Federal Deposit Insurance Corporation • Center for Financial Research

WORKING PAPER SERIES

Breaking the Word Bank: Measurement and Effects of Bank Level Uncertainty

Paul Soto
Federal Deposit Insurance Corporation

February 2019
Last Updated: December 2019

Previously circulated under the title: "Breaking the Word Bank: Effects of Verbal Uncertainty on Bank Behavior."
Forthcoming in the *Journal of Financial Services Research*

FDIC CFR WP 2019-01

fdic.gov/cfr

Breaking the Word Bank:

Measurement and Effects of Bank Level Uncertainty

Paul E. Soto*

December 12, 2019

Abstract

Banks differ from non-financial firms as banks must communicate to both regulators and shareholders. Also, unlike non-financial firms, banks possess opaque and complex balance sheets and are the main providers of credit to the real economy. In this paper, I propose a new index to detect the *idiosyncratic uncertainty* banks face at the bank-quarter level by applying natural language processing techniques to earnings conference call transcripts. The index reveals which banks at a given quarter signal more uncertainty about their balance sheets. Higher uncertainty is associated with lower lending and higher liquidity the next quarter, suggesting active management of uncertainty. The active management of uncertainty is more pronounced during periods of high aggregate volatility and for banks with more skin in the game. Using loan level data and firm fixed effects, I control for demand-side factors and find that higher bank level uncertainty is associated with lower loan issuances the following quarter.

JEL Codes: G21, G3, E5, D8

Keywords: Uncertainty, Banking, Credit, Natural Language Processing, Machine Learning

*Federal Deposit Insurance Corporation (FDIC), Washington, D.C. E-mail: paul.soto@upf.edu. I am greatly indebted to my PhD advisers, José-Luis Peydró and Andrea Polo, for extraordinary research guidance. I would like to thank Nicholas Bloom, Stephen Hansen, Michael McMahon, Xavier Freixas, Christian Fons-Rosen, Puriya Abbassi, Fillipo Ippolito, Manuel Garcia Santana, Rosalind Bennett, Christopher Kurz, Haelim Anderson, Kristoph Kleiner, Claire Brennecke, Christos Makridis, Juan Imbet, Nick Garvin, Dmitry Khametshin, Angelo Gutiérrez Daza, and participants at the seminars at the Universitat Pompeu Fabra, the University of Maryland, Rensselaer Polytechnic Institute, the Deutsche Bundesbank, and the Modelling with Big Data and Machine Learning conference for useful comments and suggestions. Special thanks to Kathy Zeidler for editorial assistance. Opinions expressed in this paper are those of the author and not necessarily those of the FDIC.

1 Introduction

The 2007 financial crisis introduced a period of prolonged uncertainty into the economy. Market participants were unsure which policies central banks and governments would introduce and how declines in asset prices would affect investment and productivity. They were also uncertain about the financial condition of banks entangled in disrupted markets. Banks play a unique and vital role in the real economy through credit provisioning (Kashyap et al. 2002), yet their fragility also can adversely affect the real economy (Bernanke et al. 1999). Evidence of how banks respond to uncertainty has been elusive given banks' opaque and complex balance sheets (Morgan 2002, Mehran et al. 2011)[1] and their tendency to window-dress assets under supervision (Abbassi et al. 2018).

Further, designing policies in the face of uncertainty remains particularly difficult for policy makers because uncertainty itself is not easy to measure.[2] Current research on uncertainty largely focuses on aggregate measures. However, time-series measures of uncertainty are difficult to disentangle from the business cycle. A more granular bank level measure of uncertainty would allow policy makers to understand which aggregate shocks affect bank level uncertainty and, more importantly, how banks respond to uncertainty while controlling for aggregate trends.

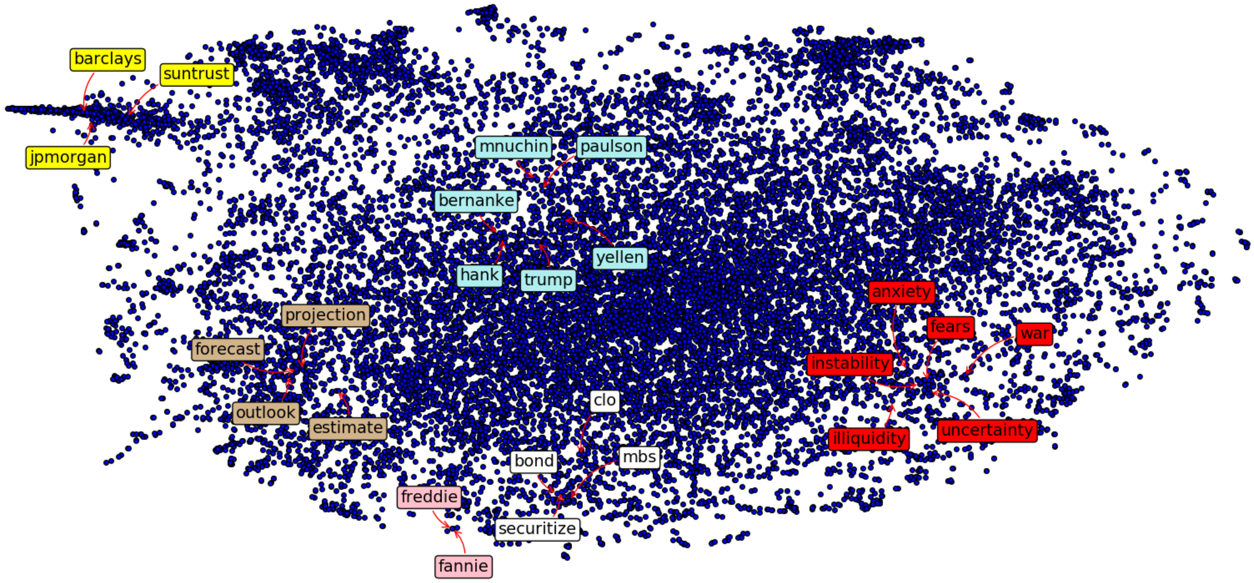
In this paper, I introduce a new method of measuring bank level uncertainty and illustrate how banks manage uncertainty through changes in balance sheet composition. I find that higher idiosyncratic uncertainty is associated with lower lending and higher liquidity the following quarter. The new measure is constructed by applying novel techniques in natural language processing and machine learning to earnings conference calls transcripts from 2002 to 2017. This paper complements extant aggregate uncertainty measures which are typically time-series with a measure at the bank level to allow for cross-sectional heterogeneity.

Creating the new measure involves two steps: (1) creating a list of *uncertainty* words and (2) counting the frequency of the *uncertainty* words in the transcripts. Many economic and finance papers using word counts involve either the researcher choosing a list of words they believe capture

[1] Healthcare and pharmaceutical companies face similar issues through required communication between regulators and shareholders. Healy et al. 2002 and David et al. 2010 discuss the interaction between regulators and pharmaceutical firms.

[2] For example, Yves Mersch, member of the Executive Board of the European Central Bank, claimed "the main problem is the difficulty the policy maker faces in distinguishing between objective and subjective uncertainty" (Mersch 2017)

Figure 1: Word Embeddings of Bank Earnings Conference Call Transcripts

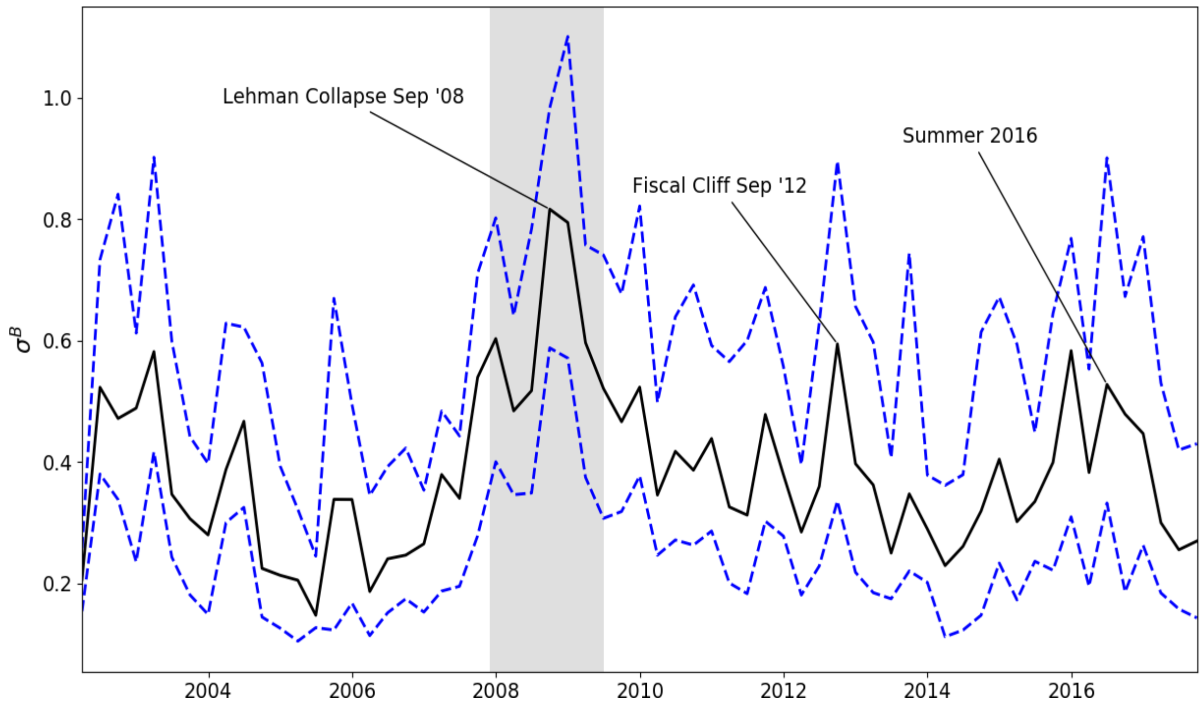


Note: This figure illustrates the vector representations of words, known as word embeddings, from bank earnings conference calls. The word vectors, initially at 300 dimensions, are projected onto two dimensions using the t-Distributed Stochastic Neighbor Embedding algorithm described in the appendix.

the topic they want to measure, or using a list of words identified from another paper. Rather than relying on an existing dictionary that may not be suited for banks, I develop a new word list using a machine learning technique known as word embeddings. Word embeddings are numeric vector representations of words. Every unique word in a set of documents has its own vector representation. The position of each word in the vector space is based on analyzing the contextual information of each unique word across documents, as similar words often appear in similar contexts. This methodology maps similar words of uncertainty into the same region of the vector space. Word embeddings provide researchers an objective framework to determine words of uncertainty based on semantic and syntactic similarities, rather than generating a list of words subjectively or with an existing lexicon from another paper. Figure 1 illustrates the two-dimensional projection of the word embeddings from the bank conference call transcripts in my sample. The rightmost region of the map shows words such as *"anxiety," "war,"* and *"fears"* alongside *"uncertainty,"* suggesting semantically similar meanings for these words within the context of bank speech. I create a new *uncertainty* word list by clustering the words in this region.

I calculate bank level uncertainty as the frequency of these words in each quarterly conference call. Importantly, by constructing a measure at the bank level, I can distinguish and control for macroeconomic variables, including aggregate uncertainty, and identify the idiosyncratic uncertainty each bank faces at a given quarter.

Figure 2: Bank Level Uncertainty



Note: This figure plots the time series median of the bank uncertainty variable, $\sigma^B_{b,t}$, I construct in this paper. $\sigma^B_{b,t}$ is the count of uncertainty terms from management responses of bank quarterly earnings conference calls, calculated at the bank-quarter level. The dashed blue lines represent the 25th and 75th percentiles.

Three results help validate the new measure as a proxy for uncertainty. First, the usage of *uncertainty* words is correlated to aggregate uncertainty measures of the economy. Though the measure is not associated to the term structure of corporate bond yields, the frequency of *uncertainty* words in the conference calls is correlated with the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) and the Economic Policy Uncertainty (EPU) index from Baker et al. 2016.[3] This correlation is reassuring because the VIX pertains to uncertainty about equity returns and the EPU reflects uncertainty about economic policy, both of which directly impact banking activities. Second, the time-series average of bank uncertainty lines up with events hampering the banking sector during the last business cycle. As seen in Figure 2, the median bank uncertainty begins to rise at the onset of the recent financial crisis, in December 2007. Bank level uncertainty remained

[3]Baker et al. 2016 use a text-based approach to measure economic policy uncertainty by counting the number of newspaper articles mentioning the words “uncertainty” and “uncertain” in certain contexts and link their index to unemployment and firm-level investment.

high following the collapse of Lehman Brothers in September 2008, peaking as regulators and banks sought solutions to the ongoing crisis. Toward the end of the crisis, uncertainty plummets. At the same time, cross-sectionally, banks differed largely in their assessment of idiosyncratic uncertainty, as shown by the dashed blue lines representing the 25th and 75th percentiles. Third, bank uncertainty is positively associated with post-call volatility, both realized and implied, but not post-call returns, suggesting that uncertainty through speech relays some form of risk and new information to market participants.

Encouraged by the evidence that the new measure is a proxy for uncertainty, I next assess how banks actively manage their balance sheets following changes in idiosyncratic uncertainty. While extant uncertainty measures are restricted to only the time dimension, the new measure allows me to exploit the cross-sectional heterogeneity of bank uncertainty at a given quarter. In particular, my measure allows me to include time fixed effects to control for interest rates, aggregate uncertainty and firm productivity, the stochastic discount factor, and other time varying characteristics related to credit and liquidity. I find that a one standard deviation increase in a bank's uncertainty is associated with a 46 basis point drop in lending relative to total assets the following quarter. The drop in lending from idiosyncratic uncertainty is not due to time-varying aggregate conditions nor several relevant controls from bank balance sheets. Indeed, while banks cutback on credit after speaking about higher uncertainty, their balance sheets the following quarter report higher liquidity. I also show that neither the widely used proxy for uncertainty, stock volatility, nor controls for individual bank responses to aggregate uncertainty are behind these results.

Next, I allow the coefficient of bank uncertainty to vary depending on bank leverage and the growth in aggregate volatility. Higher equity funding is associated with larger skin in the game, either on the part of shareholders or internally from bank executives and employees. Board members monitor and threaten bad and incompetent executives to incentivize diligent and conscientious actions. I find that declines in credit and increased liquidity from higher idiosyncratic uncertainty are more severe for banks with more skin in the game, suggesting bank executives more actively manage uncertainty under higher shareholder market discipline. Similarly, during times of high aggregate uncertainty, banks that speak using more uncertainty words tend not only to reduce credit but also to increase liquidity positions.

I then decompose the transcripts to identify topic-specific bank uncertainty. Using a technique in textual analysis and machine learning known as topic modeling, I partition the earnings conference call transcripts into two topics particularly relevant to banks: interest rates and housing. I remeasure bank level uncertainty using only the sections of the earnings call related to those topics. I find that uncertainty about interest rates leads to higher exposures of interest rate derivatives, while uncertainty about real estate is significantly associated with lower levels of real estate lending. The results are reassuring not only because they capture topic-specific uncertainty according to the established methodology, but also because they show how specific asset class exposures are partly driven by active management of uncertainty.

Finally, I provide evidence of the contraction in credit from bank level uncertainty at the loan level. A criticism of analyses done at the bank level is that firm-specific demand may drive uncertainty at banks and lower demand for credit by firms, resulting in the observed credit cutbacks. To mitigate these concerns, I use new issuances of loans at the bank-firm-quarter level and include firm-time fixed effects to control for observed and unobserved time varying firm-specific characteristics, especially firm demand. The identifying assumption requires one firm at a particular quarter to receive bank credit from two different banks. I find that banks which relay higher idiosyncratic uncertainty provide smaller-sized loans to firms, consistent with the results at the bank level. The loan level data provide further evidence that banks actively manage their uncertainty by reducing lending even after controlling for credit demand at the firm level.

This paper relates to several strands of literature. First, this paper adds to the growing literature on measuring uncertainty. Most similar to this paper is Manela and Moreira 2017 who use newspaper text to construct a measure of uncertainty starting in 1890. Two important distinctions separate Manela and Moreira 2017 with this paper. First, Manela and Moreira's measure of news implied volatility is an aggregate time series and when used as a sole regressor can not be joined with time fixed effects. This paper builds a measure at the firm-level to understand the cross-section of uncertainty facing banks while controlling for aggregate uncertainty. Second, Manela and Moreira 2017 use supervised learning to obtain their measure whereas this paper uses unsupervised learning. With supervised learning, the authors use responses of the VIX to monthly frequencies of words to identify words that are associated with high volatility. This paper applies unsupervised learning to generate uncertainty words without requiring a response variable or a training dataset. Baker et al. 2016 develop a text-based economic policy uncertainty index by counting the frequency of newspaper articles with the words "uncertainty" and "uncertain" within certain contexts. Berger et al. 2018 show that banks respond to higher economic policy uncertainty, as measured by the time series in Baker et al., by hoarding liquidity and reducing the supply of credit. While the results from Berger et al. 2018 coincide with declines in credit caused by idiosyncratic uncertainty observed in this paper, their paper does not distinguish between aggregate, bank level, and topic-specific exposure to uncertainty, which this paper attempts to do so by constructing a bank level measure of uncertainty. Jurado et al. 2015 develop another measure of uncertainty that suggests uncertainty episodes, such as those found in Baker et al. 2016, appear more infrequently but persistently. Baley and Blanco 2019 develop a firm-level uncertainty index caused by a firm's inability to disentangle temporary and permanent uncertainty shocks. Ludvigson et al. 2019 show that higher macro uncertainty is often an endogeneous response to the business cycle, while financial uncertainty is found to cause declines in real activity. Similarly, Berger et al. 2019 find that forward-looking measures of uncertainty have no real effects while measures derived from realized stock market volatility lead to contractions.

Second, this paper contributes to several studies analyzing the informativeness of earnings conference calls. Mayew 2008 and Bowen et al. 2002 show how new and valuable information is made public through these conference calls. Mayew and Venkatachalam 2012 apply voice recognition software to earnings calls to study the effect of vocal cues and emotions on stock recommendations. The informativeness of these calls affects post-call returns of public companies (Roychowdhury and Sletten 2012, Price et al. 2012). Davis et al. 2015 show that the manner in which individual managers speak affects investors' interpretation of disclosures made in conference calls.

Last, I provide applications of natural language processing and machine learning techniques to economics and finance. Loughran and McDonald 2011 illustrate the nuance of economic and financial texts compared with text from other social sciences. Text analysis has been used to analyze news (Calomiris and Mamaysky 2019, Glasserman and Mamaysky 2019, Manela and Moreira 2017), emerging risks in SEC filings (Hanley and Hoberg 2019), industry momentum (Hoberg and Phillips 2018), central bank transparency (Hansen et al. 2018), regulatory uncertainty (Gissler et al. 2016), and stock market reactions to conference call transcripts (Demers and Vega 2014, Hassan et al. 2019). Hassan et al. 2019 measure political risk in earnings conference calls transcripts and assess the impact on volatility, investment, and political donations. Kozlowski et al. 2019 use word embeddings, a technique applied in this paper, to show how gender and class co-evolved over the last century.

The remainder of this paper is organized as follows. Section 2 provides a summary of the data sources. Section 3 describes the text-based methodology I use to extract *idiosyncratic* bank uncertainty from the conference calls. Section 4 provides predictions to guide the empirical results. Section 5 illustrates how banks actively manage uncertainty. Section 6 concludes the paper.

2 Data

2.1 Text Data from Conference Calls

In 2000, the U.S. Securities and Exchange Commission (SEC) implemented Regulation Fair Disclosure requiring public companies to disclose material information pertaining to business activities. The SEC encouraged public companies to post written transcripts of calls and webcasts on their websites for an appropriate period of time each quarter. The National Investor Relations Institute reports that the percentage of companies holding earnings calls increased from 80 percent in the mid-90s to 97 percent in 2014. The calls last anywhere between 30 minutes to 60 minutes, depending on the market capitalization of the company, current events, and analyst coverage. Earnings conference calls typically begin with the CEO and the CFO providing a summary of the recent quarter and what to expect the following quarter. Next, the line opens for questions from analysts and sometimes shareholders. Conference calls represent a unique opportunity to analyze information executives would like to share with market participants along with information that investors demand, providing a setting where new and valuable information is made public (Mayew 2008).

Earnings call transcripts were downloaded from Thomson Reuters. Since implementation of Regulation Fair Disclosure in 2002, most earnings call transcripts are made public around the date of the call. The calls for banks in my database are typically only available from 2002 onward. I downloaded the available transcripts for the banks in my sample from first quarter 2002 to third quarter 2017.

2.2 Financial Data

For balance sheet data I use the Federal Reserve Y-9C forms, which all banks with more than 1 billion dollars in consolidated assets are required to file with the Federal Reserve each quarter. Data are at the bank holding company level. I retrieve the following balance sheet characteristics: total loans (*BHCK2122*), trading assets (*BHCK1773+BHCK1754+BHCK3545*), interest and non-interest bearing balances (*BHCK0081+BHCK0395+BHCK0397*), loans secured by real estate (*BHCK1410*), interest rate hedging (*BHCK8725*), non-interest income (*BHCK4079*), equity (*BHCP3210*), non-performing loans (*BHCK5525+BHCK5526*), net charge-offs (*BHCK4635 -BHCK4605*), allowance for loan & lease losses (*BHCK3123*), tier-1 capital (*BHCK7205*)[4], and total assets (*BHCK2170*). Tables A1 and A2 in the appendix provide a summary of the variables obtained from the Y-9C forms. All balance sheets variables, except for total assets, net charge-offs, and non-interest income, are calculated relative to total assets. Net charge-offs is measured relative to allowance for loan & lease losses, non-interest income is measured relative to total income, and total assets are transformed to the log of total assets.

The measures used for aggregate uncertainty are the Economic Policy Uncertainty dataset from Baker et al. 2016,[5] the CBOE volatility index VIX, and corporate bond yields from the U.S. Department of the Treasury.

I use Dealscan data for the loan level analysis. The data contains new issuances of large, syndicated, commercial loans in the United States at a quarterly frequency.

2.3 Final Dataset

I combine the FR Y-9C database with the earnings conference calls manually by matching observations using the names of the institutions. I use the *RSSD9017* identifier from the call reports to merge with the name of the bank from the earnings call. In the event of a merger, I keep only the bank holding company and drop the acquired institution from the sample starting from the quarter of the merger. The resulting dataset is joined to the aggregate uncertainty variables. The sample covers the years 2002 through 2017. I restrict the sample to banks with at least 75 percent of the conference calls available during those years. The threshold was chosen so as to maintain a

[4]This value was later changed to BHCA7205 in 2014. Following 2014, I use the latter definition.
[5]This dataset can be accessed at www.policyuncertainty.com.

high cross-section of banks, but results are robust if the sample is restricted to only those banks for which data are available for every quarter.

Table 1: Summary Statistics

	Mean	SD	Min	P25	Median	P75	Max
Dependent Variables							
Loans	62.77	16.95	4.65	59.26	67.61	73.42	96.21
Liquidity	26.92	13.98	1.93	17.67	22.93	31.64	84.33
Independent Variables and Controls							
σ^B	0.46	0.37	0.00	0.22	0.38	0.62	4.01
VIX	19.03	9.83	9.36	13.24	16.21	21.17	80.06
EPU	99.31	45.12	37.27	64.79	85.41	132.82	217.31
Corp. Bond Spread	1.55	0.83	-0.15	1.02	1.58	2.27	2.91
Log(Assets)	17.18	1.88	12.93	15.82	16.81	18.53	21.67
Equity	10.33	2.13	1.68	8.92	10.27	11.6	17.76
Profitability	29.21	15.74	-12.57	19.24	27.13	36.48	84.14
Cash	5.39	6.65	0.20	2.14	3.12	5.33	41.24
Tier 1 Ratio	13.99	2.36	8.52	12.29	13.63	15.32	24.32
Non-Performing Loans	1.19	1.33	0.00	0.37	0.73	1.57	14.46
Charge-Offs	0.34	0.53	0.00	0.06	0.16	0.38	5.18
Credit Risk	21.87	28.61	-12.34	4.64	13.05	28.39	351.56
Dealscan Sample							
Facility Amount ($MM)	1.41	2.34	0.00	0.29	0.75	1.50	36.5
Bank Amount ($MM)	0.09	0.17	0.00	0.02	0.04	0.09	2.86
Log(Assets)	20.61	1.21	16.42	20.42	21.11	21.47	21.65
Equity	8.74	1.85	5.19	7.67	8.37	9.77	13.31

Note: This table reports the summary statistics of the variables used throughout the paper. Balance sheet variables are retrieved from the FR Y-9C forms from the Federal Reserve Board. The Loans variable is total quarterly commercial and industrial, agricultural, consumer, and foreign loans over total assets. Liquidity is the sum of the interest and non-interest bearing balances, available-for-sale and hold-until-maturity portfolios, and trading assets over total assets. σ^B is the count of uncertainty terms of a bank's quarterly earnings conference call. VIX is the CBOE VIX index. EPU is the Economic Policy Uncertainty index from Baker, Bloom and Davis 2016. Corp. Bond Slope is the difference in the ten year less two year corporate bond rate. The Log(Assets) variable is the logarithm of total assets measured in thousands. Equity is total equity capital over total assets. Profitability is the non-interest net income over total net income. Cash is the interest and non-interest bearing balances over total assets. Tier-1 Ratio is the total tier-1 capital over total assets. The Non-Performing Loans variable is total non-performing loans over lagged total assets. Charge-Offs is the total amount of charge-offs over lagged total assets. Credit Risk is charge-offs less recoveries over allowance for loan & lease losses. Facility Amount is the syndicate loan amount from the loan level sample, in billions of dollars. Bank Amount is the share of the facility lent by a particular bank, in billions of dollars.

Table 1 provides the summary statistics. In total, 56 banks were analyzed from first quarter 2002 to third quarter 2017 with 2,745 bank-quarter level observations. Total assets for all commercial banks in the United States averaged \$9.5 trillion, at the industry level per quarter, during this time period. Total assets for the 56 banks in the sample averaged \$8.4 trillion, representing roughly 88 percent of total industry assets. The average asset size of the banks is roughly \$29 billion with an average equity ratio of 10.3 percent. The average lending over total assets is about 62.77 percent, and liquidity accounts for 29.92 percent on average.

For the loal level sample from Dealscan, I use data from 2002-2013. Saturating the regressions with firm-quarter fixed effects reduces the cross-section of banks in the sample to 9 banks and 329 firms. The average size of these banks is larger, nearly \$890 billion in total assets, with an equity ratio of 8.7 percent on average. These banks represent roughly 14% of total banking assets. The average size of the loans is \$1.41 billion and the median is \$750 million. The bank participation share is \$90 million on average.

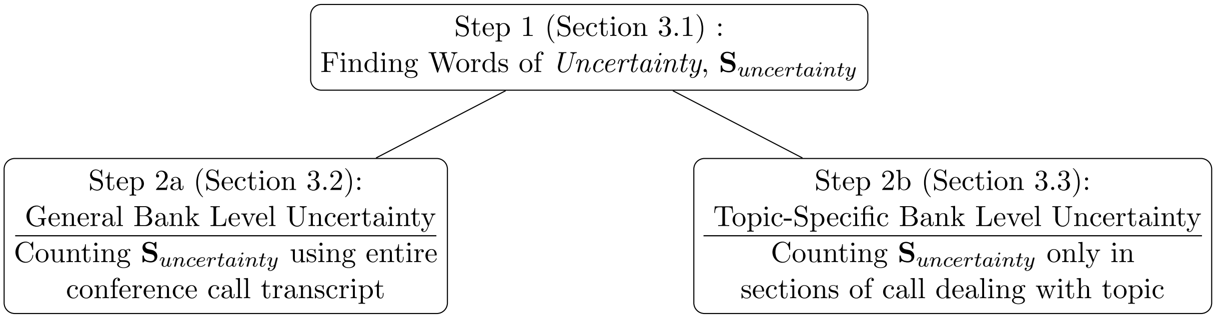
The main dependent variables used in the paper are credit, defined as the total quarterly commercial and industrial, agricultural, consumer, and foreign loans, and liquidity, defined as the sum of the interest and non-interest bearing balances, available-for-sale and hold-until-maturity portfolios, and trading assets. Both variables are scaled by total assets.

While the summary statistics report the nominal values, in all subsequent analyses, I demean the independent variables so the coefficients in regressions can be interpreted as a percentage change in the dependent variable in response to a one standard deviation increase in the independent variable. Unless otherwise noted, standard errors are clustered at the bank and quarter level.

3 Measuring Uncertainty

This section describes how bank level uncertainty is measured using text from the earnings conference calls. The process, illustrated in Figure 3, involves two steps. First, I use natural language processing and machine learning to construct a list of *uncertainty* words, $\mathbf{S}_{uncertainty}$, using the entire corpus of conference call transcripts. Second, using the new list of words, I create a general bank level uncertainty measure counting the frequency of these words in the transcript for each bank at each quarter. I also generate topic-specific measures of uncertainty by isolating parts of the call dealing with particular topics, then counting the frequency of the uncertainty words using only those sections.

Figure 3: Process of Creating Bank Level Uncertainty



3.1 Finding Words of Uncertainty, $\mathbf{S}_{uncertainty}$

Most recent economic and finance research with text analysis involves word counts where the researcher determines the words which they believe capture a specific topic (such as "uncertain" and "uncertainty" for uncertainty) or chooses a pre-made list of words carefully vetted by subject matter experts. For example, Loughran and McDonald 2011 generate a list of uncertainty words by ciphering through 10-K filings of public companies, retaining only words that occur in at least 5 percent of the filings.

To mitigate subjectivity from the researcher choosing words that share similar meaning to a particular concept, or using a pre-made dictionary that may not be tailored for banking text, I rely on recent developments in machine learning and natural language processing to more objectively and automatically identify a new word list.

3.1.1 Word Embeddings

Word embeddings, developed in Mikolov et al. 2013a, are vector representations of words where distances preserve syntactic and semantic similarities between words. The embeddings serve not only as a dimension reduction tool for representing words, but also as a way of preserving the syntactic and semantic relationship in a Euclidean space. For example, word embeddings estimated in Mikolov et al. 2013a famously predicts the relationship: i) vec(Madrid) - vec(Spain) + vec(France) is closest to vec(Paris), and ii) vec(Germany) + vec(capital) is closest to vec(Berlin). In the appendix, I provide a detailed summary of the methodology from Mikolov et al. 2013a in the context of a neural network, the parameters, and the estimation strategy.

Word embeddings can be generated through the Skip Gram model, also introduced in Mikolov et al. 2013a. For example, suppose we have three documents as follows:

Document 1 : we think uncertainty about unemployment

Document 2 : we think fears about unemployment

Document 3 : we think fears and uncertainty about unemployment

The words *uncertainty* and *fears* share several similarities, so an ideal machine learning model would group these words together. The words share similar syntax as they are both preceded at least once by the word *think* and followed by the word *about*. Semantically, *uncertainty* and *fears* both evoke feelings of worry. The Skip Gram model attempts to project both words into the same region in some abstract vector space of dimension H. In the example of the three document examples above, seven unique words lead to seven unique word embeddings:

$$\begin{bmatrix} u^1_{about} & u^1_{uncertainty} & u^1_{fears} & u^1_{we} & u^1_{unemployment} & u^1_{think} & u^1_{and} \\ u^2_{about} & u^2_{uncertainty} & u^2_{fears} & u^2_{we} & u^1_{unemployment} & u^2_{think} & u^2_{and} \\ u^3_{about} & u^3_{uncertainty} & u^3_{fears} & u^3_{we} & u^1_{unemployment} & u^3_{think} & u^3_{and} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{u_{about}} & \mathbf{u_{uncertainty}} & \mathbf{u_{fears}} & \mathbf{u_{we}} & \mathbf{u_{unemployment}} & \mathbf{u_{think}} & \mathbf{u_{and}} \end{bmatrix}$$

In this example, H has been set to 3, but this is a choice that should be determined by the researcher. Ideally, we would want the three-dimensional vector $\mathbf{u_{uncertainty}}$ to be next to $\mathbf{u_{fear}}$ as they share similar syntax and semantics. The Skip Gram model will learn the positioning of each word in the vector space by using the same technique used to learn the meaning of a new word: considering the context.

Mathematically, the Skip Gram model takes every word in every sentence and attempts to predict the words in the context. For example, with the word *uncertainty* in Document 1, the input is the word *uncertainty*, and the outputs will be the words in the context, *think* and *about*:

$$\text{we} \quad \underbrace{\text{think}}_{\text{context (output)}} \quad \underbrace{\text{uncertainty}}_{\text{word (input)}} \quad \underbrace{\text{about}}_{\text{context (output)}} \quad \text{unemployment}$$

The model maximizes the probability of the words *think* and *about* conditional on seeing the word *uncertainty*: $P(think|uncertainty) * P(about|uncertainty)$. The procedure works similarly to a multinomial logistic regression, where:

$$P(think|uncertainty) = \frac{exp(\beta'_{think}\mathbf{u_{uncertainty}})}{\sum_{j'=1}^{V} exp(\beta'_{j}\mathbf{u_{uncertainty}})}; P(about|uncertainty) \quad = \frac{exp(\beta'_{about}\mathbf{u_{uncertainty}})}{\sum_{j'=1}^{V} exp(\beta'_{j}\mathbf{u_{uncertainty}})}$$

with $\beta'_j = \mathbf{u'_j}$. Initially, the word vectors are randomly initialized. The predictions for the probabilities $P(about|uncertainty)$ and $P(think|uncertainty)$ will be incorrect, so the model will adjust all coefficients (each $\mathbf{u_j}$ and $\beta_\mathbf{j}$) through a technique known as backpropogation, such that the probabilities of these two values are closer to 1 and the prediction of all other words moves closer to 0. The process continues using the newly adjusted $\mathbf{u_j}$ word embeddings but with the next word in the sequence as input (*about*) and predicting the context (*uncertainty* and *unemployment*). This is done for every word, in every document, until a satisfactory error threshold is met.

The resulting word embeddings, $[\mathbf{u_{about}}, \mathbf{u_{uncertainty}}, ..., \mathbf{u_{and}}]$ will preserve the structure, such that words which share similar meanings and which are used the same way in a sentence will be close to each other in the vector space.

I use the K-means algorithm to cluster the word embeddings into K disjoint sets to identify the cluster containing "uncertainty" and "uncertain." K-means is an unsupervised learning algorithm that takes as input a set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_V\}$ and a hyperparameter, K, representing the number of clusters. The output is K-disjoint sets, $\{\mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_K\}$, composed of the input vectors.[6]

3.1.2 Estimation

The Skip Gram model for the earnings conference call transcripts was estimated with $H = 300$ and a context window size of 10. The dimension size of 300 is similar to the size used in Mikolov et al. 2013a. The Mikolov et al. paper uses a context window size of 5, but the authors state that a larger window "results in more training examples and thus can lead to a higher accuracy, at the expense of the training time." After inspecting several models with dimension sizes ranging from

[6]Training the algorithm works iteratively in two steps. At the onset of training, a set of vectors, $\{c_1, c_2, ..., c_K\}$ is chosen randomly as cluster centroids. The first step is cluster assignment in which each $\mathbf{u}_i$ is assigned to the cluster, which minimizes the distance between $\mathbf{u}$ and the cluster's centroid. In other words, each $\mathbf{u}_i$ is assigned to cluster k if $k = \arg\min_k \; dist(\mathbf{u}_i, c_k)$. The second step is updating the cluster centroids $\{c_1, c_2, ..., c_K\}$ so that c_k is the average of all points assigned to cluster k. These two steps are repeated until the sum of the squared errors is minimized.

100 to 300 and window sizes ranging from 5 to 10, $H = 300$, and a window size of 10 produced the most coherent representations.

I estimated the model with all available earnings conference call transcripts that contained 18,617 unique words. Figure 1 illustrates the two-dimensional projection of the 300-dimensional word embeddings using a dimension reduction technique known as t-SNE (described in the appendix). Each point represents a unique word in the vocabulary across all conference calls. The Euclidean proximity of two points proxies the similarity of the words both semantically and syntactically.

Words related to names, with "jpmorgan," "barclays," and "suntrust" identified in yellow, appear in the upper left portion of Figure 1. Due to their strong association with each other, "fannie" (from Fannie Mae) and "freddie" (from Freddie Mac) are shown nearly on top of each other in pink near the bottom middle section. The center region displays in turquoise the word embeddings of several political and economic personas, such as "mnuchin," "paulson," "yellen," and "trump." The bottom left portion displays words pertaining to forecasts, such as "forecast," "outlook," "projection," and "estimate." Last, on the right side of the figure, the word embeddings seem to associate the word "uncertainty" with "instability," "fear," "war," and "illiquidity" due to their strong syntactic and semantic similarities.

While the two-dimensional image in Figure 1 is helpful for picturing the 300-dimensional word vectors on a plane, I use the original 300-dimensional vectors to cluster the words into 350 disjoint groups.[7] The cluster containing "uncertainty" and "uncertain," which I refer to as the resulting "uncertainty" dictionary, $\mathbf{S}_{\mathbf{uncertainty}}$, is shown in Table 2. The list contains words associated with uncertainty, such as "fears," "unprecedented," and "instability." Similarly, the methodology picks up events that are associated with rises in uncertainty, such as "brexit," "terrorism," and "war." We also see words that are economic-specific and typically lead to uncertainty in financial markets, such as "illiquidity," "recessionary," and "crises." The fact that words describing downturns, such as "recessionary," "downturn," and "crises," appear in the list could explain the difficulty of measuring uncertainty separate from the business cycle. In the conference call transcripts, these words must appear in the same context as the word "uncertainty" according to their word embeddings, and

[7]The value of 350 was chosen after running a series of cross-validation tests for various cluster values and comparing residual scores.

Table 2: Uncertainty Words from Word Embeddings: $\mathbf{S_{uncertainty}}$

adapted, amid, amidst, amplified, anxiety, attacks, austerity, backdrop, benign, bipartisan, brexit, ceiling, challenges, challenging, cliff, climate, clouded, commonwealth, concerns, conditions, confluence, confronting, congress, consumption, crash, crises, currents, cycles, deficit, deficits, deflation, deflationary, downturn, dysfunction, economic, election, elections, emerged, encountered, environment, environments, eu, euro, eurozone, face, faced, faces, facing, fears, fiscal, flash, fragile, franc, geo, geopolitical, governmental, governments, gridlock, gyrations, hampering, headwinds, heightened, illiquidity, immune, impasse, instability, intervention, iraq, lackluster, legislative, legislature, lingering, looming, ltro, macroeconomic, makers, midst, midterm, monetary, myriad, nafta, navigate, navigated, navigating, paralysis, persist, persisted, persistent, persistently, persists, peso, political, posed, presidential, prevailed, prevailing, prolonged, protracted, psychology, reactions, realities, recessionary, referendum, reforms, rhetoric, rican, ripple, sars, sequester, shutdown, sluggish, society, sparked, spite, stimulative, stimulus, stressful, struggles, surrounding, swiss, tariff, tariffs, tensions, terrorism, terrorist, threat, threats, tsunami, tumultuous, turbulence, turbulent, turmoil, uncertain, uncertainty, uneven, unfolded, unprecedented, unrest, unsettled, unstable, upheaval, war, weathered, weathering, withstand

Note: This table reports the new word list I generate by clustering words close to the word embeddings of "uncertainty" and "uncertain" using the K-means algorithm.

could be perceived interchangeably by banks. Overall, the set appears to accurately resemble words related to uncertainty.

3.2 General Bank Level Uncertainty

The first measure used for bank level uncertainty is the frequency of *uncertainty* words, the cluster $\mathbf{S_{uncertainty}}$ defined in the previous section, in bank conference calls. Formally, bank level uncertainty is defined as:

$$\sigma^B = \frac{\sum_{t \in T} w_t \times \mathbb{1}(t \in \mathbf{S_{uncertainty}})}{|T|} \tag{1}$$

where T is the set of all words in the earnings conference call;[8] $\mathbb{1}(t \in \mathbf{S_{uncertainty}})$ is a dummy variable equal to 1 if word t from from the transcript is in the dictionary $\mathbf{S_{uncertainty}}$; and w_t is a term-frequency inverse document-frequency (TF-IDF) weight, common in text analysis, which provides higher weights to discriminative words in a document.

The uncertainty measure is computed at the bank-quarter frequency, as earnings conference calls

[8] As common is textual analysis, I remove frequent words from the earnings conference calls. These include English pronouns, auxiliary verbs, and articles.

Table 3: Examples of High Uncertainty Responses

Capital One Financial Q3 2008	Like all banks, we **face** increasing **cyclical economic headwinds** and market **uncertainties**. We remain well positioned to **navigate** the near term **challenges** and to realize value-creating opportunities when the time is right. Now Gary and I will answer your questions. Jeff?
SVB Financial Group Q3 2013	All that said, we're also well aware of the **challenges facing** us and the banking industry. The **lackluster** economy continues to create **uncertainty**. **Headwinds** from low interest rates continue to limit the pace of growth.
Northern Trust Q3 2016	In closing, the global **macroeconomic environment** continued to produce a difficult operating **environment** in the third quarter of 2016. Low and even negative interest rates around the globe; post-**Brexit uncertainty**; **election uncertainty**; and debate over central bank actions characterized the quarter. Despite that **challenging backdrop**, Northern Trust produced solid financial results, growing our earnings per share 13% year over year.

Note: This table reports the responses with high σ^B.

occur once per quarter per bank. It is useful, however, to look at examples for individual responses, such as how executive officers answer particular questions during the call. Table 3 shows examples of responses with high measures of uncertainty. For example, Capital One Financial discusses in third quarter 2008 *cyclical economic headwinds and uncertainties* the bank faces. Other highly uncertain responses include *challenging backdrop* and *lackluster* as descriptors of the economy.

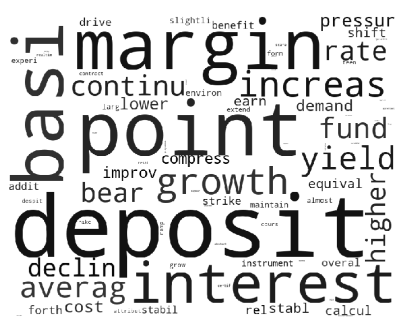
3.3 Topic-Specific Uncertainty

While σ^B represents the idiosyncratic uncertainty of the bank during the entire conference call, I am also interested in topic-specific sections of the call to better understand the uncertainty related to particular topics, such as housing and interest rates. To extract bank responses on specific topics, I implement a Latent Dirichlet Allocation (LDA) model. Introduced by Blei et al. 2003, LDA is an unsupervised algorithm that takes as input a set of documents; a hyperparameter K; and the number of topics. The LDA algorithm outputs two objects: (1) A document-topic matrix revealing the distribution over topics of each document and (2) a term-topic matrix showing a distribution over all words in the corpus for each topic. Similar to principal component analysis, the topics themselves do not have any objective meaning but rather are interpreted by the researcher. LDA models have recently been used in the economic literature to gain insight into speech data, as in Hansen et al. 2018.

Figure 4: Housing Topic (Topics 26 and 38)

Note: This figure displays the topics I use to distinguish responses pertaining to *housing*.

Figure 5: Interest Rates Topic (Topic 40)



Note: This figure displays the topics I use to distinguish responses pertaining to *interest rates*.

I ran several models with 25 to 75 topics and carefully inspected the output of each, comparing their goodness-of-fit through statistical means such as perplexity and topic coherence. I settled on a topic model of 60 topics because of the low perplexity, high coherence measures, and interpretable topics. I isolated the topics related to housing and interest rates. The word clouds for these topics are displayed in Figures 4 and 5. Figure 4 groups together distributions of words related to real estate. A response is labeled as a "real estate" response if it predominantly contains words such as "mortgage," "origin," or "purchase" (Topic 26) or "commercial," "estat," "loan," or "construct" (Topic 38). The interest rate topic identifies responses with "deposit," "margin," "interest," or "basi" (Topic 40). Similarly, I label a response as related to "interest rates" if the response contains words predominantly from the distribution of words shown in Figure 5.

Table 4: Topic-Specific Uncertainty Response Examples

Interest Rates	
Pacific Continental Q4 2012	The current **economic conditions**, the low rate **environment**, and the flat yield curve suggests future margin compression.
Boston Private Q4 2014	As we enter the early portion of 2015, we do believe that a thin yield interest rate scenario is upon us. While the interest rate **environment** is volatile and the duration of this **environment** is hard to predict, **amid concerns** regarding global economic growth and **heightened geopolitical** risk, I do want to address how we're thinking about performance in this **environment**.
Housing	
Sun Bancorp Q3 2010	Clearly this quarter presented tremendous **challenges** as our commercial real estate portfolios saw devaluation as the result of a profound slowdown in the commercial real estate market and an overall constricted **economic environment** in the markets we serve.
F.N.B. Corporation Q4 2012	Slide four shows our 2012 quarterly loan growth trends. Our team has accomplished this consistent growth despite a **challenging economic environment** and historically low line utilization. Additional **headwinds** included significant reductions in the Florida portfolio and acceleration of prepayment speeds in the residential portfolio.

Note: This table reports examples of responses with high σ^B in the interest rate and housing topics.

Using these broader categories of topics, only responses predominantly related to either housing or interest rates are retained. For each of these two broader topics, the frequency of the occurence of words from $\mathbf{S_{uncertainty}}$ is counted, adjusting the denominator T to reflect only sections of the call related to each topic. σ^B for each conference call and each topic is then recalculated using equation (1). Thus, for each conference call, two new variables are generated: idiosyncratic uncertainty about housing and idiosyncratic uncertainty about interest rates.

Table 4 displays the top responses by bank management for the two topics analyzed. Reviewing the top responses for each topic helps validate that the topic modeling is able to objectively, and in an automated manner, select the portions of the earnings call transcripts specific to each topic.

4 Empirical Predictions

To validate σ^B as a measure of uncertainty, the measure should reflect changes in the second moment of a bank's profitability. If this measure was negatively associated to stock returns, then it could be construed as another measure of bad news rather than uncertainty. Thus, if σ^B truly picks up the bank's uncertainty, it should be positively associated with the second moment of the banks returns (i.e. post conference call volatility). Because all banks in my sample are public, I use their stock price returns and volatilities to proxy asset profitability.

Second, increased idiosyncratic uncertainty of a banks balance sheet could signal more dispersion in the distribution of their return on assets, which could be limited by regulators and shareholders. In order to meet constraints, the bank might respond by reducing lending. When profitable credit opportunities erode, banks shift the composition of their balance sheets in favor of credit rationing, securities, or search for yield (Stiglitz and Weiss 1992, Abbassi et al. 2016). The effect could be exacerbated by higher equity capital at risk or during times of higher aggregate volatility. In the appendix, I develop a stylized framework illustrating these hypotheses using a risk-neutral bank that is facing shareholder and regulator constraints.

The empirical predictions of the stylized framework are twofold. First, any measure of idiosyncratic uncertainty, given a level of investment in lending and equity capital, increases the volatility of returns. Second, the optimal level of investment in lending decreases amid a bank's idiosyncratic uncertainty. The decline in lending as a result of higher idiosyncratic uncertainty is more severe for banks with larger equity capital funding and periods in which aggregate uncertainty is high.

5 Results

5.1 Validation

To intepret $\sigma^B_{b,t-1}$, I start by reviewing how the frequency counts of $\mathbf{S_{uncertainty}}$ relate to aggregate uncertainty variables. I use the following regression to gain insight into the interpretation of the

Table 5: Relationship with Aggregate Uncertainty Measures

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	$\sigma^B_{b,t}$		$Loans_{b.t}$		$Liquidity_{b.t}$	
VIX_t	0.0083***	0.0090***	0.1101***	0.1417***	-0.1408***	-0.1561***
	(0.00)	(0.00)	(0.03)	(0.03)	(0.02)	(0.03)
EPU_t	0.0010***	0.0010**	-0.0069	-0.0104*	0.0058	0.0073
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
$CorpBond_t$	-0.0094	-0.0127	-1.9047***	-1.3944***	2.0321***	1.5741***
	(0.02)	(0.02)	(0.13)	(0.33)	(0.32)	(0.38)
Bank FE	N	Y	N	Y	N	Y
N	2,745	2,745	2,745	2,745	2,745	2,745
R^2	0.0881	0.2748	0.0095	0.9169	0.0169	0.8550

Note: This table reports regression results of bank variables on aggregate uncertainty measures. $\sigma^B_{b,t}$ is the count of uncertainty terms of a banks quarterly earnings conference call. Lending is total quarterly commercial and industrial, agricultural, consumer, and foreign loans normalized by total assets. Liquidity is the sum of the interest and non-interest bearing balances, available-for-sale and hold-until-maturity portfolios, and trading assets over total assets. VIX is the CBOE VIX index in the current quarter. EPU is the Economic Policy Uncertainty index from Baker et al. 2016 in the current quarter. $CorpBond$ is the ten-year less two-year corporate yield in the current quarter. Standard errors double clustered at the bank and quarter level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

measure:

$$\sigma^B_{b,t} = \alpha + \beta_1 EPU_t + \beta_2 VIX_t + \beta_3 CorpSpread_t + \gamma_b + \epsilon_{b,t} \tag{2}$$

EPU_t captures uncertainty from policy and is measured by the quarterly Economic Policy Uncertainty (EPU) index from Baker et al. 2016. VIX_t is the quarterly VIX. $CorpSpread_t$ proxies uncertainty about the economy, measured by the ten-year less two-year corporate bond rate. Columns 1 and 2 in Table 5 show the results of the regression. σ^B is statistically and positively related to the VIX as well as the EPU. As the banking sector holds large amounts of trading assets (roughly 20% in my sample), the positive relationship with the VIX, which measures uncertainty of equity returns, is reassuring. Furthermore, as economic policies largely impact banking activities, the positive correlation with the EPU bolsters my confidence that σ^B proxies bank level uncertainty. It is also useful to see how the main variables of interest on the bank balance sheet, lending and liquidity, react to these aggregate uncertainty measures. Lending, shown in columns 3 and 4, is positively and significantly correlated to growth in the VIX, while liquidity, shown in columns 5 and 6, is negatively and significantly correlated. The growth in the EPU index is negatively associated with lending, which suggests that increasing economic policy uncertainty may lead banks to reduce credit.

In Table 6, I separate the sample into banks displaying high and low idiosyncratic bank uncertainty at each quarter using the median value as the threshold. Bank-quarter observations with high idiosyncratic bank uncertainty have higher uncertainty word counts on average (0.70 percent) than low idiosyncratic bank uncertainty (0.23 percent). Unconditionally, banks exhibiting high uncertainty report 3.45 percent fewer loans than banks with low uncertainty. In contrast, liquidity is higher among those with high $\sigma^B_{b,t-1}$, suggesting banks compensate for the reduction in credit with more liquid assets. The bank-quarter observations exhibiting high uncertainty do not correlate unconditionally with the VIX, EPU, or ten-year minus two-year yield.

Table 6: High and Low Bank Uncertainty

	Low Uncertainty (N=1,384)		High Uncertainty (N=1,361)		
	Mean	SD	Mean	SD	Diff.
σ^B	0.23	0.14	0.70	0.39	0.00
Dependent Variables					
Loans	64.48	15.61	61.03	18.05	3.45***
Liquidity	25.5	12.59	28.37	15.13	-2.87***
Independent Variables and Controls					
VIX	19.04	9.98	19.01	9.67	0.02
EPU	99.3	45.15	99.32	45.1	-0.02
CorpSpread	1.55	0.83	1.55	0.83	0.00

Note: This table divides observations in the sample between above and below the median bank uncertainty level. The rightmost column reports the differences in the mean. σ^B is the count of uncertainty terms of a banks quarterly earnings conference call. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The time-series plot of idiosyncratic uncertainty picks up several well-known events hampering the banking sector and bank uncertainty over the last business cycle. Figure 2 plots the median value of $\sigma^B_{b,t-1}$ at each quarter. As the recent financial crisis began to develop in December 2007, banks begin to exhibit larger idiosyncratic uncertainty. Bank level uncertainty remained high following the collapse of Lehman Brothers in September 2008, peaking just before the bailout of the financial system was announced. The dashed blue lines show the 25th and 75th percentiles of the uncertainty of the banks at a given quarter. Cross-sectionally, the boom period between 2004 and 2007 showed moderate disparities in uncertainty, while the beginning of the crisis in December 2007 suggests banks spoke similarly about uncertainty with small differences between the 25th and 75th percentiles. The interval rose subsequently during the peak and end of the recession, remaining wide thereafter.

Table 7: Bank Level Uncertainty and Post-Call Volatility

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:		$Volatility_{b,t:t+3}$			$Returns_{b,t:t+3}$	
$\sigma^B_{b,t}$	0.2462**	0.2293**	0.1744**	0.2706**	0.2132*	-0.1302
	(0.11)	(0.10)	(0.08)	(0.11)	(0.11)	(0.12)
Previous Volatility	0.2592***	0.2579***	0.2578***	-0.2063***	-0.2076***	-0.2069***
	(0.04)	(0.04)	(0.05)	(0.02)	(0.03)	(0.04)
Previous Return	-0.0341***	-0.0340***	-0.0340***	0.0348***	0.0349***	0.0347***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
Bank Controls	N	Y	Y	N	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y
Quarter FE	N	N	Y	N	N	Y
N	2,745	2,745	2,745	2,745	2,745	2,745
R^2	0.5436	0.5515	0.5685	0.1761	0.1856	0.2194

Note: This table shows the effect of bank uncertainty on post-call volatility. Columns 1-3 show results where volatility is the standard deviation of excess returns, as measured through a 3-factor Capital Asset Pricing Model, for $t+1$ to $t+3$ days after the conference call, while columns 4-6 show results for excess returns three days after the conference call. Previous Return (Volatility) is the excess return (volatility) during the 30 days prior to the call. $\sigma^B_{b,t}$ is the count of uncertainty terms of a bank's quarterly earnings conference call. Bank Controls include the following contemporaneous variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors double clustered at the bank and quarter level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, to validate my measure of idiosyncratic uncertainty as a form of risk, I assess how the measure is associated with post-call volatility. First, I analyze the effect of idiosyncratic uncertainty on volatility three days after the call in columns 1-3 of Table 7. Column 1 illustrates how idiosyncratic uncertainty is positively and significantly associated with post-call volatility using only bank fixed-effects. Even when controlling for size, equity, non-performing loans, charge-offs, credit risk, non-interest income, tier-1 ratio, and cash, the positive association between idiosyncratic uncertainty and short-run volatility remains statistically significant, as evident from column 2. By including the most stringent specification with quarter fixed effects, the impact of a one standard deviation increase in bank level uncertainty increases post-call three-day volatility 17 basis points, remaining statistically significant.

Columns 4-6 repeat the exercise with the cumulative absolute excess returns three days after the call as the dependent variable. In the strictest regressions with full fixed effects and controls, short-term returns are not associated with σ^B. Because σ^B is more associated with volatility than prices, which represent discounted future cash flows, the results of Table 7 suggest that *uncertainty* is not a euphemism for future losses and justifies using σ^B as a proxy for bank uncertainty.

5.2 Balance Sheet Responses to Bank Level Uncertainty

Do banks cutback more on lending in lieu of increased liquidity when they emit higher uncertainty to the market? To answer this question, I regress lending and liquidity on $\sigma^B_{b,t-1}$:

$$Y_{b,t} = \beta_1 \sigma^B_{b,t-1} + X_{b,t-1} + \gamma_b + \delta_t + \epsilon_{b,t} \tag{3}$$

Y is either lending and liquidity, and $\sigma^B_{b,t-1}$ is the count of uncertainty terms of a bank's quarterly earnings conference call.[9] I include time fixed effects to control for observed and unobserved time varying characteristics. This allows me to control for all aggregate variables related to credit, such as aggregate uncertainty, the stochastic discount factor, and aggregate firm productivity. The coefficient β_1 reports the change in $Y_{b,t}$ as a response to an increase in bank uncertainty through speech in the earnings conference calls. Time fixed effects allow β_1 to represent the change in $Y_{b,t}$ at a given quarter relative to other banks, as the time dimension is muted. Further, the vector of controls, $X_{b,t-1}$, addresses potential endogeneity stemming from the fact that the relationship between what banks speak about at $t-1$ and $Y_{b,t}$ could be confounded by bank-specific characteristics in the last quarter, such as size, equity, non-performing loans, charge-offs, credit risk, non-interest income, tier-1 ratio, and cash. Last, bank fixed effects, γ_b, control for time-invariant bank characteristics such as corporate governance structures and risk appetites.

Results are reported in columns 1 and 4 of Table 8. With the full set of controls, higher $\sigma^B_{b,t-1}$

[9] The dependent variable is scaled by total assets. In Table A3 I show estimations using the logarithm of the amounts of lending and liquidity and find similar results.

Table 8: Bank Level Uncertainty on Lending and Liquidity

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:		$Loans_{b,t}$			$Liquidity_{b,t}$	
$\sigma^B_{b,t-1}$	-0.4649**	-0.4064**	-0.4419**	0.4930**	0.4299**	0.4021**
	(0.18)	(0.18)	(0.18)	(0.21)	(0.21)	(0.19)
$\sigma^B_{b,t-1}*Equity_{b,t-1}$		-0.3923*			0.4228**	
		(0.20)			(0.19)	
$\sigma^B_{b,t-1}*HighAggVol_{t-1}$			-0.2191**			0.2113**
			(0.10)			(0.10)
Bank Controls	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y	Y	Y
N	2,745	2,745	2,745	2,745	2,745	2,745
R^2	0.9392	0.9395	0.9394	0.9024	0.9029	0.9106

Note: This table reports the effect of bank uncertainty on bank balance sheet variables. Lending is total quarterly commercial and industrial, agricultural, consumer, and foreign loans normalized by total assets. Liquidity is the sum of the interest and non-interest bearing balances, available-for-sale and hold-until-maturity portfolios, and trading assets over total assets. $\sigma^B_{b,t-1}$ is the count of uncertainty terms of a bank's quarterly earnings conference call. $Equity$ is total equity capital over total assets $HighAggVol$ is a dummy variable equal to 1 when the growth in the VIX index is above the median level in the sample and 0 otherwise. Coefficients of $Equity_{b,t-1}$ and $HighAggVol_{t-1}$ are estimated but not reported. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors double clustered at the bank and quarter level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

is associated with lower lending and more liquidity. In a given quarter, a bank with one standard deviation more uncertainty decreases lending next quarter by nearly 46 basis points, and increases liquidity by about 49 basis points, relative to total assets. In Table A3, I replace the dependent variable with the logarithm of the amounts of lending and liquidity. I find that a one standard deviation increase in uncertainty translates to a 1% decrease in the amount of lending and 2% increase in the amount of liquidity.

To better assess the extent to which uncertainty affects next-quarter credit and liquidity, I analyze the impact with two more sources of variation: equity and aggregate uncertainty. I run the following specification in Table 8:

$$Y_{b,t} = \beta_1 \sigma^B_{b,t-1} + \beta_2 \sigma^B_{b,t-1} * D_{b,t-1} + X_{b,t-1} + \gamma_b + \delta_t + \epsilon_{b,t}$$

First I use $D_{b,t-1}$ equal to equity over total assets of bank b at quarter $t-1$. By exploiting the variation of leverage at the time of the conference call, this regression allows me to assess whether banks with more skin in the game, that is higher equity, align their actions more with their words. A bank with higher equity that speaks with more uncertainty has more incentive not to confuse investors, therefore enabling the bank to respond appropriately to higher uncertainty with fewer credit extensions. Columns 2 and 5 of Table 8 suggest that banks with higher equity reduce lending and increase liquidity when they emit a noisier signal through speech.

Next, I use $D_{b,t-1}$ equal to an increase in the VIX as a source of exogenous variation to see if alignments match more when aggregate volatility is high. $HighAggVol_{t-1}$ is a dummy variable equal to 1 when quarter $t-1$ experiences an above-median increase in the growth of the VIX and 0 otherwise. The sign of β_2 is unclear. On one hand, when aggregate volatility is high, banks that are extremely uncertain may themselves be on the verge of insolvency, thus they may reach for yield in the next quarter by increasing lending to riskier borrowers. Thus β_2 could be positive. On the other hand, banks could be frank in their assessment of uncertainty, especially when uncertainty rises in the backdrop, and reduce credit more drastically when speaking with high uncertainty, leading to a negative coefficient.

Columns 3 and 6 show that the effects of $\sigma^B_{b,t-1}$ on bank behavior are stronger during times of high aggregate uncertainty. Banks that speak with more uncertainty during times of high volatility reduce credit more than the average bank. The contrary is true for liquidity as seen by the positive and significant coefficient in column 4.

5.3 Topic-Specific Uncertainty

Using the two topic-specific uncertainty measures, real estate uncertainty and interest rate uncertainty, I next assess the impact on next-quarter real estate loans and interest rate derivatives, respectively. I run the following specification:

$$Y_{b,t} = \beta_1 \sigma^{B}_{b,t-1} + \beta_2 \sigma^{B,Topic}_{b,t-1} + \beta_3 TopicAttention_{b,t-1} + \beta_4 \sigma^{B,Topic}_{b,t-1} * TopicAttention_{b,t-1} + \quad (4)$$
$$X_{b,t-1} + \gamma_b + \delta_t + \epsilon_{b,t}$$

where $\sigma^{B,Topic}_{b,t-1}$ is the TF-IDF weighted count of *uncertainty* words $\mathbf{S_{uncertainty}}$ in sections of the call dealing with a particular topic. Because the effect of uncertainty on a particular asset class might depend on emphasis of the topic during the call, I interact the topic-specific uncertainty with the proportion of the call devoted to the topic, $TopicAttention_{b,t-1}$, as estimated by the topic model.

Table 9 shows how topic-specific bank uncertainty affects particular asset classes. The dependent variable is either real estate loans (columns 1 and 2) or interest rate hedging (columns 3 and 4). The uncertainty in real estate responses is associated with reductions in loans secured by real estate. This effect increases in the attention given to real estate and housing, as seen by the negative and significant coefficient on $\sigma^{B,Topic}_{b,t-1} * TopicAttention_{b,t-1}$. Similarly, higher uncertainty of interest rates increases the exposure of bank interest rate positions, suggesting hedging activities or speculation. This effect is constant regardless of the attention paid to the topic of interest rates as seen by the insignificance of the double interaction. These results suggest topic-specific uncertainty may allow for a better lens to understand bank behavior.

5.4 Evidence at the Loan Level

The results in the previous sections suggest bank level uncertainty reduces credit the following quarter. However, data at the bank level cannot rule out the reverse causality that banks are uncertain precisely because they know firms' demand for credit will be low, resulting in a negative relationship between bank uncertainty and lending.

To mitigate this concern, I use loan level data from Dealscan. The previous analysis using the Federal Reserve Y-9C form incorporated not only commercial lending but also loans to agricultural producers, consumers, and foreign firms. Although the type of lending in Dealscan is only a subset of bank lending, two features of the data help provide evidence of active management of uncertainty through credit cutbacks. First, the Federal Reserve Y-9C reports the stock of all bank lending,

Table 9: Topic-Specific Uncertainty

	(1)	(2)	(3)	(4)
Topic:	Housing		Interest Rates	
Dependent Variable:	$RealEstate_{b,t}$		$IntRateExposure_{b,t}$	
$\sigma^{B}_{b,t-1}$	-0.1804	-0.1918	-1.0064**	-0.7014*
	(0.22)	(0.18)	(0.45)	(0.38)
$\sigma^{B,Topic}_{b,t-1}$	-0.4034**	-0.4532***	0.3582**	0.3673**
	(0.17)	(0.15)	(0.15)	(0.17)
$TopicAttention_{b,t-1}$	1.6009***	1.4763***	-1.2082	-1.1557
	(0.39)	(0.37)	(0.88)	(0.89)
$\sigma^{B,Topic}_{b,t-1} * TopicAttention_{b,t-1}$	-0.3119**	-0.3423***	-0.0797	-0.0264
	(0.10)	(0.12)	(0.29)	(0.29)
Bank Controls	N	Y	N	Y
Bank FE	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y
N	2,745	2,745	2,745	2,745
R^2	0.9314	0.9370	0.6244	0.6328

Note: This table estimates the effect of topic-specific bank uncertainty on bank balance sheet variables. Real Estate is total loans secured by real estate. Interest Exposure is total interest rate exposure over total assets. $\sigma^{B}_{b,t-1}$ is the count of uncertainty terms of a bank's quarterly earnings conference call. Columns 1 and 2 report the coefficient of $\sigma^{B}_{b,t-1}$ only for responses during the earnings call pertaining the real estate cloud of Figure 4; 3 and 4 of the interest rate cloud of Figure 5. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors clustered at the bank and quarter level reported in parentheses.

while Dealscan reports new loan issuances. Second, and perhaps more importantly, loan level data from Dealscan allows for the inclusion firm-time fixed effects. These fixed effects control for observed and unobserved characteristics at the firm level each quarter. The identifying assumption in firm-time fixed effects regression is that each firm at each time period must receive new loans from two different banks. I assume that each firm included in the estimation has a positive demand for credit during the quarter, allowing me to better estimate the effect of bank level uncertainty on the amount of the new loan issuances after controlling for demand-side factors. Summary statistics for the sample are shown in Table 1.

Table 10: Bank Level Uncertainty and Loan Level Data

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:			$Log(LoanShare)_{b,f,t}$		
$\sigma^B_{b,t-1}$	-0.0388**	-0.0438**	-0.0985**	-0.1154**	-0.1036**
	(0.01)	(0.02)	(0.04)	(0.05)	(0.04)
Bank Controls	N	N	N	Y	Y
Loan Controls	N	Y	Y	Y	Y
Bank FE	Y	Y	-	-	-
Quarter FE	Y	Y	Y	Y	-
Firm FE	Y	Y	Y	Y	-
Bank-Year FE	N	N	Y	Y	Y
Firm-Quarter FE	N	N	N	N	Y
N	1,567	1,567	1,567	1,567	1,567
R^2	0.7671	0.7799	0.7882	0.7888	0.7962

Note: This table estimates the effect of bank uncertainty on corporate loan issuances from 2002-2013. The dependent variable is the log of the amount issued by bank b to firm f at quarter t using Dealscan data. $\sigma^B_{b,t-1}$ is the count of uncertainty terms of a bank's quarterly earnings conference call. Loan Controls include dummy variables for the purpose of the loan, whether the syndicate contains multiple lead arrangers, whether the loan is a term loan or revolver loan, and whether bank b is a lead arranger. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors clustered at the bank level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I run the following specification in Table 10:

$$Log(Loan)_{b,f,t} = \beta_1 \sigma^B_{b,t-1} + X_{b,t-1} + \gamma_b + \delta_{f,t} + \epsilon_{b,f,t} \tag{5}$$

where $Log(Loan)_{b,f,t}$ is the log loan amount of b to firm f at time t and $\delta_{f,t}$ is firm-time fixed effects to control for firm demand at time t. Along with bank controls as described above, I saturate the model with an array of loan controls that include dummy variables for: the purpose of the loan, whether the syndicate contains multiple lead arrangers, whether the loan is a term loan or revolver loan, and whether bank b is a lead arranger. Column 1 excludes all bank and loan controls and shows a negative relationship between bank uncertainty and the amount of the loan. Columns 2 and 3 include loan controls and bank-year fixed effects to control for observed and unobserved bank characteristics varying by year. The coefficient on σ^B is negative and significant. In column

4, bank controls reduce the coefficient slightly to -0.1154 while remaining statistically significant. Column 5 shows the strictest regression including firm-time fixed effects. The coefficient increases slightly to -0.1036, suggesting that a one standard deviation increase in bank level uncertainty is associated with a nearly 10 percent drop in the loan amount of new credit issuances. Table 10 shows the importance of firm level fixed effects to account for demand, as column (4), which excludes the granular firm-quarter fixed effects, report a downward biased estimate on the cutbacks in credit associated with larger uncertainty. In fact, the increase in the coefficients between columns 4 and 5 where firm-quarter fixed effects are included suggests σ^B is not orthogonal to observed and unobserved characteristics of the firm, especially firm demand (Altonji et al. 2005). Thus, while firm demand tomorrow plays an important role in determining bank uncertainty today, the negative coefficient suggests that even after controlling for firm demand, σ^B still predicts lending at the loan level.

The Dealscan data are practical for understanding the decline in lending during the financial crisis of 2007 to 2009. Ivashina and Scharfstein 2010 provide evidence that banks with connections to Lehman Brothers, the investment bank that failed on September 15, 2008, was positively related to credit cutbacks. The mechanism proposed is that a bank with larger syndicates with Lehman produced more uncertainty around the financial condition of that bank, as the bank's assets could be tightly linked and similar in quality. In Table 11, I run equation (5) during a smaller window of time to incorporate the financial crisis, 2004 to 2009.[10] I interact the bank-uncertainty measure with $LehmanConnection_b$, the percentage of loans bank b syndicated with Lehman Brothers before the crisis from January 2005 to December 2006, and also a dummy variable for the crisis in years 2008 and 2009. Columns (1)-(2) repeat the previous exercise to illustrate that the smaller sample leads to qualitatively similar results. Column 3 provides the strictest specification with both firm fixed effects, bank fixed effects, and bank controls. The coefficient of $\sigma^B_{b,t-1}$ remains negative and significant similar to Table 10, and yet, the reduction in credit from higher uncertainty during the crisis is larger as seen from the negative and significant coefficient on $\sigma^B_{b,t-1} * Crisis_t$. The reduction in credit from higher bank uncertainty is larger during the crisis for banks with tighter connections to Lehman Brothers, as seen in the significantly negative coefficient on the triple interaction term.

[10] Due to the low number of observations and to preserve the degrees of freedom, this table cannot include the several loan controls used in Table 10.

Table 11: Bank Level Uncertainty and Lending During the Financial Crisis

	(1)	(2)	(3)
Dependent Variable:		$Log(LoanShare)_{b,f,t}$	
$\sigma^B_{b,t-1}$*$LehmanConnection_b$*$Crisis_t$			-0.5971***
			(0.09)
$LehmanConnection_b$*$Crisis_t$			0.4946
			(0.50)
$\sigma^B_{b,t-1}$*$LehmanConnection_b$			1.3273***
			(0.20)
$\sigma^B_{b,t-1}$*$Crisis_t$	-0.7779***	-0.2595**	-0.4099**
	(0.12)	(0.09)	(0.14)
$Crisis_t$	3.0879***	2.4604***	2.6174**
	(0.72)	(0.72)	(1.07)
$\sigma^B_{b,t-1}$	-0.4050**	-0.8585***	-0.3448***
	(0.14)	(0.18)	(0.09)
Bank Controls	Y	Y	Y
Firm FE	Y	Y	Y
Year FE	N	Y	Y
Bank FE	Y	Y	Y
N	288	288	288
R^2	0.8512	0.8545	0.8555

Note: This table reports the effect of bank uncertainty on corporate loans from 2005 to 2009. The dependent variable is new loan issuances of commercial loans from Dealscan at the bank-firm-quarter level. $\sigma^B_{b,t-1}$ is the count of uncertainty terms of a banks quarterly earnings conference call. $LehmanConnection_b$ is the percentage of loans b had syndicated with Lehman Brothers before the financial crisis during 2005-2006. $Crisis_t$ is a dummy variable for the years 2008 and 2009. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors clustered at the bank level are reported in parentheses.

In summary, Dealscan allows for an analysis at the loan level while controlling for firm demand and new issuances. While this exercise includes only a subset of banks and firms, the loan level evidence remains suggestive of active bank management of uncertainty through lower amounts of new loan issuances.

5.5 Robustness

5.5.1 Using Implied Volatility to Measure New Information

Table 7 uses realized volatility to measure the extent to which stock price volatility changes over the three-day window following the usage of uncertainty words in conference call. While realized volatility relies on past prices, *implied* volatility has the benefit of being a forward-looking metric used by options traders to calculate how volatile the stock will be in the future. In Table A4, I run similar regressions using post-call implied volatility to provide further evidence that the content of the earnings call provides new information not previously priced into the banks market value. For the subsample of 46 banks for which implied volatility data was available via Quantcha, I measure the implied volality as the mean of an at-the-money call and put for the bank stock with an expiration of n calendar days from the date of the conference call. Bank level uncertainty is positive but not signficant using implied volatility at a three-day window, as seen in column 2. However, over longer time horizons, a one standard deviation increase in bank level uncertainty increases the implied volatility positively and signficantly 41 basis points for a thirty-day window (column 4).

5.5.2 Volatility and Bank Responses to Volatility Do Not Subsume Text Measure

The relationship between bank level uncertainty from conference calls and next quarter credit and liquidity could be subsumed by controlling for the stock price volatility or individual bank responses to aggregate volatility. In Table A5, I include various controls that could potentially subsume the significance of my text-based measure of uncertainty. In column 1, I include the volatility of the stock price 30 days prior to the conference call as a control for bank level uncertainty and continue to find the negative relationship with credit in Panel A and positive relationship with liquidity in Panel B. As not every bank responds similarly to aggregate uncertainty, in columns 2-4 I include interaction terms between a dummy for each bank and aggregate uncertainty, specifically the VIX in column 2, EPU in column 3, and the corporate bond spread in column 4. In each case, signficance remains for σ^B for both credit and liquidity. In column 5, upon including volatility as well as all three bank and aggregate uncertainty interactions, the relationship persists signficantly and economically. The results of Table A5 suggest bank level uncertainty from conference call can provide added value to simpler-to-construct and existing firm-level uncertainty controls.

5.5.3 Using the Loughran and McDonald 2011 Uncertainty Word List

While the dictionary produced using the word embeddings could lead to ambiguous words possibly unrelated to uncertainty, I recompute the frequency of uncertainty words using the word list developed in Loughran and McDonald 2011. By carefully examining 10-K filings, Loughran and McDonald 2011 identify words of uncertainty that are misclassified or not present in the widely used Harvard word list. Table A6 in the appendix shows results of the regressions of returns, volatility, lending, and liquidity using the newly constructed bank uncertainty variable, $\sigma^{B,LM}$. While applying the widely used uncertainty word list from Loughran and McDonald 2011 simplifies the process of selecting a dictionary, there is no clear indication that this approach captures uncertainty or a signal of future profitability for banks. The negative sign and insignificance of $\sigma^{B,LM}$ in column 1, where the dependent variable is returns, and the large negative coefficient in column 2, where the dependent variable is three-day post call volatility, suggests this measure of uncertainty may proxy bad news as opposed to uncertainty. The lagged variable is insignificant and smaller in magnitude compared to Table 8 with regards to predicting lending and liquidity. In Table A7, I repeat the loan level analysis using the Dealscan database but do not find that using the Loughran and McDonald 2011 word list produces any economic or significant effects on next quarter loan issuances. The insignificance associated with the Loughran and McDonald 2011 uncertainty dictionary compared with using the word embeddings approach bolsters my confidence that choosing a dictionary based on machine learning and natural language processing techniques could provide for a more automated and context-specific approach to dictionary-based textual analysis.

5.5.4 Excluding the Financial Crisis

The results from Table 8 imply heightened uncertainty leads to lower credit and higher liquid assets. Considering that the Great Recession is included in the sample, it could be the case banks with high exposure to subprime mortgages in 2007 speak more about uncertainty in their conference calls and write down their losses on loans the following quarter. In Table A8, I report the main regression results excluding 2008-2010 to mitigate the effect of this time period driving the results. $\sigma^B_{b,t}$ remains positively and significantly associated, at the ten percent level, with respect to three-day post call volatility. The lagged bank uncertainty measure is negatively and significantly associated with next quarter lending and positively and significantly related with liquid assets. It does not appear that the years in which writedowns were heavily occuring are driving the results.

6 Conclusions

In this paper I propose a complimentary measure to existing uncertainty variables by using speech to quantify bank level uncertainty. Using bank earnings call transcripts, I generate a new list of uncertainty terms that is not based on existing dictionaries, but rather generated from the transcripts themselves to better capture semantic and syntactic similarities to the word "uncertainty." The measure of bank uncertainty counts the frequency of uncertainty terms within a given conference call. By developing a measure at the bank level, my analyses include not only bank fixed effects and important lagged variables but also time fixed effects to control for observed and unobserved time-varying characteristics.

Banks communicating larger uncertainty decrease future lending while simultaneously increasing liquidity. By applying topic modeling to identify topic-specific sections of the call, I find that uncertainty about housing leads to lower mortgages and uncertainty about interest rates leads to larger interest rate hedging the following quarter. Because demand-side factors may confound the effect of uncertainty on credit at the bank level, I use loan level data to control for firm demand and find that higher uncertainty is associated with lower loan issuances.

The results suggest that this new measure proxies well the idiosyncratic uncertainty facing a bank and that bank level uncertainty can explain balance sheet compositions. It is my hope that this new measure can be used to track the flow of credit to the real economy by identifying banks most likely to cut back credit during the business cycle and deviate to more liquid assets. The new bank level measure could provide another layer of transparency to an otherwise opaque financial industry.

References

Abbassi, P., Iyer, R., Peydró, J.-L., and Tous, F. R. Securities trading by banks and credit supply: Micro-evidence from the crisis. *Journal of Financial Economics*, 121(3):569–594, 2016.

Abbassi, P., Iyer, R., Peydró, J.-L., and Soto, P. E. Dressing up for the regulators: Evidence from the largest-ever supervisory review. *Mimeo*, 2018.

Altonji, J. G., Elder, T. E., and Taber, C. R. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184, 2005.

Baker, S. R., Bloom, N., and Davis, S. J. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, 2016.

Baley, I. and Blanco, J. A. Firm uncertainty cycles and the propagation of nominal shocks. *American Economic Journal: Macroeconomics*, 11(1):276–337, 2019.

Berger, A., Guedhami, O., Kim, H. H., and Li, X. Economic policy uncertainty and bank liquidity creation. 2018.

Berger, D., Dew-Becker, I., and Giglio, S. Uncertainty shocks as second-moment news shocks. *The Review of Economic Studies*, 2019.

Bernanke, B. S., Gertler, M., and Gilchrist, S. The financial accelerator in a quantitative business cycle framework. *Handbook of Macroeconomics*, 1:1341–1393, 1999.

Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

Bowen, R. R., Davis, A. K., and Matsumoto, D. A. Do conference calls affect analysts' forecasts? *The Accounting Review*, 77(2):285–316, 2002.

Buch, C. M., Buchholz, M., and Tonzer, L. Uncertainty and international banking. In *Conference paper for the conference on International Banking: Microfoundations and Macroeconomic Implications organized by the International Monetary Fund (IMF) and the Dutch National Bank (DNB)*, 2014.

Calomiris, C. and Mamaysky, H. How news and its context drive risk and returns around the world. *Journal of Financial Economics*, 133(2):299–336, 2019.

David, G., Markowitz, S., and Richards-Shubik, S. The effects of pharmaceutical marketing and promotion on adverse drug events and regulation. *American Economic Journal: Economic Policy*, 2(4):1–25, 2010.

Davis, A. K., Matsumoto, D. A., and Zhang, J. L. The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies*, 20(2):639–673, 2015.

Demers, E. A. and Vega, C. The impact of credibility on the pricing of managerial textual content. 2014.

Fahlenbrach, R. and Stulz, R. M. Bank ceo incentives and the credit crisis. *Journal of Financial Economics*, 99(1):11–26, 2011.

Froot, K. A. and Stein, J. C. Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach. *Journal of Financial Economics*, 47(1):55–82, 1998.

Gissler, S., Oldfather, J., and Ruffino, D. Lending on hold: Regulatory uncertainty and bank lending standards. *Journal of Monetary Economics*, 81:89–101, 2016.

Glasserman, P. and Mamaysky, H. Does unusual news forecast market stress? *Journal of Financial and Quantitative Analysis*, Forthcoming, 2019.

Hanley, K. W. and Hoberg, G. Text-based industry momentum. *The Review of Financial Studies*, 2019.

Hansen, S., McMahon, M., and Prat, A. Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870, 2018.

Hassan, T. A., Hollander, S., van Lent, L., and Tahoun, A. Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134(4):2135–2202, 2019.

Healy, P. M., Myers, S. C., and Howe, C. D. R&d accounting and the tradeoff between relevance and objectivity. *Journal of Accounting Research*, 40(3):677–710, 2002.

Hoberg, G. and Phillips, G. M. Text-based industry momentum. *Journal of Financial and Quantitative Analysis*, 53(6):2355–2388, 2018.

Ivashina, V. and Scharfstein, D. Bank lending during the financial crisis of 2008. *Journal of Financial Economics*, 97(3):319–338, 2010.

Jurado, K., Ludvigson, S. C., and Ng, S. Measuring uncertainty. *The American Economic Review*, 105(3):1177–1216, 2015.

Kashyap, A. K., Rajan, R., and Stein, J. C. Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking. *The Journal of Finance*, 57(1):33–73, 2002.

Kozlowski, A., Taddy, M., and Evans, J. The geometry of culture: Analyzing meaning through word embeddings. *American Sociological Review*, 84(5):905–949, 2019.

Loughran, T. and McDonald, B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.

Ludvigson, S. C., Ma, S., and Ng, S. Uncertainty and business cycles: exogenous impulse or endogenous response? Technical report, National Bureau of Economic Research, 2019.

Maaten, L. v. d. and Hinton, G. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

Manela, A. and Moreira, A. News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1):137–162, 2017.

Mayew, W. J. Evidence of management discrimination among analysts during earnings conference calls. *Journal of Accounting Research*, 46(3):627–659, 2008.

Mayew, W. J. and Venkatachalam, M. The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, 67(1):1–43, 2012.

Mehran, H., Morrison, A. D., and Shapiro, J. D. Corporate governance and banks: What have we learned from the financial crisis? *Staff Report. Federal Reserve Bank of New York, NY*, 2011.

Mersch, Y. Economic policy and the need for humility. 9 October, Available online: https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp171009.en.html, 2017.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013a.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119, 2013b.

Morgan, D. P. Rating banks: Risk and uncertainty in an opaque industry. *The American Economic Review*, 92(4):874–888, 2002.

Price, S. M., Doran, J. S., Peterson, D. R., and Bliss, B. A. Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36(4): 992–1011, 2012.

Roychowdhury, S. and Sletten, E. Voluntary disclosure incentives and earnings informativeness. *The Accounting Review*, 87(5):1679–1708, 2012.

Stiglitz, J. E. and Weiss, A. Asymmetric information in credit markets and its implications for macro-economics. *Oxford Economic Papers*, 44(4):694–724, 1992.

7 Appendix

7.1 Appendix Tables and Figures

Table A1: Description of Dependent Variables

Variable Name	Source	Description
σ^B	Reuters	TF-IDF weighted count of uncertainty words in earnings conference call transcripts
Implied Volatility	Quantcha	Mean of at-the-money call and put with n day expiration
Interest Rate Hedging	FR-Y9C	Total gross notional amount of interest rate contracts held for purposes other than trading (BHCK8725) over Total Assets (BHCK2170)
Log(LiquidityAmt)	FR-Y9C	Logarithm of the sum of Noninterest-Bearing Balances, Currency, Coin (BHCK0081)+ Interest-Bearing Balances in U.S. Offices (BHCK0395) + Interest-Bearing Balances in Foreign Offices (BHCK0397) Available-for-Sale Securities (BHCK1773) + Held-to-Matury (BHCK1754) + Trading Assets (BHCK3545)
Log(LoanAmt)	FR-Y9C	Logarithm of Total Loans and Leases (BHCK2122)
Lending	FR-Y9C	Total Loans and Leases (BHCK2122) over Total Assets (BHCK2170)
Lending (Loan Level)	Dealscan	Logarithm of Bank Allocation multiplied by Facility Amount
Liquidity	FR-Y9C	Noninterest-Bearing Balances, Currency, Coin (BHCK0081)+ Interest-Bearing Balances in U.S. Offices (BHCK0395) + Interest-Bearing Balances in Foreign Offices (BHCK0397) Available-for-Sale Securities (BHCK1773) + Held-to-Matury (BHCK1754) + Trading Assets (BHCK3545) over Total Assets (BHCK2170)
Real Estate Loans	FR-Y9C	Loans Secured by Real Estate (BHCK1410) over Total Assets (BHCK2170)
Returns	CRSP	3-Day Cumulative Absolute Returns from 3-Factor CAPM Model
Volatility	CRSP	3-Day Volatility of Excess Returns from 3-Factor CAPM

Table A2: Description of Independent Variables

Variable Name	Source	Description
σ^B	Reuters	Count of uncertainty terms of a bank's quarterly earnings conference call
Cash	FR-Y9C	Noninterest-Bearing Balances, Currency, Coin (BHCK0081)+ Interest-Bearing Balances in U.S. Offices (BHCK0395) + Interest-Bearing Balances in Foreign Offices (BHCK0397) over Total Assets (BHCK2170)
Charge-Offs	FR-Y9C	Charge-Offs (BHCK4635) over Lagged Total Assets (BHCK2170)
Credit Risk	FR-Y9C	Charge-Offs (BHCK4635)- Recoveries (BHCK4605) over Allowance for Loan & Lease Losses (BHCK3123)
Equity	FR-Y9C	Total Equity Capital (BHCP3210) over Total Assets (BHCK2170)
$HighAggVol$	CBOE	Dummy variable equal to 1 if VIX growth above median value between 2002-2017
Non-Perform. Loans	FR-Y9C	Non-Performing Assets (BHCK5525+BHCK5526) over Lagged Total Assets (BHCK2170)
Profitability	FR-Y9C	Non-interest Net Income (BHCK4079) over Total Net Income (BHCK4079+BHCK4107)
Size	FR-Y9C	Logarithm of Total Assets (BHCK2170)
Tier-1 Ratio	FR-Y9C	Tier-1 Capital (BHCK7205) over Total Assets (BHCK2170)
$TopicAttention$	Reuters	Percentage of Transcripts devoted to topic as estimated by a 60-topic topic model
$TopicUncert$	Reuters	σ^B measured with only portion of call devoted to particular topic using 60-topic topic model

Table A3: Bank Level Uncertainty on Amount of Lending and Liquidity

	(1) $Log(LoanAmt)_{b,t}$	(2) $Log(LiquidityAmt)_{b,t}$
$\sigma^B_{b,t-1}$	-0.0100***	0.0205**
	(0.00)	(0.01)
Bank Controls	Y	Y
Bank FE	Y	Y
Quarter FE	Y	Y
N	2,745	2,745
R^2	0.9971	0.9892

Note: This table reports the effect of bank uncertainty on lending and liquidity *amounts*. Log(LoanAmt) is the logarithm of the total quarterly commercial and industrial, agricultural, consumer, and foreign loans. Log(LiquidityAmt) is the logarithm of the available-for-sale and hold-until-maturity portfolios. $\sigma^B_{b,t-1}$ is the count of uncertainty terms of a bank's quarterly earnings conference call. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors double clustered at the bank and quarter level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Bank Level Uncertainty and Post-Call Implied Volatility

	(1)	(2)	(3)	(4)
Dependent Variable:		$ImpliedVolatility_{b,t:t+n}$		
	3-Day	3-Day	30-Day	30-Day
$\sigma^B_{b,t}$	1.1031**	0.2774	1.3316**	0.4109**
	(0.49)	(0.20)	(0.65)	(0.20)
Previous Volatility	0.9075***	0.7432***	0.8556***	0.7117***
	(0.11)	(0.10)	(0.13)	(0.06)
Previous Return	0.0016	-0.0360	0.0220	-0.0336
	(0.04)	(0.03)	(0.06)	(0.02)
Bank Controls	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y
Quarter FE	N	Y	N	Y
N	1,306	1,306	1,306	1,306
R^2	0.8149	0.9178	0.7697	0.9433

Note: This table shows the effect of bank uncertainty on post-call implied volatility. Impled volatility is measured as the mean of an at-the-money call and put for the bank stock with an expiration of n calendar days from the date of the conference call. n is three days for column 1, thirty days for column 2, and ninety days for column 3 after the conference call. Previous Return (Volatility) is the excess return (volatility) during the 30 days prior to the call. $\sigma^B_{b,t}$ is the count of uncertainty terms of a bank's quarterly earnings conference call. Bank Controls include the following contemporaneous variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors double clustered at the bank and quarter level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Added Information from Bank Level Uncertainty

	(1)	(2)	(3)	(4)	(5)
Panel A: $Loans_{b,t}$					
$\sigma^B_{b,t-1}$	-0.4655**	-0.4929***	-0.4868***	-0.3638**	-0.3293**
	(0.18)	(0.17)	(0.18)	(0.17)	(0.14)
Previous Volatility	0.0026	0.0029	0.0017	0.0036	0.0042
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	2,745	2,745	2,745	2,744	2,744
R^2	0.9393	0.9961	0.9961	0.9965	0.9969
Panel B: $Liquidity_{b,t}$					
$\sigma^B_{b,t-1}$	0.4935**	0.5420***	0.5294***	0.4126**	0.3959***
	(0.21)	(0.20)	(0.19)	(0.19)	(0.15)
Previous Volatility	-0.0024	-0.0027	-0.0014	-0.0036	-0.0044
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	2,745	2,745	2,745	2,744	2,744
R^2	0.9024	0.9806	0.9806	0.9820	0.9845
Bank Controls	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y	Y
Bank*VIX Controls	N	Y	N	N	Y
Bank*EPU Controls	N	N	Y	N	Y
Bank*Corp Bond Spread Controls	N	N	N	Y	Y

Note: This table reports the effect of bank uncertainty on bank balance sheet variables while controlling for volatility and heterogeneity in the banks' responses to macroeconomic variables. Previous Volatility is the volatility 30 days prior to the call. Bank*X Controls are interaction terms between a dummy variable for each bank and the macroeconomic control X. The dependent variable $Loan_{b,t}$ in Panel A is total quarterly commercial and industrial, agricultural, consumer, and foreign loans normalized by total assets. Liquidity, in Panel B, is the sum of the interest and non-interest bearing balances, available-for-sale and hold-until-maturity portfolios, and trading assets over total assets. $\sigma^B_{b,t-1}$ is the count of uncertainty terms of a bank's quarterly earnings conference call. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors double clustered at the bank and quarter level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Bank Level Uncertainty with the Uncertainty Word List from Loughran and McDonald 2011

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Returns_{b,t:t+3}$	$Volatility_{b,t:t+3}$		$Loans_{b,t}$			$Liquidity_{b,t}$	
$\sigma^{B,LM}_{b,t-1}$	-0.2343	-0.7492						
	(0.16)	(0.78)						
$\sigma^{B,LM}_{b,t-1}$			-0.3142	-0.3247	-0.3124	0.3272	0.3385*	0.3268
			(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)
$\sigma^{B,LM}_{b,t-1}{*}Equity_{b,t-1}$				-0.2200			0.2370	
				(0.35)			(0.34)	
$\sigma^{B,LM}_{b,t-1}{*}HighAggVol_{t-1}$					-0.0268			0.0037
					(0.06)			(0.07)
N	2,745	2,745	2,745	2,745	2,745	2,745	2,745	2,745
R^2	0.5689	0.2225	0.9391	0.9391	0.9391	0.9021	0.9023	0.9021
Bank Controls	Y	Y	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table reports the effect of bank uncertainty measured using Loughran and McDonald 2011 on bank variables. Column 1 shows results where volatility is the standard deviation of excess returns, as measured through a 3-factor Capital Asset Pricing Model, for $t+1$ to $t+3$ days after the conference call, while column 2 is the excess returns three days after the conference call. Lending, shown in columns 3-5, is total quarterly commercial and industrial, agricultural, consumer, and foreign loans normalized by total assets. Liquidity, shown in columns 6-8, is the sum of the interest and non-interest bearing balances, available-for-sale and hold-until-maturity portfolios, and trading assets over total assets. $\sigma^{B,LM}_{b,t-1}$ is the frequency count of the uncertainty word list from Loughran and McDonald 2011 in a bank's quarterly earnings conference call. *Equity* is total equity capital over total assets $HighAggVol$ is a dummy variable equal to 1 when the growth in the VIX index is above the median level in the sample and 0 otherwise. Coefficients of $Equity_{b,t-1}$ and $HighAggVol_{t-1}$ are estimated but not reported. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors double clustered at the bank and quarter level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Loan Level Results with the Uncertainty Word List from Loughran and McDonald 2011

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:			$Log(LoanShare)_{b,f,t}$		
$\sigma^{B,LM}_{b,t-1}$	-0.0179	-0.0137	0.0427	0.0261	0.0206
	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
Bank Controls	N	N	N	Y	Y
Loan Controls	N	Y	Y	Y	Y
Bank FE	Y	Y	-	-	-
Quarter FE	Y	Y	Y	Y	-
Firm FE	Y	Y	Y	Y	-
Bank-Year FE	N	N	Y	Y	Y
Firm-Quarter FE	N	N	N	N	Y
N	1,567	1,567	1,567	1,567	1,567
R^2	0.7670	0.7797	0.7878	0.7882	0.7958

Note: This table estimates the effect of bank uncertainty measured using Loughran and McDonald 2011 on corporate loan issuances from 2002-2013. The dependent variable is the log of the amount issued by bank b to firm f at quarter t using Dealscan data. $\sigma^{B,LM}_{b,t-1}$ is the frequency count of the uncertainty word list from Loughran and McDonald 2011 in a bank's quarterly earnings conference call. Loan Controls include dummy variables for the purpose of the loan, whether the syndicate contains multiple lead arrangers, whether the loan is a term loan or revolver loan, and whether bank b is a lead arranger. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors clustered at the bank level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Bank Level Uncertainty excluding Financial Crisis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Returns_{b,t:t+3}$	$Volatility_{b,t:t+3}$		$Loans_{b,t}$			$Liquidity_{b,t}$	
$\sigma^B_{b,t}$	-0.0891	0.2107*						
	(0.15)	(0.12)						
$\sigma^B_{b,t-1}$			-0.3969**	-0.3524*	-0.3908**	0.4429**	0.3911*	0.4379**
			(0.19)	(0.18)	(0.18)	(0.20)	(0.20)	(0.20)
$\sigma^B_{b,t-1}$*$Equity_{b,t-1}$				-0.3895*			0.4533**	
				(0.20)			(0.21)	
$\sigma^B_{b,t-1}$*$HighAggVol_{t-1}$					-0.2279*			0.1860
					(0.14)			(0.15)
N	2,103	2,103	2,103	2,103	2,103	2,103	2,103	2,103
R^2	0.2494	0.6399	0.9403	0.9405	0.9404	0.9057	0.9061	0.9058
Bank Controls	Y	Y	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table reports the effect of bank uncertainty on bank variables using a sample from 2002-2017 excluding 2008, 2009, and 2010. Column 1 shows results where volatility is the standard deviation of excess returns, as measured through a 3-factor Capital Asset Pricing Model, for $t+1$ to $t+3$ days after the conference call, while column 2 is the excess returns three days after the conference call. Lending, shown in columns 3-5, is total quarterly commercial and industrial, agricultural, consumer, and foreign loans normalized by total assets. Liquidity, shown in columns 6-8, is the sum of the interest and non-interest bearing balances, available-for-sale and hold-until-maturity portfolios, and trading assets over total assets. $\sigma^B_{b,t}$ is the frequency count of the uncertainty terms in a bank's quarterly earnings conference call. *Equity* is total equity capital over total assets $HighAggVol$ is a dummy variable equal to 1 when the growth in the VIX index is above the median level in the sample and 0 otherwise. Coefficients of $Equity_{b,t-1}$ and $HighAggVol_{t-1}$ are estimated but not reported. Bank Controls include the following lagged variables: log assets, total equity capital over total assets, non-interest net income over total net income, interest and non-interest bearing balances over total asset, tier-1 capital over total assets, credit risk, and non-performing loans and charge-offs over lagged total assets. Standard errors double clustered at the bank and quarter level reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7.2 Why Might Uncertainty Affect Bank Balance Sheet?

In this section, I illustrate reasons a bank would respond to increased idiosyncratic uncertainty by changing their balance sheet composition. I start with a stylized framework, based on Buch et al. 2014, as a useful foundation for establishing testable hypotheses in the following sections. Then, I discuss other possible mechanisms at play.

7.3 Stylized Example

A risk-neutral bank b must decide at time t the proportion of its assets to invest into lending, α_t, with the remainder $1-\alpha_t$ invested into bonds. Bonds return 1 with zero risk, while the return on lending, r_{t+1}^L is random. Macroeconomic forecasts suggest a prior over the return on lending such that $r_{t+1}^L \sim N(\mu, \sigma_\nu^2)$, where σ_ν^2 represents aggregate uncertainty. In addition, the bank guides investors by signaling its own beliefs on the uncertainty of r_{t+1}^L in a conference call, by providing another signal of r_{t+1}^L:

$$r_{t+1}^{L,call} \sim N(\mu, (\sigma^B)^2)$$

The speech during the conference call provides a new variance on the expected return of lending. $(\sigma^B)^2$ represents the bank level uncertainty. Importantly, $(\sigma^B)^2$ is composed of an aggregate component and an idiosyncratic component. Because conference calls contain important insights into the profitability of the firm, I assume $(\sigma^B)^2$ is increasing in the idiosyncratic component (Roychowdhury and Sletten 2012, Mayew 2008, Hassan et al. 2019). I abstract from strategic communication because the empirical analysis below demonstrates that banks act as if they relay truthful information in $(\sigma^B)^2$[11]. Thus, investors use Bayesian updating to form a posterior of r_{t+1}^L:

$$\tilde{r}_{t+1}^L = r_{t+1}^L | r_{t+1}^{L,call} \sim N(\mu, \tilde{\sigma}^2)$$

with the posterior variance $\tilde{\sigma}^2 = \frac{(\sigma^B)^2\sigma_\nu^2}{\sigma_\nu^2+(\sigma^B)^2}$. The expected shareholder capital, k_{t+1}, can be written as:

$$k_{t+1} = \alpha_t(1+\tilde{r}_{t+1}^L) + (1-\alpha_t) - d_t = 1 - d_t + \alpha_t\tilde{r}_{t+1}^L = k_t + \alpha_t\tilde{r}_{t+1}^L$$

[11] In an analysis of non-financial firms, Demers and Vega 2014 find managers use uncertain language to truthfully convey the uncertain prospects of their firm.

Banks seek to maximize shareholder capital tomorrow k_{t+1}. As a result, banks want to choose α as high as possible. However, a Value-at-Risk (VaR) constraint limits the amount the bank can invest in the risky technology, lending. The VaR constraint can be interpreted as limiting the probability that equity capital tomorrow is negative:

$$Pr(\tilde{r}_{t+1}^{L} < -\frac{k_t}{\alpha_t}) = 1 - p$$

Regulators also have information on the distribution of $\tilde{r}_{t+1}^{L}$, so the constraint can be rewritten as a function of the first and second moments[12]:

$$Pr(\tilde{r}_{t+1}^{L} < \mu - \phi \frac{\sigma_\nu \sigma^B}{\sqrt{\sigma_\nu^2 + (\sigma^B)^2}}) = 1 - p$$

I assume regulators are more conservative in their estimates of $\tilde{r}_{t+1}^{L}$, making $-\frac{k_t}{\alpha_t} \leq \mu - \phi \frac{\sigma_\nu \sigma^B}{\sqrt{\sigma_\nu^2 + (\sigma^B)^2}}$[13]. Because the bank wants to set α as high as possible, the inequality will bind. Equating the moments and the ratio of capital to lending leads to the optimal level of lending:

$$\alpha_t = max\left\{0, min\left\{1, \frac{k_t}{\phi \frac{\sigma_\nu \sigma^B}{\sqrt{\sigma_\nu^2 + (\sigma^B)^2}} - \mu}\right\}\right\}$$

7.4 Hypotheses

First, I derive how the posterior variance $\tilde{\sigma}^2$ is affected by increases to uncertainty $(\sigma^B)^2$:

$$\frac{d\tilde{\sigma}^2}{d\sigma^B} = \frac{2\sigma_\nu^4 \sigma^B}{(\sigma_\nu^2 + (\sigma^B)^2)^4} > 0$$

Thus, in order to validate a measure of bank level uncertainty, a positive relationship to post-call volatility would be reassuring as the posterior variance increases in σ^B. Next, I derive how optimal lending α_t changes depending on bank level uncertainty.

$$\frac{d\alpha}{d\sigma^B} = -\frac{\sigma_\nu^3 k_t \phi}{\sqrt{\sigma_\nu^2 + (\sigma^B)^2}(\sigma_\nu \sigma^B \phi - \mu\sqrt{\sigma_\nu^2 + (\sigma^B)^2})^2} < 0$$

[12] I assume the banks assets are marketable. Note that the banks I study in the empirical analysis are all publically traded.

[13] Fahlenbrach and Stulz [2011] find during the financial crisis, banks took large risks predominantly because shareholders wanted to.

As risk increases through a larger variance, banks actively manage their balance sheet compositions in favor of less risky investments in lending. The effects are more severe for banks with higher capital, as seen by differentiating with respect to k_t:

$$\frac{d^2\alpha}{d\sigma^B dk} = -\frac{\sigma_\nu^3\phi}{\sqrt{\sigma_\nu^2 + (\sigma^B)^2}(\sigma_\nu\sigma^B\phi - \mu\sqrt{\sigma_\nu^2 + (\sigma^B)^2})^2} < 0$$

Lastly, active management of bank level uncertainty is influenced by the level of aggregate uncertainty σ_ν.

$$\frac{d^2\alpha}{d\sigma^B d\sigma_\nu} = \frac{\sigma_\nu^2(\sigma^B)^2 k_t\phi(\sigma_\nu\sigma^B\phi - 3\mu\sqrt{\sigma_\nu^2 + (\sigma^B)^2})}{(\sigma_\nu^2 + (\sigma^B)^2)^{3/2}(\mu\sqrt{\sigma_\nu^2 + (\sigma^B)^2} - \sigma_\nu\sigma^B\phi)^3}$$

The cross-partial derivative will be negative as long as $\frac{\sigma_\nu\sigma^B}{\sqrt{\sigma_\nu^2+(\sigma^B)^2}} > 3\mu$. If the language in the conference call relays significant uncertainty (such that the posterior standard deviation is over 3 times larger than the prior mean), then during times of high macroeconomic uncertainty banks will reduce lending more from responses to idiosyncratic uncertainty.

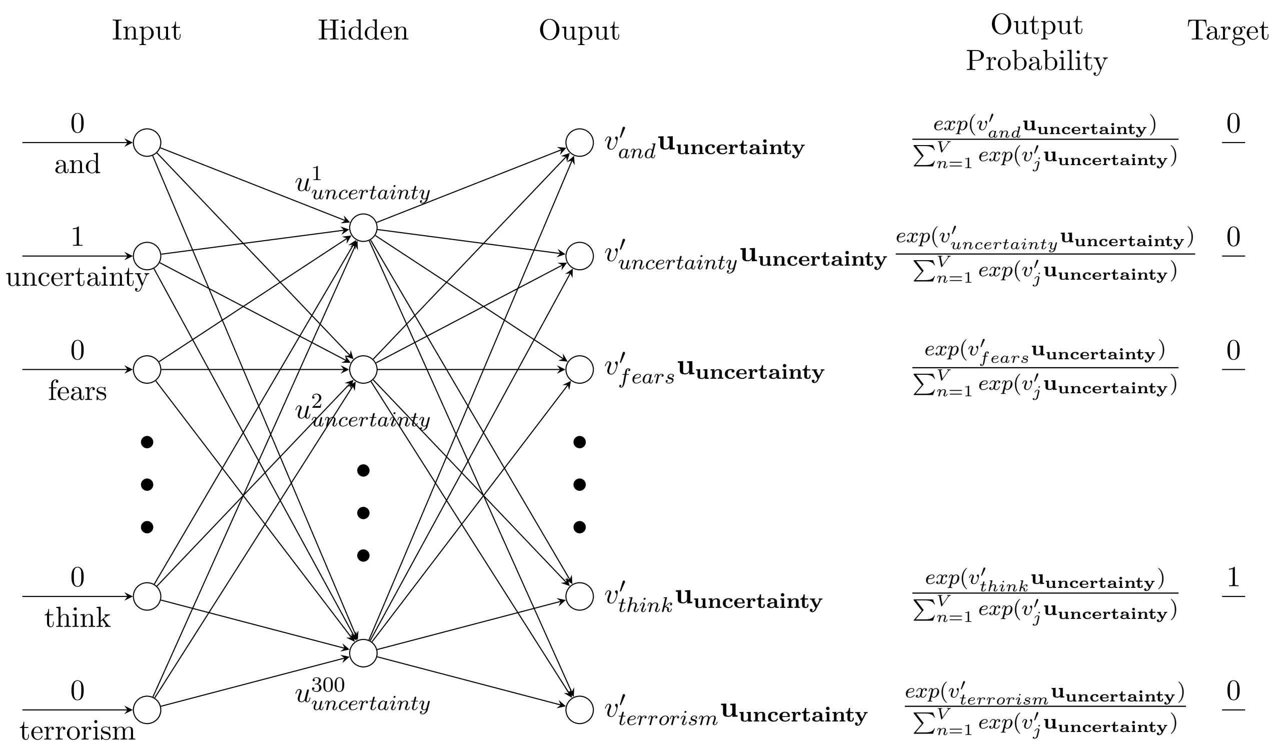
The empirical predictions of the stylized model are as follows. Any measure of idiosyncratic uncertainty, given a level of investment in lending and equity capital, increases the volatility of the returns. Second, the optimal level of investment in lending is decreasing in a bank's idiosyncratic uncertainty. The decline in lending as a result of higher idiosyncratic uncertainty is more severe for banks with larger equity capital funding and periods in which aggregate uncertainty is high.

7.4.1 Alternative Explanation

Risk-neutrality serves as a convenient benchmark for understanding the decisions of the bank as a whole. However, at the individual level, bank managers may exhibit risk aversion, for example through lending to only favorable clients or prioritizing above all else shareholder desire to avoid bankruptcy (Froot and Stein 1998). As a result, another explanation for the reduction of lending in response to higher uncertainty could be that bank managers reduce lending when their signals of r_{t+1}^L are noisy because of risk-averse preferences. This reduction in lending would reduce cash flows for investors and share price, resulting in higher volatility.

7.5 Detailed Explanation of Mikolov et al. 2013a

Figure A1: Skip Gram Model



This figure illustrates the Skip Gram model proposed by Mikolov et al. 2013a using one training example: (*uncertainty, think*). The Input Layer is a one-hot encoded vector of length V, the number of unique words across all documents. A one is placed at the index of word *uncertainty* and a zero elsewhere. Using an $H-$by$-V$ matrix, $\mathbf{U}$, the Input is transformed into an H dimension vector in the Hidden Layer, $\mathbf{u_{uncertainty}}$. Lastly, $\mathbf{u_{uncertainty}}$ is transformed back into a vector of length V using a $V-$by$-H$ matrix $\mathbf{V}$. The output vector is normalized using the softmax function to create y_o. The errors, the difference between y_o and the one-hot encoded vector of the target words in the context, are then used to update the weight matrices $\mathbf{U}$ and $\mathbf{V}$ using gradient descent.

The Skip Gram model takes as input a word represented as a one-hot encoded vector[14]. The output is the words in the context of w_i. The context is the M words to the left and M words to the right of w_i. Let w_{i+m} represent the word which lies $|m|$ words to the left (right) of word w_i for $m < 0$ ($m > 0$) in the original document.

The transformation of the input to the output happens in two steps. Let x_{w_i} be a one-hot representation of word w_i and V the total number of unique words across all documents. First, x_{w_i} is projected onto an H-Dimensional space with $\mathbf{U}$, an H-by-V matrix, to create u_{w_i}. Second, u_{w_i} is projected back onto a V dimensional space using a V-by-H matrix $\mathbf{V}$. To obtain a probability

[14] A one-hot encoded vector for word w_i is a vector of length V- the number of unique words across all documents. A value of 1 is placed at the index of w_i and 0 elsewhere.

distribution over all V words in the vocabulary, the softmax function[15] is applied to the resulting vector. The neural network, illustrated in the appendix in Figure A1 with a simplified training example, can be summarized by the following equations:

$$\begin{aligned}
\text{Input} \quad & x_{w_i} \\
\text{Hidden Layer (Word Embedding)} \quad & u_{w_i} = \mathbf{U}x_{w_i} \\
\text{Output} \quad & x_o = \mathbf{V}u_{w_i} = \begin{bmatrix} v_1'u_{w_i} & v_2'u_{w_i} & \dots & v_V'u_{w_i} \end{bmatrix}' \\
\text{Output Probabilities} \quad & y_o = softmax(x_o)
\end{aligned}$$

y_o will be a distribution over all V terms and will be the same output for all $2 * M$ words in the context of w_i.

The important object of the Skip Gram model is the matrix $\mathbf{U}$. $\mathbf{U}$ provides the word embeddings where each word will be represented as a vector in $\mathbb{R}^H$. Because $\mathbf{U}$ is an $H-$by$-V$ matrix, the word embedding of word w_i will be the column in $\mathbf{U}$ pertaining to w_i. $\mathbf{V}$ can also be interpreted as another space of word embeddings, where each row i represents the word embedding of word w_i with different valued elements. However, the semantic and syntactic differences across words are still preserved, hence typically only $\mathbf{U}$ is used for the word embeddings.

The model is trained by finding $\mathbf{U}$ and $\mathbf{V}$ such that the average probability of context words is maximized. The objective function to maximize is $\frac{1}{Z}\sum\limits_{i=1}^{Z}\sum\limits_{\substack{-M\leq m\leq M\\ m\neq 0}} log(p(w_{i+m}|w_i))$, where Z is the total number of words across all documents and $p(w_{i+m}|w_i)$ is the element in y^o pertaining to word w_{i+m}. Taking just one word w_i, we can change the objective function to a loss function to minimize:

$$\begin{aligned}
E &= -\sum_{\substack{-M\leq m\leq M\\ m\neq 0}} log(p(w_{i+m}|w_i)) = -\sum_{\substack{-M\leq m\leq M\\ m\neq 0}} log\frac{exp(v'_{w_{i+m}}u_{w_i})}{\sum\limits_{j'=1}^{V} exp(v'_j u_{w_i})} \\
&= -\sum_{\substack{-M\leq m\leq M\\ m\neq 0}} v'_{w_{i+m}}u_{w_i} + log(\sum_{j'=1}^{V} exp(v'_j u_{w_i}))
\end{aligned}$$

[15]The softmax function maps a vector of K elements to a range of [0,1] with elements summing to 1. If z_i is element i of vector $\mathbf{z}$, the softmax funtion is given by $softmax(\mathbf{z})_i = \frac{exp(z_i)}{\sum\limits_{k=1}^{K} exp(z_k)}$.

where $v_{w_{i+m}}$ is the row of $\mathbf{V}$ pertaining to w_{i+m}.

Training the model requires obtaining optimal estimates of $\mathbf{U}$ and $\mathbf{V}$ through gradient descent. Basically, given some values of $\mathbf{U}$ and $\mathbf{V}$, we observe how far off the current estimates of the probabilities for the words in the context of w_i were. Then, depending on the error we move around the elements of $\mathbf{U}$ and $\mathbf{V}$ such that the error improves.

More formally, first, every word in every sentence is fitted with the model to obtain the prediction y_o. Second, the errors are calculated using the gradient of E with respect to $\mathbf{U}$ and $\mathbf{V}$ to update the two matrices:

$$\mathbf{V}^{new} = \mathbf{V}^{old} - \alpha \frac{\partial E}{\partial \mathbf{V}^{old}}$$
$$\mathbf{U}^{new} = \mathbf{U}^{old} - \alpha \frac{\partial E}{\partial \mathbf{U}^{old}}$$

While normal gradient descent could be applied to discover optimal parameters, Mikolov et al. 2013b discuss significant improvements to training with respect to sampling for parameter updates. Mikolov et al. 2013b address the computational complexity of using basic gradient descent by describing two techniques known as *negative sampling* and *subsampling of frequent words*.

For each training example, the output will be a sparse vector with 1 only at the index of the context word and 0 in the tens of thousands of indices which are not in the context. *Negative sampling* means only updating the weights of a sample of the words (columns) in $\mathbf{V}$ which should output a 0, or negative sample.[16] The positive word, the target word in the context whose output should be 1, is also updated in $\mathbf{V}$. The sampling distribution for the negative words is given by:

$$P(w_i) = \frac{f(w_i)^{\frac{3}{4}}}{\sum_{v=0}^{V} f(w_v)^{\frac{3}{4}}}$$

where $f(w_i)$ is the frequency of word w_i.[17] Mikolov et al. suggest choosing 5-20 negative samples so only 0.04 percent of the millions of weights need to be updated for each training example.

Another improvement to the computational speed of estimating the neural network is *subsampling*

[16]Note the update in $\mathbf{U}$ will only be the word embedding of the input word w_i. This is evident when looking at the update function of $\mathbf{U}^{new}$: $\frac{\partial E}{\partial \mathbf{U}^{old}} = \mathbf{V}^T[x^o \circ \sigma']x^h x^i T$. The input vector x^{iT} is a sparse vector with a 1 only at the index of word w_i.

[17]The authors mention trying different functional forms of the sampling distribution and find this one to perform the best.

of frequent words. The Skip Gram model can learn much from the co-occurrence of words such as *terrorism* and *uncertainty*, as these are relatively infrequent. However, words such as *the* are relatively uninformative as a multitude of words can precede or follow them. Thus, each input word in the training set are kept with probability $P(w_i) = \sqrt{\frac{t}{f(w_i)}}$, where t is a chosen threshold, typically 0.001.[18]

[18]This functional form was chosen by Mikolov et al. because it attributes higher probabilities to more frequent words while preserving the ranking of the frequencies. The C-code for estimation provided by the authors, however, uses $P(w_i) = (\sqrt{\frac{f(w_i)}{t}} + 1)\frac{t}{f(w_i)}$, which is a slightly more convex form of the initial suggestion.

7.6 t-Distributed Stochastic Neighbor Embedding (Maaten and Hinton 2008)

The goal of t-SNE is to map a set of V vectors, $X = \mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_v}$, from $\mathbb{R}^M$ into another set of vectors, $Y = \mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_v}$, from $\mathbb{R}^N$ such that $N < M$. First, distances are measured between each vector in X as probabilities. This is done by centering a Gaussian distribution at each vector, $\mathbf{x_i}$ to compute a probability for every other vector in X. That is, the probability/distance measure is calculated as follows:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2V} \qquad \text{where} \qquad p_{j|i} = \frac{exp(-||x_i - x_j||^2/2\sigma_i^2)}{\sum\limits_{j' \neq i} exp(-||x_i - x_{j'}||^2/2\sigma_i^2)}$$

Each p_{ij} provides a measure of probability between points which is proportional to their similarity. If two points are close together in $\mathbb{R}^M$, then the probability is high. t-SNE seeks to find the set Y whose distances are similar to all p_{ij}s. The distances in $\mathbb{R}^N$ are calculated similarly, except the t-distribution density is used.

$$q_{ij} = \frac{q_{j|i} + q_{i|j}}{2V} \qquad \text{where} \qquad q_{j|i} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum\limits_{k} \sum\limits_{l' \neq k} (1 + ||y_l - y_k||^2)^{-1}}$$

Ideally, the p_{ij}s and q_{ij}s will be similar to each other. If so, then distances in the high-dimensional space would be preserved in the low-dimensional space. As a result, the objective function to minimize is the divergence between the p and q distributions, which is commonly computed as the Kullback-Leibler divergence:

$$KL(P||Q) = \sum i \sum j \neq i p_{ij} log \frac{p_{ij}}{q_{ij}}$$

The Kullback-Leibler divergence is well suited for this task as large penalties are produced when p_{ij} is large and q_{ij} is small. Thus, it will aim to associate large ps with large qs, hence the difference between PCA and t-SNE. t-SNE will seek to preserve vectors close together (high ps), while PCA preserves vectors further apart.

t-SNE is then estimated via gradient descent. The points in Y will move around until the Kullback-Leibler is sufficiently small. Maaten and Hinton 2008 provide more details about the estimation and algorithm for optimal performance.