Bank Information Production Over the Business Cycle*

Cooper Howes[†] Gregory Weitzner[‡]

January 2023

Abstract

The information banks have about borrowers drives their lending decisions and macroeconomic outcomes, but this information is inherently difficult to analyze because it is private. We construct a novel measure of bank information quality from confidential regulatory data that include banks' private risk assessments for US corporate loans. Information quality improves as local economic conditions deteriorate, particularly for newly originated loans and loans with larger potential losses. Our results provide empirical support for theories of countercyclical information production in credit markets.

*We thank Hassan Afrouzi, Cynthia Balloch, Javier Bianchi, Olivier Coibion, Mariela Dal Borgo, Adolfo De Motta, Miguel Faria-e-Castro, Daniel Greenwald, Stefan Jacewitz, Gustavo Joaquim, Yueran Ma, Blake Marsh, Johannes Matschke, Karel Mertens, Atanas Mihov, Lars Norden, Guillermo Ordoñez, Pablo Ottonello, Samuel Rosen, Kasper Roszbach, Guillaume Roussellet, Jane Ryngaert, Garrett Schaller, Padma Sharma, Lee Smith, Wenting Song, Andrea Vedolin, Yufeng Wu and Choongryul Yang as well as seminar and conference participants at the Bank of Canada, Bentley University (Economics), Federal Reserve Bank of Kansas City, Federal Reserve Board, IBEFA Annual Meeting, McGill University (Finance), Norges Bank, University of Notre Dame (Economics), University of Oklahoma (Economics), Oxford Saïd – Risk Center at ETH Zürich Macro-finance Conference, George Washington University (Finance), University of Virginia (Finance), CAFRAL, CICF, FDIC Bank Research Conference, the Finance Forum, Fixed Income and Financial Institutions Conference, NFA, SEA 2021 annual meetings, and NAMES 2022 summer meetings for helpful comments and discussions. These views are those of the authors and do not reflect the views of the Federal Reserve Board of Governors or the Federal Reserve System.

[†]Federal Reserve Board of Governors. Email: cooper.a.howes@frb.gov.

[‡]McGill University. Email: gregory.weitzner@mcgill.ca.

1 Introduction

A fundamental role of banks is to produce information about prospective borrowers.[1] Banks use this information to determine the recipients and terms of financing; hence, their information production decisions can affect real economic activity and financial stability through the supply of credit to firms. If the returns to distinguishing between different types of borrowers change with economic conditions, banks' incentives to produce information can affect and be affected by business cycles (e.g., Dang, Gorton, and Holmström (2012) and Gorton and Ordonez (2014)). Despite policymaker interest and an extensive theoretical literature[2] emphasizing the importance of banks' information, there is little evidence of its empirical properties.

The key empirical challenge to testing theories of bank information production is that banks' information is intrinsically private, and therefore unobservable to the econometrician. Because of this data limitation, researchers often rely on indirect evidence; however, without access to banks' private information, researchers are severely constrained in their ability to test these theories. In this paper, we address this challenge using confidential regulatory data that contain banks' private risk assessments for the vast majority of corporate bank loans in the US. We first create a measure of bank information quality based on how well banks' private risk assessments predict realized defaults. Next, we use county-level variation in unemployment rates to show that information quality is countercyclical, i.e., it improves as local economic conditions worsen. Finally, consistent with banks actively producing more information when their incentives to do so are higher, we find that the sensitivity of information quality to the business cycle is concentrated in loans which theory predicts to be more information sensitive: new loans and loans with higher potential losses. Overall, our results provide empirical support for theories in which banks' have stronger incentives to produce information during economic downturns.

Our analysis uses the Federal Reserve's Y-14Q Schedule H.1 data that include all

[1] e.g., Leland and Pyle (1977), Diamond (1984) and Boyd and Prescott (1986).

[2] A non-exhaustive list of theoretical analyses of information production in credit markets includes: Gorton and Pennacchi (1990), Thakor (1996), Fulghieri and Lukin (2001), Dang, Gorton, and Holmström (2012), Chemla and Hennessy (2014), Yang and Zeng (2019), Yang (2020) and Weitzner (2019).

corporate loans larger than one million dollars extended by large bank holding companies (BHCs) beginning in 2011. In addition to detailed loan and borrower characteristics, qualified BHCs[3] are required to report their internal estimate of the borrower's probability of default (PD) for each loan. Because the data also reveal if and when loans ultimately default, these PDs—which incorporate both "hard" and "soft" information—allow us to quantitatively analyze bank information quality.

We first show that banks' PD estimates are statistically and economically significant predictors of realized default even after controlling for a rich set of loan- and firm-level controls. This result suggests that banks' risk assessments contain private information that is i) relevant for predicting default, and ii) not captured by other observables, including the loan's interest rate. We define our measure of information quality as the size of the PD coefficient in OLS regressions predicting future realized default for newly originated loans.

Next, we test the cyclicality of banks' information quality. We overcome the identification challenges arising from the short time dimension of our data by exploiting rich geographical variation in local economic conditions within the US. Specifically we compare the information quality for two similar loans given by the same bank, at the same time, to similar borrowers located in counties with different local unemployment rates. Using this approach, we find that banks' information quality improves as local economic conditions deteriorate. Specifically, our estimates imply that a one percentage point increase in the local unemployment rate increases the sensitivity of realized default to PD by roughly one third of its average level. This result is consistent with theories in which banks produce more information when economic conditions weaken as the returns to distinguishing between different borrower types increase.[4]

While this result is consistent with banks producing more information about borrowers during downturns, it is also possible that banks' information quality varies exogenously over the business cycle. We perform several tests to disentangle these two channels.

[3]Details about participating institutions can be found here: https://www.federalreserve.gov/supervisionreg/stress-tests-capital-planning.htm.

[4]See Ruckes (2004), Dell'Ariccia and Marquez (2006), Dang, Gorton, and Holmström (2012) and Gorton and Ordonez (2014).

First, we analyze how the R^2 obtained from regressing default on PD and firm/loan characteristics changes over the business cycle. During periods of elevated unemployment, the total R^2 of regressions predicting default is smaller, suggesting that default does not become easier to predict during downturns. However, we also find that the marginal contribution of PD to the regression's R^2 is higher. As an additional check we show that PDs estimated purely based on observable characteristics do not exhibit any statistically significant cyclicality in their ability to predict default. Together, these results suggest that PD—which incorporates banks' private information—becomes *more* useful for predicting default at the same time other observable characteristics become *less* useful, and is consistent with banks exerting greater effort to distinguish between borrowers during downturns.

Next, we analyze how bank information quality varies across loans based on their information sensitivity, which Dang, Gorton, and Holmström (2013) define to be the incentive that banks have to produce information about a loan based on its characteristics. First, banks' information production incentives should be more sensitive to the business cycle for new loans because they are risking additional capital, while for already issued loans that capital is already sunk. Consistent with this prediction, we find that the cyclical sensitivity of information is almost entirely driven by newly originated loans. We also find that the dispersion of risk assessments across banks for the same borrower narrows during periods of higher unemployment for new loans, but not for previously issued loans, suggesting that banks' information converges as they produce information about new borrowers in bad times.

Second, we test whether banks have higher-quality information for loans with higher potential losses. According to several theories of information production in credit markets such as Dang, Gorton, and Holmström (2012) and Gorton and Ordonez (2014), banks should produce more information for loans with larger potential losses during downturns, as these are the loans for which the returns to distinguishing between borrowers will be highest. We calculate a measure of potential losses by multiplying the loss given default (LGD), which is defined as the share of the loan that the bank would expect to lose in the event of a default, times the amount of the loan. We then estimate regressions that include interactions between a loan's PD, its potential losses, and the local unemployment rate. We find that the coefficient is positive and statistically significant, suggesting that information quality is more cyclically sensitive for loans with higher potential losses. Overall, these results provide additional support for the information production channel, and to our knowledge, are the first in the literature highlighting how a loan's potential losses affect its information sensitivity over the business cycle.

Taken together, our results suggest that banks produce more information in bad times as the returns to distinguishing across borrower types increase. These findings provide direct empirical support for models featuring endogenous bank information production and have important implications for policymakers, as many of these models emphasize the benefits of policy interventions that limit information production incentives during downturns. By directly analyzing the quality of banks' private information, we view our paper as an important step in understanding the critical link between information production, bank lending, and real economic activity.

Literature review. Our paper relates to the empirical literature on bank information production. A subset of this literature focuses on banks' monitoring over the life of loans (e.g., Ono and Uesugi (2009), Cerqueiro, Ongena, and Roszbach (2016), Gustafson, Ivanov, and Meisenzahl (2020)). Gustafson, Ivanov, and Meisenzahl (2020) create a measure of monitoring based on the number of visits banks take to firms. In contrast, we are focused on banks' information about borrower risk at loan origination.

Other papers analyze information production in the primary market. For example, Keys et al. (2010) and Keys, Seru, and Vig (2012) analyze screening in the consumer loan market, while Iyer et al. (2016) analyze information production in an online peer lending platform. Lisowsky, Minnis, and Sutherland (2017) show that banks collected less information from construction firms in the run-up to the 2008-09 financial crisis. Bedayo et al. (2020) analyze the time to originate over the cycle in Spain. They find that banks spend more time originating loans in downturns, which is consistent with our results to the extent that this increased effort leads to more accurate PD estimates.

The key difference in methodology between the aforementioned papers and our approach is we directly use banks' private information to analyze banks' information production decisions.[5]

The paper whose empirical approach is closest to ours is Becker, Bos, and Roszbach (2020), who find that bank credit ratings perform better at predicting default in bad economic times. There are several key differences in both our analysis and the interpretation of our results. First, their data are restricted to a single Swedish bank. Because of this, they rely on a single time series measure of aggregate economic conditions. In contrast, our paper exploits variation in economic conditions across the US at each point in time. This allows us to rule out supply-side effects at the bank level because we compare information quality across loans issued by the same bank across different regions with different economic conditions. Second, their data are at the firm level rather than the loan level. This difference allows us to explore the relationship between loan characteristics and information production, as well as how this relationship changes over the business cycle. Finally, we provide evidence that the countercyclicality of information quality is driven by endogenous bank information production by showing that the effects are almost entirely concentrated in new loans and loans with higher potential losses, both of which are difficult to rationalize solely through exogenous fluctuations in information precision over the business cycle.

While we focus on information production, our work also relates to the empirical literature on the cyclicality of lending standards (e.g., Asea and Blomberg (1998), Lown and Morgan (2006), Dell'Ariccia, Igan, and Laeven (2012), Jiménez et al. (2014), Bassett et al. (2014), Dell'Ariccia, Laeven, and Suarez (2017), Maddaloni and Peydró (2011), Rodano, Serrano-Velarde, and Tarantino (2018), Mariathasan and Zhuk (2018), and Dempsey, Ionescu et al. (2019)). For the most part, existing studies find that lending standards are countercyclical. Our results complement these findings by suggesting banks are producing

[5]Our paper also relates to empirical work analyzing bank internal risk-measures (e.g., Treacy and Carey (2000), Grunert, Norden, and Weber (2005), Mester, Nakamura, and Renault (2007), Agarwal and Hauswald (2010), Qian, Strahan, and Yang (2015), Behn, Haselmann, and Vig (2016), Berg and Koziol (2017), Dell'Ariccia, Laeven, and Suarez (2017), Plosser and Santos (2018), Nakamura and Roszbach (2018), Becker, Bos, and Roszbach (2020), Adelino, Ivanov, and Smolyansky (2019) and Beyhaghi, Fracassi, and Weitzner (2020)).

more information when economic conditions deteriorate.

Our paper also relates to the theoretical work analyzing the cyclicality of information production in credit markets. This includes an extensive theoretical literature in which information production is countercyclical (e.g., Ruckes (2004), Gorton and He (2008), Gorton and Ordonez (2014) Gorton and Ordonez (2020), Fishman, Parker, and Straub (2020), Dell'Ariccia and Marquez (2006), Petriconi (2015), Farboodi and Kondor (2020) and Asriyan, Laeven, and Martin (2022)).[6]

Finally, our work complements research analyzing the cyclicality of attention in macroeconomic settings more broadly. One closely related example is Cao et al. (2022), who provide evidence of countercyclical due diligence in syndicated lending markets. Other related empirical work includes Coibion and Gorodnichenko (2015), who show that forecast quality for macroeconomic aggregates such as inflation is countercyclial, and Song and Stern (2021) and Flynn and Sastry (2021), who show that firm attention to macroeconomic news is countercyclical. Other theoretical research analyzing the causes and consequences of information production decisions in macroeconomic settings includes Mäkinen and Ohl (2015), Benhabib, Liu, and Wang (2016), Chiang (2021). Our work connects these research agendas to those from the finance literature by providing direct empirical evidence of the relationship between information production decisions in the banking sector and macroeconomic conditions.

The paper is structured as follows. Section 2 describes our data. Section 3 presents empirical evidence that information quality is countercyclical and evidence that this is driven by banks' endogenous information production decisions. Section 4 discusses the broader implications of our results through the lens of several theories featuring endogenous bank information production. Section 5 concludes.

[6]Several of our tests also appeal to predictions in the theoretical literature relating security design and information production such as Boot and Thakor (1993), Gorton and Pennacchi (1990), Fulghieri and Lukin (2001), Dang, Gorton, and Holmström (2012), Gorton and Ordonez (2014), Dang, Gorton, and Holmstrom (2019), Yang and Zeng (2019), Yang (2020), and Weitzner (2019).

2 Data

Our main source of data is Schedule H.1 of the Federal Reserve's Y-14Q filings. The Federal Reserve began collecting these data to support the Dodd-Frank mandated stress tests and the Comprehensive Capital Analysis and Review (CCAR). The sample includes commercial and industrial (C&I) loans from banks with \$50bn or more in total assets[7], accounting for 85.9% of all assets in the banking sector (Frame, McLemore, and Mihov (2020)). Qualified banks are required to report detailed quarterly loan-level data on corporate loans of at least \$1mm in size. The universe of loans we analyze is large: Bidder, Krainer, and Shapiro (2020) show that the Y-14Q data cover 70% of all commercial and industrial loan volume. The data include detailed loan characteristics (such as interest rates, maturity, amount, collateral, and purpose) and performance measures (defaults, past-due payments, non-accruals, and charge-offs). They also include income, balance sheet, and geographic information about borrowers. Crucially, banks are also required to report their internal estimates of the borrower's probability of default (PD) for each loan to the Federal Reserve on their Y-14Q filings.

Because we are focused on banks' information production incentives at the time financing is committed, our baseline results only include newly originated loans. We exclude demand loans, which can be recalled by the borrower at any time, as well as loans with government guarantees, tax-exempt loans, loans to foreign borrowers, and loans to firms in the finance, insurance, and real estate (FIRE) sectors. We drop loans with negative interest rates, or interest rates over 100%, as well as those with missing company identifiers, PD, or loan amount at origination. We follow Brown, Gustafson, and Ivanov (2021) and exclude loans to companies with under \$100k in reported assets at origination; given that the minimum reporting thresholds for loans is \$1 million, these observations are likely reporting errors. We also drop loans with PDs that are above the 99th percentile at origination to minimize the effects of outliers and reporting errors. Finally, we drop firms with assets above the 99th percentile and publicly traded firms, as these firms are

[7]In 2019, this threshold was increased to \$100bn.

likely to be more geographically diverse and thus less sensitive to changes in local economic conditions. Our sample period starts in 2014Q4, which is when the PD variable first becomes well populated. We exclude data after 2020Q1 due to the exceptional circumstances surrounding the pandemic, and thus include new loans through 2019Q1 to allow at least one year for loans to default.

We define the following firm-level financial variables: profitability (EBITDA/assets), size (log assets), tangibility (tangible assets/assets), and leverage (debt/assets), which we winsorize at the 1% and 99% level. Our main measure of loan performance is default, which is a dummy variable that equals 1 if the borrower defaults within two years after origination. Focusing on a two-year default window strikes a balance between the limited time series dimension of our data and the fact that the median loan maturity is close to five years. The full details of the variable construction as well as the sources, purpose, and properties of the PD estimates are described in Appendix A.

Table 1 includes firm, loan, and county summary statistics. Panel A shows summary statistics at the loan level for newly originated loans, where the average and median loan size is approximately \$12.7mm and \$3.4mm, respectively. To calculate the firm-level statistics in Panel B, we average each reported measure at the firm-quarter level across all outstanding loans. The median firm has \$20.4mm in assets and a leverage ratio of 0.27. These loan and firm sizes are small relative to other sources of loan data such as DealScan, reflecting the fact that our sample contains substantially more small, nonpublic firms. Over our sample period, 0.40% of loans default within two years after loan origination. This compares to an average ex-ante expected PD of 1.40%, suggesting that economic conditions were relatively benign during this period relative to banks' ex-ante expectations. Finally, Panel C shows characteristics calculated at the county level. The median number of loans outstanding for each county is 6, while the median new loan volume in each quarter is about \$41.3mm.

This dataset is uniquely suited to analyze the dynamics of banks' information quality over the business cycle for several reasons. First, as mentioned earlier, it covers the vast majority of commercial bank loans in the US. Second, it includes both realized and expected default rates at the loan level, which allows us to create an empirical measure of bank information quality and analyze its properties over the business cycle.

Figure 1 shows the distributions of both PD and log(PD). If PD contains information useful for predicting default, then there should be a positive correlation between PD and future realized default. Figure 2 confirms this relationship holds in our data. Each column corresponds to a PD quintile, with the number below the column representing the average level of PD for that bucket of loans, while the vertical axis represents the average realized default rate for loans in that bucket. There is a clear positive relationship between PDs and realized defaults, suggesting that PD has useful information for predicting default. In the next section, we show this formally in regressions and implement our approach to measuring bank information quality.

3 Empirical Results

This section contains our main empirical results. We first justify our approach to measuring information quality by showing that PD is a statistically and economically significant predictor of realized default in Section 3.1. In Section 3.2, we show that PD becomes a better predictor of default as the local unemployment rate increases. In Section 3.3 we conduct several tests which suggest that the cyclicality of banks' information quality is driven by endogenous information production.

3.1 Predicting Default

We first formally confirm that banks' PD estimates predict default. To do so we estimate the following linear regression:

$$Default_i = \beta PD_i + \Omega X_i + \delta_{b,t} + \gamma_{j,t} + \sigma_{b,c} + \epsilon_i, \tag{1}$$

where i, b, t, j and c, index loan, bank, quarter, industry and county, respectively. $Default_i$ is a dummy variable that equals 1 if loan i defaults within eight quarters following origination. PD_i is defined as the banks' estimate of PD; however, as robustness checks, we show very similar results using both log(PD) and the percentile rank of PD within each bank/quarter for our main results in the Appendix. X_i is a vector of firm and loan characteristics which include firm size (log of total assets), leverage ratio (total debt to total assets), profitability ratio (EBITDA to total assets), and tangibility ratio (tangible assets to total assets), log loan size, the log of the original loan maturity in months, the bank's estimate of loss given default per dollar of debt (LGD), as well as loan type fixed effects. We include bank-quarter fixed effects ($\delta_{b,t}$) to absorb any differences in banks' risk assessment models and cost of capital, industry-quarter fixed effects ($\gamma_{j,t}$) to absorb variation in average loan performance across industries, and bank-county fixed effects ($\sigma_{b,c}$) to absorb persistent differences in risk assessment models or credit analysts across geographies. Throughout all the regressions we cluster standard errors at the county level.

The results are shown in Table 2. The primary coefficient of interest is β, which represents the expected increase in realized default (measured in percentage points) from a one percentage point increase in a loan's PD. In Column (1), the coefficient estimate is 0.243, which means that an increase in PD of 1pp increases the probability of realized default by about 24bps. In Column (2), we display the results with firm and loan characteristics, and in Column (3) we include the interest rate as an additional regressor.[8] These specifications show similar results, suggesting that PD captures information that is useful for predicting default even after controlling for the interest rate and other observable characteristics.

3.2 Information Quality Over the Business Cycle

In this section, we test the cyclicality of bank information quality by analyzing how changes in local economic conditions affect PD's regression coefficient and R^2 when predicting default. Our measure of county-level economic conditions is the unemployment rate from the BLS. The Y-14Q data use ZIP codes as geographical identifiers, so we first use the ZIP-to-county crosswalks from the Department of Housing and Urban Develop-

[8]The number of observations drops once we add interest rate variables because banks do not report interest rates for undrawn credit lines.

ment to assign a county to each zip code before merging it with the unemployment rate data. While the aggregate unemployment rate declined steadily from 5.7% to 3.9% between 2014Q4 and 2019Q1, Figure 6 highlights the substantial cross-sectional variation in the changes in county-level unemployment rates over this period, with roughly one quarter of counties experiencing an increase. Figure 7, which shows a histogram of defaults across county unemployment rates, suggests that the variation in defaults is not coming solely from high-unemployment areas, while Appendix Table B.12 shows that most counties experienced meaningful variation in the unemployment rate during our sample period.

To test if the ability of PD to predict realized default changes across different economic conditions, we estimate separate predictive regressions based on whether the county-level unemployment rate is above or below that county's median during our sample. The first two columns of Table 3 show the results of regressing default on controls and fixed effects alone. The R^2 is higher (0.310 versus 0.269) when the unemployment rate is below the county's median. To test the statistical significance of this difference, we bootstrap the regression 250 times and estimate a t-statistic of -9.8. This soundly rejects the null hypothesis that the R^2 values are equal across periods of high and low unemployment, suggesting that observables do a better job at explaining realized default in good economic times.

In Columns (3) and (4) we report the results of regressions that include PD as the sole independent variable without any controls or fixed effects. While the coefficient on PD is statistically significant for both high and low unemployment periods, it is about three times larger during periods of high unemployment. Furthermore, the R^2 is higher during high-unemployment periods, and the bootstrapped t-statistic of 41.3 indicates that this difference is statistically significant. These results suggest PD becomes a better predictor of default as local economic conditions deteriorate.

Finally, in Columns (5) and (6) we include both PD and controls/fixed-effects. The coefficient on PD remains much higher during periods of high unemployment. Moreoever, the marginal contribution of PD to the total R^2 is higher during periods of high unemployment (going from 0.269 to 0.274) than during periods of low unemployment (going from 0.310 to 0.312). Taken together, these results suggest that PD becomes a more useful predictor of default during periods of high unemployment, while other firm and loan characteristics become less useful. This result is consistent with banks putting more effort into distinguishing between borrowers during downturns and inconsistent with default simply being easier to forecast during these periods.

Next, we directly test whether the sensitivity of realized default to PD varies over the business cycle by estimating the following regression:

$$Default_i = \beta_0 PD_i + \beta_1 UR_{c,t} + \beta_2 (PD_i \times UR_{c,t}) + \Omega X_i + \delta_{b,t} + \gamma_{j,t} + \sigma_{b,c} + \epsilon_i. \quad (2)$$

This regression is similar to Equation (1) with the addition of an interaction term between predicted default and the county-level unemployment rate ($PD_i \times UR_{c,t}$). The coefficient of interest is β_2, which represents the change in the sensitivity of actual to predicted default given a one percentage point increase in the unemployment rate.[9] Our use of bank-by-quarter fixed effects means that our results will not be driven by supply-side factors affecting lending decisions at the bank level, such as changes in a bank's cost of capital or bank-level risk appetite.

The results are displayed in Table 4. The first two columns show the results with and without the inclusion of controls. Across both of these specifications, we find a positive and statistically significant coefficient for β_2, which suggests that banks' PDs are better at predicting default in bad times. Column (1) shows a 1pp increase in the unemployment rate increases the coefficient on PD by about 8 basis points, which represents about one third of the average effect of PD estimated in Section 3.1. Column (2) adds interactions between firm and loan-level controls and PD and shows that the estimates become slightly larger in magnitude and remain statistically significant. This result suggests that our results are not being driven simply by changes in loan and firm characteristics over the business cycle. Columns (3) and (4) add county-quarter fixed effects, which absorb level

[9] Appendix Table B.6 shows very similar results using the lagged, rather than the contemporaneous, unemployment rate.

differences in activity across different counties in each quarter, and show very similar coefficient estimates.

Next, we construct a "predicted" PD by taking the fitted values from regressing PD on observable characteristics, i.e., the same controls and fixed effects used in Equation (1). We then compare the estimated interaction coefficients for these predicted PDs to those calculated using banks' actual reported PDs. The results, which are displayed in Table 5, show that predicted PDs do not display any statistically significant changes in their ability to predict default over the business cycle. However, Columns (2) and (4) show that the interaction coefficients between the unemployment rate and banks' actual PDs displays similar magnitudes and levels of statistical significance to our baseline results even when the predicted PDs are included in the same regression.

Overall, these results suggest that increases in unemployment are associated with statistically and economically significant improvements in bank information quality. Hence, we conclude that bank information quality is countercyclical. In the next section, we provide further evidence for the mechanisms driving our results.

3.3 Mechanisms

While our empirical results are consistent with banks producing more information about borrowers during downturns, we cannot directly observe banks' information production decisions, only the ability of banks' PDs to predict default. Hence, it is possible that our empirical results are simply driven by exogenous variation in information quality. In this section, we develop several additional tests in order to distinguish between these channels.

First, we compare the cyclicality of bank information quality for newly issued loans to those which were issued in prior quarters. Intuitively, the marginal value of information about a borrower's quality should be highest prior to the capital being sunk. If banks' incentives are driving them to produce more information about their loans in bad times, we would thus expect these effects to be concentrated in newly originated loans rather than loans which were previously originated.

To test this hypothesis, we extend our sample to include all observations of each loan, rather than focusing exclusively on the quarter of origination as we have throughout the rest of our analysis. We then estimate a modified version of Equation (2) that includes interactions with *New Loan*, which is a dummy variable that equals one if the loan is originated in that quarter.

The results are shown in Table 6. Because these regressions include previously issued loans on banks' balance sheets in addition to newly originated loans, the sample size for these regressions is much larger. First, note that the interaction term between PD and the new loan indicator is negative and statistically significant. This result suggests that PD becomes a better predictor of default after origination, which is not surprising given that banks likely learn more about a borrower's risk over time. However, the triple interaction term between PD, the new loan indicator, and the unemployment rate is positive and statistically significant across all specifications, which implies that information quality is more sensitive to economic conditions for new loans than for loans issued in prior quarters. In addition, the coefficient capturing the interaction between PD and the unemployment rate is consistently much smaller than the triple interaction term, suggesting that information quality is less sensitive to local economic conditions for previously issued loans. This result provides support for endogenous information production as a driving force behind the cyclicality of information quality we observe in the data.

We also plot how both the sensitivity of PD to realized default and its cyclicality evolve over the life of the loan in Figures 3 and 4. These figures plot regression coefficients from a modified version of Equation (2) that includes additional interactions between PD, the unemployment rate, and dummy variables capturing how many years since the loan was issued. As seen in Figure 3, the sensitivity of realized default increases over the life of the loan. In contrast, the cyclicality of the sensitivity of realized default, displayed in Figure 4 decreases by over 0.1 (compared to a base of 0.08) in the year after origination and remains lower after origination. The fact that bank information quality simultaneously improves and becomes less cyclical after origination is difficult to reconcile with theories in which variation in information quality over the business cycle is driven by purely exogenous factors.

To provide additional evidence that our results are driven by new loans, we leverage the fact that we can see different banks evaluating the same borrowers at the same time in the data. If banks are producing relatively more information about new borrowers in areas experiencing higher unemployment, we would expect their evaluations to converge for new loans, but not for existing loans. This would manifest as a reduction in the dispersion in banks' PD estimates of the same borrower for newly issued loans.[10] To test this hypothesis, we create a firm-level measure of PD dispersion based on the range of PD estimates across banks within firm and regress this measure on the unemployment rate. The results shown in Table 7 support our hypothesis. A one percentage point increase in the unemployment rate decreases the range of banks' PD evaluations by about 20 basis points for new loans, while PD dispersion for the full sample of loans does not appear to exhibit any statistically significant cyclicality. Because Columns (3) and (4) include only firms which receive multiple new loans from different banks in the same quarter, they are based on a much smaller sample than our main results. Even in this limited sample, however, our results are consistent with the idea that information production is more cyclically sensitive for new loans.

We next test whether banks' information quality is higher among loans for which their incentives to produce information are higher, i.e., loans that are more information sensitive. Several theories of endogenous information production such as Dang, Gorton, and Holmström (2012) predict that lenders will have higher incentives to produce information for loans in which lenders' potential losses are higher.[11] Intuitively, as lenders face larger potential losses, they gain more from learning more about the borrower's type.

If banks do indeed have more precise information about loans with higher potential losses, we would expect a positive coefficient estimate for the interaction between PD and the total dollar loss given default ($LGD), defined as the log of the product of loan size and loss given default (which is recorded as a ratio of the loan commitment). This

[10] See Brancati and Macchiavelli (2019) who show that analyst forecast dispersion regarding banks' ROAs became smaller during the 2008/2009 financial crisis.

[11] See also Manove, Padilla, and Pagano (2001).

measure represents the bank's estimate of the dollar value that would be lost *conditional on the borrower defaulting* and is calculated independently of PD. We estimate a modified version of Equation (1) once again using only loans at origination:

$$Default_i = \beta_0 PD_i + \beta_1 \$LGD + \beta_2 \left(PD_i \times \$LGD\right) + \Omega X_i + \delta_{b,t} + \gamma_{j,t} + \sigma_{b,c} + \epsilon_i. \quad (3)$$

The results are shown in Table 8. The first row shows the interaction between PD and \$LGD is positive and statistically significant and suggests that a one standard deviation increase in \$LGD (about 1.41) increases the sensitivity of realized default to PD by about 0.12, which amounts to just under half of the unconditional effect shown in Table 2. This result is consistent with banks producing more information about loans with larger potential losses.

We next examine how the sensitivity of information quality to \$LGD evolves over the business cycle. Dang, Gorton, and Holmström (2012) and Biswas (2022) show that lenders' incentives to produce information about loans are more sensitive to potential losses following negative aggregate shocks. We test this hypothesis by estimating a modified version of Equation (2) that also includes a triple interaction term between PD, the unemployment rate, and \$LGD. Table 9 shows that the interaction coefficient is positive, suggesting that bank information is more sensitive to potential losses in downturns.

Finally, we expect that the cyclicality of information production will be greater for industries whose cash flows are more sensitive to local economic conditions. We test this by comparing firms in tradable and nontradable industries. Because firms in nontradable industries are more likely to operate primarily in local markets, the same change in local economic conditions should have a larger effect on their underlying businesses, and as a result we would expect banks to produce more information about these firms as local conditions worsen. In Table 10, we test this prediction using regressions that interact PD and the local unemployment rate with dummy variables if the firm is a nontradable industry.[12] Consistent with our hypothesis, the cyclical sensitivity of PD to default is

[12] The list of nontradable industries includes utilities, construction, wholesale trade, retail trade, transportation, accommodation, food services, information and communication, and professional services.

only statistically significant for nontradable firms.

Because we cannot directly observe banks' information production decisions, we cannot entirely rule out the possibility that banks exogenously receive more precise information about their borrowers in bad times, as argued in Becker, Bos, and Roszbach (2020). However, it is difficult for this channel alone to jointly rationalize that the cyclicality of bank information production is driven by: i) new loans, ii) loans with higher potential losses, and iii) loans to firms in nontradable industries. Overall, we believe our results are consistent the framework of Dang, Gorton, and Holmström (2012) and other models in which the endogenous information production decisions of financiers vary over the business cycle and across different types of securities.

4 Implications and Discussion

Collectively, our evidence is consistent with banks endogenously producing more information in downturns because they have stronger incentives to do so. In this section, we discuss some implications of this mechanism. First, we analyze how local business cycles affect lending volume. Specifically, while the characteristics of new loans in a county do not meaningfully change as the local unemployment rate rises, the number and volume of new loans decline sharply. Next, we discuss these findings through the lens of several theoretical models that assume bank information is endogenously countercyclical. These models predict that policy interventions seeking to promote bank lending in response to a crisis are most effective if they can make loans more information insensitive; to the extent our empirical results confirm the underlying assumptions of these models, they provide further support for their conclusions.

4.1 Lending Outcomes Over the Business Cycle

In this section, we analyze the relationship between business cycles and lending outcomes. We first estimate the following regression across different outcome variables y_i:

$$y_i = \beta UR_{c,t} + \Omega X_i + \delta_{b,t} + \gamma_{j,t} + \sigma_{b,c} + \epsilon_i. \tag{4}$$

This regression includes the same firm-level characteristics in X_i and fixed effects that we use in our baseline specification; however, we exclude loan characteristics as controls and instead include them as dependent variables. The coefficient β reflects how each of these characteristics changes with the local unemployment rate. As in previous specifications, we cluster standard errors by county. Table 11 displays the results.

Loan amounts and loan maturities do not seem to vary over the business cycle in a statistically significant way. Moreoever, we find that interest rates and PDs are only marginally higher in bad times: a 1pp increase in the unemployment rate increases interest rates and PDs by about 3bps and 4bps (statistically significant at the 10% and 5% level, respectively). While the pool of potential borrowers is likely to be riskier in downturns, the pool of loans actually granted does not seem substantially riskier.

Instead, changes in lending behavior at the county level seem to be driven by the extensive margin of lending. We aggregate the number and total volume of loans to the county level, take logs, and then regress these measures on the county unemployment rate and county fixed effects. The results are reported in Table 12, which shows a decline in both the number and total volume of loans in a county as its economic conditions worsen. Specifically, a 1pp increase in the unemployment rate is associated with a 1.7% decrease in the number of loans and 7% decrease in total loan volume. Together, these results suggest that local downturns primarily affect the number and volume of loans banks issue, rather than the composition of loan types or borrowers.

4.2 Implications

The results in Section 4.1 are consistent with several theories of countercyclical information production in credit markets including Dang, Gorton, and Holmström (2012), Gorton and Ordonez (2014), Fishman, Parker, and Straub (2020), Farboodi and Kondor (2020), and Asriyan, Laeven, and Martin (2022). In these models, following a negative aggregate shock, the returns to distinguishing between different borrowers increase. In turn, banks produce more information about borrowers and lend to a smaller subset of higher-quality potential borrowers. In other words, while the average quality of *potential* borrowers drops in downturns, the average quality of those granted credit may not, which is consistent with what we show in Section 4.1.[13]

A common mechanism in these theories is that increased information production by lenders can exacerbate reductions in lending during downturns. Moreoever, they highlight some challenges standard policies may face when they interact with banks' information production incentives. One example of this is Fishman, Parker, and Straub (2020) who show that higher capital constraints can slow recoveries because banks' have a higher incentive to produce information following a negative shock when their capital constraints are binding. An implication of this model is that regulators could potentially speed up recoveries by either reducing capital requirements in bad times or recapitalizing banks (as emphasized in Holmstrom (2015)).

Policies that raise collateral values are another potential intervention that can offset the increased incentives of banks to produce information during downturns. As Manove, Padilla, and Pagano (2001) show, collateral can act as a substitute for screening because it reduces lenders' potential losses. Our results are consistent with this idea because banks' information quality is higher for loans with higher potential losses, and the cyclicality of banks' information quality is driven by loans with higher potential losses. In many cases, policymakers can directly target the value of the underlying collateral used in these loans, such as real estate as in Chaney, Sraer, and Thesmar (2012). If policies like asset

[13] Unfortunately, we cannot fully distinguish to what extent changes in loan volume over the cycle could also be due to changes in loan demand.

purchases boost the market value of collateral, banks would face lower potential losses and their incentives to produce information would be reduced, thereby leading to increased loan volume.

A more blunt tool would be for the government to fully guarantee loans, which in theory would completely deter banks from producing information. One recent example of this type of policy was the Paycheck Protection Program (PPP) implemented in 2020 following the COVID-19 pandemic. Past work such as Autor et al. (2022), Bartik et al. (2020), Granja et al. (2022), Joaquim and Wang (2022), and Marsh and Sharma (2021) finds generally positive effects of the PPP on employment, particularly in light of the massive uncertainty surrounding the economy at the time. These papers also emphasize that there is no "free lunch" when it comes to reducing information sensitivity, as the program also came at a relatively high cost and channeled funds to many firms which were either unwilling or unable to obtain more traditional financing arrangements. Nonetheless, their findings are consistent with the idea that limiting the incentives for banks to produce information about borrowers can be the fastest way to push funds to businesses through the banking system.

It is important to caveat that our analysis cannot speak to the welfare consequences of such policies. While information production can reduce total credit, it can also serve a valuable social function by improving the allocation of capital to firms. Nonetheless, many policy interventions during downturns explicitly focus on increasing lending. To the extent that our empirical results provide support for the fundamental mechanisms underlying the models discussed in this section, they suggest that interventions seeking to stimulate lending should focus on making loans less information-sensitive.

5 Conclusion

Information plays a crucial role in banks' lending decisions and in turn macroeconomic outcomes, but it is difficult to analyze empirically. In this paper, we construct a novel measure of bank information quality from confidential regulatory data containing banks' private risk assessments of their borrowers. Using county-level variation in unemployment rates, we find that information quality improves as local economic conditions worsen. We argue that these results are consistent with theories of endogenous information production by showing that our results are driven by newly originated loans and loans with higher potential losses. To our knowledge, our findings are the first in the empirical banking literature providing evidence of countercyclical information *production*. These findings have important implications for policymakers because banks' information production decisions affect the volume of credit granted to firms, and thus the efficacy of many policy tools may critically depend on economic conditions.

References

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.

Asea, Patrick K and Brock Blomberg, 1998, Lending cycles, *Journal of Econometrics* 83, 89–128.

Asriyan, Vladimir, Luc Laeven, and Alberto Martin, 2022, Collateral booms and information depletion, *The Review of Economic Studies* 89, 517–555.

Autor, David, David Cho, Leland D Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz, 2022, An evaluation of the paycheck protection program using administrative payroll microdata, *Journal of Public Economics* 211, 104664.

Bartik, Alexander W, Zoe B Cullen, Edward L Glaeser, Michael Luca, Christopher T Stanton, and Adi Sunderam, 2020, The targeting and impact of paycheck protection program loans to small businesses, Working paper, National Bureau of Economic Research.

Bassett, William F, Mary Beth Chosak, John C Driscoll, and Egon Zakrajšek, 2014, Changes in bank lending standards and the macroeconomy, *Journal of Monetary Economics* 62, 23–40.

Becker, Bo, Marieke Bos, and Kasper Roszbach, 2020, Bad times, good credit, *Journal of Money, Credit and Banking* 52, 107–142.

Bedayo, Mikel, Gabriel Jiménez, José-Luis Peydró, and Raquel Vegas Sánchez, 2020, Screening and loan origination time: lending standards, loan defaults and bank failures .

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

Benhabib, Jess, Xuewen Liu, and Pengfei Wang, 2016, Endogenous information acquisition and countercyclical uncertainty, *Journal of Economic Theory* 165, 601–642.

Berg, Tobias and Philipp Koziol, 2017, An analysis of the consistency of banks' internal ratings, *Journal of Banking & Finance* 78, 27–41.

Beyhaghi, Mehdi, Cesare Fracassi, and Gregory Weitzner, 2020, Bank loan markups and adverse selection, *Available at SSRN 3733932* .

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

Biswas, Sonny, 2022, Collateral and bank screening as complements: A spillover effect, *Working Paper* .

Boot, Arnoud WA and Anjan V Thakor, 1993, Security design, *The Journal of Finance* 48, 1349–1378.

Boyd, John H and Edward C Prescott, 1986, Financial intermediary-coalitions, *Journal of Economic theory* 38, 211–232.

Brancati, Emanuele and Marco Macchiavelli, 2019, The information sensitivity of debt in good and bad times, *Journal of Financial Economics* 133, 99–112.

Brown, James R, Matthew T Gustafson, and Ivan T Ivanov, 2021, Weathering cash flow shocks, *The Journal of Finance* 76, 1731–1772.

Cao, Qingqing, Marco Di Pietro, Sotirios Kokas, and Raoul Minetti, 2022, Liquidity and discipline. bank due diligence over the business cycle, *Journal of the European Economic Association* .

Cerqueiro, Geraldo, Steven Ongena, and Kasper Roszbach, 2016, Collateralization, bank loan rates, and monitoring, *The Journal of Finance* 71, 1295–1322.

Chaney, Thomas, David Sraer, and David Thesmar, 2012, The collateral channel: How real estate shocks affect corporate investment, *American Economic Review* 102, 2381–2409.

Chemla, Gilles and Christopher A Hennessy, 2014, Skin in the game and moral hazard, *The Journal of Finance* 69, 1597–1641.

Chiang, Yu-Ting, 2021, Strategic uncertainty over business cycles, *Mimeo* .

Coibion, Olivier and Yuriy Gorodnichenko, 2015, Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* 105, 2644–78.

Dang, Tri Vi, Gary Gorton, and Bengt Holmström, 2012, Ignorance, debt and financial crises, *Unpublished, Yale SOM* .

Dang, Tri Vi, Gary Gorton, and Bengt Holmström, 2013, The information sensitivity of a security, *Unpublished working paper, Yale University* 39–65.

Dang, Tri Vi, Gary B Gorton, and Bengt R Holmstrom, 2019, The information view of financial crises, Working paper, National Bureau of Economic Research.

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.

Dempsey, Kyle, Felicia Ionescu et al., Lending standards and consumption insurance over the business cycle, *2019 Meeting Papers*, 1428 (Society for Economic Dynamics 2019).

Diamond, Douglas W, 1984, Financial intermediation and delegated monitoring, *The review of economic studies* 51, 393–414.

Farboodi, Maryam and Peter Kondor, 2020, Rational sentiments and economic cycles, Working paper, National Bureau of Economic Research.

Fishman, Michael J, Jonathan A Parker, and Ludwig Straub, 2020, A dynamic theory of lending standards, Working paper, National Bureau of Economic Research.

Flynn, Joel P and Karthik Sastry, 2021, Attention cycles, *Mimeo* .

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

Fulghieri, Paolo and Dmitry Lukin, 2001, Information production, dilution costs, and optimal security design, *Journal of Financial Economics* 61, 3–42.

Gorton, Gary and Guillermo Ordonez, 2014, Collateral crises, *American Economic Review* 104, 343–78.

Gorton, Gary and Guillermo Ordonez, 2020, Good booms, bad booms, *Journal of the European Economic Association* 18, 618–665.

Gorton, Gary and George Pennacchi, 1990, Financial intermediaries and liquidity creation, *The Journal of Finance* 45, 49–71.

Gorton, Gary B and Ping He, 2008, Bank credit cycles, *The Review of Economic Studies* 75, 1181–1214.

Granja, João, Christos Makridis, Constantine Yannelis, and Eric Zwick, 2022, Did the paycheck protection program hit the target?, *Journal of financial economics* 145, 725–761.

Grunert, Jens, Lars Norden, and Martin Weber, 2005, The role of non-financial factors in internal credit ratings, *Journal of Banking & Finance* 29, 509–531.

Gustafson, Matthew, Ivan Ivanov, and Ralf R Meisenzahl, 2020, Bank monitoring: Evidence from syndicated loans, *Available at SSRN 2831455* .

Holmstrom, Bengt, 2015, Understanding the role of debt in the financial system .

Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo FP Luttmer, and Kelly Shue, 2016, Screening peers softly: Inferring the quality of small borrowers, *Management Science* 62, 1554–1577.

Jiménez, Gabriel, Steven Ongena, José-Luis Peydró, and Jesús Saurina, 2014, Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking?, *Econometrica* 82, 463–505.

Joaquim, Gustavo and J Christina Wang, 2022, What do 25 million records of small businesses say about the effects of the ppp? .

Keys, Benjamin J, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2010, Did securitization lead to lax screening? evidence from subprime loans, *The Quarterly journal of economics* 125, 307–362.

Keys, Benjamin J, Amit Seru, and Vikrant Vig, 2012, Lender screening and the role of securitization: evidence from prime and subprime mortgage markets, *The Review of Financial Studies* 25, 2071–2108.

Leland, Hayne E and David H Pyle, 1977, Informational asymmetries, financial structure, and financial intermediation, *The journal of Finance* 32, 371–387.

Lisowsky, Petro, Michael Minnis, and Andrew Sutherland, 2017, Economic growth and financial statement verification, *Journal of Accounting Research* 55, 745–794.

Lown, Cara and Donald P Morgan, 2006, The credit cycle and the business cycle: new findings using the loan officer opinion survey, *Journal of Money, Credit and Banking* 1575–1597.

Ma, Yueran, Teodora Paligorova, and José-Luis Peydro, 2021, Expectations and bank lending, *University of Chicago Working Paper* .

Maddaloni, Angela and José-Luis Peydró, 2011, Bank risk-taking, securitization, supervision, and low interest rates: Evidence from the euro-area and the us lending standards, *the review of financial studies* 24, 2121–2165.

Mäkinen, Taneli and Björn Ohl, 2015, Information acquisition and learning from prices over the business cycle, *Journal of Economic Theory* 158, 585–633.

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

Mariathasan, Mike and Sergey Zhuk, 2018, Attention allocation and counter-cyclical credit quality, *Available at SSRN 2933476* .

Marsh, W Blake and Padma Sharma, 2021, Government loan guarantees during a crisis: The effect of the ppp on bank lending and profitability, *Federal Reserve Bank of Kansas City Working Paper* .

Mester, Loretta J, Leonard I Nakamura, and Micheline Renault, 2007, Transactions accounts and loan monitoring, *The Review of Financial Studies* 20, 529–556.

Nakamura, Leonard I and Kasper Roszbach, 2018, Credit ratings, private information, and bank monitoring ability, *Journal of Financial Intermediation* 36, 58–73.

Ono, Arito and Iichiro Uesugi, 2009, Role of collateral and personal guarantees in relationship lending: Evidence from japan's sme loan market, *Journal of money, credit and banking* 41, 935–960.

Petriconi, Silvio, 2015, Bank competition, information choice and inefficient lending booms, *Information Choice and Inefficient Lending Booms (December 9, 2015)* .

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

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.

Rodano, Giacomo, Nicolas Serrano-Velarde, and Emanuele Tarantino, 2018, Lending standards over the credit cycle, *The Review of Financial Studies* 31, 2943–2982.

Ruckes, Martin, 2004, Bank competition and credit standards, *Review of Financial Studies* 17, 1073–1102.

Song, Wenting and Samuel Stern, 2021, Firm inattention and the transmission of monetary policy: A text-based approach, *Mimeo* .

Thakor, Anjan V, 1996, Capital requirements, monetary policy, and aggregate bank lending: theory and empirical evidence, *The Journal of Finance* 51, 279–324.

Treacy, William F and Mark Carey, 2000, Credit risk rating systems at large us banks, *Journal of Banking & Finance* 24, 167–201.

Weitzner, Gregory, 2019, Debt maturity and information production, *Working Paper* .

Yang, Ming, 2020, Optimality of debt under flexible information acquisition, *The Review of Economic Studies* 87, 487–536.

Yang, Ming and Yao Zeng, 2019, Financing entrepreneurial production: security design with flexible information acquisition, *The Review of Financial Studies* 32, 819–863.

6 Figures

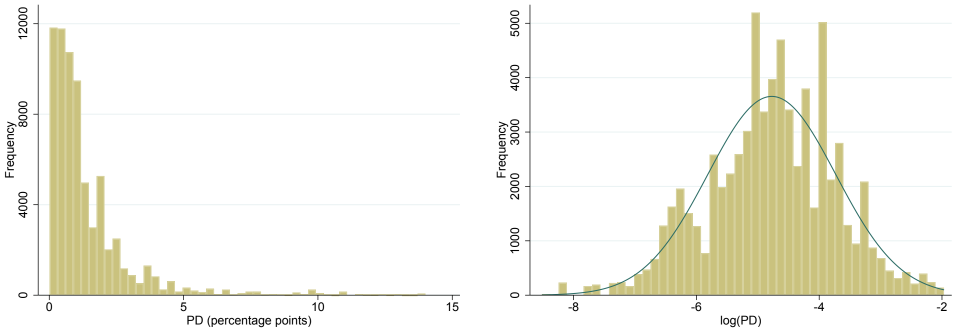

Figure 1: Frequency distributions of PD (left) and log(PD) (right)

This figure shows the frequency distributions of PD (left) and log(PD) (right) at origination for our sample. The sample construction is described in Section 2.

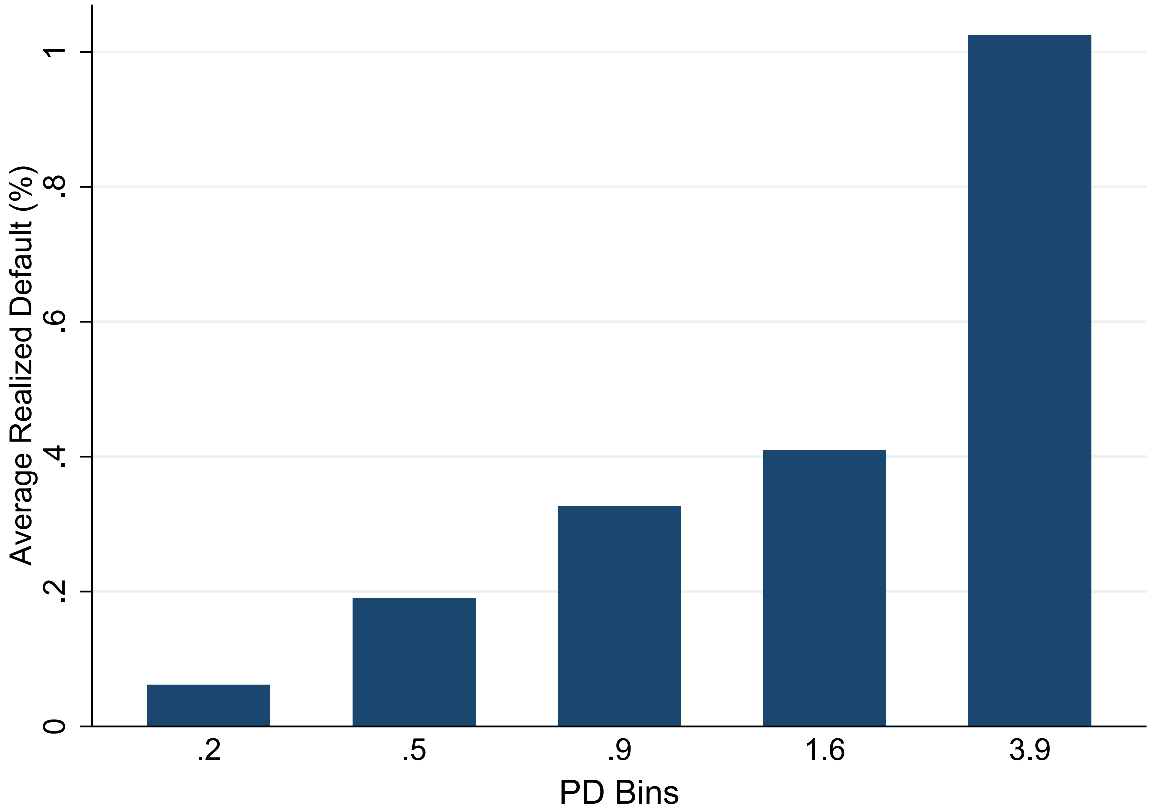

Figure 2: Realized Default Rates Across PD Quintiles

This figure shows default rates by quintiles of PD. The y-axis shows the realized default rate for each quintile while the numbers on the x-axis underneath each bar correspond to the average value included in the qunitile (rounded to the nearest 0.1pp). All variables are measured in percentage points.

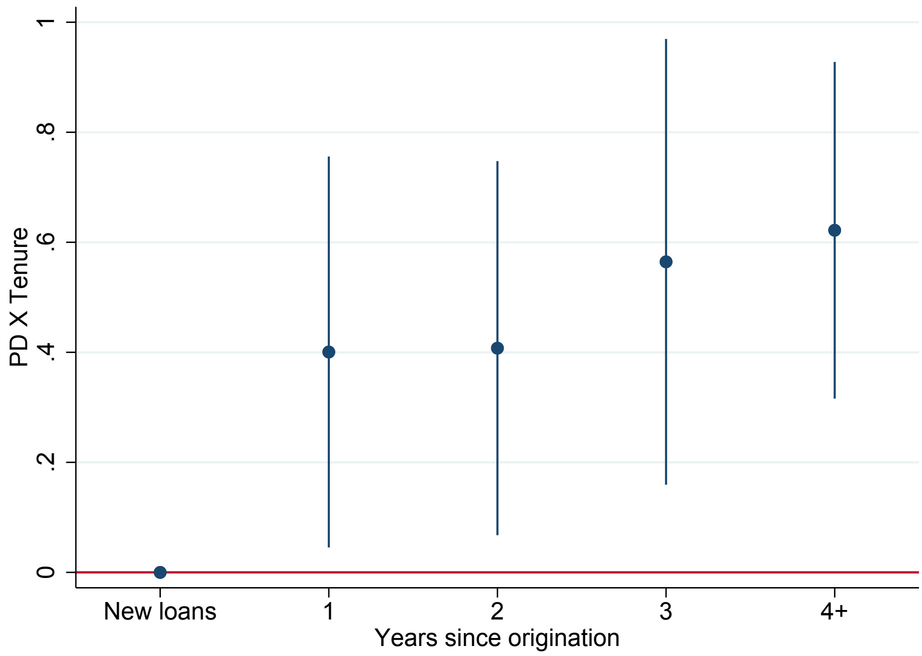

Figure 3: Information quality over the life of a loan

This figure shows estimates of the coefficient on PD x Tenure from a modified version of Equation 2 that includes an additional interaction between PD, UR, and dummy variables for the number of years since the loan was issued. PD is in levels and multiplied by 100. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within the subsequent eight quarters. Coefficient estimates show effects relative to the excluded group, which is comprised of new loans issued in that quarter. Vertical bars represent 95% confidence intervals from standard errors clustered by county.

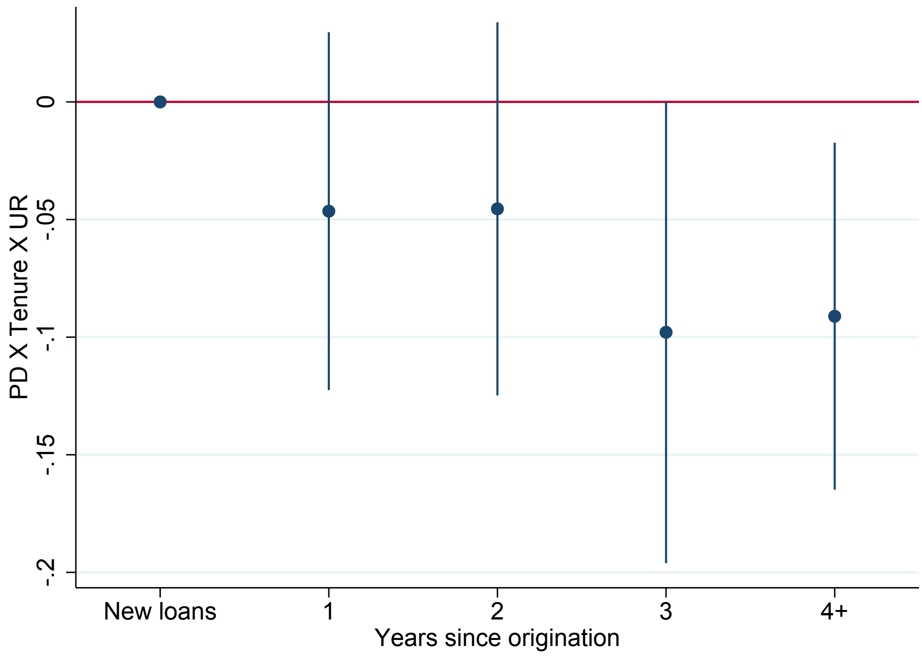

Figure 4: Information sensitivity over the life of a loan

This figure shows estimates of the coefficient on PD x Tenure x UR from a modified version of Equation 2 that includes an additional interaction between PD, UR, and dummy variables for the number of years since the loan was issued. PD is measured in percentage points and the dependent variable is $Default$, an indicator for whether each loan defaults within eight quarters after origination, multiplied by 100. Coefficient estimates show effects relative to the excluded group, which is comprised of new loans issued in that quarter. Vertical bars represent 95% confidence intervals from standard errors clustered by county.

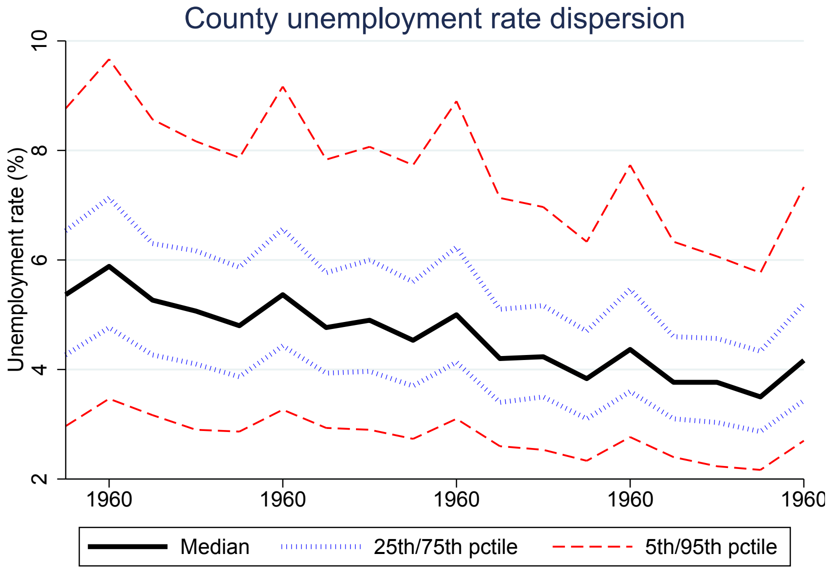

Figure 5: Unemployment rate dispersion over time

This figure displays the range of the county-level unemployment rates in our sample period for all county/quarter observations with at least one loan.

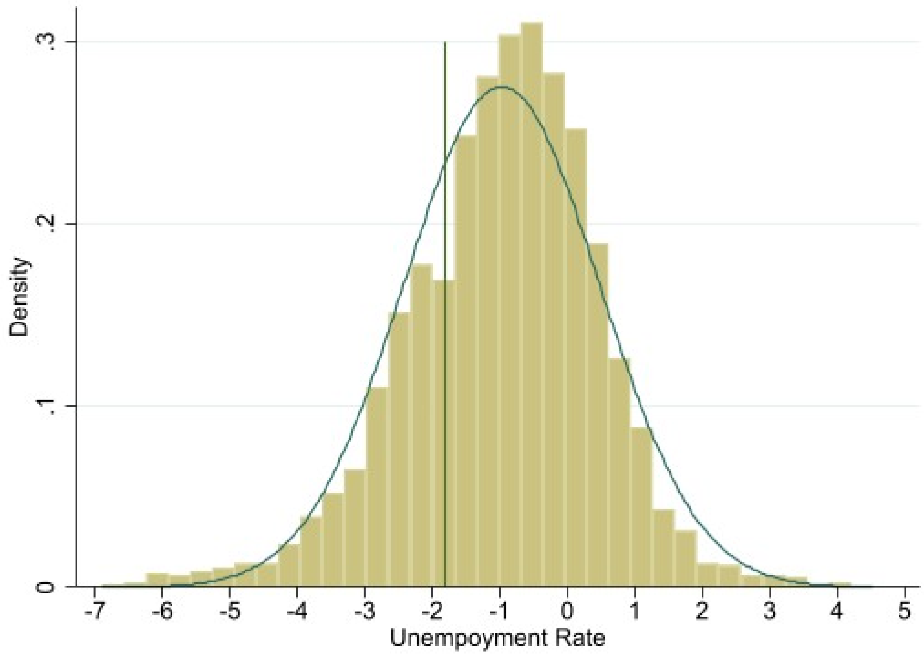

Figure 6: Changes in local unemployment rates across counties

This figure displays a histogram with a normal density curve of the changes in unemployment rates across all US counties from 2014Q4 - 2019Q1. The green vertical line is the change in the national unemployment rate over this period.

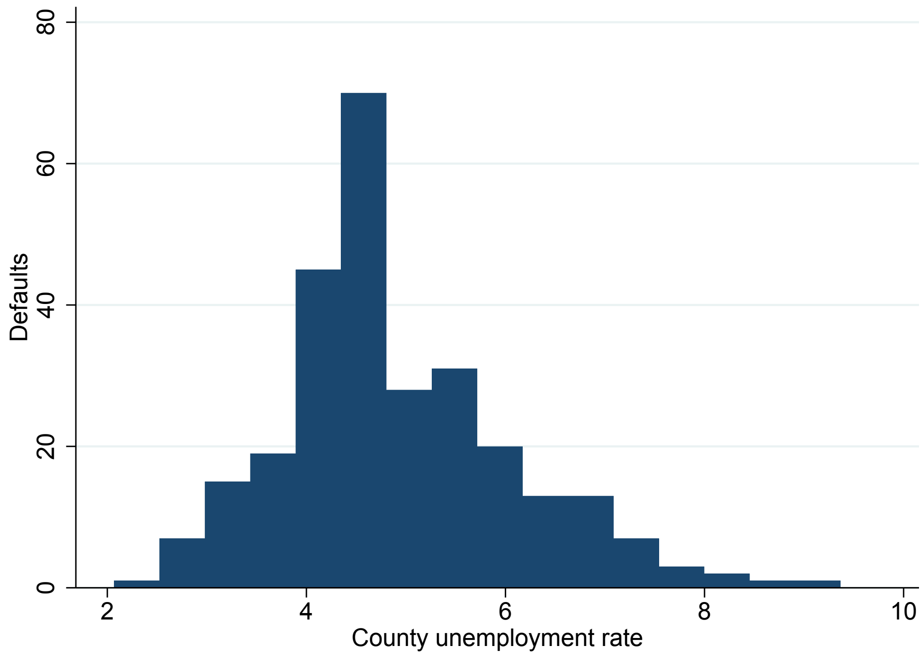

Figure 7: Default frequency across unemployment rates

This figure shows the frequency distribution of defaults within two years of origination in our sample based on the county-level unemployment rate at origination. For readability, the figure excludes a single default for a loan issued in a county with an unemployment rate of more than 15%.

7 Tables

Table 1: Summary Statistics

This table contains summary statistics for our sample. Panel A includes loan characteristics, Panel B firm characteristics and Panel C county characteristics. Section 2 describes our sample and Appendix A describes how the variables are constructed.

	Mean	Median	5%	95%	SD	N
Panel A: Loan Characteristics						
Interest rate (pp)	3.02	3.25	0.00	6.00	1.97	70,107
PD (pp)	1.40	0.91	0.15	4.34	1.66	70,107
LGD (ratio)	0.36	0.38	0.09	0.62	0.16	68,632
Realized default (pp)	0.40	0.00	0.00	0.00	6.27	70,107
Maturity (months)	47.05	58.00	7.00	88.00	30.59	70,107
Loan size ($ mil)	12.70	3.41	1.00	50.00	38.32	70,107
Revolver (indicator)	0.38	0.00	0.00	1.00	0.49	70,107
Term loan (indicator)	0.40	0.00	0.00	1.00	0.49	70,107
Floating rate (indicator)	0.55	1.00	0.00	1.00	0.50	70,107
Panel B: Firm Characteristics						
Sales ($ mil)	786.54	42.04	2.55	1,355.75	41,392.38	689,659
Assets ($ mil)	1,602.61	20.36	1.41	1,369.07	332,739.54	689,190
Leverage (ratio)	0.31	0.27	0.00	0.81	0.27	669,622
Profitability (ratio)	0.18	0.13	-0.03	0.56	0.23	680,177
Tangibility (ratio)	0.90	0.99	0.40	1.00	0.19	679,311
Nontradeable (indicator)	0.59	1.00	0.00	1.00	0.49	781,778
PD (pp)	2.75	0.88	0.14	9.82	9.13	726,998
Total number of loans	1.61	1.00	1.00	4.00	5.12	781,778
Number of new loans	0.09	0.00	0.00	1.00	0.46	781,778
Number of banks	1.16	1.00	1.00	2.00	0.70	781,778
Total loan volume ($ mil)	19.91	4.00	1.00	80.00	141.84	781,778
Panel C: County Characteristics						
Unemployment rate (pp)	4.88	4.57	2.63	8.00	1.88	34,328
Number of new loans	2.04	0.00	0.00	10.00	7.41	34,363
Number of total loans	36.70	6.00	1.00	162.00	122.43	34,363
Total new loan volume ($ mil)	452.91	41.30	1.50	2,020.14	1,962.85	34,363

Table 2: Predicting Default

This table shows the results of estimating Equation (1). The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination, multiplied by 100. Interest rates, interest rate spreads and probability of default (PD) are measured in percentage points. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Default		
	(1)	(2)	(3)
PD	0.243***	0.305***	0.300***
	(0.040)	(0.062)	(0.062)
Interest rate			3.553
			(3.102)
Controls	N	N	N
Bank-quarter FE	Y	Y	Y
Industry-quarter FE	Y	Y	Y
Bank-county FE	Y	Y	Y
Observations	66,821	52,416	52,416
R^2	0.195	0.216	0.217

Table 3: Predicting Default Over the Business Cycle

This table shows the results of estimating Equation 3 across separate samples depending on whether the unemployment rate in the county of issuance was above or below its median value for that county over the sample (2014Q4 through 2019Q1) at the time the loan was issued. The unemployment rate and PD are measured in percentage points. Bootstrapped t-statistics testing the differences in R^2 across high and low unemployment periods are shown in the row below each pair of values and are calculated based on 250 draws. Appendix A describes how the variables are constructed and Section 2 describes our sample. Standard errors are clustered at the county level and shown in parentheses.

	UR above median	UR below median	UR above median	UR below median	UR above median	UR below median
	(1)	(2)	(3)	(4)	(5)	(6)
PD			0.356***	0.127***	0.409***	0.202**
			(0.060)	(0.046)	(0.081)	(0.078)
Controls	Y	Y	N	N	Y	Y
Bank-quarter FE	Y	Y	N	N	Y	Y
Industry-quarter FE	Y	Y	N	N	Y	Y
Bank-county FE	Y	Y	N	N	Y	Y
Observations	25,539	25,055	35,253	34,576	25,539	25,055
Total R^2	0.269	0.310	0.007	0.002	0.274	0.312
Bootstrap t-statistic	-9.754		41.343		-9.129	
Within R^2	0.002	0.001			0.009	0.003
Bootstrap t-statistic	12.303				25.340	

Table 4: Information Quality Over the Business Cycle

This table shows coefficient estimates from Equation 2 with and without interactions between PD and the firm- and loan-level controls. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination, multiplied by 100. The unemployment rate (UR) and probability default (PD) are measured in percentage points. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Default			
	(1)	(2)	(3)	(4)
PD × UR	0.083***	0.103***	0.083***	0.085**
	(0.027)	(0.036)	(0.032)	(0.041)
PD	-0.130	-1.634**	-0.119	-1.268
	(0.118)	(0.797)	(0.138)	(0.853)
UR	-0.042	-0.070		
	(0.168)	(0.184)		
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	66,821	52,416	62,667	48,564
R^2	0.196	0.219	0.283	0.331

Table 5: Information Quality Using Predicted PD

This table shows the results of estimating a modified version of Equation (3) that uses a predicted value of PD generated from regressions including the same controls and fixed effects as in our default specification. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination, multiplied by 100. Both actual and predicted probability of default (PD) are measured in percentage points. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Default			
	(1)	(2)	(3)	(4)
Predicted PD × UR	0.177	0.100	0.135	0.065
	(0.150)	(0.158)	(0.154)	(0.166)
PD × UR		0.103***		0.087*
		(0.039)		(0.046)
Predicted PD	-0.550	-0.502	-0.364	-0.356
	(0.637)	(0.695)	(0.634)	(0.713)
PD		-0.158		-0.085
		(0.181)		(0.213)
UR	-0.166	-0.211		
	(0.306)	(0.311)		
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	52,416	52,416	48,564	48,564
R^2	0.212	0.217	0.326	0.330

Table 6: Information Quality Over the Business Cycle: New Versus Old Loans

This table tests whether the sensitivity of PD to realized default is higher for new loans when economic conditions deteriorate. The regression is estimated using a modified version of Equation 2 that also includes an triple interaction term between PD, the unemployment rate, and an indicator representing whether the loan was issued in that quarter. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination, multiplied by 100. The unemployment rate (UR) and probability default (PD) are measured in percentage points. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Default			
	(1)	(2)	(3)	(4)
PD × New loan	-0.500***	-0.565***	-0.466***	-0.526***
	(0.111)	(0.152)	(0.111)	(0.154)
PD × New loan × UR	0.056**	0.080**	0.046*	0.070*
	(0.028)	(0.037)	(0.027)	(0.036)
PD × UR	0.023*	0.024	0.030**	0.031**
	(0.012)	(0.015)	(0.013)	(0.015)
PD	0.480***	0.474***	0.461***	0.458***
	(0.047)	(0.056)	(0.048)	(0.057)
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	1,164,359	997,463	1,158,213	991,293
R^2	0.381	0.378	0.397	0.396

Table 7: PD Dispersion over the Business Cycle

This table shows how the range of PDs assigned to firms across all banks change over the business cycle. The dependent variable is the range between the highest and lowest PDs recorded across all banks in each quarter for a given firm and is measured in percentage points. This variable is regressed on the county-level unemployment rate and county fixed effects. Columns (2) and (4) also include the firm-level controls used in our baseline specification. Columns (1) and (2) use all loans and include only firms with multiple outstanding loans in a given quarter. Columns (3) and (4) use only new loans and include only firms with multiple new loans in each quarter. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	PD Dispersion			
	All loans		New loans	
	(1)	(2)	(3)	(4)
UR	-0.014	0.033	-0.209**	-0.189**
	(0.133)	(0.134)	(0.091)	(0.092)
Firm controls	N	Y	N	Y
County FE	Y	Y	Y	Y
Observations	53,805	51,828	1,666	1,577
R^2	0.090	0.112	0.115	0.138

Table 8: Information Quality and Potential Losses

This table tests whether \$LGD (defined as the log of the product of loan and loss given default) affects the sensitivity of realized default to PD (Equation 3). The dependent variable is a dummy variable indicating whether each loan defaults within eight quarters after origination, multiplied by 100. Probability of default (PD) is measured in percentage points. Firm size and loan size are measured in standard deviations of logs while leverage is a ratio. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Default	
	(1)	(2)
PD × \$LGD	0.087***	0.079***
	(0.026)	(0.027)
PD	-0.977***	-0.861**
	(0.366)	(0.383)
\$LGD	-0.025	-0.026
	(0.029)	(0.032)
Controls	Y	Y
Bank-quarter FE	Y	Y
Industry-quarter FE	Y	Y
Bank-county FE	Y	Y
County-quarter FE	N	Y
Observations	65,222	61,103
R^2	0.198	0.286

Table 9: Information Quality and Potential Losses Over the Business Cycle

This table tests whether the relationship between PD and \$LGD (defined as the log of the product of loan and loss given default) in predicting default intensifies when economic conditions deteriorate. The regression is estimated using a modified version of Equation 2 that also includes an triple interaction term between PD, the unemployment rate, and \$LGD. The dependent variable is a dummy variable indicating whether each loan defaults within eight quarters after origination, multiplied by 100. The unemployment rate (UR) and probability default (PD) are measured in percentage points. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Default	
	(1)	(2)
PD × UR × \$LGD	0.047***	0.061***
	(0.017)	(0.020)
PD × \$LGD	-0.117	-0.184**
	(0.075)	(0.083)
PD × UR	-0.572**	-0.771***
	(0.232)	(0.268)
PD	1.504	2.472**
	(1.046)	(1.172)
UR	-0.566	
	(0.415)	
Controls	Y	Y
Bank-quarter FE	Y	Y
Industry-quarter FE	Y	Y
Bank-county FE	Y	Y
County-quarter FE	N	Y
Observations	65,222	61,103
R^2	0.199	0.287

Table 10: Information Quality and Tradability

This table tests whether the cyclicality of information quality is concentrated in nontradable industries. Industry classification is based on two-digit NAICS codes; nontradables include firms in utilities, construction, wholesale trade, retail trade, transportation, accommodation, food services, information and communication, and professional services (NAICS codes 22-23, 42, 44-45, 48-49, 51, 54, and 72). Default is multiplied by 100 and the unemployment rate (UR) and probability default (PD) are measured in percentage points. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Default	
	(1)	(2)
PD × UR × Nontradeable	0.180**	0.209**
	(0.073)	(0.087)
PD × UR	0.001	-0.041
	(0.044)	(0.056)
PD × Nontradeable	-0.485	0.024
	(1.735)	(1.800)
PD	-1.315	-1.163
	(1.458)	(1.571)
UR	-0.008	
	(0.212)	
Controls	Y	Y
Bank-quarter FE	Y	Y
Industry-quarter FE	Y	Y
Bank-county FE	Y	Y
County-quarter FE	N	Y
Observations	52,416	48,564
R^2	0.221	0.334

Table 11: Loan Characteristics Over the Business Cycle

This table analyzes the relationship between the unemployment rate on loan characteristics. The dependent variable in each regression is shown at the top of each column. The unemployment rate (UR), PD, default, and interest rate are measured in percentage points. Maturity is measured in log months and loan size is measured in log dollars. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Loan size	Interest rate	Maturity	Default	PD
UR	-1.018	0.030*	1.284	0.077	0.036**
	(1.195)	(0.018)	(1.110)	(0.174)	(0.017)
Controls	Y	Y	Y	Y	Y
Bank-quarter FE	Y	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y	Y
Observations	53,690	53,693	53,677	53,693	53,693
R^2	0.547	0.625	0.429	0.211	0.358

Table 12: County-Level Lending Over Business Cycle

This table analyzes the relationship between the unemployment rate and county-level loan volume. Data are aggregated at the county level. The dependent variable in each regression is shown at the top of each column and both are in logs. The local unemployment rate (UR) and aggregate unemployment rate are measured in percentage points. Standard errors are clustered at the county level and are shown below the parameter estimates in parenthesis. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Loan Count	Loan Volume
	(1)	(2)
County UR	-0.017**	-0.070***
	(0.007)	(0.016)
County FE	Y	Y
Observations	12,173	12,173
R^2	0.779	0.637

Appendix A. Additional Data Description

A.1. Probability of Default Estimates

This section describes in more detail the probability of default (PD) estimates that we use in our analysis. The primary purposes of these estimates are stress testing and capital risk weight calculations. According to the Basel Committee on Banking Supervision, internal estimates of PD "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)." This instruction suggests that banks must report their best estimates of PD based on any information they have. Moreover, the instructions also state "PD estimates must be a long-run average of one-year default rates for borrowers in the grade'.[14] See Treacy and Carey (2000) for an excellent overview of how large US banks develop their internal ratings.

Banks have strong incentives to ensure that these estimates are accurate. Consistently underestimating default rates will attract regulatory scrutiny and can lead to additional restrictions on banks' activities.[15] Following supervisory exams, for example, a bank's models can be flagged by regulators as falling under Matters Requiring Attention (MRAs) or Matters Requiring Immediate Attention (MRIAs). While not as severe as other enforecment penalties, if left unresolved they can escalate into more severe penalties. Inadequate models can also be used by regulators as justification to force banks to recognize (or provision for) additional losses, which can lead to embarrassment and financial losses for the bank. Regulators can also prevent the banks which are unable to accurately model their losses from paying dividends.

Evaluating these estimates is complicated by the fact they are driven in part by other factors that may affect all other loans at the bank, county, or time level. For example, some counties might contain more small firms throughout our sample period; some banks may systematically focus on lending to less risky borrowers; and a nationwide recession means that all defaults may be higher in some quarters. These factors will all affect the average *level* of defaults for a county, bank, or quarter, respectively, without necessarily affecting the *relative* risk between loans within each of these groups. Given this issue, the models producing these default forecasts are often evaluated by both banks and regulators in relative (rather than absolute) terms. This aligns closely with our empirical approach where we use a rich set of fixed effects, which tells us *given the same loan, borrower and lender characteristics* whether loans that have higher PDs are more likely to default. Our

[14]In our main analysis we include a default horizon of two years to capture as many defaults as possible. Our results are robust to alternative default horizons and definitions.

[15]For more details regarding the regulatory approach to model evaluation, see the Federal Reserve Board's Bank Holding Company Supervision Manual (`https://www.federalreserve.gov/publications/files/bhc.pdf`) and the Basel II framework (`https://www.govinfo.gov/content/pkg/FR-2007-12-07/pdf/07-5729.pdf`).

approach will thus be unaffected by systematic misestimation of the *level* of the default rate.

Figure A.1 shows the dispersion of both PD and log(PD) over time in our sample. The left panel shows that the median values are quite stable. This is consistent with the instructions given to banks that ask for "through-the-cycle" estimates that ask for default forecasts averaged across a range of potential future business cycle conditions. This interpretation is supported by Ma, Paligorova, and Peydro (2021), who show that banks with more pessimistic forecasts of economic conditions do not necessarily have higher PDs. The rightmost columns show residuals from regressing each measure on the set of fixed effects used in our main specification. This suggests that our empirical approach, which we describe in the next section, is driven by cross-sectional variation in the ability of PD to predict default.

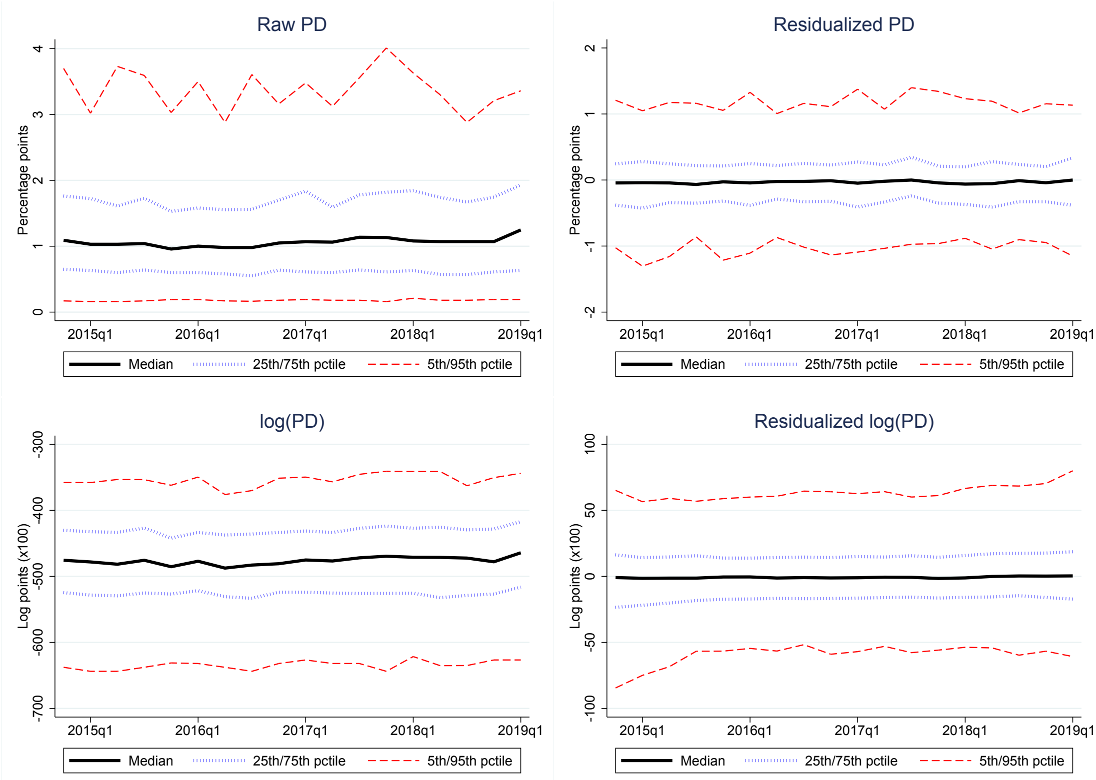

Figure A.1: PD dispersion over time

This figure shows the ranges of PD over our sample. The top row uses the standard PD, while the bottom row uses log(PD). The left column shows the raw series, while the right column shows the residuals after regressing each PD measure on bank-time, county-time, bank-county, industry-time, and loan type fixed effects.

A.2. Variable Definitions

Aggregate UR: United States national unemployment rate, from FRED.

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

Default: Dummy variable that equals one if the firm defaults within the first 8 quarters following the origination of the loan multiplied by 100, from Y-14Q.

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

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

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

LGD: The bank's estimated loss given default per dollar of debt in percentage points, from Y-14Q.

$LGD: LGD $\times$ Loan Size, from Y-14Q.

Loan Count: The number of new loans granted in the county in the quarter, from Y-14Q.

Loan Size: Log of committed amount of loan, from Y-14Q.

Loan Volume: The volume of new loans granted in the county in the quarter in logs, from Y-14Q.

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

New Loan: Dummy variable that equals one if the loan is newly originated in the quarter, from Y-14Q.

PD: The bank's expected long-run average default rate, trimmed if it equals zero or above the 99th percentile,from Y-14Q.

PD Dispersion: The range in PD estimates within firm/quarter across banks, from Y-14Q.

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

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

Nontradable: Dummy variable that equals one if the firm is in a nontradable industry. Nontradables industries include utilities, construction, wholesale trade, retail trade, transportation, accommodation, food services, information and communication, and professional services (NAICS codes 22-23, 42, 44-45, 48-49, 51, 54, and 72), from Y-14Q.

Tenure: Number of years elapsed since the loan was issued, from Y-14Q.

Total Debt: The sum of long-term debt and short-term debt, from Y-14Q.

UR: The county-level quarterly unemployment rate in percentage points from BLS.

Appendix B. Extensions and Robustness Checks

Table B.1: Predicting Default: PD Percentile Rank

This table tests whether PD predicts realized default beyond other loan and firm characteristics (Equation 1). The dependent variable in each regression is $Default$, an indicator for whether each loan defaults within eight quarters after origination, multiplied by 100. PD represents the percentile rank within a bank-quarter pair for each PD and takes values in $(0, 100]$. Section 2 describes our sample. Standard errors are clustered at the county level and shown in parentheses.

	Default		
	(1)	(2)	(3)
PD percentile	0.011***	0.012***	0.012***
	(0.001)	(0.002)	(0.002)
Interest rate			4.119
			(3.110)
Controls	N	N	N
Bank-quarter FE	Y	Y	Y
Industry-quarter FE	Y	Y	Y
Bank-county FE	Y	Y	Y
Observations	66,821	52,416	52,416
R^2	0.194	0.215	0.215

Table B.2: Predicting Default: Log(PD)

This table shows the results of estimating Equation 1. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination, multiplied by 100. PD is reported in logs and multiplied by 100. Interest rates and interest rate spreads are measured in percentage points. Appendix A describes how the variables are constructed and Section 2 describes our sample. Standard errors are clustered at the county level and shown in parentheses.

	Default		
	(1)	(2)	(3)
log(PD)	0.003***	0.004***	0.004***
	(0.000)	(0.001)	(0.001)
Interest rate			3.854
			(3.127)
Controls	N	N	N
Bank-quarter FE	Y	Y	Y
Industry-quarter FE	Y	Y	Y
Bank-county FE	Y	Y	Y
Observations	66,821	52,416	52,416
R^2	0.194	0.215	0.215

Table B.3: Alternate Loan Performance Measures

This table shows the results of estimating Equation 1 with alternative measures of loan performance. The dependent variable in each regression is a dummy variable corresponding to the column heading, multiplied by 100. "Any Default" measures whether a loan is recorded as defaulting at any point in our sample period. "Average Default" divides the "Any Default" measure by the number of years in which the loan is observed to generate an annual average; if a loan defaults within one quarter after origination, this variable will take on a value of 2, while if the loan defaults eight quarters after origination, this variable will take on a value of 0.5. "1Y Default" is an indicator for whether the loan defaults within four quarters of origination. "Delinquency" is an indicator for whether the loan is reported as delinquent within eight quarters after origination. "Chargeoff" is an indicator representing whether a bank records a chargeoff for that loan within eight quarters after origination. Probability of default (PD) is measured in percentage points. Standard errors are clustered at the county level and shown in parentheses.

	Any default	Average default	1Y default	Delinquency	Chargeoff
	(1)	(2)	(3)	(4)	(5)
PD	0.428***	0.294***	0.137***	0.099***	0.081***
	(0.075)	(0.062)	(0.041)	(0.024)	(0.024)
Controls	Y	Y	Y	Y	Y
Bank-quarter FE	Y	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y	Y
Observations	52,416	52,416	52,416	52,416	52,416
R^2	0.225	0.241	0.213	0.145	0.196

Table B.4: Information Quality Over the Business Cycle: PD Percentile Rank

This table tests whether the local unemployment rate affects the sensitivity of realized default to PD (Equation 2). The dependent variable in each regression is *Default*, an indicator for whether each loan defaults within eight quarters after origination expressed in percentage points. The unemployment rate is measured in percentage points. Probability of default (PD) is measured in percentage points. The "Control interactions" means that interaction terms between the controls and unemployment rate are included in the regression. Appendix A describes how the variables are constructed and Section 2 describes our sample. Standard errors are clustered at the county level and shown in parentheses.

	Default			
	(1)	(2)	(3)	(4)
PD percentile × UR	0.005***	0.006***	0.004***	0.005***
	(0.001)	(0.001)	(0.001)	(0.002)
PD percentile	-0.011***	-0.068***	-0.009*	-0.052*
	(0.004)	(0.026)	(0.005)	(0.032)
UR	-0.180	-0.228		
	(0.157)	(0.172)		
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	66,821	52,416	62,667	48,564
R^2	0.195	0.216	0.282	0.329

Table B.5: Information Quality over the Business Cycle: Log(PD)

This table shows coefficient estimates from Equation 2 with and without interactions between PD and the firm- and loan-level controls. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination expressed in percentage points. The unemployment rate is measured in percent. log(PD) is the log of PD multiplied by 100. Appendix A describes how the variables are constructed and Section 2 describes our sample. Standard errors are clustered at the county level and shown in parentheses.

	Default			
	(1)	(2)	(3)	(4)
log(PD) × UR	0.001***	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
log(PD)	-0.003**	-0.018**	-0.002	-0.014
	(0.001)	(0.008)	(0.001)	(0.009)
UR	0.726***	0.878***		
	(0.228)	(0.284)		
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	66,821	52,416	62,667	48,564
R^2	0.195	0.216	0.282	0.330

Table B.6: Information Quality over the Business Cycle Using Lagged Unemployment Rate

This table shows coefficient estimates from a modified version of Equation 2 that uses the unemployment rate lagged by one quarter with and without interactions between PD and the firm- and loan-level controls. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination expressed in percentage points. UR_{t-1} corresponds to the previous quarter's unemployment rate and is measured in percent. Probability of default (PD) is measured in percentage points. Appendix A describes how the variables are constructed and Section 2 describes our sample. Standard errors are clustered at the county level and shown in parentheses.

	Default			
	(1)	(2)	(3)	(4)
PD × Lagged UR	0.059**	0.062**	0.061**	0.053
	(0.024)	(0.029)	(0.029)	(0.036)
PD	-0.029	-1.467*	-0.032	-1.146
	(0.108)	(0.776)	(0.127)	(0.837)
Lagged UR	0.010	0.040		
	(0.163)	(0.182)		
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	66,821	52,416	62,667	48,564
R^2	0.196	0.218	0.283	0.331

Table B.7: Information quality excluding oil and gas firms

This table shows estimates from Equation 2 that exclude all loans to companies in mining, quarrying, and oil and gas extraction (NAICS sector 21). The dependent variable in each regression is an indicator for whether each loan defaults within eight quarters, multiplied by 100. Probability of default (PD) is measured in percentage points. Standard errors are clustered at the county level and shown in parentheses.

	Default			
	(1)	(2)	(3)	(4)
PD × UR	0.076***	0.101***	0.081**	0.091**
	(0.027)	(0.035)	(0.032)	(0.043)
PD	-0.152	-0.668	-0.160	-0.325
	(0.117)	(0.765)	(0.142)	(0.815)
UR	-0.129	-0.136		
	(0.105)	(0.102)		
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	64,552	50,884	60,460	47,093
R^2	0.190	0.210	0.278	0.325

Table B.8: Information quality excluding high-PD loans

This table shows estimates from Equation 2 that exclude all loans with PDs above the 90th percentile in a given bank-quarter. The dependent variable in each regression is an indicator for whether each loan defaults within eight quarters, multiplied by 100. Probability of default (PD) is measured in percentage points. Standard errors are clustered at the county level and shown in parentheses.

	Default			
	(1)	(2)	(3)	(4)
PD × UR	0.223***	0.226***	0.223***	0.232***
	(0.064)	(0.069)	(0.074)	(0.083)
PD	-0.597**	-1.440	-0.570*	-1.586
	(0.256)	(1.224)	(0.298)	(1.393)
UR	-0.210	-0.194		
	(0.164)	(0.171)		
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	60,223	47,555	56,276	43,849
R^2	0.191	0.209	0.269	0.305

Table B.9: Information quality excluding syndicated loans

This table shows estimates from Equation 2 that exclude all syndicated loans. The dependent variable in each regression is an indicator for whether each loan defaults within eight quarters, multiplied by 100. Probability of default (PD) is measured in percentage points. Standard errors are clustered at the county level and shown in parentheses.

	Default			
	(1)	(2)	(3)	(4)
PD × UR	0.073***	0.070*	0.068**	0.053
	(0.028)	(0.036)	(0.032)	(0.040)
PD	-0.227**	-0.691	-0.209	-0.150
	(0.111)	(0.934)	(0.130)	(1.017)
UR	-0.147	-0.133		
	(0.119)	(0.111)		
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	44,765	34,355	40,538	30,462
R^2	0.215	0.239	0.325	0.375

Table B.10: Information quality and lending relationships over the business cycle

This table shows results from a modified version of Equation 2 that includes an additional interaction between PD and a "New match" indicator. This indicator is equal to 1 for the first loan that we observe between a firm and bank, and zero otherwise. Probability of default (PD) is measured in percentage points. Standard errors are clustered at the county level and shown in parentheses.

	Default			
	(1)	(2)	(3)	(4)
PD	-0.175	3.152*	-0.197	5.669**
	(0.179)	(1.901)	(0.199)	(2.264)
PD X New match	0.110	0.308	0.169	0.394
	(0.172)	(0.226)	(0.202)	(0.285)
PD X UR	0.107**	-0.945**	0.112**	-1.450***
	(0.043)	(0.428)	(0.048)	(0.520)
PD X UR X New match	-0.052	-0.103*	-0.061	-0.118
	(0.044)	(0.058)	(0.052)	(0.073)
Controls	N	Y	N	Y
Bank-quarter FE	Y	Y	Y	Y
Industry-quarter FE	Y	Y	Y	Y
Bank-county FE	Y	Y	Y	Y
County-quarter FE	N	N	Y	Y
Observations	66,821	52,416	62,667	48,564
R^2	0.196	0.221	0.283	0.334

Table B.11: Information Quality over the Business Cycle Using Aggregate US Unemployment Rate

This table shows coefficient estimates from a modified version of Equation 2 that uses the total US unemployment rate. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination expressed in percentage points. UR_t corresponds to the total US unemployment rate and is measured in percent. Probability of default (PD) is measured in percentage points. Appendix A describes how the variables are constructed and Section 2 describes our sample. Standard errors are clustered at the county level and shown in parentheses.

	Default	
	(1)	(2)
PD × Aggregate UR	0.297***	0.300***
	(0.060)	(0.059)
PD	-1.135***	-2.552***
	(0.256)	(0.872)
Aggregate UR	0.121	0.145
	(0.126)	(0.133)
Controls	N	Y
Bank-quarter FE	N	N
Industry-quarter FE	N	N
Bank-county FE	Y	Y
County-quarter FE	N	N
Observations	66,983	52,500
R^2	0.154	0.171

Table B.12: Unemployment variation within counties

This table shows measures of variation in the unemployment rate within counties that had at least one outstanding loan over our sample period (2014Q4 through 2019Q1). Columns under the "Range" heading show the difference between the highest and lowest unemployment rates observed within each county. Columns under the "Standard deviation" heading show the standard deviation for the unemployment rate within each county. Results are split based on how many times each county showed up in the data during the sample period: the "$\geq 2/4$" columns show results for all counties that had outstanding loans in least two/four quarters, respectively, while the "All" column restricts the results to only counties which had at least one observation in every quarter throughout the sample. The last row shows the number of counties used in each calculation.

	Range			**Standard deviation**		
County-quarters observed	≥ 2	≥ 4	All	≥ 2	≥ 4	All
5th percentile	0.37	0.83	1.13	0.19	0.27	0.29
25th percentile	1.17	1.50	1.70	0.44	0.48	0.46
Median	1.87	2.07	2.25	0.66	0.67	0.6
Mean	1.97	2.20	2.26	0.74	0.72	0.63
75th percentile	2.53	2.67	2.80	0.90	0.87	0.78
95th percentile	5.93	3.90	3.47	1.51	1.32	0.99
Number of counties	1,417	1,035	186	1,417	1,035	186

Appendix C. Simple Theoretical Framework

In this section we present a simple model that highlights how the business cycle can affect bank information production incentives.

There is a single borrower seeking funds from a bank at $t = 0$ for a project that pays off at $t = 1$. The borrower and bank are risk neutral and there is no discounting. There are two types of borrowers $\theta \in \{G, B\}$ (Good, Bad) where θ is initially unknown to all and the prior probability of the borrower being good is λ.[16] The borrower has an investment opportunity that requires an initial investment of I at $t = 0$ and yields a cash flow at $t = 1$ of $R > I$ with probability π_θ and 0 otherwise where $\pi_G > \pi_B$. Although the borrower's type θ is initially unknown, the bank can pay a cost $c > 0$ to learn θ before committing funds at $t = 0$. The borrower offers the bank a loan contract that raises I at $t = 0$ and promises to repay D at $t = 1$. To simplify the analysis, we take the terms of the contract, i.e., D, as given.[17]

We assume the average project is NPV positive, i.e., $(\lambda\pi_G + (1-\lambda)\pi_B)R > I$, while the bad project is NPV negative, i.e., $\pi_B R < I$. Moreoever, we make the following assumptions so that the bank's participation constraint always holds

$$\lambda\pi_G D + (1-\lambda)\pi_B D - I \geq 0 \tag{5}$$
$$\lambda(\pi_G D - I) \geq c \tag{6}$$

The bank then decides whether to produce information based on the following inequality

$$\lambda(\pi_G D - I) - c \geq \lambda\pi_G D + (1-\lambda)\pi_B D - I, \quad \Longrightarrow \underbrace{(1-\lambda)(I - \pi_B D)}_{\text{Value of Information}} \geq c. \tag{7}$$

Intuitively, (7) says that the bank's profits from producing information and only financing the good borrower must be higher than the profits from not producing information and financing the borrower regardless of its type. We interpret a recession as either a decrease in the probability of the project being good λ or a decrease in the expected cash flow of bad borrowers, i.e., a decrease in π_B. For both of these cases, the value of information in (7) increases, thereby increasing the incentives of the bank to produce information.

[16] The borrower can know its type and the results would not change as there is no potential for signaling and the borrower's outside option is zero so there is no adverse selection problem on the borrower side.

[17] This allows us to abstract away from the bargaining process. See Dang, Gorton, and Holmström (2012) and Weitzner (2019) for cases in which the face value of debt is endogenous.