Sentiment in Bank Examination Reports and Bank Outcomes

Maureen Cowhey[*], Seung Jung Lee[†], Thomas Popeck Spiller[‡], Cindy M. Vojtech[§¶]

December 15, 2021

Abstract

We investigate whether the bank supervisory process provides useful insight into bank future outcomes. We do this by conducting textual analysis on about 5,400 small to medium-sized commercial bank examination reports from 2004 to 2016. These confidential examination reports provide textual context to each component of the supervisory CAMELS ratings: capital adequacy, asset quality, management, earnings, liquidity, and sensitivity to market risk. Each component is given a categorical rating, and each bank is then designated an overall composite CAMELS rating along the same scale, which are used to determine the safety and soundness of banks. We find that controlling for a variety of factors, including the ratings themselves, the sentiment supervisors express in describing most of the components predict future bank outcomes. The sentiment conveyed in the capital, asset quality, management, and earnings sections provides significant information in predicting future outcomes for capital levels, problem loans, supervisory actions, and profitability, respectively. This suggests that bank supervisors play a meaningful role in the surveillance of the banking system.

JEL classification: G21; G28.

Keywords: Bank examination reports; private supervisory information; Natural Language Processing.

[*]E-mail: maureen.r.cowhey@frb.gov

[†]E-mail: seung.j.lee@frb.gov

[‡]E-mail: thomas.w.spiller@frb.gov

[§]E-mail: cindy.m.vojtech@frb.gov

[¶]All authors are at the Board of Governors of the Federal Reserve System. We thank Matthew Deininger, Nathan Mislang, Ariel Polani, Mariya Pominova, Sarah Ridout, Ashley Sexton, Krista Stapleford, and Nicholas Stewart for help cleaning the data and for research assistance. We thank participants at the Financial Stability Workshop and International Banking and Finance Lunch Brownbag for helpful suggestions. Disclaimer: The material here does not represent the views of the Board of Governors of the Federal Reserve System or its staff.

1 Introduction

This paper investigates the effectiveness of supervision in maintaining the safety and soundness in the banking system in the United States. Supervisors play a vital role in the surveillance of banks by actively engaging in bank examinations and off-site monitoring. Many studies on the effectiveness of bank supervision have relied on the categorical CAMELS ratings of banks' safety and soundness. We take a close look at additional information—the examination reports based on on-site exams on small and medium-sized banks—to gauge the usefulness of more detailed information in this surveillance process. In particular, we study how sentiment in these reports conveys important information on future bank outcomes.

Banks are important intermediaries in the U.S. financial system. While offering a plethora of financial services, their main function is to receive and manage deposits in order to originate loans and invest in securities. Partly due to the maturity mismatch between assets (which are more long-term) and liabilities (which are more short-term), banks are susceptible to runs. Bank runs, in turn, could render the entire financial system unstable. Deposit insurance helps to prevent some of these unwarranted runs, but creates limited liability for its stakeholders which may motivate excess risk-taking. Risk taking also has financial stability implications, if occurring on a large scale. Therefore, banks are also subject to various forms of capital requirements that ensure banks' ability to absorb nontrivial shocks to their earnings and balance sheets.

In this context, bank supervision, along with regulation, plays a key role in maintaining the banking system's safety and soundness by establishing a good understanding of banks' capital adequacy, asset quality, management effectiveness, earnings prospects, liquidity positions, and sensitivity to market risk. Supervisors convey their assessments on these measures through CAMELS ratings. These ratings summarize both public data and private supervisory information gathered during on-site bank exams and have been found to contain information useful to the supervisory monitoring of commercial banks (Berger and Davies (1998) and Berger, Davies and Flannery (2000)).

However, past studies have relied on these categorical ratings (of integers from 1 to 5) in showing the effectiveness of supervision and not from the more comprehensive supervisory process itself, for example, coming from the information content in the bank examination reports. To achieve a more thorough and complete understanding of the usefulness of the supervisory process, it is important to look at whether the reports also convey extra value-added information beyond what is conveyed in the discrete ratings.

In this paper, we look at about 5,400 bank examination reports from 2004 to 2016 and calculate sentiment scores based on the text in these documents. We look at language associated with five of the components of the CAMELS; capital adequacy, asset quality, management, earnings, and liquidity, and calculate the sentiment associated with each section in the reports.[1] We use this information to see if it has any additional predictive power for determining various bank outcomes related to each component. We also test the informational content of sentiment given a particular setting. For example, we compare information content when a bank has a weak CAMELS score or not, when a bank receives a worse rating in the future or not, or by splitting the sample into the GFC period (up to 2011) and thereafter.

We find that the sentiment conveyed in different components of the CAMELS ratings have varying impacts in different circumstances. In particular, we find that controlling for a variety of factors, the sentiment supervisors express in describing many of the components predict future bank outcomes. More specifically, the sentiment conveyed in the capital, asset quality, management, and earnings sections provides significant information in predicting future outcomes for capital levels, problem loans, supervisory actions, and profitability, respectively. Evidence suggests, this relationship is driven by banks with better ratings when it comes to management, and banks with worse ratings when it comes to asset quality and earnings. This suggests that bank supervisors play a meaningful role in the surveillance of the banking system.

The rest of the paper is as follows. Section 2 provides a literature review and Section 3 provides a more detailed description of the bank examination process. Section 4 briefly describes the data and Section 5 investigates which sentiment score to use for our analysis. Then we provide our main econometric specification and results, followed by a conclusion.

2 Literature

Our paper mainly contributes to two strands of literature. The first strand is related to the information created though the bank examination process, mainly through the determination of supervisory ratings and related enforcement actions. Our paper also contributes to the more recent and growing literature on extracting sentiment information from financial text and finding out how such sentiment matters for predicting macro or financial outcomes.

Supervisory safety and soundness bank ratings and related supervisory actions have been shown to have useful private information. For example, most recently Gaul and Jones (2021) find that CAMELS ratings (and the Management component rating) have singificant predictive power fo future bank performance and risk measures relevant to bank regulators and supervisors. DeYoung et al. (2001) show that on-site commercial bank examinations produce value-relevant information

[1]Sensitivity to market risk was added to the framework in 1995. The text associated with this score is not widely available in the sample of banks we have, so we skip the analysis on this particular category.

about the future safety and soundness of banks reflected in bond prices of parent holding companies. Berger, Davies and Flannery (2000) note that supervisory assessments immediately following a bank examination generally contribute substantially to forecasting future problem loans and bank earnings, often exceeding the contribution of the market assessments. In addition, Berger and Davies (1998) show that CAMELS rating downgrades have a significant relationship with abnormal returns. Similarly, Jordan, Peek and Rosengren (2000) show that the announcement of formal supervisory actions have stock market reactions; only banks in the same region as the announcing bank, with similar exposures, are found to be affected. Finally, at a more aggregate level, going beyond just looking at publicly listed large commercial banks or bank holding companies, Peek, Rosengren and Tootell (1999) show that the percentage of commercial bank assets associated with the worst CAMELS ratings helps provide more accurate forecasts of macroeconomic variables such as the unemployment rate and inflation than can be predicted by Federal Reserve Board staff. Indeed, supervisory ratings are also used to determine FDIC deposit insurance premiums and examination frequency, given their high informational content. Our paper tries to add to this literature by providing more granular evidence of private information creation during the bank examination process. This involves extracting additional information from the bank examination reports through sentiment scoring.

Up to now, sentiment analysis in the economics and finance literature has generally been used on three different types of publicly available text: economic news, central bank communications, and corporate financial filings or earnings calls. Recent papers have analyzed new ways to combine commonly used lexicons with machine learning techniques to construct sentiment scores that accurately extract signal from economic news text to predict future changes in macroeconomic and financial cycle indicators. (See Nyman et al. (2018), Shapiro, Sudhof and Wilson (2020), and Kalamara et al. (2020).) In addition, other research combines these same techniques and utilizes them to determine how central bank communications (in the form of internal and external reports, FOMC meetings, and/or internal communications) accurately predict financial crises and future policy decisions or drive changes in certain macroeconomic indicators, such as inflation. (See Correa et al. (2017), Shapiro and Wilson (2019), and Hubert and Labondance (2017).) Most relevant to our work are the papers that look at earnings call transcripts and corporate financial filings. For example, several papers such as Jiang et al. (2019) and Price et al. (2012) show that the tone of the earnings calls lead to significant changes in stock market prices. More relevant to the banking industry, a few papers build off of Loughran and McDonald (2011) by using their finance lexicon to look at how sentiment expressed by corporate managers in official filings can predict future financial distress at banks. (See Gandhi, Loughran and McDonald (2019), Nopp and Hanbury (2015), and Gupta, Simaan and Zaki (2018).)

Our analysis seeks to expand the literature on sentiment analysis by looking at the relationship between the sentiment expressed in bank supervision examination reports used in supervisory assessments and a bank's CAMELS ratings, as well as a host of other quantitative factors that are traditionally indicative of a bank's financial soundness. While Goldsmith-Pinkham, Hirtle and Lucca (2016) use computational linguistic methods to categorize Matters Requiring Attention—one text product of bank supervision reports—into a number of topics, our study differs in that we actually calculate a sentiment score of examiners text, and we do so for the full-scope report that accompanies the CAMELS scoring.

3 Bank Examination Process

Every commercial bank undergoes a comprehensive bank exam about once a year. This process is led by one bank regulator, and sometimes in partnership with another regulator. In the case of State-Member Banks (SMB), examinations are performed alternating being led by the Federal Reserve Bank responsible for the SMB and the state-level financial regulator. The goal of these "full scope" exams is to assess the safety and soundness of the bank by reviewing any problems that were identified last round, scoring the bank on six areas, and generating an overall composite score. This scoring system is called CAMELS and is an acronym of the subscores:

- Capital Adequacy: Representing the ability of the bank to absorb losses
- Asset Quality: Representing the known and likelihood of losses the bank might face
- Management: Representing the quality of the management team, compliance function, audit function, and business strategy
- Earnings: Representing the ability of the bank to provide returns on their activities
- Liquidity: Representing the ability of the bank to absorb short term funding difficulties
- Sensitivity to Market Risk: Representing the bank's exposure to markets such as interest rate changes and marketable securities

Each subscore and the overall composite is rated from 1 (strongest) to 5 (weakest). And depending on the composite rating, the period between comprehensive exams will change. Banks with ratings of 3, 4, or 5 are considered "weak" banks or banks with weak ratings and banks with ratings of 1 or 2 are considered "strong" banks or banks with strong ratings. Weak banks are examined every 6 months, and strong banks may have their examinations up to once every 18 months.[2]

[2]Prior to The Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) effective in December 1992, the three federal regulators and many state banking departments examined banks on frequencies that varied

To finish the exam, the examination team writes a Report of Examination, which is vetted through the leadership of the regulator, and the team presents their findings to the executive team of the bank. Because of this vetting and presentation, the language of the examination is important. The examiner must be able to justify their rating both with financial ratios and other information, as explored in Bassett, Lee and Spiller (2015), and in the text of the examination.

The CAMELS ratings themselves are private supervisory information and weaker ratings mean that banks are subject to, not only more frequent examinations, but also restrictions on certain activities such as mergers and acquiitions, dividend payouts, or new activities.

Figure 1 shows the number of exams for each quarter in the sample. Figure 2 shows the distribution of composite scores in the sample through time. Most banks are strong at a given point in time, but over the GFC many banks received a weak composite rating.

4 Extracting Sentiment from Bank Exams

4.1 Sentiment Score Design

This section evaluates three lexicons and three methods to characterize the sentiment of the text within the bank exam reports. A lexicon refers to a pre-defined list of words assigned to positive or negative values, classifying each word with positive or negative sentiment, respectively. The three methods use different techniques to weight those positive and negative values.

Proper selection of the lexicon is important. As Loughran and McDonald (2011) highlight in their construction of a new finance-based lexicon, words often have different meanings in different domains. For example, as insinuated in the title of their seminal piece on the issue, while "liability" is generally considered a negative word in English vernacular, in finance or economics, the word is neutral. As a result, we explore three different lexicons to construct scores for the sentiment expressed by bank examiners and compare our results across lexicons. The Loughran and McDonald (LM) lexicon was created specifically for finance and economics-related text. The Financial Stability (FS) lexicon (Correa, Garud, Londonoy and Mislang, 2017) was created to analyze the sentiment of central bank financial stability reports. Finally, Hu and Liu's (Hu and Liu, 2004, QDAP) opinion lexicon was constructed with broader sentiment analysis in mind and are not tailored to specific text in finance/economics.[3]

---

by banks condition or past ratings. FDICIA established a uniform criteria based on size and risk profile, which has, since then, changed over time. The threshold for the risk profile last changed as part of The Riegle Community Development and Regulatory Improvement Act of 1994 and the threshold for size last changed in 2007 as part of the Interim Rules to Implement the Examination Amendments of 2006, which require annual examinations of banks with assets greater than $500 million. See Rezende (2014) for more details.

[3]Other lexicons include the Harvard-IV-4 psychosocial dictionary that is also broad-based, and Elaine Henry's

We apply these dictionaries using three methods. The first sentiment score approach is oftentimes called bag-of-words, as it neglects the role that grammar, syntax, and other context-specific traits play in the overall sentiment of text and merely bases the analysis on individual words. The score is based on the following index calculation:

$$\text{sentiment index}_{i,t} = \frac{\#PositiveWords - \#NegativeWords}{\#NegativeWords + \#PositiveWords},$$

where i represents either the report as a whole or the subsections of the report, which correspond to the different components of the CAMELS rating.

However, merely utilizing a raw word count is likely insufficient; a particular word's frequency is directly related to the length of the document, not its importance. Moreover, a word that occurs infrequently likely carries more weight to the average reader than a word that is commonplace in the document, and raw word count does not account for this. To address this, we apply a sentiment score method as our main method that adjusts the sentiment scores for term frequency and inverse document frequency, a weighting mechanism commonly referred to as tf-idf. Terms that are used frequently are down-weighted using these formulas

$$\begin{aligned}
\text{term frequency: } tf_{t,d} &= \frac{\text{number of term}}{\text{total words}}, \\
\text{inverse term frequency: } idf_{t,D} &= log\Big(\frac{N}{|d \in D : t \in d|}\Big), \\
tfidf_{t,d,D} &= tf_{t,d} * idf_{t,D},
\end{aligned}$$

where N is the number of documents in the corpus and $|d \in D : t \in d|$ is the number of documents where the term t appears (i.e., $tf(t, d) \neq 0$).

Finally, in order to overcome the bag-of-words method's failure to incorporate context into its calculations, we consider a third method that also takes valence shifters into account. Valence shifters are words that either change the polarity of the word (i.e., negators like "not" and "never") or amplify the word (i.e., words like "slightly" or "extremely"). We use an R package called -sentimentR-, which is designed to consider four words before and two words after a dictionary-weighted word to search for valence shifters and to weight words differently based on their association with various amplifiers.

(Henry, 2008, HE) lexicon was created to develop sentiment measures for earnings press releases.

4.2 Evaluating Sentiment Scores

Because there are numerous ways to capture sentiment, the first step is to select a few methods that seem best suited to capture the information content of a bank exam. We focus on testing three of the dictionaries including two that are most tied to financial information: FS, LM, and QDAP.

To assess the performance of the sentiment scores, we compare how well sentiment measures explain CAMELS scores. In particular, we regress CAMEL scores on sentiment. We do this for the composite score and sentiment scores built using all of the text, and also for scores of individual sections and sentiment scores based on only that section of the exam text. The exact specification is

$$\text{CAMELS score}_s = \text{Sentiment score}_{s,m,l},$$

where s is the section of the exam, m is the sentiment score method, and l is the sentiment score lexicon.

Table 1 summarizes the results by reporting the R-squared and adjusted R-squared for each specification. The three methods are listed in the first column: polar, weighted polar (tf-idf), and valence. The second column is the bank exam section of text used, and the three lexicons are listed across the top: FS, LM, and QDAP. The best performer based on R-squared or adjusted R-squared are listed in the last two columns. Overall, the LM lexicon is the best performer. Comparing the level of the R-squared or R-squared score, the regular polar generally performs better than the weighted polar (tf-idf). The valence method generally performs better than the regular polar. We retain the two best methodologies to use for the main analysis in next section: LM regular polar and LM valence.

Indeed, a simple chatterplot example in Figure 3 illustrates that the LM dictionary has nice differentiating properties when it comes to the language used in bank exam reports. Positive words in the LM dictionary are are used more often in Earnings sections of stronger bank exam reports, while negative words in the LM dictionary are more frequently used in weaker bank exam reports.

To provide a better idea of how the LM polar and LM valence-based sentiment scores are associated with CAMELS ratings in geneneral, Figure 4 shows that in general these two methods show differentiating properties when it comes tov stronger vs. weaker banks. We show that a more granular distinction of stronger vs. weaker banks also relate to helping predict bank outcomes in the next Section. We note here that the polar sentiment score ranges from negative one to positive one. The weighting in the valence method makes the range of that score a bit narrower.

5 Data

Our examination data primarily consist of state member banks (SMBs) because SMBs are directly regulated by the Federal Reserve. In addition to the text of the bank exams, we also have information on financial performance and supervisory activity. More specifically, we have counts of how many times regulators cited matters requiring attention (MRA) or matters requiring immediate attention (MRIA) in a bank exam. The remaining bank performance data come from Consolidated Reports of Condition and Income, also referred to as the Call Reports, which are quarterly financial statements that commercial banks are required to report (FFIEC 031 and 041).

Within the bank exam report, the data of the financial statements used is indicated. We use that date for merging the exam and Call Report data. The final sample used for testing has about 5,400 observations between 2004:Q1 and 2016:Q2.

Table 2 summarizes the data. The top panel reports the sentiment scores that are discussed in the previous section. The bottom panel reports the control variables, some of which also serve as outcome variables. To control for outliers, all the various financial ratios are winsorized at 1 and 99 percent. However, to maintain confidentiality, we show the 5th and 95th percentiles of all the variables in our sample in Table 2.

6 Econometric Specification

This section will explore the information content of bank exams and how that soft information is related to future bank developments. Section 4.2 evaluated how well the various sentiment methods captured the CAMELS ratings. Now that the two best methods have been established, we will test to see what additional information, beyond the CAMELS score, is contained in the text.

For each CAMELS score category, we identify performance measures related to that category: capital, asset quality, management, earnings, liquidity, and the composite score. We then test to see if sentiment from an exam is related to a performance measure one year later after controlling for other observables, including the ratings themselves. Any statistically significant results that link sentiment to bank outcomes will be that much more profound, as we use the sentiment methods that are the most highly correlated with the CAMELS scores.

The first set of regressions uses the full sample between 2004:Q1 and 2016:Q2. Additional tests will subset the data in various ways. The baseline specifications follow this structure:

$$\begin{aligned} \text{outcome}_{i,t} &= \rho \text{ outcome}_{i,t-1} + \beta \text{ sentiment}_{i,c,t-1} + \gamma \; log(\text{assets}_{i,t}) \\ &\quad + \Sigma_{n=1}^{4} \; \psi_n \text{ CAMEL dummy}_{i,n,t-l} + \theta_i + \phi_t + \epsilon_{i,t}, \end{aligned}$$

for bank i, in period t, for bank exam component c, and where θ_i and ϕ_t are bank and time fixed effects, respectively. Table 2 lists the outcome variables for each of the exam sections. For data that comes from the Call Report (all variables except MRA/MRIA), the regression is testing outcomes one year after the exam. Because MRAs/MRIAs get completely refreshed at a "full scope" exam, these regressions are run using adjacent exams. Exams are generally 12 to 18 months apart.

The lagged value of the outcome variable (value at the time of the exam with the sentiment score) is included because many of these performance metrics are persistent, and again, in order to assess the information content of the exam text, we are trying to strip out simple observable information. The coefficient of interest is β, the loading on the sentiment score. Notice that the sentiment score is subscripted by c, the exam component. The sentiment score can be based on all of the text in the case of testing with the composite CAMELS score, or the sentiment score could be based on text from one component such as asset quality that is then tested with the asset quality CAMELS score. By including the set of CAMELS composite score dummies, we are also controlling for the general performance of the bank. The specification design is testing whether the information content of the exam text is associated with future bank outcomes.

7 Main Results

This section provides a description of the results of our baseline regressions by different categorical ratings, including the composite.

Composite

Table 3 shows the results of the regressions using a sentiment score based on all of the exam text. One outcome variable is based on the number of matters requiring attention (MRA) and matters requiring immediate attention (MRIA). The second outcome variable takes a value of one if there are any MRAs/MRIAs and a value of zero otherwise. More of these "matters" are a sign that the bank is not operating properly and has supervisory concerns requiring attention. The coefficient on the lag sentiment score is negative as expected and is highly significant across all specifications. Higher or more positive sentiment in the exam text is associated with a lower number of MRA/MRIAs at the next comprehensive exam. To put the magnitude of the coefficient in perspective, a one-standard deviation change in either the polar or valence LM sentiment (0.26 and 0.12, respectively) leads to about three-quarters of a decrease in an MRA or MRIA or increases the likelihood of an

MRA or MRIA by about 10 percentage points. Given that there is just over five MRA/MRIAs per exam on average, these are significant results. Current MRAs or MRIAs and CAMELS composite ratings themselves are also highly relevant to the future number and likelihood of MRAs or MRIAs.

Capital Adequacy

Table 4 shows the results of using the text in the capital adequacy section. The coefficient of interest is positive and significant across all specifications. Higher sentiment is associated with higher (stronger) capital ratios one year later. Applying a one-standard deviation change in either the polar or valence LM sentiment existing in the capital section of exams (0.55 and 0.17, respectively), more positive sentiment leads to an increase of a little over 15 basis points in tier 1 or CET1 capital ratios. These relatively small (though highly statistically significant) effects compared to the standard deviation of capital ratios themselves (both over 10 percentage points) are consistent with a bank's capital adequacy status being well-informed simply by using regulatory ratios. However, the sentiment score still picks up information related to future capital ratios. Interestingly, only when the CAMELS capital-specific rating is a 5, the worst rating, does the rating have a statistically significant (negative) effect on future capital ratios.

Asset Quality

Table 5 shows the results of testing measures of asset quality. Three ratios are tested: loan loss provisions, four quarters of net charge-offs (NCOs), and delinquencies. All measures are defined relative to all loans. Note that in all cases a higher ratio is associated with worse asset quality. The coefficient of interest on lag sentiment is negative and statistically significant across all specifications, indicating that more positive sentiment is associated with improved asset quality one year later. A one-standard deviation improvement in either the polar or valence LM sentiment existing in the asset quality section of bank exams (0.32 and 0.15, respectively) leads to about 5 basis point decrease in loan loss provision rates, about 2 basis point decrease in net charge-off rates, and more than 10 basis point decrease in delinquency rates one year afterwards, which are all economically meaningful results. In addition, unlike in the capital results, all CAMELS asset quality-specific ratings are also highly significant.

Management

Tests on the management text use the same outcome variables as the composite score and are reported in table 6. The coefficient of interest is negative as expected and is significant across all specifications. Positive sentiment on the text in the management section is associated with lower MRAs and MRIAs at the next comprehensive exam. Note that the magnitudes of the coefficients are much lower than in the composite regressions. The R-squareds are slightly lower as well. This decreased explanatory power is consistent with MRAs/MRIAs not all necessarily being associated with management specific issues. The results, however, are still highly statistically and economically significant. Indeed, the CAMELS management-specific ratings are also highly significant.

Earnings Quality

Table 7 shows the results of testing measures of profitability: return on assets (ROA) and pre-provision net revenue (PPNR) divided by assets. Higher values for both of these ratios are associated with higher earnings and profitability. The coefficient of interest on lag sentiment is positive across all specifications, indicating that higher sentiment is associated with improved profitability one year later. A one-standard deviation improvement in either the polar or valence LM sentiment existing in the Asset Quality section of bank exams (0.48 and 0.20, respectively) leads to about 6 basis point increase in ROA and PPNR rates. These results are both statistically and economically significant results. In fact, these effects are high compared to the standard deviations of ROA and PPNR rates, which are 39 and 53 basis points, respectively.

Liquidity

Table 8 shows the results of testing liquidity measures. The first measure is securities as a share of assets. The second measure is securities and cash as a share of assets. The coefficient of interest changes signs and is not statistically different from zero. While these measures are clearly related to liquidity, it is harder to determine a bank's liquidity needs. Supervisors would look at both levels and needs when assessing this CAMELS score. Since the financial crisis, internal liquidity stress tests are also reviewed. These data is not available in the Call Reports. In addition, the word counts for this section are generally low, making it hard for the sentiment score to capture additional information.

8 Results using Subsamples

In this section, we provide more results based on subsamples of the bank exam data.

8.1 Strong Score Versus Weak Score

The next set of regressions relate bank outcomes and sentiment but separately test banks that had a strong or weak rating in the prior exam for that section of the exam. Most banks have a strong score (1 or 2). When a bank has a lower quality score (3, 4, or 5), especially for the composite score, the bank is subject to more restrictions. As a result, banks have strong incentives to improve the score while examiners are likely focused on any further deterioration.

The sentiment score regression is run separately depending on the CAMELS score.

$$\begin{aligned}
\text{outcome}_{i,t} &= \rho_g \text{ outcome}_{t-1} + \beta_g \text{ sentiment}_{t-1} + \gamma_g \, log(\text{assets}_t) \\
&\quad + \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n} + \theta_i + \phi_t + \epsilon_{i,t}, \text{if } t-1 \text{ component rating } \in [1,2] \\
\text{outcome}_{i,t} &= \rho_b \text{ outcome}_{t-1} + \beta_b \text{ sentiment}_{t-1} + \gamma_b \, log(\text{assets}_t) \\
&\quad + \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n} + \theta_i + \phi_t + \epsilon_{i,t}, \text{if } t-1 \text{ component rating } \in [3,4,5]
\end{aligned}$$

The coefficient of interest is still β, and the test results are summarized in table 9.

Composite

For the composite score test that uses all of the exam text, β is negative and significant across all specifications. As sentiment increases, MRAs are likely to decrease regardless of the composite score. The coefficient magnitudes are very similar to the results in table 3.

Capital Adequacy

The sentiment scores do not load in the tests of the capital section. Recall, the results for the full sample were economically small. The capital information is likely well-informed by ratios alone.

Asset Quality

The β coefficients are still negative and generally statistically significant for the asset quality section. The results are stronger for banks with weak scores. This result is consistent with a mean reversion story. Weaker banks should have higher ratios that can be lowered when checked one year later. The result can also imply that supervisors tend to focus on conveying more granular information through their write-ups, especially in times of weak asset quality at banks.

Management

The β coefficients are negative and often statistically significant for the management regressions. However, the results are stronger for the strong banks. This is consistent with strong banks being better (faster) at correcting MRA/MRIAs. The results can also imply that supervisors pay more attention to management when banks are strong.

Earnings Quality

The earnings results for weak banks are much stronger. This is consistent with another reversion to the mean story. Banks with low earnings improve those metrics over the year. Weak banks have more room to improve. Similarly to the Asset Quality, the result can also imply that supervisors tend to convey more granular information in times of weak earnings at banks.

Liquidity

The liquidity measures continue to not load.

8.2 Same or Better Score versus Worse Score

The next set of regressions relate bank outcomes and sentiment but separately test banks that had the same or a better rating between $t-1$ and t and banks that received a worse score.

The sentiment score regression is run separately.

$$
\begin{aligned}
\text{outcome}_t &= \rho_w \text{ outcome}_{t-1} + \beta_w \text{ sentiment}_{t-1} + \gamma_w \; log(\text{assets}_t) \\
&\quad + \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n} + \theta_i + \phi_t + \epsilon_{i,t}, \text{if rating change} > 0 \\
\text{outcome}_t &= \rho_{sb} \text{ outcome}_{t-1} + \beta_{sb} \text{ sentiment}_{t-1} + \gamma_{sb} \; log(\text{assets}_t) \\
&\quad + \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n} + \theta_i + \phi_t + \epsilon_{i,t}, \text{if rating change} \leq 0
\end{aligned}
$$

The coefficient of interest is still β, and the test results are summarized in table 10.

Composite

The β coefficients for the composite score regressions continue to be negative and statistically significant in many of the specifications. Results are slightly stronger for banks with stable or improving scores. This is again consistent with better run banks being able to resolve MRAs/MRIAs faster.

Asset Quality

The asset quality results are mixed. The β coefficients are generally negative across the specifications, but statistical significance only occurs when testing banks with stable or improving scores. It is somewhat surprising that banks that received a worse score did not have a positive coefficient. Supervisors that downgrade a bank on the asset quality must be identifying deeper issues than evident from standard performance metrics.

Management

The β coefficients on the management score regressions remain negative, and results are stronger among banks with a stable or improving management CAMELS score. Again, it is a bit surprising that the downgraded banks did not have a positive coefficient.

Earnings Quality

The results for earnings continue to be positive and statistically significant. Interestingly, the results are about the same across the samples. Good performers do no better than banks that got downgraded.

Capital Adequacy and Liquidity

The capital adequacy measures and the liquidity measures do not load.

8.3 GFC versus post-GFC

The next set of regressions relate bank outcomes and sentiment but separately for exams during the financial crisis (2004-2011) and after the financial crisis (2012-2016).

The sentiment score regression is run separately.

$$\begin{aligned} \text{outcome}_t &= \rho_{GFC}\ \text{outcome}_{t-1} + \beta_{GFC}\ \text{sentiment}_{t-1} + \gamma_{GFC}\ log(\text{assets}_t) \\ &\quad + \Sigma_{n=1}^{4}\ \psi_n\ \text{CAMEL dummy}_{i,n} + \theta_i + \phi_t + \epsilon_{i,t}, \text{if } t \geq 2012 \\ \text{outcome}_t &= \rho_{post}\ \text{outcome}_{t-1} + \beta_{post}\ \text{sentiment}_{t-1} + \gamma_{post}\ log(\text{assets}_t) \\ &\quad + \Sigma_{n=1}^{4}\ \psi_n\ \text{CAMEL dummy}_{i,n} + \theta_i + \phi_t + \epsilon_{i,t}, \text{if } t \leq 2011 \end{aligned}$$

The coefficient of interest is still β. The results are summarized in table 11.

Composite
The β coefficients for the composite score regressions continue to be negative and statistically significant in many of the specifications. Results are slightly stronger during the financial crisis. This is consistent with MRAs/MRIAs generally having higher magnitudes during the crisis.

Capital Adequacy
The results for Capital Adequacy are mixed, though the coefficients are all consistently positive. When it comes to the valence sentiment, it appears to be driven by the post-GFC period.

Asset Quality
The β coefficients remain negative in the asset quality tests. The statistical significance is slightly higher during the financial crisis. This is again consistent with generally higher magnitudes during the crisis, more room to improve.

Earnings Quality
The β coefficients on the earnings tests remain positive and significant across specifications. The magnitudes are higher during the crisis consistent with more room to improve.

Management and Liquidity
The management tests and the liquidity measures do not load.

9 Conclusion

In this paper, we analyze whether the bank supervisory process provides useful insight into bank future outcomes by using textual analysis on commercial bank examination reports. In particular, we find that controlling for a variety of factors, the sentiment supervisors express in describing many of the components predict future bank outcomes. More specifically, the sentiment conveyed in the capital, asset quality, management, and earnings sections provides significant information in predicting future outcomes for capital levels, problem loans, supervisory actions, and profitability, respectively. We show that this relationship is driven by banks with better ratings when it comes to management, and banks with worse ratings when it comes to asset quality and earnings. Our results on the relationship between positive sentiment and future positive bank outcomes is most striking when it comes to bank earnings. All of this suggests that bank supervisors play a meaningful role in the surveillance of the banking system by creating and sharing information that is embedded in bank examination reports through the bank examination process.

There are several caveats to our analysis, however. First, we may be capturing the effects of other types of information in the bank examination process rather than the sentiment in the exams itself. However, even if this were true, this still implies that meaningful information is created, documented, and shared in the supervisory process. In turn, this is important for understanding the role that supervision plays in maintaining the safety and soundness of the banking system. Second, we only show that the supervisory process appears to help in monitoring future bank outcomes and the banking system as a whole. However, we have nothing to say about the efficiency or the supervisory process, for example, if the added-value of bank supervision significantly outweighs the costs of maintaining a large number of personnel and resources in this process. Third, our analysis is largely for small to medium sized banks. The degree in which supervisory information is useful for bank outcomes for large bank holding companies may be different.

References

Bassett, William F., Seung Jung Lee and Thomas Popeck Spiller. 2015. "Estimating changes in supervisory standards and their economic effects." *Journal of Banking and Finance* 60:21–43.

Berger, Allen N. and Sally M. Davies. 1998. "The Information Content of Bank Examinations." *Journal of Financial Services Research* 14(2):117–144.

Berger, Allen N., Sally M. Davies and Mark J. Flannery. 2000. "Comparing Market and Supervisory Assessments of Bank Performance: Who Knows What When?" *Journal of Money, Credit and Banking* 32(3):641–667.

Correa, R., K. Garud, J. Londonoy and N. Mislang. 2017. "Sentiment in central banks financial stability reports." *Federal Reserve Board International Finance Discussion Paper* pp. 811–841.

DeYoung, Robert, Mark J. Flannery, William W. Lang and Sorin M. Sorescu. 2001. "The Information Content of Bank Exam Ratings and Subordinated Debt Prices." *Journal of Money, Credit and Banking* 33(4):900–925.

Gandhi, P., T. Loughran and B. McDonald. 2019. "Using Annual Report Sentiment as a Proxy for Financial Distress in U.S. Banks." *Journal of Behavioral Finance* 20(4):424–436.

Gaul, Lewis and Jonathan Jones. 2021. "CAMELS rAtings and Their Information Content." *OCC Working Papers* .

Goldsmith-Pinkham, P., B. Hirtle and D. Lucca. 2016. "Parsing the Content of Bank Supervision." *Federal Reserve Bank of New York Staff Reports* (770).

Gupta, A., M. Simaan and M. Zaki. 2018. "When Positive Sentiment is Not So Positive: Textual Analytics and Bank Failure." *Working Paper* .

Henry, E. 2008. "Are investors influenced by how earnings press releases are written?" *Journal of Business Communication* 45(4):363–407.

Hu, M. and B. Liu. 2004. "Mining Opinion Features in Customer Reviews." *proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* p. 168177.

Hubert, P. and F. Labondance. 2017. "Central bank sentiment and policy expectations." *Bank of England Staff Working Paper* .

Jiang, Fuwei, Joshua Lee, Xiumin Martin and Guofu Zhou. 2019. "Manager sentiment and stock returns." *Journal of Financial Economics* 132(1):126 – 149.

Jordan, John S., Joe Peek and Eric S. Rosengren. 2000. "The Market Reaction to the Disclosure of Supervisory Actions: Implications for Bank Transparency." *Journal of Financial Intermediation* 9(3):298 – 319.

Kalamara, E., A. Turrell, C. Redl, G. Kapetanios and S. Kapadia. 2020. "Making text count: economic forecasting using newspaper text." *Bank of England Working Paper* .

Loughran, T. and B. McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance* 66(1):35–65.

Nopp, C. and A. Hanbury. 2015. "Detecting Risks in the Banking System by Sentiment Analysis." *Working paper* .

Nyman, R., S. Kapadia, D. Tuckett, D. Gregory, P. Ormerod and R. Smith. 2018. "News and narratives in financial systems: exploiting big data for systemic risk assessment." *Bank of England Working Paper* .

Peek, Joe, Eric S. Rosengren and Geoffrey M. B. Tootell. 1999. "Is Bank Supervision Central to Central Banking?*." *The Quarterly Journal of Economics* 114(2):629–653.

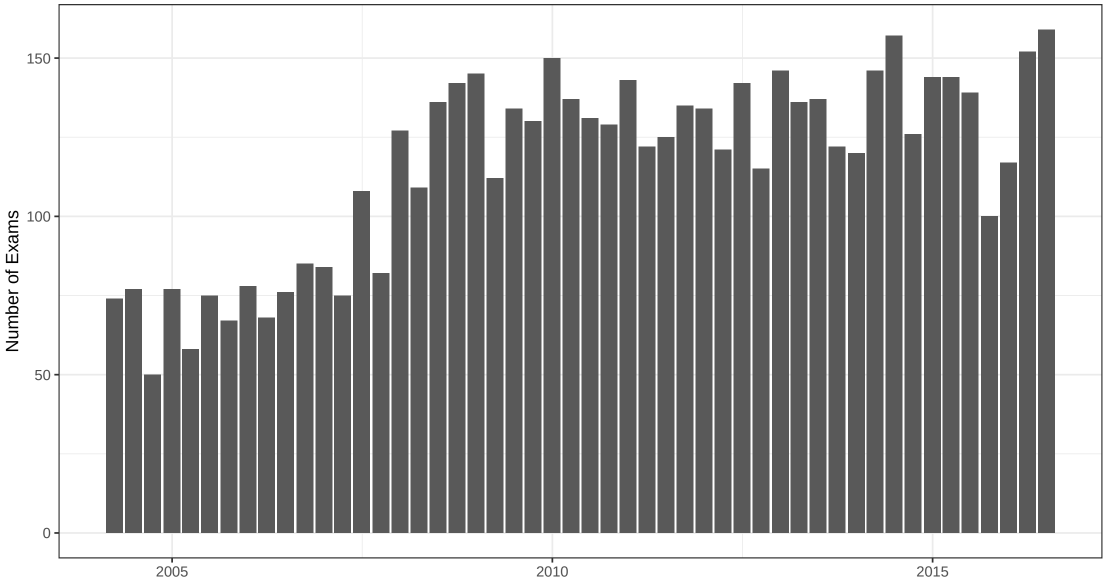
Price, S. McKay, James S. Doran, David R. Peterson and Barbara A. Bliss. 2012. "Earnings conference calls and stock returns: The incremental informativeness of textual tone." *Journal of Banking & Finance* 36(4):992 – 1011.

Rezende, Marcelo. 2014. "The Effects of Bank Charter Switching on Supervisory Ratings." *Finance and Economics Discussion Series* (2014-20).

Shapiro, A.H. and D. Wilson. 2019. "Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives using Text Analysis." *Federal Reserve Bank of San Francisco Working Paper* .

Shapiro, A.H., M. Sudhof and D. Wilson. 2020. "Measuring News Sentiment." *Federal Reserve Bank of San Francisco Working Paper* .

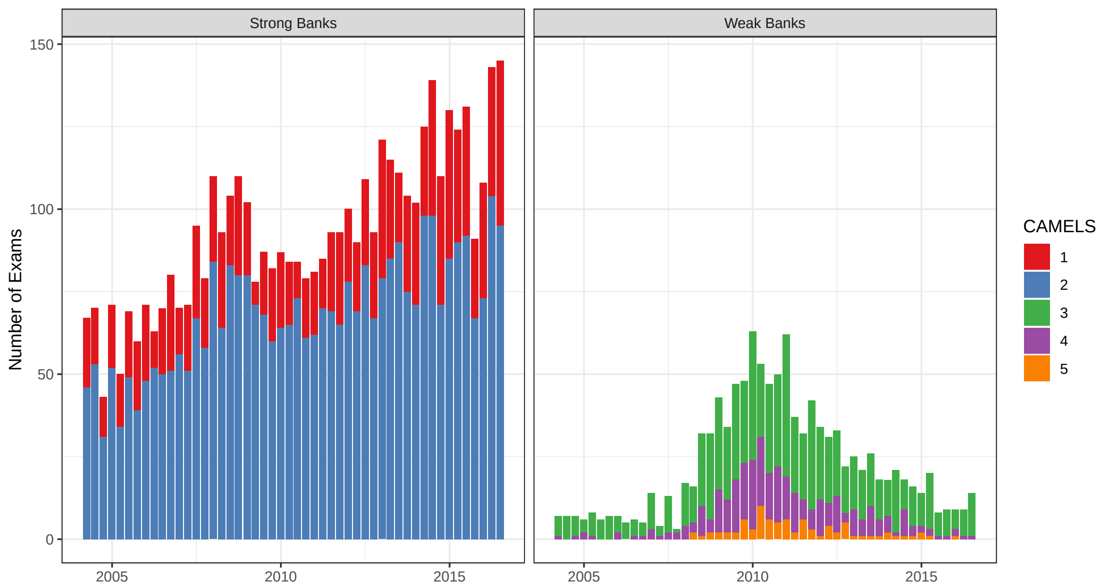
Figure 1: Exams in the Sample



Note: Data between 2004:Q1 and 2016:Q2.

Source: Confidential bank exams.

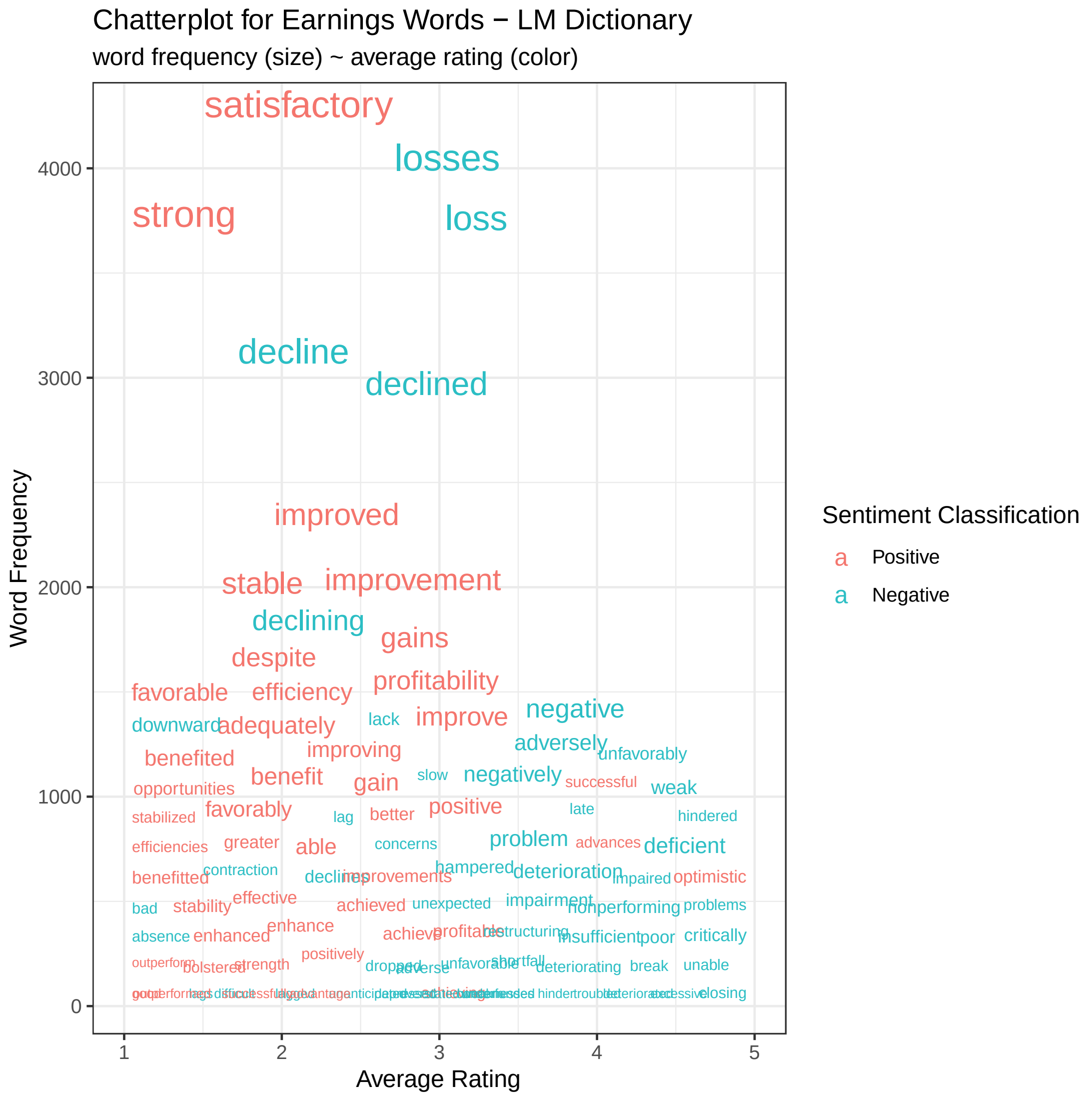
Figure 2: Distribution of CAMELS Scores–Composite Score



Note: Data between 2004:Q1 and 2016:Q2.

Source: Confidential bank exams.

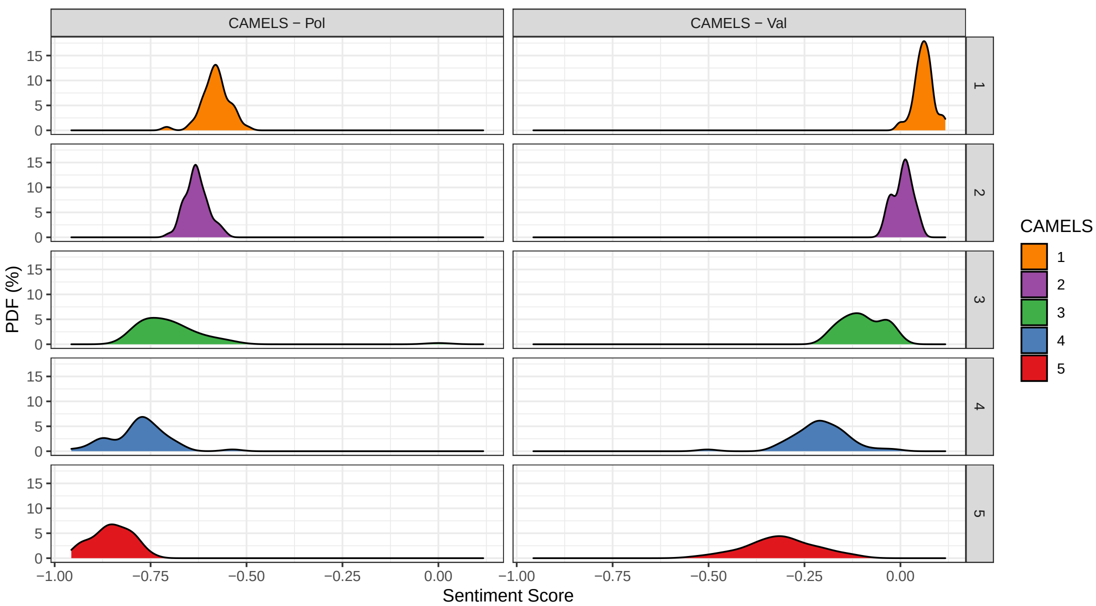
Figure 3: Frequencies and Average Ratings associated with Words in the LM Dictionary in Earnings Sections of Bank Exams



Note: Data between 2004:Q1 and 2016:Q2.

Source: Confidential bank exams.

Figure 4: LM Polar and Valence Sentiment Score Distributions by Composite CAMELS Score



Note: Data between 2004:Q1 and 2016:Q2. Distributions are kernel densities.

Source: Confidential bank exams.

Table 1: Sentiment Score Performance: Composite and Section CAMELS Score

This table reports the model performance (R-squared) from the specification

$$\text{CAMELS score}_s \quad = \quad \text{Sentiment score}_{s,m,l},$$

where s is the section of the exam, m is the sentiment score method, and l is the sentiment score lexicon. The methods are listed in the first column: polar, polar (weighted, tf-idf), and valence. The second column is the bank exam section of text used, and the three lexicons are list across the top: FS, LM, and QDAP. The best performer based on R-squared or adjusted R-squared are listed in the last two columns.

	Exam	FS		LM		QDAP		Highest	
	section	R-squared	Adj. R-sqd.	R-squared	Adj. R-sqd.	R-squared	Adj. R-sqd.	R-squared	Adj. R-sqd.
Polar, regular	Total	0.16	0.16	0.25	0.25	0.10	0.10	LM	LM
	Capital	0.06	0.06	0.21	0.21	0.15	0.15	LM	LM
	Asset quality	0.03	0.03	0.10	0.10	0.08	0.08	LM	LM
	Management	0.13	0.13	0.14	0.14	0.12	0.12	LM	LM
	Earnings	0.20	0.20	0.30	0.30	0.32	0.32	QDAP	QDAP
	Liquidity	0.02	0.02	0.09	0.09	0.10	0.10	QDAP	QDAP
Polar, tf-idf	Total	0.15	0.14	0.10	0.10	0.14	0.14	FS	FS
	Capital	0.13	0.13	0.13	0.13	0.11	0.11	LM	LM
	Asset quality	0.05	0.05	0.07	0.06	0.05	0.05	LM	LM
	Management	0.15	0.15	0.10	0.10	0.13	0.13	FS	FS
	Earnings	0.23	0.23	0.24	0.24	0.25	0.25	QDAP	QDAP
	Liquidity	0.02	0.02	0.05	0.05	0.05	0.05	QDAP	QDAP
Valence	Total	0.20	0.20	0.33	0.33	0.04	0.04	LM	LM
	Capital	0.12	0.12	0.29	0.28	0.20	0.20	LM	LM
	Asset quality	0.08	0.08	0.18	0.18	0.12	0.12	LM	LM
	Management	0.14	0.14	0.18	0.18	0.16	0.16	LM	LM
	Earnings	0.24	0.24	0.35	0.35	0.37	0.37	QDAP	QDAP
	Liquidity	0.03	0.03	0.14	0.14	0.13	0.13	LM	LM

Table 2: Summary Statistics

This table reports summary statistics of variables used in the baseline regression tests. Control variables that are ratios have been winsorized at 1 and 99 percent. However, we show the 5th and 95th percentiles of all the variables to maintain confidentiality of the data.

Variable			Obs	Mean	Std. Dev.	5th Percent.	95th Percent.
Sentiment Scores							
	Polar						
		Composite	5,321	-0.038	0.263	-0.500	0.387
		Capital	5,401	0.272	0.545	-0.625	1
		Asset quality	5,401	-0.365	0.315	-0.818	0.167
		Management	5,321	0.209	0.482	-0.600	1
		Earnings	5,401	0.160	0.476	-0.636	1
		Liquidity	5,401	0.224	0.489	-0.538	1
	Valence						
		Composite	5,321	-0.009	0.123	-0.241	0.173
		Capital	5,401	0.073	0.169	-0.237	0.314
		Asset quality	5,401	-0.147	0.150	-0.385	0.101
		Management	5,321	0.073	0.179	-0.253	0.330
		Earnings	5,401	0.061	0.195	-0.293	0.345
		Liquidity	5,401	0.068	0.142	-0.176	0.290
Control Variables							
		MRIA/MRA sum	5,321	2.899	5.323	0	14
		MRIA/MRA dummy	5,321	0.390	0.488	0	1
		Tier 1 ratio	5,401	15.853	10.537	9.419	28.387
		CET1 ratio	5,398	15.763	10.172	9.361	28.436
		Loan loss provisions/loans	5,398	0.304	0.580	0	1.366
		4-qtr net charge-offs/assets	5,329	0.110	0.188	-0.010	0.486
		Delinquent loans/loans	5,398	2.636	2.633	0.078	7.878
		4-qtr ROA	5,401	0.312	0.390	-0.348	0.800
		4-qtr PPNR/assets	5,401	0.784	0.533	-0.043	1.526
		Securities/assets	5,401	21.141	13.853	2.327	48.153
		(Cash+securities)/assets	5,401	28.408	14.776	8.625	56.529
		CAMEL dummy 2	5,401	0.166	0.373	0	1
		CAMEL dummy 3	5,401	0.057	0.233	0	1
		CAMEL dummy 4	5,401	0.022	0.146	0	0
		CAMEL dummy 5	5,401	0.006	0.076	0	0
		Ln(assets)	5,401	12.494	1.495	10.417	15.145

Table 3: Composite Score Regressions

This table shows the regression results from testing

$$\begin{aligned} \text{outcome}_{i,t} &= \rho \text{ outcome}_{i,t-1} + \beta \text{ sentiment}_{i,t-1} + \gamma \, log(\text{assets}_{i,t}) \\ &+ \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n,t-l} + \theta_i + \phi_t + \epsilon_{i,t}, \end{aligned}$$

where outcome is either the summation of all matters requiring [immediate] attention (MRA/MRIA) for a bank or a dummy variable that takes the value of one if any MRAs or MRIAs exist for a bank and zero otherwise. The sentiment score is based on all exam text.

	(1)	(2)	(3)	(4)
	MRA/MRIA Sum		MRA/MRIA Dummy	
VARIABLES	Polar	Valence	Polar	Valence
Lag sentiment	-3.181***	-6.174***	-0.403***	-0.812***
	(0.345)	(0.826)	(0.0299)	(0.0665)
Lag MRA/MRIA Sum	-0.344***	-0.350***		
	(0.0222)	(0.0221)		
Lag MRA/MRIA dummy			-0.506***	-0.519***
			(0.0181)	(0.0180)
CAMELS 2 dummy	1.211***	1.256***	0.0867***	0.0901***
	(0.224)	(0.224)	(0.0209)	(0.0209)
CAMELS 3 dummy	4.575***	4.632***	0.160***	0.161***
	(0.427)	(0.428)	(0.0301)	(0.0301)
CAMELS 4 dummy	6.488***	6.501***	0.162***	0.154***
	(0.697)	(0.703)	(0.0425)	(0.0431)
CAMELS 5 dummy	6.988***	6.909***	0.236***	0.215***
	(1.094)	(1.095)	(0.0547)	(0.0559)
Ln(total assets)	0.787***	0.767***	0.0162	0.0141
	(0.279)	(0.280)	(0.0290)	(0.0289)
Constant	-6.786*	-6.484*	0.431	0.468
	(3.483)	(3.496)	(0.363)	(0.362)
Observations	5,321	5,321	5,321	5,321
Fixed effects	0.515	0.513	0.611	0.608
R-squared	bank & year	bank & year	bank & year	bank & year
Adj. R-squared	0.399	0.396	0.518	0.514

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Capital Score Regressions

This table shows the regression results from testing

$$\begin{aligned} \text{outcome}_{i,t} &= \rho \text{ outcome}_{i,t-1} + \beta \text{ sentiment}_{i,t-1} + \gamma \, log(\text{assets}_{i,t}) \\ &+ \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n,t-l} + \theta_i + \phi_t + \epsilon_{i,t}, \end{aligned}$$

where outcome is either the tier 1 ratio or the common equity tier 1 (CET1) ratio. The sentiment score is based on exam text in the capital section.

	(1)	(2)	(3)	(4)
	Tier 1 Ratio		CET1 Ratio	
VARIABLES	Polar	Valence	Polar	Valence
Lag sentiment	0.282***	1.032***	0.265***	0.975***
	(0.0940)	(0.318)	(0.0919)	(0.315)
Lag Tier 1 ratio	0.344***	0.343***		
	(0.0621)	(0.0625)		
Lag CET1 ratio			0.351***	0.350***
			(0.0628)	(0.0631)
CAMELS 2 dummy	0.125	0.128	0.0982	0.101
	(0.0930)	(0.0932)	(0.0933)	(0.0935)
CAMELS 3 dummy	-0.0970	-0.0860	-0.106	-0.0949
	(0.172)	(0.167)	(0.176)	(0.172)
CAMELS 4 dummy	-0.322	-0.264	-0.329	-0.274
	(0.276)	(0.269)	(0.278)	(0.271)
CAMELS 5 dummy	-2.144***	-2.045***	-2.151***	-2.056***
	(0.371)	(0.359)	(0.381)	(0.370)
Ln(total assets)	-1.875***	-1.878***	-1.850***	-1.853***
	(0.494)	(0.495)	(0.472)	(0.473)
Constant	33.56***	33.60***	33.09***	33.13***
	(6.687)	(6.704)	(6.409)	(6.425)
Observations	5,401	5,401	5,397	5,397
R-squared	0.922	0.922	0.916	0.916
Fixed effects	bank & year	bank & year	bank & year	bank & year
Adj. R-squared	0.903	0.903	0.896	0.896

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Asset Quality Score Regressions

This table shows the regression results from testing

$$\begin{aligned}\text{outcome}_{i,t} &= \rho\ \text{outcome}_{i,t-1} + \beta\ \text{sentiment}_{i,t-1} + \gamma\ log(\text{assets}_{i,t}) \\ &+ \Sigma_{n=1}^{4}\ \psi_n\ \text{CAMEL dummy}_{i,n,t-l} + \theta_i + \phi_t + \epsilon_{i,t},\end{aligned}$$

where outcome is loan loss provisions/loans, net charge-offs/loans, or delinquencies/loans. The sentiment score is based on exam text in the asset quality section.

	(1)	(2)	(3)	(4)	(5)	(6)
	Loan Loss Provisions/Loans		4-qtr Net Charge-offs/Loans		Delinquencies/Loans	
VARIABLES	Polar	Valence	Polar	Valence	Polar	Valence
Lag sentiment	-0.135***	-0.381***	-0.0483***	-0.141***	-0.396***	-1.158***
	(0.0201)	(0.0505)	(0.00738)	(0.0184)	(0.0830)	(0.203)
Lag loan loss provisions/loans	0.215***	0.210***				
	(0.0272)	(0.0271)				
Lag 4-qtr net charge-offs/loans			0.248***	0.240***		
			(0.0280)	(0.0280)		
Lag wdelinq_rate					0.479***	0.474***
					(0.0226)	(0.0226)
CAMELS 2 dummy	-0.0721***	-0.0735***	-0.0374***	-0.0380***	-0.368***	-0.372***
	(0.0232)	(0.0231)	(0.00736)	(0.00736)	(0.0833)	(0.0836)
CAMELS 3 dummy	0.146***	0.141***	0.0529***	0.0512***	0.305**	0.292**
	(0.0404)	(0.0404)	(0.0131)	(0.0132)	(0.129)	(0.130)
CAMELS 4 dummy	0.283***	0.273***	0.171***	0.167***	1.397***	1.371***
	(0.0829)	(0.0825)	(0.0315)	(0.0313)	(0.310)	(0.308)
CAMELS 5 dummy	0.651***	0.636***	0.353***	0.347***	2.258***	2.218***
	(0.198)	(0.197)	(0.0595)	(0.0592)	(0.741)	(0.738)
Ln(total assets)	0.146***	0.144***	0.0455***	0.0449***	0.300*	0.294*
	(0.0464)	(0.0462)	(0.0132)	(0.0130)	(0.154)	(0.153)
Constant	-1.672***	-1.661***	-0.510***	-0.504***	-2.538	-2.469
	(0.580)	(0.577)	(0.165)	(0.163)	(1.923)	(1.909)
Observations	5,398	5,398	5,211	5,211	5,398	5,398
R-squared	0.538	0.541	0.546	0.549	0.707	0.709
Fixed effects	bank & year	bank & year	bank & year	bank & year	bank & year	bank & year
Adj. R-squared	0.426	0.430	0.435	0.439	0.637	0.638

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Management Score Regressions

This table shows the regression results from testing

$$\begin{aligned} \text{outcome}_{i,t} &= \rho \text{ outcome}_{i,t-1} + \beta \text{ sentiment}_{i,t-1} + \gamma \, log(\text{assets}_{i,t}) \\ &\quad + \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n,t-l} + \theta_i + \phi_t + \epsilon_{i,t}, \end{aligned}$$

where outcome is either the summation of all matters requiring [immediate] attention (MRA/MRIA) for a bank or a dummy variable that takes the value of one if any MRAs or MRIAs exist for a bank and zero otherwise. The sentiment score is based on exam text in the management section.

	(1)	(2)	(3)	(4)
	MRA/MRIA Sum		MRA/MRIA Dummy	
VARIABLES	Polar	Valence	Polar	Valence
Lag sentiment	-0.946***	-2.419***	-0.121***	-0.292***
	(0.168)	(0.477)	(0.0136)	(0.0377)
Lag MRA/MRIA Sum	-0.348***	-0.349***		
	(0.0222)	(0.0222)		
Lag MRA/MRIA dummy			-0.526***	-0.529***
			(0.0184)	(0.0183)
CAMELS 2 dummy	1.449***	1.475***	0.126***	0.130***
	(0.209)	(0.209)	(0.0206)	(0.0205)
CAMELS 3 dummy	4.818***	4.801***	0.190***	0.190***
	(0.373)	(0.373)	(0.0278)	(0.0278)
CAMELS 4 dummy	7.423***	7.345***	0.228***	0.222***
	(0.666)	(0.669)	(0.0423)	(0.0422)
CAMELS 5 dummy	10.37***	10.25***	0.299***	0.289***
	(1.371)	(1.372)	(0.0631)	(0.0641)
Ln(total assets)	0.714**	0.687**	0.00762	0.00370
	(0.282)	(0.281)	(0.0296)	(0.0298)
Constant	-5.803	-5.491	0.555	0.599
	(3.526)	(3.514)	(0.370)	(0.372)
Observations	5,321	5,321	5,321	5,321
Fixed effects	0.515	0.515	0.597	0.595
R-squared	bank & year	bank & year	bank & year	bank & year
Adj. R-squared	0.399	0.398	0.500	0.498

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Earnings Score Regressions

This table shows the regression results from testing

$$\begin{aligned} \text{outcome}_{i,t} &= \rho \text{ outcome}_{i,t-1} + \beta \text{ sentiment}_{i,t-1} + \gamma \, log(\text{assets}_{i,t}) \\ &+ \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n,t-l} + \theta_i + \phi_t + \epsilon_{i,t}, \end{aligned}$$

where outcome is either ROA or PPNR/assets. The sentiment score is based on exam text in the earnings section.

	(1)	(2)	(3)	(4)
	Weighted 4-qtr ROA		Weighted 4-qtr PPNR/Assets	
VARIABLES	Polar	Valence	Polar	Valence
Lag sentiment	0.117***	0.332***	0.133***	0.350***
	(0.00969)	(0.0272)	(0.0111)	(0.0291)
Lag weighted 4-qtr ROA	0.357***	0.340***		
	(0.0243)	(0.0247)		
Lag weighted 4-qtr PPNR/assets			0.417***	0.412***
			(0.0290)	(0.0292)
CAMELS 2 dummy	0.0237**	0.0229**	0.00232	0.00114
	(0.0108)	(0.0108)	(0.0141)	(0.0141)
CAMELS 3 dummy	0.00979	0.00954	-0.0224	-0.0229
	(0.0202)	(0.0201)	(0.0230)	(0.0231)
CAMELS 4 dummy	-0.120***	-0.112***	-0.0977***	-0.0891***
	(0.0312)	(0.0312)	(0.0271)	(0.0271)
CAMELS 5 dummy	-0.428***	-0.414***	-0.216***	-0.200***
	(0.0546)	(0.0544)	(0.0401)	(0.0403)
Ln(total assets)	0.0848***	0.0849***	0.247***	0.247***
	(0.0286)	(0.0290)	(0.0384)	(0.0387)
Constant	-0.866**	-0.864**	-2.651***	-2.645***
	(0.355)	(0.359)	(0.470)	(0.474)
Observations	5,401	5,401	5,401	5,401
R-squared	0.708	0.712	0.786	0.787
Fixed effects	bank & year	bank & year	bank & year	bank & year
Adj. R-squared	0.637	0.642	0.734	0.736

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Liquidity Score Regressions

This table shows the regression results from testing

$$\begin{aligned} \text{outcome}_{i,t} &= \rho \text{ outcome}_{i,t-1} + \beta \text{ sentiment}_{i,t-1} + \gamma \, log(\text{assets}_{i,t}) \\ &+ \Sigma_{n=1}^{4} \psi_n \text{ CAMEL dummy}_{i,n,t-l} + \theta_i + \phi_t + \epsilon_{i,t}, \end{aligned}$$

where outcome is either securities/assets or (cash+securities)/assets. The sentiment score is based on exam text in the liquidity section.

	(1)	(2)	(3)	(4)
	Securities/Assets		(Cash+Securities)/Assets	
VARIABLES	Polar	Valence	Polar	Valence
Lag sentiment	0.184	0.577	-0.131	-1.027*
	(0.145)	(0.507)	(0.169)	(0.587)
Lag securities/assets	0.668***	0.668***		
	(0.0177)	(0.0177)		
Lag cash+securities/assets			0.643***	0.644***
			(0.0178)	(0.0178)
CAMELS 2 dummy	-0.181	-0.179	-0.162	-0.171
	(0.179)	(0.179)	(0.223)	(0.223)
CAMELS 3 dummy	-0.976**	-0.973**	-0.544	-0.579
	(0.410)	(0.410)	(0.477)	(0.479)
CAMELS 4 dummy	-0.0499	-0.0415	-1.309	-1.388
	(0.614)	(0.614)	(1.097)	(1.097)
CAMELS 5 dummy	2.677**	2.724**	0.730	0.566
	(1.248)	(1.247)	(1.558)	(1.534)
Ln(total assets)	-0.117	-0.116	-0.564	-0.557
	(0.390)	(0.390)	(0.567)	(0.567)
Constant	8.375*	8.363*	17.50**	17.45**
	(4.938)	(4.944)	(7.286)	(7.286)
Observations	5,401	5,401	5,401	5,401
R-squared	0.930	0.930	0.914	0.914
Fixed effects	bank & year	bank & year	bank & year	bank & year
Adj. R-squared	0.912	0.912	0.893	0.893

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Regressions–Weak/Strong Separately

This table shows the coefficients and standard errors on the lagged sentiment score. Banks are classified as "strong" or "weak" based on rating in the prior exam for that section of the exam. Most banks have a strong score (1 or 2). Weak is defined as 3, 4, or 5.

		1	2	3	4
		Polar		Valence	
		Weak	Strong	Weak	Strong
Composite					
	MRA/MRIA Sum	-3.489***	-3.288***	-5.974***	-7.103***
		(1.147)	(0.371)	(2.184)	(0.940)
	MRA/MRIA Dummy	-0.380***	-0.376***	-0.737***	-0.782***
		(0.0881)	(0.0318)	(0.162)	(0.0753)
Capital					
	Tier 1 Ratio	0.122	-0.0554	-0.301	0.464
		(0.322)	(0.149)	(0.735)	(0.463)
	CET1 Ratio	0.115	-0.0667	-0.313	0.440
		(0.326)	(0.149)	(0.746)	(0.465)
Asset Quality					
	Loan Loss Provisions/Loans	-0.372***	-0.0166	-0.946***	-0.0153
		(0.139)	(0.0482)	(0.317)	(0.102)
	4-qtr Net Charge-offs/Loans	-0.158**	-0.0276**	-0.376***	-0.0494
		(0.0674)	(0.0133)	(0.119)	(0.0307)
	Delinquency Rate	-1.767***	0.0456	-3.254**	-0.179
		(0.610)	(0.162)	(1.363)	(0.392)
Management					
	MRA/MRIA Sum	-0.713	-1.012***	-2.014	-2.988***
		(0.658)	(0.176)	(1.409)	(0.527)
	MRA/MRIA Dummy	-0.0391	-0.111***	-0.117	-0.299***
		(0.0423)	(0.0142)	(0.0852)	(0.0420)
Earnings					
	Weighted 4-qtr ROA	0.216***	0.0110	0.467***	0.0340
		(0.0496)	(0.0154)	(0.106)	(0.0403)
	Weighted 4-qtr PPNR/Assets	0.168***	0.0329	0.325***	0.0845
		(0.0484)	(0.0229)	(0.107)	(0.0638)
Liquidity					
	Securities/Assets	-1.264	0.0547	-2.999	0.841
		(1.187)	(0.360)	(3.482)	(1.331)
	(Cash+Securities)/Assets	0.227	-0.137	1.212	0.229
		(1.590)	(0.425)	(4.075)	(1.423)

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Regressions–Same/Better and Worse Separately

This table shows the coefficients and standard errors on the lagged sentiment score.

		1	2	3	4
		Polar		Valence	
		Worse	**Same/Better**	**Worse**	**Same/Better**
Composite					
	MRA/MRIA Sum	-5.561*	-3.458***	-11.23*	-7.180***
		(3.247)	(0.334)	(6.731)	(0.811)
	MRA/MRIA Dummy	-0.425***	-0.397***	-0.853***	-0.815***
		(0.145)	(0.0322)	(0.299)	(0.0733)
Capital					
	Tier 1 Ratio	-0.176	0.0760	-0.478	0.406
		(0.288)	(0.114)	(0.900)	(0.322)
	CET1 Ratio	-0.144	0.0544	-0.438	0.336
		(0.305)	(0.110)	(0.930)	(0.312)
Asset Quality					
	Loan Loss Provisions/Loans	-0.348*	-0.0408*	-0.662	-0.108**
		(0.203)	(0.0212)	(0.471)	(0.0470)
	4-qtr Net Charge-offs/Loans	-0.0367	-0.0173**	-0.0917	-0.0519***
		(0.0626)	(0.00672)	(0.137)	(0.0159)
	Delinquency Rate	-0.0895	-0.169**	0.349	-0.495***
		(0.589)	(0.0745)	(1.283)	(0.172)
Management					
	MRA/MRIA Sum	-1.828**	-1.028***	-3.384	-2.669***
		(0.864)	(0.167)	(2.474)	(0.481)
	MRA/MRIA Dummy	-0.0972*	-0.122***	-0.196	-0.286***
		(0.0558)	(0.0150)	(0.150)	(0.0424)
Earnings					
	Weighted 4-qtr ROA	0.0375	0.0509***	0.0814	0.159***
		(0.0384)	(0.00785)	(0.0890)	(0.0241)
	Weighted 4-qtr PPNR/Assets	0.0736	0.0596***	0.123	0.158***
		(0.0511)	(0.0119)	(0.119)	(0.0322)
Liquidity					
	Securities/Assets	-0.0115	0.0423	0.0810	0.1000
		(0.729)	(0.169)	(2.889)	(0.612)
	(Cash+Securities)/Assets	0.0787	-0.342*	-0.228	-1.493**
		(1.037)	(0.203)	(4.002)	(0.706)

Robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 11: Regressions–GFC and Post-GFC Separately

This table shows the coefficients and standard errors on the lagged sentiment score. GFC is defined as 2004–2011. Post GFC is defined as 2012–2016:Q2.

		1	**2**	**3**	**4**
		Polar		**Valence**	
		GFC	**Post GFC**	**GFC**	**Post GFC**
Composite					
	MRA/MRIA Sum	-2.862**	-1.886	-6.966**	-2.017
		(1.418)	(1.280)	(2.970)	(2.875)
	MRA/MRIA Dummy	-0.265**	-0.239*	-0.574***	-0.483*
		(0.104)	(0.126)	(0.210)	(0.267)
Capital					
	Tier 1 Ratio	0.266**	0.265*	0.601	1.486**
		(0.123)	(0.143)	(0.414)	(0.652)
	CET1 Ratio	0.244**	0.249*	0.528	1.445**
		(0.120)	(0.141)	(0.407)	(0.646)
Asset Quality					
	Loan Loss Provisions/Loans	-0.199***	-0.0436**	-0.585***	-0.0696
		(0.0333)	(0.0186)	(0.0806)	(0.0492)
	4-qtr Net Charge-offs/Loans	-0.0595***	-0.0202**	-0.189***	-0.0439**
		(0.0120)	(0.00870)	(0.0300)	(0.0192)
	Delinquency Rate	-0.542***	-0.00521	-1.480***	-0.167
		(0.125)	(0.121)	(0.297)	(0.291)
Management					
	MRA/MRIA Sum	-0.671	0.112	-1.864	1.381
		(0.561)	(0.694)	(1.500)	(1.641)
	MRA/MRIA Dummy	-0.00861	-0.0672	-0.00136	-0.0841
		(0.0420)	(0.0594)	(0.110)	(0.144)
Earnings					
	Weighted 4-qtr ROA	0.158***	0.0683***	0.431***	0.224***
		(0.0151)	(0.0131)	(0.0407)	(0.0379)
	Weighted 4-qtr PPNR/Assets	0.168***	0.0779***	0.434***	0.224***
		(0.0167)	(0.0123)	(0.0447)	(0.0330)
Liquidity					
	Securities/Assets	0.295	0.161	1.127	0.188
		(0.231)	(0.176)	(0.785)	(0.614)
	(Cash+Securities)/Assets	-0.417*	0.198	-2.137**	0.178
		(0.251)	(0.235)	(0.871)	(0.802)

Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1