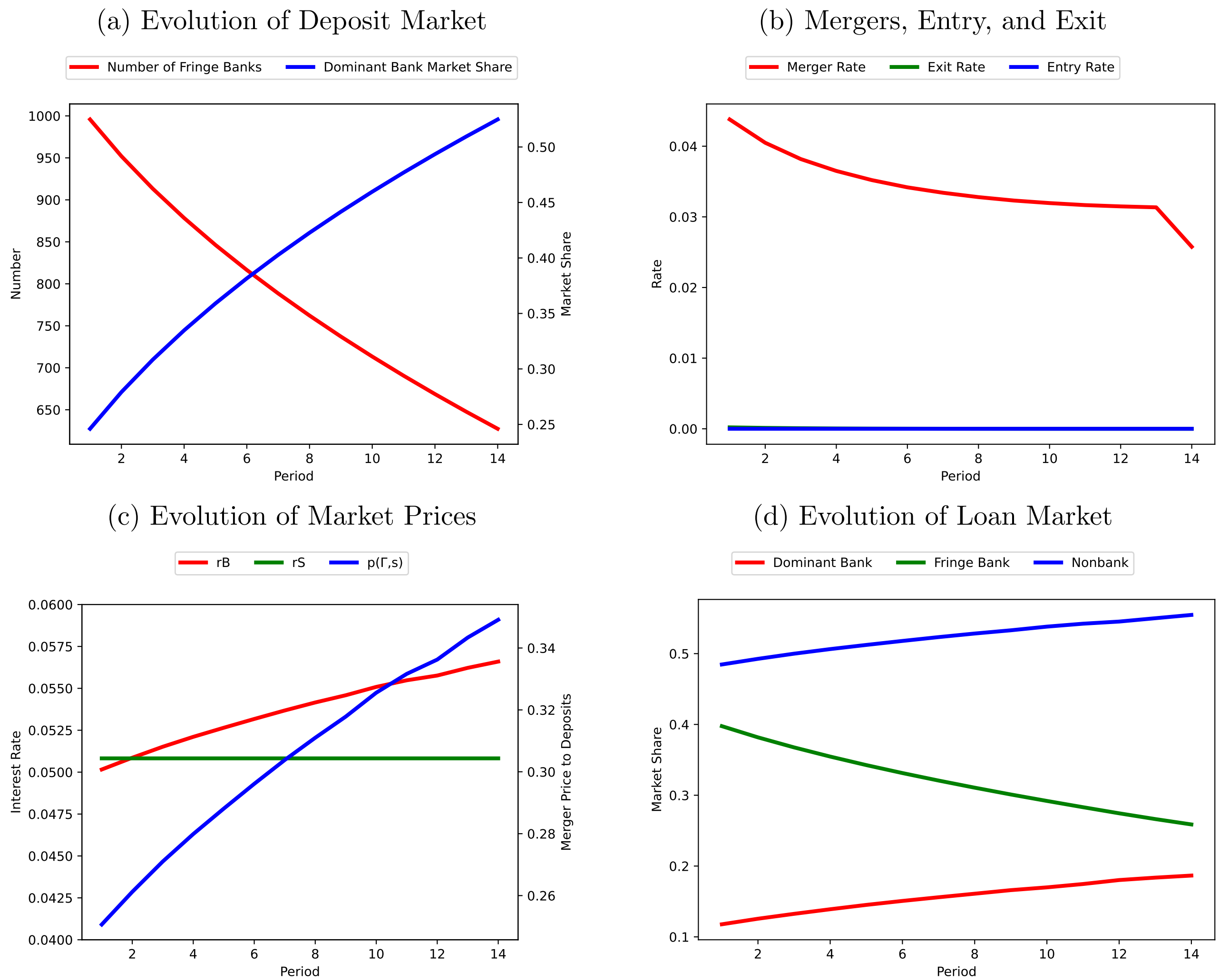
Note: Data Pre corresponds to the period 1984 - 1994 and Data Post to the period 2011 to 2019. Moments below the line are not-targeted moments.

6 Results

6.1 Validation and Transition Path

To illustrate that our model captures the appropriate merger dynamics in the transition period, we present the evolution of the endogenous state variables during the 15 year transition period between Riegle-Neil and the post-crisis Dodd-Frank regulations.

Figure 4: Model Transition



In Figure 4a, we illustrate the growth of the dominant bank and the decline in the number of fringe banks, matching the transition dynamics in panel (i) of Figure 1. The dominant bank grows from 25% deposit market share to 55% over 14 years, driven by an average 2% merger rate (as opposed to 1.84% in the data). The number of fringe banks declines substantially.[13]

In Figure 4b, we present the merger, exit, and entry rates, calculated as $\frac{(\gamma-\Gamma)}{\gamma}$, $\frac{\gamma_x}{\gamma}$, and $\frac{\gamma_e}{\gamma}$, respectively. The merger rate declines throughout the transition for three reasons. First, due to regulation (i.e. $H(\Gamma, s)$), the cost of acquiring fringe banks increases as the dominant bank gains market share and the deposit market HHI rises. Second, as in Gowrisankaran and Holmes (2004), fringe banks become more profitable as the prevailing interest rate r_B^L increases. As a result, the price to acquire fringe banks $p(\Gamma, s)$ rises. These two effects are both illustrated in Figure 4c. As the price of acquisition rises, the dominant bank chooses to slow acquisitions.

[13]Given that the dominant bank is matching data from the 10 largest banks, the initial size of the fringe (1000 banks) is close to the initial size of the fringe in the data (12000 banks)

Finally, given that we have entry and exit in the model, the threat of fringe entry may discourage the dominant bank from acquiring more fringe banks, as new entrants would prevent the dominant bank from gaining substantial market power. In Figure 4d, we illustrate the growth of the nonbank sector as a result of banking mergers. Since the dominant bank uses its market power to decrease the supply of loans and increase the prevailing interest rate, potential borrowers switch to nonbanks. As a result, the growth of the nonbank sector can partially be explained by mergers in the banking industry.

Comparing model results from the pre-Riegle-Neal long-run equilibrium to the post-Dodd-Frank long-run equilibrium suggests that the combination of mergers, changing bank fixed costs, and changing nonbank marginal costs, have lead to greater dominant bank market shares in both deposit and loan markets, higher interest margins, greater dominant bank value, lower fringe bank value, and had minimal effect on borrower value. In order to decompose these various changes driven by bank regulation, below we run policy counterfactuals. These counterfactuals allow us to answer the following questions i) what would happen if Riegle-Neal was never implemented? ii) How much of the nonbank sectors growth can be attributed to post-Riegle-Neal mergers as opposed to technological improvement?

7 Policy Counterfactuals

7.1 What if Riegle-Neal never happened?

We can use our model as a laboratory to run counterfactuals. One interesting counterfactual is, what would have happened to competition, market efficiency, and stability if Riegle-Neal was never implemented? To run this counterfactual, we obtain a new long-run equilibrium of the model as parameterized for the post-reform period but we restrict mergers as in the pre-reform period by setting $H(\Gamma, s) = \infty$. Since there are other differences between the pre-reform and the post-reform parameterization other than merger restrictions (i.e. (i) increased fixed costs for both dominant and fringe banks to reflect rising fixed costs in the data associated with Dodd -Frank bank regulation and (ii) declining nonbank marginal costs which can be thought of as technological innovation as a partial explanation for rising nonbank market share in the data), we need to decompose the numerous sources of change. The new long-run equilibrium associated with those changes are listed in "Post - No Mergers" in Table 8.

The solution to this version of the model with merger restrictions but a more efficient nonbank sector (relative to the pre-reform period) induces a significant increase in the share of credit originated by nonbank (as measured by one minus the ratio of bank loans to total loans). The decline in the post-reform model without merger restrictions equals 11.8 percentage points implying that the consolidation of the banking sector reduced the growth of nonbank lending by close to 58% (as the full decline from pre-reform to post-reform with merger restrictions equals 23.7 percentage points).[14] The more efficient nonbank sector together with the prevalence of a large set of competitive banks leads to a smaller increase in loan interest rates and interest margins and a smaller markup for both bank types (all relative to the changes observed between

[14]The change from pre-reform to post-reform represents 42.4% (11.8/23.7) and the rest corresponds to the potential gains by the nonbank sector absent bank mergers.

the pre-reform and the post-reform). Lower margins result in a decline in the value of the banks (as measured by Tobin's Q Value to Asset ratio) and an increase in failure rates. Since our focus is on how deposit growth funds bank lending, unlike papers solely focusing on effects on the deposit market, nonbank competition in the loan market is particularly relevant here. .

In addition to our main counterfactual experiment, we also evaluate how much of the changes between the pre-reform and the post-reform can be explained by the increase in efficiency in the nonbank sector. We implement this decomposition by obtaining the long-run equilibrium of the model as paramterized in the post-reform period but by keeping the marginal cost of the nonbank sector as in the pre-reform period. The results for that equilibrium are listed in "Post - c_N Same" in Table 8. We observe in this case that bank loans to total credit declines by less than the decline observed in the post-reform period. We can interpret this result as showing that the increase in concentration coupled with a decline in the efficiency of the banking sector (as banks' costs increase in the post-reform period) explain about 64% of the percentage point decline in bank lending.[15] The Post-Nonbank Same shows a larger increase in loan bank concentration which leads to higher markups, Tobin's Q, and lower failure rates (relative to the change observed in the post-reform). This comes at a substantial cost to borrowers as their utility declines by 3.1%. This also illustrates that the rise of nonbank lending can be partially attributed to bank mergers.

Table 7: Model With Mergers vs Model With Merger Restriction

Name	Pre	Post	Post - c_N Same	Post - No Mergers
Deposit Market Share Top 10 to Fringe	24.57%	55.27%	54.57%	29.18%
Loan Market Share Top 10 to Fringe	21.51%	40.68%	43.29%	27.85%
Bank Loans to Total Loans Ratio	51.52%	39.70%	43.92%	27.85%
Interest Rate	4.93%	5.49%	5.68%	5.08%
Net Interest Margin	4.64%	5.18%	5.37%	4.79%
Bank Failure Rate Top 10	0.00%	0.00%	0.00%	0.00%
Bank Failure Rate Fringe	0.03%	0.44%	0.00%	2.02%
Loan Markup Top 10	218.16%	258.73%	282.34%	211.93%
Loan Markup Fringe	204.25%	135.03%	139.78%	125.02%
Tobin's Q (Dominant)	27.08%	34.04%	38.26%	30.85%
Tobin's Q (Fringe)	23.40%	6.42%	34.54%	2.15%
Total Bank Sector Value	0.322	0.282	0.482	0.117
Borrower Value	0.00579	0.00580	0.00562	0.00590

Note: Model "Post" has no restriction on the size of the dominant bank. Model "Post - cN Same" does not alter the marginal cost of nonbanks. Model "Post - No Mergers" maintains the pre-Riegle-Neal restriction that the dominant bank cannot merge to greater than 25% deposit market share. Tobin's Q is computed as $\frac{V_i}{D_i}$ for banks of type i.

In sum, our results illustrate the costs and benefits of Riegle-Neal. Following the Riegle-Neal induced merger wave, bank profitability rose and financial stability improved. This came at the

[15]The difference between pre-reform and Post-Nonbank same equals 7.6pp which represents 64.3% of the decline observed between pre and post-reform (11.8pp).

cost of higher interest margins and higher markups for borrowers. However, both the costs and benefits were mitigated by the growing nonbank sector.

7.2 Allocative Efficiency

In order to understand optimal merger policy, we measure allocative efficiency under a variety of merger policies. Given large fixed costs, as illustrated in Table 6, mergers can improve allocative efficiency due to increasing returns to scale. However, due to market power and potential reductions in loan quantity, mergers could also exacerbate these fixed costs. We use the following decomposition of weighted average bank-level cost, as proposed by Olley and Pakes (1996).

$$\hat{c} \equiv \sum_{i \in \{D,F\}} C_i(L_i)\omega(L_i) = \bar{c} + Cov(C_i(L_i), \omega(L_i)) \tag{24}$$

where $\omega(L_i)$ is the loan market share of bank i. The loan weighted average cost can be decomposed into two terms: the un-weighted average of bank-level cost and a covariance term between loan shares and cost. A smaller value for the covariance term captures an improvement in allocative efficiency as a larger share of loans are provided by banks with lower costs.

Table 8: Allocative Efficency

Moment	Pre Riegle-Neal	Post Dodd-Frank	No Merger Regulation $h = 0$
Avg. (loan-weighted) cost $\hat{c}$	0.0293	0.0403	0.0352
Avg. cost $\bar{c}$	0.0300	0.0479	0.0479
$Cov(c, \omega)$	-0.0006	-0.0076	-0.0128
Total Bank Loans $L_d + \Gamma L_f$	1.2000	0.9174	0.7594

Note: Model "Pre Riegle-Neal" requires that the dominant bank cannot merge to greater than 25% market share. "Post Dodd-Frank" removes that restriction, but retains the other regulatory merger costs of $H(\Gamma, s)$. "No Merger Regulation" sets $h = 0$, removing all regulatory merger costs.

We find evidence for improved allocative efficiency follow Riegle-Neal as measured by the covariance of c and ω. The measure becomes more negative following Riegle-Neal as mergers decrease total fixed costs for the banking sector. We find that under a no-regulation equilibrium (i.e. $h = 0$) allocative efficiency further improves due to greater mergers. However, this allocative efficiency comes at the cost of a substantial reduction in bank lending.

7.3 Granular Regressions

We also can use our model to study how an idiosyncratic shock to the dominant bank can generate nontrivial aggregate fluctuations. Specifically, we implement an unexpected shock to dominant bank deposits before the merger stage, so that $D_{d,t} = D'_{d,t-1} + \epsilon_{dt}$. We then construct the *granular residual* as $\Gamma_t^{\hat{D}_d} = \frac{L_{d,t-1}}{L_{d,t-1}+\Gamma L_{f,t-1}}\epsilon_{dt}$. We then estimate how much aggregate fluctuations in credit can be explained by $\Gamma_t^{\hat{D}_d}$. To do so, we regress total lending growth on

$\Gamma_t^{\hat{D}_d}$ and its lags. We run these regressions under different regulatory regimes to understand how regulation affects the explanatory power of granular residuals. We use the same sequence of shocks for all regressions. Table 9 presents the regression results.

Table 9: Model based explanatory power of granular regressions

Coefficient	Pre Riegle-Neal	Post Dodd-Frank $h > 0$	No Merger Regulation $h = 0$
(intercept)	-0.0001	-0.0002	-0.0001
$\Gamma_t^{\hat{D}_d}$	0.4730	0.0013	0.0114
$\Gamma_{t-1}^{\hat{D}_d}$	-0.4673	0.0031	-0.0050
$\Gamma_{t-2}^{\hat{D}_d}$	0.4757	-0.0030	0.0059
N	190	190	190
R^2	0.0366	0.0114	0.0320

Note: Model "Pre Riegle-Neal" requires that the dominant bank cannot merge to greater than 25% market share. "Post Dodd-Frank" removes that restriction, but retains the other regulatory merger costs of $H(\Gamma, s)$. "No Merger Regulation" sets $h = 0$, removing all regulatory merger costs.

We find from Table 9 that the explanatory power is highest in the Pre Riegle-Neal period and in a counterfactual environment with no merger regulation. The reasons for this are two-fold. First, in the Pre Riegle-Neal environment, the dominant bank lends out a substantially higher fraction of its deposits (84% to 55%). As a result, shocks to the dominant bank's deposits have a greater effect on dominant bank lending. Second, in the absence of merger regulation, the dominant bank is much larger, and so shocks to the dominant bank translate to much larger aggregate shocks to lending.

7.4 Optimal Merger Policy

One important counterfactual left to compute as future work is to assess optimal dynamic regulatory merger policy along the lines of Nocke and Whinston (2010)? That is, what should $H(\Gamma, s)$ be to maximize market efficiency and minimize financial instability? Given our dynamic model, we can consider whether regulators should consider how mergers today affect the market structure not just tomorrow, but well into the future. The results in Pakes, Whinston, and Zheng (2024) show that consumer (household and firm) welfare fell significantly, economically and statistically, in single-merger counties compared to no-merger counties. Our model has the potential to quantify the optimal level of mergers by incorporating allocative efficiency and financial stability in addition to measures of consumer/borrower welfare.

8 Conclusion

In this paper, we study the effect of Riegle-Neal on banking mergers. We provide a model laboratory consistent with the ensuing large merger wave and, in the framework of Gabaix (2011), can contribute to spillovers by which idiosyncratic shocks to dominant (nonatomistic) banks can have substantial aggregate effects. Given that most bank mergers are done by acquirers that are substantially larger than their targets, we believe a dominant-fringe model suitably extended to capture essential elements of the banking sector is appropriate to study the effects of Riegle-Neal. The calibrated model allows us to perform counterfactuals to assess the effects of Riegle-Neal and increased nonbank competition on financial stability, bank profitability, and borrower surplus. We find that Riegle-Neal improved bank profitability and financial stability, at the cost of higher interest rates (lower efficiency) for borrowers. Our counterfactuals also show that rising non-bank competition has mitigated these effects. Our quantitative laboratory can provide regulators with a tool to set optimal merger policy which trades off possible market inefficiency, misallocation of capital, and financial instability that might arise as a result of merger activity. Since mergers are endogenous outcomes of the economic environment, it avoids the Lucas critique associated with reduced form empirical approaches.

References

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2015. “Systemic Risk and Stability in Financial Networks.” *American Economic Review* 105 (2):564–608.

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

———. 2025. “A quantitative model of banking industry dynamics.” *forthcoming Journal of Political Economy Macroeconomics* .

Gabaix, Xavier. 2011. “The granular origins of aggregate fluctuations.” *Econometrica* 79 (3):733–772.

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.

Nocke, Volker and Michael D. Whinston. 2010. “Dynamic Merger Review.” *Journal of Political Economy* 118 (6):1201–1251.

———. 2022. “Concentration Thresholds for Horizontal Mergers.” *American Economic Review* 112 (6):1915–48.

Olley, G. Steven and Ariel Pakes. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica* 64 (6):1263–1297.

Pakes, Ariel, Michael Whinston, and Fanyin Zheng. 2024. “The Consumer Welfare Effects of Bank Mergers.” *mimeo* .

A-1 Supplementary Appendix

A-1.1 Pricing of Loans and Deposits by Bank Size

Table A.1 describes the structure of asset returns and funding costs across the bank size distribution.[16]

Table A.1: Loan Interest Rates, Deposit Costs, and Markup by bank size (1984-2019)

	All Banks		Top 10		Fringe (no Top 10)	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Charge-off Rate %	0.85	0.55	1.08 †	0.71	0.61	0.38
Default freq. %	2.14	1.27	2.80 †	1.82	1.66	0.92
Cost Deposits %	0.03	1.49	0.08	1.62	0.05	1.45
Loan Interest Rate %	4.85	1.80	4.93	1.86	4.80	1.80
Markup	1.08	0.54	1.27 †	0.71	0.96	0.43

Note: Loan interest income as well interest expenses on deposits are deflated using the CPI. All estimates correspond to annualized values (computed using quarterly data). Top 10 Banks refers to the Top 10 banks when sorted by assets. Fringe Banks refers to all banks outside the top 10. † denotes statistically significant difference (at 1% level) between the average for Top 10 banks and the average of fringe banks. To test for statistical significance we regress the corresponding variable of a dummy that takes value equal to 1 if the value corresponds to a bank in the Top 10 and 0 otherwise. Data correspond to commercial banks in the U.S. between 1984 and 2019. Source: Consolidated Reports of Condition and Income

[16] Following the notation of the model we present, the charge off rate corresponds to $(1-\theta')\lambda'$ where $(1-\theta')$ 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 $(1-\theta)$ 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 interest rate is computed by taking the ratio of interest income from loans over loans and dividing by one minus the default frequency. The loan markup is computed as in Corbae and D'Erasmo (2021).

A-1.2 Other Moments

Table A.2: Moments Data

	Pre-reform 1984 - 1993			Post-reform 2010 - 2019		
	All Banks	Top 10	Fringe	All Banks	Top 10	Fringe
Cost Deposits % r_D	1.57	1.94	1.46	-1.31	-1.37	-1.22
Return on Securities % r_A	4.17	4.48	4.03	2.96	2.26	4.06
Default freq. % θ	3.10	4.64	2.49	2.48	3.03	1.65
Default freq. Autocorrel.	0.700	0.711	0.700	0.797	0.974	0.797
Std Dev Error Default freq. (σ_{u^θ})	0.008	0.006	0.008	0.004	0.002	0.004
Charge-off Rate %	0.96	1.22	0.86	0.94	1.18	0.55
Loan Loss Rate %	31.42	28.66	32.63	31.04	34.36	25.75
Std Dev Loan Loss Rate %	15.41	14.12	15.94	12.54	11.47	14.09
Net Mg Expenses %	1.78	1.15	2.00	1.48	1.35	1.69
Elasticity Net Mg Expenses	0.83	0.95	0.78	0.95	1.03	0.84
Fixed Cost / Loans %	0.96	0.89	0.99	2.81	0.78	5.83
Avg. Equity Issuance / Assets %	0.11	0.07	0.13	0.03	0.01	0.04
Avg. Dividends / Assets %	0.38	0.36	0.39	0.71	0.74	0.66
Interest Rate % r_B^L	6.69	6.98	6.58	3.18	3.27	3.03
Interest spread % $r_B^L - r_D$	5.25	5.24	5.25	4.50	4.67	4.22
Interest margin % $\theta r_B^L - r_D$	4.94	4.77	5.00	4.35	4.50	4.11
Loan markups % $\frac{\theta r_B^L}{r_D + c'(L^*)}$	49.08	56.29	46.64	175.58	205.37	149.73

Note: Data Pre corresponds to the period 1984 - 1993 and Data Post to the period 2010 to 2019. All moments reported correspond to asset weighted averages. Source: Call Reports

Table A.3: Moments Data (cont.)

	Pre-reform 1984 - 1993	Post-reform 2010 - 2019
Number of Fringe Banks (National)	10,392	5,183
Deposit HHI (national)	93.90	541.99
Number Top 10 Banks (state)	1.471	4.109
Number Fringe Banks (state)	167.667	116.876
Deposit HHI (state)	1019.39	1555.86
Number Top 10 Banks (county)	0.444	1.069
Number Fringe Banks (county)	5.428	6.495
Deposit HHI (county)	2191.70	2575.50
Deposit Market Share Top 10	24.77	57.79
Loan Market Share Top 10	28.55	52.86
Relative Size Top 10 / Fringe (Deposits)	324.79	688.47
Relative Size Top 10 / Fringe (Loans)	394.64	556.01
Bank Entry (denovo) rate	5.00	1.25
Bank Failure Rate	0.76	0.44
Bank Failure Rate Top 10	0.00	0.00
Bank Failure Rate Fringe	0.76	0.44
Bank Merger Rate	1.27	2.69
Deposit to Output Ratio	39.01	57.19
Bank Loans to Output Ratio[#]	32.46	43.72
Bank Loans to Output Ratio[†]	43.15	60.22
Bank Loans to Output Ratio[‡]	51.77	50.53
Bank Loans to Total Loans Ratio[*]	44.54	33.28

Note: Data Pre corresponds to the period 1984 - 1993 and Data Post to the period 2010 to 2019. [#] Loans and Leases in Bank Credit (All Commercial Banks) [†] Bank Credit, All Commercial Banks [‡] Credit to Private Non Financial Sector by Banks [*] Credit to Private Non-Financial Sector by Banks / Total Credit to Private Non-Financial Sector

Table A.4 presents the results when x_{it} corresponds to real deposits and aggregate fluctuations are measured using the growth of detrended real gdp. We find that R^2 are smaller but still significant with granular bank idiosyncratic shocks explaining upt to 4% of cyclical fluctuations in real gdp.

Table A.4: Explanatory power of granular residuals of dominant banks on aggregate output (R^2)

	Dep. Var. $\Delta\log(gdp_t)$					
	Top 10 banks		Top 35 banks		Top 100 banks	
(intercept)	-0.000133	0.000207	-0.000169	0.000127	-0.000109	0.000136
	(0.866)	(0.795)	(0.829)	(0.872)	(0.888)	(0.861)
Γ_t^x	-0.0182	-0.0204	-0.0280	-0.0324	-0.0395	-0.0468
	(0.550)	(0.497)	(0.449)	(0.377)	(0.312)	(0.226)
Γ_{t-1}^x	-0.0282	-0.0276	-0.0414	-0.0405	-0.0323	-0.0270
	(0.354)	(0.359)	(0.263)	(0.270)	(0.407)	(0.484)
Γ_{t-2}^x		0.0754**		0.0869**		0.0966**
		(0.013)		(0.019)		(0.013)
N	191	190	191	190	191	190
R^2	0.006	0.039	0.010	0.039	0.009	0.041
Adjusted R^2	0.000	0.024	0.000	0.023	0.000	0.025

Note: Table presents the R^2 from a regression of the corresponding dependent variable on Γ_t^x, Γ_{t-1}^x, and Γ_{t-2}^x when x is real deposits. "$\Delta\log(gdp_t)$" refers to growth rate of detrended real gdp. Top 10 Banks, Top 35 Banks, and Top 100 Banks correspond to the cases when $K = 10$, $K = 35$, and $K = 100$, respectively. We let $X_{it} = \overline{g}_t = N^{-1}\sum_{i=1}^N g_{it}^x$, so $\hat{\Gamma}_t^x = \sum_{i=1}^K \frac{\ell_{it-1}}{L_{t-1}}(g_{it}^x - \overline{g}_t)$. Source: Call Reports