Bank Information Production Over the Business Cycle*

Cooper Howes† Gregory Weitzner‡

September 2021

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 loan-level risk assessments for US corporate loans. We show that our measure of information quality improves as local economic conditions deteriorate, particularly among loans with greater information sensitivity. To alleviate endogeneity concerns, we also use unexpected snowfall as an exogenous shock to local economic conditions. Taken together, our results provide support for theories in which economic conditions and security design decisions drive information production in credit markets. Our findings also suggest that policies designed to stimulate macroeconomic activity through the banking sector may be less effective in recessions.

*We thank Hassan Afrouzi, Olivier Coibion, Stefan Jacewitz, Blake Marsh, Johannes Matschke, Adolfo De Motta, Guillaume Roussellet, Garrett Schaller, Padma Sharma, Lee Smith, and Choongryul Yang as well as seminar participants at the Federal Reserve Bank of Kansas City for the helpful comments and discussions. These views are those of the authors and do not reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

†Federal Reserve Bank of Kansas City. Email: cooper.howes@kc.frb.org.

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

1 Introduction

A key role of banks is to produce information about prospective borrowers (e.g., Leland and Pyle (1977) and Diamond (1984)). Because banks' information influences both the recipients and terms of financing, their information production decisions can affect real economic activity and financial stability through the supply of credit to firms. Moreoever, banks' incentives to produce information may vary over the business cycle. Despite policymaker interest and an extensive theoretical literature[1] emphasizing the importance of bank information, there is little evidence of its empirical properties.

In this paper, we create a new measure of bank information quality using confidential loan-level regulatory data for corporate bank loans in the US. We then use this measure to examine the cyclicality of banks' information quality. Exploiting variation in county-level unemployment rates and shocks to local economic conditions in the form of abnormally severe snowfall, we find that banks' information quality improves when economic conditions deteriorate. We also find that banks have higher quality information about loans that theory predicts to be more information sensitive (e.g., Dang, Gorton, and Holmström (2012)), and that this relationship intensifies during periods of high unemployment. Overall, our results provide empirical support for theories in which banks' information production decisions hinge on aggregate economic conditions and security design considerations. Our results can also provide insights to policymakers and regulators about how banks respond to policy interventions meant to stimulate lending, as these types of policies may be less effective in recessions when banks screen their borrowers more intensively.

The main empirical challenge to testing theories of information production is that lenders' information is generally private and unobservable. We overcome this hurdle by using loan-level risk assessments from Y-14Q regulatory filings that banks report to the Federal Reserve. This dataset includes the universe of corporate loans over $1 million

[1]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).

on the balance sheets of US bank holding companies (BHCs) with at least $50 billion in total assets. In addition to detailed loan and borrower characteristics, qualified BHCs are required to report their internal estimate of the loan's probability of default (PD).

We first establish that banks' PDs are statistically and economically significant predictors of realized default even after controlling for a rich set of loan and firm-level controls, suggesting that banks' risk assessments contain private information that is both: 1) relevant for predicting default, and 2) not captured by other observables.[2] We create a measure of information quality based on the size of the PD coefficient in OLS regressions predicting realized default. Next, we use this approach to evaluate how information quality varies over the business cycle.

Several theories predict that banks produce more information when economic conditions weaken as the returns to distinguishing between different borrower types increase[3]. Consistent with countercyclical information production, we find that a one percentage point increase in the local unemployment rate increases the sensitivity of realized default to PD by roughly 50% of its average level. We provide evidence that this improved information quality is a response to, rather than the cause of, economic downturns by showing that a measure of abnormally severe snowfall, developed in Brown, Gustafson, and Ivanov (2021), leads to both higher unemployment rates and better information quality. Hence, exogenous shocks to economic conditions lead to banks' having higher quality information regarding borrowers.

The richness of our data allows us to rule out several other alternative explanations for our empirical findings. Because we compare loans given by the same bank in the same quarter in counties with different unemployment rates, our results cannot be due to differences in costs of capital either across banks or over time. By the same reasoning, our results are not driven by US-wide aggregate supply or demand effects.

There are multiple channels that can cause banks' information quality to vary over

[2]This result is consistent with the literature using the Y-14Q data to analyze the predictability of banks' PDs for default (Adelino, Ivanov, and Smolyansky (2019) and Beyhaghi, Fracassi, and Weitzner (2020).

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

the business cycle. One possibility is that banks simply receive information which is exogenously more precise during bad times. Alternatively, banks may choose to produce more information about borrowers during downturns. To attempt to distinguish between these channels, we first analyze how bank information quality varies across different proxies for the information sensitivity of a loan (Dang, Gorton, and Holmström (2013)). Specifically, we hypothesize that banks will have higher quality information regarding larger loans. Intuitively, as long as there is a fixed cost to information production, banks will have more incentives to produce information about larger loans.[4] We also hypothesize that higher leverage will induce higher information quality as the higher a firm's leverage, the less banks can expect to recover in default per unit of debt, thereby raising their incentives to learn the borrower's quality.[5] Finally, we expect the information asymmetry problem to be more severe for smaller firms (e.g., Chae (2005)), which incentivizes more information production by banks. To test these hypotheses, we estimate regressions that include interactions between PD and these characteristics. Consistent with our hypotheses, we find positive interaction coefficients with loan size and leverage, and a negative interaction coefficient with firm size. These results suggest that banks produce more information about loans with greater information sensitivity.

Next, we analyze how the relationship between banks' information and loan and firm characteristics varies over the business cycle. Dang, Gorton, and Holmström (2012) show that information production decisions become more sensitive to the size of the loan following negative aggregate shocks. If information frictions are relatively more severe for small and highly levered firms in bad times, we would expect the relationship between firm characteristics and information quality to be amplified in bad times. Consistent with these predictions, we find that larger loans, higher leverage, and smaller firm size all amplify the effects of unemployment on information quality. These results provide additional support for the information production channel, and to our knowledge, are the first in the literature highlighting how the business cycle can affect the information

[4] E.g., see Dang, Gorton, and Holmström (2013) and Weitzner (2019).

[5] This is a direct implication of Dang, Gorton, and Holmström (2013) who show that lower expected recovery values increase lenders' information production incentives.

sensitivity of loans.

Our findings have important consequences for policymakers because a bank's lending response to policy interventions may vary depending on the information production incentives that bank faces. One policy-relevant example is that programs designed to stimulate bank lending, including monetary policy, may become less effective in downturns. During good times, banks may have lower incentives to screen borrowers, and as a result more firms will receive credit. However, during downturns, banks will screen their potential borrowers more thoroughly, leading to a smaller but higher quality pool of loan recipients. This can help explain why some research has found modest effects of stimulus policies enacted in the wake of the Great Recession, particularly for the riskiest firms.[6] Ultimately, the empirical properties of bank information production are crucial for understanding the link between bank lending and real economic activity, as well as how this link changes over the business cycle. We view this paper as an important step in analyzing these properties.

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 by directly using banks' information they report to the Federal Reserve.

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 2008/2009 financial crisis. Our paper also relates to the body of empirical work analyzing bank internal risk-measures

[6]For example, Andrade et al. (2019) analyze the bank lending response to the Eurosystem's LTRO program and find an increase in overall lending but no increase in lending to riskier firms. In the case of QE, Butt et al. (2014) find no evidence that it stimulated bank lending in the UK.

(e.g., Agarwal and Hauswald (2010), Qian, Strahan, and Yang (2015), Behn, Haselmann, and Vig (2016), Dell'Ariccia, Laeven, and Suarez (2017), Plosser and Santos (2018), Nakamura and Roszbach (2018), Becker, Bos, and Roszbach (2020), Adelino, Ivanov, and Smolyansky (2019) and Beyhaghi, Fracassi, and Weitzner (2020)).

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 can compare information quality across loans in different regions given by the same bank. Second, their data are at the firm level rather than the loan level. This difference allows us to explore the relationship between loan/firm characteristics and information production and how the intensity of this relationship varies 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), Dempsey, Ionescu et al. (2019)). For the most part, existing studies find that lending standards are countercyclical. Our results compliment this finding by suggesting banks are producing more information when economic conditions deteriorate. While we do not observe the composition of loan applicants and thus cannot directly analyze lending standards, the theoretical channels we use to interpret our results are consistent with countercyclical lending standards.

Finally, our paper relates to work analyzing the cyclicality of information production. This includes an extensive theoretical literature in which information production is countercyclical in credit markets[7], including 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)). Moreover, our results are also consistent with several 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 Holmstrom (2009), Gorton and Ordonez (2014), Dang, Gorton, and Holmstrom (2019), Yang and Zeng (2019), Yang (2020), and Weitzner (2019).

The paper is structured as follows. Section 2 describes our data. Section 3 presents empirical evidence that information quality is countercyclical as well as evidence on potential mechanisms. Section 4 discusses implications for policymakers. Section 5 concludes.

2 Data

Our main source of data is Schedule H.1 of the Federal Reserve's Y-14Q data. The Federal Reserve began collecting this data in June 2012 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, accounting for 85.9% of all assets in the banking sector (Frame, McLemore, and Mihov (2020)).[8] Qualified banks are required to report detailed quarterly loan level data on corporate loans that exceed $1mm in size. The universe of loans we analyze is large: Bidder, Krainer, and Shapiro (2020) show that the Y-14Q data makes up 70% of all commercial and industrial loan volume.

The data include detailed loan characteristics (such as interest rate, maturity, amount, collateral, credit guarantee, purpose) and performance measures (past due payments, non-accruals, charge-offs). They also include income, balance sheet, and geographic information about borrowers. Crucially, banks are also required to report their internal

[7]It also complements research in non-financial settings, including Coibion and Gorodnichenko (2015), who find empirical evidence consistent with countercyclical information quality in forecasts of macroeconomic aggregates such as inflation.

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

estimates of probability of default (PD) for each loan to the Federal Reserve on their Y-14Q filings.

We restrict or sample to include only US borrowers. Because we are focused on banks' information production incentives at loan origination, 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, 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 as these observations are likely reporting errors. We 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 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, and extends through 2019Q1 to allow at least one year for loans to default (the data are available through 2020Q1). The full details of the sample construction as well as the sources, purpose, and properties of the PD estimates are described in Appendix A.

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.

Table 1 includes summary statistics for these measures. The average and median loan size is approximately $12.9mm and $3.5mm, respectively, and over 90% of loans are less than \$30mm. The median firm has \$47.4mm in assets and a leverage ratio of 0.31. These loan and firm sizes are small relative to other sources of loan data such as DealScan, reflecting the fact that our sample contains many more small and nonpublic firms. The loan sample is approximately evenly split among credit lines and term loans and the median interest rate is 3.25%. The rightmost column of Table 1 shows summary statistics for the PD estimates in our sample. Over our sample period, 0.41% of firms default within the first two years after loan origination. This compares to an average ex-ante expected PD of 1.47%, suggesting that economic conditions were relatively benign during this period relative to banks' expectations.

Overall, this dataset is uniquely suited to analyze the dynamics of banks' information quality over the business cycle. First, as mentioned earlier, it covers the vast majority of commercial bank loans in the US. Second, it includes both realized and predicted default at the loan level, which allows us to create an empirical measure of bank information precision 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, than there should be a positive correlation between predicted default and actual default. A simple graphical illustration of this is shown in Figure 2. The left panel of Figure 3 displays the relationship between PD and realized default in our data. Each column corresponds to a PD quintile, while the vertical axis represents the average default rate for loans in that bucket. This panel shows a clear relationship: higher PDs correspond to higher realized defaults. For comparison, the right panel shows a similar breakdown, but with quintiles of interest rates instead of PDs. If default risk were the only factor driving differences across interest rates, these columns should increase monotonically. Instead, the relationship between interest rates and realized default is far weaker. This suggests that PD has useful information for predicting default beyond what is reflected in interest rates. In the next section, we implement this approach to measure bank information quality in the data.

3 Empirical Results

3.1 Predicting Default

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

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

where i, t, b, c, and j index loan, quarter, bank, county, and industry, respectively. D_i is a dummy variable that equals 1 if loan i defaults within eight quarters following origination. PD_i is the percentile rank (scaled to $[0, 100]$) within each bank and quarter of the PD for loan i at origination. We use this measure rather than the raw PD level for several reasons. First, the distribution of PD has a very large number of very small observations with a long right tail, leading to a potential outsized influence of outliers (see Figure 1). Second, ranking each PD within a given bank-quarter also ensures that all comparisons are between loans issued by the same group of bank decisionmakers. Third, the ranked approach avoids potential biases in the level of PD that might change over the business cycle. Finally, focusing on the ordinal ranking of PDs rather than their level is closer conceptually to the tools used by regulators to evaluate bank default models. As a robustness check, we show very similar results using log(PD) in Appendix C.

X_i is a vector of firm and loan characteristics which include 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 loan's interest rate, the bank's estimate of loss given default (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 of our 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 unit increase in a loan's PD percentile rank. For example, in Column (1) the coefficient estimate is 0.0117, which means that, holding all else equal, going from the 25th percentile of PD to the 75th percentile will increase the probability of realized default by about $0.0117 \times (75 - 25) = 0.585$ percentage points. This effect is both statistically and economically significant considering the unconditional default rate in our sample is 0.4%. In Column (2), we display the results with firm and loan characteristics, and in Column (3), we include the loan interest rate as an additional regressor. Both of these specifications show similar results, suggesting that PD reflect information which is useful for predicting default even after controlling for the interest rate and other observable characteristics.[9]

Henceforth, we treat our estimate of β, the sensitivity of realized default to PD, as our measure of information quality. In the next section, we test how this measure of information quality varies over the business cycle.

3.2 Information Quality Over the Business Cycle

3.2.1 Local Unemployment Rate and Information Quality

In this section, we test the cyclicality of bank information quality. Our measure of county-level economic conditions is the unemployment rate from the BLS.[10] We test whether the sensitivity of realized default to PD varies over the business cycle by estimating the following regression:

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

[9] These results are consistent with Adelino, Ivanov, and Smolyansky (2019) and Beyhaghi, Fracassi, and Weitzner (2020) who also show PDs predict loan performance even after controlling for interest rates.

[10] 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 Development to assign a county to each zip code before merging it with the unemployment rate data.

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.[11] 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 3. 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 default sensitivity by about 0.5 basis points; this is statistically significant and represents about 45% of the average effect of PD estimated in Section 3.1. Column (2) adds interactions between the employment rate and other firm and loan controls and shows that the estimates become slightly larger in magnitude and remain statistically significant. Columns (3) and (4) add county-quarter fixed effects, which absorb level differences in activity across different counties in each quarter, and show very similar coefficient estimates. Overall, these results suggest that increases in unemployment have a statistically and economically significant relationship with improvements in bank information quality. Hence, we argue that bank information quality is countercyclical.

On its own, our approach cannot tell us whether higher unemployment rates cause banks to have higher information quality. This is an important distinction because we would ideally test whether information production is an endogenous response to changes in economic conditions. For example, if improvements in bank information quality lead to higher screening, reduced loan volumes, and thus higher unemployment, then our evidence cannot be interpreted as evidence for these theories. To address this issue, we next exploit exogenous variation in local economic conditions caused by severe weather.

[11] Appendix C shows very similar results using the lagged, rather than contemporaneous, unemployment rate.

3.2.2 Abnormal Snowfall and Information Quality

In this section, we exploit abnormal snowfall as an exogenous shock to economic conditions. To construct the abnormal snow series, we follow the approach of Brown, Gustafson, and Ivanov (2021). We first obtain data on daily snow cover from the National Oceanic and Atmospheric Administration website and calculate the average value of snow cover across weather stations for each day and county pair. Next, we calculate the average snow cover in each county-quarter from 2000-2020. We create a rolling 10-year average snowfall cover measure for each county-quarter and create a variable *AbnormalSnow* which is the difference between the current county-quarter average snowfall and the trailing 10-year average.[12] To minimize the impact of extreme outliers, we winsorize the top and bottom 1% of observations across the entire sample period. For these tests, we only use data in the first and fourth quarter of the calendar year because there is rarely any snowfall outside of these periods.

We begin by showing that excess snowfall leads to an increase in a county's unemployment rate. To do so we estimate the following regression at the county-quarter level:

$$UR_{c,t} = \beta AbormalSnow_{c,t} + \alpha_c + \delta_t + \epsilon_{c,t}, \tag{3}$$

where α_c are county fixed effects and δ_t are quarter fixed effects. The results are displayed in Table 4, with and without county fixed-effects. For instance, without county fixed effects an additional inch of snow leads to a statistically significant 0.05pp increase in that county's unemployment rate. The standard deviation of *AbnormalSnow* is 3.7 inches and the average unemployment rate of 6.3%. Hence, a one standard deviation increase in abnormal snowfall leads to a 3% increase in the unemployment rate. This result is consistent with the main findings in Brown, Gustafson, and Ivanov (2021), who show that abnormal snowfall leads to decreases in firms' cash flows.

After establishing that abnormal snowfall leads to increases in unemployment, we estimate a modified version of Equation (2) where the unemployment rate is replaced

[12]The trailing average only considers snowfall in the calendar quarter of interest, so in Q1 of each year snowfall is compared to the average over the previous ten Q1s.

with this measure of abnormal snowfall.[13] These results are shown in Table 5. We estimate that an additional inch of abnormal snow increases the sensitivity of realized to predicted default by 0.18bp. This measure is both statistically and economically significant; as an additional inch of snow increases the sensitivity of realized default to PD by approximately 15% of the unconditional estimate shown in Table 2.

3.3 Mechanisms

In this section we explore potential mechanisms that can generate countercyclical information quality. Specifically, we analyze the relationship between information quality and loan and firm characteristics and how this relationship evolves over the business cycle. If banks are producing information endogenously, we would expect their information to be of higher quality for loans for which their incentives to produce information are higher. Based on this logic, we develop predictions regarding information quality and loan and firm characteristics.

Our first prediction is that banks should have higher quality information about larger loans. As the size of a loan increases, the bank puts more of its capital at risk, hence its gains from learning more about the borrower's type increases, while the cost of producing information should not vary with the loan size since the information is about the borrower as a whole.[14] We also predict that banks should produce more information about more highly levered borrowers. Intuitively, the asymmetric information problem becomes exacerbated by increasing the sensitivity of a security's payoff to firm quality e.g., (Heider (2003)).[15] Finally, we also predict that banks will have greater incentives to produce information about smaller firms because the asymmetric information problem is more severe (e.g., Chae (2005)). Finally, Dang, Gorton, and Holmström (2013) predicts that following a negative aggregate shock the relationship between loan size and information production intensifies. Although, they do not specifically analyze this, a direct implica-

[13]One might consider a two-stage least squares specification given that abnormal snow is likely exogenous; however, we do not because it is unlikely abnormal snow only affects information quality through the unemployment rate.

[14]See Dang, Gorton, and Holmström (2013) and Weitzner (2019).

[15]Moreoever, higher leverage means lower recovery values in the event of default, so banks will have a stronger incentive to avoid losses for these types of firms (e.g., Dang, Gorton, and Holmström (2013)).

tion of their theory is that the relationship between leverage and information production will become stronger in bad times.[16]

These predictions imply particular signs of the interaction coefficients between PD and firm or loan characteristics. For instance, a positive coefficient estimate for the interaction between PD and a characteristic will imply that PD's realized default predictability increases for larger values of that characteristic, or equivalently, banks have more precise information about loans with higher values of that characteristic. We estimate a modified version of Equation (1) that includes these interactions:

$$D_i = \beta PD_i + \Omega X_i + \rho \left(X_i \times PD_i\right) + \delta_{b,t} + \gamma_{j,t} + \sigma_{b,c} + \epsilon_i. \tag{4}$$

The results are shown in Table 6. The first row shows the interaction between PD and loan size, which is measured as the standard deviation after taking logs. This coefficient suggests that a one standard deviation increase in the log loan size increases the sensitivity of realized default to PD by 0.7bp, or about 59% of the unconditional effect shown in Table 2. The second row displays the interaction between PD and firm size and shows that a one standard deviation decrease in a firm's log assets increases the PD coefficient by approximately 34%.[17] Finally, the third row shows the interaction coefficient between PD and leverage. This coefficient is positive, but not quite statistically significant. Overall, these results are consistent with banks producing more information as their incentives vary across firm and loan types.

For further evidence, we examine interactions between the effects of business cycles and these firm and loan characteristics. Dang, Gorton, and Holmström (2012) show that lenders' incentives to produce information about loans are more sensitive to security design decisions following negative aggregate shocks. We test this hypothesis by estimating a modified version of Equation 2 that also includes triples interaction terms between PD,

[16]This follows directly from the fact that the expected recovery is lower for more highly levered firms.

[17]The fact the size interaction is negative may be puzzling at first blush because there is likely more public information about larger firms. However, this result can be rationalized if the asymmetric information problem is more severe for smaller firms which induces banks to produce more information. For instance, this might occur if the distribution of cash flows across smaller firms has a higher variance than larger firms, increasing the incentives for the bank to produce information.

the unemployment rate, and firm/loan characteristics. Following the predictions of Dang, Gorton, and Holmström (2013), we expect the triple interaction coefficients should have the same sign as the interaction coefficients shown in Table 6 as banks respond more strongly to these characteristics in downturns.

The results are shown in Table 7. Consistent with our hypothesis, we find that the triple interaction coefficients for loan size, leverage, and firm size all have the same sign as the interactions with predicted default in Table 6. The coefficients are positive for loan size and leverage, suggesting that bank information quality is more sensitive to the business cycle for large loans and loans to highly leveraged companies. The interaction coefficient for firm size is negative, though it is not quite statistically significant. Taken together, these results provide support for the theoretical framework of Dang, Gorton, and Holmström (2012), in which the countercyclical behavior of bank information quality is the result of endogenous information production decisions.

Our evidence does not 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 to rationalize why information quality is higher for larger loans, smaller firms and more highly levered firms, especially after controlling for other firm and loan characteristics. Moreoever, an alternative channel would also have to explain why the sensitivity between these characteristics and information quality increases in bad times. Even in the presence of some form of exogenous aggregate component of information quality that might vary across the business cycle, our results suggest that information production still increases *relatively* more for certain types of securities in a manner consistent with the framework of Dang, Gorton, and Holmström (2013).

4 Policy Implications

The evidence we provide in this paper suggests that countercyclical information quality is driven by endogenous information production by banks. In this section we consider the implications of this mechanism for policies designed to stimulate bank lending. Many governments and central banks around the world responded to the global financial crisis by implementing monetary or fiscal stimulus measures. These measures included policies explicitly focused on promoting bank lending, such as the U.K.'s Funding for Lending program (see Churm et al. (2012)). Supporting credit markets was also often mentioned as a motivation behind more other policy instruments such as interest rate cuts, liquidity facilities, and asset purchases.[18]

Because economic conditions affect banks' screening efforts, they will also impact the transmission of these policies. As bank screening intensity increases during downturns, some firms which would have been able to receive credit in good times might suddenly find themselves excluded from borrowing. This means the marginal borrowers in recessions and expansions—and thus the ultimate beneficiaries of policies designed to stimulate lending—will be different. In other words, recessions will be periods in which fewer but higher quality firms receive credit. This mirrors the findings of Ates and Saffie (2021), who show that financial factors can explain why fewer firms enter during periods of financial distress, but the firms that do enter are more productive.

More generally, these results can also shed light on the potential underlying mechanisms in research analyzing the state dependent effects of monetary policy. Past work such as Gertler and Gilchrist (1994), Bernanke and Gertler (1995), and Bernanke, Gertler, and Gilchrist (1999) has argued that bank lending is an important transmission channel for monetary policy. Separate and more recent work including Tenreyro and Thwaites (2016) has showed that monetary policy is less effective in recessions. Our results suggest that changes in banks' information production incentives can lead to changes in the number and composition of firms that receive financing following changes in monetary policy, and thus help explain the state dependence observed in the data. This is also consistent the findings of Wieland and Yang (2020), who show that loan retrenchment by bank holding companies during downturns diminishes the efficacy of monetary policy.

[18] See https://www.federalreserve.gov/newsevents/speech/bernanke20090113a.htm.

5 Conclusion

Information plays a crucial role in banks lending behavior to firms and in turn macroeconomic outcomes, but is difficult to analyze empirically. In this paper, we construct a novel measure of bank information quality from confidential regulatory data and analyze its properties. Using county-level variation in unemployment rates, we find that information quality improves as economic conditions worsen. We argue that these results are consistent with theories of endogenous information production by providing additional empirical evidence linking information quality to firm and loan characteristics as well as exogenous shocks to local economic conditions. These findings have important implications for policymakers because banks' information production decisions affect the volume of credit available to firms, and thus the usefulness of different policy tools may critically depend on aggregate 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.

Andrade, Philippe, Christophe Cahn, Henri Fraisse, and Jean-Stéphane Mésonnier, 2019, Can the provision of long-term liquidity help to avoid a credit crunch? evidence from the eurosystem's ltro, *Journal of the European Economic Association* 17, 1070–1106.

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

Ates, Sina T and Felipe E Saffie, 2021, Fewer but better: Sudden stops, firm entry, and financial selection, *American Economic Journal: Macroeconomics* 13, 304–56.

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.

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

Bernanke, Ben S and Mark Gertler, 1995, Inside the black box: the credit channel of monetary policy transmission, *Journal of Economic perspectives* 9, 27–48.

Bernanke, Ben S, Mark Gertler, and Simon Gilchrist, 1999, The financial accelerator in a quantitative business cycle framework, *Handbook of macroeconomics* 1, 1341–1393.

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 de-risking? how banks cope with loss, *Review of Economic Dynamics* .

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

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

Butt, Nicholas, Rohan Churm, Michael F McMahon, Arpad Morotz, and Jochen F Schanz, 2014, Qe and the bank lending channel in the united kingdom .

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

Chae, Joon, 2005, Trading volume, information asymmetry, and timing information, *The journal of finance* 60, 413–442.

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

Churm, Rohan, Amar Radia, Jeremy Leake, Sylaja Srinivasan, and Richard Whisker, 2012, The funding for lending scheme, *Bank of England Quarterly Bulletin* Q4.

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 Holmstrom, 2009, Opacity and the optimality of debt for liquidity provision, *Manuscript Yale University* .

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* .

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.

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.

Gertler, Mark and Simon Gilchrist, 1994, Monetary policy, business cycles, and the behavior of small manufacturing firms, *The Quarterly Journal of Economics* 109, 309–340.

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.

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

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

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.

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.

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.

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

Tenreyro, Silvana and Gregory Thwaites, 2016, Pushing on a string: Us monetary policy is less powerful in recessions, *American Economic Journal: Macroeconomics* 8, 43–74.

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

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

Wieland, Johannes F and Mu-Jeung Yang, 2020, Financial dampening, *Journal of Money, Credit and Banking* 52, 79–113.

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

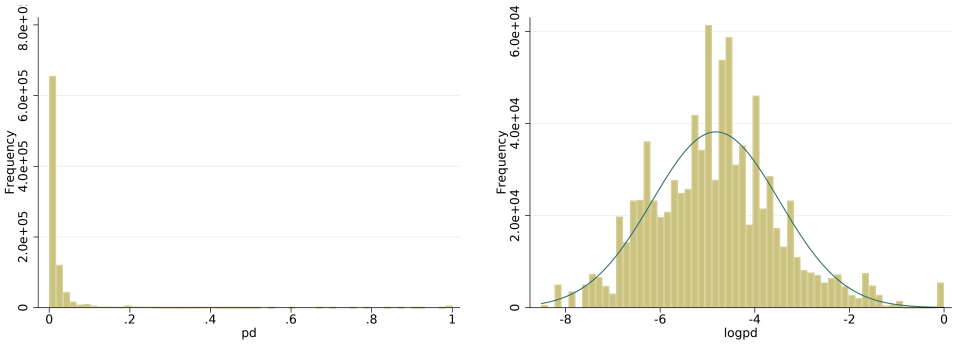

Figure 1: Frequency distributions of PD (left) and log(PD) (right)

This figure shows the frequency distributions of PD (left) and log(PD) (right) for our sample.

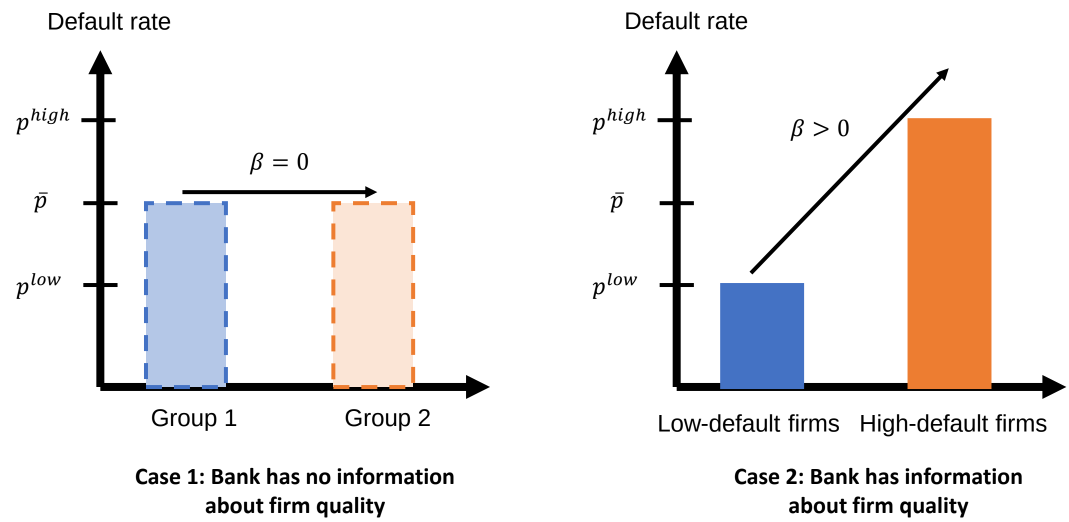

Figure 2: Information Precision

This figure illustrates our empirical approach to measuring information quality. The left panel shows a bank which has no information about firm quality, which means their default rates for each group of firms will simply be the unconditional average default rate $\bar{p}$. The right panel shows a bank with information about firm quality. Because the bank is able to identify which firms have ex-ante higher or lower default rates, the sensitivity of actual to predicted default will be positive.

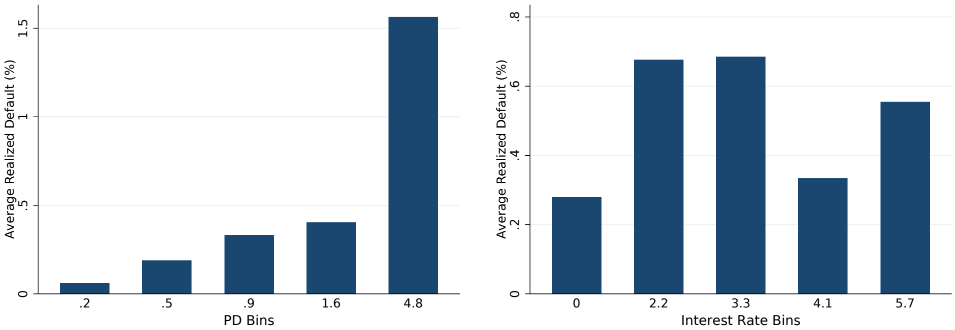

Figure 3: Default rates across quintiles of PD (left) and interest rate (right)

This figure shows default rates by quintiles of PD (left) and interest rates (right). Both variables are measured in percent. The y-axis shows the realized default rate for each bucket in percent, while the numbers on the x-axis underneath each bar correspond to the lowest value for included in bucket.

7 Tables

Table 1: Summary Statistics

This table contains summary statistics for our sample. Appendix A describes the sample construction and reports all variable definitions.

	Sales ($mn)	Assets ($mn)	Leverage	Maturity (m)	Loan ($mn)	PD (%)
Mean	2,286.3	3,921.0	0.34	47.3	12.9	1.47
Median	77.1	47.4	0.31	58.0	3.5	0.91
5th pct	2.7	1.8	0.00	7.0	1.0	0.15
95th pct	4,250.4	4,797.3	0.80	88.0	50.0	4.44
SD	38,233.0	81,803.4	0.26	30.5	38.7	2.03
N	58,393	58,379	57,268	71,048	71,048	71,048

Table 2: Predicting Realized Default

This table tests whether PD predicts realized default beyond other loan and firm characteristics (Equation 1). The dependent variable in each regression is an indicator for whether each loan defaults within eight quarters after origination and then multiplied by 100 so that coefficients are interpreted in percentage points. The interest rate is measured in percent. PD represents the percentile rank within a bank-quarter pair for each PD and takes values in $(0, 100]$. Appendix A describes the sample construction and reports all variable definitions. Standard errors are clustered at the county level and shown in parentheses.

	(1)	(2)	(3)
PD	0.0117***	0.0132***	0.0126***
	(0.00139)	(0.00216)	(0.00218)
Interest rate (%)			0.128**
			(0.0493)
Controls	N	Y	Y
Observations	67,759	53,107	42,542
R^2	0.194	0.210	0.245

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Information Quality Over the Business Cycle

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 an indicator for whether each loan defaults within eight quarters after origination expressed in percentage points. The unemployment rate is measured in percentage points. PD represents the percentile rank within a bank-quarter pair for each PD and takes values in $(0, 100]$. The "Control interactions" means that interaction terms between the controls and unemployment rate are included in the regression. Appendix A describes the sample construction and reports all variable definitions. Standard errors are clustered at the county level and shown in parentheses.

	(1)	(2)	(3)	(4)
PD	-0.0118^{***}	-0.0701^{***}	-0.00855^{*}	-0.0528^{*}
	(0.00426)	(0.0267)	(0.00475)	(0.0313)
UR	-0.208	-0.262		
	(0.167)	(0.187)		
PD × UR	0.00534^{***}	0.00610^{***}	0.00482^{***}	0.00557^{***}
	(0.000995)	(0.00125)	(0.00111)	(0.00143)
Control interactions	N	Y	N	Y
County-quarter FE	N	N	Y	Y
Observations	67,759	53,107	63,587	49,284
R^2	0.195	0.212	0.280	0.322

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: The Effect of Abnormal Snowfall on Unemployment

This table tests whether abnormal snowfall leads to higher local unemployment rates (Equation 3). The construction of *AbnormalSnow* is described in Section 3.2.2. Standard errors are clustered at the county level and shown in parentheses.

	(1)	(2)
AbnormalSnow	0.0047**	0.013***
	(0.0019)	(0.0021)
Quarter FE	Y	Y
County FE	N	Y
Observations	30,957	30,939
R^2	0.535	0.885

Table 5: The Effect of Abnormal Snowfall on Information Quality

This table tests whether abnormal snowfall affects the sensitivity of realized default to PD. The estimated regression is a modified version of Equation 2 in which the unemployment rate is replaced with *AbnormalSnow* which is measured in inches. The dependent variable is a dummy indicating whether each loan defaults within eight quarters after origination expressed in percentage points. PD represents the percentile rank within a bank-quarter pair for each PD and takes values in $(0, 100]$. Standard errors clustered by county are shown in parentheses.

PD	-0.0861^{*}
	(0.0473)
AbnormalSnow	-0.0934^{***}
	(0.0360)
PD $\times$ AbnormalSnow	0.00175^{**}
	(0.000809)
Observations	27,442
R^2	0.239

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Information Quality Across Loan and Firm Characteristics

This table tests whether certain firm and loan characteristics affect the sensitivity of realized default to PD (Equation 4). PD represents the percentile rank within a bank-quarter pair for each PD and takes values in $(0, 100]$. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination and then multiplied by 100 so that coefficients are interpreted in percentage points. Firm size and loan size are measured in standard deviations of logs while leverage is a ratio. Appendix A describes the sample construction and reports all variable definitions. Standard errors are clustered at the county level and shown in parentheses.

PD × (Loan size)	0.00704***
	(0.00187)
PD × (Firm size)	−0.00391**
	(0.00192)
PD × Leverage	0.00167
	(0.00129)
Observations	53,107
R^2	0.212

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Sensitivity of Firm and Loan Characteristics to the Business Cycle

This table tests whether firm and loan characteristics increase the sensitivity of realized default to PD when economic conditions deteriorate. The regression is estimated using a modified version of Equation 2 that also includes triple interaction terms between PD, the unemployment rate, and firm/loan characteristics. PD represents the percentile rank within a bank-quarter pair for each PD and takes values in $(0, 100]$. The dependent variable in each regression is a dummy variable indicating whether each loan defaults within eight quarters after origination and then multiplied by 100 so that coefficients are interpreted in percentage points. Appendix A describes the sample construction and reports all variable definitions. Standard errors are clustered at the county level and shown in parentheses.

PD × UR × (Loan size)	0.00468***
	(0.00137)
PD × UR × (Firm size)	-0.00240
	(0.00252)
PD × UR × Leverage	0.00373**
	(0.00181)
Observations	53,107
R^2	0.213

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix A. 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'.[19]

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.[20] 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 approach will thus be unaffected by systematic misestimation of the *level* of the default

[19]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.

[20]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).

rate.

Figure 4 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.

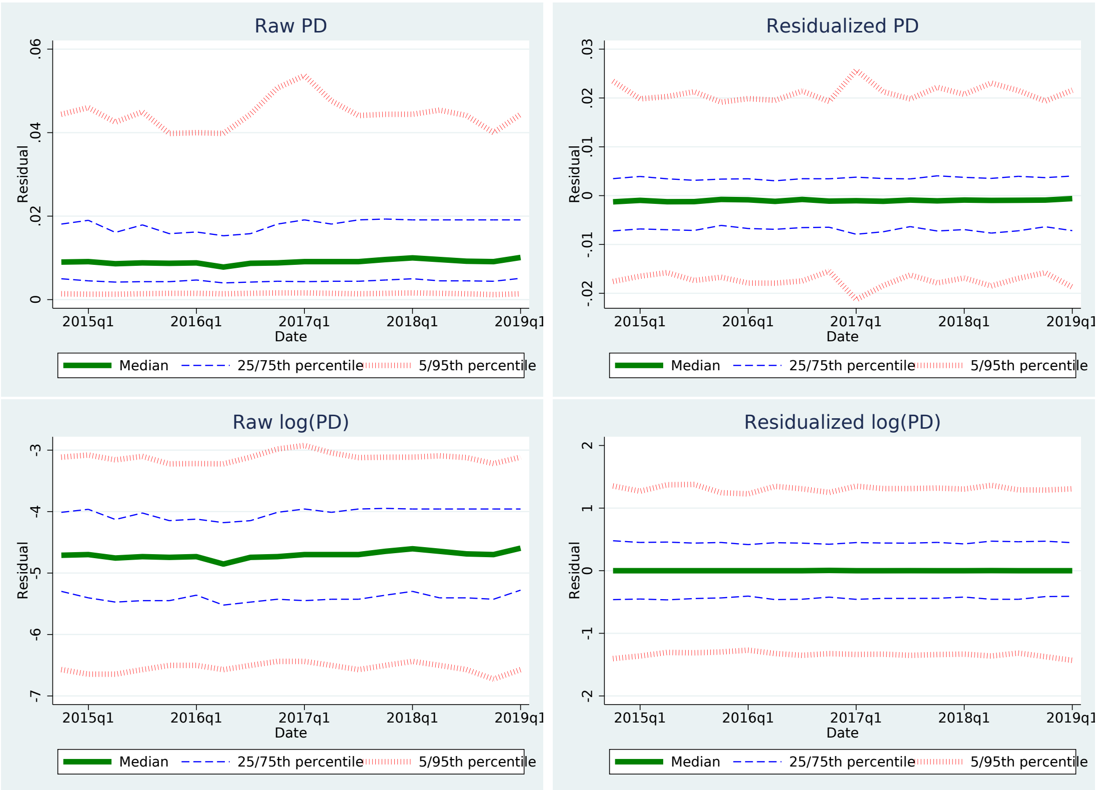

Figure 4: PD dispersion over time

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

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

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

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

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

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

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

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

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

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

PD: The bank's expected annual default rate over the life of the loan, trimmed if equals zero or above the 99th percentile, from Y-14Q.

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

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

Abnormal Snow: The difference between the current quarter's realized average daily snow cover and the trailing ten year average quarterly snowfall during the first quarter and fourth quarter of each year (see Brown, Gustafson, and Ivanov (2021) for more details), from NOAA.

Appendix B. Evaluating Information Quality

This section describes our approach to evaluating bank information quality in more detail. Our measure is based on the idea that after controlling for observables, the predicted probability of default (PD) will have higher correlation with realized default for banks with better information. We provide some theoretical intuition for this approach and show that it can be calculated empirically as the coefficient on PD in regressions with default as the dependent variable. Higher coefficients on PD in these regressions imply a greater elasticity of actual default to predicted default and thus more precise information.

To provide a more general illustration of this concept, consider a simple model of forecasters trying to predict the likelihood that a loan will default. The default probability p is drawn from a known distribution $F(p)$. Forecasters do not observe p, but instead receive a signal $s = p + \epsilon$, where ϵ is a conditionally independent noise term with mean zero and variance σ^2. Default probability forecasts will be weighted averages of the signal and the unconditional default probability: $\hat{p} = \alpha s + (1 - \alpha)\bar{p}$, where α is the weight banks place on their signal. The optimal weight α^* will be decreasing in σ^2. If $\sigma^2 = 0$, the optimal forecast will be the signal realization, so $\hat{p} = s = p$. In the extreme case in which σ^2 is infinite, the signal s_i provides no information and the optimal forecast is simply the unconditional average $\bar{p}$.

These weights can be estimated empirically using a linear regression of default on PD. Figure 5 shows this approach graphically using simulated data for three different signal qualities. The left panel shows a high-precision signal, the middle panel shows a low-precision signal, and the right panel shows a signal that has zero correlation with p. The vertical axes correspond to bins of realized default probabilities, with the red dotted line representing the unconditional average default rate, while the horizontal axes correspond to bins of predicted default probabilities. For example, the point (0.20,0.25) would mean that the set of all loans classified by forecasters as having a default probability of 20% actually defaulted 25% of the time.

The slope of the solid black regression line through these points represents the elasticity of actual default to predicted default. This line can be thought of as the optimal default forecast as a function of PD. A steeper slope means that predicted and actual default will have a stronger correlation, and therefore that more weight should be put on PD. When the variance of the noise surrounding the signal is low as in the left panel, the information contained in PD will be more precise and the regression line will be steeper. In this case, a 1 percentage point increase in predicted default corresponds to an increase of around 0.74 percentage points in actual default.

As the variance of the noise term increases, the signal becomes less precise and the estimated coefficient attenuates. This can be seen in the middle panel, where a noisier signal leads to a flatter slope and the same 1 percentage point increase in predicted default leads to an increase of just 0.26 percentage points in actual default. Finally, in the extreme case in which the signal provides no useful information, the elasticity of actual default to predicted default will be zero and the optimal forecast of default probability will simply correspond to the unconditional average default rate regardless of the signal's realization. Together, these images provide graphical intuition for why larger regression coefficients on PD correspond to more precise information.

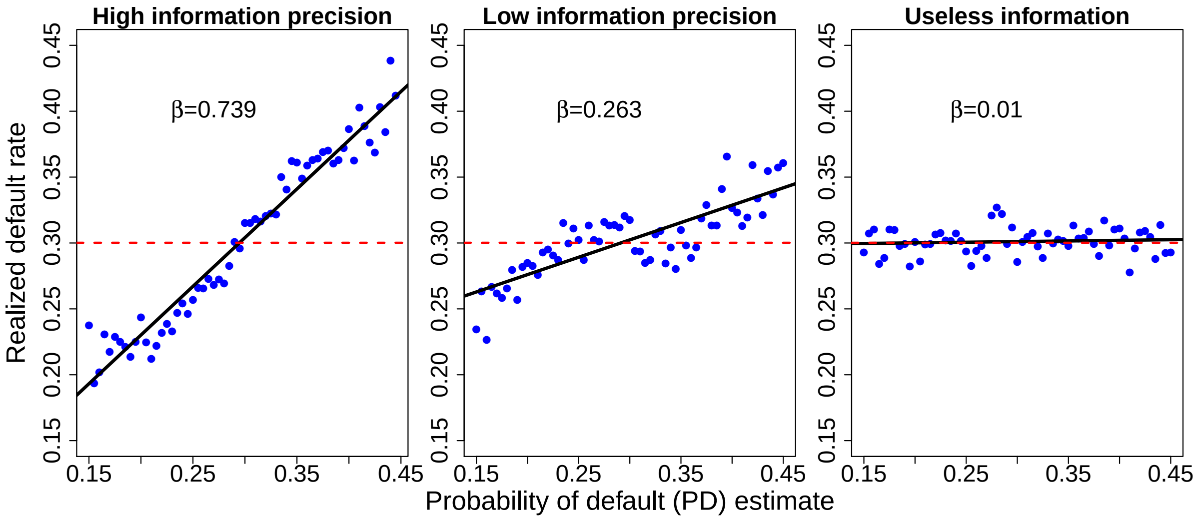

Figure 5: Simulated information precision

This simple illustration is most easily interpreted as a single bank making a large number of loans to a fixed distribution of borrowers. In practice, heterogeneity across firms, banks, and economic conditions can lead to differences across these groups that can complicate the process of estimating the relationship between expected and actual defaults. This could occur, for example, if one bank consistently over-estimated its default probabilities. In that case, a naive approach which pooled all observations together would not accurately estimate the correct elasticity between predicted and actual default.

This can be seen graphically in Figure 6. The left panel shows a similar exercise to the one shown in Figure 5, but for two banks. Bank 1 (shown as the orange circles) has estimates which are on average equal to the true underlying default rate. Bank two (shown as the purple squares) has default estimates which are consistently too conservative. Despite this bias in the *level* of its default estimates, the elasticity of actual default to predicted default is still positive for Bank 2. Both banks have the same signal variance, so in the limit the estimates of β should both be equal to one. When the elasticities are estimated separately for each bank, the estimates are similar to those of the high-information case shown previously ($\beta^1 = 0.63$ and $\beta^2 = 0.75$).

When all of the estimates are pooled, however, the coefficient attenuates to $\beta = 0.22$. This is shown in the middle panel and occurs because many of the predicted default rates with the same level correspond to different portions of the distribution of each bank's estimates: a PD of 0.4 is on the low end of what Bank 2 forecasts, but on the high end of what Bank 1 forecasts. Not accounting for this leads to a deterioration in the relationship between predicted and actual default and lowers the coefficient estimate relative to the true marginal effect.

The use of fixed effects allow accurate estimation of the average elasticity of actual to predicted default in the presence of this type of persistent heterogeneity. An illustration of this approach is shown in the right panel of Figure 6. In this figure both the dependent and independent variables are de-meaned within each bank, so that positive values of the x-axis correspond to above-average predicted default rates. Using this approach, the estimated elasticity is the average of the elasticities obtained from the separate approach in the left panel and confirms that bank default forecasts are useful for predicting actual default. This approach is equivalent to evaluating banks' PD forecasts on a relative (rather than absolute) basis.[21]

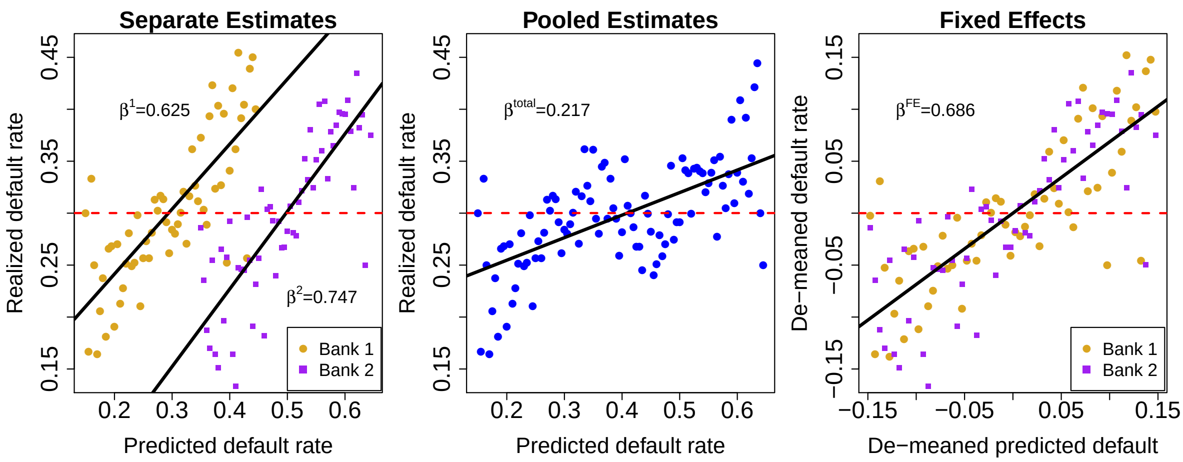

Figure 6: Simulated information precision with bank heterogeneity

In summary, the elasticity of actual default to predicted default is a useful measure of the underlying quality of bank information, and the use of fixed effects for different groups of loans allows us to estimate changes in information quality even in the presence of persistent biases in the *level* of default forecasts.

[21] As discussed in Appendix A.1, this approach is consistent with how regulators evaluate banks' risk models.

Appendix C. Extensions and Robustness Checks

Table 8: 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 and then multiplied by 100 so that coefficients are interpreted in percentage points. The interest rate is measured in percent. log(PD) is measured measured in standard deviations calculated across our entire sample. Appendix A describes the sample construction and reports all variable definitions. Standard errors are clustered at the county level and shown in parentheses.

	(1)	(2)	(3)
log(PD)	0.398***	0.471***	0.478***
	(0.0492)	(0.0765)	(0.0860)
Interest rate (%)			0.111***
			(0.0466)
Controls	N	Y	Y
Observations	67,759	53,107	42,542
R^2	0.195	0.211	0.246

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: 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 measured measured in standard deviations calculated across our entire sample. Appendix A describes the sample construction and reports all variable definitions. Standard errors are clustered at the county level and shown in parentheses.

	(1)	(2)	(3)	(4)
PD	-0.338^{**}	-2.03^{**}	-0.243	-1.56^{*}
	(0.139)	(0.00795)	(0.00155)	(0.00903)
UR	0.813^{***}	0.964^{***}		
	(0.238)	(0.287)		
PD × UR	0.167^{***}	0.205^{***}	0.152^{***}	0.174^{***}
	(0.0995)	(0.125)	(0.111)	(0.143)
Control interactions	N	Y	N	Y
County-quarter FE	N	N	Y	Y
Observations	67,759	53,107	63,587	49,284
R^2	0.195	0.213	0.280	0.323

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D. Simple Theoretical Framework

In this section we present a very simple model that highlights how the business cycle can affect bank information production incentives. It is important to stress that we view this as a possible mechanism, not the exclusive one, to generating the empirical prediction that banks' produce more information when aggregate economic conditions deteriorate.

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.[22] 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 firm 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.[23]

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

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

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.

[22]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 0 so there is no adverse selection on the borrower side.

[23]This allows us to abstract away from the bargaining process. See Dang, Gorton, and Holmström (2013) and Weitzner (2019) for cases in which the face value of debt is endogenous.