Market discipline in banking: the role of financial analysts

Andreas Barth[124] Sasan Mansouri[125] Fabian Woebbeking[23]

[1]Goethe University Frankfurt
{barth, mansouri}@finance.uni-frankfurt.de

[2]IWH - Leibniz Institute for Economic Research
fabian.woebbeking@iwh-halle.de

[3]Martin Luther University Halle-Wittenberg

[4]Saarland University Saarbrücken

[5]University of Groningen

January 15, 2024

Abstract

Using DebtBERT, a novel measure for analysts' attention during earnings conference calls, this paper studies how markets discipline banks. We consider two groups of banks: a treated group with implicit bail-out guarantees and an untreated group without such guarantees. Our analysis focuses on the information that analysts request from banks, which we classify using DebtBERT, a specifically trained large language model. We find that analysts increased their scrutiny post-global financial crisis. This increased attention affects banks' abnormal stock returns in the short-term and leverage in the long-term, which is especially notable in banks lacking bail-out guarantees. This result suggests a moral hazard problem, where investors strategically discipline banks based on their perception of bail-out guarantees. Our findings have important implications for regulatory and policy decisions aimed at promoting transparency and stability in the banking sector.

Keywords: bank, lending, risk, market discipline, too big to fail, DebtBERT, bidirectional encoder representations from transformers, large language model, machine learning, natural language processing

JEL classification: G01; G21; G28; M41; M48

1 Introduction

The link between accounting and financial stability has been of major importance since the 2007/09 financial crisis. Particular attention is given to the production of information via bank active disclosure as an important prerequisite for market discipline [Bischof et al., 2021]. However, hardly anything is known about the disciplining role of financial analysts approaching bank managers directly during earnings conference calls. We provide evidence that the information production by analysts serve an additional important function for market discipline when directly approaching bank managers in earnings conference calls.

Employing DebtBERT as a machine learning tool, we identify the extent to which analysts' questions are related to the traditional banking activity of providing debt to firms. Not surprisingly, there is a significant difference in the level of these banking-related questions between banks and non-financial companies, suggesting that the natural language processing model correctly classifies analysts' questions. However, we observed that after the financial crisis, analysts have become even more interested in banks' asset business. Zooming in on financial institutions, we found that while banks labeled as systemically important financial institutions (SIFIs) were asked fewer banking-related questions compared to non-SIFI banks before the financial crisis, there is a strong increase in the demand for information about SIFIs' assets by financial analysts after the financial crisis.

These results suggest that there was minimal information production by analysts regarding SIFIs' assets before the crisis, as these banks largely benefited from implicit guarantees. With the regulatory changes in the post-crisis years, particularly the numerous regulatory actions focusing on the resolution of too-big-to-fail banks, analysts had an increasing interest in the asset business of SIFIs. They started to produce important information during conference calls as a prerequisite for a market disciplining mechanism.

We further show in the cross-section of banks that for non-SIFIs, more analysts' questions related to bank's asset business translate into lower abnormal stock returns following the conference call, to a reduction in leverage and to an increase in the liquidity ratio. Thus, the production of information by analysts helps to impose market discipline for non-SIFI banks. For SIFIs, however, despite an increased attention to bank's asset business by analysts in conference calls, it seems that markets do not discipline these institutions.

The third pillar of the Basel II regulation asks market participants to discipline banks for misbehavior. Market discipline should arise from the thread of a bank run by depositors [Calomiris and Kahn, 1991, Diamond and Rajan, 2001], and by pricing in information in bank securities [Barth and Schnabel, 2013, Hett and Schmidt, 2017, Berndt et al., 2023]. In order to do so, market participants need to have access to the relevant information, and regulators forced the availability of this information by mandatory disclosure requirements for banks' risk exposures. In addition to mandatory disclosures, earnings conference calls serve as an additional disclosure event during which managers make presentations to and answer questions from various market participants. As most investors of non-financial corporations provide equity to the firm, most discussions in these conference calls relate to earnings. Banks, however, are special due to their risk-transforming business. Hence, financial analysts in bank conference calls might, in the absence of implicit guarantees, not only care about earnings but also about risk-management or other bank asset-related issues.

However, despite some prominent bank failures, bank business was regarded as safe until the financial crisis,[1] and banks – particularly large and systemically important institutions – benefited from implicit guarantees.[2] While market participants had little incentives to actively monitor the banking sector due to these guarantees, it is an open question whether financial analysts in conference calls were asking the right questions to produce information that was necessary to impose market discipline. Understanding the content of questions during conference calls of financial versus non-financial firms allows us to shed light on this questions. We show that analysts have asked more bank business-

[1]The most prominent bank failures prior to the crisis were attributed to some idiosyncratic issues. For example, the failure of Continental Illinois National Bank and Trust Company in 1984 was attributed to its heavy exposure to the energy sector, and the Savings and Loan crisis was primarily seen as a crisis in financial institutions specializing in mortgage lending.

[2]Problems associated with banks being too big to fail have been extensively discussed in the literature, see, e. g. Boyd and Gertler [1994], Stern and Feldman [2004].

related questions only from the financial crisis onward. They have done so particularly in conference calls of systemically important institutions, and have asked these questions using a more negative sentiment. While this information production mechanism led to more market discipline for non-SIFI banks, it did fail to do so for SIFI banks, where markets still

This paper adds to the literature in several ways. First, we add to the financial economics literature discussing the concept market discipline. Starting in the late 1990s, the notion that private investors can influence the behavior of financial institutions has gained growing attention from financial regulatory bodies. Following this growing attention, market discipline has been established in the regulatory framework as one of the three pillars in the Basel II Accords.[3] Bliss and Flannery [2002] emphasize that market discipline comes with two distinct components. On the one hand, investors need to be able to evaluate the true condition of a bank (monitoring), and on the other hand, bank managers must be responsive to investor feedback (influence). Our paper focuses particularly on the first component, the possibility to evaluate bank risk, and whether and how financial analysts contribute to this task during conference calls.

We further add to the discussion on the prevalence of implicit guarantees for systemically important banks before and after the financial crisis, and the implication of bailout guarantees for market discipline. Several articles have shown that systemically important banks benefited from implicit bailout guarantees and were regarded as being too systemic to fail, resulting in a lack of market discipline for these banks [Ueda and Di Mauro, 2013, Barth and Schnabel, 2013, Santos, 2014, Hett and Schmidt, 2017, Acharya et al., 2022]. However, little is known about the role that financial analysts play in this respect, i.e. whether analysts were 'blinded' by the implicit guarantees for systemically important banks and thus asked different questions in conference calls towards banks versus non-financial institutions, and in particular towards systemically important financial institutions.

[3] The Basel II Accords states that "[T]he Committee aims to encourage market discipline by developing a set of disclosure requirements which will allow market participants to assess key pieces of information on the scope of application, capital, risk exposures, risk assessment processes, and hence the capital adequacy of the institution" [Basel Committee on Banking Supervision, 2004].

Finally, we contribute to the discussion of how financial accounting, in particular disclosure practices, can contribute to achieving financial stability. Acharya and Ryan [2016], for example, emphasize the importance of the quality of banks' financial reporting for the effectiveness of market discipline and non-market mechanisms in limiting banks' debt and risk overhangs during good economic times and mitigating their adverse consequences for financial system stability during downturns. The paper by Bischof et al. [2021] examines banks' disclosures and loss recognition during the 2007-2009 financial crisis, identifying several core issues for the link between accounting and financial stability. They find that banks' disclosures about relevant risk exposures were relatively sparse before the crisis, with more detailed disclosures emerging only after significant concerns about banks' exposures had arisen in the markets. Bischof et al. [2021] further provide evidence that shielding regulatory capital from accounting losses through prudential filters can dampen banks' incentives for corrective actions. Overall, their analysis reveals several significant challenges for accounting and financial reporting to contribute to financial stability. We add to this literature by not only focusing on bank disclosure and the effect of accounting rules on financial stability, but stressing the importance of financial analysts to provide additional information to markets.

The rest of the paper is structured as follows. In Section 2, we discuss our machine learning approach for identifying debt related questions. Section 3 covers the variables used in our empirical analysis, while Section 4 presents the empirical study. This section is further divided into three parts that investigate analysts' role in monitoring debt related questions, their influence on market response, and the market reaction to increasing discipline. Finally, we summarize our findings and discuss their implications in Section 5.

2 DebtBERT: Identifying bank-business questions

2.1 Training and validation

The Bidirectional Encoder Representations from Transformers (BERT) model, introduced by Devlin et al. [2018], is a natural language processing (NLP) model that has significantly advanced the state of the art in various NLP tasks. BERT is pre-trained using a combination of two unsupervised learning objectives: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). In MLM, a percentage of input tokens are randomly masked, and the model is trained to predict the original token based on the context provided by the remaining tokens. In NSP, the model learns to predict whether two input sentences are consecutive in the original text.

BERT utilizes the Transformer architecture [Vaswani et al., 2017] for its underlying structure, and the pre-training process allows it to capture bidirectional contextual information, resulting in improved performance on downstream tasks, such as – in our case – text classification. More specifically, this paper uses the 'bert-base-uncased' variant, a BERT model that was pre-trained on a large corpus of English text, including the BooksCorpus with 800 million words [Zhu et al., 2015] as well as English Wikipedia with $2,500$ million words. The contextualized representations generated by the transformer layers are then used as input during fine-tuning for specific NLP tasks.

Fine-tuning in the context of BERT refers to the process of taking a pre-trained BERT model and adapting it to a specific NLP task by training it further on a smaller labeled dataset specific to that task. When fine-tuning a BERT model, the pre-trained weights of the model are kept fixed, and only the weights of the task-specific output layer are trained using the labeled data for the specific task.

The DebtBERT model introduced in this paper is trained exclusively on labeled data from non-financial firms' earnings conference calls, ensuring a strict separation between the training data and the empirical analysis. Our training strategy is anchored on the rationale that the assets side of banks closely mirrors a significant portion of a corporation's debt structure. Essentially, we train DebtBERT to identify and understand debt related questions in the context of non-financial firms. Once trained, the model is then applied to an independent dataset comprising banks' earnings conference calls. In this new context, owing to its large base model, DebtBERT is capable of adapting to the use case of bank lending, despite being fine-tuned on a sample of corporate debt related questions.[4]

Given the earnings conference calls from non-financial firms, the fine-tuning task at hand is the classification of questions into debt related and equity related. The labeled data comes from a sub-set of firms that host independent earnings calls that cater specifically to debt or equity investors. We collect said specialized calls, extract the question and answer (Q&A) pairs therein and label each pair according to its origin into 'debt' or 'equity'.[5] This procedure leaves us with $18,973$ Q&A, of which roughly 50% are labeled debt. Before training, we split the data by randomly selecting Q&A pairs into a training set (80%), a validation set (10%) and a test set (10%). The machine learning algorithm only observes the training and validation set, while the remaining data is retained for independent testing.

In its final state, the model achieves an accuracy[6] of 79.40%. Training the model for additional epochs[7] yields a higher accuracy in the training set, however, decreases accuracy in the test set. The shift in accuracy between the two sets indicates that the model starts to over fit the training set after the second epoch. For the remaining analysis, we therefore make the modelling choice to train for two epochs. The so-called confusion matrix in Figure 1 provides a detailed review of the final model's accuracy and type one/two errors on predictions made in the test set. Most prominently, the algorithm was able to correctly classify 76% of equity questions and 77% of debt questions.

[Figure 1 about here]

The entire trained model is available for replication and future research at hugging-

[4]See Section 2.2 ('Explainable AI') for evidence that corroborates this claim.

[5]We ignore observations with less than 10 words, in order to focus on sentences long enough to extract a meaningful context.

[6]This paper measures accuracy using a metric called 'token-level accuracy'. This metric measures the percentage of tokens in the test dataset that are correctly predicted by the model.

[7]An epoch is a complete pass through the entire training set, where the BERT model looks at each training example once and updates its parameters based on the error it made in its predictions.

face.co/Econlinguistics/debtbert.

2.2 Explainable AI

Any BERT model is a black-box model where the reasoning is often difficult to explain due to its complex internal structure. It is therefore imperative to leverage explainability techniques to shed light on the model's decision-making process. Understanding the rationale behind a particular classification not only helps in validating the model's performance, but also assists in providing insights into what features are significant for distinguishing debt related from other questions.

We use Multinomial Inverse Regression (MNIR), a method developed by Taddy [2013], to make the output of our DebtBERT model explainable. The idea behind MNIR is to reduce the complexity of high dimensional text data to a smaller set of topic or feature representations, much like other dimension reduction techniques such as Least Absolute Shrinkage and Selection Operator (LASSO) or Principal Component Analysis (PCA). Using MNIR in our setting follows a three step procedure:

1. Document Feature Matrix: First, the text data is transformed into a document-feature matrix, where each row represents a document (in our case, a question) and each column represents a unique word or n-gram in the dataset. The entries in this matrix represent the frequency of each word or n-gram in each question.

2. Inverse Regression: Instead of using the fitted model to predict labels from word frequencies (as is typically done in regression), the inverse regression approach uses the fitted model to predict word frequencies from labels. Essentially, it estimates which words or n-grams are most likely to appear in a question given that it is labeled as 'debt'.

3. Interpretation: The estimated word or n-gram frequencies for each label provide an interpretable representation of what characterizes debt related questions. Words or n-grams that are estimated to be more frequent in debt questions are indicative of debt related topics, and vice versa for other questions.

The MNIR model extracts textual features that are most predictive of debt related questions, thereby providing an interpretable explanation for the predictions of the DebtBERT model. More specifically, each token inside the document-feature-matrix can be ranked according to its importance based on their likelihood of influencing the 'debt' classification. The higher the frequency of a token in documents classified under 'debt', the more it contributes to a document being labelled as such, and vice versa for non-debt questions.

Figure 2 presents the MNIR results, where the size of a token indicates its importance in predicting the 'debt' label of a question. It is evident that corporate debt related questions frequently associate with financial terminology such as: 'loans', 'covenants', and 'banks'.[8] For banks, tokens like 'borrowers', lending', and issuance' are prominent as well. Additionally, DebtBERT shifts its focus towards terminology that is very specific to bank-lending, including terms such as 'basel', 'repo', and 'lcr'. This highlights that, owing to its large base model, DebtBERT is well capable of adapting to the use case of bank lending, despite being fine-tuned on a sample of corporate debt related questions.

[**Figure 2 about here**]

It is important to note that this method does not offer perfect transparency. Although MNIR can determine which tokens are most influential in classifying the 'debt' label, it does not necessarily know how they interact with each other within DebtBERT to reach a final classification decision. However, the insights from MNIR remain crucial for validating DebtBERT, while also allowing us to better understand the decisions made by the black-box model.

3 Data and Summary Statistics

Our sample comprises of all banks listed in the Bank Regulatory Database and spans over the time period 2002 - 2020. For these banks, we collect quarterly earnings conference calls from StreetEvents. This results in 8,834 quarterly earnings conference calls for 260

[8] In total, the model extracts 1201 tokens that identify corporate debt-labeled questions.

financial institutions. We merge this data to Compustat, where we obtain quarterly balance sheet characteristics as well as bank's stock returns. In addition to balance sheet characteristics, we collect stock returns around the earnings conference calls from CRSP and calculated for each conference call day the cumulative abnormal return, $CAR_{(-1,1)}$, defined as the difference between the actual return and the expected return based on the Fama-French three-factor model with momentum [Carhart, 1997], cumulated over the day before the conference call to the day after the call. We further collect information about analysts' earnings expectations from the Institutional Brokers' Estimate System (IBES), and following Barth et al. [2022], we measure earnings surprises in groups of ten deciles (five for negative and five for positive) ranging from -5 (the most negative surprise) to +5 (the most positive surprise). We also analyze the language used in earnings conference calls to assess the sentiment and uncertainty expressed by the management team. Management tone reflects the overall sentiment and attitude during these calls, while management uncertainty captures the extent of uncertainty conveyed by the team. To calculate tone, we compute the difference between the ratios of positive and negative words in the management's responses. Similarly, uncertainty is determined by calculating the ratio of uncertain words to the total words spoken by management. The lists of positive, negative, and uncertain words are based on Loughran and McDonald [2011].

We present descriptive statistics of these variables in Table 1.

[Table 1 about here]

The $Book-to-Market$ (BTM) ratio is defined as a bank's book value of equity to its market value of equity. In our sample, the average BTM ratio amounts 81%, indicating a slightly higher market equity compared to the book value . *Leverage* is defined as the ratio of a bank's long and short-term debts over the total stakeholders' equity. Naturally, banks finance their assets with a large portion of the debt, resulting in a relatively high average leverage of 1.7. The *Liquidity Ratio* calculates the proportion of a bank's most liquid assets to its short-term liabilities, and thus, determines the ability of a bank to meet its short-term obligations. We employ two measures of bank size, $Market\,Capitalization$, which refers to the total value of a bank's outstanding shares of stock in the market, and

Total Assets on bank's balance sheet. Both size measures were used in logs. *Tobin's Q* measures the ratio of a bank's market value to the replacement cost of its assets. A Q value greater than one suggests that the bank's market value is greater than the cost of replacing its assets, indicating the presence of intangible assets, such as brand value or management expertise. We further flag whether a bank is considered a Systemically Important Financial Institution with a *SIFI* dummy.[9]

Our main focus is on debt related questions asked by financial analysts in earnings conference calls. We identify debt related questions as described in section 2. In total, the average share of debt related questions in conference calls amounts 0.5 with a notable variation in the cross-section and over time.[10] The top panel of Figure 3 shows the share of debt related questions over time for all banks in our sample. In the bottom panel of Figure 3, we split the sample into systemically important financial institutions (SIFI) and non-systemically important institutions. Interestingly, the share of debt related questions was quite similar before the financial crisis, but increased tremendously for systemically important institutions over the financial crisis.

[Figure 3 about here]

4 Empirical Analysis

4.1 Analysts and bank disclosure: monitoring

In our empirical analysis, we investigate the two distinct components of market discipline as described in Bliss and Flannery [2002]. First, we aim to identify the monitoring component and analyze whether financial analysts add to the production of relevant information.

[Figure 3 about here]

[9] Banks were classified as systemically important financial institutions based on the first announcement by the Financial Stability Board in 2011. While the list was published only in 2011, the banks on this list were considered too systemic to fail before publication.

[10] Financial institutions, in general, face more debt related questions. **??** in the appendix shows that the probability of recieving debt related questions in the earnings calls of a non-financial firm is almost half of the financial firms.

We start with descriptive evidence in Figure 3. The top panel shows bank business-related questions over time, separately for financial and non-financial institutions. Not surprisingly, the share of bank business-related questions is on average larger for financial versus non-financial firms.[11] However, starting with the financial crisis in 2007, there is a strong increase in the number of questions dealing with bank business-related topics for financial firms, while the same is not true for non-financial corporations.

In order to rule out that this observation is purely driven by an increase in bank business-related issues but related to an increase in analysts' monitoring of banks, we zoom in into the cross section of financial institutions. In particular, we make use of the fact that large systemically important institutions benefited from implicit bailout guarantees prior to the financial crisis, which got removed after the crisis, while smaller institutions always bore the risk of a default (although they, too, to a smaller degree before the crisis). Thus, if analysts help to provide information for market discipline, they should have asked more monitoring questions particularly to systemically important institutions when they came at risk after the crisis. Descriptive evidence of bank business-related questions in conference calls for SIFIs versus non-SIFIs is shown in the bottom panel of Figure 3. It is striking to see that analysts approach particularly SIFI banks with more bank business-related questions, while the increase in bank business-related questions for non-SIFIs is only moderate. While the observation in Figure 3 is purely descriptive, we investigate this question further and run the following regression:

$$\begin{aligned}\text{DebtBERT}_{it} = {} & \alpha_t + \text{Non-SIFI}_i + \text{PostCrisis}_t \\ & + \beta \cdot \text{Non-SIFI}_i \cdot \text{PostCrisis}_t + \gamma X_{it} + \epsilon_{it}. \qquad (1)\end{aligned}$$

DebtBERT_{it} measures the share of bank business-related questions that all analysts ask to the management of bank i in the conference call at time t. The Non-SIFI_i dummy flags whether bank i has not been classified as a systemically important financial institu-

[11]Note again that bank business-related questions are closely related corporate borrowing (see Section 2.2 for a detailed discussion).

tion by the Financial Stability Board in the first report in 2011. PostCrisis_t is a dummy that indicates the time after the financial crisis, i.e. the dummy equals 1 from September 2007 onward and zero otherwise. As financial institutions were strongly affected by the financial crisis, we exclude the crisis years throughout the analysis, that is, the months between 09/2007 and 06/2009. In our baseline specification, we absorb variation that is common across all banks by time fixed effects. In further specifications, we include industry code fixed effects to control for bank business models or even saturate the model with bank fixed effects. We additionally control for time-varying bank characteristics with the vector X_{it}, including all balance sheet variables shown in section 3. We allow for a potential serial correlation of analyst questions within each bank and within each quarter and employ two-way clustering of standard errors [Cameron et al., 2011] at the bank and year-quarter dimensions. We are mostly interested in the coefficient of the interaction $\text{Non-SIFI}_i \cdot \text{PostCrisis}_t$, which captures the differential treatment of SIFIs versus non-SIFIs, pre- and post crisis. If financial analysts add a disciplinary component, we would expect more monitoring questions towards SIFIs and thus, β to have a negative sign.

[Table 2 about here]

Results are shown in Table 2. We observe a negative and significant coefficient for the diff-in-diff term, indicating that financial analysts asked less bank business-related questions to non-SIFI banks post crisis, relative to SIFI banks. That is, analysts understood that regulatory changes removed implicit guarantees such that bank resolution became more likely.

Bliss and Flannery [2002] describe that market discipline comes with the two distinct components that investors need to be able to evaluate the true condition of a bank (monitoring), and bank managers must be responsive to investor feedback (influence). Our finding so far suggests that analysts started to provide important information to bank stakeholders, which serves as a necessary prerequisite for market discipline. We will now investigate whether the additional information provided by analysts in conference calls translates to market discipline, i.e. whether markets started to discipline banks more the more bank business-related information is produced by analysts in conference calls and whether banks eventually react to the increased demand for bank asset-related information.

4.2 Market and bank reaction

As analysts started to care about bank assets and asked more bank business-related questions in conference calls after the financial crisis, the first component of market discipline as described in Bliss and Flannery [2002], i.e., investors need to be able to evaluate the true condition of a bank, is satisfied. However, this does not necessarily imply that financial markets or bank mangers react to the increase in information production through financial analysts. To analyze whether analysts were successful in increasing market discipline, we now investigate market reactions following earnings conference calls. In particular, we analyze the cumulative abnormal return of bank i around conference call in quarter t as a function of bank business-related questions. Investors, who generally consider themselves exposed to bank business risk, should update their believes based on the additional information that arrives due to additional scrutiny. Ultimately, this re-calibration of investor beliefs should be instantaneously reflected in the equilibrium price. We analyze market reaction employing the following regression model:

$$\begin{aligned}\mathrm{CAR}_{it}^{[-1,1]} = \alpha_t &+ \mathrm{DebtBERT}_{it} \cdot \mathrm{PostCrisis}_t + \mathrm{DebtBERT}_{it} \cdot \text{Non-SIFI}_i \\ &+ \beta \cdot \mathrm{DebtBERT}_{it} \cdot \mathrm{PostCrisis}_t \cdot \text{Non-SIFI}_i + \gamma X_{it} + \eta Z_{it} + \epsilon_{it}. \end{aligned} \tag{2}$$

$\mathrm{CAR}_{it}^{[-1,1]}$ measures the cumulative abnormal return of bank i in quarter t. $\mathrm{DebtBERT}_{it}$ measure the share of bank business-related questions that all analysts ask to the management of bank i in the conference call in quarter t. The Non-SIFI_i dummy flags whether bank i has not been classified as a systemically important financial institution by the Financial Stability Board in the first report in 2011. $\mathrm{PostCrisis}_t$ is a dummy that indicates the time after the financial crisis, i.e. the dummy equals 1 from September 2007 onward and zero otherwise. We again exclude the crisis years throughout the analysis, that is, the months between 09/2007 and 06/2009.

We absorb variation that is common across all banks by time fixed effects and further include industry code fixed effects to control for bank business models or even saturate the model with bank fixed effects. We again control for time-varying bank characteristics with the vector X_{it}, which includes all balance sheet variables that were shown in section 3, and further include the sentiment and uncertainty score as text-based measures for conference call of bank i in quarter t. We again allow for a potential serial correlation of analyst questions within each bank and within each quarter and employ two-way clustering of standard errors [Cameron et al., 2011] at the bank and time(quarters) dimensions.

[Table 3 about here]

We present results in Table 3. We observe a negative and highly significant coefficient for the triple interaction term, indicating that financial markets discipline non-SIFIs more with increasing information production by analysts after the crisis compared to SIFI banks. This result suggests that markets exercise less discipline towards SIFIs per additional information produced by analysts. That is, despite the awareness of analysts and the additional information about the true condition of a bank, markets seem to still believe in bailout guarantees for systemically important institutions

Finally, we investigate whether analysts' questions in conference calls influence banks' behavior. In particular, we investigate whether banks that were asked more debt related questions build up capital or liquidity buffer in the quarter (year) following the respective conference calls. Our study focuses on bank leverage and liquidity as critical determinant of both profitability and risk in the banking sector. Unlike other types of firms, banks are uniquely motivated to maximize their leverage and minimize liquid assets, approaching the thresholds established by regulatory authorities. However, when banks opt to reduce their leverage below or increase their liquidity positions above the regulatory limits, they engage in a strategic trade-off between profitability and risk minimization. This interplay is central to our analysis, as it sheds light on the strategic decisions banks make in response to external market pressure.

Using the global financial crisis as an attention shock as before, we investigate the effect of analysts' scrutiny pre- and post-financial crisis to see whether the increased monitoring of analysts translate into bank behavior. Again, we separate between banks with (SIFI) and without (Non-SIFI) bailout guarantees prior to the crisis, following the rationale that banks perceive market discipline only if they consider themselves at risk. We run the following regression:

$$
\begin{aligned}
Y_{it} = \alpha_t &+ \text{DebtBERT}_{it} \cdot \text{PostCrisis}_t + \text{DebtBERT}_{it} \cdot \text{Non-SIFI}_i \\
&+ \beta \cdot \text{DebtBERT}_{it} \cdot \text{PostCrisis}_t \cdot \text{Non-SIFI}_i + \gamma X_{it} + \eta Z_{it} + \epsilon_{it}.
\end{aligned} \tag{3}
$$

Y_{it} measures interchangeably the leverage ratio or the ratio of liquidity coverage ratio of bank i in quarter t. DebtBERT_{it} measures the share of bank business related questions that all analysts ask to the management of bank i in the conference call in quarter t. The Non-SIFI_i dummy flags whether bank i has not been classified as a systemically important financial institution by the Financial Stability Board in the first report in 2011. PostCrisis_t is a dummy that indicates the time after the financial crisis, i.e. the dummy equals 1 from September 2007 onward and zero otherwise. We again exclude the crisis years throughout the analysis, that is, the months between 09/2007 and 06/2009.

In our baseline specification, we absorb variation that is common across all banks by time fixed effects and further include industry code fixed effects to control for bank business models or even saturate the model with bank fixed effects. We additionally control for time-varying bank characteristics with the vector X_{it}, which includes all balance sheet variables that were shown in section 3. We allow for a potential serial correlation of analyst questions within each bank and within each quarter and employ two-way clustering of standard errors [Cameron et al., 2011] at the bank and time dimensions. We use a time lag of one quarter (year) for all right-hand-side variables.

[**Table 4 about here**]

[**Table 5 about here**]

We show results in Table 4 and Table 5. We observe a negative (positive) and significant coefficient for the triple interaction term in the leverage ratio (liquid asset) regression, indicating that non-SIFIs reduce their leverage and increase their liquid assets in the first and fourth quarter following the earnings conference call the more detailed analysts are approaching the business of a bank. Thus, banks that were asked more bank business-related questions increased their liquidity position in the quarter (year) following the conference call in the years following the financial crisis. These results hold not only in the cross-section of banks, but also in the within-bank variation, controlling for bank fixed effects. These results suggest that analysts' questions in conference calls influence banks' behavior more for non-SIFI banks and thus, provided a disciplinary mechanism mostly for non-SIFIs, while analysts were not successful in disciplining SIFIs.

5 Conclusion

This paper investigates the role of analysts in disciplining financial institutions. Starting with analysts' inquiries into bank business models, we identify a channel that initially affects investors' behavior and ultimately influences banks' risk-taking.

We focus on the role of financial analysts who directly approach bank managers during earnings conference calls, examining whether these analysts can perform an additional important function in market discipline. Utilizing DebtBERT, a specifically fine-tuned large language model, we assess the extent to which analysts' questions relate to bank business activities and analyze the frequency of these inquiries over time.

Our findings reveal that post-crisis, financial analysts pose more business related questions to banks, especially to systemically important financial institutions (SIFIs). These inquiries have facilitated better investor evaluation of a bank's true condition, enabling more effective monitoring, a crucial element for market discipline.

We observe that the information produced by analysts has been only partially successful in strengthening market discipline. Our analysis shows lower abnormal stock returns for non-SIFI banks following conference calls with numerous bank business-related questions, compared to SIFI banks. Additionally, non-SIFI banks have demonstrated more significant deleveraging and improved liquidity compared to SIFI banks after such conference calls. This indicates that while analysts directed questions at both types of banks, investors were less inclined to apply the same level of discipline to too-big-to-fail (SIFI) banks. This disparity might suggest that investors strategically discipline banks based on their perception of bail-out guarantees.

References

V. V. Acharya and S. G. Ryan. Banks' financial reporting and financial system stability. Journal of Accounting Research, 54(2):277–340, 2016.

V. V. Acharya, D. Anginer, and A. J. Warburton. The end of market discipline? investor expectations of implicit government guarantees. Technical report, 2022.

A. Barth and I. Schnabel. Why banks are not too big to fail–evidence from the cds market. Economic Policy, 28(74):335–369, 2013.

A. Barth, S. Mansouri, and F. Woebbeking. "let me get back to you": A machine learning approach to measuring nonanswers. Management Science, 2022.

Basel Committee on Banking Supervision. Basel ii: International convergence of capital measurement and capital standards: A revised framework - comprehensive version, 2004.
https://www.bis.org/publ/bcbs128.htm.

A. Berndt, D. Duffie, and Y. Zhu. The decline of too big to fail. Technical report, 2023.

J. Bischof, C. Laux, and C. Leuz. Accounting for financial stability: Bank disclosure and loss recognition in the financial crisis. Journal of Financial Economics, 141(3): 1188–1217, 2021.

R. R. Bliss and M. J. Flannery. Market discipline in the governance of us bank holding companies: Monitoring vs. influencing. Review of Finance, 6(3):361–396, 2002.

J. H. Boyd and M. Gertler. The role of large banks in the recent u.s. banking crisis. Federal Reserve Bank of Minneapolis Quarterly Review, 18(1):2–21, 1994.

C. W. Calomiris and C. M. Kahn. The role of demandable debt in structuring optimal banking arrangements. The American Economic Review, pages 497–513, 1991.

A. C. Cameron, J. B. Gelbach, and D. L. Miller. Robust inference with multiway clustering. Journal of Business and Economic Statistics, 29(2):238–249, 2011.

M. M. Carhart. On persistence in mutual fund performance. The Journal of Finance, 52 (1):57–82, 1997.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

D. W. Diamond and R. G. Rajan. Liquidity risk, liquidity creation, and financial fragility: A theory of banking. Journal of Political Economy, 109(2):287–327, 2001.

F. Hett and A. Schmidt. Bank rescues and bailout expectations: The erosion of market discipline during the financial crisis. Journal of Financial Economics, 126(3):635–651, 2017.

T. Loughran and B. McDonald. When is a Liability not a Liability? Journal of Finance, 66(1):35 – 65, 2011. ISSN 00221082. doi: 10.1111/j.1540-6261.2010.01625.x.

J. A. Santos. Evidence from the bond market on banks"too-big-to-fail'subsidy. Economic Policy Review, 20(2):29–39, 2014.

G. Stern and R. Feldman. Too big to fail: the hazards of bank bailouts. G - Reference, Information and Interdisciplinary Subjects Series. Brookings Institution Press, 2004. ISBN 9780815781523. URL http://books.google.com/books?id=XlPjOF1W-_UC.

M. Taddy. Multinomial inverse regression for text analysis. Journal of the American Statistical Association, 108(503):755–770, 2013.

K. Ueda and B. W. Di Mauro. Quantifying structural subsidy values for systemically important financial institutions. Journal of Banking & Finance, 37(10):3830–3842, 2013.

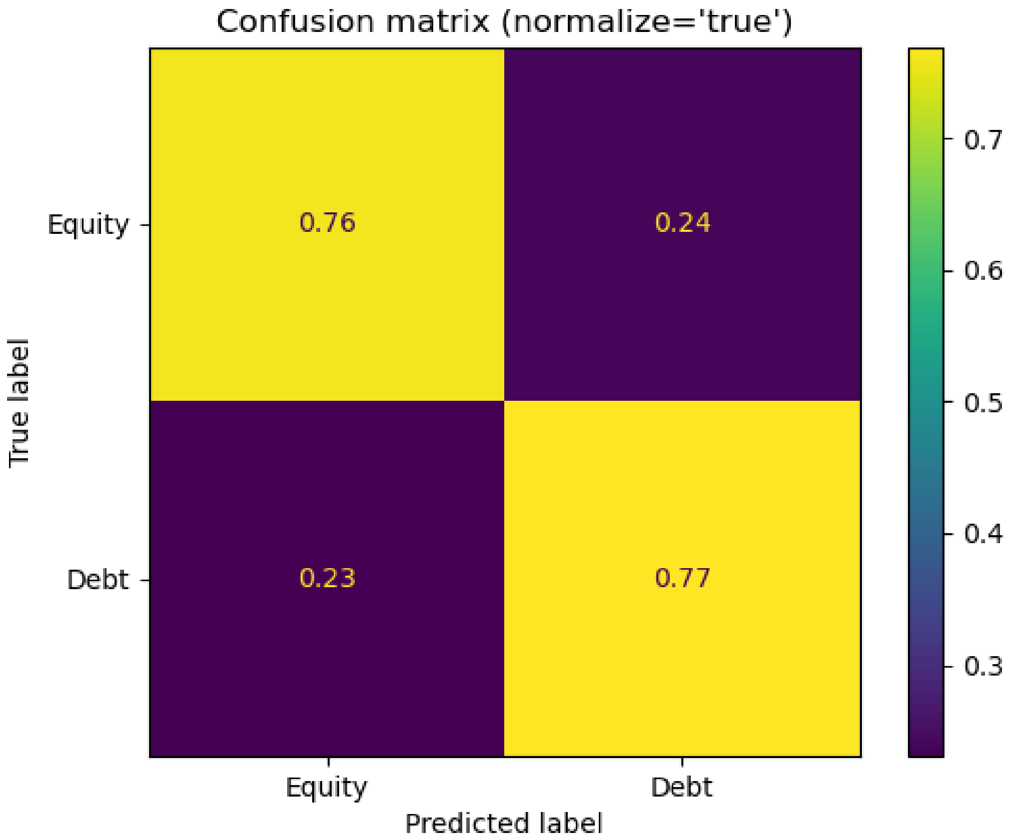
A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.

Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In Proceedings of the IEEE international conference on computer vision, pages 19–27, 2015.

Figures and Tables

Figure 1: **DebtBERT confusion matrix**

This figure presents the confusion matrix for DebtBERT, our model designed to analyze and classify debt related questions during earnings calls. The confusion matrix visually displays the model's performance by comparing its predictions against the actual labels. Columns represent model predictions, whereas rows represent the true labels in the test set. The diagonal cells indicate correct predictions, while off-diagonal cells represent miss-classifications. The overall accuracy and other performance metrics can be derived from the confusion matrix to evaluate the effectiveness of DebtBERT in identifying and categorizing debt related questions.



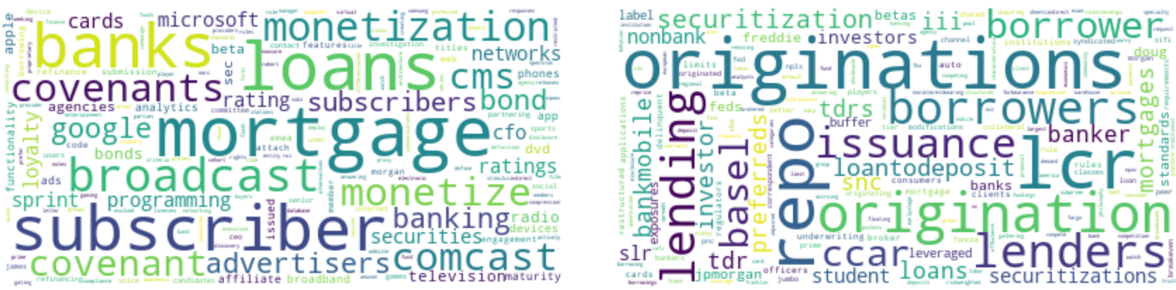


Figure 2: These word clouds display tokens that are associated with the classifications made by the DebtBERT model. The font size represents the MNIR factor load, indicating that a larger font corresponds to a stronger (positive) association with the DebtBERT classification. The word cloud on the **left** showcases words associated with corporate debt, while the **right** side highlights words related to bank lending.

Figure 3: **Bank-business questions over time - financials vs non-financials**

This figure shows the line plot illustrating the portion of debt related questions in earnings calls during our sample period at a yearly frequency, distinguishing between financials and non-financial firms.

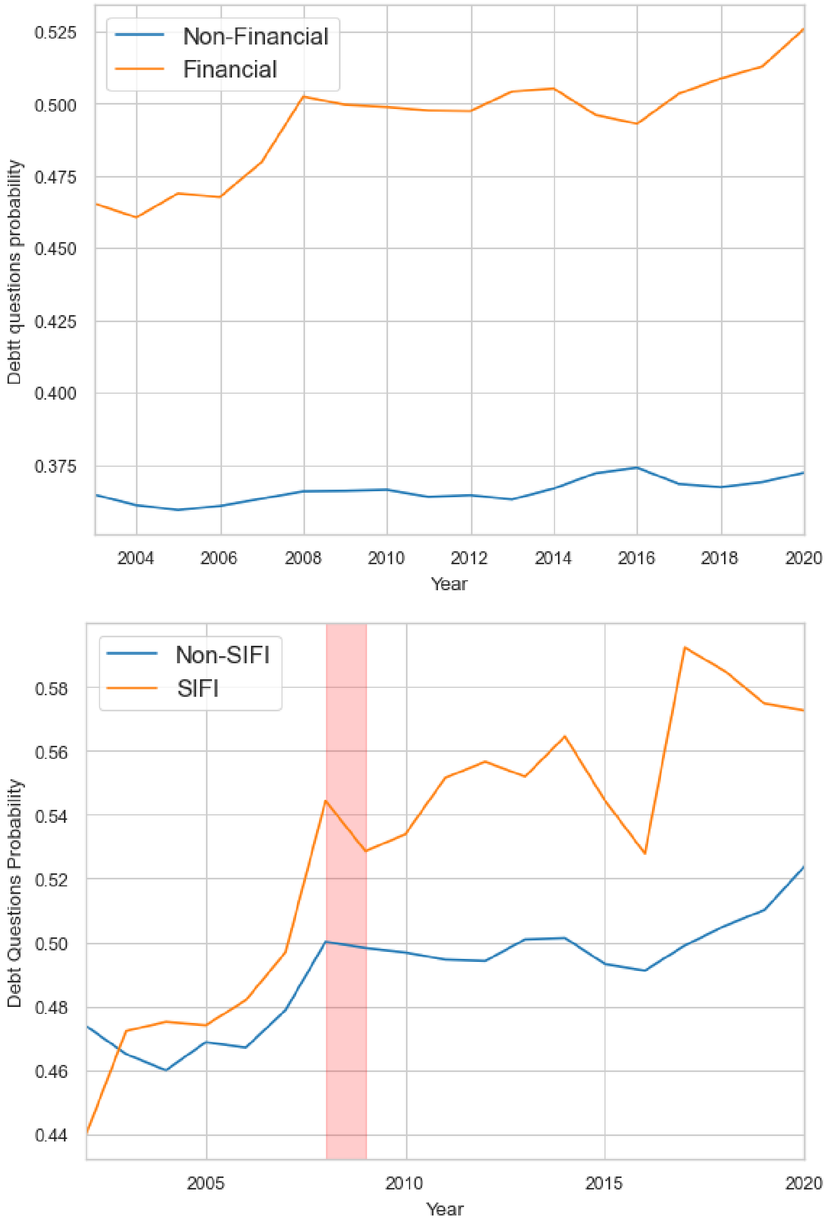


Table 1: **Descriptive statistics**

This table shows descriptive statistics of the variables used in the analysis. The variables are sorted alphabetically. The sample consists of 7,931 observations. All variables are defined in Table A1.

	N	Min	P25	Mean	P50	P75	Max	Std. Dev.
Book-to-Market	7,930	-1.4	.53	.81	.72	.95	10	.49
$CAR_{(-1,1)}$	7,961	-.87	-.023	.003	.0022	.029	.83	.056
Earnings Surprise	7,961	-5	0	.27	0	0	5	1.7
DebtBERT	7,961	.14	.45	.5	.49	.54	.92	.074
Leverage	7,647	0	.51	1.7	.96	1.7	44	3.5
Liquidity Ratio	6,499	.0013	.022	.052	.033	.061	.48	.055
Management Tone	7,961	-.42	-.018	-.0085	-.0095	.000099	.16	.028
Management Uncertainty	7,961	0	.0081	.012	.011	.015	.2	.0071
Market Cap.	7,930	1.9	6.3	7.5	7.3	8.4	13	1.7
Non-SIFI (d)	7,961	0	1	.95	1	1	1	.22
Tobin's Q	7,930	.87	1	1.1	1	1.1	18	.68
Total Assets	7,930	4.9	8.2	9.4	9.1	10	15	1.7

Table 2: **Debt Questions After Crisis**

This table presents the regression results for Equation 1. The dependent variable is the ratio of *Debt Questions* in the banks' earnings conference calls. All variables are defined in Table A1. t-statistics are given in parentheses. Standard errors are clustered at the bank and quarter level. ***, **, * indicate significance at the 1%, 5% and 10% levels.

	DebtBERT					
	(1)	(2)	(3)	(4)	(5)	(6)
AfterGFC (d)	0.083***	0.099***	0.000	0.000	0.000	0.000
	(6.96)	(8.35)	(.)	(.)	(.)	(.)
AfterGFC (d) × Non-SIFI (d)	-0.051***	-0.060***	-0.051***	-0.058***	-0.052***	-0.056***
	(-4.26)	(-5.17)	(-4.33)	(-5.11)	(-3.95)	(-4.15)
Non-SIFI (d)	-0.009	0.000	-0.009	0.000	0.025*	0.000
	(-1.05)	(0.00)	(-1.03)	(.)	(1.93)	(.)
Total Assets					0.030***	0.016**
					(5.25)	(2.04)
Book-to-Market					-0.013**	-0.007
					(-2.48)	(-1.61)
Tobin's Q					0.012***	0.008***
					(4.82)	(2.88)
Market Cap.					-0.024***	-0.019***
					(-4.47)	(-2.90)
Leverage					-0.001	-0.002***
					(-1.61)	(-2.66)
Observations	7930	7921	7930	7921	7647	7638
R^2	0.072	0.268	0.085	0.281	0.113	0.282
SIC2 FE	Yes	Implied	Yes	Implied	Yes	Implied
Bank FE	No	Yes	No	Yes	No	Yes
Year FE	No	No	Yes	Yes	Yes	Yes

Table 3: **Debt Questions After Crisis - Market Reaction**

This table presents the regression results for the market reaction to the earnings conference calls. The dependent variable is the cumulative abnormal return from the day before the earnings call to the day after it. Abnormal returns are calculated based on the Fama-French three-factor model with momentum [Carhart, 1997]. All variables are defined in Table A1. t-statistics are provided in parentheses, and standard errors are clustered at the bank and quarter level. Significance levels are indicated by ***, **, and * for the 1%, 5%, and 10% levels, respectively.

	$CAR_{(-1,1)}$		
	(1)	(2)	(3)
DebtBERT	0.003	-0.183***	-0.144**
	(0.31)	(-4.98)	(-2.44)
DebtBERT × AfterGFC (d)		0.170***	0.138***
		(6.10)	(3.67)
DebtBERT × Non-SIFI (d)		0.135***	0.096*
		(4.22)	(1.77)
DebtBERT × AfterGFC (d) × Non-SIFI (d)		-0.105***	-0.088***
		(-6.13)	(-3.27)
Non-SIFI (d)		-0.062***	0.000
		(-3.84)	(0.00)
AfterGFC (d) × Non-SIFI (d)		0.053***	0.047***
		(4.51)	(2.95)
Management Tone	0.080***	0.081***	0.066**
	(5.03)	(3.83)	(2.16)
Management Uncertainty	0.077	0.084	0.112
	(0.71)	(0.68)	(0.85)
Earnings Surprise	0.007***	0.007***	0.007***
	(9.70)	(9.80)	(9.00)
Total Assets	-0.000	0.000	0.001
	(-0.13)	(0.05)	(0.21)
Book-to-Market	0.009	0.009	0.013
	(1.55)	(1.43)	(1.18)
Tobin's Q	0.001	0.001	-0.000
	(0.72)	(0.59)	(-0.14)
Market Cap.	-0.001	-0.001	-0.009*
	(-0.49)	(-0.46)	(-1.86)
Observations	7930	7930	7921
R^2	0.056	0.058	0.104
SIC2 FE	Yes	Yes	Implied
Bank FE	No	No	Yes
Year FE	Yes	Yes	Yes

Table 4: **Debt Questions After Crisis - Leverage**

This table presents the regression results for Equation 1. The dependent variable in columns (1)-(2) represents the leverage value of the bank in the subsequent quarter, while in columns (3)-(4), it represents the bank's leverage in the following year. All variables are defined in Table A1. t-statistics are provided in parentheses, and standard errors are clustered at the bank and quarter level. Significance levels are indicated by ***, **, and * for the 1%, 5%, and 10% levels, respectively.

	Leverage$_{t+1}$		Leverage$_{t+4}$	
	(1)	(2)	(3)	(4)
DebtBERT	-21.231**	-11.709*	-26.418**	-16.563
	(-2.38)	(-1.89)	(-2.06)	(-1.64)
AfterGFC (d) × DebtBERT	19.465**	12.301*	25.212**	17.592*
	(2.22)	(1.94)	(2.01)	(1.71)
DebtBERT × Non-SIFI (d)	24.055**	13.872**	30.000**	18.300*
	(2.62)	(2.23)	(2.33)	(1.82)
AfterGFC (d) × DebtBERT × Non-SIFI (d)	-24.187***	-16.513**	-30.320**	-20.973**
	(-2.70)	(-2.51)	(-2.41)	(-2.02)
Non-SIFI (d)	-11.546**	0.000	-14.558**	0.000
	(-2.19)	(.)	(-2.06)	(0.00)
AfterGFC (d) × Non-SIFI (d)	13.340***	9.792**	16.386**	11.961**
	(2.72)	(2.57)	(2.44)	(2.06)
Total Assets	4.714***	4.156**	4.724***	4.114**
	(4.13)	(2.14)	(4.05)	(2.02)
Book-to-Market	-2.439***	-1.023*	-2.903***	-1.299*
	(-2.91)	(-1.68)	(-3.12)	(-1.80)
Tobin's Q	1.497***	0.740*	1.508***	0.689*
	(3.06)	(1.79)	(3.03)	(1.72)
Market Cap.	-4.283***	-2.197***	-4.304***	-2.095**
	(-4.20)	(-2.71)	(-4.10)	(-2.26)
Observations	7641	7632	7219	7211
R^2	0.464	0.819	0.475	0.828
SIC2 FE	Yes	Implied	Yes	Implied
Bank FE	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes

Table 5: **Debt Questions After Crisis - Liquidity Ratio**

This table presents the regression results for Equation 1. The dependent variable in columns (1)-(2) represents the liquidity ratio of the bank in the subsequent quarter, while in columns (3)-(4), it represents the bank's liquidity ratio in the following year. All variables are defined in Table A1. t-statistics are provided in parentheses, and standard errors are clustered at the bank and quarter level. Significance levels are indicated by ***, **, and * for the 1%, 5%, and 10% levels, respectively.

	Liq. Ratio$_{t+1}$		Liq. Ratio$_{t+4}$	
	(1)	(2)	(3)	(4)
DebtBERT	0.021	0.239	-0.038	0.180
	(0.35)	(1.10)	(-0.45)	(0.83)
AfterGFC (d) × DebtBERT	-0.358**	-0.057	-0.298*	0.008
	(-2.42)	(-0.39)	(-1.87)	(0.05)
DebtBERT × Non-SIFI (d)	-0.110*	-0.250	-0.058	-0.192
	(-1.71)	(-1.14)	(-0.63)	(-0.88)
AfterGFC (d) × DebtBERT × Non-SIFI (d)	0.467***	0.073	0.413**	0.010
	(3.05)	(0.49)	(2.48)	(0.07)
Non-SIFI (d)	0.053	0.000	0.025	0.000
	(1.08)	(0.00)	(0.39)	(.)
AfterGFC (d) × Non-SIFI (d)	-0.344***	-0.074	-0.319***	-0.042
	(-3.31)	(-0.95)	(-2.99)	(-0.58)
Total Assets	0.026*	0.018*	0.022	0.008
	(1.80)	(1.76)	(1.41)	(0.89)
Book-to-Market	-0.007	-0.007	-0.005	-0.002
	(-0.74)	(-1.49)	(-0.37)	(-0.51)
Tobin's Q	0.276***	0.023	0.273***	0.006
	(3.69)	(0.71)	(3.66)	(0.17)
Market Cap.	-0.024**	-0.022***	-0.019	-0.012**
	(-2.09)	(-2.99)	(-1.56)	(-2.31)
Observations	6482	6472	6393	6384
R^2	0.247	0.770	0.273	0.770
SIC2 FE	Yes	Implied	Yes	Implied
Bank FE	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes

Appendix

Table A1: **Variable definitions.**

Variable	**Definition**
AfterGFC (d)	Dummy for the period before and after the financial crisis: Accounts for the impact of the financial crisis period between September 1, 2007, and June 1, 2009.
Book-to-Market	The ratio of a bank's book value of equity to its market value of equity, providing insight into the market's valuation of the bank compared to its accounting value.
$CAR_{(-1,1)}$	Cumulative Abnormal Returns: Abnormal returns based on the Fama-French model with momentum from the day before the earnings call to the day after, capturing market reactions to earnings announcements.
Earnings Surprise	Ten deciles for earnings surprises, ranging from -5 (the most negative surprise) to +5 (the most positive surprise), measuring the degree of deviation from analysts' expectations as in Barth et al. [2022]
DebtBERT	The number of debt-related questions asked during the earnings calls, reflecting investor concerns and focus on the bank's debt management.
Leverage	Proportion of total assets financed by debt: Indicates a bank's reliance on borrowed funds to finance its operations and its exposure to financial risk.
Liquidity Ratio	The proportion of a bank's most liquid assets to its short-term liabilities, helping assess the bank's ability to meet short-term obligations without facing financial distress.
Management Tone	This variable captures the overall sentiment and attitude of the management team during earnings conference calls. It is calculated as the difference between the ratio of positive words and the ratio of negative words used during the call. The list of positive and negative words is sourced from Loughran and McDonald [2011].
Management Uncertainty	Captures the level of uncertainty expressed by the management team in earnings conference calls, indicating potential risks and challenges faced by the bank. The list of uncertainty-relevant words is sourced from Loughran and McDonald [2011].
Market Cap.	The total value of a bank's outstanding shares of stock in the market, representing a key measure of a bank's size and the market's perception of its value.

SIFI (d)	Dummy for Systemically Important Financial Institutions: Indicates whether a bank is considered systemically important, posing a higher risk to the financial system if it were to fail.
Tobin's Q	The ratio of a bank's market value to the replacement cost of its assets, indicating the presence of intangible assets such as brand value or management expertise when the value is greater than one.
Total Assets	The log value of the total value of a bank's assets, including cash, loans, securities, and fixed assets, serves as an indicator of a bank's size and financial capacity.