Adverse Selection in Corporate Loan Markets

Journal of Finance (forthcoming)

Mehdi Beyhaghi, Cesare Fracassi, and Gregory Weitzner*

*Mehdi Beyhaghi (corresponding author, mehdi.beyhaghi@frb.gov) is at the Board of Governors of the Federal Reserve System. Cesare Fracassi (cesare.fracassi@mccombs.utexas.edu) is at the University of Texas at Austin. Gregory Weitzner (gregory.weitzner@mcgill.ca. is at McGill University. We thank Andres Almazan, Olivier Darmouni, Adolfo De Motta, Mark Flannery, Janet Garufis, Jeff Gerlach, Christoph Herpfer, Yunzhi Hu, Jiro Kondo, Brittany Lewis, Elena Loutskina, Robert Marquez, David Martinez-Miera, Atanas Mihov, Carola Müller, Ettore Panetti, Raghu Rajan, Uday Rajan, Farzad Saidi, Philipp Schnabl, Antoinette Schoar (the editor), Denis Sosyura, Phil Strahan, Sheridan Titman, Yufeng Wu, Zhanbing Xiao, Anthony Lee Zhang, two anonymous referees and associate editor, and seminar participants at Arizona State, the Federal Reserve Board, McGill, Richmond Fed, and UT Austin, as well as the participants at the AFA, AFBC, CICF, Community Bank Research Conference, FDIC Bank Research Conference, FTG Bridging Theoretical and Empirical Research Conference, IBEFA Summer Meeting, MFA, SFS Cavalcade, Utah Winter Finance Conference and the Corporate Finance Conference at Washington University in St Louis for the helpful comments and discussions. A previous version of this paper circulated under the title "Bank Loan Markups and Adverse Selection". The views expressed in this article are solely those of the authors. They do not necessarily reflect the views of the Federal Reserve Board or the Federal Reserve System. Beyhaghi's employer, the Federal Reserve, had the right to review the manuscript prior to submission. Fracassi and Weitzner have nothing to disclose.

ABSTRACT

Theories of competition typically predict a positive relationship between market concentration and prices. However, in loan markets, adverse selection can reverse this relationship as riskier borrowers become more likely to receive funding. Using supervisory data, we show that interest rates, borrower risk, and lending volume are higher in markets with more banks. We also create a novel measure of markup that is orthogonal to borrower risk, and, consistent with adverse selection, find that markups are higher after repeated borrowing relationships. Finally, we use a shock to large banks' lending costs to provide further support for the adverse selection channel.

Traditional models of product market competition predict that higher levels of market concentration lead to higher prices and reduced supply. For this reason, the antitrust division of the US Department of Justice and other regulatory agencies commonly use various measures of market concentration as important criteria to approve or block mergers. Indeed, in the banking industry, many studies find a positive relationship between prices and market concentration in the deposit market.[1]

However, unlike deposit markets, credit markets are plagued by two levels of asymmetric information: 1) borrowers are often better informed than lenders about their own creditworthiness, and 2) some lenders know more about certain borrowers' quality than other lenders. This latter form of asymmetric information can create a positive relationship between the num-

[1]E.g., Hannan (1991), Neumark and Sharpe (1992) and Drechsler, Savov, and Schnabl (2017).

ber of banks and interest rates in local banking markets.[2] Intuitively, when there are more banks in a market, individual banks become more concerned that borrowers approaching them for a loan have been previously rejected, creating an adverse selection problem similar to the "winner's curse" (e.g., Broecker, 1990). This mechanism can lead to higher average interest rates and lower quality borrowers receiving financing *conditional on observables.* Anecdotally, adverse selection appears to be an important issue for banks in practice. For instance, Shaffer (1998) states:

> "The chief financial officer of a new bank once told the author that 'as soon as you open your doors, every deadbeat in town lines up to try to borrow from you'... Bankers and bank examiners alike are very familiar with this phenomenon."

Despite the importance of adverse selection in credit markets, identifying it is challenging because it is driven by banks' private information, which is typically unobservable. For instance, if we were to compare two loans with similar observable characteristics but different interest rates, we cannot determine whether the loans' interest rates are different due to differences in the bank's risk assessment or due to differences in bank market power. More generally, without access to agents' private information, it is extremely difficult to test for adverse selection in financial markets.

In this paper, we address this challenge by using a confidential supervisory dataset that includes banks' private assessments regarding the riskiness of

[2] E.g., Broecker (1990), Riordan (1995), Shaffer (1998), Dell'Ariccia (2001), Marquez (2002), Dell'Ariccia and Marquez (2006) and Axelson and Makarov (2016).

their loans. Consistent with the adverse selection channel, we find a positive relationship between the number of banks in local markets and interest rates, borrower risk and loan volume. We also analyze market power by creating a novel measure of loan markup that incorporates banks' private information regarding the riskiness of their loans and find that markups are higher in markets with more banks and that banks with an informational advantage due to prior relationships with borrowers charge higher markups. Finally, we address endogeneity concerns regarding market structure by using a shock that reduced lending by the largest banks in the US. Taken together, our results provide support for adverse selection driving lending outcomes and market power in local loan markets.

Our analysis uses the Federal Reserve's Y-14Q Schedule H.1 data, which includes all corporate loans over one million dollars extended by large bank holding companies (BHC) in the United States. We restrict the sample to private, non-syndicated borrowers, resulting in a median asset size and revenue of \$23.6mm and \$46mm. Because these firms are private and on the smaller size, they are likely more subject to issues of local asymmetric information.

In addition to detailed firm and loan characteristics, a key advantage of the data is that banks are required to report their internal measures of probability of default (PD) and loss given default (LGD) for each eligible loan on their balance sheets. First, we show that these risk assessments are strong predictors of both interest rates and future loan non-performance and default, even after controlling for observable loan characteristics. Second, once we control for banks' internal risk assessments, interest rates do not predict future loan performance at all. This result implies that banks' internal risk assessments provide a more accurate measure of the lender's risk than the interest rate, giving further validity to the use of these assessments as measures of banks' private information.

In Broecker (1990) and other models of adverse selection, as the number of banks in a market increases, firms are able to approach more banks for loans after they have been previously rejected, increasing the likelihood that lower-quality firms ultimately receive funding. Hence, the adverse selection channel predicts a positive relationship between the number of banks in a market and interest rates, borrower risk, and lending volume. Consistent with this channel, we find that interest rates are higher in counties with more banks. A one standard deviation increase in the number of banks operating in a county (about 6 banks) is associated with a 7bp increase in interest rates which compares to an average credit spread of about 150bps over the sample period. Moreover, the higher interest rates are at least partially due to higher borrower risk: we find that banks' estimated PDs are higher in counties with more banks. Importantly, we estimate the difference in risk across loans that is not already being reflected in observable characteristics.[3] Finally, we aggregate the amount of lending volume at the county-quarter level and find that lending volume is higher in counties with more banks. Taken together, these results are consistent with more low-quality borrowers receiving financing in markets with more banks, a central prediction of models of adverse selection in credit markets. Also consistent with adverse selection, we show that 1) the effects are stronger for loans issued to firms with fewer tangible

[3]If the difference in risk was already reflected in observables, then by definition, it could not be driven by adverse selection.

assets, 2) loans are more heavily collateralized in markets with more banks, 3) firms rarely switch banks and do so at the same frequency across market structures, and 4) banks' PDs after granting the loan are more predictive of default in markets with more banks as successfully extending the loan conveys more information about the borrower's creditworthiness (e.g., Milgrom (1979)).

One concern with our analysis is that our measure of the number of banks only includes large banks, not small banks or online lenders. We thus define an alternative measure of the number of banks in the county based on the number of FDIC-insured bank branches. We show that this measure, which is not biased towards large banks, is highly correlated with our main measure of the number of banks (0.84) and that our main results hold if we use this alternative measure in our tests. Second, we also use survey data from Wiersch et al. (2022) to show that firms of similar size rarely apply for loans to online lenders, credit unions, and finance companies.

Interest rates and borrower risk are higher in markets with more banks; however, differences in loan processing costs, funding costs could also partially explain the difference in interest rates across these market structures. First, we control for bank by time fixed effects, which absorb any differences in costs of capital across banks or over time.[4] Second, we also control for county-level characteristics such as population density and local wages in the financial industry that may affect the costs of processing loans. Another

[4]Bank by time fixed effects also alleviate concerns that strategic underreporting of risk assessments could be driving our results (e.g., Behn, Haselmann, and Vig (2016) and Plosser and Santos (2018)).

possibility that we explore next is that differences in market power may also affect interest rates across markets.

While standard theories of competition predict reduced market power in markets with more banks, theories of adverse selection predict that market power can be higher if banks are better informed about certain borrowers.[5] In order to disentangle these forces, we estimate predictive regressions with interest rate as the dependent variable, controlling for the bank's assessed risk of the loan. By stripping out the variation in interest rates due to borrower risk, which in all studies up to this point has been unobservable, we are better able to isolate the markup on the loan due to market power. We find empirical support for this argument because, as discussed earlier, once we control for banks' internal risk assessments, interest rates do not predict future loan performance at all.

Consistent with the adverse selection channel, we find that our measure of markup increases with the number of banks in a county. Hence, the higher interest rates exhibited in markets with more banks can be explained not only by higher average borrower risk but also by higher market power induced by adverse selection.

That markups increase in the number of banks suggests that adverse selection may be a source of market power for banks. To further explore this idea, we develop additional tests based on theories of adverse selection and market power. A common theme in these models is that a bank's superior information about certain borrowers increases its market power. In practice,

[5] E.g., Dell'Ariccia and Marquez (2006), Fishman and Parker (2015) and Bolton, Santos, and Scheinkman (2016).

banks that have existing relationships with firms are likely to have better information than other banks (e.g., Sharpe, 1990 and Rajan, 1992). Hence, we test whether banks charge their existing clients higher markups than firms that switch banks. Consistent with this hypothesis, we find that firms that stay with their existing banks face 9bps higher markups on their loans.

While our evidence supports the adverse selection channel, the number of banks is not randomly allocated across counties. Although we control for a rich set of firm, loan, industry, bank, and county characteristics, our results could still be explained by unobserved drivers of lending behavior that are correlated with the number of banks in the county. To address these concerns, we use capital surcharges that were imposed on global systemically important banks (GSIBs) in 2016 as an exogenous shock to the lending costs of those banks. Favara, Ivanov, and Rezende (2021) show that following the imposition of these capital surcharges, GSIBs reduced their lending relative to other banks. We first show that a greater presence of GSIBs in a county prior to the imposition of capital surcharges leads to a reduction in both aggregate lending and the total number of banks servicing the county after the imposition of the surcharges. While standard models of competition predict that a reduction in the supply of capital should lead to higher prices, this forced reduction in lending could also reduce the adverse selection problem in local markets. Intuitively, it becomes less of a bad signal if a particular firm does not receive a loan from a GSIB because the bank may have denied the firm credit (or not even considered the firm at all) because it was forced to cut back its lending, not because it deemed them a lower quality borrower. Hence, the adverse selection channel predicts that interest rates, PDs, and markups would drop in regions with more GSIBs present prior to the surcharges. We find evidence consistent with all three of these predictions; however, while interest rates and PDs are statistically significant, markup is not.

This paper contributes to the literature that analyzes the relationship between market structure and lending outcomes in banking markets. To our knowledge, this is the first paper documenting a positive relationship between the number of banks and interest rates in local corporate loan markets. The existing literature finds either a positive or no relationship between loan interest rates and market concentration (e.g., Hannan, 1991, Petersen and Rajan, 1995, Cyrnak and Hannan, 1999, Sapienza, 2002, Cavalluzzo, Cavalluzzo, and Wolken, 2002, Berger, Rosen, and Udell, 2007 and Rice and Strahan, 2010).[6] Petersen and Rajan (1995) find that in concentrated markets firms borrow at lower rates earlier in their life but then borrow at higher rates later in their life, but do not find a difference in unconditional borrowing costs across market structures.[7] Finally, to our knowledge, this is the

[6]Berger et al. (2004) and Degryse and Ongena (2008) survey the literature.

[7]Other papers using different arguments to cast doubt on concentration as a proxy for competitiveness in banking markets are Berger (1995), Rhoades (1995), Hannan (1997), Claessens and Laeven (2005), and Carbo-Valverde, Rodriguez-Fernandez, and Udell (2009). Although we focus on the effect of market power on interest rates, other papers highlight benefits of increased competition unrelated to the interest rates of loans (e.g., Jayaratne and Strahan, 1996, Cetorelli, 2002, Bertrand, Schoar, and Thesmar, 2007, Liebersohn, 2017 and Saidi and Streitz, 2018). Increased competition can also reduce efficiency if it causes banks' charter values to decrease, thereby inducing an increase in risk taking (Keeley, 1990). See Granja, Leuz, and Rajan (2022) and Carlson, Correia, and Luck (2022) for empirical evidence supporting this channel. Because we use bank by

first paper to analyze how underlying borrower quality systematically varies across market structures as measured by banks' private risk assessments and use these risk assessments to measure loan markups.

Why do our results on the relationship between market concentration and interest rates differ from those in the existing literature? Much of the existing literature uses data containing many sole proprietorships and partnerships, to which banks typically lend based on the commercially available credit score of the owner (e.g., FICO score), which is available to all lenders.[8] In contrast, our sample contains only corporate loans for which banks assess the credit quality entirely based on their own internal risk ratings. For example, the Fed distinguishes between corporate loans (which are in our sample) versus sole-proprietorships and partnerships, which the Fed regards as "small businesses" (which are outside our sample).[9]

Asymmetric information across banks is likely severe among the corporate loans in our sample for two reasons. First, while credit bureaus maintain standardized credit histories for individuals, no equivalent system exists for corporations. The lack of publicly available information on past interactions is critical to the type of asymmetric information problem in the theories

time fixed effects throughout our analysis, we absorb any aggregate bank-level risk-taking effects.

[8]For example, in the most recent Survey of Small Business Finance, about two-thirds of the firms are sole-proprietorships or partnerships (see The 2003 Survey of Small Business Finances Methodology Report).

[9]See Online Appendix Section 2 for the specific Fed instructions. This distinction may not be as stark for smaller banks outside our sample, which are not subject to the same reporting requirements.

that motivate our paper. Second, without a common credit score, banks must rely on their own internal evaluations of borrowers' creditworthiness, increasing the potential for information asymmetries between lenders. Increased asymmetric information across banks offers a potential explanation for why we find different results on the relationship between market structure and interest rates than the existing literature.[10]

Our paper also contributes to the empirical literature testing the relationship between bank market structure and asymmetric information (e.g., Cetorelli and Gambera (2001), di Patti and Dell'Ariccia (2004), Crawford, Pavanini, and Schivardi (2018) and Yannelis and Zhang (2022)).[11] Compared to these papers, ours is the only paper to directly analyze banks' private risk assessments and how they vary across market structures, conditional on observables.

Our paper also relates to the broader literature testing asymmetric information in credit markets (e.g., Stroebel, 2016, Botsch and Vanasco, 2019,

[10]It is also worth mentioning that the vast majority of papers use measures of deposit market concentration rather than measures of loan market concentration due to a lack of data. A problem with this approach is that deposit concentration may not line up with loan market concentration. Indeed, in our sample, the correlation between the number of banks and deposit HHI is -0.27 (see Table 2). Moreover, in Online Appendix Table 14, we do not find a statistically or economically significant relationship between deposit HHIs and loan rates.

[11]Yannelis and Zhang (2022) find a negative relationship between market concentration and interest rates in the auto loan industry among high-risk borrowers, but do not analyze bank risk assessments or markups. Auto loans are also quite different from corporate loans, as banks typically assess the credit risk of auto loans based on commercially available credit scores.

Darmouni, 2020, Weitzner and Howes, 2021, DeFusco, Tang, and Yannelis, 2022, and Beyhaghi, Howes, and Weitzner (2023)). Because agents' information is not observable, the most common approach in this literature is to rely on proxies of asymmetric information or assume agents' decisions imply certain distributions of outcomes and test whether these outcomes bear out in the data. In contrast, our data allows us to directly analyze banks' private information to see how borrowers' risk varies across market structures *conditional on observables*, thereby allowing us to more directly test for adverse selection.

Finally, this is the first paper to estimate loan markups by orthogonalizing the interest rate to the bank's perceived risk of the loan. Using banks' private risk assessments to estimate markup is critical because, as we show, observable characteristics do not fully account for the underlying risk of loans. Furthermore, we show that loan riskiness varies across regions, even after controlling for observable characteristics. This makes it extremely challenging to estimate the relationship between market structure and market power without access to banks' private information. Hence, our methodology could be useful to both regulators and researchers who have access to the Y-14Q data to estimate markups and thereby better understand market power in corporate loan markets.

I. Theoretical Background

In homogeneous product markets, standard theories of competition predict that fewer competitors (or higher concentration) lead to higher prices.

For example, in static Cournot models, in which firms compete via quantities, firms better internalize the impact that their production has on prices when there are fewer firms, leading to higher markups and firm profits. If firms compete through prices, à la Bertrand, prices are competitive as soon as there are two firms. In dynamic settings, higher concentration facilitates collusion, which can also lead to higher markups (e.g., Abreu, 1986).

However, in credit markets plagued by asymmetric information, the relationship between market concentration and prices can reverse due to adverse selection. Intuitively, if banks cannot see whether firms have been previously rejected by other banks, they face a winner's curse problem similar to a common value auction (e.g., Broecker, 1990, Riordan, 1995, Flannery, 1996, Shaffer, 1998, Axelson and Makarov, 2016, He, Huang, and Zhou, 2020 and Goldstein, Huang, and Yang, 2022). The greater the number of banks, the more concerned banks are that firms have been previously rejected by other banks, forcing them to charge higher rates to borrowers. In a related, but slightly different, mechanism based on adverse selection, as the number of banks increases, information becomes more dispersed (e.g., Marquez, 2002 and Dell'Ariccia and Marquez, 2006), as banks have a harder time determining whether borrowers have been rejected by other banks. These two mechanisms both predict that more banks lead to i) higher interest rates, ii) higher borrower risk, and iii) higher aggregate loan volume. Importantly, these results hold after conditioning on observable characteristics, a requirement of adverse selection.

Adverse selection can also affect the relationship between market structure and market power. If some banks know more about certain borrowers than other banks, either because they are better at screening (e.g., He, Huang, and Zhou, 2020), or because they have access to private information through ongoing relationships with those borrowers (e.g., Sharpe, 1990 and Rajan, 1992), this can cause banks' market power to increase with the number of banks in the market. Intuitively, informed banks can charge their borrowers interest rates higher than marginal costs because those borrowers would be pooled with lower quality borrowers if they tried to borrow from another bank. Because having more banks in the market worsens the adverse selection problem and, in turn, borrowers' outside options, informed banks are able to extract more information rents from their borrowers. This mechanism, which appears in Dell'Ariccia and Marquez (2006), Fishman and Parker (2015), and Bolton, Santos, and Scheinkman (2016), causes a positive relationship between the number of banks and markups.

While in the aforementioned models the relationship between the number of banks in a market and interest rates and markups is unambiguous, in reality, both of these effects might be present. Hence, when analyzing interest rates and markups, our estimates inevitably reflect the net effect of these two channels. For instance, it is possible that the standard forces of competition first cause interest rates and markups to decrease as the marginal impact of an additional bank on the degree of competition is high. However, after a certain point, the marginal impact of the standard competition channel could diminish, and adverse selection effects begin to dominate. Because standard models of competition, such as Cournot and Bertrand, do not speak to the relationship between the number of banks and borrower risk, we expect a monotonically increasing relationship between the number of banks and borrower risk under the adverse selection channel.

Finally, while we cannot directly observe which banks are better informed about other borrowers, past lending relationships can create information advantages.[12] Through the information advantages that develop over time, we expect that borrowers that stay with their existing banks face higher markups than those that switch to different banks (e.g., Sharpe (1990) and Rajan (1992)).

II. Data

Our main source of data is Schedule H.1 of the Federal Reserve's Y-14Q data. The Federal Reserve began collecting this data in 2011 to support the Dodd-Frank mandated stress tests and the Comprehensive Capital Analysis and Review (CCAR). Qualified BHCs are required to report detailed quarterly loan-level data on all corporate loans that exceed one million dollars in size. These loans represent 70% of all commercial and industrial loan volume from US BHCs (Bidder, Krainer, and Shapiro, 2020).

The data include detailed loan characteristics (such as interest rate, maturity, and amount), quarterly loan performance, the ZIP code of the borrowers' headquarters, and firm financials. Importantly for our analysis, banks are also required to report their internal estimates of probability of default (PD)

[12]See Berger and Udell (1998) for a detailed discussion of how this information can be obtained. Empirically, Beyhaghi, Howes, and Weitzner (2023) show that changes in banks' risk assessments, i.e., PD and LGD, predict stock returns, bond returns, and earnings announcement surprises, suggesting banks can obtain an information advantage through lending relationships.

and loss given default (LGD) for each loan to the Federal Reserve on their Y-14Q filings. According to the Basel Committee on Banking Supervision, internal estimates of PD and LGD "must incorporate all relevant, material and available data, information and methods. A bank may utilize internal data and data from external sources (including pooled data)."[13]

Following Brown, Gustafson, and Ivanov (2017), we restrict the sample to domestic borrowers and remove financial firms, government entities, individual borrowers, foreign entities, and nonprofit organizations. In addition, we drop loans to special purpose entities, loans with government guarantees, demandable loans, loans with prepayment penalty clauses, loans that are tax-exempt, and loans that are contractually subordinated. We include these additional screens to make the loans in our sample as comparable as possible, thereby allowing us to accurately compare interest rates. To keep focus on issues of local information asymmetry, we also drop publicly traded firms (firms with a valid ticker information) and syndicated loans because they are usually sourced nationally rather than locally. To ensure that our results are not affected by the sample of public firms with unreported ticker information, we trim the sample on borrower size at the 99th percentile.

To correct for reporting inconsistencies, we exclude loans that are less than \$1 million because banks are only required to report loans in which the total commitment amount is \$1 million or more. Additionally, we drop borrowers with reported total assets less than the loan amount, loans with interest rates equal to or below 0% or above 100%,[14] loans with missing

[13] The most recent instructions are available at Calculation of RWA for credit risk.

[14] In cases where an interest rate floor or an interest rate ceiling is specified for the

maturities or with maturities more than 30 years. Finally, we drop loans with PDs that are missing, equal zero, or are greater than the 99th percentile. Interest rates are reported only in the quarter in which the borrower makes a payment on a loan, otherwise the loan's interest rate is reported as zero. For credit lines, this has a material impact because firms may not draw them immediately. Hence, when the interest rate field is zero, we take the interest rate from the next quarter that is populated. Credit lines that are not utilized within two quarters after initiation are dropped from the sample. As a loan might remain on the bank's balance sheet for multiple quarters, we only keep the first appearance of a loan in the data (i.e., new loans). Finally, we drop loans from borrowers that have more than the 99th percentile in the total number of new loans over the sample period, as these tend to be financial subsidiaries of industrial firms. After these filters, we are left with 21,924 new loans originated from 2014Q4 to 2019Q4 by 23 BHCs.[15]

We define the following firm-level financial variables: profitability (EBITDA/assets), firm size (log assets), tangibility (tangible assets/assets), and leverage (debt/assets), all of which, except for firm size, are winsorized at the 1% and 99% level. Furthermore, we use two measures of loan performance: 1) non-performance, which is a dummy variable equal to one if the bank reports the loan as 90 days past due or non-accrual, or reports a positive net cumulative charge-off amount, or reports specific reserve for an

loan, we keep only loans with reported interest rates between the reported floor and the reported ceiling. If an interest rate spread is reported for variable-rate loans, we ensure that the total interest rate is at least as much as the interest rate spread.

[15]We provide a more detailed explanation of the data filters in Online Appendix Section 1.

impaired loan for the loan within the 12 months following the origination of the loan, or if the bank considers the borrower as defaulted as defined below; and 2) realized default, which is a dummy variable that equals 1 if the borrower defaults within one year since origination. We use a window of one year because banks' PD estimates are required to reflect long-run annual default rates. For regressions predicting non-performance or default, our sample ends one year earlier to ensure all loans have the same one-year window to potentially default.

Finally, the data includes ZIP codes corresponding to each borrower's headquarters. To control for county characteristics, we collect annual population estimates from the Census and quarterly wage data from BLS. As is typical in the banking literature, we use counties as our measure of local markets (e.g., Drechsler, Savov, and Schnabl (2017). In order to create market concentration measures at the county level, we obtain the ZIP code to county crosswalks from the Housing and Urban Development (HUD). After merging the county data into the Y-14Q dataset, we construct a measure of market concentration based on the number of banks that operate in that county in the sample. Specifically, we consider a bank to operate in a county if it gives a loan at any point in the sample.[16] We use the number of banks as our measure of concentration because what matters in models of adverse selection is the number of banks that firms can approach for a loan. However, in Table II we show that the loan Herfindahl-Hirschman Index (HHI) is highly correlated with the number of banks (-0.90) and in Online Appendix

[16] In Section VII, we also measure the number of banks at the annual level to explore variation in the number of banks following a shock to banks' cost of lending.

Table 14, we show our main result regarding interest rates holds if we use loan HHI instead. The average loan HHI is also much larger than deposit HHI (0.50 versus 0.20). This could be partially due to the fact that we do not have all US banks in our sample, but also banks may need more expertise to make corporate loans compared to taking deposits. The appendix contains detailed definitions of all of our variables.

Table I includes summary statistics. The average and median loan size are approximately \$7.2mm and \$2.7mm, respectively and over 90% of loans are less than \$16.2mm. The loan sample is approximately evenly split among credit lines and term loans, and the median interest rate is 3.66%, which corresponds to about a 150bp credit spread over the average 5-year swap rate. The median firm has \$23.6mm in assets, 7% profitability, and 31% book leverage. The fact that the majority of the loans and firms are relatively small is important for testing the local effects of asymmetric information, as larger firms are usually able to source their loans nationally.

Over the sample period, 0.82% of firms default within the first year after loan origination. This compares to an average ex-ante PD of 1.34%. This discrepancy is likely due to realized aggregate economic conditions in the US over the sample period being positive relative to banks' ex-ante expectations.

There is quite a bit of cross-sectional variation in market concentration, with the bottom ten percentile of loans being in counties with 4 or fewer banks (loan HHI of 0.91) and the top ten percentile of loans being in counties with 20 or more banks (loan HHI of 0.22). Online Appendix Figure 1 plots the full distribution of loans across the number of banks.

A. *Validity of Bank Risk Assessments*

For the risk measures to reflect banks' private information, they must reflect the actual risk of the loans. In this section, we verify that banks' risk assessments predict interest rates and ex-post loan performance.

First, we test the relationship between bank risk assessments and interest rates by estimating the following regression:

$$IR_l = \beta_0 PD_l + \beta_1 LGD_l + \beta_2 (PD_l \times LGD_l) + \Gamma X_l + \delta_{b,t} + \alpha_{i,t} + u_l, \quad (1)$$

where the unit of observation is each loan l in industry i originated by bank b in quarter t. The outcome variable IR_l is loan l's interest rate and PD, LGD and the Expected Loss ($PD \times LGD$) are our key explanatory variables. In addition, X_l is a vector of loan-level controls, which include the log of the maturity in months, the log of the loan amount, whether the loan has a guarantor or not, loan purpose fixed effects and loan type fixed effects (combination of credit facility type (term/line), interest rate variability (floating/fixed), $\delta_{b,t}$ is bank by quarter fixed effects, and $\alpha_{i,t}$ is industry by quarter fixed effects. Bank by quarter fixed effects allow us to control for any differences in internal risk models across banks or within bank over time. Importantly, this allows us to control for any potential aggregate underreporting of risk assessments by banks in order to reduce their capital requirements (e.g., Behn, Haselmann, and Vig (2016) and Plosser and Santos (2018)). Moreover, by evaluating loans given by the same bank in the same quarter, we also absorb any differences in banks' cost of capital that may affect interest rates. Throughout the paper, we cluster the standard errors by county.

The results are displayed in Table III. In Column (1), where we only include loan-level characteristics and fixed effects, the adjusted R-squared is 52%. In Column (2), we include PD, LGD, and their interaction term (Expected Loss) in the regression. Consistent with banks acting upon their risk assessments, risk assessments strongly predict interest rates even after controlling for loan characteristics and many fixed effects. The adjusted R-squared increases from 52% to 55%, confirming that these bank assessments can explain a large portion of the heterogeneity in interest rates.[17] The effect of risk assessments on interest rates is not only statistically significant, but also economically relevant. For example, a 1pp increase in PD (about two-thirds of a standard deviation) leads to a 7.7bp increase in interest rates.

Second, we evaluate the univariate correlation between realized default and interest rate, comparing it to the correlation between realized default and PD. In Figure 1A we place loans into five equal-sized bins sorted on interest rate and plot their average realized default rate. While the overall correlation is positive, there is a minimal increase in realized default rates as the interest rate increases across the first three bins. On the other hand, when we place loans into five equal-sized PD buckets, we see a much clearer positive and monotonic relationship between average realized defaults and PD than interest rates (Figure 1B). Bank risk assessments appear more strongly correlated with performance than interest rates, suggesting that interest rates may include substantial non-risk components (e.g., markups due to market

[17] The F-value of joint significance of the three risk measures is 178.28, allowing us to clearly reject the null hypothesis that risk assessments do not explain interest rates.

power).

Third, we test whether the degree to which bank risk assessments and interest rates separately explain loan performance in a multivariate setting. We estimate the following regression:

$$y_l = \beta_0 IR_l + \beta_1 PD_l + \beta_2 LGD_l + \beta_3 (PD_l \times LGD_l) + \Gamma X_l + \delta_{b,t} + \alpha_{i,t} + u_l, \quad (2)$$

where the outcome variable y_l is either Non-Performance or Realized Default and X_l contains the same loan-level controls as in (1). The results are displayed in Table IV.[18] In Columns (1) and (4), we only include the interest rate as the main independent variable and find that, consistent with Figure 1A, interest rates are strongly related to loan performance. In Columns (2) and (5), we only include the risk assessments as the main independent variables and find that PD strongly predicts future loan performance. In particular, a 1pp increase in PD increases the likelihood that the firm defaults by about 0.89pp. The adjusted R-squared increases from 0.09 to 0.10 and 0.07 to 0.08 when the risk assessments alone are included as compared to the interest rate alone. Finally, in Columns (3) and (6), we include interest rates and risk assessments together as independent variables. Interestingly, while the risk assessments still strongly predict future loan performance, the coefficient on interest rate drops substantially (from 0.527 to 0.101 on non-performance and 0.354 to 0.093 on realized default) and becomes statistically insignificant for both specifications. These results are consistent with bank

[18] We have slightly fewer observations than in Table III, because we do not include loans in 2019 because the default horizon is less than a year.

risk assessments fully capturing the underlying risk of the loans. Moreover, they suggest that interest rates may be affected by factors other than risk, such as market power, which we explore further in Section IV. We cannot formally test whether interest rates are truly orthogonal to loan performance after controlling for banks' risk assessments, as the inability to reject the null hypothesis does not mean that the null is true. However, as we just showed, the point estimate on interest rate is both economically small (less than one-fifth of its level when we exclude risk assessments in predicting non-performance) and statistically insignificant.

III. Loan Market Structure and Adverse Selection

We have established that bank risk assessments contain substantial information about the riskiness of borrowers beyond interest rates and observable characteristics. In this section, we test how lending terms and borrower risk vary across local market structures. First, we test the relationship between market concentration and interest rates by estimating the following loan-level regression:

$$IR_l = \beta NOB_c + \Gamma_0 X_l + \Gamma_1 Z_{f,t} + \delta_{b,t} + \alpha_{i,t} + u_l, \tag{3}$$

where NOB_c is the the number of banks in the county where the firm is headquartered, X_l is a vector of loan characteristics as in (1), and $Z_{f,t}$ is a vector of firm characteristics, which include profitability, firm size, tangibility, and leverage. The results are displayed in Table V. In Column (1), we estimate the regression without firm characteristics. The point estimate for the number of banks is 0.012 and statistically significant, i.e., if the county has an additional bank, this is associated with a 1.2bp higher interest rate. This result suggests that a one standard deviation increase in the number of banks (5.86) increases the credit spread by about 7bp compared to an average credit spread over 5-year swap rates of about 150bps over the sample period. In Column (2), we include firm characteristics, and the coefficient on the number of banks remains stable. Finally, in Column (3) we show that the coefficient remains positive and statistically significant after controlling for population density, average wages, and average wages in the financial industry (all in logs) in the county that may be related to the costs of administering the loan.

In terms of economic magnitudes, Degryse and Ongena (2008) survey the literature on the relationship between market structure and interest rates and find that a 10pp increase in HHI (measured by deposit HHI for studies in the US), which is about 1.25 SD in our data, leads to anywhere between not statistically different than 0 to 61bps increase in loan rates. However, interest rates were much higher during these samples, so our effect should be considered slightly larger on a relative scale. Moreover, as shown below, the effects are non-monotonic, which reduces the average effect.

We test for non-monotonicity by plotting the coefficients for various numbers of banks operating in the county in Figure 2. Specifically, we estimate differences in interest rates across dummies of number of banks, from 1 (reference group) to 2, 3-7, 8-12, 13-17, and 18-23. The coefficients exhibit a U-shape in which interest rates first drop, but then increase as the number of banks increases. This result is consistent with the standard competition force dominating at first, but the adverse selection channel dominating after a certain number of banks are in the market. This result seems natural, particularly if banks are competing on price à la Bertrand, as the marginal impact of an additional bank on the degree of competition begins to diminish.

Next, in Table VI we estimate the same regressions in Table V but use PD as the dependent variable rather than interest rate to test how borrower risk varies based on the number of banks in the county. In many models of adverse selection, the average risk of borrowers increases with the number of banks operating in a market, as lower quality firms have more opportunities to receive funding from banks (e.g., Broecker, 1990). Across all specifications, increasing the number of banks in a county is associated with higher PDs. In Column (2) with firm characteristics, PDs increase by 1.1bp with one additional bank in the county. This compares to an average PD of 134bps. This result not only lends support to the adverse selection channel but also highlights the potential risks of making inference purely based on observable firm and loan characteristics, as the risk of loans can systematically vary across different market structures *conditional on observables*.

We again plot the coefficients for various numbers of banks operating in the county in Figure 3. In contrast to interest rates, PDs are monotonically increasing in the number of banks, which is consistent with standard models of adverse selection (e.g., Broecker, 1990) where lower quality borrowers have a higher chance of receiving funding when there are more banks in the market.

Another key prediction of adverse selection models is that lending volume increases in the number of banks as more low-quality borrowers receive financing. To test this prediction, we aggregate the data at the county quarter level and estimate the following regression:

$$Volume_{c,t} = \beta_0 NOB_c + \Gamma_0 X_{c,t} + \delta_t + u_{c,t}, \tag{4}$$

where $Volume_{c,t}$ is the log of total dollar loan volume or the total number of loans granted in county c in quarter t, $X_{c,t}$ are county level controls, which include population density, average wages, finance industry wages, and the total population (all in logs) and δ_t are year-quarter fixed effects. The results are displayed with and without county-level controls in Table VII. Across all specifications, the coefficient is positive and statistically significant. For instance, in Column (2), the coefficient for lending volume is 0.14, implying that one additional bank is associated with about 14% higher lending volume.

Similar to the behavior of PDs, Figure 4 shows that loan volume monotonically increases with the number of banks in a market. Of course, higher lending volume in markets with more banks is not a unique prediction of adverse selection models. Many models of competition, such as Cournot, make the same prediction. However, these models do not predict higher interest rates and a riskier pool of borrowers as we see in the data.

Lastly, we perform three tests of ancillary predictions of models of adverse selection. First, if the winner's curse becomes more severe in markets with more banks, we would expect that banks' PDs would more accurately predict subsequent non-performance. Intuitively, when the winner's curse is more severe, a bank "winning" a borrower reveals more information about that borrower's quality.[19] To test this prediction, we use the area under the receiver operating characteristic curve (AUC) as a measure of the accuracy of banks' internal risk assessments.[20] Figure 5 shows that the AUC is higher in markets with more banks (above median) for realized default (0.769 versus 0.694) and also for non-performance (0.705 versus 0.630).[21] These results suggest that banks' PDs are more predictive of loan performance and provide further support for the adverse selection mechanism.

Second, Dell'Ariccia and Marquez (2006) show that in markets with more banks, as the asymmetric information problem becomes more severe, this can result in an increase in collateralization so that banks can screen borrowers.[22] We test this prediction by estimating the same specification as (4), but with the variable *Secured by Blanket Lien* as the dependent variable. This is a dummy variable that equals one if the bank has a claim on all unencumbered collateral of the borrower and proxies for the degree of collateralization of

[19] In fact, Milgrom (1979) shows that under certain conditions the price paid in a common-value auction converges to the true value of the good when the number of bidders gets large.

[20] The AUC is the most common approach to measure the discriminatory ability of binary classifier variables. Online Appendix Section 9 provides a detailed explanation of the AUC.

[21] These differences are statistically significant based on the standard Delong test (DeLong, DeLong, and Clarke-Pearson (1988)).

[22] Specialization could be another way to avoid adverse selection. In Online Appendix Table 3, we find no difference in specialization (as defined by Paravisini, Rappoport, and Schnabl (2023)) across markets with more or fewer banks. Thus, it seems that collateralization, rather than specialization, is the more prevalent way banks protect themselves from adverse selection.

the loan.[23] Table VIII shows that loans are more likely to have blanket liens in markets with more banks. Specifically, a one standard deviation increase in the number of banks (about 6) increases the likelihood of a blanket lien by 1.2pp (3.2%).

Finally, in a frictionless competitive market, we would expect a strong positive relationship between the frequency of firms changing banks and the number of banks in the market.[24] In contrast, as mentioned in Section I, if banks can gain an information advantage over other banks through their lending relationships, this can create adverse selection, making it very difficult to switch banks even if there are many banks in the market. To test this hypothesis, we create a variable *Stay Bank*, a dummy that equals one when a firm borrows from an existing bank on the current loan. For this analysis, we restrict the sample to firms with more than one loan over the sample period and analyze all loans that follow their first loan.[25] In Figure 6, we assign loans to categories based on the number of banks in the county and plot the average of *Stay Bank* in each category. Firms stay with their existing banks around 75% of the time, with minimal changes in that rate as the number of banks in the county increases. Hence, even though there are more possible banks to borrow from, firms rarely change banks. We also formally test

[23]The vast majority of loans in our sample have some form of collateral; hence, we instead rely on variation in the type/degree of collateralization.

[24]For example, suppose there are N banks in a market, each of which receives an i.i.d. signal about the borrower's creditworthiness. The probability that a firm switches to a new bank is $\frac{N}{N-1}$, which is increasing in N.

[25]For this filter, we use the data from 2011 up to the beginning of our sample to determine whether a loan follows the firm's first loan.

this in Figure 7 by estimating regressions with *Stay Bank* as the dependent variable and the same controls as in (4). There is a decrease in the likelihood of firms' staying with their existing banks; however, none of the point estimates are statistically different than zero and are economically small (e.g., in markets with 18-23 banks, firms change banks about 3pp more often than in markets with 2 banks). The fact that firms still barely switch banks in markets with many banks is consistent with adverse selection plaguing these markets. In Section VI, we explore how information obtained through these existing lending relationships can create market power for banks.[26]

One concern with our measure of the number of banks in a county is that we do not have loans originated by small banks and nonbank financial institutions such as online FinTech lenders, as our data only covers US banks with over $50bn in assets. To address the issue of a lack of small banks, we also create an alternative measure of the number of banks in each county based on branch-level data from the FDIC Summary of Deposits, which includes all FDIC-insured banks. The measure has a correlation of 0.84 with our main measure (Table II), and when we reproduce our main tables using this alternative measure of concentration, we find very similar qualitative results (See Online Appendix Tables 11 - 13). These results suggest that our measure of the number of banks is not simply picking up the degree of dominance of large banks in a county.

[26]In Online Appendix Table 4, we also show that new borrowers tend to be *riskier* and receive *better* terms on their loans relative to older borrower, which suggests banks anticipate the value of relationships and future market power from adverse selection in markets with more banks.

Second, Gopal and Schnabl (2022) show that after the financial crisis online lenders or "FinTechs" and finance companies, increased their market share in small business lending dramatically.[27] However, among our sample of firms, these types of lenders appear less important. According to a survey conducted by the Federal Reserve Bank of Cleveland, firms rarely apply for loans of over \$1mm from online lenders (Wiersch et al. (2022)). Specifically, they find that 3.7% of firms applying for loans over \$1mm apply for loans from online lenders at all.[28] Moreover, among the 3.7% that applied to online lenders, almost two-thirds also applied to banks.

There has also been an increase in nonbank lending at the middle market level (e.g., Chernenko, Erel, and Prilmeier (2022) and Davydiuk, Marchuk, and Rosen (2024)).[29] Middle market firms are typically defined as having revenue of \$10mm to \$1bn. Moreover, the sample of loan-level data in Chernenko, Erel, and Prilmeier (2022) includes much larger, publicly traded firms.[30]

Finally, even to the extent that nonbank lenders are prevalent in these markets, Begley, Purnanandam, and Traweek (2023) find a positive relationship between the presence of nonbank lenders and bank branches, suggesting that nonbanks are not simply a pure substitute for banks.

[27] Begley and Srinivasan (2022) show that nonbanks have also taken market share from banks in the mortgage market.

[28] See Online Appendix Table 37 for more details.

[29] See also Chen, Hanson, and Stein (2017) for evidence that nonbank lenders partially filled in for the reduction in small business lending by large banks after the GFC.

[30] The median firm size in our sample is less than one-fifth the size of the firms in Chernenko, Erel, and Prilmeier (2022) (23.6mm versus 126mm for firms that receive nonbank loans).

IV. Loan Market Structure and Market Power

In Section III, we find that markets with more banks issue loans to riskier borrowers at a higher interest rate. However, it is not clear whether the higher interest rate is purely due to the borrowers' higher risk or if it also includes higher profits, or markups, that banks can extract from borrowers. As discussed in Section I, standard theories of competition, e.g., Cournot, would predict reduced markups in markets with more banks. On the other hand, if more banks exacerbate the adverse selection problem, this can lead to higher market power for banks that are better informed about certain borrowers.

To better analyze market power, we attempt to isolate the marginal cost of each loan based on borrower risk, bank funding costs, and loan processing costs. To do this, we estimate regressions predicting interest rates where we again control for bank-level funding costs through bank by time fixed effects and loan processing costs with county-level characteristics and wage data, but also control for the bank's perceived risk of the loan. The remaining portion of the interest rate, which is unexplained by risk, we refer to as the markup on the loan. Effectively, this approach allows us to compare two loans from the same bank at the same time, with the same level of risk. Importantly, as shown in Table IV, once we control for the risk of the loan, interest rates no longer predict loan performance. Hence, our measure of markup arguably is not driven by the ex-ante riskiness of the loan.

To test the relationship between local bank concentration and bank markups, we estimate the following regression:

$$
\begin{aligned}
IR_l = {} & \beta_0 NOB_c + \beta_1 PD + \beta_2 LGD + \beta_3 (PD \times LGD) \\
& + \Gamma_0 X_l + \Gamma_1 Z_{f,t} + \delta_{b,t} + \alpha_{i,t} + u_l, \qquad (5)
\end{aligned}
$$

This regression is the same as (4) but contains banks' internal risk assessments as controls and hence orthogonalizes the interest rate to the underlying risk of the loan. It is important to note that our measure of markup is relative, and not absolute as we are comparing the relative interest rate across loans after controlling for the risk of those loans. However, our main focus is analyzing relative markups across counties with different market structures.

Table IX shows that markups are higher in regions with more banks. In Column (2), where we control for firm characteristics, the coefficient on number of banks is 0.012 and statistically significant, implying that markups increase by 1.2bps for each additional bank in the county.

This result is consistent with adverse selection driving market power; however, as documented in Section III, interest rates do not monotonically increase in the number of banks. We thus test if markups exhibit a similar pattern by plotting the coefficients for various numbers of banks operating in the county in Figure 8. Consistent with our results on interest rates, the coefficients exhibit a U-shape in which markups first drop as the number of banks increases, but eventually begin to increase.

V. Cross-Sectional Analysis and Robustness Checks

We have thus far treated firms as homogenous; however, for certain types of firms, the asymmetric information problem may be less severe. In particular, firms with more tangible assets may be less subject to asymmetric information because tangible assets can be more easily valued (e.g., Frank and Goyal (2009)). Firms with more tangible assets also tend to use more collateral (e.g., Rajan and Zingales (1995) and Almeida and Campello (2007)), which makes it easier for banks to lend without having to collect extensive information (e.g., Manove, Padilla, and Pagano (2001)).

In Table X we split firms into groups based on whether their tangible assets are above the median in a given quarter (i.e., high tangibility) and test the relationship between number of banks and interest rates, PDs and markups separately for each of these groups.[31] For firms with high tangible assets (Columns (1), (3), and (5)), the relationship between the number of banks and interest rates, PDs, and markups is basically zero. Conversely, for firms with low tangible assets (Columns (2), (4), and (6)), the effect is strong and larger than our baseline effects (Column (3) of Tables V, VI,

[31]The level of tangibility is fairly high in our sample (as shown in Table I the median is 0.99) for two main reasons. First, for smaller firms, the majority of their assets are tangible. Second, the definition of tangible assets in Y-14Q is very broad and includes any asset with a physical presence, including cash and accounts receivables. However, many firms do have meaningful intangible assets (e.g., goodwill, trademarks and patents): for example, in the tenth percentile of firms, one-third of their assets are intangibles. Hence, these tests effectively compare firms with intangible assets to those without.

and IX). Hence, for firms for which unobservables are important, we see a strong relationship between the number of banks and interest rates, PDs, and markups, which we argue further supports the adverse selection channel.

Furthermore, in Online Appendix Tables 6 - 8, we also conduct placebo tests on public firms, for which we expect asymmetric information across banks in local markets to be less relevant. We find no relationship between the number of banks in the county and interest rates, PDs, and markups among publicly traded firms.[32]

One concern is that linear controls may not fully capture the relationship between county characteristics and markups. To address this issue, in Online Appendix Table 10, we show that our results are robust to matching counties based on population, population density, total wages, and financial industry wages using the methodology in Scharfstein and Sunderam (2016) and suggested by Imbens (2015). Relatedly, in Online Appendix Tables 22 - 24, we also show that our main results are robust to including interaction terms between the number of banks and other observable characteristics, and in Online Appendix Tables 27 - 29, we show that our results are robust to including non-linear terms for all control variables.

Loan demand could also be a confounding factor affecting both interest rates and the number of banks in a region. For example, banks may respond to higher loan demand and interest rates by entering markets.[33] However,

[32]In Online Appendix Tables 40 - 42, we also find similar results when we restrict firms to those classified as small businesses with less than $25 million in revenue. This threshold is based on the small-business classification used in the Tax Cuts and Jobs Act of 2017 (TCJA), as described by Sanati and Beyhaghi (2024).

[33]In contrast, it is not obvious why higher loan demand would lead to riskier borrowers

given that these are large banks with funds all around the country, we argue that the existing banks in the region should be able to move capital to the most profitable regions with high loan demand (e.g., Gilje, Loutskina, and Strahan, 2016). Specifically, without any information frictions, in a pure loan demand story, banks should funnel money to regions with the higher interest rate/markup markets until the marginal cost of capital at the aggregate bank-level equates the marginal profits of lending. However, the fact that we do see in equilibrium interest rates and markups vary across counties, suggests that information frictions can be preventing this from happening. In particular, adverse selection implies that private information regarding borrowers prevents banks from simply entering markets with high average markups because banks end up being stuck with the worst borrowers.[34] Indeed, our result showing firms rarely change banks, even in markets with many banks, is entirely consistent with this channel. Nonetheless, in Online Appendix Tables 19 - 21, we show that our main results are robust to controlling for several proxies of loan demand. Moreover, in Section VII we exploit a shock that differentially affects counties' lending conditions to further mitigate these various concerns.

VI. Markups and Switching Banks

The evidence in Section IV suggests that adverse selection is a source of market power for banks. In models of adverse selection, market power arises when a lender has superior information about a specific borrower. In this

receiving funding, as we observe in the data.

[34]See Dell'Ariccia, Friedman, and Marquez (1999) for a theory exploring this idea.

section, we explore this idea further in the context of repeated relationships. If banks gain an information advantage over other lenders through their lending relationships, they can hold up their borrowers and extract information rents.

To capture the information effect of repeat borrowers, we again restrict the sample to firms with more than one loan over the sample period and analyze all loans that follow their first loan. After making these restrictions, we estimate the following regression:

$$
\begin{aligned}
IR_l = {} & \beta_0 Stay\ Bank_l + \beta_1 PD + \beta_2 LGD + \beta_3 (PD \times LGD) \\
& + \Gamma_0 X_l + \Gamma_1 Z_{f,t} + \delta_{b,t} + \alpha_{i,t} + \lambda_{c,t} + u_l,
\end{aligned} \tag{6}
$$

where *Stay Bank* is a dummy that equals one when firms stay with their existing banks on their new loan, $\lambda_{c,t}$ is county by quarter fixed effects to control for unobserved differences in markups across regions and time. As in Table IX, we control for the banks' risk assessments to capture the variation in interest rates due to markups. The results, which we estimate with and without firm characteristics, are displayed in Table XI. Consistent with banks extracting information rents, we find that the estimated coefficient of Stay Bank is positive in all specifications and statistically significant except for Column (1) without firm-level controls, which is borderline statistically significant. For example, in Column (2), when we include firm characteristics, firms face 9bp higher markups when they remain with their existing bank.

If banks are indeed holding up their existing borrowers, we would expect this problem to be less severe if firms i) are borrowing from multiple banks or ii) borrowed from other banks in the past. In Column (3), we interact *Stay Bank* with the number of additional banks the firm has borrowed from in the sample. We find that the interaction coefficient is -0.052 and statistically significant. This result suggests that the information holdup problem becomes less severe the more competition there is among informed banks.

These results provide additional support for adverse selection driving market power in local loan markets. Next, we analyze a plausibly exogenous shock to banks' lending costs that affects local market structures and lending behavior.

VII. Capital Surcharges as a Shock to Bank Lending

We have found a positive relationship between the number of banks in a county and interest rates, markups, borrower risk, and lending volume. However, the number of banks is not randomly allocated to counties, and there could be omitted variables correlated with the number of banks and the lending outcomes. In Section IV we discuss how controlling for banks' private information, funding/operating costs, and proxies for loan demand can ameliorate these issues, but lingering concerns about the causality of the relationship may still exist. In this section, we use the capital surcharges imposed on global systemically important banks (GSIBs) as a shock to large banks' lending costs in local banking markets. Following the global financial crisis, regulators imposed additional surcharges on the largest global banks to strengthen their capital positions.[35] These capital surcharges were approved in July 2015 and went into effect in 2016.[36]

Several papers find that higher capital requirements reduce bank lending (e.g., Behn, Haselmann, and Vig, 2016 and Fraisse, Lé, and Thesmar, 2020). Similarly, Favara, Ivanov, and Rezende (2021) show that after these capital surcharges were phased in, affected banks decreased their lending relative to other banks. Hence, we hypothesize that counties with more GSIBs in 2015 will experience drops in the number of banks operating in the county and aggregate lending volume as these large banks find it more costly to lend. To test this hypothesis, we estimate the following difference-in-differences regression:

$$y_{t,c} = \beta_0 + \beta_1 NOG_{2015,c} \times Post_t + \Gamma_0 X_l + \gamma Z_{f,t} + \delta_{b,t} + \gamma_{b,c} + \alpha_{i,t} + u_{t,c}, \tag{7}$$

where the dependent variable is either the number of banks lending, calculated at the annual level, or the log of county-level loan volume calculated quarterly, $NOG_{2015,c}$ is the number of GSIBs in county c in 2015, and $Post_t$ is a dummy variable that equals one if the observation's year is greater than or equal to 2016. We measure the number of banks at the annual level to

[35] The eight US banks that are identified as GSIBs are: Bank of America Corporation, The Bank of New York Mellon Corporation, Citigroup, Inc., The Goldman Sachs Group, Inc., JPMorgan Chase & Co., Morgan Stanley, State Street Corporation, and Wells Fargo & Company.

[36] The implementation of the surcharges was staggered over time, beginning with 25% in 2016, followed by 25% increments each year until hitting 100% in 2019. See Favara, Ivanov, and Rezende (2021) for more institutional details regarding the capital surcharges.

minimize noise as best we can while still obtaining a measure that varies over time. Because our measure is annual, we also exclude 2014Q (the first quarter in our sample) from our sample in any regressions that include the number of banks at the annual level. We also include the same loan-level controls as in Table XI as well as bank by county fixed effects $\gamma_{b,c}$. We estimate these regressions at the loan-level even though the number of banks and loan volume are not measured at the loan-level because below we use $NOG_{2015,c} \times Post_t$ as an instrumental variable in a two-stage least squares regression. However, in Online Appendix Table 19, we show that the results regarding the number of banks and loan volume are robust to estimating these regressions at the county-year and county-quarter level, respectively. As before, we cluster our standard errors by county. The main coefficient of interest is β_1, which tells us how the outcome variable evolved differentially after the imposition of the surcharges across areas with different initial GSIB presences. Our approach is essentially a Bartik-style difference-in-differences (e.g., Goldsmith-Pinkham, Sorkin, and Swift, 2020).

For the GSIB surcharges to be a valid instrument, there must be no other shocks at that time that differentially affect outcome variables across counties with a different number of initial GSIB banks. Although this identifying assumption is not directly testable, when possible, we test for parallel trends prior to the imposition of the surcharges.

First, we validate that capital surcharges, and thus the presence of GSIBs in a region in 2015, significantly impact the local market structure. In Columns (1) and (2) of Table XII, the coefficients are both negative and statistically significant. An additional GSIB bank in a county in 2015 results in a

0.20 drop in the number of total banks in that county and about an 8.4% drop in lending volume following the imposition of the surcharges. We also plot the time series of annual interaction coefficients (i.e., $NOG_{2015,c} \times$ Year FE) in Figures 9 and 11. Importantly, Figure 11 does not exhibit any apparent pretrends.[37] Finally, we show that this drop in number of banks is specifically driven by a drop in GSIBs, not non-GSIBs in Figure 9 and Online Appendix Table 31.[38]

At first glance, the lending volume result seems contradictory to Favara, Ivanov, and Rezende (2021) as they find an increase in lending from non-GSIBs, which offsets the decrease in lending from GSIBs. However, their sample includes large corporate loans from publicly traded firms and syndicated loans, while ours does not. Consistent with their analysis, when we reestimate Figure 11 but include all syndicated loans and loans to public firms, we find no drop in aggregate lending (Online Appendix Figure 2). Hence, while total lending does not change, lending to smaller firms drops more in areas where GSIBs were more prevalent in 2015. If information frictions are more severe for these smaller private firms, it might make it difficult for other banks to simply take up the slack from the reduced GSIB lending. Indeed, many parts of our analysis are consistent with severe information frictions in these markets.

While standard models of competition predict that fewer banks and re-

[37]We cannot test for pretrends for the number of banks in Figure 9 since our sample begins in 2015 and the surcharges were imposed in 2016.

[38]In Online Appendix Table 32, we also show that the drop in lending volume is driven by GSIBs as well.

duced lending volume should reduce competition and lead to higher interest rates and markups, the adverse selection channel predicts that fewer banks lead to lower interest rates as the average quality of borrowers increases. Moreover, several theories of adverse selection and our earlier results suggest that more banks can also lead to higher markups.

We have shown that following the imposition of surcharges, there were drops in the number of banks and total lending in areas with a high number of GSIBs prior to the surcharges. Under the adverse selection channel, other banks in those markets should be less worried (or not worried at all) that a potential borrower has been denied a loan from one of these GSIBs, thereby reducing the adverse selection problem. If this were the case, we would first expect interest rates to decrease via a lower adverse selection discount. Consistent with this hypothesis, in Column (3) of Table XII we find a statistically significant decrease in interest rates. Second, with fewer potential competitors in the market and reduced adverse selection, lower quality borrowers now have fewer chances to find a lender who is optimistic about them (e.g., Broecker, 1990). This effect should cause a reduction in borrower risk in markets with more GSIBs. Also consistent with this hypothesis, in Column (4) we find a statistically significant drop in PD. Finally, with reduced adverse selection problems, the information rents banks can extract from their borrowers should drop. In Column (5), markups drop; however, the estimate is not quite statistically significant. We again plot the time series of annual interaction coefficients in Figures 12 - 14, which also do not exhibit any apparent pretrends. Taken together, these results support the channel through which the forced reduction in lending by GSIBs caused a reduction in adverse selection.

To better compare these reduced-form difference-in-differences estimates to the results in the first part of the paper, we next take an instrumental variables approach. Specifically, we use $NOG_{2015,c} \times Post_t$ as an instrument for the number of banks in the county (measured at the annual level) to test the relationship between the number of banks and lending volume, interest rates, PDs, and markups. The first stage is the same as Column (1) of Table XII.[39] To calculate the number of banks based on full years of data, the sample period for all instrumental variables specifications begins in 2015Q1 rather than 2014Q4.

The second stage results are displayed in Table XIII. Consistent with the reduced form analysis, the coefficients for loan volume, interest rates, and PD are all positive and statistically significant, while the coefficient for markups is positive but not quite statistically significant. More noticeably, the size of the IV coefficients are much larger than the OLS estimates we obtained in the main analysis. First, this is common in instrumental variable settings, particularly if our variable for the number of banks is measured with noise (e.g., Jiang, 2017 and Pancost and Schaller, 2021). Second, in this section, we measure the number of banks at the annual level, which is lower than the number of banks that ever lend in a county over the entire sample, mechanically increasing the magnitude of the coefficients. Third, while our main analysis presents the average effect across all counties with a wide range

[39]For the markup specifications, i.e., those in which we control for the risk assessments, the first stage is slightly different than Column (1) since it also includes PD, LGD, and Expected Loss.

of number of banks, inevitably, the GSIB shock is concentrated in counties with more banks in which our earlier results suggest the adverse selection channel dominates. Table II shows that the correlation between the number of banks in a county over the whole sample and the number of GSIBs in the county in 2015 is 0.79. Nonetheless, we view the economic magnitudes of the main variables of interest as reasonable: one additional bank leads to a 18bp higher interest rate and 40bp higher probability of default.[40]

One concern is that these results can be explained by GSIBs simply cutting back lending to the riskiest borrowers following a supply shock. In Online Appendix Table 33, we find no evidence that GSIBs cut their lending to high-risk borrowers, relative to non-GSIBs.[41] Moreover, Online Appendix Tables 34 and 35 show that our results hold if we restrict the sample to loans granted by non-GSIBs.

VIII. Conclusion

Adverse selection is a central problem in credit markets, but it is difficult to analyze because it is driven by the private information of both borrowers

[40]The coefficient for markup is 12bps, although as mentioned earlier, it is not statistically significant. The coefficient for loan volume is large and suggests one additional bank increases the amount of loan volume by over 50%; however, our IV approach assumes all changing in lending occurs at the extensive margin, i.e., the choice of banks to lend, where clearly the intensive margin, i.e., how much to lend, may also be affecting lending volume.

[41]In that table, we find that GSIBs cut lending relatively more to larger borrowers after the imposition of the surcharges. While exploring the mechanism for this is beyond the scope of this paper, one possibility is that banks earn lower rents on these loans, as large firms have better outside options and face fewer adverse selection problems.

and lenders. In this paper, we show that in markets with more banks, more low-quality borrowers receive financing, resulting in higher interest rates, banks' private risk assessments, and loan volume. We also develop a new methodology for measuring loan markups by controlling for the underlying risk of the borrower and provide evidence that adverse selection can be a source of market power for banks.

While we do not provide a welfare analysis of market structure, our results suggest that standard models of competition do not fully capture the subtleties of the impact of market structure on competition, borrower risk, and interest rates in corporate loan markets. Because of this, a potential unintended consequence of antitrust policies is that by making banking markets less concentrated, these policies may also lead to higher interest rates and riskier borrowers receiving financing.

REFERENCES

Abreu, Dilip, 1986, Extremal equilibria of oligopolistic supergames, *Journal of Economic Theory* 39, 191–225.

Almeida, Heitor and Murillo Campello, 2007, Financial constraints, asset tangibility, and corporate investment, *The Review of Financial Studies* 20, 1429–1460.

Axelson, Ulf and Igor Makarov, *Informational black holes in financial markets* (LSE Financial Markets Group 2016).

Begley, Taylor A, Amiyatosh Purnanandam, and Virginia Traweek, 2023, Shadows on main street: Mapping america's financial architecture from the yellow pages of the 20th century, *Available at SSRN 4639955* .

Begley, Taylor A and Kandarp Srinivasan, 2022, Small bank lending in the era of fintech and shadow banks: a sideshow?, *The Review of Financial Studies* 35, 4948–4984.

Behn, Markus, Rainer FH Haselmann, and Vikrant Vig, 2016, The limits of model-based regulation .

Berger, Allen and Gregory Udell, 1998, The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle, *Journal of Banking and Finance* 22, 613–673.

Berger, Allen N, 1995, The profit-structure relationship in banking–tests of market-power and efficient-structure hypotheses, *Journal of Money, Credit and Banking* 27, 404–431.

Berger, Allen N, Asli Demirgüç-Kunt, Ross Levine, and Joseph G Haubrich, 2004, Bank concentration and competition: An evolution in the making, *Journal of Money, credit and Banking* 433–451.

Berger, Allen N, Richard J Rosen, and Gregory F Udell, 2007, Does market size structure affect competition? the case of small business lending, *Journal of Banking & Finance* 31, 11–33.

Bertrand, Marianne, Antoinette Schoar, and David Thesmar, 2007, Banking deregulation and industry structure: Evidence from the french banking reforms of 1985, *The Journal of Finance* 62, 597–628.

Beyhaghi, Mehdi, Cooper Howes, and Gregory Weitzner, 2023, The information advantage of banks: Evidence from their private credit assessments .

Bidder, Rhys M, John R Krainer, and Adam Hale Shapiro, 2020, Deleveraging or de-risking? how banks cope with loss, *Review of Economic Dynamics* .

Bolton, Patrick, Tano Santos, and Jose A Scheinkman, 2016, Cream-skimming in financial markets, *The Journal of Finance* 71, 709–736.

Botsch, Matthew and Victoria Vanasco, 2019, Learning by lending, *Journal of Financial Intermediation* 37, 1–14.

Broecker, Thorsten, 1990, Credit-worthiness tests and interbank competition, *Econometrica: Journal of the Econometric Society* 429–452.

Brown, James R, Matthew Gustafson, and Ivan Ivanov, 2017, Weathering cash flow shocks, *Available at SSRN 2963444* .

Carbo-Valverde, Santiago, Francisco Rodriguez-Fernandez, and Gregory F Udell, 2009, Bank market power and sme financing constraints, *Review of Finance* 13, 309–340.

Carlson, Mark, Sergio Correia, and Stephan Luck, 2022, The effects of banking competition on growth and financial stability: Evidence from the national banking era, *Journal of Political Economy* 130, 462–520.

Cavalluzzo, Ken S, Linda C Cavalluzzo, and John D Wolken, 2002, Competition, small business financing, and discrimination: Evidence from a new survey, *The Journal of Business* 75, 641–679.

Cetorelli, Nicola, 2002, Life-cycle dynamics in industrial sectors. the role of banking market structure .

Cetorelli, Nicola and Michele Gambera, 2001, Banking market structure, financial dependence and growth: International evidence from industry data, *The Journal of Finance* 56, 617–648.

Chen, Brian S, Samuel G Hanson, and Jeremy C Stein, 2017, The decline of big-bank lending to small business: Dynamic impacts on local credit and labor markets, Working paper, National Bureau of Economic Research.

Chernenko, Sergey, Isil Erel, and Robert Prilmeier, 2022, Why do firms borrow directly from nonbanks?, *The Review of Financial Studies* 35, 4902–4947.

Claessens, Stijn and Luc Laeven, 2005, Financial dependence, banking sector competition, and economic growth, *Journal of the European Economic Association* 3, 179–207.

Crawford, Gregory S, Nicola Pavanini, and Fabiano Schivardi, 2018, Asymmetric information and imperfect competition in lending markets, *American Economic Review* 108, 1659–1701.

Cyrnak, Anthony W and Timothy H Hannan, 1999, Is the cluster still valid in defining banking markets? evidence from a new data source, *The Antitrust Bulletin* 44, 313–331.

Darmouni, Olivier, 2020, Informational frictions and the credit crunch, *The Journal of Finance* 75, 2055–2094.

Davydiuk, Tetiana, Tatyana Marchuk, and Samuel Rosen, 2024, Direct lenders in the us middle market, *Journal of Financial Economics* 162, 103946.

DeFusco, Anthony A, Huan Tang, and Constantine Yannelis, 2022, Measuring the welfare cost of asymmetric information in consumer credit markets, *Journal of Financial Economics* 146, 821–840.

Degryse, Hans and Steven Ongena, 2008, Competition and regulation in the banking sector: A review of the empirical evidence on the sources of bank rents, *Handbook of financial intermediation and banking* 2008, 483–554.

Dell'Ariccia, Giovanni, 2001, Asymmetric information and the structure of the banking industry, *European Economic Review* 45, 1957–1980.

Dell'Ariccia, Giovanni, Ezra Friedman, and Robert Marquez, 1999, Adverse selection as a barrier to entry in the banking industry, *The RAND Journal of Economics* 515–534.

Dell'Ariccia, Giovanni and Robert Marquez, 2006, Lending booms and lending standards, *The Journal of Finance* 61, 2511–2546.

DeLong, Elizabeth R, David M DeLong, and Daniel L Clarke-Pearson, 1988, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, *Biometrics* 837–845.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2017, The deposits channel of monetary policy, *The Quarterly Journal of Economics* 132, 1819–1876.

Favara, Giovanni, Ivan Ivanov, and Marcelo Rezende, 2021, Gsib surcharges and bank lending: Evidence from us corporate loan data, *Journal of Financial Economics* 142, 1426–1443.

Fishman, Michael J and Jonathan A Parker, 2015, Valuation, adverse selection, and market collapses, *The Review of Financial Studies* 28, 2575–2607.

Flannery, Mark J, 1996, Financial crises, payment system problems, and discount window lending, *Journal of money, credit and banking* 28, 804–824.

Fraisse, Henri, Mathias Lé, and David Thesmar, 2020, The real effects of bank capital requirements, *Management Science* 66, 5–23.

Frank, Murray Z and Vidhan K Goyal, 2009, Capital structure decisions: which factors are reliably important?, *Financial management* 38, 1–37.

Gilje, Erik P, Elena Loutskina, and Philip E Strahan, 2016, Exporting liquidity: Branch banking and financial integration, *The Journal of Finance* 71, 1159–1184.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, 2020, Bartik instruments: What, when, why, and how, *American Economic Review* 110, 2586–2624.

Goldstein, Itay, Chong Huang, and Liyan Yang, 2022, Open banking with depositor monitoring, *Unpublished manuscript* .

Gopal, Manasa and Philipp Schnabl, 2022, The rise of finance companies and fintech lenders in small business lending, *The Review of Financial Studies* 35, 4859–4901.

Granja, João, Christian Leuz, and Raghuram G Rajan, 2022, Going the extra mile: Distant lending and credit cycles, *The Journal of Finance* 77, 1259–1324.

Hannan, Timothy H, 1991, Bank commercial loan markets and the role of market structure: Evidence from surveys of commercial lending, *Journal of Banking & Finance* 15, 133–149.

Hannan, Timothy H, 1997, Market share inequality, the number of competitors, and the hhi: An examination of bank pricing, *Review of Industrial Organization* 12, 23–35.

He, Zhiguo, Jing Huang, and Jidong Zhou, 2020, Open banking: credit market competition when borrowers own the data, Working paper, National Bureau of Economic Research.

Imbens, Guido W, 2015, Matching methods in practice: Three examples, *Journal of Human Resources* 50, 373–419.

Jayaratne, Jith and Philip E Strahan, 1996, The finance-growth nexus: Evidence from bank branch deregulation, *The Quarterly Journal of Economics* 111, 639–670.

Jiang, Wei, 2017, Have instrumental variables brought us closer to the truth, *The Review of Corporate Finance Studies* 6, 127–140.

Keeley, Michael C, 1990, Deposit insurance, risk, and market power in banking, *The American economic review* 1183–1200.

Liebersohn, Jack, 2017, How does competition affect bank lending? quasi-experimental evidence from bank mergers, Working paper, Technical report.

Manove, Michael, A Jorge Padilla, and Marco Pagano, 2001, Collateral versus project screening: A model of lazy banks, *Rand journal of economics* 726–744.

Marquez, Robert, 2002, Competition, adverse selection, and information dispersion in the banking industry, *The Review of Financial Studies* 15, 901–926.

Milgrom, Paul R, 1979, A convergence theorem for competitive bidding with differential information, *Econometrica: Journal of the Econometric Society* 679–688.

Neumark, David and Steven A Sharpe, 1992, Market structure and the nature of price rigidity: evidence from the market for consumer deposits, *The Quarterly Journal of Economics* 107, 657–680.

Pancost, N. Aaron and Garrett Schaller, 2021, Measuring measurement error, *Working Paper* .

Paravisini, Daniel, Veronica Rappoport, and Philipp Schnabl, 2023, Specialization in bank lending: Evidence from exporting firms, *The Journal of Finance* 78, 2049–2085.

di Patti, Emilia Bonaccorsi and Giovanni Dell'Ariccia, 2004, Bank competition and firm creation, *Journal of Money, Credit and Banking* 225–251.

Petersen, Mitchell A and Raghuram G Rajan, 1995, The effect of credit market competition on lending relationships, *The Quarterly Journal of Economics* 110, 407–443.

Plosser, Matthew C and Joao AC Santos, 2018, Banks' incentives and inconsistent risk models, *The Review of Financial Studies* 31, 2080–2112.

Rajan, Raghuram G, 1992, Insiders and outsiders: The choice between informed and arm's-length debt, *The Journal of finance* 47, 1367–1400.

Rajan, Raghuram G and Luigi Zingales, 1995, What do we know about capital structure? some evidence from international data, *The journal of Finance* 50, 1421–1460.

Rhoades, Stephen A, 1995, Market share inequality, the hhi, and other measures of the firm-composition of a market, *Review of industrial organization* 10, 657–674.

Rice, Tara and Philip E Strahan, 2010, Does credit competition affect small-firm finance?, *The Journal of Finance* 65, 861–889.

Riordan, Michael H, 1995, 11 competition and bank performance: a theoretical perspective, *Capital markets and financial intermediation* 328.

Saidi, Farzad and Daniel Streitz, 2018, Bank concentration and product market competition, *Work. Pap., Stockh. Sch. Econ., Stockholm Google Scholar Article Location* .

Sanati, Ali and Mehdi Beyhaghi, 2024, How does removing the tax benefits of debt affect firms? evidence from the 2017 us tax reform, *Available at SSRN 3878363* .

Sapienza, Paola, 2002, The effects of banking mergers on loan contracts, *The Journal of finance* 57, 329–367.

Scharfstein, David and Adi Sunderam, 2016, Market power in mortgage lending and the transmission of monetary policy, *Unpublished working paper. Harvard University* 2.

Shaffer, Sherrill, 1998, The winner's curse in banking, *Journal of Financial Intermediation* 7, 359–392.

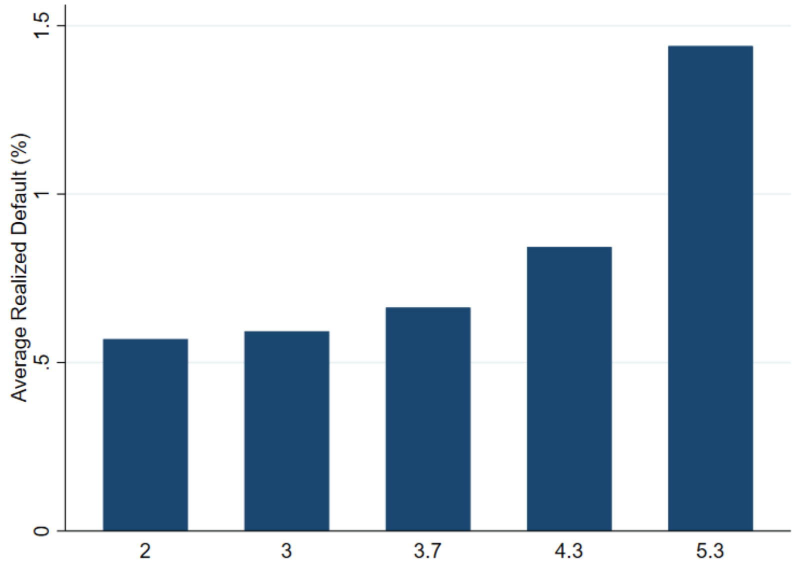
Sharpe, Steven A, 1990, Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships, *The journal of finance* 45, 1069–1087.

Stroebel, Johannes, 2016, Asymmetric information about collateral values, *The Journal of Finance* 71, 1071–1112.

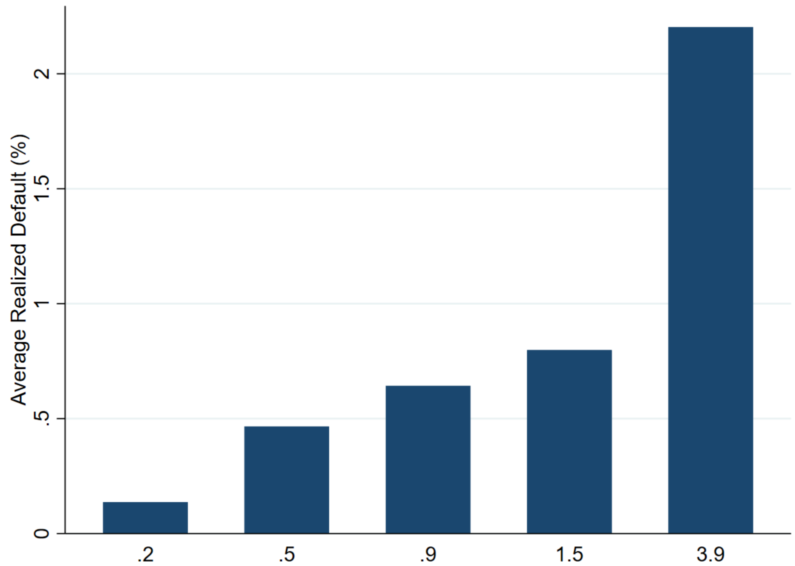
Weitzner, Gregory and Cooper Howes, 2021, Bank information production over the business cycle, *Available at SSRN* .

Wiersch, Ann Marie, Barbara J Lipman, Kim Wilson, and Lucas Misera, 2022, Clicking for credit: Experiences of online lender applicants from the small business credit survey, *Community Development Reports* .

Yannelis, Constantine and Anthony Lee Zhang, 2022, Competition and selection in credit markets, Working paper, National Bureau of Economic Research.



Panel A. Average Realized Default Rate Across Interest Rate Bins



Panel B. Average Realized Default Rate Across PD Bins

Figure 1. Average Realized Default Rates. Figure 1A plots the average realized default rates over the twelve months following origination across five interest rate bins. Figure 1B plots the average realized default rates across five PD bins. The average interest rate or PD in each bin is listed below each bar.

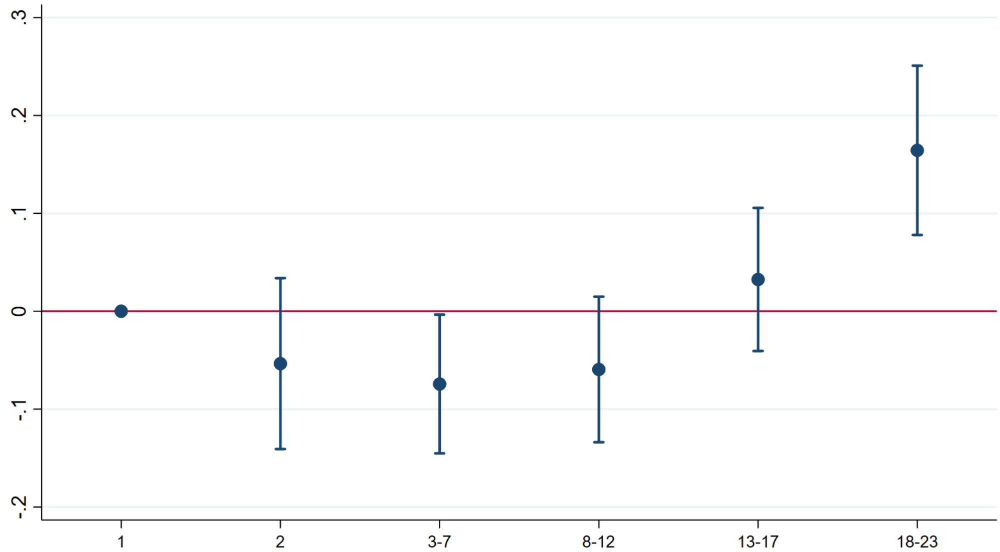


Figure 2. Number of Banks in the County and Interest Rates. This figure plots estimated coefficients, with 90% confidence intervals, for regressions of interest rates on different number of bank group dummies as well as the same controls as in Column (2) of Table V. The number of banks in each group is listed below, where the reference number of banks is 1. Standard errors are clustered by county.

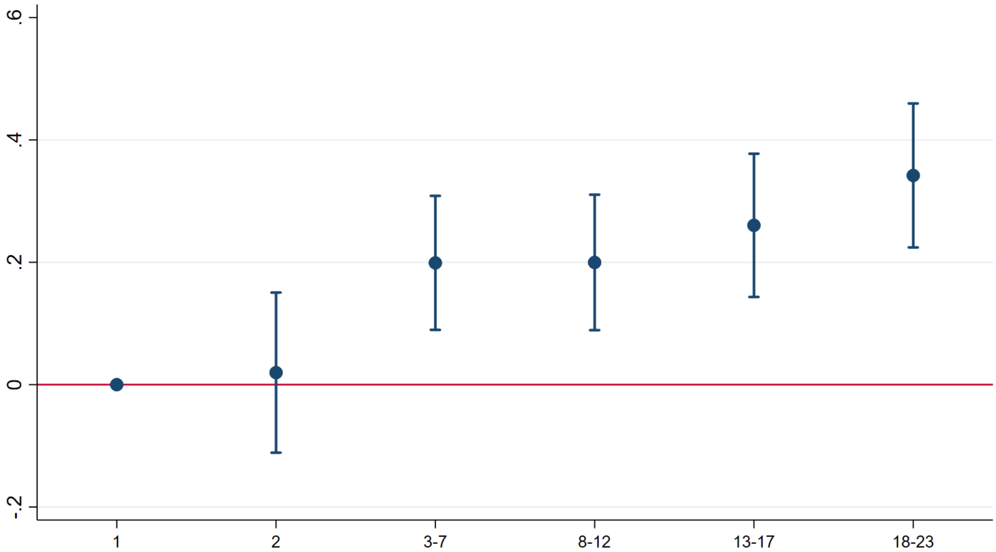


Figure 3. Number of Banks in the County and PDs. This figure plots estimated coefficients, with 90% confidence intervals, for regressions of PD on different number of bank group dummies as well as the same controls as in Column (2) of Table VI. The number of banks in each group is listed below, where the reference number of banks is 1. Standard errors are clustered by county.

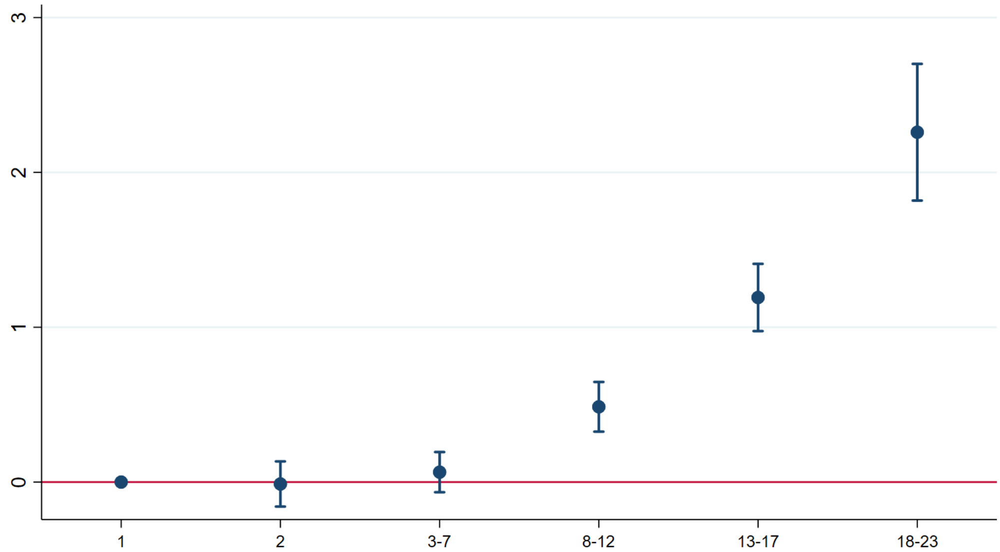
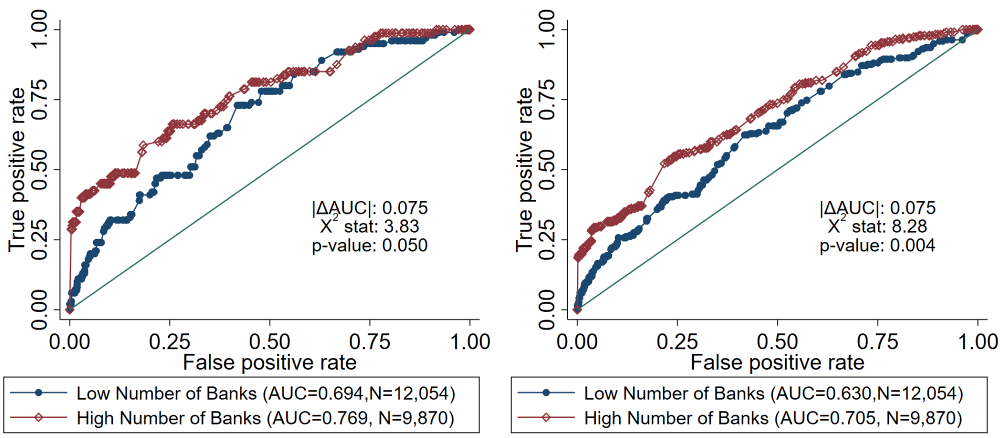


Figure 4. Number of Banks in the County and Loan Volume. This figure plots estimated coefficients, with 90% confidence intervals, for regressions of loan volume on different number of bank group dummies as well as the same controls as in Column (2) of Table VII. The number of banks in each group is listed below, where the reference number of banks is 1. Standard errors are clustered by county.



Panel A. Realized Default

Panel B. Nonperformance

Figure 5. Market Structure and PD Accuracy. The left panel plots ROC curves split by whether the number of banks is above the median over the sample, with PD as the predictor and Realized Default as the outcome variable. The right panel uses PD as the predictor and Nonperformance as the outcome variable. The area under each ROC curve (AUC) is reported along with the number of observations in the legend. |ΔAUC| reports the difference between the two AUCs. Below |ΔAUC|, the DeLong, DeLong, and Clarke-Pearson (1988) statistics are reported: the χ^2 test statistic and its corresponding p-value are reported, which test the null hypothesis that the difference between the two AUC values equals zero.

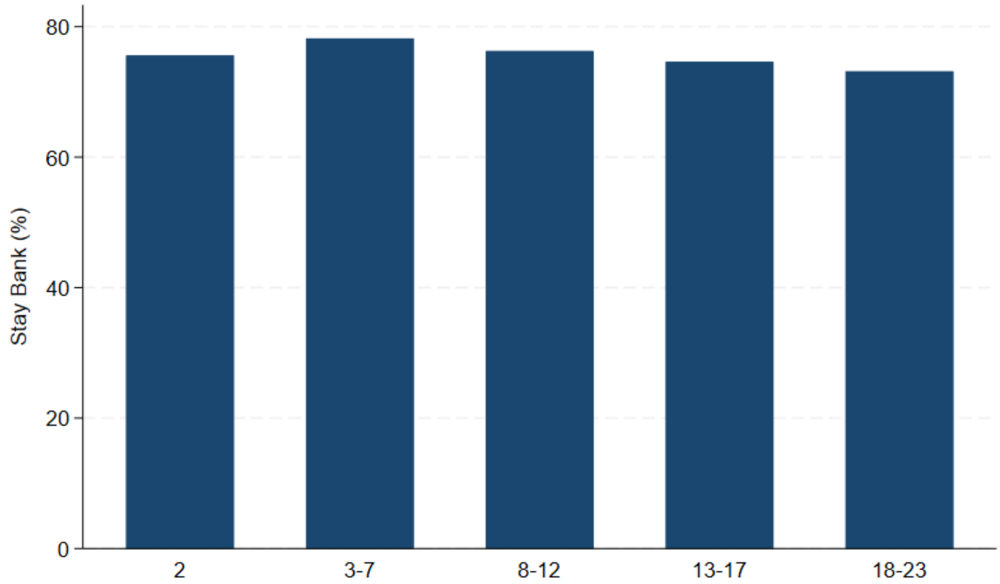


Figure 6. The Frequency of Firms Staying With Their Existing Banks. This figure plots the frequency of firms staying with one of their existing banks, i.e., *Stay Bank* across counties with different numbers of banks. Loans from counties with one bank are excluded.

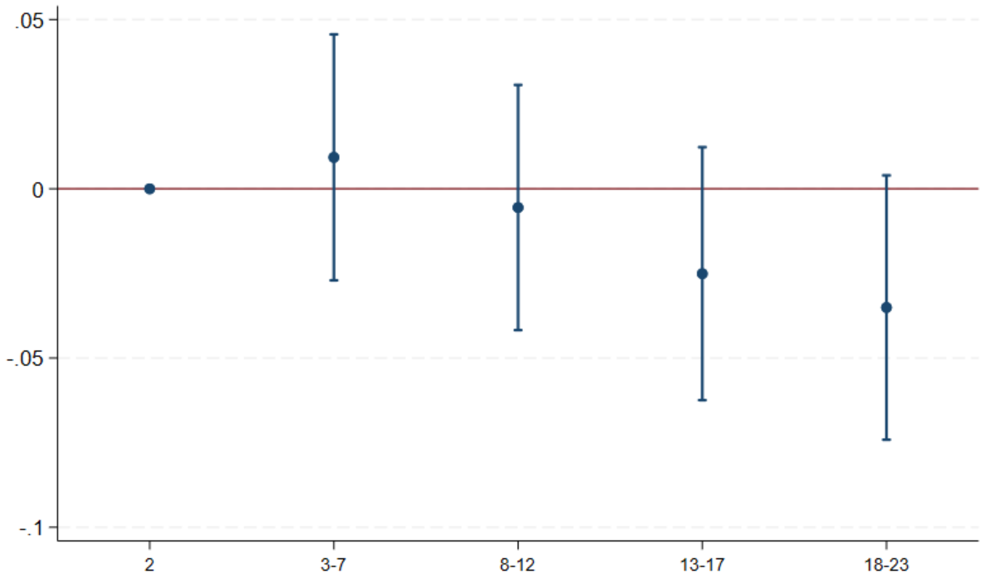


Figure 7. The Frequency of Firms Staying With Their Existing Banks (Regression Analysis). This figure plots estimated coefficients, with 90% confidence intervals, for regressions of *Stay Bank* on different number of bank group dummies as well as the same controls as in Column (2) of Table VI. The number of banks in each group is listed below, where the reference number of banks is 2. Loans from counties with one bank are excluded. Standard errors are clustered by county.

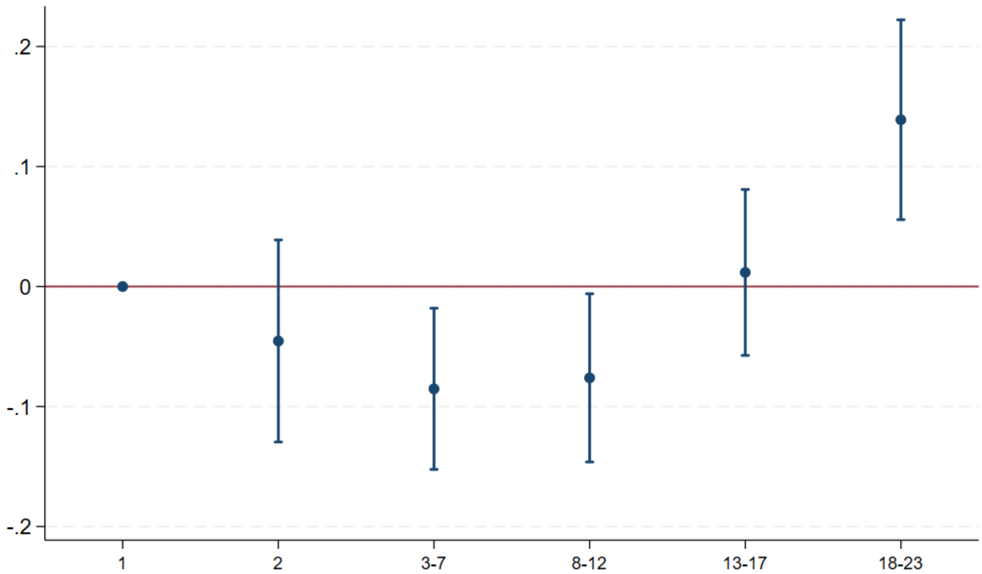


Figure 8. Number of Banks in the County and Markups. This figure plots estimated coefficients, with 90% confidence intervals, for regressions of markup on different number of bank group dummies as well as the same controls as in Column (2) of Table IX. We refer to markups as any variation in interest rates after controlling for the risk of the loan. The number of banks in each group is listed below, where the reference number of banks is 1. Standard errors are clustered by county.

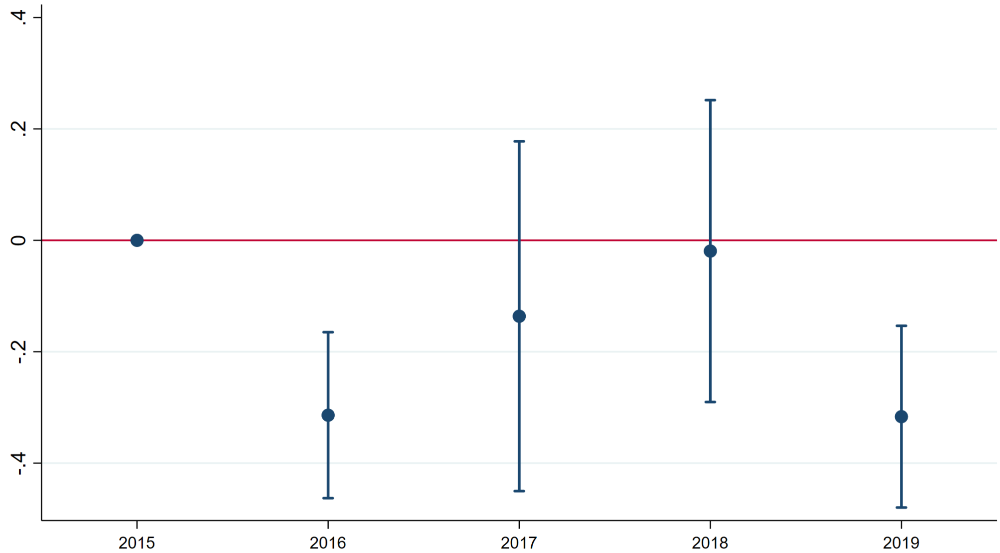


Figure 9. The Effect of GSIB Surcharges on the Number of Banks. This figure plots estimated regression coefficients with 90% confidence intervals from a version of (7) with annual interaction terms and the number of banks (calculated annually) as the dependent variable. Standard errors are clustered by county.

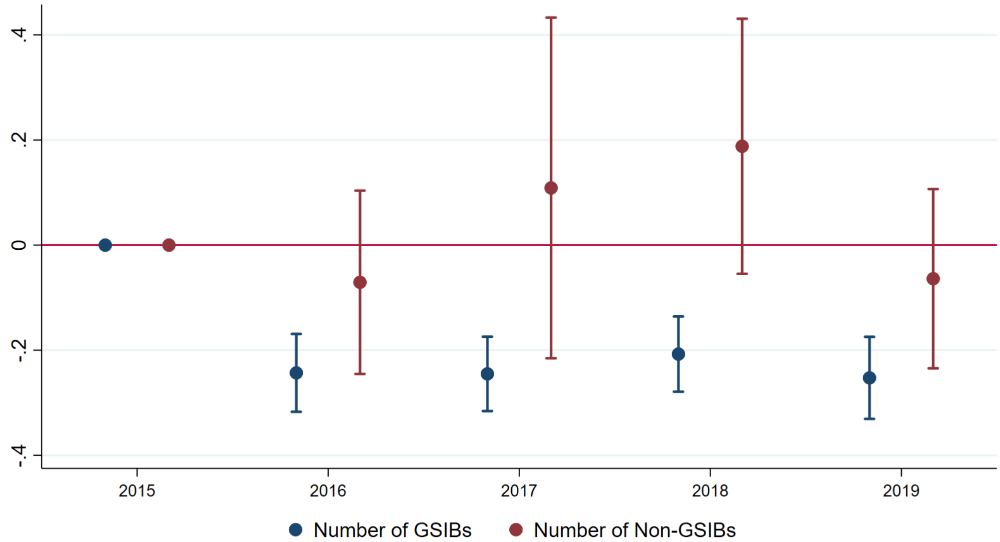


Figure 10. The Effect of GSIB Surcharges on the Number of GSIB and Non-GSIBs. This figure plots estimated regression coefficients with 90% confidence intervals from a version of (7) with annual interaction terms and either the number of GSIBs or non-GSIBs (both calculated annually) as the dependent variable. Standard errors are clustered by county.

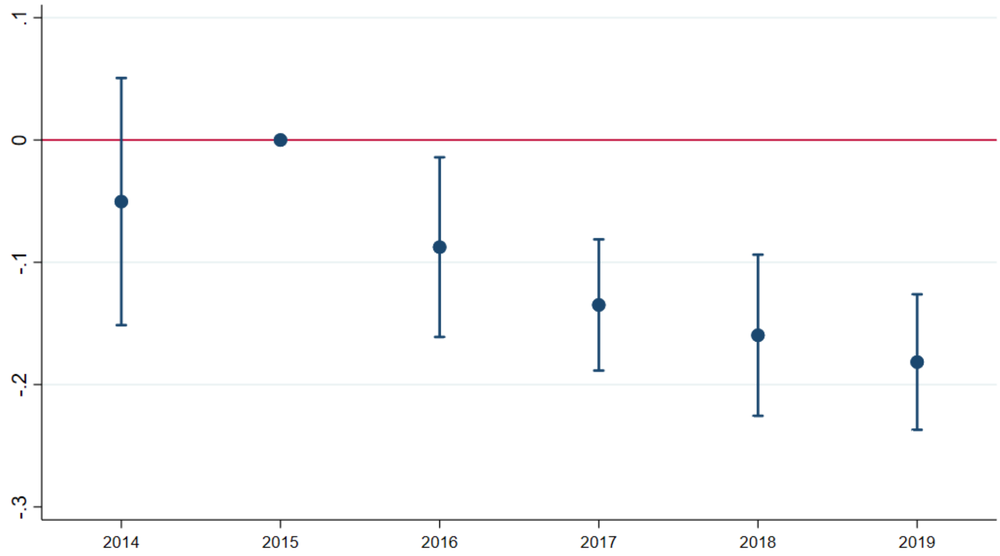


Figure 11. The Effect of GSIB Surcharges on Lending Volume. This figure plots estimated regression coefficients with 90% confidence intervals from a version of (7) with annual interaction terms and the log of loan volume as the dependent variable. Standard errors are clustered by county.

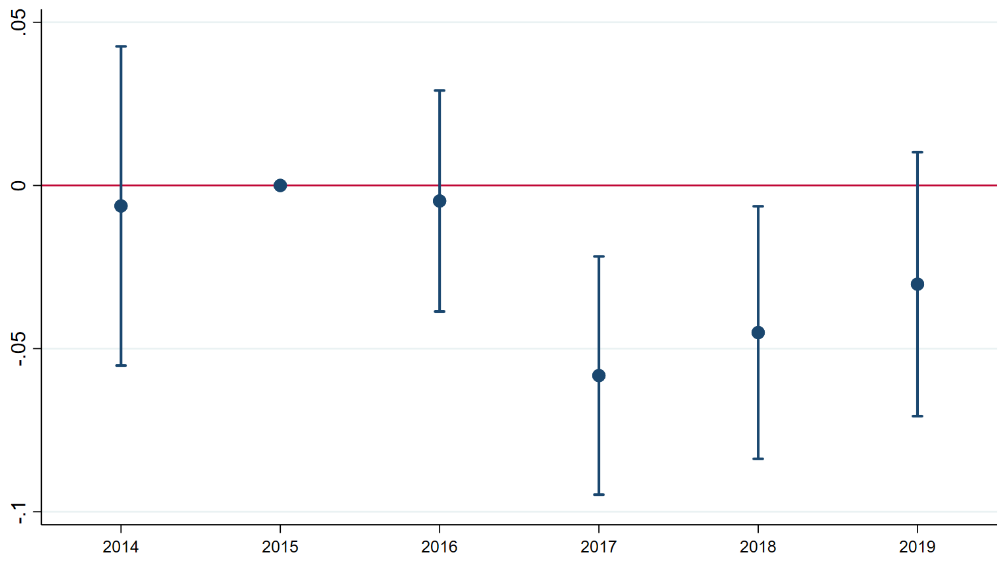


Figure 12. The Effect of GSIB Surcharges on Interest Rates. This figure plots estimated regression coefficients with 90% confidence intervals from a version of (7) with annual interaction terms and Interest Rate (%) as the dependent variable. Standard errors are clustered by county.

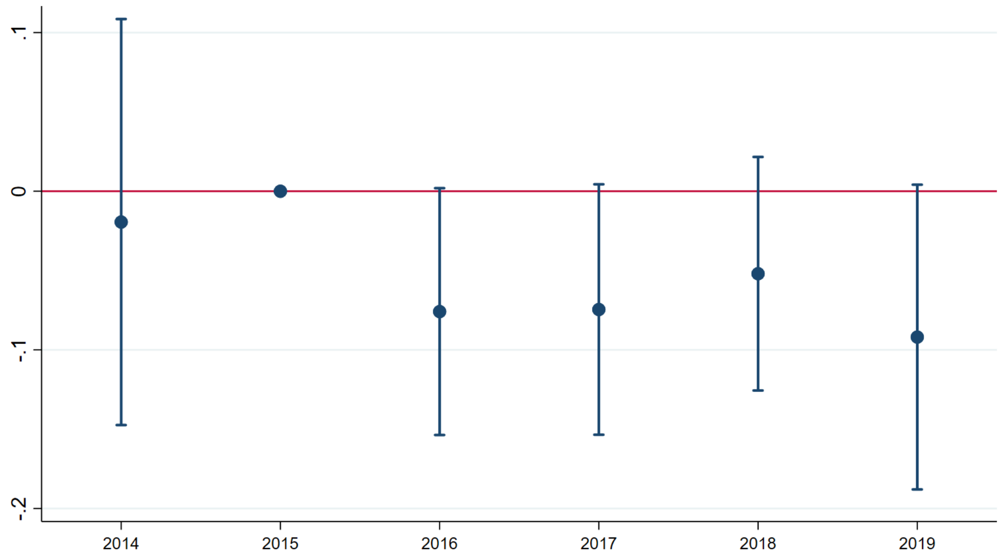


Figure 13. The Effect of GSIB Surcharges on PDs. This figure plots estimated regression coefficients with 90% confidence intervals from a version of (7) with annual interaction terms and Probability of Default (%) as the dependent variable. Standard errors are clustered by county.

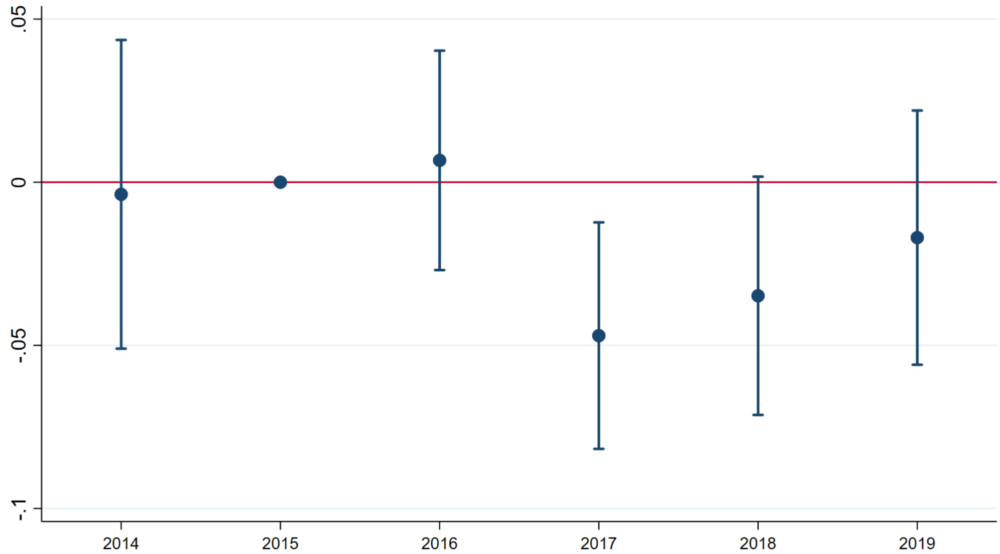


Figure 14. The Effect of GSIB Surcharges on Markups. This figure plots estimated regression coefficients with 90% confidence intervals from a version of (7) with annual interaction terms and Interest Rate (%) as the dependent variable, while controlling for the risk of the loan (i.e., PD, LGD and Expected Loss). Standard errors are clustered by county.

Table I. Summary Statistics

This table contains summary statistics for loan-level, firm and geographic characteristics. The appendix includes detailed definitions of all of our variables and Section II explains our filters.

	Mean	SD	10%	Median	90%	N
Loan Characteristics						
Amount (million USD)	7.17	14.74	1.06	2.66	16.19	21,924
Interest Rate (%)	3.66	1.17	2.16	3.66	5.24	21,924
Probability of Default (%)	1.34	1.70	0.20	0.87	2.79	21,924
Loss Given Default (%)	35.11	14.68	15.00	36.00	50.00	21,924
Expected Loss (%)	0.45	0.58	0.06	0.27	0.97	21,924
Floating Interest Rate	0.79	0.41	0.00	1.00	1.00	21,924
Guaranteed	0.50	0.50	0.00	0.00	1.00	21,924
Maturity (months)	41.70	30.25	11.00	36.00	84.00	21,924
Non-Performance (%)	2.05	14.18	0.00	0.00	0.00	21,924
Number of Prior Lenders	0.82	1.76	0.00	0.00	2.00	21,924
New Borrower	0.27	0.44	0.00	0.00	1.00	21,924
Line of Credit	0.50	0.50	0.00	0.00	1.00	21,924
Realized Default (%)	0.82	9.02	0.00	0.00	0.00	21,924
Secured	0.91	0.29	1.00	1.00	1.00	21,924
Secured by Blanket Lien	0.38	0.49	0.00	0.00	1.00	21,923
Stay Bank	0.76	0.43	0.00	1.00	1.00	16,087
GSIB	0.42	0.49	0.00	0.00	1.00	21,924
Firm Characteristics						
Assets (million USD)	153.50	706.61	4.30	23.59	207.67	21,924
Net Sales (million USD)	229.50	1,247.45	9.22	45.98	337.85	21,854
Leverage	0.34	0.25	0.02	0.31	0.68	21,480
Profitability	0.12	0.18	-0.00	0.07	0.28	21,924
Tangibility	0.91	0.17	0.67	0.99	1.00	21,866
Geographic Characteristics						
Number of Banks	11.64	5.86	4.00	12.00	20.00	21,924
Number of Banks (Annual)	5.74	3.77	1.00	5.00	11.00	21,924
Number of GSIBs (2015)	2.19	1.25	0.00	2.00	4.00	20,420
Number of All Banks	31.06	24.27	9.00	24.00	71.00	21,904
Population Density	6.83	1.54	4.72	7.05	8.30	21,924
Wages	9.50	0.27	9.19	9.48	9.81	21,907
Financial Industry Wages	9.84	0.39	9.39	9.81	10.28	21,901
Population	13.28	1.36	11.42	13.48	14.82	21,924
Deposit HHI	0.20	0.08	0.11	0.18	0.29	21,924
Loan HHI	0.50	0.25	0.22	0.43	0.91	21,924

Table II. Correlation Between Measures of Market Concentration

This table contains a correlation matrix containing different measures of market concentration as well as the number of GSIBs present in 2015 at the county level.

Variables	(i)	(ii)	(iii)	(iv)	(v)
(i) Number of Banks	1.00				
(ii) Deposit HHI	-0.27	1.00			
(iii) Loan HHI	-0.90	0.23	1.00		
(iv) Number of GSIBs (2015)	0.79	-0.16	-0.77	1.00	
(v) Number of Banks (Branches)	0.84	-0.32	-0.69	0.70	1.00

Table III. Interest Rates and Risk Assessments

This table tests whether banks' internal risk assessments predict loan interest rates. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Interest Rate (%)	
	(1)	(2)
Probability of Default (%)		0.077***
		(4.660)
Loss Given Default (%)		0.003***
		(4.366)
Expected Loss (%)		0.155***
		(3.542)
Log(maturity)	−0.001	0.016*
	(0.154)	(1.842)
Log(Amount)	−0.159***	−0.151***
	(20.598)	(20.563)
Guaranteed	0.062***	0.061***
	(3.255)	(3.416)
Loan Purpose FE	YES	YES
Loan Type FE	YES	YES
Bank-Quarter FE	YES	YES
Industry-Quarter FE	YES	YES
Observations	21,853	21,853
Adj. R-squared	0.52	0.55

Table IV. Interest Rates, Risk Assessments and Loan Performance

This table tests whether interest rates and banks' internal risk assessments predict non-performance and default after controlling for loan characteristics and fixed effects. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Non-Performance (%)			Realized Default (%)		
	(1)	(2)	(3)	(4)	(5)	(6)
Interest Rate (%)	0.527***		0.101	0.354***		0.093
	(3.699)		(0.541)	(3.346)		(0.620)
Probability of Default (%)		1.392**	1.384**		0.888***	0.880**
		(2.259)	(2.206)		(2.624)	(2.524)
Loss Given Default (%)		0.028	0.028		0.014	0.014
		(1.501)	(1.450)		(1.469)	(1.375)
Expected Loss (%)		−1.229	−1.244		−0.816	−0.830
		(0.918)	(0.940)		(1.262)	(1.307)
Loan Controls	YES	YES	YES	YES	YES	YES
Bank-Quarter FE	YES	YES	YES	YES	YES	YES
Industry-Quarter FE	YES	YES	YES	YES	YES	YES
Observations	18,246	18,246	18,246	18,246	18,246	18,246
Adj. R-squared	0.09	0.10	0.10	0.07	0.08	0.08

Table V. Market Structure and Interest Rates

This table tests the relationship between the number of banks and interest rates. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Interest Rate (%)		
	(1)	(2)	(3)
Number of Banks	0.012***	0.013***	0.009**
	(6.338)	(6.318)	(2.433)
Log(Assets)		−0.151***	−0.152***
		(20.304)	(20.735)
Leverage		0.204***	0.207***
		(6.366)	(6.559)
Tangibility		−0.671***	−0.672***
		(14.876)	(14.922)
Profitability		−0.388***	−0.382***
		(9.027)	(9.070)
Population Density			−0.008
			(0.369)
Wages			0.142
			(1.479)
Financial Industry Wages			0.037
			(0.569)
Loan Controls	YES	YES	YES
Bank-Quarter FE	YES	YES	YES
Industry-Quarter FE	YES	YES	YES
Observations	21,853	21,388	21,348
Adj. R-squared	0.52	0.54	0.54

Table VI. Market Structure and Borrower Risk

This table tests the relationship between the number of banks and probability of default (PD). T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Probability of Default (%)		
	(1)	(2)	(3)
Number of Banks	0.008***	0.011***	0.009**
	(3.470)	(4.970)	(2.330)
Log(Assets)		−0.138***	−0.137***
		(10.845)	(10.764)
Leverage		0.963***	0.961***
		(11.562)	(11.648)
Tangibility		−0.235***	−0.232***
		(2.757)	(2.715)
Profitability		−1.822***	−1.828***
		(24.588)	(24.871)
Population Density			0.045***
			(3.164)
Wages			−0.169
			(1.536)
Financial Industry Wages			−0.034
			(0.327)
Loan Controls	YES	YES	YES
Bank-Quarter FE	YES	YES	YES
Industry-Quarter FE	YES	YES	YES
Observations	21,853	21,388	21,348
Adj. R-squared	0.17	0.23	0.23

Table VII. Market Structure and Loan Volume

This table tests the relationship between the number of banks and loan volume. Loan volume is measured in logs and is aggregated at the county-quarter level. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Log(loan volume)		
	(1)	(2)	(3)
Number of Banks	0.144***	0.140***	0.125***
	(21.235)	(14.662)	(11.707)
Population Density		−0.040*	−0.065**
		(1.821)	(2.535)
Wages		0.161	0.213
		(0.945)	(1.255)
Financial Industry Wages		0.173*	0.160
		(1.766)	(1.624)
Population			0.086**
			(2.164)
Quarter FE	YES	YES	YES
Observations	8,060	8,026	8,026
Adj. R-squared	0.26	0.27	0.27

Table VIII. Market Structure and Collateralization

This table tests the relationship between the number of banks and whether loans contain a blanket lien. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Secured by Blanket Lien		
	(1)	(2)	(3)
Number of Banks	0.002***	0.002***	0.002**
	(2.977)	(3.116)	(2.246)
Log(Assets)		−0.017***	−0.017***
		(7.530)	(7.535)
Leverage		−0.033***	−0.034***
		(2.685)	(2.743)
Tangibility		−0.076***	−0.078***
		(4.070)	(4.192)
Profitability		0.078***	0.076***
		(4.999)	(4.904)
Population Density			−0.003
			(0.656)
Wages			−0.029
			(0.976)
Financial Industry Wages			0.023
			(1.143)
Loan Controls	YES	YES	YES
Bank-Quarter FE	YES	YES	YES
Industry-Quarter FE	YES	YES	YES
Observations	21,853	21,388	21,348
Adj. R-squared	0.56	0.56	0.56

Table IX. Market Structure and Markups

This table tests the relationship between the number of banks and markups. We refer to markups as any variation in interest rates after controlling for the risk of the loan. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Interest Rate (%)		
	(1)	(2)	(3)
Number of Banks	0.011***	0.012***	0.008**
	(5.948)	(5.844)	(2.343)
Probability of Default (%)	0.076***	0.064***	0.064***
	(4.615)	(3.875)	(3.988)
Loss Given Default (%)	0.003***	0.002***	0.002***
	(4.360)	(2.962)	(2.950)
Expected Loss (%)	0.155***	0.143***	0.142***
	(3.532)	(3.303)	(3.330)
Log(Assets)		−0.132***	−0.133***
		(17.901)	(18.270)
Leverage		0.108***	0.111***
		(3.474)	(3.630)
Tangibility		−0.624***	−0.625***
		(14.289)	(14.322)
Profitability		−0.200***	−0.192***
		(4.721)	(4.627)
Population Density			−0.012
			(0.620)
Wages			0.165*
			(1.789)
Financial Industry Wages			0.032
			(0.494)
Loan Controls	YES	YES	YES
Bank-Quarter FE	YES	YES	YES
Industry-Quarter FE	YES	YES	YES
Observations	21,853	21,388	21,348
Adj. R-squared	0.55	0.56	0.56

Table X. Adverse Selection and Firm Tangibility

This table tests the relationship between the number of banks and interest rates, PDs and markups for high- and low-tangibility firms. In each quarter, we sort loan observations into two groups based on borrower tangibility: above-median (high-tangibility) and below-median (low-tangibility). T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Interest Rate (%)		Probability of Default (%)		Interest Rate (%)	
	High Tangibility	Low Tangibility	High Tangibility	Low Tangibility	High Tangibility	Low Tangibility
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Banks	0.003	0.013***	0.000	0.014**	0.003	0.011***
	(0.681)	(3.769)	(0.101)	(2.466)	(0.808)	(3.234)
Probability of Default (%)					0.051***	0.101***
					(2.653)	(6.528)
Loss Given Default (%)					0.003***	0.003***
					(2.856)	(2.841)
Expected Loss (%)					0.140***	0.087*
					(2.700)	(1.869)
Log(Assets)	−0.203***	−0.105***	−0.161***	−0.127***	−0.184***	−0.086***
	(17.683)	(10.326)	(7.647)	(8.672)	(17.435)	(8.336)
Leverage	0.170***	0.279***	1.064***	0.703***	0.079*	0.193***
	(3.729)	(6.638)	(9.905)	(8.736)	(1.808)	(4.652)
Tangibility	7.593	−0.680***	23.017**	−0.345***	5.867	−0.622***
	(1.309)	(12.106)	(2.221)	(3.998)	(1.001)	(11.449)
Profitability	−0.257***	−0.624***	−1.594***	−2.251***	−0.114**	−0.348***
	(4.990)	(8.981)	(17.261)	(17.249)	(2.307)	(4.617)
Population Density	−0.000	−0.016	0.057**	0.020	−0.006	−0.018
	(0.022)	(0.843)	(2.558)	(1.066)	(0.297)	(0.992)
Wages	0.080	0.221**	−0.129	−0.124	0.100	0.235**
	(0.735)	(1.995)	(0.804)	(0.799)	(0.945)	(2.264)
Financial Industry Wages	0.086	−0.038	−0.177	0.029	0.088	−0.041
	(1.234)	(0.458)	(1.023)	(0.268)	(1.251)	(0.546)
Loan Controls	YES	YES	YES	YES	YES	YES
Bank-Quarter FE	YES	YES	YES	YES	YES	YES
Industry-Quarter FE	YES	YES	YES	YES	YES	YES
Observations	10,563	10,709	10,563	10,709	10,563	10,709
Adj. R-squared	0.55	0.56	0.32	0.19	0.57	0.59

Table XI. Switching Banks and Markups

This table tests whether firms that stay with their existing banks face higher markups. We refer to markups as any variation in interest rates after controlling for the risk of the loan. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Interest Rate (%)		
	(1)	(2)	(3)
Stay Bank	0.040	0.090***	0.160***
	(1.647)	(3.637)	(4.925)
N. of Prior Lenders			0.037***
			(2.701)
Stay Bank × N. of Prior Lenders			−0.052***
			(3.504)
Probability of Default (%)	0.102***	0.089***	0.088***
	(4.522)	(4.003)	(3.945)
Loss Given Default (%)	0.006***	0.005***	0.005***
	(5.047)	(4.654)	(4.630)
Expected Loss (%)	0.061	0.044	0.045
	(0.954)	(0.706)	(0.724)
Log(Assets)		−0.131***	−0.127***
		(12.135)	(10.789)
Leverage		0.187***	0.193***
		(4.110)	(4.246)
Tangibility		−0.677***	−0.676***
		(9.379)	(9.300)
Profitability		−0.288***	−0.285***
		(5.006)	(5.010)
Loan Controls	YES	YES	YES
County-Quarter FE	YES	YES	YES
Bank-Quarter FE	YES	YES	YES
Industry-Quarter FE	YES	YES	YES
Observations	12,257	11,915	11,915
Adj. R-squared	0.61	0.62	0.62

Table XII. GSIB Surcharges and Market Outcomes (Reduced Form Difference-in-Differences)

This table contains reduced form difference-in-differences regressions testing whether the number of GSIBs induces changes in market outcomes following the imposition of capital surcharges. The sample period is 2014Q4 - 2019Q4 except for Column (1) which is 2015Q1 - 2019Q4 in order to calculate the number of banks based on a full year. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	N. of Banks	Loan Volume (log)	Interest Rate (%)	Probability of Default (%)	Interest Rate (%)
	(1)	(2)	(3)	(4)	(5)
Post × N. of GSIBs (2015)	-0.201^{**}	-0.084^{***}	-0.031^{*}	-0.069^{*}	-0.021
	(2.443)	(3.158)	(1.957)	(1.676)	(1.317)
Probability of Default (%)					0.078^{***}
					(3.756)
Loss Given Default (%)					0.005^{***}
					(4.592)
Expected Loss (%)					0.103^{*}
					(1.771)
Loan Controls	YES	YES	YES	YES	YES
Bank-County FE	YES	YES	YES	YES	YES
Bank-Quarter FE	YES	YES	YES	YES	YES
Industry-Quarter FE	YES	YES	YES	YES	YES
Observations	12,697	13,640	13,640	13,640	13,640
Adj. R-squared	0.91	0.72	0.61	0.29	0.63

Table XIII. GSIB Surcharges and Market Outcomes (Two-Stage Least Squares)

This table contains two-stage least-squares regressions of market outcomes on the number of banks in the county. The excluded instrument is Number of GSIBs $(2015)_c \times \text{Post}_c$. The first stage is shown in Column (1) of Table XII. The sample period is 2015Q1 - 2019Q4 in all specifications. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by county. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Loan Volume (log)	Interest Rate (%)	Probability of Default (%)	Interest Rate (%)
	(1)	(2)	(3)	(4)
(5)				
Number of Banks (Annual)	0.525**	0.179*	0.396*	0.119
	(2.345)	(1.851)	(1.664)	(1.373)
Probability of Default (%)				0.082***
				(3.911)
Loss Given Default (%)				0.004***
				(4.442)
Expected Loss (%)				0.091
				(1.583)
Loan Controls	YES	YES	YES	YES
Bank-County FE	YES	YES	YES	YES
Bank-Quarter FE	YES	YES	YES	YES
Industry-Quarter FE	YES	YES	YES	YES
Observations	12,697	12,697	12,697	12,697

Appendix A. Variable Definitions

Deposit HHI: The average annual Deposit HHI (sum of squared bank market shares) from each county, from Drechsler, Savov, and Schnabl (2017).

Expected Loss: $PD \times LGD$, from Y-14Q.

Financial Industry Wages: Average, quarterly county-level wages in the financial industry in logs, from BLS.

Firm Size: log(assets) trimmed at the 99th percentile, from Y-14Q.

Floating Interest Rate: Dummy variable that equals one if the loan is floating rate, from Y-14Q.

GSIB: Dummy variable that equals one if the loan is from a global systemically important bank (GSIB), from Y-14Q.

Guarantee: Dummy variable that equals one if the loan is fully or partially guaranteed, from Y-14Q.

High Tangibility: Dummy variable that equals one if the firm's tangibility, as defined as tangible assets over total assets, is higher than the median within quarter, from Y-14Q.

Interest Rate: Loan interest rate in percentage points, trimmed at [0,1), from Y-14Q.

Loan HHI: The average annual county HHI (sum of squared bank market shares) for the county over the entire sample period, from Y-14Q.

Loan Purpose: Categorical variable for loan purpose. Popular loan purposes are general corporate purposes, working capital, and capital expenditures. For a complete list of possible loan purposes, see page 178 of the FR Y-14Q Instructions, from Y-14Q.

Loan Volume: The average quarterly loan volume in the county, from Y-14Q.

Leverage: total debt/assets, winsorized at [1%, 99%], from Y-14Q.

LGD: The bank's estimated loss given default, from Y-14Q.

Line of Credit: Dummy variable that equals one if the loan is a line of credit, from Y-14Q.

Maturity: Log of loan maturity in months, from Y-14Q.

MSA Loan HHI: The average annual MSA HHI (sum of squared bank market shares) for each MSA over the entire sample period, from Y-14Q.

Net Sales: Trailing twelve-month gross sales minus trade discounts, and returned sales and allowances for which credit is given to customers, less returns and allowances, freight out, and cash discounts, from Y-14Q.

Non-Performance: Dummy variable that equals one if the bank reports the loan as 90 days past due or non-accrual, or reports a positive net cumulative charge-off amount, or reports specific reserve for an impaired loan for the loan within the 12 months following the origination of the loan, or if the bank considers the borrower as defaulted as defined by Realized Default below, from Y-14Q.

Number of Banks: Number of unique banks to have given a loan in the county at any point over the entire sample, from Y-14Q.

Number of Banks (Annual): Number of unique banks to have given a loan in the county in the calendar year, from Y-14Q.

Number of Banks (Branch): Number of unique banks that have a commercial lending branch in a county in the calendar year, from the FDIC Summary of Deposits.

Number of GSIBs (Annual): Number of unique GSIBs to have given a loan in the county in the calendar year, from Y-14Q.

Number of GSIBs (2015): Number of unique GSIBs to have given a loan in the county in 2015, from Y-14Q.

Number of non-GSIBs (Annual): Number of unique non-GSIBs to have given a loan in the county in the calendar year, from Y-14Q.

Number of Prior Lenders: Number of additional past lenders beyond the borrower's current lender. If the firm borrows from n multiple lenders we include $n - 1$ in this count, from Y-14Q.

Probability of Default (PD): The bank's expected annual default rate over the life of the loan, trimmed if $PD = 0$ or above the 99th percentile, from Y-14Q.

Population: Annual county-level estimate from the US Census Bureau, calculated by adjusting the previous year's population (or most recent decennial census) to account for births, deaths, and net migration, from Census

Population Density: Annual county-level population per square mile in logs, from Census.

Profitability: EBITDA/assets, winsorized at [1%, 99%], from Y-14Q.

Rent: Average county level residential rent in logs, from Zillow.

Realized Default: Dummy variable that equals one if the borrower is rated D (defaulted) or is assigned a PD=100% by the lending bank within one year after the origination of the loan, from Y-14Q.

Secured: Dummy variable that equals one if the loan is secured, from Y-14Q.

Secured by Blanket Lien: Dummy variable that equals one if the lending bank has a claim on all unencumbered collateral of the borrower, from Y-14Q.

Stay Bank: Dummy variable that equals one if the firm has previously borrowed from one of the banks it is currently borrowing from, from Y-14Q.

Tangibility: tangible assets/assets, winsorized at [1%, 99%], from Y-14Q.

Wages: Average, quarterly, county-level wages in logs, from BLS.