Bank Sentiment and Loan Loss Provisioning

Junsung Bae* Allen N. Berger† Hyun-Soo Choi‡ Hugh H. Kim§

February 2024

Abstract

The recent shift in regulatory policy towards forward-looking discretion-based loan loss provisioning hinges on the assumption of rational optimizing behavior by bank managers. This paper challenges this premise by investigating the influence of bank managerial sentiment on loan loss provisioning. Leveraging large-language models such as BERT and GPT, we extract sentiment indicators from banks' 10-K filings, distinct from fundamental-based beliefs and borrower-side sentiments. Our analysis reveals that banks exhibiting more negative sentiment tend to increase their loan loss provisions beyond what is justified by economic fundamentals and future loan charge-offs. Furthermore, banks with more excessive sentiment-driven provisions reduce their lending in the future. The impact of bank sentiment is more pronounced during the recessionary periods, suggesting that the sentiment can amplify the counter-cyclicality of loan loss provisions and the pro-cyclicality of bank lending. To mitigate endogeneity concerns related to bank sentiment, we employ exogenous weather conditions as instrumental variables. Overall, our results suggest that the sentiment-driven discretion in loan loss provisioning may exacerbate the pro-cyclicality of bank lending practices.

JEL classification: G21, G40, M41
Keywords: bank sentiment, loan loss provision, lending, textual analysis, banking industry, financial disclosure

*Korea Advanced Institute of Science and Technology (e-mail: kihons125@kaist.ac.kr)
†Darla Moore School of Business, University of South Carolina (e-mail: aberger@moore.sc.edu)
‡Korea Advanced Institute of Science and Technology (e-mail: hschoi19@kaist.ac.kr)
§Darla Moore School of Business, University of South Carolina (e-mail: hugh.kim@moore.sc.edu)

1. Introduction

Loan loss provision directly influences a bank's capital adequacy and its ability to lend. An important academic and regulatory debate is about how to mitigate the counter-cyclicality of loan loss provisions and the resulting pro-cyclicality of bank lending. In the traditional incurred loss (IL) model, banks are often late in recognizing loan losses and thus excessively reduce their lending during economic downturns (Beatty and Liao (2011); Bushman and Williams (2012); Bushman and Williams (2015)). A recent regulatory change to forward-looking provisioning rules—expected credit loss (ECL) or current expected credit loss (CECL) model—gives bank managers more discretion on how much to set aside as provisions for their future losses (The Financial Stability Forum (2009); U.S. Treasury (2009)). This proposal is based on the assumption that forward-looking rational bank managers will promptly choose their optimal level of provisions, reducing the counter-cyclicality of loan loss provisions and the pro-cyclicality of bank lending. However, whether bank managers actually behave in this manner is subject to debate.

In this paper, we assess the assumption of objective optimizing behaviors of bank managers by testing the impact of their sentiment on loan loss provisions. We argue that bank managers' sentiment, broadly defined as their belief about current and future economic conditions, is likely to influence their loan loss provisioning. Even under the IL model, loan loss provisions crucially depend on bank managers' discretion in their risk assessments. By setting aside a portion of their earnings to cover potential future loan defaults, bank managers inherently make statements about their expectations for the creditworthiness of their borrowers, the banks' conditions, and the economic environment at large. In this paper, we focus on banks' managerial sentiment independent of their fundamental-based beliefs and other economic agents' sentiments. To the extent that such sentiment affects loan loss provisions, the implementation of more discretion-based accounting rules such as CECL can potentially amplify the cyclicality of bank lending, contrary to regulators' intentions.

Loan loss provisioning swayed by bank sentiment may have adverse effects on the real economy. At the onset of recessions, some banks may be overly pessimistic about the future economic environment and set aside excessive loan loss provisions. Over-provision can lead to a reduction in credit availability, potentially deepening the recession and stifling economic growth (Laeven and Majnoni (2003); Beatty and Liao (2011)). Similarly, during economic booms, overly optimistic banks may set aside too little for their future loan defaults, which can lead to an oversupply of credit, possibly creating asset bubbles and contributing to economic instability (Acharya and Naqvi (2012)).

To answer our research question, we utilize large language models (LLMs) and extract a bank sentiment measure from the annual reports (Form 10-K) filed by all public bank holding companies (hereafter, called banks) in the U.S. We build the bank sentiment measure distinct from economic fundamentals and the sentiments of investors, consumers, and corporate managers by implementing the two-step approach of Hribar et al. (2017) and Berger et al. (Forthcoming). First, we calculate the net negative sentence ratio of annual reports at a bank-year level by employing *FinBERT* fine-tuned by Huang et al. (2023).[1] Second, we estimate a regression of the net negative sentence ratio on the state-year fixed effects, absorbing the impact of economic fundamentals and other macro-level sentiment shocks from consumers (Carroll et al. (1994)), investors (Baker and Wurgler (2006)), and corporate managers (Jiang et al. (2019)). We then capture the residuals of the estimated regression model, the portion of negative tone in annual reports not explained by the economic fundamentals and other economic agents' sentiments, calling it "unexplained negative belief" (*Neg-BankSentiment*).

We analyze the annual reports of the U.S. banks from 1995 to 2019 before the implementation of the CECL.[2] We parse the whole parts of annual reports because full textual

[1] *FinBERT* is a large language model adapted for the finance domain. It takes in a sentence and extracts the most likely sentiment of the given sentence: negative, positive, or neutral. For example, when the sentence "The increase in net interest income in 2015 predominantly reflected higher average loan balances and lower interest expense on deposits" is given to the model, *FinBERT* predicts its sentiment as positive. We utilize a *FinBERT* model fine-tuned by Huang et al. (2023) for sentence classification tasks. Compared to the conventional bag-of-words approach using Loughran and McDonald (2011) dictionary, *FinBERT* shows superior performance in sentence classification tasks (Huang et al. (2023)).

[2] We focus on the sample period before the CECL implementation to avoid the impact of major accounting rule changes. Even under the incurred loss (IL) model, bank managers still exercise their judgement to estimate probable loan losses but with a lesser degree compared to the CECL model.

information on Form 10-Ks provides a comprehensive and detailed overview of a bank's financial performance and potential risks. In our robustness checks, we also consider only the MD&A section as a textual source for deriving sentiment measures. We additionally employ alternative textual analysis techniques such as the GPT language model and the traditional bag-of-word approach based on Loughran and McDonald (2011) dictionary. Importantly, we include the future realized net charge-offs as a control variable in all our regressions. By directly controlling for the actual realization of loan losses, we can identify the sentiment-driven loan loss provision distinct from the effect of a bank's private information about the creditworthiness of its borrowers, which is unobservable to researchers.

To preview the analysis results, we find that a bank's negative sentiment is positively and significantly correlated with loan loss provision. The result holds even after we control for future net charge-offs and other variables that can account for the regulatory portion of loan loss provision. The result supports our hypothesis that bank managers with negative sentiment conservatively project future economic conditions, leading to increased provisions for loan losses. The result also holds across different size groups of banks. Importantly, the relation between negative bank sentiment and loan loss provision is more pronounced during recessions, suggesting that bank sentiment may play a role in amplifying the counter-cyclicality of loan loss provisions—too little during good times and too much during bad times. As the sentiment measure is net of all macro variables, this result is not driven by recession-related macroeconomic fundamentals.

We address potential endogeneity concerns about the bank sentiment measure by using exogenous weather variations near the headquarters of banks. Exogenous weather variation near the headquarters is an attractive instrumental variable because it would influence bank sentiment (Lerner et al. (2015); Dehaan et al. (2017); Berger et al. (Forthcoming)), but it is not likely to affect loan loss provisions directly. Our sample is of large public banks with business exposures to diverse geographical locations, and their loan loss provision cannot be solely driven by weather-related local economic conditions. Following Berger et al. (Forthcoming), we utilize LASSO (Least Absolute Shrinkage and Selection Operator) for choosing instrumental variables from a large number of IV candidates to overcome the over-fitting and hand-picking issues. We find that cloud coverage is the most probable instrument variable—bank sentiment is more negative when there are more consecutive cloudy days during the periods prior to the filing of annual reports compared to the last years. The instrumental variable analysis confirms that the negative bank sentiment increases loan loss provisions above and beyond the level warranted by the economic fundamentals.

A crucial question following the above analysis is whether sentiment-driven loan provision affects a bank's lending behavior. In our additional analysis, we estimate regression models of a bank's future lending growth on sentiment-driven loan loss provisions. We find that the coefficient on the sentiment-driven loan loss provision is negative and statistically significant, suggesting that sentiment-driven over-provisioning can reduce credit provided by banks to the economy. Combining this result with the above analysis of the counter-cyclical relation between negative bank sentiment and loan loss provision (i.e., the negative sentiment increases loan loss provisions more during recessions), we can infer that bank lending can be stifled more by the sentiment-driven loan loss provisions during recessions. In a similar vein, banks with positive sentiment during economic booms might set aside too little provisions and extend too much credit, adding to economic instability in the future.

This paper contributes to the literature investigating the determinants of loan loss provisioning and its countercyclicality. The current literature focus on the earnings management or capital management incentives of bank managers (Moyer (1990), Collins et al. (1995); Beatty et al. (1995); Kim and Kross (1998)). Other studies explore the effect of countercyclical loan loss provisioning on bank lending and risk-taking (Beatty and Liao (2011); Bushman and Williams (2012); Bushman and Williams (2015)). There is, however, little research exploring the effect of behavioral aspects on loan loss provisions, which could not be fully explained by the manager's incentives. We argue that the bank sentiment is an important driver of the countercyclicality of loan loss provisioning and the resulting procyclicality of bank lending.

Given that regulatory changes are increasingly bestowing more discretion to bank managers (Cohen and Edwards (2017)), our paper also helps understand the potentially adverse economic effects of discretionary loss provisioning under the new ECL standards.

Our paper also relates to the literature on the sentiment—beliefs or attitudes unjustified by economic fundamentals— of economic agents. There are ample evidences about the impact of investor sentiment (Baker and Wurgler (2006); Lemmon and Portniaguina (2006)), corporate manager sentiment (Brown et al. (2012); Jiang et al. (2019)), and consumer sentiment (Ludvigson (2004)). Berger et al. (Forthcoming)) study the impact of bank sentiment on liquidity hoarding. But there is little research studying the effect of bank sentiment on loan loss provisioning. Hribar et al. (2017) is a closely related paper focusing on the sentiment of general corporate managers (i.e., borrowers), not the bank managers (i.e., lenders). Our paper shows that the sentiment of bank managers, even after controlling for time-varying local economic conditions and the corporate manager sentiment, has a distinct effect on the loan loss provision. We also find that the resulting sentiment-driven loan loss provision can influence bank lending.

The remainder of the paper is organized as follows. In Section 2, we develop our main hypothesis by reviewing prior studies in the literature. Section 3 explains the key variables of interest, especially the bank sentiment measure. In Section 4, we empirically test our hypothesis and verify the result in various ways. Section 5 concludes by discussing the implication of the paper's findings.

2. Related Literature and Hypothesis Development

It has long been believed that the sentiment of economic agents influences the real economy and financial markets in a way that is distinct from economic fundamentals (Keynes (1937)). Many studies have considered the impact of behavioral aspects of economic agents, such as investors (Baker and Wurgler (2006); Lemmon and Portniaguina (2006)), corporate managers

(Jiang et al. (2019); Hribar et al. (2017)), and consumers (Carroll (1997); Carroll et al. (1994); Batchelor and Dua (1998)). Although the exact definition and nuance of sentiment varies across contexts, an overarching feature of sentiment is an unjustified belief of economic agents about current and future economic conditions. The influence of sentiment on economic activities is based on evidence from psychology and behavioral economics literature. Previous research documents that negative sentiment can heighten the perceived likelihood of adverse events, shaping an individual's expectations about the future (Johnson and Tversky (1983); Wright and Bower (1992); Wegener and Petty (1994)). Moreover, individuals' risk aversion can be swayed by sentiment (Zuckerman (1984); Wong and Carducci (1991); Horvath and Zuckerman (1993); Tokunaga (1993); Bassi et al. (2013)). Therefore, negative sentiment can induce economic agents to be overly pessimistic about future economic conditions and reduce their risk appetite.

In the domain of banking research, the perceptions held by senior management significantly influence key strategic decisions (Rajan (1994)). When top executives are held up by negative sentiment, they might form overly negative views on their future economic conditions and credit-worthiness of their borrowers, resulting in an overly conservative loan loss provisioning beyond the level warranted by the current and future economic conditions. Therefore, we postulate the following hypothesis:

> ***Hypothesis 1-A***: Banks with more negative sentiment have more loan loss provisions.

An alternative hypothesis is that bank managers are objectively optimizing and their provisioning is not swayed by their sentiment.

> ***Hypothesis 1-B***: Loan loss provisions are not related to bank sentiment.

When economic conditions are worsened and uncertain, the effect of negative sentiment would have a greater impact on bank managers' behavior as it can additionally heighten the perceived likelihood of adverse events (McLean and Zhao (2014); Hribar et al. (2017)). Therefore, we hypothesize that the effect of negative bank sentiment on loan loss provision would be more pronounced during recessions.

> ***Hypothesis 2***: Banks with more negative sentiment have more loan loss provisions by a larger margin during recessions than other times.

3. Data and Key Variables

3.1. Bank Sentiment Measure

Measuring bank manager sentiment is challenging as it reflects the beliefs, attitudes, and emotions of bank managers, which are usually unobservable. Recent studies, however, find that qualitative components in corporate disclosure documents can be useful sources for acquiring firms' unobservable information (Campbell et al. (2014); Hanley and Hoberg (2019); Berger et al. (Forthcoming)). Moreover, as regulators require more extensive and accurate information in corporate disclosure documents (e.g., Sarbanes-Oxley Act of 2002), the embedded textual information is likely to contain relevant information about top corporate executives, including managerial team-level sentiment. In this paper, we build our measure of bank management sentiment from the textual information of annual reports (Form 10-K) filed by all bank holding companies in the U.S.

We develop our bank sentiment measure using the complete textual content of banks' annual reports, distinguishing it from other measures that concentrate on the transcripts of earnings conference calls, press releases, or solely the Management Discussion and Analysis (MD&A) section of 10-K filings. While the textual contents in earnings conference calls can directly reveal a bank manager's sentiment (Davis et al. (2015)), the calls are voluntary and suffer from a selection bias. Similarly, press releases suffer a different selection bias as banks carefully select the information content to be delivered to the shareholders (Davis and Tama-Sweet (2012)). As the MD&A section of an annual report contains the commentaries from the management team, it can be a concise source of textual information to extract managers' sentiments within 10-Ks, when the capacity of processing entire textual information is limited (Brown and Tucker (2011); Muslu et al. (2015)). We analyze the whole parts of annual reports because 10-Ks provide a comprehensive and detailed overview of a bank's financial performance and potential risks, but we also construct a measure solely based on the MD&A section for the robustness of our results.

We first construct a measure for the tone of annual reports. From each 10-K report, we sort all sentences in the report into negative, positive, and neutral sentences by using *FinBERT*, a large language model fine-tuned by Huang et al. (2023). We then calculate the tone of the annual reports as follows:

$$Net\ Negative\ Sentence\ Ratio_{i,t} = \frac{\#\ of\ Neg.\ Sentence_{i,t} - \#\ of\ Pos.\ Sentence_{i,t}}{\#\ of\ Total\ Sentence_{i,t}} \tag{1}$$

for a bank i on year t. Similarly, we can separately build the positive and negative tone (rather than the net negative tone), which we employ in our robustness check section.

We use a variant of *BERT* model (Devlin et al. (2019)) rather than a traditional bag-of-words approach. While the bag-of-words approach is simple and straightforward, there can be a potential concern about the erroneous classification of sentences due to a lack of contextual consideration. A large language model (LLM), such as *BERT* or *GPT*, is less prone to such errors. We adopt *FinBERT*, a variant of *BERT* model pre-trained and fine-tuned by Huang et al. (2023), to sort sentences in the context of financial documents. We later provide the robustness of our results using another large-language model, *GPT* fine-tuned for sentence classification tasks, and the bag-of-words approach based on Loughran and McDonald (2011) dictionary.

To answer our research question about the impact of bank sentiment, it is important to extract the element of an annual report's tone that is independent of the economic fundamentals. We decompose the tone of annual reports into the *explainable* segment, which can be rationalized by economic conditions, and the *unexplainable* segment. Following a widely adopted approach in the literature (e.g., Lemmon and Portniaguina (2006) and Hribar et al. (2017)), we regress our tone measure on a granular set of state-by-year fixed effects and consider the fitted value as *explainable* segment rationalized by the time-varying local economic fundamentals. The residual values from the regression are considered as the *unexplainable* segment or *sentiment* as these are independent of the economic fundamentals. Since the state-by-year fixed effects encompass all macroeconomic variables, this method enables us to derive a bank sentiment measure that is independent of macro-level sentiment measures such as consumers (Carroll (1997)), investors (Baker and Wurgler (2006)), and corporate managers (Jiang et al. (2019)).

More specifically, we estimate the following regression of an annual report's tone on state-time fixed effects:

$$Net\ Negative\ Sentence\ Ratio_{i,t} = \gamma + \rho State_i \times Year_t + \epsilon_{i,t} \tag{2}$$

Our unit of analysis is at a bank-year level, so all macroeconomic variables and yearly-varying variables of banks are absorbed by the fixed effects. Thus, the regression model accounts for the changes in the tone of annual reports driven by all macroeconomic changes such as monetary policy, financial market conditions, industry conditions, consumer conditions, and other macro-level sentiment measures on the demand side (investors, consumers, and corporate managers).

We use the residuals of the estimated regression as *unexplainable* segment, and we call it *unexplainable negative belief* (*Neg-BankSentiment*), which is our main independent variable of bank sentiment. By construction, a higher *Neg-BankSentiment* indicates that the sentiment measured in the annual reports of a bank is more negative above and beyond the level rationalized by the economic fundamentals and the demand-side sentiment.

3.2. Other Variables and Summary Statistics

We obtain variables of banks from the Compustat Bank database. Compustat Bank provides annual accounting data on banks[3] and the identifier of the banks, the Central Index Key (CIK), that we use for merging the Compustat Bank database to the variables we construct from 10-K reports in the SEC's EDGAR system. We remove bank observations in the year of the changes in the fiscal year-end date to avoid duplicate observations. Our final sample consists of 9,290 bank-year observations from 1995 to 2019. We select this sample period because the first year with available tier 1 capital ratio is 1995, and in 2020, most large banks in the U.S. adopted a new accounting standard for loan loss provision, the Current Expected Credit Loss (CECL).

We calculate the main dependent variable, $\textit{Loan Loss Provision}_t$, as the amount of provision for loan losses ("pll" in Compustat) normalized by the lagged amount of total loans ("lntal" in Compustat). In Panel A of Table 1, we report the summary statistics of our variables of interest. $\textit{Loan Loss Provision}_t$ has a mean of 0.6% with a standard deviation of 0.9%. That is, banks set the provision about 0.6% of total loans.

Panel A also reports the summary statistics of the bank sentiment measures, which are our main independent variables, constructed as above. *Neg-BankSentiment*, our main measure of bank sentiment has a mean of -0.001, and a standard deviation of 0.024. We also construct additional sentiment measures focusing on either the positive or negative part of the sentiment. $\textit{BankSentiment\_OnlyNegative}_t$ is defined as the fitted residual from the regression of (2) using *Negative Sentence Ratio* as the dependent variable, which has a mean of 0.000 with a standard deviation of 0.019. Similarly, $\textit{BankSentiment\_OnlyPositive}_t$ is defined as the residual from the regression of (2) using *Positive Sentence Ratio* as the dependent variable, which has a mean of 0.001 with a standard deviation of 0.019.

[3] We follow the bank definition in the Compustat Bank database, which uses firms' SIC codes. Compustat Bank defines a firm as a bank if the SIC code of a firm is one of the following: 6020 (Commercial banks), 6021 (National commercial banks), 6022 (State commercial banks), 6029 (NEC commercial banks), 6035 (Saving institutions, Fed-chartered), and 6036 (Savings institutions, not Fed-charted).

The level of loan loss provision can be affected by both the fundamental status of loans and the discretionary decision by bank management. To control for the effect of fundamental loan status, we include the realized net charge-offs in the future as a control variable in all our regression analyses. We calculate *Net Charge-offs*$_{t+1}$ as the future amount of net charge-offs ("nco" in Compustat) normalized by the lagged amount of total loans.[4] Panel A of Table 1 reports that *Net Charge-offs*$_{t+1}$ has a mean of 0.5% with a standard deviation of 0.8%. That is, on average, about 0.5% of bank loans are net charged off in the following year. Note that the average amount of net charge-offs matches the average amount of loan loss provisions. We also include other bank characteristics that might affect the level of banks' loan loss provisions. *Chg. in Non-performing Loans*$_{t-1}$ is the change of non-performing loans (NPLs) from year *t-2* to year *t-1* and *Chg. in Non-performing Loans*$_t$ is the change of NPLs from year *t-1* to year *t*. On average, we find 0.1% growth in NPLs in our sample.

$1_{Size=Middle}$ is a dummy variable that equals 1 if the gross total assets (GTA) at year *t-1* is greater than \$1B and smaller than or equal to \$3B, where GTA is defined as the sum of total assets and the allowance for loan and the lease losses. $1_{Size=Large}$ is a dummy variable that equals 1 if GTA at year *t-1* is greater than \$3B. About 30% of banks in our sample are in the middle size range and another 30% of banks are in the large size range. *Chg. in Total Loans*$_t$ is the growth rate of total loans from year *t-1* to year *t*. In our sample period, average banks show 11.4% growth in total loan size.

Earnings Before Provision$_t$, the amount of earnings before provision at year *t* scaled by total loans at year *t-1*, is included in our analysis to account for earning management incentives. Panel A of Table 1 reports that it has a mean of 2.5% and a standard deviation of 1.6%. *Tier 1 Capital Ratio*$_{t-1}$ is the ratio of core tier 1 capital to its risk-weighted assets at year *t-1*, included to account for capital management incentives. The average tier 1 capital ratio is 12.1% with a standard deviation of 3.5%. *Loan Loss Reserve*$_{t-1}$ is the amount of allowances for loan losses at year *t-1*, scaled by total loans at year *t-1*. The average loan loss reserve is

[4]The raw variable of net charge-offs in Compustat Bank is negatively signed when losses exceed recoveries. We multiply -1 to the variable so that higher values correspond to larger net charge-offs.

1.4% of the total loans outstanding.

Panel B and Panel C report the summary statistics for the variables in our analysis of bank lending. In Panel B, the sample consists of 1,018 banks from 1995 to 2019. The dependent variable for analysis on the extensive margin of bank lending is *Loan Growth*, which is the ratio of new credit extended in year *t+1* to the total amount of loans. We additionally include the factors that might affect banks' lending. *Deposits* is the total customer deposits at *t-1* scaled by total asset. *Net Income* is net income of a bank at year *t-1* scaled by total asset.

In Panel C, the sample consists of 30 lead banks and 2,948 borrowers (firms) in DealScan from 1998 to 2016. The dependent variable for analysis on the intensive margin of bank lending is *Credit Spread*, which is the annual interest-only spread paid over LIBOR by a firm to a bank at origination year *t+1*. We additionally include loan-specific characteristics in the analysis. *Maturity* is the maturity of the facility in months. $1_{Loantype=Line\ of\ Credit}$ is a dummy variable that equals 1 if the loan type of the facility is the line of credit. *Facility Amount* is the amount of the facility in million U.S. dollars. *Borrower's Cash* is the cash plus short-term investment of a firm at year t scaled by total asset. *Borrower's Long-term Debt* is the total long-term debt of a firm at year t scaled by total asset. *Borrower's Tangible Asset* is the net property, plant, and equipment of a firm at year t scaled by total asset.

A detailed description of the variables is in Appendix Table A. All continuous variables are winsorized at the 1% and 99% levels.

4. Empirical Results

4.1. The effect of bank sentiment on loan loss provision

To examine the effect of bank sentiment on loan loss provision, we estimate the following regression model:

$$Loan\ Loss\ Provision_{i,t} = \alpha + \beta Neg\text{-}BankSentiment_{i,t} + \Gamma \cdot X_{i,t} + \eta_i + \tau_t + \epsilon_{i,t}, \qquad (3)$$

where i indexes a bank and t indexes year. The dependent variable, *Loan Loss Provision*, is the loan loss provision normalized by the lagged amount of total loans. *Neg-BankSentiment* is a bank managerial sentiment, defined as the fitted residuals of the regression (2). $X_{i,t}$ is a set of bank-level control variables. Importantly, it includes the realized net charge-offs in the future (*Net Charge-offs*$_{t+1}$) to account for the fundamental-driven loan loss provision because it is intended to buffer loan defaults in the future. We also control the growth of non-performing loans (*Chg. in Non-performing Loans*$_t$; *Chg. in Non-performing Loans*$_{t-1}$), dummy variables for the size of banks ($\mathit{1}_{Size=Middle}$; $\mathit{1}_{Size=Large}$), the growth of total loans (*Chg. in Total Loans*$_{i,t}$), which can possibly influence provisions for loan losses. *Earnings Before Provision*$_t$ and *Tier 1 Capital Ratio*$_{t-1}$ are also included in the model to control for incentives of managers for earning management and capital management. Finally, the lagged allowance for loan losses (*Loan Loss Reserve*$_{t-1}$) is included to account for the possibility that if banks recognize sufficiently high provisions in the past, then the current provisions for loan loss may be lower (Beatty and Liao (2014)). η_i represents bank fixed effects and τ_t represents year fixed effects.

Our main independent variable (*Neg-BankSentiment*) is the estimated residuals from the regression model (2), and hence, standard errors may not be correctly estimated by the conventional clustering method. We alternatively adopt a wild bootstrapping method where clusters of residuals are resampled to estimate the standard error (Cameron et al. (2008)). We use this bootstrapping method throughout the paper to control for potential bias. We report p-values of the regression coefficients based on the standard errors estimated with bank-level clustering and year-level clustering, bootstrapped with 1,000 iterations.

We present our regression results in Table 2. The variable of interest is *Neg-BankSentiment*, which is the measure of bank managerial sentiment extracted from annual reports (Form 10-K). In an univariate regression model (Column (1)), the estimated coefficient on *Neg-BankSentiment* is positive and statistically significant, implying that a bank manager with negative sentiment makes more provisions than their counterparts. In Column (2), we additionally control for variables related to loan losses. Importantly, we control for the future charge-off (*Net Charge-offs*$_{t+1}$), which the loan loss provision is supposed to cover. By doing so, we can separate out the marginal effect of bank sentiment on the loan loss provision, not explained by the fundamental-based reasons. In addition, bank managers may use past and current information about non-performing loans when they estimate the expected level of loss recognition. Thus, we additionally control for the past and current non-performing loan ratios (*Chg. in Non-performing Loans*$_{t-1}$ and *Chg. in Non-performing Loans*$_{t}$). The estimated coefficient on *Neg-BankSentiment* remains positive and statistically significant.

In Column (3) of Table 2, we add more control variables related to the level of regulatory scrutiny proxied by bank size, lending growth, earnings, and capital ratio. In Column (4), we additionally include the previous level of loan loss reserves to control for a potential confounding effect from a bank's target loan loss reserves level. We still observe a positive and statistically significant coefficient on *Neg-BankSentiment.*

Overall, the result in Table 2 supports our main hypothesis that banks with more negative sentiment are likely to set aside more capital as loan loss provision than the level warranted by their key fundamental economic conditions.

In Table 3, we include an interaction term between bank size and the bank sentiment measure (*Neg-BankSentiment*) to check if the sentiment effect is more pronounced in a subgroup of banks. The estimated coefficient on the main variable of interest remains positive and statistically significant. However, the estimated coefficients on the interaction terms are not statistically significant, implying that the sentiment-driven loan loss provision is observed across all size groups of banks.

Importantly, we test whether the effect of bank sentiment on loan loss provision is more pronounced during recessionary periods (*Hypothesis 2*). Table 4 reports the estimated regression model of (3) with the interaction terms between *Neg-BankSentiment* and *Recessions*, a binary variable equal to one for the NBER recessions and zero otherwise. We find that the estimated coefficient on the interaction term (*Neg-BankSentiment* $\times$ *Recessions*) is positive and statistically significant, implying that banks with more negative sentiment increase their loan loss provisions during recessionary periods. This result suggests bank sentiment can be a potential driver of the counter-cyclical loan loss provisions, which is well documented in the literature (Beatty and Liao (2011); Bushman and Williams (2012); Basel Committee on Banking Supervision (2021)). During recession periods, the effect of negative sentiment on loan loss provisions is more pronounced, and banks set aside more capital as a buffer against loan losses, which can affect their lending behavior.

4.2. Instrumental Variable Analysis

There can be potential endogeneity concerns regarding the *Neg-BankSentiment*. The sentiment measure could be affected by other unobservable economic conditions, which could be correlated with the loan loss provisions.

To address these concerns, we use exogenous local weather conditions near the bank headquarters as instrument variables for bank sentiment. Prior research shows that exposure to inclement weather can have a long-lasting effect on a human being's emotional state (Cunningham (1979); Schwarz and Clore (1983); Lerner et al. (2015); Kamstra et al. (2003)). The recent literature also finds that weather-induced sentiment can influence key corporate decisions such as investment and hiring (Chhaochharia et al. (2019); Zolotoy et al. (2019)). Professional workers are also influenced by weather-related sentiment, such as bank loan officers (Cortés et al. (2016)) and professional stock investors (Saunders (1993); Hirshleifer and Shumway (2003)). Thus, the relevance condition is likely to hold—we test it below. And the exclusion restriction condition would reasonably hold. Our sample is of large public banks with business exposures to diverse geographical locations, and their loan loss provision cannot be solely driven by weather-related local economic conditions.

We obtain a broad set of weather information from the Integrated Surface Data-Lite (ISD-Lite) database maintained by the National Oceanic and Atmospheric Administration (NOAA). The ISD-Lite database offers a wide range of weather information at an hourly frequency per weather station: air temperature, dew point temperature, sea level pressure, wind direction, wind speed rate, sky condition total coverage, and liquid precipitation depth dimension (one- and six-hour duration). We observe hourly weather measures for each of the 2,358 U.S. weather stations from 1995 to 2019. We exclude six-hour precipitation and wind direction because six-hour precipitation is often redundant with one-hour precipitation.

For each weather station, we count the number of instances in a quarter where a specific type of extreme weather event occurs consecutively more than 10 times. These events include extreme cloudy days (Okta above 7), extreme heat days (temperature above 30°C), and rainy days. We primarily focus on the prolonged lack of sunlight (i.e., consecutive cloudy days) because there are evidences that sunlight affects the emotional state of decision-makers (e.g., Kamstra et al. (2003)). We de-seasonalize the weather variables by calculating their differences from the same quarter over the last year, thereby capturing "unanticipated" weather shocks. As a result, we obtain six types of inclement weather variables for each weather station.

We match each bank's headquarters with its neighboring weather stations. We define "local" weather conditions as the average value of inclement weather conditions observed by weather stations located within a 50 km radius of the bank headquarters.[5] To reduce the effect of potential outliers, all weather conditions are winsorized at the 1% and 99% levels. Then, we create dummy variables for each weather variable to capture the non-linear effects of weather conditions (Gilchrist and Sands (2016), Berger et al. (Forthcoming)). We create dummy variables with 1 instance bin for each of the weather variables. In total, we construct 46 potential dummy variables as potential instrument variables for the bank management sentiment.

To avoid overfitting and data-mining concerns, we implement the LASSO procedure (Least Absolute Shrinkage and Selection Operator) for selecting the best instrumental variable out of the 46 candidates (Belloni et al. (2011); Gilchrist and Sands (2016)). LASSO provides a principled search for instruments and offers well-performing results compared to other robustness

[5]About 70% (6,416 out of 9,290) of bank headquarters are matched with local weather stations.

procedures for instrumental variables. We find that the LASSO-chosen instrumental variable is the seasonally-adjusted cloudiness variable, indicating two or three additional instances of 10 consecutive cloudy days during a quarter compared to the last year. This choice is largely consistent with prior studies using cloud coverage as a driver of sentiment (e.g., Goetzmann et al. (2015); Chhaochharia et al. (2019); Kamstra et al. (2003)).

Table 5 reports the first-stage (Column (1)) and the second-stage regressions (Column (2)). The LASSO-chosen instrumental variable is highly correlated with the bank sentiment measure, satisfying the relevance condition of IV. In Column (2), the second-stage regression shows that the negative bank sentiment instrumented by the weather conditions increases the loan loss provision. Because the weather condition is unlikely to influence a bank's loan loss provision through channels other than the sentiment effect, Table 5 allows a causal interpretation of the effect of negative bank sentiment on banks' loan loss provision.

4.3. Robustness Tests

As the bank sentiment is an unobservable construct, there is no definite way to measure it. We conduct additional robustness checks using two alternative textual analysis techniques to measure the bank sentiment.

First, we fine-tune the pre-trained *GPT* model for the sentiment classification (see the Online Appendix for the detailed fine-tuning process) and classify the sentiment of all sentences in annual reports using the fine-tuned *GPT* model. We then re-construct the bank sentiment measure as we do with the *FinBERT* model.

Second, we employ the traditional bag-of-words approach that utilizes the word list of Loughran and McDonald (2011) (LM dictionary). It is a widely adopted approach in the finance and accounting literature (Rogers et al. (2011), Ertugrul et al. (2017), Engelberg et al. (2012), Berger et al. (Forthcoming)). We classify a sentence as a negative sentence if it contains more negative words than positive words. A positive sentence is similarly defined. We re-construct the bank sentiment following the same procedure as above to mitigate the time-varying economic fundamentals and macro shocks.

In our robustness check tests, we replace the *FinBERT*-based bank sentiment measure with the GPT- and LM-based sentiment measures in the baseline regression model (3). Table 6 reports the result. We continue to observe a positive and statistically significant coefficient on the alternative bank sentiment measures.

In an additional analysis, we narrow down the part of the annual reports for textual analysis, focusing on the MD&A section. Although the whole 10-K gives more comprehensive and holistic information about the sentiment of bank managers, the MD&A section can be considered as a more relevant part for bank managers to discuss their future conditions and economic fundamentals. We build the sentiment measure from the textual information from the MD&A section only and replicate the main regression results. Table 7 reports the result. We still find that the negative bank sentiment increases the loan loss provisions.

Instead of the net negative bank sentiment (*Neg-BankSentiment*), we can use the positive sentiment and negative sentiment separately using *BankSentiment_OnlyNegative* and *BankSentiment_OnlyPositive*. We report the results in Appendix Table B.1. The specification in Panel A is similar to Table 2 but replaces *Neg-BankSentiment* with *BankSentiment_OnlyNegative* and *BankSentiment_OnlyPositive*. We find that the negative bank sentiment increases the loan loss provisions but the positive bank sentiment decreases the loan loss provisions. Panel B reports the results using the positive and negative sentiment measures constructed from the fine-tuned *GPT* model and Loughran and McDonald (2011) dictionary, and Panel C reports the results using the *FinBERT*-classified positive and negative sentiment measures constructed from the MD&A section only. Our results remain the same.

4.4. Bank Lending and Sentiment-Driven Loan Loss Provision

A crucial question is whether sentiment-driven loan loss provision impacts the real economy via a lending channel (Basel Committee on Banking Supervision (2021)). To answer this question, we empirically test whether a bank with a higher level of sentiment-driven LLP reduces its loans in the future. To do this, we first derive a sentiment-driven measure based on the estimated regression model of loan loss provisions on bank sentiment and other controls (Equation 3). We use its expected values as the sentiment-driven LLP (*Sentiment-Driven LLP*). We then estimate a regression model of loan growth on *Sentiment-Driven LLP*. We also include the bank sentiment measure (*Neg-BankSentiment*) in the regression model to control for the direct effect of bank sentiment on bank lending.

$$\begin{aligned} \textit{Loan Growth}_{i,t+1} = \alpha + \beta_1 \textit{Sentiment-Driven LLP}_{i,t} + \beta_2 \textit{Neg-BankSentiment}_{i,t} \\ + \Gamma \cdot X_{i,t} + \eta_i + \tau_t + \epsilon_{i,t} \end{aligned} \tag{4}$$

where i, t indexes for a bank and year, respectively.

Table 8 reports the estimated coefficients of the regression model 4. The estimated coefficient on the sentiment-driven LLP is negative and statistically significant. The result holds across all specifications with different sets of control variables. Combined with the counter-cyclicality of the loan loss provision in Table 4 (i.e., more sentiment-driven over-provision during recessions), the negative effect of the sentiment-driven LLP on bank lending implies that the banks with negative sentiment reduce their lending more during recessions. In other words, the sentiment-driven loan loss provision can be a potential driver of the pro-cyclical lending behavior of banks, an important concern of banking regulators.

To substantiate that the reduced lending is driven by the supply side of credits (i.e., banks), we check if the loan pricing increases following the sentiment-driven LLP increases. To the extent that the reduced lending is driven by the supply side (demand side), the sentiment measure should be positively (negatively) correlated with the credit spread. Using loan pricing data from DealScan, we estimate a regression of credit spread on sentiment-driven LLP, controlling for the direct effect of bank sentiment.

$$\begin{aligned} \textit{Credit Spread}_{i,j,t+1} = \alpha + \beta_1 \textit{Sentiment-Driven LLP}_{i,t} + \beta_2 \textit{Neg-BankSentiment}_{i,t} \\ + \Gamma \cdot X_{i,t} + \eta_i + \tau_t + \epsilon_{i,t}, \end{aligned} \tag{5}$$

where i,j,t indexes for a bank (specifically, lead bank), a borrowing firm, and year, respectively. We also control for the features of loan facilities, capital ratio of lead banks and borrower characteristics.

Table 9 reports the results. We find that the estimated coefficient on the sentiment-driven LLP is positive and statistically significant, suggesting that the reduced bank lending is mainly driven by banks rather than the reduced demand of borrowers.

We check the robustness of the results using our alternative sentiment measures constructed from the fine-tuned *GPT* model and Loughran and McDonald (2011) dictionary. The results are in Appendix Table B.2—Panels A and B replicate Tables 8 and Table 9 with the *GPT* model, respectively. Panels C and D replicate Tables 8 and 9 by using Loughran and McDonald (2011) dictionary. We find that our results are robust.

5. Conclusion

This paper studies whether bank management sentiment influences the level of loan loss provisions, a buffer against future loan losses. To answer this question, we use a bank sentiment measure independent of key economic fundamentals and other economic agents' sentiments, utilizing various large-language models (LLMs) applied to the annual reports of all banks in the U.S. We document that banks with more negative sentiment are likely to increase their loan loss provisions. This result holds even after we control for the actual net charge-offs in the future, implying that the sentiment-driven loan loss provision can significantly deviate from the fundamental-based level. To address endogeneity concerns on the sentiment measure, we use exogenous weather conditions near bank headquarters as instrumental variables and find that the negative bank sentiment increases loan loss provisions.

We find that the sentiment effect on loan loss provisions is more pronounced during economic downturns; when swayed by sentiment, banks can be overly conservative in setting their loan loss provisions during recessions. Importantly, we find that banks with more sentiment-driven loan provisions reduce their lending in the future. These results suggest that bank sentiment can be an important channel through which the banks can amplify the business cycle instead of absorbing the shocks.

The accounting standard for estimating loan loss provision has been in transition from the incurred loss (IL) model to the current expected credit loss (CECL) model since January 1, 2020.[6] These changes give more discretion to managers. The rationale is to mitigate the "too little, too late" problem under the IL model regime. Policymakers and regulators posit that the "backward-looking" practice of loan loss recognition under the IL model contributes to the pro-cyclicality of bank lending, resulting in excessive economic growth during upturns and deeper recessions during downturns (The Financial Stability Forum (2009); U.S. Treasury (2009)). However, the new rule can actually worsen the pro-cyclicality of bank lending when bankers' sentiment amplifies the counter-cyclicality of loan loss provision as this paper documents. In this paper, unfortunately, we cannot directly test whether the sentiment-driven effect on loan loss provision has been exacerbated under the ECL standard due to a limited sample period since 2020. There are some recent studies showing the effectiveness of the new accounting standard (Chen et al. (2023); Kim et al. (2023)), but future research with longer sample periods will offer a more comprehensive analysis.

[6]In the IL standard, loan losses are recognized only after loss events have occurred prior to the reporting date that are likely to result in future non-payment of loans. This accounting standard does not allow loss recognition of future expected losses based on economic trends suggestive of additional future losses. Under the CECL standards, however, banks are required to recognize loan losses projected not to be repaid in the future.

References

Acharya, V., and H. Naqvi. 2012. The seeds of a crisis: A theory of bank liquidity and risk taking over the business cycle. *Journal of Financial Economics* 106:349–366.

Baker, M., and J. Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance* 61:1645–1680.

Basel Committee on Banking Supervision. 2021. The procyclicality of loan loss provisions: A literature review. Working Paper 29, Bank for International Settlements.

Bassi, A., R. Colacito, and P. Fulghieri. 2013. 'O Sole Mio: An experimental analysis of weather and risk attitudes in financial decisions. *The Review of Financial Studies* 26:1824–1852.

Batchelor, R., and P. Dua. 1998. Improving macro-economic forecasts: The role of consumer confidence. *International Journal of Forecasting* 14:71–81.

Beatty, A., S. L. Chamberlain, and J. Magliolo. 1995. Managing financial reports of commercial banks: The influence of taxes, regulatory capital, and earnings. *Journal of Accounting Research* 33:231–261.

Beatty, A., and S. Liao. 2011. Do delays in expected loss recognition affect banks' willingness to lend? *Journal of Accounting and Economics* 52:1–20.

Beatty, A., and S. Liao. 2014. Financial accounting in the banking industry: A review of the empirical literature. *Journal of Accounting and Economics* 58:339–383. 2013 Conference Issue.

Belloni, A., V. Chernozhukov, and C. Hansen. 2011. LASSO methods for gaussian instrumental variables models.

Berger, A. N., H. H. Kim, and X. F. Ma. Forthcoming. Bank sentiment and liquidity hoarding. *Contemporary Accounting Research* .

Brown, N. C., T. E. Christensen, W. B. Elliott, and R. D. Mergenthaler. 2012. Investor sentiment and pro forma earnings disclosures. *Journal of Accounting Research* 50:1–40.

Brown, S. V., and J. W. Tucker. 2011. Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research* 49:309–346.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33:1877–1901.

Bushman, R. M., and C. D. Williams. 2012. Accounting discretion, loan loss provisioning, and discipline of banks' risk-taking. *Journal of Accounting and Economics* 54:1–18.

Bushman, R. M., and C. D. Williams. 2015. Delayed expected loss recognition and the risk profile of banks. *Journal of Accounting Research* 53:511–553.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics* 90:414–427.

Campbell, J. L., H. Chen, D. S. Dhaliwal, H.-m. Lu, and L. B. Steele. 2014. The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies* 19:396–455.

Carroll, C. D. 1997. Buffer-stock saving and the life cycle/permanent income hypothesis. *The Quarterly Journal of Economics* 112:1–55.

Carroll, C. D., J. C. Fuhrer, and D. W. Wilcox. 1994. Does consumer sentiment forecast household spending? If so, why? *The American Economic Review* 84:1397–1408.

Chen, J., Y. Dou, S. G. Ryan, and Y. Zou. 2023. Does the Current Expected Credit Loss Approach Decrease the Procyclicality of Banks' Lending? Evidence from the COVID-19 Recession. *SSRN Electronic Journal* 56 Pages Posted: 18 May 2022 Last revised: 12 Jul 2023.

Chhaochharia, V., D. Kim, G. M. Korniotis, and A. Kumar. 2019. Mood, firm behavior, and aggregate economic outcomes. *Journal of Financial Economics* 132:427–450.

Cohen, B. H., and G. A. Edwards. 2017. The new era of expected credit loss provisioning. *BIS Quarterly Review* .

Collins, J. H., D. A. Shackelford, and J. M. Wahlen. 1995. Bank differences in the coordination of regulatory capital, earnings, and taxes. *Journal of Accounting Research* 33:263–291.

Cortés, K., R. Duchin, and D. Sosyura. 2016. Clouded judgment: The role of sentiment in credit origination. *Journal of Financial Economics* 121:392–413.

Cunningham, M. R. 1979. Weather, mood, and helping behavior: Quasi experiments with the sunshine samaritan. *Journal of Personality and Social Psychology* 37:1947–1956.

Davis, A., W. Ge, D. Matsumoto, and J. L. Zhang. 2015. The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies* 20:639–673.

Davis, A. K., and I. Tama-Sweet. 2012. Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research* 29:804–837.

Dehaan, E., J. Madsen, and J. D. Piotroski. 2017. Do weather-induced moods affect the processing of earnings news? *Journal of Accounting Research* 55:509–550.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Engelberg, J. E., A. V. Reed, and M. C. Ringgenberg. 2012. How are shorts informed?: Short sellers, news, and information processing. *Journal of Financial Economics* 105:260–278.

Ertugrul, M., J. Lei, J. Qiu, and C. Wan. 2017. Annual Report Readability, Tone Ambiguity, and the Cost of Borrowing. *Journal of Financial and Quantitative Analysis* 52:811–836.

Gilchrist, D. S., and E. G. Sands. 2016. Something to talk about: Social spillovers in movie consumption. *Journal of Political Economy* 124:1339–1382.

Goetzmann, W. N., D. Kim, A. Kumar, and Q. Wang. 2015. Weather-induced mood, institutional investors, and stock returns. *The Review of Financial Studies* 28:73–111.

Hanley, K. W., and G. Hoberg. 2019. Dynamic interpretation of emerging risks in the financial sector. *The Review of Financial Studies* 32:4543–4603.

Hirshleifer, D., and T. Shumway. 2003. Good day sunshine: Stock returns and the weather. *The Journal of Finance* 58:1009–1032.

Horvath, P., and M. Zuckerman. 1993. Sensation seeking, risk appraisal, and risky behavior. *Personality and Individual Differences* 14:41–52.

Hribar, P., S. J. Melessa, R. C. Small, and J. H. Wilde. 2017. Does managerial sentiment affect accrual estimates? Evidence from the banking industry. *Journal of Accounting and Economics* 63:26–50.

Huang, A. H., H. Wang, and Y. Yang. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research* 40:806–841.

Jiang, F., J. Lee, X. Martin, and G. Zhou. 2019. Manager sentiment and stock returns. *Journal of Financial Economics* 132:126–149.

Johnson, E., and A. Tversky. 1983. Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology* 45:20–31.

Kamstra, M. J., L. A. Kramer, and M. D. Levi. 2003. Winter blues: A SAD stock market cycle. *The American Economic Review* 93:324–343.

Keynes, J. M. 1937. The general theory of employment. *The Quarterly Journal of Economics* 51:209–223.

Kim, M.-S., and W. Kross. 1998. The impact of the 1989 change in bank capital standards on loan loss provisions and loan write-offs. *Journal of Accounting and Economics* 25:69–99.

Kim, S., S. Kim, A. Kleymenova, and R. Li. 2023. Current expected credit losses (CECL) standard and banks' information production. *Finance and Economics Discussion Series* 2023-063. Washington: Board of Governors of the Federal Reserve System.

Laeven, L., and G. Majnoni. 2003. Loan loss provisioning and economic slowdowns: Too much, too late? *Journal of Financial Intermediation* 12:178–197. Special Issue for the Federal Reserve Bank of Boston.

Lemmon, M., and E. Portniaguina. 2006. Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies* 19:1499–1529.

Lerner, J. S., Y. Li, P. Valdesolo, and K. S. Kassam. 2015. Emotion and decision making. *Annual Review of Psychology* 66:799–823. PMID: 25251484.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66:35–65.

Ludvigson, S. C. 2004. Consumer confidence and consumer spending. *The Journal of Economic Perspectives* 18:29–50.

Malo, P., A. Sinha, P. Korhonen, J. Wallenius, and P. Takala. 2014. Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology* 65:782–796.

McLean, R. D., and M. Zhao. 2014. The business cycle, investor sentiment, and costly external finance. *The Journal of Finance* 69:1377–1409.

Moyer, S. E. 1990. Capital adequacy ratio regulations and accounting choices in commercial banks. *Journal of Accounting and Economics* 13:123–154.

Muslu, V., S. Radhakrishnan, K. R. Subramanyam, and D. Lim. 2015. Forward-looking MD&A disclosures and the information environment. *Management Science* 61:931–948.

Radford, A., K. Narasimhan, T. Salimans, I. Sutskever, et al. 2018. Improving language understanding by generative pre-training .

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1:9.

Rajan, R. G. 1994. Why bank credit policies fluctuate: A theory and some evidence. *The Quarterly Journal of Economics* 109:399–441.

Rogers, J. L., A. V. Buskirk, and S. L. C. Zechman. 2011. Disclosure Tone and Shareholder Litigation. *The Accounting Review* 86:2155–2183.

Saunders, E. M. 1993. Stock prices and wall street weather. *The American Economic Review* 83:1337–1345.

Schwarz, N., and G. Clore. 1983. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology* 45:513–523.

The Financial Stability Forum. 2009. Report of the financial stability forum on addressing procyclicality in the financial system .

Tokunaga, H. 1993. The use and abuse of consumer credit: Application of psychological theory and research. *Journal of Economic Psychology* 14:285–316.

U.S. Treasury. 2009. Financial regulatory reform: A new foundation .

Wegener, D., and R. Petty. 1994. Mood management across affective states: The hedonic contingency hypothesis. *Journal of personality and social psychology* 66:1034–48.

Wong, A., and B. J. Carducci. 1991. Sensation seeking and financial risk taking in everyday money matters. *Journal of Business and Psychology* 5:525–530.

Wright, W. F., and G. H. Bower. 1992. Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes* 52:276–291.

Zolotoy, L., D. O'Sullivan, and Y. Chen. 2019. Local religious norms, corporate social responsibility, and firm value. *Journal of Banking Finance* 100:218–233.

Zuckerman, M. S. 1984. Experience and desire: A new format for sensation seeking scales. *Journal of Behavioral Assessment* 6:101–114.

Table 1: Summary Statistics

We report the summary statistics of variables used in the analysis. In Panel A, the main sample consists of 1,018 banks from 1995 to 2019. The main dependent variable in the analysis is *Loan Loss Provision*, which is the amount of provision for loan losses at year *t*. *Neg-BankSentiment* measures net negative bank sentiment disclosed by annual reports at year *t*, defined as fitted residuals in (2), using the sentence-level analysis by *FinBERT*. Similarly, *BankSentiment_OnlyNegative* and *BankSentiment_OnlyPositive* measure negative and positive bank sentiment, defined as fitted residuals in (2), respectively. *Net Charge-offs* is the amount of gross charge-offs net of the amount of recoveries at year *t+1*. *Chg. in Non-performing Loans*$_{t-1}$ is the change of non-performing loans (NPLs) from year *t-2* to year *t-1*. *Chg. in Non-performing Loans*$_t$ is the change of NPLs from year *t-1* to year *t*. $1_{Size=Middle}$ is a dummy variable that equals 1 if the gross total assets (GTA) at year *t-1* is greater than \$1B and smaller than or equal to \$3B, where GTA is defined as the sum of total assets and the allowance for loan and the lease losses. $1_{Size=Large}$ is a dummy variable that equals 1 if GTA at year *t-1* is greater than \$3B. *Chg. in Total Loans* is the change in total loans from year *t-1* to year *t*. *Earnings Before Provision* is the amount of earnings before provision at year *t*, which is included to account for earning management incentives. *Tier 1 Capital Ratio* is the ratio of core tier 1 capital to its risk-weighted assets at year *t-1*, included to account for capital management incentives. *Loan Loss Reserve* is the amount of allowances for loan losses at year *t-1*. In Panel B, the sample consists of 1,018 banks from 1995 to 2019. The dependent variable for the analysis on the extensive margin of bank lending is *Loan Growth*, which is the amount of new credit to the economy in year *t+1*. *Deposits* is the total customer deposits at *t-1*. *Net Income* is net income of a bank at year *t-1*. In Panel C, the sample consists of 30 lead banks and 2,948 borrowers (firms) from 1998 to 2016. The dependent variable for the analysis on the intensive margin of bank lending is *Credit Spread*, which is the annual interest spread paid over LIBOR by a firm to a bank at origination year *t+1*. *Maturity* is the maturity of the facility in months. $1_{Loantype=Line\ of\ Credit}$ is a dummy variable that equals 1 if the loan type of the facility is the line of credit. *Facility Amount* is the amount of the facility in million dollars. *Borrower's Cash* is the cash plus short-term investment of a firm at year *t*. *Borrower's Long-term Debt* is the total long-term debt of a firm at year *t*. *Borrower's Tangible Asset* is the net property, plant, and equipment of a firm at year *t*. All variables are defined in the Appendix Table A in more detail.

Panel A: *Loan Loss Provision* Variables	Obs.	Mean	Std. Dev.	25^{th} pct.	Median	75^{th} pct.
Dependent variable						
Loan Loss Provision$_{i,t}$	9,290	0.006	0.009	0.001	0.003	0.006
Main independent variables						
Neg-BankSentiment$_{i,t}$	9,290	-0.001	0.024	-0.015	0.001	0.016
BankSentiment_OnlyNegative$_{i,t}$	9,290	0.000	0.019	-0.011	0.000	0.012
BankSentiment_OnlyPositive$_{i,t}$	9,290	0.001	0.019	-0.011	-0.002	0.010
Control variables						
Net Charge-offs$_{i,t+1}$	9,290	0.005	0.008	0.001	0.002	0.006
Chg. in Non-performing Loans$_{i,t-1}$	9,290	0.001	0.013	-0.003	0.000	0.003
Chg. in Non-performing Loans$_{i,t}$	9,290	0.001	0.014	-0.003	0.000	0.004
$1_{Size=Middle}$	9,290	0.289	0.453	0.000	0.000	1.000
$1_{Size=Large}$	9,290	0.283	0.451	0.000	0.000	1.000
Chg. in Total Loans$_{i,t}$	9,290	0.114	0.184	0.018	0.079	0.163
Earnings Before Provision$_{i,t}$	9,290	0.025	0.016	0.017	0.024	0.032
Tier 1 Capital Ratio$_{i,t-1}$	9,290	0.121	0.035	0.099	0.117	0.138
Loan Loss Reserve$_{i,t-1}$	9,290	0.014	0.008	0.010	0.013	0.017

Panel B: *Bank Lending* (Extensive Margin)						
Variables	Obs.	Mean	Std. Dev.	25^{th} pct.	Median	75^{th} pct.
Dependent variable						
$Loan\ Growth_{i,t+1}$	9,290	0.125	0.211	0.015	0.082	0.178
Additional control variables						
$Deposits_{i,t-1}$	9,290	0.761	0.093	0.710	0.780	0.830
$Net\ Income_{i,t-1}$	9,290	0.007	0.008	0.006	0.009	0.011

Panel C: *Bank Lending* (Intensive Margin)						
Variables	Obs.	Mean	Std. Dev.	25^{th} pct.	Median	75^{th} pct.
Dependent variable						
$Credit\ Spread_{i,j,t+1}$	17,122	210.374	122.512	125.000	190.000	275.000
Additional control variables						
$Maturity_{i,j,t+1}$	17,122	53.376	17.026	43.000	60.000	60.000
$1_{Loantype=Line\ of\ Credit}$	17,122	0.695	0.460	0.000	1.000	1.000
$Borrower's\ Cash_{j,t}$	17,122	0.085	0.106	0.014	0.045	0.115
$Borrower's\ Long\text{-}term\ Debt_{j,t}$	17,122	0.287	0.224	0.119	0.258	0.407
$Borrower's\ Tangible\ Asset_{j,t}$	17,122	0.300	0.255	0.090	0.217	0.469

Table 2: The Effect of Bank Sentiment on Loan Loss Provision

We report the panel regressions estimates of the effect of *Neg-BankSentiment* on *Loan Loss Provision*. We use bank-year observations from 1995 to 2019. The dependent variable is *Loan Loss Provision*, which is the amount of provision for loan losses. The main independent variable is *Neg-BankSentiment*, which measures net negative bank sentiment extracted from annual reports using *FinBERT*. Column (1) reports the univariate results with bank and year-fixed effects. In Column (2), we additionally control the future charge-off (*Net Charge-offs*$_{t+1}$) and the past and current non-performing loan ratios (*Chg. in Non-performing Loans*$_{t-1}$ and *Chg. in Non-performing Loans*$_{t}$). In Column (3), we expand our control to include bank size ($\mathit{1}_{Size=Middle}$ and $\mathit{1}_{Size=Large}$), lending growth (*Chg. in Total Loans*$_{t}$), earnings (*Earnings Before Provision*$_{t}$), and capital ratio (*Tier 1 Capital Ratio*$_{t-1}$). In Column (4), we also control the previous level of loan loss reserves (*Loan Loss Reserve*$_{t-1}$). All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

	(1)	(2)	(3)	(4)
	Dep. Variable = *Loan Loss Provision*$_t$			
Neg-BankSentiment$_t$	0.043***	0.028***	0.023***	0.018***
	(<0.000)	(<0.000)	(<0.000)	(<0.000)
Net Charge-offs$_{t+1}$		0.442***	0.433***	0.407***
		(<0.000)	(<0.000)	(<0.000)
Chg. in Non-performing Loans$_{t-1}$		0.110***	0.107***	0.109***
		(0.002)	(<0.000)	(0.001)
Chg. in Non-performing Loans$_t$		0.034	0.040	0.061*
		(0.219)	(0.156)	(0.050)
$\mathit{1}_{Size=Middle}$			0.000	0.000
			(0.352)	(0.293)
$\mathit{1}_{Size=Large}$			0.001	0.001**
			(0.134)	(0.040)
Chg. in Total Loans$_t$			-0.001	-0.001
			(0.227)	(0.266)
Earnings Before Provision$_t$			-0.044***	-0.039***
			(0.006)	(0.008)
Tier 1 Capital Ratio$_{t-1}$			-0.005	-0.006
			(0.260)	(0.184)
Loan Loss Reserve$_{t-1}$				0.154***
				(0.009)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290

Table 3: Heterogeneity Analysis by the Size of Banks

We report the panel regressions estimates of the effect of *Neg-BankSentiment* on *Loan Loss Provision*, by the size of banks. We use bank-year observations from 1995 to 2019. The dependent variable is *Loan Loss Provision*, which is the amount of provision for loan losses. In Column (1), the main independent variables are *Neg-BankSentiment*, which is the measure of net negative bank sentiment extracted from annual reports using *FinBERT*, and its interactions with two bank-size dummies of $1_{Size=Middle}$ and $1_{Size=Large}$. We include bank and year-fixed effects. In Column (2), we additionally control the future charge-off (*Net Charge-offs*$_{t+1}$) and the past and current non-performing loan ratios (*Chg. in Non-performing Loans*$_{t-1}$ and *Chg. in Non-performing Loans*$_t$). In Column (3), we expand our control to include lending growth (*Chg. in Total Loans*$_t$), earnings (*Earnings Before Provision*$_t$), and capital ratio (*Tier 1 Capital Ratio*$_{t-1}$). In Column (4), we also control the previous level of loan loss reserves (*Loan Loss Reserve*$_{t-1}$). All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

	(1)	(2)	(3)	(4)
	Dep. Variable = *Loan Loss Provision*$_t$			
Neg-BankSentiment$_t$	0.044***	0.029***	0.024***	0.016***
	(<0.000)	(<0.000)	(<0.000)	(0.003)
Neg-BankSentiment$_t$ × $1_{Size=Middle}$	-0.010	-0.007	-0.006	-0.002
	(0.280)	(0.390)	(0.449)	(0.751)
Neg-BankSentiment$_t$ × $1_{Size=Large}$	0.004	0.004	0.003	0.009
	(0.717)	(0.632)	(0.692)	(0.236)
$1_{Size=Middle}$	0.001***	0.000	0.000	0.000
	(0.004)	(0.218)	(0.341)	(0.240)
$1_{Size=Large}$	0.003***	0.001*	0.001	0.001**
	(0.005)	(0.098)	(0.131)	(0.033)
Net Charge-offs$_{t+1}$		0.440***	0.433***	0.407***
		(<0.000)	(<0.000)	(<0.000)
Chg. in Non-performing Loans$_{t-1}$		0.109***	0.107***	0.109***
		(0.002)	(<0.000)	(0.001)
Chg. in Non-performing Loans$_t$		0.034	0.040	0.061*
		(0.218)	(0.155)	(0.050)
Chg. in Total Loans$_t$			-0.001	-0.001
			(0.230)	(0.271)
Earnings Before Provision$_t$			-0.044***	-0.039***
			(0.006)	(0.009)
Tier 1 Capital Ratio$_{t-1}$			-0.005	-0.006
			(0.268)	(0.180)
Loan Loss Reserve$_{t-1}$				0.154***
				(0.008)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290

Table 4: The Effect of Bank Sentiment on Loan Loss Provision during Recessions

We report the panel regressions estimates of the effect of *Neg-BankSentiment* on *Loan Loss Provision*, by the NBER recessionary period. We use bank-year observations from 1995 to 2019. The dependent variable is *Loan Loss Provision*, which is the amount of provision for loan losses. In Column (1), the main independent variables are *Neg-BankSentiment*, which is the measure of net negative sentiment extracted from annual reports using *FinBERT*, and its interactions with a dummy of *Recessions*. We include bank and year-fixed effects. In Column (2), we additionally control the future charge-off ($Net\ Charge\text{-}offs_{t+1}$) and the past and current non-performing loan ratios ($Chg.\ in\ Non\text{-}performing\ Loans_{t-1}$ and $Chg.\ in\ Non\text{-}performing\ Loans_{t}$). In Column (3), we expand our control to include bank size ($1_{Size=Middle}$ and $1_{Size=Large}$), lending growth ($Chg.\ in\ Total\ Loans_t$), earnings ($Earnings\ Before\ Provision_t$), and capital ratio ($Tier\ 1\ Capital\ Ratio_{t-1}$). In Column (4), we also control the previous level of loan loss reserves ($Loan\ Loss\ Reserve_{t-1}$). All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

	(1)	(2)	(3)	(4)
	Dep. Variable = $Loan\ Loss\ Provision_t$			
$Neg\text{-}BankSentiment_t \times Recessions_t$	0.052*	0.029*	0.024*	0.026*
	(0.090)	(0.062)	(0.079)	(0.056)
$Neg\text{-}BankSentiment_t$	0.035***	0.024***	0.019***	0.014***
	(<0.000)	(<0.000)	(<0.000)	(<0.000)
$Net\ Charge\text{-}offs_{t+1}$		0.441***	0.432***	0.406***
		(<0.000)	(<0.000)	(<0.000)
$Chg.\ in\ Non\text{-}performing\ Loans_{t-1}$		0.109***	0.106***	0.108***
		(0.003)	(<0.000)	(0.001)
$Chg.\ in\ Non\text{-}performing\ Loans_t$		0.033	0.040	0.060*
		(0.222)	(0.162)	(0.051)
$1_{Size=Middle}$			0.000	0.000
			(0.337)	(0.278)
$1_{Size=Large}$			0.001	0.001**
			(0.129)	(0.036)
$Chg.\ in\ Total\ Loans_t$			-0.001	-0.001
			(0.234)	(0.275)
$Earnings\ Before\ Provision_t$			-0.043***	-0.038**
			(0.006)	(0.010)
$Tier\ 1\ Capital\ Ratio_{t-1}$			-0.005	-0.006
			(0.265)	(0.189)
$Loan\ Loss\ Reserve_{t-1}$				0.154***
				(0.008)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290

Table 5: Instrumental Variable Analysis

We report instrumental variable analysis of the effect of *Neg-BankSentiment* instrumented by *Cloud Coverage* on *Loan Loss Provision* with the same specifications as in Column (4) of Table 2. We use bank-year observations from 1995 to 2019. The dependent variable is *Loan Loss Provision*, which is the amount of provision for loan losses. The main independent variable is *Neg-BankSentiment*, which is the measure of net negative sentiment extracted from annual reports. The instrumental variable is *Cloud Coverage* near the bank holding company's headquarters, ensuring that the weather station is located within 50km radius of the headquarters each year. Column (1) reports coefficient estimates from the first-stage regression of *Neg-BankSentiment* on *Cloud Coverage* near the headquarters. Column (2) reports coefficient estimates from the second-stage regression of *Loan Loss Provision* on *Neg-BankSentiment* instrumented by *Cloud Coverage*. All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

	(1)	(2)
Dep. Variable =	$Neg\text{-}BankSentiment_t$	$Loan\ Loss\ Provision_t$
$Cloud\ Coverage_t$	0.005***	
	(0.003)	
$Neg\text{-}BankSentiment_t$		0.237*
		(0.077)
$Net\ Charge\text{-}offs_{t+1}$	0.178***	0.354***
	(<0.000)	(<0.000)
$Chg.\ in\ Non\text{-}performing\ Loans_{t-1}$	0.023	0.102***
	(0.224)	(<0.000)
$Chg.\ in\ Non\text{-}performing\ Loans_t$	0.021	0.057*
	(0.327)	(0.082)
$1_{Size=Middle}$	0.003	-0.000
	(0.141)	(0.319)
$1_{Size=Large}$	0.003	0.000
	(0.249)	(0.806)
$Chg.\ in\ Total\ Loans_t$	-0.010***	0.002
	(<0.000)	(0.100)
$Earnings\ Before\ Provision_t$	-0.188***	-0.004
	(<0.000)	(0.884)
$Tier\ 1\ Capital\ Ratio_{t-1}$	0.018	-0.013***
	(0.202)	(0.009)
$Loan\ Loss\ Reserve_{t-1}$	0.443***	0.028
	(0.001)	(0.711)
F-statistic	15.50	
Bank F.E.	YES	YES
Year F.E.	YES	YES
Observations	6,416	6,416

Table 6: Robustness Tests of Alternative Language Models

We report the panel regressions estimates of the effect of *Loan Loss Provision* on *Neg-BankSentiment*. We use bank-year observations from 1995 to 2019. The dependent variable is *Loan Loss Provision*, which is the amount of provision for loan losses. The main independent variable is *Neg-BankSentiment*, which is the measure of net negative bank sentiment extracted from annual reports by employing alternative language models. We utilize another language model, a fine-tuned *GPT* model for sentence classification tasks, and conventional Loughran and McDonald (2011)'s dictionary-based approach. We construct the main independent variable from the *GPT* model, and we reports the univariate results with bank and year-fixed effects in Column (1). In Column (2), we replicate Column (4) of Table 2 with the *GPT* model for robustness tests. In Column (3), we construct sentiment measure from the Loughran and McDonald (2011) dictionary and report the univariate results with bank and year-fixed effects. In Column (4), we replicate Column (4) of Table 2 with the dictionary-based sentiment measure. All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

Language Model Used	GPT Model		LM Model	
	(1)	(2)	(3)	(4)
	Dep. Variable = *Loan Loss Provision*$_t$			
Neg-BankSentiment$_t$	0.036***	0.016***	0.029***	0.014***
	(0.002)	(0.001)	(0.002)	(<0.000)
Net Charge-offs$_{t+1}$		0.407***		0.407***
		(<0.000)		(<0.000)
Chg. in Non-performing Loans$_{t-1}$		0.109***		0.109***
		(0.001)		(<0.000)
Chg. in Non-performing Loans$_t$		0.061**		0.060*
		(0.049)		(0.050)
$1_{Size=Middle}$		0.000		0.000
		(0.311)		(0.205)
$1_{Size=Large}$		0.001**		0.001**
		(0.042)		(0.039)
Chg. in Total Loans$_t$		-0.001		-0.001
		(0.268)		(0.247)
Earnings Before Provision$_t$		-0.039***		-0.039***
		(0.007)		(0.007)
Tier 1 Capital Ratio$_{t-1}$		-0.006		-0.007
		(0.181)		(0.168)
Loan Loss Reserve$_{t-1}$		0.154***		0.153***
		(0.009)		(0.007)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290

Table 7: Robustness Tests with Bank Sentiment in MD&A Sections Only

We report the panel regressions estimates of the effect of *Loan Loss Provision* on *Neg-BankSentiment*. We use bank-year observations from 1995 to 2019. The dependent variable is *Loan Loss Provision*, which is the amount of provision for loan losses. The main independent variable is *Neg-BankSentiment*, which is the measure of net negative bank sentiment extracted only from the MD&A section of Form 10-K using the *FinBERT* model. Column (1) reports the univariate results with bank and year-fixed effects. In Column (2), we additionally control the future charge-off ($\textit{Net Charge-offs}_{t+1}$) and the past and current non-performing loan ratios ($\textit{Chg. in Non-performing Loans}_{t-1}$ and $\textit{Chg. in Non-performing Loans}_{t}$). In Column (3), we expand our control to include bank size ($\mathit{1}_{Size=Middle}$ and $\mathit{1}_{Size=Large}$), lending growth ($\textit{Chg. in Total Loans}_t$), earnings ($\textit{Earnings Before Provision}_t$), and capital ratio ($\textit{Tier 1 Capital Ratio}_{t-1}$). In Column (4), we also control the previous level of loan loss reserves ($\textit{Loan Loss Reserve}_{t-1}$). All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

	(1)	(2)	(3)	(4)
	Dep. Variable = $\textit{Loan Loss Provision}_t$			
$\textit{Neg-BankSentiment}_t$	0.017***	0.013***	0.010***	0.008***
	(<0.000)	(<0.000)	(<0.000)	(<0.000)
$\textit{Net Charge-offs}_{t+1}$		0.424***	0.422***	0.393***
		(<0.000)	(<0.000)	(<0.000)
$\textit{Chg. in Non-performing Loans}_{t-1}$		0.106***	0.102***	0.105***
		(0.007)	(0.007)	(0.005)
$\textit{Chg. in Non-performing Loans}_{t}$		0.036	0.041*	0.063**
		(0.130)	(0.097)	(0.023)
$\mathit{1}_{Size=Middle}$			0.000	0.000
			(0.407)	(0.252)
$\mathit{1}_{Size=Large}$			0.001	0.001**
			(0.123)	(0.024)
$\textit{Chg. in Total Loans}_t$			-0.002	-0.001
			(0.289)	(0.301)
$\textit{Earnings Before Provision}_t$			-0.040*	-0.036*
			(0.076)	(0.078)
$\textit{Tier 1 Capital Ratio}_{t-1}$			-0.006	-0.007
			(0.146)	(0.126)
$\textit{Loan Loss Reserve}_{t-1}$				0.172***
				(0.006)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	6,743	6,743	6,743	6,743

Table 8: Sentiment-Driven Loan Loss Provision and Bank Lending—Extensive Margin

We report the panel regressions estimates of the effect of *Sentiment-Driven LLP* on *Loan Growth* in the future. We use bank-year observations from 1995 to 2019. The dependent variable is $Loan\ Growth_{t+1}$, which is the amount of new credit to the economy in the future. The main independent variable is *Sentiment-Driven LLP*, which measures the additional loan loss provision due to the *FinBERT*-classified net negative bank sentiment. We also control for net negative bank sentiment for the direct effect. Column (1) reports the univariate results with bank and year-fixed effects. In Column (2), we additionally control the amount of deposits ($Deposits_{t-1}$) and net income ($Net\ Income_{t-1}$). In Column (3), we expand our control to include the past and current non-performing loan ratios ($Chg.\ in\ Non\text{-}performing\ Loans_{t-1}$ and $Chg.\ in\ Non\text{-}performing\ Loans_t$) and bank size ($1_{Size=Middle}$ and $1_{Size=Large}$). In Column (4), we also control the capital ratio ($Tier\ 1\ Capital\ Ratio_{t-1}$). All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

	(1)	(2)	(3)	(4)
	Dep. Variable = $Loan\ Growth_{t+1}$			
$Sentiment\text{-}Driven\ LLP_t$	-9.954***	-9.299***	-10.042***	-9.657***
	(<0.000)	(<0.000)	(<0.000)	(<0.000)
$Neg\text{-}BankSentiment_t$	-0.424***	-0.368**	-0.324**	-0.358**
	(0.005)	(0.013)	(0.024)	(0.012)
$Deposits_{t-1}$		0.148**	0.111	0.143**
		(0.036)	(0.103)	(0.030)
$Net\ Income_{t-1}$		1.742***	1.743***	1.458***
		(<0.000)	(<0.000)	(<0.000)
$Chg.\ in\ Non\text{-}performing\ Loans_{t-1}$			0.730**	0.714**
			(0.015)	(0.021)
$Chg.\ in\ Non\text{-}performing\ Loans_t$			-0.082	-0.120
			(0.663)	(0.537)
$1_{Size=Middle}$			-0.036**	-0.032**
			(0.022)	(0.039)
$1_{Size=Large}$			-0.088***	-0.080***
			(0.001)	(0.001)
$Tier\ 1\ Capital\ Ratio_{t-1}$				0.559***
				(0.002)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290

Table 9: Sentiment-Driven Loan Loss Provision and Bank Lending—Intensive Margin

We report the panel regressions estimates of the effect of *Sentiment-Driven LLP* on *Credit Spread* in the future. We use bank-year observations from 1998 to 2016. The dependent variable is *Credit* $Spread_{i,j,t+1}$, which is the annual interest spread paid by firm j to lead bank i at origination year *t+1*. The main independent variable is *Sentiment-Driven LLP*, which measures the loan loss provision due to the *FinBERT*-classified net negative bank sentiment. We also control for net negative bank sentiment for the direct effect. Column (1) reports the univariate results with bank, firm and year-fixed effects. In Column (2), we additionally control the features of the facility, including the maturity ($Maturity_{i,j,t+1}$), the loan type ($1_{LoanType=Line\ of\ Credit}$), and the amount of the facility (*Facility* $Amount_{i,j,t+1}$). In Column (3), we expand our control to include the capital ratio (*Tier 1 Capital* $Ratio_{t-1}$). In Column (4), we also control the borrower's characteristics, which include cash and short-term investment (*Borrower's* $Cash_{j,t}$), long-term debt (*Borrower's Long-term* $Debt_{j,t}$), and tangible assets (*Borrower's Tangible* $Asset_{j,t}$). All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

	(1)	(2)	(3)	(4)
	Dep. Variable = *Credit* $Spread_{i,j,t+1}$			
Sentiment-Driven $LLP_{i,t}$	2012.646*	1956.949*	2041.702*	1992.697*
	(0.073)	(0.085)	(0.083)	(0.091)
$Neg\text{-}BankSentiment_{i,t}$	152.935	124.229	124.087	115.173
	(0.118)	(0.219)	(0.221)	(0.265)
$Maturity_{i,j,t+1}$		-0.027	-0.028	-0.021
		(0.850)	(0.844)	(0.889)
$1_{LoanType=Line\ of\ Credit}$		-49.468***	-49.472***	-48.526***
		(0.001)	(<0.000)	(0.001)
Facility $Amount_{i,j,t+1}$		-0.014**	-0.014**	-0.014**
		(0.014)	(0.015)	(0.014)
Tier 1 Capital $Ratio_{i,t-1}$			176.420	158.369
			(0.391)	(0.424)
Borrower's $Cash_{j,t}$				9.221
				(0.659)
Borrower's Long-term $Debt_{j,t}$				71.876***
				(0.007)
Borrower's Tangible $Asset_{j,t}$				41.033**
				(0.039)
Bank F.E.	YES	YES	YES	YES
Firm F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	17,122	17,122	17,122	17,122

Appendix Table A: Variable Definitions

Panel A: BHC-level Variables Variable	Description
Loan Loss Provision$_{i,t}$	Loan loss provision (Compustat "pll") at year t, scaled by total loans (Compustat "lntal") at year *t-1*.
Neg-BankSentiment$_{i,t}$	Net negative bank sentiment disclosed by an annual report at year t, defined as the fitted residuals in the regression (2).
Net Negative Sentence Ratio$_{i,t}$	The ratio of classified sentences to total sentences of an annual report, which is calculated as the number of sentences classified as negative minus the number of sentences classified as positive, divided by the total number of sentences at year t.
BankSentiment_OnlyNegative$_{i,t}$	Negative bank sentiment disclosed by an annual report at year t, defined as the fitted residuals in the regression (2).
Negative Sentence Ratio$_{i,t}$	The ratio of classified sentences to total sentences of an annual report, calculated as the number of sentences classified as negative, divided by the total number of sentences at year t.
BankSentiment_OnlyPositive$_{i,t}$	Positive bank sentiment disclosed by an annual report at year t, defined as the fitted residuals in the regression (2).
Positive Sentence Ratio$_{i,t}$	The ratio of classified sentences to total sentences of an annual report, calculated as the number of sentences classified as positive, divided by the total number of sentences at year t.
Net Charge-offs$_{i,t+1}$	Net charge-offs (Compustat "nco") at year *t+1*, scaled by total loans (Compustat "lntal") at year *t-1*. In Compustat, net charge-offs are reported as negative if losses exceed recoveries. We adjust net charge-offs by multiplying -1 so that higher values correspond to larger net charge-offs.
Chg. in Non-performing Loans$_{i,t-1}$	Change in non-performing loans (Compustat "npat") from year *t-2* to year *t-1*, scaled by total loans (Compustat "lntal") at year *t-1*.
Chg. in Non-performing Loans$_{i,t}$	Change in non-performing loans (Compustat "npat") from year *t-1* to year t, scaled by total loans (Compustat "lntal") at year *t-1*.
$1_{Size=Middle}$	An indicator variable that equals one, if the total assets (Compustat "at") at year *t-1* plus the allowance for loan and the lease losses (Compustat "rcl") at year *t-1* is greater than \$1B and smaller than or equal to \$3B.
$1_{Size=Large}$	An indicator variable that equals one, if the total assets (Compustat "at") at year *t-1* plus the allowance for loan and the lease losses (Compustat "rcl") at year *t-1* is greater than \$3B.

Chg. in Total Loans$_{i,t}$	Change in total loans (Compustat “lntal”) from year *t-1* to year *t*, scaled by total loans (Compustat “lntal”) at year *t-1*.
Earnings Before Provision$_{i,t}$	Pre-tax income (Compustat “pi”) at year *t* plus the provision for loan losses (Compustat “pll”) at year *t*, scaled by total loans (Compustat ”lntal”) at year *t-1*.
Tier 1 Capital Ratio$_{i,t-1}$	The ratio of core tier 1 capital to its total risk-weighted assets (Compustat “capr1”) at year *t-1*, normalized by 100.
Loan Loss Reserve$_{i,t-1}$	The allowance for loan and the lease losses (Compustat “rcl”) at year *t-1*, scaled by total loans (Compustat “lntal”) at year *t-1*.
Loan Growth$_{i,t+1}$	Total loans (Compustat “lntal”) at year *t+1* net of total loans at year *t*, scaled by total loans at year *t-1*.
Deposits$_{i,t-1}$	Total customer deposits (Compustat “dptc”) at year *t-1*, scaled by total assets (Compustat “at”) at year *t-1*.
Net Income$_{i,t-1}$	Net income (Compustat “ni”) at year *t-1*, scaled by total assets (Compustat “at”) at year *t-1*.

Panel B: Facility-level Variables

Variable	Description
Credit Spread$_{i,j,t+1}$	Total (fees and interest) annual spread paid over LIBOR for each dollar drawn from the loan (DealScan “AllInDrawn”) in basis point at the loan origination year *t+1*.
Maturity$_{i,j,t+1}$	Maturity of the facility in months (DealScan “Maturity”) at the loan origination year *t+1*.
$1_{Loantype=Line\ of\ Credit}$	An indicator variable if loan type (DealScan “LoanType”) of the facility is “Revolver/Line >= 1 Yr.” or “Revolver/Line < 1 Yr.” at the loan origination year *t+1*.
Facility Amount$_{i,j,t+1}$	The facility amount (DealScan “FacilityAmt”) at the loan origination year *t+1*, scaled by \$1M dollars.

Panel C: Firm-level (Borrower) Variables

Variable	Description
Borrower's Cash$_{j,t}$	Cash and short-term investment (Compustat “che”) at year *t*, scaled by the total assets (Compustat “at”) at year *t*.
Borrower's Long-term Debt$_{j,t}$	Total long-term debt (Compustat “dltt”) at year *t*, scaled by the total assets (Compustat “at”) at year *t*.
Borrower's Tangible Asset$_{j,t}$	Net property, plant, and equipment (Compustat “ppent”) at year *t*, scaled by the total assets (Compustat “at”) at year *t*.

Appendix Table B.1: The Effect of Positive and Negative Bank Sentiment on Loan Loss Provision

We report the panel regressions estimates of the effect of *BankSentiment_OnlyNegative* and *BankSentiment_OnlyPositive* on *Loan Loss Provision*. We use bank-year observations from 1995 to 2019. The dependent variable is *Loan Loss Provision*, which is the amount of provision for loan losses. In Panel A, the main independent variables are *BankSentiment_OnlyNegative* and *BankSentiment_OnlyPositive*, which are the measures of negative bank sentiment and positive bank sentiment, respectively, using the *FinBERT* model. In Panel B, the main independent variables are from alternative language models. We adopt another language model, a fine-tuned *GPT* model for sentence classification tasks, and conventional Loughran and McDonald (2011)'s dictionary-based approach. In Panel C, *BankSentiment_OnlyNegative* and *BankSentiment_OnlyPositive* are extracted only from the MD&A section of annual reports using the *FinBERT* model. All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

Panel A: Positive and Negative Sentiments (FinBERT)				
	(1)	(2)	(3)	(4)
	Dep. Variable = *Loan Loss Provision*$_t$			
BankSentiment_OnlyNegative$_t$	0.047***	0.033***	0.027***	0.020***
	(<0.000)	(<0.000)	(<0.000)	(<0.000)
BankSentiment_OnlyPositive$_t$	-0.039***	-0.024***	-0.019***	-0.016***
	(0.002)	(0.002)	(<0.000)	(<0.000)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290

(Cont'd) Appendix Table B.1: The Effect of Positive and Negative Bank Sentiment on Loan Loss Provision

Panel B: Positive and Negative Sentiments (Alternative Language Model)				
Language Model Used	GPT Model		LM Model	
	(1)	(2)	(3)	(4)
	Dep. Variable = $Loan\ Loss\ Provision_t$			
$BankSentiment\_OnlyNegative_t$	0.044***	0.017***	0.037***	0.020***
	(0.002)	(0.002)	(0.001)	(<0.000)
$BankSentiment\_OnlyPositive_t$	-0.030***	-0.014***	-0.009	-0.002
	(0.003)	(0.003)	(0.273)	(0.677)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290

Panel C: Positive and Negative Sentiments (MD&A)				
	(1)	(2)	(3)	(4)
	Dep. Variable = $Loan\ Loss\ Provision_t$			
$BankSentiment\_OnlyNegative_t$	0.020***	0.017***	0.014***	0.010***
	(0.002)	(0.001)	(<0.000)	(0.003)
$BankSentiment\_OnlyPositive_t$	-0.014**	-0.008	-0.006	-0.005
	(0.018)	(0.103)	(0.146)	(0.258)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	6,743	6,743	6,743	6,743

Appendix Table B.2: Robustness Test of Sentiment-Driven LLP and Bank Lending—Extensive Margin and Intensive Margin

We report the panel regressions estimates of the effect of *Sentiment-Driven LLP* on *Loan Loss Provision*. We use bank-year observations from 1995 to 2019 in the extensive margin analysis (Panel A and Panel C). In Panel A and Panel C, the dependent variable is *Loan Growth*, which is the amount of new credit to the economy in the future. We use bank-firm-year observations from 1998 to 2016 in the intensive margin analysis (Panel B and Panel D). In Panel B and Panel D, the dependent variable is *Credit Spread*, which is the annual interest spread paid by firm j to lead bank i at origination year $t+1$. The main independent variable is *Sentiment-Driven LLP*, which measures the additionally provisioned amount for loan losses due to the net negative bank sentiment. For Panel A and Panel B, we construct bank sentiment from the *GPT* model, and we construct bank sentiment from the Loughran and McDonald (2011)'s dictionary-based approach for Panel C and Panel D. All continuous variables are winsorized at the 1% and 99% levels. Coefficient estimates are reported with p-values in parentheses based on standard errors with bank cluster and year cluster bootstrapping with 1,000 iterations, with ***, **, and * respectively denoting statistical significance at the 1%, 5%, and 10% levels.

Panel A: Extensive Margin				
Language Model Used	GPT Model			
	(1)	(2)	(3)	(4)
	Dep. Variable = $Loan\ Growth_{t+1}$			
$Sentiment\text{-}Driven\ LLP_t$	-9.895***	-9.266***	-9.932***	-9.553***
	(<0.000)	(<0.000)	(<0.000)	(<0.000)
$Neg\text{-}BankSentiment_t$	-0.491***	-0.432***	-0.383***	-0.410***
	(0.001)	(0.001)	(0.002)	(<0.000)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290

Panel B: Intensive Margin				
Language Model Used	GPT Model			
	(1)	(2)	(3)	(4)
	Dep. Variable = $Credit\ Spread_{i,j,t+1}$			
$Sentiment\text{-}Driven\ LLP_{i,t}$	2214.647*	2105.628*	2186.550*	2140.904*
	(0.059)	(0.071)	(0.062)	(0.071)
$Neg\text{-}BankSentiment_{i,t}$	49.880	48.594	52.133	42.096
	(0.617)	(0.640)	(0.626)	(0.685)
Bank F.E.	YES	YES	YES	YES
Firm F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	17,122	17,122	17,122	17,122

(Cont'd) Appendix Table B.2: Robustness Test of Sentiment-Driven LLP and Bank Lending—Extensive Margin and Intensive Margin

Panel C: Extensive Margin				
Language Model Used	LM Model			
	(1)	(2)	(3)	(4)
	Dep. Variable = *Loan Growth*$_{t+1}$			
Sentiment-Driven LLP$_{i,t}$	-9.977***	-9.300***	-10.048***	-9.651***
	(<0.000)	(<0.000)	(<0.000)	(<0.000)
Neg-BankSentiment$_{i,t}$	-0.358***	-0.341***	-0.305***	-0.334***
	(<0.000)	(<0.000)	(<0.000)	(<0.000)
Bank F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	9,290	9,290	9,290	9,290
Panel D: Intensive Margin				
Language Model Used	LM Model			
	(1)	(2)	(3)	(4)
	Dep. Variable = *Credit Spread*$_{i,j,t+1}$			
Sentiment-Driven LLP$_{i,t}$	2325.019**	2208.561**	2294.854**	2226.045**
	(0.041)	(0.042)	(0.040)	(0.041)
Neg-BankSentiment$_{i,t}$	3.012	3.774	1.143	2.221
	(0.943)	(0.917)	(0.981)	(0.957)
Bank F.E.	YES	YES	YES	YES
Firm F.E.	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES
Observations	17,122	17,122	17,122	17,122

Online Appendix: Fine-tuning GPT for Sentiment Classification

Overview of the GPT classifier

Classifying a sentence's sentiment with the large language model (LLM) consists of two steps. The first step is to "pre-train" a model on an extensive dataset of general languages, and the second step is to "fine-tune" the pre-trained language model for a specific task. The pre-training process enables the model to learn general patterns and characteristics of the language dataset. Leveraging the learned patterns, a language model can be "fine-tuned" for a smaller and specific task-related dataset. In other words, the pre-trained language model is suited for general tasks and the fine-tuning process is required to adapt it for a specific task such as sentence classification, question/answering, text summarization, or translation. Using a pre-trained model saves time and resources than training a model from scratch on a large dataset.

GPT (Generative Pre-trained Transformer) is one of the pre-trained language models. During the pre-training step, the GPT model learns syntactic and semantic relations from a large corpus of texts by maximizing the likelihood of the next word in a sequence (Radford et al. (2018); Radford et al. (2019); Brown et al. (2020)). More formally, the objective function to be maximized in the pre-training is as follows:

$$L_1(\mu_1, ..., \mu_n; \Theta) = \Sigma_i logP(\mu_i|\mu_{i-k}, ..., \mu_{i-1}; \Theta) \tag{6}$$

where μ_i is an i^{th} token from unsupervised corpus and Θ is learnable parameters in a neutral network.

The GPT itself is an unsupervised model, predicting only general patterns of sentences. To tailor the general model to our specific task of sentiment classification, we further need to "fine-tune" the GPT model on a smaller and task-specific dataset. Because we are interested in sentiment analysis, we fine-tune the GPT model for sentiment-based sentence classification in the finance context. During fine-tuning, the model makes minor adjustments to its "hyperparameters" (i.e., task-specific parameters), leveraging the parameters learned during the pre-training phase.

Fine-tuning GPT for the sentiment classification

We select the open-sourced GPT-2 model by OpenAI (available on the Hugging Face library). With the pre-trained GPT model, we fine-tune the model for the sentiment classification using the dataset of labeled sentences provided by Malo et al. (2014). Malo et al. (2014) make sentence data pool from English news articles on all listed companies in the OMX Helsinki index. 10,000 articles are randomly sampled from the pool, correcting biases from the company size, industry, and news sources. From the selected articles, about 5,000 sentences are randomly chosen to represent the overall news database. The selected sentences are labeled by 16 annotators. Each annotator manually classifies about 1,500 sentences into positive, negative, or neutral sentiment. For each sentence, five to eight annotators independently label it to minimize human errors in the labeling process (Malo et al. (2014)). We use the labeled pair that at least 66% of annotators agreed on for fine-tuning the GPT model with 5 epochs.[7]

[7] Malo et al. (2014) provide labeled datasets with four options: 1) at least 50% of annotators agreed on, 2) 66% of annotators agreed on, 3) 75% of annotators agreed on, and 4) all of annotators agreed on.