Bank Loan Markups and Adverse Selection*

Mehdi Beyhaghi† Cesare Fracassi‡ Gregory Weitzner§

November 2021

Abstract

We analyze the relationship between loan pricing and market concentration in the US corporate loan market by creating a measure of markup using banks' internal loan risk assessments. Our risk-adjusted measure of markup is orthogonal to the subsequent performance of loans, while a measure that excludes banks' private risk assessments strongly predicts performance. Consistent with theories in which asymmetric information across banks creates adverse selection which drives markups, we find that markups are higher in less concentrated regions. We provide further support for the adverse selection channel by showing that markups are higher among firms that are more subject to asymmetric information and when firms stay with their existing banks. Finally, higher local markups are associated with lower loan volume and higher levels of collateralization. Our findings suggest that adverse selection drives markups, loan volume and lending standards in local bank markets and have implications for antitrust policy.

*We thank Andres Almazan, Aydoğan Altı, Paul Beaumont, Adolfo De Motta, Jeff Gerlach, Cooper Howes, David Martinez-Miera, Thomas Rivera, Farzard Saidi, David Schumacher, Sheridan Titman and seminar participants at McGill, the SFS Cavalcade, UT Austin for the helpful comments and discussions. The views expressed in this this article are solely those of the authors. They do not necessarily reflect the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

†Federal Reserve Bank of Richmond. Email: mehdi.beyhaghi@rich.frb.org.

‡University of Texas at Austin. Email: cesare.fracassi@mccombs.utexas.edu.

§McGill University. Email: gregory.weitzner@mcgill.ca.

1 Introduction

Market concentration has risen significantly across most industries in the United States over the last decades (Grullon, Larkin, and Michaely (2019)). In the banking sector, the top 10 largest bank holding companies have doubled their market share of total assets since 1990 (Fernholz and Koch (2020)). Traditional models of competition typically predict that higher levels of market concentration lead to higher prices and reduced supply in product markets. In fact, the antitrust division of the U.S. Department of Justice commonly uses market concentration as an important criteria to approve or block mergers, arguing that higher market concentration reduces competition, thereby harming consumers with excessively high prices.

However, unlike typical product markets, credit markets are plagued by two levels of asymmetric information: borrowers are often better informed than lenders about their own creditworthiness, and some lenders know more about certain borrowers' credit quality than other lenders. In fact, several models predict that higher concentration leads to lower prices when certain banks know more about borrower quality than others (e.g., Broecker (1990), Riordan (1995), Shaffer (1998), Dell'Ariccia (2001) Marquez (2002) and Dell'Ariccia and Marquez (2006)). Intuitively, when markets are less concentrated, banks have a more difficult time distinguishing which borrowers have been previously rejected by other banks which exacerbates the adverse selection problem and leads to higher markups. Because these two sets of theories have opposite predictions, the relationship between local market structure and loan prices is ultimately an empirical question. In this paper, we propose a novel measure of loan markup that incorporates banks' private information, and find a negative relationship between market concentration and markups, providing support for the adverse selection channel of market power. These results have important implications not only for antitrust policy, but also the evolution of lending standards and loan volume over time.

Despite the theoretical importance of market power in credit markets, measuring loan markups accurately is challenging because the marginal cost of loans is unobservable to the econometrician. The most common approach in the literature is to assume that after conditioning on observable firm and loan characteristics, any difference in interest rates across loans is due to market power. However if a loan's interest rate is affected by asymmetric information, this information is inherently unobservable and will not be reflected in observable characteristics. In other words, we cannot attribute the unobserved component of interest rate to markups unless we control for the bank's private information about the borrower's risk.

We address this problem by using Federal Reserve's Y-14Q Schedule H.1 data that includes all corporate loans over $1mm extended by large bank holding companies (BHCs) in the United States. A key advantage of the data is that BHCs are required to report their internal measures of probability of default (PD) and loss given default (LGD) for each loan on their balance sheets. We first show that these risk assessment measures (PD and LGD) are strong predictors of future loan delinquency and default, even after controlling for other determinants of firm performance. We then estimate markup as the unexplained residual of a regression of loan interest rate on banks' private risk assessment measures and loan-level controls. We also include industry by quarter, loan type and bank by quarter fixed effects as controls in our estimate to account for differences across banks in their internal risk assessments at each point in time. Our analysis relies on the identifying assumption that after controlling for loan characteristics and banks' loan risk assessments, the variation in interest rates on loans given by the same bank in the same quarter is not driven by the ex-ante riskiness of that loan. We provide support for this assumption by showing that the risk-adjusted estimate of markup does not predict ex-post non-performance or default. Highlighting the importance of controlling for banks' private information, when we exclude these risk measures from the model, the baseline markup predicts future loan performance, making inferences of the determinants of markup confounded by risk factors.

After having established the validity of our risk-adjusted markup, we test whether local market structure and firm characteristics drive markups. In the absence of asymmetric information or other frictions, higher bank concentration can lead to higher markups because banks can more easily collude or better internalize their price impact as in a standard Cournot setting. However, higher concentration can also reduce adverse selection in the presence of asymmetric information about borrower quality between banks (e.g., Broecker (1990), Riordan (1995), Shaffer (1998), Marquez (2002) and Dell'Ariccia and Marquez (2006)).[1] In order to compare markups across regions, we calculate loan-level Herfindahl-Hirschman Indices (HHIs) at both the county and MSA level. Consistent with the adverse selection channel, we find that markups are higher in regions with lower levels of concentration. Increasing the local HHI from 0 to 1 lowers the markup on a loan by 12bp, or 15% of a standard deviation, which compares to an all-in average interest rate of 3.65pp. Also consistent with adverse selection driving markups, we find that markups are larger for firms likely facing a high degree of asymmetric information (smaller, low profitability, highly-levered and low tangibility firms).

To further support the adverse selection channel, we develop additional tests motivated by the theoretical predictions in Dell'Ariccia and Marquez (2006). In their model, adverse selection arises because banks cannot distinguish firms that have been previously rejected by other banks among the pool of borrowers that approach them for loans. This allows banks that know the quality of firms to charge an information rent, or markup, to high quality borrowers. This effect is exacerbated in less concentrated markets because banks have a harder time determining whether borrowers have been already rejected by other banks. In practice, banks that have existing relationships with firms are likely to have better information than other banks (e.g., Greenbaum, Kanatas, and Venezia (1989), Sharpe (1990) and Rajan (1992)). Hence, we test whether banks charge their existing clients higher markups than firms that switch banks. Consistent with this channel, we find that within region and quarter-of-origination, firms that stay with their banks receive higher markups (approximately 13bps for both MSA and county-level regressions). Furthermore, we find some evidence that this effect is stronger in markets with lower levels of market concentration. Finally, we find that markups are higher for firms borrowing

[1]For a excellent summary of these issues see Carbo-Valverde, Rodriguez-Fernandez, and Udell (2009). Concentration can also affect the relative markups between younger and older firms Petersen and Rajan (1995).

from banks with higher market shares in the local region. These additional tests provide further evidence in support of the adverse selection channel driving markups.

We next test two main aggregate predictions from Dell'Ariccia and Marquez (2006). A main feature of the model is that lending booms arise when new borrowers of unknown quality apply for loans, thereby mitigating the adverse selection problem among banks. Lending booms result in both lower aggregate markups, higher loan volume and reduced screening in the form of lower levels of collateralization. We thus aggregate the loan-level data at both the MSA and county level, and exploit time-series variation in market conditions within geographic region. We find that when MSA and counties exhibit higher aggregate markups, aggregate loan volumes are higher and aggregate collateralization is lower. These results are consistent with banks applying looser lending standards when the asymmetric information across banks is less severe. A caveat to our aggregate analysis is that other unobserved factors could explain the relationship between markups and lending standards. However, the correlations that we find in the data are consistent with Dell'Ariccia and Marquez (2006) and other theories in which asymmetric information about borrower quality across banks drives markups.

A possible concern with our findings is that our measure of markups might be influenced by unobserved costs of originating, processing, administering, and monitoring loans, which vary within banks, and across loans and regions. For instance, these costs may be systematically lower in more concentrated regions, thus explaining why markups seem to be lower in these regions. However, these costs are likely to be higher in rural areas (Bottomley (1963)), where loan-level HHIs are higher. Hence, if these costs vary significant across regions, this should attenuate our results. Furthermore, even after controlling for geographic region and quarter of origination, we find that markups vary across loans given by the same bank depending on whether the firm switched banks or not.[2]

Another potential concern with our analysis is that banks may not report their risk measures truthfully (Begley, Purnanandam, and Zheng (2017), Plosser and Santos (2018),

[2]An additional concern could be that advertising is more expensive in less concentrated areas. In the context of mortgages, a more homogenous market than corporate loans, Gurun, Matvos, and Seru (2016) find that heavily advertised mortgages are more expensive. However, to our knowledge banks do not advertise corporate loans as they do consumer loans.

Behn, Haselmann, and Vig (2016)). For example in the context of the German loan market, Behn, Haselmann, and Vig (2016) find that banks systematically underreport default probabilities on loans that impact their capital requirements. While we cannot rule out that banks underreport their average risk measures, aggregate misreporting should not bias our results because our analysis includes bank by quarter fixed effects. Our results could also be biased if banks differentially misreport across loans; however, if this were the case we would expect our estimate of markup to predict the future performance of loans, which it does not.

Finally, one limitation of our analysis is that our measure of markup is the residual of a regression, which by construction has mean of zero. Thus our measure of markup is only relative to other loans in the sample. Hence, we cannot identify the absolute level of bank markups, only how markups vary across loans.

We contribute to the empirical literature testing on the relationship between bank market structure and asymmetric information. Zarutskie et al. (2003) find that firms are more likely to get loans in more concentrated regions. Cetorelli and Gambera (2001) and di Patti and Dell'Ariccia (2004) find that higher concentration is associated with higher growth in sectors that are highly dependent on external finance. Consistent with the predictions of Dell'Ariccia and Marquez (2006), Dell'Ariccia, Igan, and Laeven (2012) find that that denial rates on mortgages are lower in areas that experience faster credit demand growth and that lenders attached less weight to applicants' loan-to-income ratios. Crawford, Pavanini, and Schivardi (2018) use a structural model to analyze the effect of asymmetric information between individual banks and borrowers and find evidence that market power moderates the effect of adverse selection. A major hurdle to identifying the effects of asymmetric information in the banking literature is that banks' private information is usually unobservable. We arguably overcome this hurdle by directly using banks' internal risk assessments.

Our paper also contributes to the empirical literature testing information hold up problems between borrowers and banks (e.g., Santos and Winton (2008), Hale and Santos (2009), Schenone (2010) and Ioannidou and Ongena (2010)). These papers generally argue that around event studies (e.g., IPOs and new bond offerings) interest rate changes are due to changes in market power and not changes in risk. In contrast, we directly measure the magnitude of the information rents banks extract from borrowers by controlling for the risk of the loan.

A large existing literature analyzes the effect of market concentration on loan prices (e.g., Hannan (1991), Petersen and Rajan (1995), Sapienza (2002), Cavalluzzo, Cavalluzzo, and Wolken (2002) Rice and Strahan (2010)). Most papers find either a positive or no relationship between loan interest rates and market concentration.[3] There are a few key differences between our analysis than the aforementioned papers. First, the existing papers in the literature focus on small-business loans, while our sample only includes corporate loans of at least $1 million in size. Second, the vast majority of papers use deposit HHI rather than loan HHI.[4] A problem with this approach is deposit HHIs may not line up with loan HHIs. Indeed we do not find a statistically or economically significant relationship between deposit HHIs and loan markups. Third, our data includes extensive loan and firm level characteristics. Fourth, and most importantly, to our knowledge this is the first paper to systematically measure markups using banks' internal risk measures, thereby controlling for banks' private information regarding the loan. We argue that this information is critical to properly measure the effect of bank concentration on markups.[5] For instance, there could be more screening in a particular regions which leads to systematically lower interest rates conditional on observables (e.g., Dell'Ariccia and Marquez (2006)). In fact, this mechanism is consistent with our analysis as we find a positive (negative) relationship between collateralization (loan volume) and average markups at

[3] Degryse and Ongena (2008) survey the literature. A related literature analyzing deposits, generally finds a positive relationship between bank concentration and deposit rates (e.g., Hannan (1991), Neumark and Sharpe (1992), Prager and Hannan (1998) Driscoll and Judson (2013), Drechsler, Savov, and Schnabl (2017), Wang et al. (2018)).

[4] An exception is Sapienza (2002) in the context of Italian small-business loans.

[5] 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), Carbo-Valverde, Rodriguez-Fernandez, and Udell (2009)).

the local level.[6].

A growing literature analyzes the macroeconomic implications of the rise of concentration in the United States (Gutiérrez and Philippon (2017), Barkai (2019), De Loecker, Eeckhout, and Unger (2020) and Liu, Mian, and Sufi (2019)). While the aforementioned papers highlight the negative effects of higher concentration, we find a benefit in banking markets in the form of reduced adverse selection. Furthermore, our paper relates to the macroeconomic literature analyzing the cyclicality of markups (e.g., Rotemberg and Saloner (1986), Rotemberg and Woodford (1991), Chevalier and Scharfstein (1995), Rotemberg and Woodford (1992), Olivero (2010), Aliaga-Díaz and Olivero (2011), Nekarda and Ramey (2013) and Anderson, Rebelo, and Wong (2018)). The negative relationship we find between loan volume and markups is consistent with countercyclical markups in bank loan markets.

Our paper also relates to the literature analyzing bank internal risk-measures (e.g., Agarwal and Hauswald (2010), Qian, Strahan, and Yang (2015), Behn, Haselmann, and Vig (2016), Dell'Ariccia, Laeven, and Suarez (2017), Plosser and Santos (2018), Becker, Bos, and Roszbach (2018)), Adelino, Ivanov, and Smolyansky (2019)). Adelino, Ivanov, and Smolyansky (2019) also use Y-14Q data and show that interest rates have minimal predictive power on loan performance after controlling for PDs, which is consistent with our analysis. However, they do not explore the variation in interest rates that is unexplained by risk, which is the main focus of our paper.

2 Theoretical Background

In homogeneous product markets, standard models predict that higher concentration leads to higher prices. For example, in static Cournot models, where firms compete via quantities, firms better internalize the impact that their production has on prices when

[6]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)).

markets are more concentrated, leading to higher markups and firm profits.[7] In dynamic settings, higher concentration facilitates collusion, which can also lead to higher markups (e.g., Stigler (1964), Friedman (1971) and Abreu (1986)).

However, in credit markets plagued by adverse selection, the relationship between market concentration and markups can flip. There are two main forms of asymmetric information in credit markets. First, borrowers are often better informed about their own creditworthiness than the banks that lend to them. Second, some banks might know more about certain borrowers than other banks, either because they are better at screening, or because they have access to private information through ongoing relationships with their existing clients. The latter form of asymmetric information can lead to the classic hold-up problem, where banks charge prices higher than marginal cost to high quality borrowers because those borrowers cannot find more competitive prices because other lenders pool them with low-quality applicants.[8] Moroever, higher concentration can limit the adverse selection and hold-up problems. Intuitively, as the number of banks decreases, the winner's curse problem becomes less severe (e.g., Broecker (1990), Riordan (1995) and Shaffer (1998)) or information becomes less dispersed (e.g., Marquez (2002) and Dell'Ariccia and Marquez (2006)), as banks have an easier time determining whether borrowers have been been rejected by other banks.[9] A natural consequence of adverse selection, as stated by Marquez (2002), is:

> "..focusing exclusively on the number of banks may not provide a very good indicator of the competitiveness of a market... markets composed of many small banks may actually have higher expected interest rates in equilibrium than markets composed of a few large banks."

Furthermore, this form of adverse selection anecdotally appears to an issue in practice for banks. 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

[7] In contrast, if firms compete via prices à la Bertrand in a static setting, prices are competitive as soon as there are multiple firms.

[8] Greenbaum, Kanatas, and Venezia (1989), Sharpe (1990) and Rajan (1992) provide microfoundations for this phenomenon in dynamic settings in which banks learn the quality of their borrowers over time.

[9] Bolton, Santos, and Scheinkman (2016) and Fishman and Parker (2015) also show how informed investors profits can increase as more informed investors enter.

> from you" and that the only solution to this problem was to hire "superior" loan officers. Bankers and bank examiners alike are very familiar with this phenomenon."

In summary, two opposing forces can affect the relationship between market concentration and markups in credit markets: on the one hand, higher market concentration can increase banks' market power and loan markups, while on the other hand, higher market concentration can reduce banks' market power and loan markups by alleviating adverse selection concerns.

We also provide additional tests for the adverse selection channel, based on the predictions in Dell'Ariccia and Marquez (2006). In their model, banks have knowledge about an existing pool of borrowers. When a borrower applies for a loan at a new bank, the bank does not know whether other banks have already rejected the applicant or not, creating adverse selection. If higher bank concentration leads to lower markups, we would expect that borrowers that stay with their existing banks face higher markups than those that switch to different banks as existing banks are likely to know more about their borrowers than other banks. Moroever, we would expect this effect to be more pronounced in less concentrated markets where the adverse selection problem is more severe. Finally, we also expect that markups will be higher for banks with higher market shares as the adverse selection problem is more severe for those firms to switch to banks with lower market shares.

Dell'Ariccia and Marquez (2006) also show that as new borrowers enter the market, the asymmetric information problem becomes mitigated, as the likelihood that another bank already knows the quality of that applicant decreases. This reduces the need for banks to screen applicants, which in turn leads to an increase in loan volume and a decrease in collateralization.[10] Hence, we would expect that aggregate loan volume is negatively correlated with aggregate markups and average collateralization levels are positively correlated with aggregate markups at the county/MSA level.

[10] In the model (as well as several others in the theory literature), posting collateral is a form of screening high type borrowers.

3 Empirical Design

3.1 Estimating Markup

The interest rate (IR) charged by a bank to a firm can be decomposed into three parts: (i) the marginal cost due to the credit risk that the banks faces in case the firm defaults (MC_{risk}); (ii) the marginal cost of originating, administering, and monitoring the loan ($MC_{non-risk}$), and (iii) the markup (MU), which by construction is the difference between the price and the marginal cost.

$$IR = MC_{risk} + MC_{non-risk} + MU, \tag{1}$$

In the banking literature, markups are commonly estimated by absorbing both risk and non-risk marginal costs using a host of controls and fixed effects based on bank and loan characteristics. However, the premise of the adverse selection channel, as described in Section 2, is that banks have private information that is not observed by others, including the econometrician. Any measure of markup that does not directly control for the ex-ante risk assessment of banks could be contaminated by unobserved risk factors, biasing the estimation of the determinants of loan markups. In fact, we show below that markup predicts ex-post loan performance and realized default if it is estimated without explicitly controlling for banks' ex-ante risk assessments. One of the main contributions of this paper is to use a proprietary dataset of corporate loans that include banks' ex-ante assessments of the probability of default and the recovery rate of each loan.

We estimate bank loan markup using a two-stage procedure. First we estimate the following linear regression:

$$IR_l = \beta_0 PD_l + \beta_1 LGD_l + \beta_2 \left(PD_l \times LGD_l\right) + \gamma X_l + \delta_{b,t} + \alpha_{i,t} + u_l, \tag{2}$$

where the unit of observation is loan l in industry i originated by bank b in quarter t. The outcome variable IR_l is the loan's interest rate and X_l are loan-level controls, $\delta_{b,t}$

is bank by quarter fixed effects, and $\alpha_{i,t}$ is industry by quarter fixed effects. Our first-stage includes all variables that we believe may influence a loan's marginal cost. Most importantly, we include the bank's estimate of the loan's probability of default PD_l and loss given default LGD_l.

Next, we estimate markup by decomposing the interest rate into two components using the coefficient estimates from Equation (2): the predicted interest rate and the residual. The predicted interest rate represents the marginal cost (both the risk and non-risk component), while the residual is our estimate of loan markup. If the interest rate decomposition is valid, the predicted interest rate component should reflect the marginal cost of the loan, and thus predict future loan performance while the markup should not as it is orthogonal to the risk of the loan. As shown below, we will directly perform this test.

It is important to note that throughout our analysis we will use bank by quarter fixed effects to control for any bank specific factors that may affect the marginal cost of a loan, e.g., difference in cost of capital, regulatory costs, monitoring skills, etc. Moreoever, if banks use difference risk models, this approach should absorb such heterogeneity.

We do not include firm characteristics when we estimate markup for two reasons: first, as long as we control for the riskiness and characteristic of the loan, firm characteristics should not affect the risk-based component of the interest rate. Second, we do not want to control for variables, such as the size of the firm, that may be related to asymmetric information and thereby drive markups. In the Appendix, we show that markup predicts loan performance when we include firm characteristics but not banks' risk assessments in our estimate of markup.

One concern with our approach is that bank and loan characteristics could proxy not only for the marginal cost of a loan, but also for the degree of asymmetric information of the loan. For example, the adverse selection problem may be less severe for a revolving credit line than a term loan because the quality of the borrower may be less important if the loan remains undrawn for a period. By absorbing these characteristics using control variables and fixed effects, our approach will attribute all variation in the interest rate driven by that control variable to marginal cost. However, this potential mis-attribution would only attenuate our results by making our estimate of markup noisier.

A second concern is that the independent variables we include to estimate markup do not absorb all the differences in non-risk components of marginal costs, such as originating, administering, and monitoring costs. For example, monitoring costs could be higher in rural areas, where the distance between banks and borrowers is usually larger. However, as we show later we find that regions with *higher* concentration exhibit *lower* markups. Hence, if present, this effect should only attenuate our results.

An alternative approach to estimating markups is to use a structural model (e.g., De Loecker and Warzynski (2012)). However, this approach requires estimating a production function, which may not be well-suited for a bank loans where the main input is the information regarding the borrower. Moreover, as far as we are aware, it cannot deal with differences in marginal cost across loans due to banks' private information. Hence we view our approach as flexible enough to capture what we view as the main driver of markups in bank loans.

3.2 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).[11] The sample includes corporate loans from all bank holding companies (BHCs) with $50bn or more in total assets, accounting for 85.9% of all assets in the U.S. banking sector as of 2018:Q4 (Frame, McLemore, and Mihov (2020)). Qualified BHCs are required to report detailed quarterly loan level data on all corporate loans that exceed $1mm in size. These loans represent 70% of all commercial and industrial loan volume in the U.S. (Bidder, Krainer, and Shapiro (2020)).

The data include detailed loan characteristics (such as interest rate, maturity, amount,

[11] Other papers that use Y-14Q data include: Bidder, Krainer, and Shapiro (2020), Brown, Gustafson, and Ivanov (2017), Balasubramanyan, Berger, and Koepke (2019), Ivanov, Pettit, and Whited (2020), Abdymomunov, Curti, and Mihov (2020), Beyhaghi (2020) and Greenwald, Krainer, and Paul (2020).

collateral, credit guarantee, purpose), quarterly loan performance (past due payments, non-accruals, charge-offs), the ZIP code of the borrowers' headquarters as well as firm financials (balance sheet and income statement). Importantly for our analysis, banks are also required to report their internal estimates of probability of default (PD) 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)."[12]

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-exempted, 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 estimate markups. 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 reporting errors, we drop loans with interest rates equal to or below 0% or above 100% and loans with PDs missing, zero, or 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 it is populated. Loans that are not utilized within two quarters after initiation are dropped from the sample as no interest rate is reported for these loans. As a loan might remain on the bank's

[12]Most recent instructions available at Calculation of RWA for credit risk.

balance sheet for multiple quarters, we only keep the first appearance of a loan in the data (i.e., new loans). Finally, as some firms have an abnormally large number of loans in the sample, we remove loans in which the borrowing firm has more than the 99th percentile in total new loans over the sample. After these filters, we are left with 28,000 new loans originated from 2014Q4 to 2020Q3 by 23 BHCs.[13]

We define the following firm-level financial variables: profitability (EBITDA/assets), firm size (log assets), tangibility (tangible assets/assets), and leverage (debt/assets), winsorized at the 1% and 99% level. Furthermore, we use two measures of loan performance: (i) 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 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 (ii) default, which is a dummy variable equal to 1 if the borrower defaults within one year since origination, defined as a borrower rated D (defaulted) or is assigned a PD=100% by the lending bank within one year after the origination of the loan. We use a window of one year because banks' PD estimates are required to reflect one year default rates.

Finally, the data includes ZIP codes corresponding to each borrower's headquarters. In order to create concentration measures at the county and MSA levels, we obtain the ZIP code to county crosswalks from the Housing and Urban Development (HUD) and a county to MSA crosswalk from Center for Medicare and Medicaid Services (CMS). After merging the county and MSA data into the Y-14Q dataset, we construct annual loan-level Herfindahl-Hirschman indices (HHI) for each county or MSA. We then proceed to average the HHI for each region over the entire sample period.[14] We also use population density data from the Census as an additional proxy for market concentration.

In Section A of the Appendix we include detailed definitions of all of our variables.

[13]In the early part of the sample, PDs were not reported consistently; hence, our sample begins in 2014Q4.

[14]This is a standard procedure to calculate HHI (e.g., Drechsler, Savov, and Schnabl (2017))

3.3 Descriptive Statistics

Table 1 includes summary statistics for the variables used in the paper. The average and median loan size is approximately \$7mm and \$2.6mm, respectively and over 90% of loans are less than \$16mm. The fact that the majority of the loans and firms are relatively small is also important for testing the effect of geographic concentration on markups, as larger firms often source their loans nationally.

The loan sample is approximately evenly split among credit lines and term loans and the median interest rate is 3.66%. The median firm has \$20mm in assets, 8% profitability, and 30% book leverage. The loans in our sample thus make up a large portion of firms' capital.

Over our sample period, 0.81% of firms default within the first year after loan origination. This compares to an average ex-ante expected PD of 1.38%. This discrepancy is likely due to the fact that the aggregate economic conditions in the U.S. over the sample period were positive relative to banks' expectations.

3.4 Validity of Bank Risk Assessments and Estimation of Markup

In order to properly estimate markup, banks' reported risk measures must reflect the actual risk of the loans. Therefore, we verify that the ex-ante banks' risk metrics predict ex-post performance (delinquency and defaults) and interest rates.

First, we compare the univariate relationship that realized default has with interest rate and PD. In Figure 2a 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, the trend is not monotonic: average default rates increase from bins 2-5 but not from bin 1-2. On the other hand, when we place loans into five PD buckets we see a much clearer positive and monotonic relationship between average realized defaults and PD than interest rates (Figure 2b). The preliminary evidence shows that bank risk measures are more strongly correlated to performance than interest rates, suggesting that interest rates may include substantial non-risk components to interest rates, namely non-risk marginal costs and markups.

Second, we formally test whether the bank risk metrics explain loan performance and interest rates after we control for loan characteristics and add a host of fixed effects. We thus formally estimate the following multivariate regression

$$y_l = \beta_0 PD_l + \beta_1 LGD_l + \beta_2 \left(PD_l \times LGD_l\right) + \gamma X_l + \delta_{b,t} + \alpha_{i,t} + u_l, \tag{3}$$

where the unit of observation is each loan l of firm f in industry i originated by bank b in quarter t. The outcome variable y_l is either Non-Performance, Realized Default or Interest Rate and X_l are loan-level controls, which include: log(Maturity), log(Amount), Guarantee and loan type fixed effects, $\delta_{b,t}$ 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. Furthermore, by always evaluating two loans given by the same bank in the same quarter, we absorb any differences in banks' cost of capital or financial constraints that may affect interest rates. We add an interaction term between PD and LGD to explicitly take into account the expected loss of each loan. To adjust standard errors for correlations in the residuals, we cluster the standard errors by firm.

The results are displayed in Table 2. In Columns (1) and (3), we estimate the predictive power of loan-level controls and fixed effects in explaining non-performance and realized default. These baseline regressions do not include banks' risk assessment measures. The adjusted R-squared are 8% and 5%, respectively. In Columns (2) and (4) we add PD, LGD, and PDLGD. Consistent with Figure 2b, banks' PDs strongly predict future non-performance and realized default, even after controlling for a host of loan characteristics and fixed effects. For example, a 1pp increase in PD implies an 0.63pp increase in realized default rates. The adjusted R-squared of the regressions also increase significantly to 10% and 7%. Overall, we conclude that the ex-ante bank risk assessments of loan credit risk predict ex-post loan performance, and that the information included in these measures is not fully absorbed by loan and bank characteristics. It is thus crucial to control for such variables when estimating markups.

We then turn to predicting interest rates and estimating markups in Table 3. In Column (1), we establish a baseline model where we predict interest rates using only loan-level characteristics and fixed effects. The adjusted R-squared is 49%. In Column (2), we include PD, LGD, and their interaction term (Expected Loss) to the regression. Consistent with the results in Table 2, bank risk measures also strongly predict interest rates. Even after controlling for loan, bank, and industry characteristics, loans that have greater probability of default, loss given default, and expected loss are charged higher interst rate. The adjusted R-squared increases from 49% to 52%, confirming that these bank assessments can explain a large portion of the heterogeneity in interest rates. The effect of the risk assessment on interest rates is not only statistically significant, but also economically relevant: a 10pp increase in PD leads to a 0.75pp increase in interest rate.

Third, we estimate loan markups by decomposing the interest rate into two components using the coefficient estimates from Equation (3): the predicted interest rate and the residual, which will be our proxy for loan markup. We define two measures of markup: a baseline markup, which uses the residual from the estimates in Column (1) that does not include banks' risk assessment measures, and a risk-adjusted markup, which uses the residual from the estimates in Column (2), that takes into account the banks private risk assessment. It is important to note that our measure of markup is relative, and not absolute. By construction, the markup is a residual with a mean of zero. However, as shown in Table 1, the standard deviation of markup is 0.8pp, which is very similar to the standard deviation of the predicted interest rate. In other words, half of the difference in interest rates charged by banks to firms is driven by observable factors, and half by unobservable factors.

Finally, we test the validity of the baseline and risk-adjusted markups. For the residual to be a plausible measure of markup, it should be unrelated to the future performance of the loan. We directly test the relationship between markup and loan non-performance and default in the following regression:

$$y_l = \beta_0 \widehat{IR}_l + \beta_1 \widehat{MU}_l + \gamma X_l + \delta_{b,t} + \alpha_{i,t} + u_l, \tag{4}$$

where the outcome variable, y_l, is either Non-Performance or Default, $\widehat{IR}_l$ and $\widehat{MU}_l$ are the predicted interest rate and markup estimated from Equation (3). We estimate Equation (4) with the same set of fixed effects from Equation (3) and we cluster our standard errors by firm. The results, which we estimate excluding bank risk measures (baseline) and including bank risk measures (risk-adjusted) are displayed in Table 4 with non-performance as the dependent variable in Panel A and realized default as the dependent variable in Panel B. For reference, we also include regressions with the actual interest rate as an independent variable rather than its decomposition into predicted interest rate and markup.

Columns (1) and (4) of Panel A and B in Table 4 confirm the univariate results of Figure 2a, showing that a higher interest rate predicts loan non-performance and default with or without the use of fixed effects. In Columns (2) and (5) we decompose the interest rate into the predicted interest rate and markup (residual) for the baseline model without banks' risk assessment measures. In all specifications, the baseline markup still predicts loan performance, suggesting that controlling for loan and bank characteristics does not absorb drivers of loan default risk.[15] In other words, not adjusting for banks' private information leads to a biased measure of markup that correlated with loan performance, and any inference using such a measure of markup could be confounded by risk factors.

In Columns (3) and (6), we repeat the interest rate decomposition using the risk-adjusted model estimated using banks' private risk assessments. Unlike the baseline markup, the risk-adjusted markup does not appear to predict ex-post default or non-performance, while the predicted interest rate is positive and statistically significant. It is important to note that we cannot formally tests whether risk-adjusted markup is

[15] In the Online Appendix, we show that even if we include firm characteristics in the estimate of the baseline model, the baseline markup still strongly predicts loan performance. We conduct this exercise to give the baseline markup the best chance to capture the variation in interest rates due to the riskiness of the loan without using bank risk assessments.

completely orthogonal to loan performance, as the inability to reject the null hypothesis does not mean that the null is true. However, the point estimate on the risk-adjusted markup is both economically small (less than one order of magnitude of the baseline markup) and statistically insignificant. Overall, the drastic differences we see in the baseline and risk-adjusted markup performance highlights the importance of adjusting for banks' risk assessments to have an unbiased measure of markup that is not contaminated by credit risk.

After having established the validity of the risk-adjusted markup, in the following section we analyze the relationship between market concentration, firm characteristics and markups.

3.5 Markups, Market Concentration and Firm Characteristics

In this section we test how risk-adjusted markup relates to market concentration and firm characteristics. We thus estimate the following loan-level regression:

$$\widehat{MU}_l = \beta_0 HHI_c + \gamma Z_{f,t} + \delta_{b,t} + \alpha_{i,t} + u_l, \tag{5}$$

where risk-adjusted markup ($\widehat{MU}_l$) for loan l originated by bank b to firm f in 2-digit NAICS industry i is the dependent variable, HHI_c is the HHI of the MSA or county where the firm is headquartered, $Z_{f,t}$ is a vector of firm characteristics, and $\alpha_{i,t}$ are industry-quarter fixed effects. We also cluster the standard errors by MSA or county.

The results are displayed in Table 5. In Columns (1) - (3) we use MSAs to compute market concentration, while in Columns (4) - (6) we use counties. In Columns (1) and (4) we analyze the univariate relationship between HHI and markup. The coefficient is negative and statistically significant: regions that are more concentrated exhibit lower markups.

We next further test the adverse selection channel by adding firm characteristics to the regression in Columns (2) and (5). The coefficient on HHI does not change. Furthermore, firms that the literature suggests suffer less from asymmetric information problem, i.e., larger, more profitable, with low leverage and with higher tangible assets, receive loans with lower markups.[16] Lastly, in Columns (3) and (6) we find that results are robust even after controlling for bank by quarter, industry by quarter, and loan fixed effects.

Across the specifications, the coefficient on HHI is between $-0.13 and -0.18$, implying that increasing HHI from 0 to 1 reduces the markup by 13 - 18bps, or 16 - 23% of a standard deviation. This magnitude is larger than doubling firm size or going from no leverage to a fully levered firm.

A concern with our HHI calculation is we do not have all loans from all banks in the region because our data only covers U.S. banks with over $50bn in assets. However, as noted earlier, the Y-14Q eligible banks make up the vast majority of corporate volume. Nonetheless, in the Appendix we show that our qualitative results are unchanged if we use a dummy variable that equals one if the local HHI is above the median HHI or if we use population density rather than loan HHI.[17]

3.6 Markups and Switching Banks

After establishing that markups are negatively related to concentration and are higher among firms more likely to be subject to asymmetric information, we now attempt to provide further evidence for the adverse selection channel. In particular, we test whether firms face higher markups when they remain with their existing banks on new loans, a direct prediction of several theories of adverse selection in banking markets.

In order to capture the information effect of repeat borrowers, we restrict the sample to firms with more than one loan and analyze all loans that follow their first loan.[18] After

[16] Firm size should not be related to transaction costs from loans because we control for loan size in our estimate of markup. Higher asset tangibility can reduce asymmetric information if the payoff of the assets are easier to observe (Almeida and Campello (2007)). Higher leverage can exacerbate the asymmetric information problem by increasing the sensitivity of a security's payoff to firm quality e.g., (Heider (2003)).

[17] Intuitively, sparser populations limit the potential of adverse selection as it becomes more difficult for firms to visit many banks.

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

making these restrictions, we estimate the following regression:

$$\widehat{MU}_l = \beta_0 StayBank_l + \gamma X_{f,t} + \delta_{b,t} + \alpha_{i,t} + \lambda_{c,t} + u_l, \tag{6}$$

where $StayBank$ is a dummy that equals one when firms stay with their existing banks on their new loan, $\lambda_{c,t}$ is either county or MSA by quarter-of-origination fixed effects to control for any unobserved differences in markups across regions and time. We also cluster the standard errors by MSA or county depending on if MSA or county HHI is used in the regression. The results, which we estimate (6) with and without our main set of fixed effects and firm characteristics are displayed in Table 6. Consistent with our hypothesis, we find that the estimated coefficient of $StayBank$ is positive and statistically significant for all MSA level regressions and county level regressions without controls. For example as shown in Column (1), at the MSA-level firms face 9bp higher markups when they remain with their existing bank.

In Columns (3) and (6), we interact $StayBank$ with MSA and County HHI, respectively. If the information hold up problem is more severe when the region is less concentrated, the coefficient on the interaction term should be negative. Consistent, with this hypothesis, both coefficients are negative; however, only the MSA specification (Column (3)) is statistically significant. Nonetheless, the sign of the coefficient is consistent with reduced concentration exacerbating the information hold-up problem at the local level.

Next, we test whether a bank's market share affects the markups its borrowers receive in the local market. If a firm tries to leave its existing bank, if that bank has a higher market share, the other banks in the region will have smaller market shares and the adverse selection problem will be more severe because those banks know the quality of fewer borrowers in the region.[19] To test this prediction, we estimate regressions with bank market share as the independent variable predicting risk-adjusted markup. Throughout our tests we include include MSA or county by quarter-of-origination fixed effects. The results are displayed in Table 7. Consistent with our prediction, the larger a bank's

[19] Although there is no heterogeneity among banks in their model, this is a direct implication of Dell'Ariccia and Marquez (2006).

market share, the higher the markups it charges on its loans.

Together these results suggest that variation in local markups arise from asymmetric information problems across banks. Hence, these results also provide additional support for the adverse selection channel of markups.

3.7 The Relationship Between Markups, Loan Volume and Lending Standards

In this section we analyze the relationship between markups, loan volume and lending standards. If asymmetric information across banks is driving markups, Dell'Ariccia and Marquez (2006) predicts that aggregate markups should be negatively correlated with loan volume and positively correlated with collateralization levels. To test these predictions we create the following variables at the county and MSA level: average markup, total loan volume and average number of loans that are collateralized. We then estimate the following time-series regression:

$$y_{t,c} = \beta_0 AverageMarkup_{t,c} + \lambda_c + u_{t,c}, \tag{7}$$

where $y_{t,c}$ is either the log of total loan volume or the average number of loans collateralized in the region/quarter of origination, $AverageMarkup_{t,c}$ is the average markup in the region/quarter of origination and λ_c is MSA or county fixed effects. We also cluster the standard errors by MSA or county depending on the unit of observation. The results are displayed in Table 8. Consistent with our hypothesis we find that collateralization is lower and aggregate loan volume is higher when markups are higher within regions. For example as shown in Columns (1) and (3) a 1pp increase in markup is associated with a 5.8pp decrease in loan volume and a 1.3pp increase in collateralization at the MSA level. These results provide further support that asymmetric information across banks drives markups, lending volume and lending standards in bank loan markets.

A caveat to this analysis is that other unobserved factors could explain the relationship between markups, collateralization and loan volume. However, the correlations we find are consistent with the predictions of Dell'Ariccia and Marquez (2006) and thus provide further support for the adverse selection channel.

4 Conclusion

In this paper we provide evidence that asymmetric information across banks drives markups in the corporate bank loan market. Critically, we estimate markups using banks' private information about borrower quality in the form of the risk metrics they report to Federal Reserve. While the existing literature finds evidence that more concentrated banking markets raise deposit rates, we find the opposite in loan markets. Hence, a potential unintended consequence of antitrust policies is that by making banking markets less concentrated these policies may also raise interest rates on local bank loans.

We also provide further support for the adverse selection mechanism by showing that i) markups are higher for firms that remain with their existing banks and ii) higher aggregate markups are associated with lower loan volume and higher degrees of collateralization at the local level. Overall, our results provide support for theories in which asymmetric information across lenders drives the evolution of markups, lending volume and lending standards over time. Although our empirical setting is restricted to corporate bank loans, we argue that these types of effects are likely to be present in other markets in which lenders can approach multiple lenders for loans.

References

Abdymomunov, Azamat, Filippo Curti, and Atanas Mihov, 2020, Us banking sector operational losses and the macroeconomic environment, *Journal of Money, Credit and Banking* 52, 115–144.

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

Adelino, Manuel, Ivan Ivanov, and Michael Smolyansky, 2019, Humans vs machines: Soft and hard information in corporate loan pricing, *Available at SSRN 3596010* .

Agarwal, Sumit and Robert Hauswald, 2010, Distance and private information in lending, *The Review of Financial Studies* 23, 2757–2788.

Aliaga-Díaz, Roger and María Pía Olivero, 2011, The cyclicality of price-cost margins in banking: An empirical analysis of its determinants, *Economic Inquiry* 49, 26–46.

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

Anderson, Eric, Sergio Rebelo, and Arlene Wong, 2018, Markups across space and time, Working paper, National Bureau of Economic Research.

Balasubramanyan, Lakshmi, Allen N Berger, and Matthew M Koepke, 2019, How do lead banks use their private information about loan quality in the syndicated loan market?, *Journal of Financial Stability* 43, 53–78.

Barkai, Simcha, 2019, Declining labor and capital shares, *Journal of Finance, Forthcoming* .

Becker, Bo, Marieke Bos, and Kasper Roszbach, 2018, Bad times, good credit, *Swedish House of Finance Research Paper* .

Begley, Taylor A, Amiyatosh Purnanandam, and Kuncheng Zheng, 2017, The strategic underreporting of bank risk, *The Review of Financial Studies* 30, 3376–3415.

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

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.

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, 2020, Third-party credit guarantees and the cost of debt: Evidence from corporate loans, *Available at SSRN 3492106* .

Bidder, Rhys M, John R Krainer, and Adam Hale Shapiro, 2020, De-leveraging 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.

Bottomley, Anthony, 1963, The cost of administering private loans in underdeveloped rural areas, *Oxford Economic Papers* 15, 154–163.

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.

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.

Chevalier, Judith A and David S Scharfstein, 1995, Liquidity constraints and the cyclical behavior of markups, *The American Economic Review* 85, 390–396.

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.

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

De Loecker, Jan and Frederic Warzynski, 2012, Markups and firm-level export status, *American economic review* 102, 2437–71.

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, Luc Laeven, and Gustavo A Suarez, 2017, Bank leverage and monetary policy's risk-taking channel: evidence from the united states, *the Journal of Finance* 72, 613–654.

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

Dell'Ariccia, Giovanni, Deniz Igan, and Luc UC Laeven, 2012, Credit booms and lending standards: Evidence from the subprime mortgage market, *Journal of Money, Credit and Banking* 44, 367–384.

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

Driscoll, John C and Ruth Judson, 2013, Sticky deposit rates, *Available at SSRN 2241531* .

Fernholz, Ricardo T. and Christoffer Koch, 2020, Why are big banks getting bigger?, *Unpublished working paper.* .

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

Frame, W Scott, Ping McLemore, and Atanas Mihov, 2020, Haste makes waste: Banking organization growth and operational risk .

Friedman, James W, 1971, A non-cooperative equilibrium for supergames, *The Review of Economic Studies* 38, 1–12.

Greenbaum, Stuart I, George Kanatas, and Itzhak Venezia, 1989, Equilibrium loan pricing under the bank-client relationship, *Journal of Banking & Finance* 13, 221–235.

Greenwald, Daniel L, John Krainer, and Pascal Paul, The credit line channel (Federal Reserve Bank of San Francisco 2020).

Grullon, Gustavo, Yelena Larkin, and Roni Michaely, 2019, Are us industries becoming more concentrated?, *The Review of Finance* 23, 697–743.

Gurun, Umit G, Gregor Matvos, and Amit Seru, 2016, Advertising expensive mortgages, *The Journal of Finance* 71, 2371–2416.

Gutiérrez, Germán and Thomas Philippon, 2017, Declining competition and investment in the us, Working paper, National Bureau of Economic Research.

Hale, Galina and Joao AC Santos, 2009, Do banks price their informational monopoly?, *Journal of Financial Economics* 93, 185–206.

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.

Heider, Florian, 2003, Leverage and asymmetric information about risk and value, *Unpublished working paper. New York University* .

Ioannidou, Vasso and Steven Ongena, 2010, "time for a change": loan conditions and bank behavior when firms switch banks, *The Journal of Finance* 65, 1847–1877.

Ivanov, Ivan, Luke Pettit, and Toni M Whited, 2020, Taxes depress corporate borrowing: Evidence from private firms, *Available at SSRN 3694869* .

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

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.

Liu, Ernest, Atif Mian, and Amir Sufi, 2019, Low interest rates, market power, and productivity growth, Working paper, National Bureau of Economic Research.

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

Nekarda, Christopher J and Valerie A Ramey, 2013, The cyclical behavior of the price-cost markup, Working paper, National Bureau of Economic Research.

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.

Olivero, Maria Pia, 2010, Market power in banking, countercyclical margins and the international transmission of business cycles, *Journal of International Economics* 80, 292–301.

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.

Prager, Robin A and Timothy H Hannan, 1998, Do substantial horizontal mergers generate significant price effects? evidence from the banking industry, *The Journal of Industrial Economics* 46, 433–452.

Qian, Jun, Philip E Strahan, and Zhishu Yang, 2015, The impact of incentives and communication costs on information production and use: Evidence from bank lending, *The Journal of Finance* 70, 1457–1493.

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

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.

Rotemberg, Julio and Garth Saloner, 1986, A supergame-theoretic model of price wars during booms, *New Keynesian Economics* 2, 387–415.

Rotemberg, Julio J and Michael Woodford, 1991, Markups and the business cycle, *NBER macroeconomics annual* 6, 63–129.

Rotemberg, Julio J and Michael Woodford, 1992, Oligopolistic pricing and the effects of aggregate demand on economic activity, *Journal of political Economy* 100, 1153–1207.

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

Santos, Joao AC and Andrew Winton, 2008, Bank loans, bonds, and information monopolies across the business cycle, *The Journal of Finance* 63, 1315–1359.

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

Schenone, Carola, 2010, Lending relationships and information rents: Do banks exploit their information advantages?, *The Review of Financial Studies* 23, 1149–1199.

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.

Stigler, George J, 1964, A theory of oligopoly, *Journal of political Economy* 72, 44–61.

Wang, Yifei, Toni M Whited, Yufeng Wu, and Kairong Xiao, 2018, Bank market power and monetary policy transmission: Evidence from a structural estimation, *Available at SSRN 3049665* .

Zarutskie, Rebecca et al., Does bank competition affect how much firms can borrow? new evidence from the us, *Federal Reserve Bank of Chicago Proceedings*, 856 (2003).

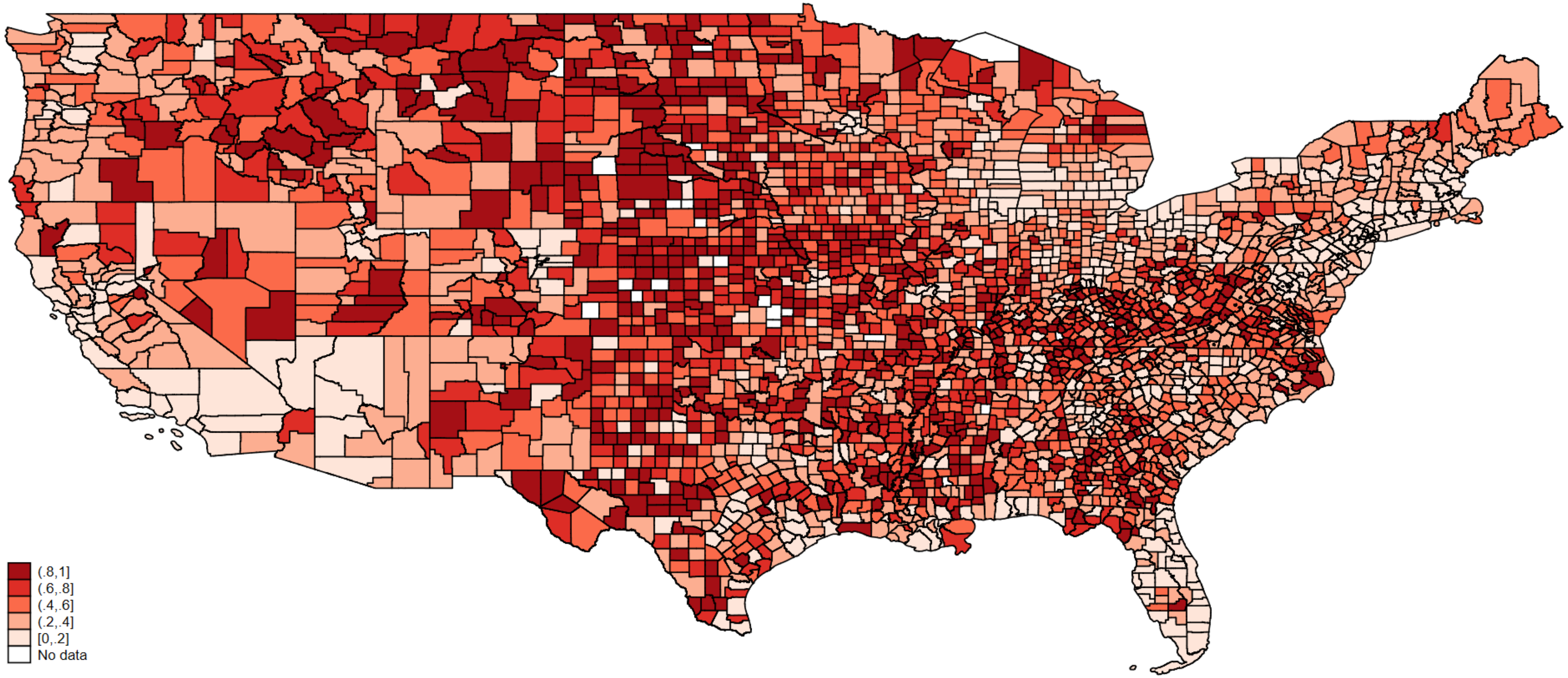
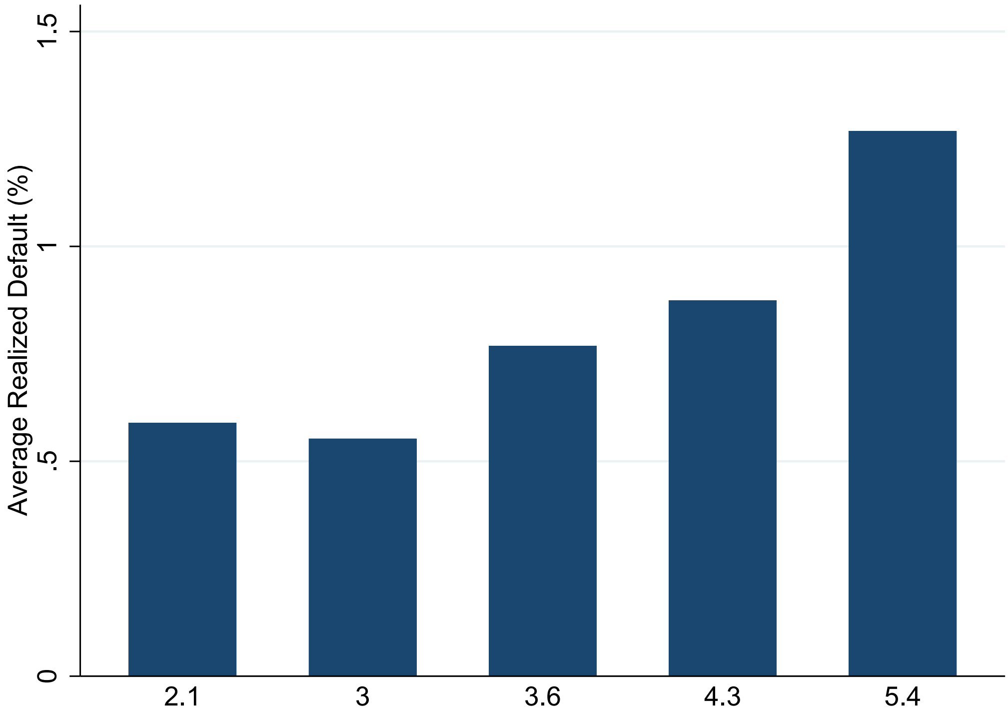
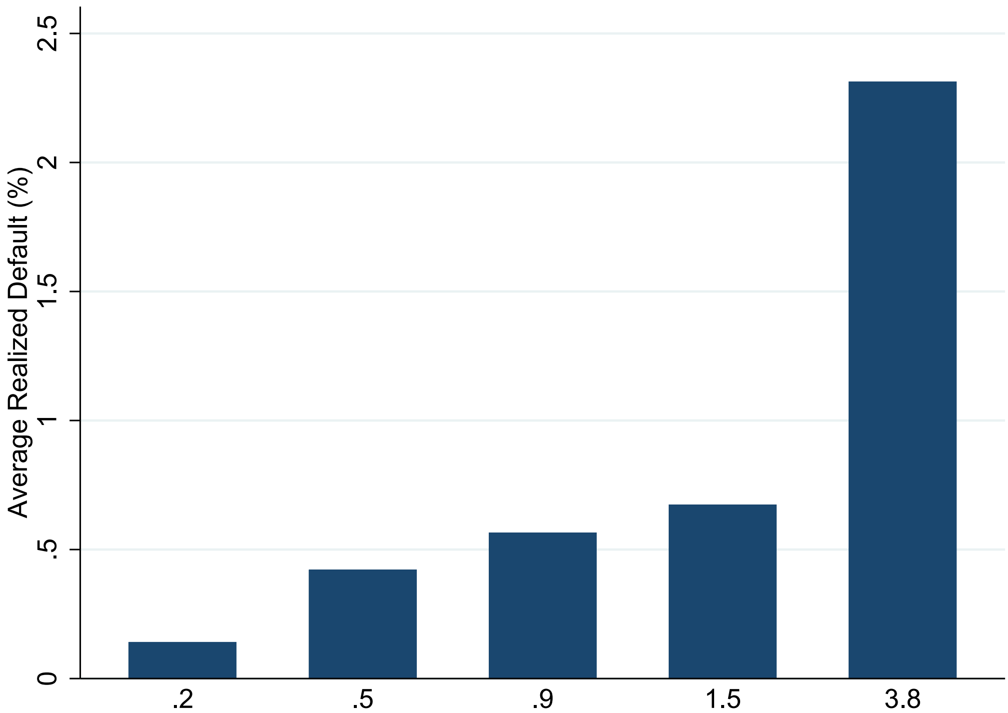


Figure 1: Loan HHI Heat Map

32



(a) Average Realized Default Rate Across Interest Rate Bins



(b) Average Realized Default Rate Across PD Bins

Figure 2a plots the average realized default rates over the twelve months following origination across five interest rate bins. Figure 2b plots the average realized default rates over the twelve months following origination across five PD bins. The average interest rate or PD in each bin is listed below each bar.

Table 1: Summary Statistics

This table contains summary statistics for loan-level, firm and geographic characteristics. Section A of the Appendix includes detailed definitions of all of our variables and filters.

	Mean	SD	10%	Median	90%	N
Loan Characteristics						
Amount (million USD)	6.97	14.15	1.04	2.55	15.77	28,000
Collateral	0.90	0.30	0.00	1.00	1.00	28,000
Non-Performance (%)	2.01	14.04	0.00	0.00	0.00	28,000
Floating Interest Rate	0.79	0.41	0.00	1.00	1.00	28,000
Guaranteed	0.49	0.50	0.00	0.00	1.00	28,000
Interest Rate (%)	3.66	1.17	2.16	3.63	5.25	28,000
LGD (%)	35.29	14.89	15.00	36.00	50.50	28,000
Markup (Baseline) (%)	-0.00	0.82	-0.94	-0.08	1.05	28,000
Markup (Risk-Adjusted) (%)	0.00	0.80	-0.91	-0.07	1.02	28,000
Maturity (months)	41.28	31.49	10.00	36.00	84.00	28,000
PD (%)	1.38	1.74	0.21	0.90	2.83	28,000
PDLGD (%)	0.46	0.61	0.06	0.29	0.99	28,000
Predicted IR (Baseline) (%)	3.66	0.84	2.62	3.62	4.81	28,000
Predicted IR (Risk-Adjusted) (%)	3.66	0.86	2.58	3.62	4.83	28,000
Revolver/Line of Credit	0.51	0.50	0.00	1.00	1.00	28,000
Realized Default (%)	0.81	8.97	0.00	0.00	0.00	28,000
Firm Characteristics						
Assets (million USD)	109.37	410.16	2.71	20.03	175.64	28,000
Leverage	0.33	0.26	0.01	0.30	0.69	27,474
Profitability	0.14	0.26	-0.01	0.08	0.33	28,000
Tangibility	0.91	0.17	0.67	0.99	1.00	27,930
Geographic Characteristics						
County HHI (county level)	0.85	0.21	0.52	1.00	1.00	1,505
County HHI (loan level)	0.48	0.24	0.23	0.41	0.89	28,000
MSA HHI (MSA level)	0.72	0.23	0.37	0.76	1.00	496
MSA HHI (loan level)	0.34	0.21	0.16	0.24	0.69	25,535

Table 2: Predictiveness of Risk Assessments on Loan Performance

This table tests whether banks internal risk assessments predict non-performance and default. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Non-Performance (%)		Realized Default (%)	
	(1)	(2)	(3)	(4)
PD (%)		1.455***		0.627***
		(5.963)		(2.855)
LGD (%)		0.023**		-0.001
		(2.341)		(0.169)
PDLGD (%)		-0.824		0.412
		(1.192)		(0.640)
Loan Characteristics Controls	YES	YES	YES	YES
Bank-Quarter FE	YES	YES	YES	YES
Industry-Quarter FE	YES	YES	YES	YES
Loan Type FE	YES	YES	YES	YES
Loan Purpose FE	YES	YES	YES	YES
Observations	28,000	28,000	28,000	28,000
Adj. R-squared	0.08	0.10	0.05	0.07

Table 3: Estimating Markup

This table tests whether banks internal risk assessments predict loan interest rates and is used to calculate both baseline and risk-adjusted markup. Columns (1) and (2) contains the estimation of the baseline markup and risk-adjusted markup, respectively. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Interest Rate (%)	
	(1)	(2)
PD (%)		0.075***
		(7.184)
LGD (%)		0.002***
		(3.184)
PDLGD (%)		0.161***
		(5.356)
Loan Characteristics Controls	YES	YES
Bank-Quarter FE	YES	YES
Industry-Quarter FE	YES	YES
Loan Type FE	YES	YES
Loan Purpose FE	YES	YES
Observations	28,000	28,000
Adj. R-squared	0.49	0.52

Table 4: Validity of Markup and Predicted Interest Rate

This table tests the validity of the baseline and risk-adjusted markup. Panel A uses Non-Performance as the dependent variable, while Panel B uses Realized Default. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Non-Performance (%)						
	(1)	(2)	(3)	(4)	(5)	(6)
Interest Rate (%)	0.264***			0.607***		
	(3.262)			(5.429)		
Markup (%) (Baseline)		0.603***			0.603***	
		(4.997)			(5.116)	
Predicted IR (%) (Baseline)		-0.065			0.673	
		(0.569)			(1.602)	
Markup (%) (Risk-Adjusted)			0.085			0.085
			(0.669)			(0.706)
Predicted IR (%) (Risk-Adjusted)			0.417***			5.045***
			(3.535)			(10.189)
Bank-Quarter FE				YES	YES	YES
Industry-Quarter FE				YES	YES	YES
Loan Type FE				YES	YES	YES
Loan Purpose FE				YES	YES	YES
Observations	28,000	28,000	28,000	28,000	28,000	28,000
Adj. R-squared	0.00	0.00	0.00	0.07	0.07	0.08

Panel A: Realized Default (%)						
	(1)	(2)	(3)	(4)	(5)	(6)
Interest Rate (%)	0.211***			0.365***		
	(3.824)			(4.586)		
Markup (%) (Baseline)		0.407***			0.407***	
		(4.653)			(4.814)	
Predicted IR (%) (Baseline)		0.021			-0.482*	
		(0.307)			(1.868)	
Markup (%) (Risk-Adjusted)			0.068			0.068
			(0.739)			(0.774)
Predicted IR (%) (Risk-Adjusted)			0.335***			2.899***
			(4.085)			(6.832)
Bank-Quarter FE				YES	YES	YES
Industry-Quarter FE				YES	YES	YES
Loan Type FE				YES	YES	YES
Loan Purpose FE				YES	YES	YES
Observations	28,000	28,000	28,000	28,000	28,000	28,000
Adj. R-squared	0.00	0.00	0.00	0.05	0.05	0.06

Table 5: Do Firm Characteristics and Market Concentration Drive Markups?

This table tests whether firm characteristics and market concentration drive risk-adjusted markups. The dependent variable is risk-adjusted markup. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by county or MSA depending on whether county or MSA HHI is used in the regression. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	MSA-Level			County-Level		
	(1)	(2)	(3)	(4)	(5)	(6)
MSA HHI	-0.143***	-0.131**	-0.153***			
	(2.588)	(2.439)	(2.884)			
County HHI				-0.150***	-0.145***	-0.175***
				(3.756)	(3.500)	(4.128)
Log(Assets)		-0.061***	-0.093***		-0.059***	-0.091***
		(9.281)	(15.467)		(14.030)	(18.433)
Leverage		0.077*	0.091**		0.084***	0.101***
		(1.826)	(2.403)		(2.829)	(3.633)
Tangibility		-0.339***	-0.492***		-0.319***	-0.471***
		(8.048)	(10.963)		(8.487)	(12.022)
Profitability		-0.049	-0.066**		-0.044*	-0.059**
		(1.604)	(2.117)		(1.675)	(2.197)
Bank-Quarter FE			YES			YES
Industry-Quarter FE			YES			YES
Loan Type FE			YES			YES
Loan Purpose FE			YES			YES
Observations	25,535	25,037	25,035	28,000	27,454	27,452
R-squared	0.00	0.03	0.04	0.00	0.02	0.04

Table 6: Switching Banks and Markups

This table tests whether firms that stay with their existing banks face higher markups. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by county or MSA depending on whether county or MSA HHI is used in the regression. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Risk-Adjusted Markup					
	(1)	(2)	(3)	(4)	(5)	(6)
Stay Bank	0.091**	0.113**	0.231***	0.136***	0.133**	0.187*
	(2.188)	(2.556)	(2.865)	(2.642)	(2.478)	(1.758)
Stay Bank × MSA HHI			-0.475*			
			(1.796)			
Stay Bank × County HHI						-0.163
						(0.557)
Log(Assets)		-0.084***	-0.085***		-0.066***	-0.067***
		(4.693)	(4.701)		(4.627)	(4.633)
Leverage		0.289***	0.289***		0.314***	0.317***
		(4.944)	(4.977)		(4.318)	(4.331)
Tangibility		-0.563***	-0.568***		-0.387**	-0.389**
		(4.584)	(4.633)		(2.104)	(2.114)
Profitability		-0.156**	-0.157**		-0.141*	-0.140*
		(2.210)	(2.233)		(1.665)	(1.653)
MSA-Quarter FE	YES	YES	YES			
County-Quarter FE				YES	YES	YES
Bank-Quarter FE		YES	YES		YES	YES
Industry-Quarter FE		YES	YES		YES	YES
Loan Type FE		YES	YES		YES	YES
Loan Purpose FE		YES	YES		YES	YES
Observations	4,662	4,457	4,457	4,178	3,929	3,929
R-squared	0.30	0.49	0.49	0.43	0.64	0.64

Table 7: Loan-level markups and bank market share in the MSA/county

This table tests whether markups are higher for loans from banks with higher market shares. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by county or MSA depending on whether county or MSA HHI is used in the regression. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Risk Adjusted Markup (%)			
	(1)	(2)	(3)	(4)
Bank Market Share in MSA	0.122***	0.172***		
	(3.474)	(4.694)		
Bank Market Share in County			0.109***	0.162***
			(3.731)	(5.237)
Log(Assets)	-0.068***	-0.098***	-0.061***	-0.091***
	(12.811)	(16.310)	(11.560)	(14.796)
Leverage	0.074**	0.085**	0.071**	0.080**
	(2.213)	(2.577)	(2.202)	(2.381)
Tangibility	-0.409***	-0.577***	-0.379***	-0.519***
	(8.998)	(11.874)	(8.519)	(10.965)
Profitability	-0.068**	-0.082**	-0.062**	-0.067**
	(2.147)	(2.419)	(2.052)	(2.157)
MSA-Quarter FE	YES	YES		
County-Quarter FE			YES	YES
Bank-Quarter FE		YES		YES
Industry-Quarter FE		YES		YES
Loan Type FE		YES		YES
Loan Purpose FE		YES		YES
Observations	22,951	22,947	22,491	22,471
R-squared	0.19	0.21	0.29	0.33

Table 8: The Aggregate Relationship between Markups, Loan Volume and Collateralization

This table tests whether firms that stay with their existing banks face higher markups. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by county or MSA depending on whether county or MSA HHI is used in the regression. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Loan Volume		Average Collateralized	
	(1)	(2)	(3)	(4)
Average MSA Markup	-0.058*		0.013**	
	(1.762)		(2.038)	
Average County Markup		-0.072***		0.017***
		(3.167)		(3.649)
MSA FE	YES		YES	
County FE		YES		YES
Quarter FE	YES	YES	YES	YES
Observations	4,969	9,390	4,969	9,390
R-squared	0.56	0.45	0.32	0.33

Appendix A. Variable Definitions

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

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

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

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

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

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

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

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

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

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.

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 above, from Y-14Q.

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.

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

PDLGD: $PD \times GD$, from Y-14Q.

Predicted Interest Rate: The predicted interest rate from regression (3) in percentage points, the baseline model excludes PD, LGD and PDLGD, while the risk-adjusted model includes PD, LGD and PDLGD, from Y-14Q.

Markup: The estimated residual from Equation (3) in percentage points, the baseline model excludes PD, LGD and PDLGD, while the risk-adjusted model includes PD, LGD and PDLGD, from Y-14Q.

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

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

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

Stay Bank: Dummy variable that equals one if the firm borrows from the bank it received its previous loan from, from Y-14Q.

Average Markup: Average quarterly risk-adjusted markup at the county/MSA level in percentage points, from Y-14Q.

Loan Volume: Log total quarterly loan volume at the county/MSA level, from Y-14Q.

Average Collateralized: The quarterly average number of loans that are collateralized at the county/MSA, from Y-14Q.

High HHI (County): a dummy variable that equals one if the County HHI is above the median in the sample, from Y-14Q.

High HHI (MSA): a dummy variable that equals one if the MSA HHI is above the median in the sample, from Y-14Q.

Population Density: Average county population per square mile, from Census.

Appendix B. Robustness Tests

Table B1: Robustness of Alternative Measures of Concentration on Markup

This table tests the relationship between alternative measures of concentration and risk-adjusted markup. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by county or MSA depending on whether county or MSA measures of concentration are used in the regression. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Risk-Adjusted Markup (%)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High HHI (MSA)	-0.058***	-0.074***						
	(2.713)	(3.161)						
High HHI (County)			-0.089***	-0.103***				
			(3.957)	(4.466)				
Log(Population Density)					0.027***	0.034***		
					(2.795)	(3.281)		
Deposit HHI							-0.083	-0.105
							(0.839)	(0.969)
Log(Assets)	-0.059***	-0.090***	-0.059***	-0.090***	-0.059***	-0.091***	-0.059***	-0.091***
	(13.901)	(18.201)	(13.995)	(18.294)	(14.136)	(18.568)	(13.945)	(18.195)
Leverage	0.079***	0.098***	0.086***	0.102***	0.085***	0.103***	0.075**	0.094***
	(2.658)	(3.503)	(2.848)	(3.628)	(2.841)	(3.744)	(2.541)	(3.333)
Tangibility	-0.324***	-0.475***	-0.321***	-0.473***	-0.317***	-0.472***	-0.325***	-0.476***
	(8.549)	(12.053)	(8.478)	(11.974)	(8.425)	(11.966)	(8.512)	(12.032)
Profitability	-0.040	-0.058**	-0.044*	-0.059**	-0.045*	-0.058**	-0.037	-0.056**
	(1.555)	(2.146)	(1.703)	(2.218)	(1.697)	(2.171)	(1.435)	(2.088)
Bank-Quarter FE		YES		YES		YES		YES
Industry-Quarter FE		YES		YES		YES		YES
Loan Type FE		YES		YES		YES		YES
Loan Purpose FE		YES		YES		YES		YES
Observations	27,454	27,452	27,454	27,452	27,454	27,452	27,454	27,452
R-squared	0.02	0.04	0.03	0.04	0.03	0.04	0.02	0.04

Table B2: Baseline Markup Estimated with Firm Characteristics

This table tests whether an estimate of markup using firm characteristics but not banks' private risk assessments predicts loan performance. T-statistics are shown below the parameter estimates in parenthesis and are calculated using robust standard errors clustered by firm. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Non-Performance (%)	Realized Default (%)
	(1)	(2)
Markup (%)	0.632***	0.418***
	(5.088)	(4.713)
Predicted IR (%)	0.571*	0.032
	(1.846)	(0.169)
Bank-Quarter FE	YES	YES
Industry-Quarter FE	YES	YES
Loan Type FE	YES	YES
Loan Purpose FE	YES	YES
Observations	27,285	27,285
Adj. R-squared	0.07	0.05