Credit Market Expectations and the Business Cycle: Evidence from a Textual Analysis Approach *

Mardoqueo Arteaga†

December 13, 2022

Abstract

I study the interplay between errors in credit spread expectations and macroeconomic indicators between 1948 and 2022. I approach this by deriving a proxy for credit market sentiment using textual analysis on Wall Street Journal title pages that aid me in filling in historical credit spread expectations from 1919 onward, using the Survey of Professional Forecasters as my training data. After validating my textual analysis approach, I find that one-standard deviation jumps in credit spread forecast errors are associated with declines in economic activity, at most with a 3% decline in GDP growth during the sample period. This finding supports the behavioral models highlighting the importance of sentiment in the credit market as a key driver in cycles in the macroeconomy.

*Draft of Chapter 2. This version: December 2022. I thank Arunima Sinha, Johanna Francis, Ralf Hepp, and other seminar participants at Fordham University. Special thanks to to Roger Ferguson, Francesco Montagnani, and Luis Armona who provided comments and feedback. This paper makes extensive use of the Survey of Professional Forecasters, © 2010-2022 Federal Reserve Bank of Philadelphia. The data are available without charge at https://www.philadelphiafed.org/surveys-and-data/data-files; FRBNY disclaims any responsibility for this analysis and interpretation of the data herein. Additionally, it makes usage of Google's word2vec model to compute vector representation of words; Google makes the repository available at https://code.google.com/archive/p/word2vec/. The usual caveat on error applies.

†Economics Department, Fordham University, Bronx, NY 10458. Email: arteaga@fordham.edu

1 Introduction

. . . [My] results, which come from a sample spanning the period from January 1973 to December 2012, are striking. Upward moves in excess bond premium – again, those corresponding to a widening of credit spreads–are very informative about the future evolution of the real economy . . . I have to believe that our macro models will ultimately be more useful as a guide to policy if they build on a more empirically realistic foundation with respect to the behavior of interest rates and credit spreads.

Governor Jeremy C. Stein, March 2014 (Stein, 2014)

The Great Recession led to renewed interest in the relationship between credit expansion and the macroeconomy, with various parties attempting to confirm or refute competing narratives about the causes and propagation of financial crises. Economists noted that developed countries often experience alternating periods of growth and decline in real and financial activity. These so called "boom-bust cycles" are characterized by high levels of investment, output, and leverage, as well as low credit spreads; conversely, what often follows is a reversal in which credit spreads rise and investment, output growth, and leverage decline (Schularick and Taylor, 2012; López-Salido et al., 2017). In addition, credit spreads are widely used by economists, policymakers, and market practitioners as a measure of financial strain, and changes in credit spreads are often seen as a leading indicator of future economic activity, which is conveyed as much in the introductory quote.

Traditionally, theories on the causes of financial instability have often centered on the amplification of shocks that can sometimes be traced back to underlying fundamental factors, but may also arise from financial shocks like a spike in required returns or increased uncertainty (Bernanke and Gertler, 1989; Bianchi, 2011; Eggertsson and Krugman, 2012; Arellano et al., 2019). Specifically, in models with financial market frictions, changes in credit spreads can reflect shifts in the effective supply of funds, which can then affect future economic outcomes. A disruption in the financial market, for instance, could lead to a reduction in the supply of credit, causing credit spreads to widen and leading to a subsequent reduction in spending and production.

Dissent abounds. From both theorists and empiricists, a competing explanation in understanding financial instability emphasizes the role of non-rational beliefs, such as excessive optimism during good times, leading to overexpansion of credit and investment (Minsky, 1977; Gennaioli et al., 2016; Bordalo et al., 2018, 2019, 2020a,b). When beliefs subsequently cool off, credit markets tighten and real activity declines, leading to increased default rates. Credit spreads play a critical role in shaping investor expectations about future credit defaults, and changes in credit spreads can reflect shifts in investor sentiment (Jordá et al., 2013; Greenwood and Hanson, 2013; Baron and Xiong, 2017). In these models, excessively narrow credit spreads can lead to expansions of credit and increased real economic activity, but these patterns will typically reverse when future economic outcomes disappoint investors. This recent research has found that the measured expectations of a broad range of economic agents systematically deviate from rationality, tending to be overly optimistic during good times and then reverting. The key difference between these two types of theories is whether agents know the objective probability distribution in equilibrium. Rational agents should understand the cyclical nature of the credit market and take this into account when forming their expectations. This is why it is crucial to study expectations data directly.

In this paper, I provide empirical evidence on how the errors between realized credit spreads and their forecasts can predict future macroeconomic outcomes. To do so, I first predict a series of past credit spread expectations from 1919Q1 to 2022Q3 by applying textual analysis through natural language processing and topic models in statistical machine learning on front pages of the Wall Street Journal. Then, following the methodology for textual factors in the natural language processing space, I derive the credit spread expectation error which I use as the independent variable in a series of quarterly predictive regressions for GDP, unemployment, and private domestic investment from 1948Q1 until 2022Q3.

I focus on the BAA credit spread which is the difference in yield between the risk-free 10-Year Treasury Bond Yield and the BAA Corporate Bond Yield. Credit spread dynamics refer to the changes in the size of the credit spread over time and it can provide useful information about investors' perceptions of risk and the overall health of the economy because the size of the credit spread is directly related to the level of risk that investors are willing to take on. Wide spreads indicate that investors are demanding a higher return to compensate for the increased risk (less willing to take on risk), while narrow spreads indicate that investors are willing to accept a lower return for the potential reward (more willing to take on risk). The former can be a sign of economic uncertainty or instability as investors are more cautious about the potential risks and rewards of different investments while the latter can be a sign of economic growth and stability, as investors are more confident in the potential returns of different investments.

Credit spread expectations refer to the anticipated changes in the size of the credit spread over a time. These expectations can be based on a variety of factors, such as changes in the level of risk in the economy, changes in interest rates, or changes in investor sentiment. For example, if investors expect the level of risk in the economy to increase, they may anticipate that the credit spread for risky bonds will widen, and they may adjust their investment strategies accordingly. Similarly, if investors expect interest rates to rise, they may anticipate that the credit spread for risky bonds will narrow, and they may adjust their investment strategies accordingly. In theory, analyzing credit spread expectations allow investors and policymakers to gain insight into the market's expectations for changes in the level of risk and the overall health of the economy. Unfortunately, this data is largely collected from surveys and has had limited collection in the past. For instance, the widely used Blue Chip Financial Forecasts only started soliciting forecasts of credit spreads in 1999, covering just two recessions.

By deriving my own proxy for forecasts of credit spreads through a textual analysis approach, I can analyze their empirical relationship with a variety of macroeconomic indicators and the business cycle. I find that overly optimistic sentiment in credit spreads, associated with higher expectation errors, predict a decline in economic activity across three macroeconomic indicators up to four quarters ahead. Specifically, a one-standard deviation jump in the error for credit spread expectations is associated with, at most, a predicted 3% decline in GDP growth, a 1.68% increase in the unemployment rate, and a 2.9% decline in private domestic investment growth. These results are robust to a variety of controls that are common in the literature and include lagged values of the respective indicators, the most recent values for the BAA credit spread and CPI inflation rate, and the changes associated with the 3-month and 10-Year Treasury Yields. The findings empirically corroborate the story in the behavioral models which posit that elevated sentiment predates declines in economic activity while also suggesting that the textual analysis approach in creating historical credit spread expectations and their errors through time has value and provides a proxy for sentiment.

Outline. In the next section, I present the context from the literature most closely aligns to this investigation. In Section 3, I present the data acquired from various sources as well as the method by which I construct my own expectations data using statistical machine learning. Section 4 presents the empirical specification that set out to predict macroeconomic outcomes using expectations errors. I present the baseline results in Section 5, and finally Section 6 concludes.

2 Context In Literature

This paper contributes to two active research areas.The first is regarding behavioral approaches to predicting business cycles using credit spreads. Whether it be from a theoretical perspective as in Kubin et al. (2019) or by creating a factor that summarizes credit sentiment as in Leiva-León et al. (2022), this literature is most poignantly summarized in Bordalo et al. (2018). There, credits spreads play a critical role as expectations about future credit defaults are over influenced by current news. Bordalo et al. use the Blue Chip Financial Forecast data to document predictability in credit spread expectation errors and revisions between 1999Q1 and 2014Q4. As a follow up to their analysis, and as a motivational exercise for my textual analysis approach, I perform the same test on data from the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters (SPF) from 2010Q1 to 2022Q3. Under the assumption of rational expectations (and knowledge of the data generating process), the forecast errors (actual credit spread minus expectation of credit spread) from the professional forecasters should be unpredictable from past data but Figure 1 suggests otherwise. Similar to Bordalo et al.'s findings using Blue Chip Financial Forecasts, the SPF corroborates the narrative that when the current spread is low, the expected future spread is too low (actual > forecast = positive error), and when the current spread is high, the expected future spread is too high (actual < forecast = negative error). This visual motivates a simple econometric test of predictability with results in Table 1 wherein I regress the actual credit spread (averaged over the next four quarters), the current forecaster expectation (averaged over the next four quarters), and the forecast error (actual spread minus spread expectation) all on the current spread. The evidence does not fit well with the idea that people have rational expectations and instead suggests that analysts' forecasts tend to go through cycles of growth and decline. Column (3) shows that when the current spread is high, the higher the forecast relative to the realization is (thus producing a negative error); the same is true in reverse. When bond markets are doing well (low spreads), expectations are too optimistic and tend to go back to more realistic levels in the future, which can lead to a downturn in the bond markets.

In this space, my paper is most similar to Gilchrist and Zakrajsek (2012) and López-Salido et al. (2017). The first finds that credit spreads contain valuable information about the economy, and that much of this information comes from changes in excess bond risk premiums. While some of the variation these premiums can be explained by frictions in the traditional rational expectations approach, it could also be seen as sentiment. Gilchrist and Zakrajsek use a "bottom-up" approach to construct a credit spread index allows them to accurately measure investors' expectations of future economic outcomes, despite the presence of time-varying risk premiums, spanning back to mid-1970s. They use a sample of US non-financial firms covered by the S&P's Compustat database and the Center for Research in Security Prices (CRSP). López-Salido et al. (2017) measure credit-market sentiment based on the expected return to bearing credit risk from 1929 to 2015; more specifically, they use the ex-ante predictable component of corporate bond returns as a proxy for this sentiment and find that elevated sentiment from $t-2$ years ago is associated with a decline in economic activity in years t and $t+1$.[1] López-Salido et al. address concerns that changes in economic activity are driven by external non-financial factors, and that credit spreads simply reflect these changes in advance. This predictable component of changes in credit spreads therefore reflects past shifts in investor sentiment. The current analysis combines the construction of a proxy for credit spread expectations approach from Gilchrist and Zakrajsek with time span and the predictive regression framework from López-Salido et al.. By contrast, I employ textual analysis from a news source to construct my historical credit spread expectations, which I assume can be used as a direct measure of credit-market sentiment. This sentiment in the corporate bond market is likely to be a key channel of economic transmission that is closely tied to perceptions of credit risk in the financial system. In order to train my machine learning model, I also assume that the dynamics of the observable credit spread expectations remain consistent throughout the entire sample. This revised version expresses the same ideas as the original sentence, but uses different language and sentence structure to convey these ideas.

The second strand concerns itself with applications of text analysis and machine learning to the fields of economics and finance. For a contemporary introduction, see Gentzkow et al. (2019) who detail the overarching steps by which most of the research in this space, including my own analysis, follows. In essence, Step 1 is to represent the text $\mathcal{D}$ as a numerical array $\mathcal{W}$. Step 2 is to map $\mathcal{W}$ to the predicted values $\tilde{\mathcal{V}}$ of unknown outcomes $\mathcal{V}$. Step 3 is to use the predicted $\tilde{\mathcal{V}}$ in either descriptive or causal analysis. There are various methods by which to use each step, and their applications have spanned widely.

From using text-mining techniques to extract sentiment from financial statements and examining asset returns (Ke et al., 2019; Yue and Jing, 2022), to quantifying political stance and risk using newspaper prints and digital media (Giavazzi et al., 2020; Caldara and Iacoviello, 2022), to even analyzing text from FOMC announcements to study the effects of transparency and measuring

[1]In their analysis, elevated sentiment means that the expected return to bearing credit risk is low.

monetary policy surprises (Hansen et al., 2018; Shapiro and Wilson, 2019; Doh et al., 2020; Gorodnichenko et al., 2021), the text analysis approach is gaining traction in economics due to computational efficiency. I contribute to this area by following a number of key papers to predict expectation error of credit spreads through time. First, following Doh et al. (2020), I use vector embedding and employ the cosine similarity approach to represent similar texts within the front pages of the Wall Street Journal (WSJ). This choice of newspaper is guided by similar studies filling in historical variables using text from WSJ such as Manela and Moreira (2017) who fill in the CBOE S&P 500 Volatility Index (aka VIX), and Kelly et al. (2021) who do the same for nonfarm payroll employment and housing starts. Next, following the Textual Factors model from Cong et al. (2020), I employ Locality Sensitive Hashing (LSH) to cluster similar embeddings that most closely predict expectation errors; LSH is a method borrowed from the neural information processing literature and has been used in a variety of social science applications (Andoni et al., 2015). Lastly, I use a Latent Dirichlet Allocation technique to uncover topics in the unstructured WSJ text data without linking themes to particular word lists prior to my estimation. This approach is used in Hansen et al. (2018) to uncover latent themes in the text database of FOMC transcripts and is meant to produce a number of textual factors by which I can quantitatively calculate loadings, or how much a document represents a certain topic, that can serve as a proxy for the sentiment regarding credit spreads over time.

3 Data and Construction of Explanatory Variables

After describing the acquired data from public databases, I proceed with the construction of my explanatory variables which are textual factors that can be thought of as proxies for sentiment through time. These proxies are generated using front page articles from the Wall Street Journal which have been used in a variety of economic studies I describe in Section 2. The acquired data on BAA credit spread dates back to January 1919 and is available until the end of the third quarter of 2022 (1919Q1 to 2022Q3), totaling 1245 months, or 415 quarters. The Survey of Professional Forecasters began acquiring BAA credit spread expectations in January 2010 until present day; matching the end of the third quarter limit from the realized credit spread means it spans 2010Q1 to 2022Q3 and covers a total of 153 months, or 51 quarters.[2] The macroeconomic data obtained

[2]In comparison, the longest survey that includes credit spread expectations is the proprietary Blue Chip Financial Forecasts which have span from 1999Q1. Studies that have used it for credit spread analysis, such as Bordalo et al. (2018), analyze up until 2014Q4 which spans 192 months, or 64 quarters, or about 25% more quarterly data on BAA credit spread expectations than the SPF as of this analysis.

from other public data sources such as FRED span, at most, 1948Q1 to 2022Q3, which is 897 months or 299 quarters.

3.1 Acquired Data

My entire analysis relies on publicly available data for which historical realized values are plentiful. Sources such as the Federal Reserve Economic Database (FRED) and Capital Markets Data provide the realized values of my key indicator of credit market conditions which is the difference between the Moody's seasoned BAA Corporate Bond Yield and the 10-Year Treasury Bond Yield (interchangeably referred to as the BAA credit spread). This measure can be calculated from data ranging back to January 1919 until the most recent complete quarter of 2022 as of writing. Much less plentiful are data on the forecast values of the BAA Corporate Bond Yield, which, when looked at its difference with the 10-Year Treasury Bond Yield, can be referred to as the forecast of the BAA credit spread (or BAA credit spread expectation). Proprietary sources, such as Moody's historical BAA Corporate Bond yield forecasts and the forecasts from the Blue Chip Financial Forecasts services, have been often used in the aforementioned literature. One advantage of using the Blue Chip Financial Forecasts, for instance, is that information is solicited monthly from 40 panelists working in major financial institutions such as S&P Global, Goldman Sachs, and Citibank; the Blue Chip data, and expectation data on this variable in general, is limited and for this service is available from January 1999 onward. Forecasts for the current quarter up to five future quarters are obtained and averaged to derive a consensus forecast that can be used to derive a BAA credit spread expectation; the difference between this expectation and the realized BAA credit spread is called the expectation error.

I opt to use a similar forecast solicitation survey through the Federal Reserve Bank of Philadelphia which, since January 2010, has been collecting forecasts of the BAA Corporate Bond Yield in their Survey of Professional Forecasters (SPF). The SPF consists of a limited panel of professional forecasters (anywhere from 30 to 45 respondents per survey round) and they give their quarterly projections on major macroeconomic indicators up to four quarters ahead. Most forecasters use a mathematical model, adjusted with their subjective judgments, to ascertain their projection while many update forecasts at a monthly frequency (Stark, 2013). I obtain the average point forecasts of the BAA Corporate Bond Yield that are forecasters' projections of the current quarter t to up to four quarters $(t + 4)$ ahead forecast, given that they have up until $t - 1$ quarter's information. Mathematically, the SPF provides credit spread $CS_{t+\tau|t-1}$ for $\tau = 0, 1, 2, 3, 4$. Coupled with the realized 10-Year Treasury Bond Yield, I can calculate a measure of BAA credit spread expectation in much the same way as I would if I were using proprietary data.

For the analysis of forecasting GDP and other macroeconomic variables, I follow standard sources such as FRED and the ArchivaL Federal Reserve Economic Data (ALFRED) from the St. Louis Federal Reserve Bank. The most historical vintage on GDP is from 1929 and is provided on a yearly basis. Given that the SPF uses quarterly data, I decide to limit my analysis on the most common vintage for GDP data which begins in 1948. On a quarterly basis, I take GDP, unemployment, the CPI Inflation Rate, the 3-Month Treasury Yield, and the 10-Month Treasury Yield from 1948Q1 to 2022Q3, which is 897 months or 299 quarters.

3.2 Constructed Data — *The Wall Street Journal*

Most newspapers, magazines, and other sources of information where historical expectations data could be reported are found in scanned images that archives and other historical data repositories allow. To create a more thorough range of data on credit spread expectations, I would need to access these images and extract the data from them. With recent advances in technology, we now have optical character recognition (OCR) software that allows me to perform this function efficiently. I use Textract, a machine learning process hosted by Amazon Web Services (AWS), to create my own textual factor loadings from inputted text data from the Wall Street Journal front pages on expectations for the implied BAA credit spread from January 1919 (1919Q1) to September 2022 (2022Q3). This will complement the Survey of Professional Forecasters data by the Federal Reserve Bank of Philadelphia which covers from 2010Q1 until 2022Q3 by creating a proxy for sentiment on credit spread expectations; I will use the textual factor loadings from the realized credit spread expectations from 2010Q1 to 2022Q3 as the training data and then predict those past credit spread expectations from 1919Q1 to 2009Q4. Since the Wall Street Journal is printed six days a week (excluding 8 federal holidays), I simplify the analysis by averaging the textual factors within a given month (a median of 26 observations per month) or within a given quarter (a median of 78 observations per quarter).

Sections 3.2.1 to 3.2.3 follow the Textual Factors model from Cong et al. (2020) which quantitatively represents text data while preserving interpretability through its informational structure; I leave the main derivations to their paper and discuss the main points below. The goal is to create a series of textual factors that can be loaded into a regression model to estimate the expectation error through time. These factors are created as vectors that quantitatively explain main variations in texts using locality of words and similarities across texts as the main drivers. In some sense, they can also be thought of as a proxy for sentiment across time. This process is computed in three steps using text data as an input. I represent the process through a simple diagram in Figure 2 and go into the high level thought process of each step below before describing the algorithmic details in the proceeding subsections.

First, as I explain in Section 3.2.1, I will create vector representations of words that account for their semantic and syntactic meanings. These "word embedded" vectors will result in a rather large data set where words in various documents are represented by a vector according to distance and similarity with each other whose fixed dimension is set by a neural network; this process is guided by the assumption that words with similar meanings are often used together. For a simple example, consider a scenario where the rule is to create a vector representation using average frequency of neighboring words, and the fixed dimension is the top 5 most frequent words. In a list of words such as *credit*, *stimulus*, and *Senate*, we might find that the words most frequently neighboring *credit* and *stimulus* are *money*, *card*, and *check*. For *Senate*, it might be *decision* and *bill*. Following our rule, the word *credit* might be represented by $[4, 8, 5, 2, 1]$ which means *money* neighbors the word an average of 4 times, *card* neighbors the word an average of 8 times, *check* neighbors the word an average of 5 times, *decision* neighbors the word an average of 2 times, and *bill* neighbors the word an average of 1 time. In the same way, *stimulus* might be represented by $[5, 2, 4, 2, 2]$ and *Senate* might be represented by $[0, 0, 0, 9, 10]$. Note that the actual rule will be more complex than just frequency and rely on the context that the word is being used in. A large collection (often billions of words) gets fed to the model and a sliding window is used to capture the words that lie on either side of the word to determine its context; this context is then represented by its embedding vector that is controlled by a negative sampling process to update weights given to surrounding words. Because words with a similar context usually have closely-linked meanings, such words will end up having similar embedding vectors too. Even in the above example, dimensionality is already a concern; with just three words and with a fixed dimension of 5 words, we have created a 3 x 5 matrix that represents a document; in this case, a document is the front page of a Wall Street Journal volume which would have W number of words to represent.

Second, as I explain in Section 3.2.2, I first tackle the high dimensionality problem from Step 1 in Step 2 by clustering similar words together using a method called Locality Sensitive Hashing. Cascitti et al. (2022) provide a timely and layman-friendly article in the computational science space but the main idea behind this method is to use a hash function that takes an arbitrary amount of data and approximately categorizes it according to bins; these bins are chosen according to the context of the word space it is analyzing based on locality, or distance, to other words. Then, the clustering is more likely to happen for input text values that are close together rather than for inputs that are far apart. For example, consider the word *amazon*. This word, and its attached vector representation, could mean the worlds largest natural rainforest if words such as *jungle* and *biome* were nearby versus referring to the multinational tech company if words such as *alexa* and *.com* were nearby. The word would then be clustered differently based on its locality. Following the simple example from earlier, the three words *credit*, *stimulus*, and *Senate* would be clusters if their locality of neighboring words in that particular Wall Street Journal page were similar enough, and they would become the representative clusters for that document. In essence, I take the W number of words and divides them into K clusters.

Lastly, with these word vector embeddings from Step 1 clustered in different bins according to their locality in Step 2, I employ a topic model to reduce dimensionality by representing different words according to similar topics per document. This process is done via a Latent Dirichlet Allocation and will allow me to find documents according to whatever query I desire. For example, if I want to query the phrase *credit spread*, LDA will search for that phrase amongst the topics it has attributed different documents to. This is where the clustering from Step 2 serves it purpose: the clusters guide the topic search so that the algorithm does not need to comb through all the words. In essence, the K word clusters get fed into the algorithm as an educated guess of what the topics representing that document means, reducing computational time to finding the most optimal topic representation of that document. From the earlier example, a query with the term *credit spread* will pick up the document represented by the three clusters of *credit*, *stimulus*, and *Senate* since *credit* is a match, but it will rank that document less than one whose three clusters are, say, *credit*, *spread*, and *recession*. Thus, the algorithm will identify topics, or textual factors, that are represented by a set of words and the relative frequency distribution of their frequency in a given body of documents. Given that I am looking for specific topics in a body of words related to the BAA credit spread expectation, I can require that the query searches through clusters which have relevant words such as *BAA, bond, Treasury, yield, expectations, spread,* etc.

3.2.1 Word Embedding

The initial stage in any textual analysis is to summarize or represent the words that are present in the texts; this is called embedding. Less complex and count-based statistical models for textual analysis in the social sciences frequently use the one-hot vector encoding representation where words (dubbed N-grams) are treated as extremely high dimensional vectors/indices over a vocabulary with only one 1 and lots of 0s, omitting any consideration of the semantic relations among words. The flaw here is that we end up with words treated as independent units; *economic boom, upturn, strong markets* would all be treated as unrelated, which is inaccurate. To overcome this, the literature turns to semantic vector-space models with real-valued vector representations to create one-hidden-layer neural-network models; starting with Mikolov et al. (2013), this approach has gained traction in economics research to create a proxy for sentiment through these vector representations (Cheng et al., 2022; Dubovik et al., 2022; Fano and Toschi, 2022). Specifically, I follow the Word2Vec version of word embedding, pre-trained by Google on its Google News dataset, to let me filter out common typos, words and phrases that are too frequent (words such as *the, it, then*), and other common words not associated with the sentiment I want to capture (for example, *table of contents, what's news, DOW JONES, wsj.com* and other recurring words on the front page). This processing method for newspaper data is similar to the one done in Manela and Moreira (2017). Once fed documents, the Word2Vec model loops on the words therein to map them to a real-valued p-dimensional vector less than the size of the document vocabulary $\mid V \mid$ to a create a learned embedding vector $w \in \mathbb{R}^{p\times V}$. It will then calculate the distances between vectors w_i and w_j with a metric called cosine similarity, defined as

$$\text{similarity}(w_i, w_j) = \arccos\frac{\langle w_i, w_j\rangle}{\mid\mid w_i \mid\mid \mid\mid w_j \mid\mid} = \frac{\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j}{\sqrt{\sum_{i=1}^{n} w_i^2}\sqrt{\sum_{j=1}^{n} w_j^2}}$$

where $w_{i,j}$ are the two vectors being compared with their different indices and where a higher cosine similarity implies a more similar vector representation (and thus a more similar semantic meaning in the context of the sentences it appears in). This produces distances in a range $[-1, 1]$ denoting total opposites to exactly the same. For example, if the algorithm is trained on the words *Apple, Bill Gates*, it might spit out similarity$(Microsoft, SteveJobs) = 0.781$. I pre-process the words as per Manela and Moreira (2017) to reduce the dictionary of words I consider and then, following Mikolov et al. (2013), use the Word2Vec model to generate a 300×1 vectors for each remaining word that appears in the data from the front pages of the Wall Street Journal through time.

3.2.2 Clustering

To cluster the $w \in \mathbb{R}^{p\times V}$ word vectors, I turn to Locality Senstitive Hashing which returns the nearest-neighbor information through constructing a series of hash functions H that assert the similarity of items in order to put them into bins. Conceptually, on a 2D space, you can think of these bins as being created by random lines the algorithm is generating to cut up the observations based on similarity of the observations. In essence, the hash functions are generally claiming that vectors are similar when they are close together. More specifically, for any random element $h(\cdot) \in H$, which is our case are the word embedded vectors,

$$P[h(w_i) = h(w_j)] = p_1, \quad \text{for any } w_i, w_j \text{ such that } d(w_i, w_j) \leq d_1$$
$$P[h(w_i) = h(w_j)] = p_2, \quad \text{for any } w_i, w_j \text{ such that } d(w_i, w_j) \geq d_2$$

For any well defined LSH family $H(d_1, d_2, p_1, p_2)$, p_1 is the probability of retrieving points that are close to a query point and thus I would like to maximize this value; a query point, in this analysis, is a word such as *credit*. We can think of $1 - p_1$ as the occurrence of false negatives meaning that some points that are closer than distance d_1 to the query point won't be retrieved in the sample. Meanwhile, p_2 is the probability of retrieving points that are further than desired to our query point; p_2 is the probability of false positives which I want to minimize. With this in mind, then we can extrapolate that if similarity(w_i, w_j) is high, then $P[h(w_i) = h(w_j)]$ is high as well. Following the tech leader example, if similarity$(Microsoft, SteveJobs) = 0.781$, then we can infer that the probability that the two words are in the same bin is relatively high, too. To define how the hash function family is generated to make the bins, I turn to the random hyperplane projection method which produces a spherically symmetric random vector r of unit length from the $p-$dimensional space using a signum function; in my analysis, since I follow Mikolov et al. (2013), $p = 300$. More concretely, the hash function family for vectors w can be described as

$$h_r(w) = sgn(\langle r, w \rangle), \quad \text{for } r \text{ randomly sampled from the unit sphere } S^{p-1}$$

Using this method, I reduce the highly-dimensional vectors of word embeddings by generating hash functions with good performance in finding the nearest neighbors of a query word, thus creating my bins, or clusters, K.

3.2.3 Topic Modeling

Topic modeling is a statistical technique that uses the distribution of words in a provided texts to identify their underlying semantic structure and generate a set of topics. For example, in my analysis, a Wall Street Journal front page following Black Monday might contain words like "trade deficit," "financing mergers," and "SEC Chairman" more frequently than other front pages, while a front page about Apple's valuation might contain words like "cash generating," "visionary," and "Tim Cook" more often. The Wall Street Journal may have common topics through time such as crisis, wars and conflicts, natural disasters and healthcare, the Federal Reserve or monetary policy, taxes or Congress, technology, and employment. In practical research, the amount of documents in a given area makes it near impossible to manually categorize them by topic. However, topic modeling can help us automatically identify the topics discussed in each document by observing their word distributions; for this, I follow the Latent Dirichlet Allocation (LDA).

Following the economics literature using computational linguistics such as Bholat et al. (2015), Hansen et al. (2018), and Tobback et al. (2017), I assume a simple, two-distribution data-generating process where each of my Wall Street Journal front pages are generated from a distribution over a collection of topics. In turn, each topic is generated from a distribution of words in a dictionary; many authors opt to feed this dictionary the words most relevant to their analysis by creating something based on frequency while others use pre-trained dictionaries such as the aforementioned one used by Google. Given the rich vocabulary often used in the Wall Street Journal, the computation of the distributions would be onerous. However, by feeding the clusters from Section 3.2.2 as the topic dictionary, I significantly reduce the search complexity of the topic word distributions. Mathematically, and following Cong et al. (2020), let the notation $\beta_k \sim \text{Dirichlet}(\eta)$ be a multinomial distribution over the dictionary of words for each topic and $\theta_d \sim \text{Dirichlet}(\alpha)$ be a multinomial distribution over K topics for a particular document d. Then, the word-generating process for any document is such that I sample a specific topic $z_{di} \in (1, 2, \ldots, K)$ with $z_{di} \sim \theta_d$ and then sample the observed word $w_{di} \sim \beta_{z_{di}}$ from the entire document vocabulary V. In an expression, this takes the form of

$$[\Theta B]_{dw} \coloneqq P(w_{di} = w \mid [\theta_d, \beta_1, \beta_2, \ldots, \beta_K]) = \sum_k \theta_{dk}\beta_{kw}$$

where I am calculating the probability of word w_{di} to be equal to word w in the defined dictionary space, and where $\Theta = [\theta_1, \theta_2, \ldots, \theta_D]' \in \mathbb{R}^{D \times K}$ and $B = [\beta_1, \beta_2, \ldots, \beta_K]' \in \mathbb{R}^{K \times V}$. Then, denote the product $[\Theta B]_{dw} = N_{dw} \in \mathbb{R}^{D \times V}$ as the number of times a word w appears in a document d such that, in different topics, different words are assigned different weights. This implicitly assumes that topics that allocate similar weights to words are related more closely to each other than those that do not. Then, each document fulfills being described by the two distributions above: the probability that a document covers a certain topic and the probability that topic itself has certain words assigned to it. I follow Mikolov et al. (2013) and limit the size to the top 300 topics the method identifies as important.

3.2.4 Validation of Factors

The output from the above subsections are the K textual factors which are represented by the word support S of the word cluster i (the words which are being used to denote similarity to other topics), the real-valued vector representing the textual factor F_i, and the factor importance d_i (the similarity measure). Together, they are a triplet of information by which Cong et al. (2020) use the following projection to create the loadings of the textual factor i from:

$$x_i^d := \frac{\langle N_{S_i}^d, F_i \rangle}{\langle F_i, F_i \rangle} \tag{1}$$

A thought experiment most easily helps understand these textual factor loadings. Unlike structured data documents, which list quantifiable amounts for relevant topics such as stock prices or the 10-Year Treasury Bond Yield, unstructured data has text which may use a certain vocabulary to define the overall sentiment of the topic. For instance, texts from the Wall Street Journal front page surrounding the 2008 Financial Crisis might have centered around discussions covering things like *financial crisis, investor confidence, fiscal policy,* and *bankruptcy*. The x_i^d obtained in Equation 1 allow me to assign a quantitative measure to how much the document "loads" on that topic; each document d can then be represented by the loadings $x_1^d, x_2^d, \ldots x_K^d \in \mathbb{R}^K$.

For a visualization of how these loadings translate to analysis, consider the checks on accuracy via the word cluster and loadings chart for the word *recession*. In Figure 3, I find that words like *financial*, *depression*, *inflation*, *GDP*, and *unemployment* all have a close proximity to the main word *recession* in the vector space measured by their cosine similarity. Additionally, in sub-figure (b), the textual factors that are generated follow consistent patterns with real-time events that would have those relevant Wall Street Journal front pages related to the topic, such as during the Great Depression, Post WWII spending declines, the oil crises of the 70s and 80s, and the Great Recession.[3]

[3]To further decrease the noise from the textual loading generation process, I follow Manela and Moreira (2017) and omit other words and phrases found regularly in the Wall Street Journal such as *"business and finance", "world wide", "what's news", "table of contents", "masthead", "other", "no title", and "financial diary"*.

4 Methodology for Credit Spreads and Macroeconomic Indicators

My methodology for this section can be broken into two parts. First, I predict past credit spreads through the entire span of available realized BAA credit spreads by first following the predictive regression methodology for credit spreads in Bordalo et al. (2018) and then enhancing that with my textual factor loadings. I predict the BAA credit spread expectations from 1919Q1 to 2009Q4 by first training my model with the SPF data from 2010Q1 to 2022Q3. Then, with the predicted credit spread, I can generate my series of the credit spread expectation error by subtracting out the realized credit spread. With the predicted credit spread expectation error series, I am then able to run predictive regressions on changes in macroeconomic indicators across different time horizons. In short, I use the textual factors from Equation 1 to predict BAA credit spread expectations in Equation 2, which in turn allow me to calculate the expectation error that I treat as the independent variable in Equation 3.

4.1 Historical Credit Spread Expectations

Similar to Manela and Moreira (2017) who use regressions to predict a Volatility Index (VIX) using n-gram frequencies from their body of text, I will do the same to predict BAA Credit Spread Expectations using the textual factor loadings from my body of text. Given the high number of topics that will be involved in the analysis, I opt to use penalized Lasso regressions so that only the topics that most contribute to the credit spread expectations stand out.[4] Then, using the BAA Credit Spread Expectations data from the SPF covering 2010Q1 through 2022Q3, I estimate the following:

$$E[CS]_t = \alpha + \mathbf{\Gamma}\mathbf{x}_t^D + \eta_t, \quad t = 1, 2, \ldots, T \tag{2}$$

where the BAA credit spread expectations $E[CS]_t$ for quarter t are being calculated by the textual factor loadings $\mathbf{x}_t^D$ estimated during the training time period; $\mathbf{\Gamma}$ is a K vector of regression coefficients which, from the LDA using the inputted text, uses 300 topics, and T is the training time period for $T = 51$ observations. I compare this with the Bordalo et al. (2018) model from Section 2 where $E[CS]_t = \alpha + \gamma[CS]_t + \epsilon_t$. Next, I perform the same analysis as in Equation 2 but this time predicting the variable over the entire span of the textual factor loadings from 1919Q1 to 2009Q4.

[4] Lasso stands for Least Absolute Shrinkage and Selection Operator. This approach works well with my assumption that there are a select number of significant variables that influence credit spread expectations while the rest are close to zero.

From this, I can calculate the predicted credit spread expectation error $\widehat{E[CSE]}_t$ for the entire time span by comparing it to the realized BAA credit spread.

4.2 Predicting Macroeconomic Indicators

Similar to López-Salido et al. (2017) who use the expected return to bearing credit risk to predict macroeconomic indicators, I follow their methodology to predict changes in GDP, unemployment, and domestic investment using the changes in predicted credit spread expectation error as follows:

$$\Delta y_{t+h} = \beta_0 + \beta_1 \Delta \widehat{E[CSE]}_t + \boldsymbol{\gamma}' \mathbf{x_{t-1}} + \upsilon_t \tag{3}$$

where Δy_{t+h} will be, in different variants, the log-difference of real GDP per capita, the change in unemployment rates, and the log-difference in domestic investment, all over the course of quarter t to horizon h.[5] $\Delta \widehat{E[CSE]}_t$ is the change of the predicted credit spread expectation error from quarter $t-1$ to t, and the controls $\mathbf{x_{t-1}}$ include the credit spread in quarter $t-1$, the log-difference of each predicted indicator y_t from quarter $t-2$ to $t-1$, the CPI inflation rate in quarter $t-1$, and the changes in both the 3-month and 10-year Treasury yields from quarter $t-2$ to $t-1$. Newey-West standard errors are estimated with the Newey and West (1994) automatic lag-selection that allows correction for heteroskedasticity and autocorrelation.

This regression approach is merely predictive and cannot pinpoint the casual relationship between the expectation error and the macroeconomic indicators. Instead, we can interpret the coefficients as those that will predict a certain path for the indicator in the future. For example, a negative and statistically significant coefficient on the change in predicted credit spread expectation error is predicting that a positive expectation error can predict negative GDP growth in the corresponding time horizon.

5 Results

I break down my results into first a discussion about the textual factors and the validity of how I fill in the historical values for credit spread expectation error, and then the results from my predictive regressions for macroeconomic indicators.

The time series of the actual and predicted errors are in Figure 4. As a comparison of the strength of my textual factors for BAA credit spread expectation errors, I take the Bordalo et al. (2018) test

[5] As per Subsection 3.1, the horizons are defined as $h = 0, 1, 2, 3, 4$

as a basis to see how the model with all historical data fits into the finding that current spreads overly predict expectation error (Column (3) in Table 1, panel A, where $E[CSE]_t = \beta_0 + \beta_1[CS]_t + \varepsilon_t$). When filling in the historical expectation error for credit spreads calculated with the textual factors, I find that the current credit spread is a significant contributor to the error with a coefficient of 0.456 (standard error of 0.079). In comparison, without the textual factors, the current credit spread is a significant contributor to the error with a coefficient of 0.564 (standard error of 0.093). Since this was a machine learning exercise, I look to the in-sample and out of sample R^2 as a means to express variation through time. To do this, I follow the standard machine learning literature and conduct a k-fold cross-validation to estimate the skill my model had on the unobserved credit spread expectation error. By convention, I leave $k = 10$ and split my sample into 10 groups and find that the in-sample R^2 without textual factors is 0.57 but with the factors it jumps to 0.69, a 21.5% increase. The more important test is the out-of-sample R^2 as it will show how the model can predict data it has not yet seen. In this case, my out-of-sample R^2 without textual factors is 0.41 but with the factors it jumps to 0.52, a 26.8% increase!

I present the results from the predictive regression for growth of Real GDP in Table 2. The columns each signify which quarter horizon is being calculated, with $h = 0$ meaning the current forecast, or nowcast, and each subsequent column denoting that many quarters ahead up to four quarters, or one full year, ahead. I find that the change in my predicted credit spread expectation error is statistically significant at the one year horizon, with a negative predictor that implies a one standard deviation jump in expectation error is associated with a predicted real GDP growth decline of 0.03 percentage points. This corroborates the finding in Bordalo et al. (2018) that an increase in expectation error (when the forecasts of credit spreads get higher than the actual realized values, or when sentiment is overly optimistic) predates a decline in economic activity. Similar to López-Salido et al. (2017), I also find that the most recent credit spread and changes in the 10-Year Treasury Yield positively impact real GDP growth at the one year horizon.

I present the results from the predictive regression for changes in the unemployment rate in Table 3. The columns each signify which quarter horizon is being calculated, with $h = 0$ meaning the current forecast, or nowcast, and each subsequent column denoting that many quarters ahead up to four quarters, or one full year, ahead. I find that the change in my predicted credit spread expectation error is statistically significant in the nowcast as well as throughout the quarter forecasts up until the one year ahead, all positively significant. In other words, when the expectation error continues to increase (when sentiment is elevated), there is an associated increase in the unemployment rate.

With magnitudes, a one-standard deviation jump in the credit spread expectation error leads to an average of 1.68% increase in the unemployment rate over the next year (an average calculated using the coefficients associated with the variable for the nowcast and the subsequent h quarters ahead). The story complements the one told in Table 2. I find that the most recent spread has a negative association with a one standard deviation in the change from the most recent spread leading to a -0.51% drop in the unemployment rate. No other controls have statistically significant predictive power across the horizons.

Lastly, I present the results from the predictive regression for growth in private domestic investment in Table 4. The columns each signify which quarter horizon is being calculated, with $h = 0$ meaning the current forecast, or nowcast, and each subsequent column denoting that many quarters ahead up to four quarters, or one full year, ahead. I find that the change in my predicted credit spread expectation error is statistically significant for the one year horizon, suggesting that a one-standard deviation jump in credit spread expectation error predicts a 2.9% decline in one year. Similarly, the coefficients for the most recent credit spread show a negative association with the domestic investment growth. This result suggests that changes in private domestic investment respond more strongly following periods associated with overly optimistic sentiment in the market. No other controls have statistically significant predictive power across the horizons. In all, the three aforementioned results are consistent with overly optimistic sentiment in the credit market predating declines in economic activity.

The results here corroborate the story of the behavioral models described in the introduction where there are systematic biases in expectations and whereby errors to the credit spread expectations of the future are seemingly correlated with macroeconomic indicators. However, there is little that my textual factors approach can say regarding the different channels of transmission that these errors are playing into. That said, the historical filling in of errors in credit spread expectations through time remain consistent with the shorter expectations data we do have in the recent decades.

6 Conclusion

In this paper, I analyze how credit spread expectation errors can predict future macroeconomic indicators through different time horizons. To do so, I focus on the BAA credit spread which provides useful information about investors' perceptions of risk and the overall health of the economy. Current research looking at the interplay between credit markets and the macroeconomy provides two approaches where one focuses on shock amplification that affects underlying fundamental factors and the other emphasizes the role of beliefs, such as excessive optimism, as a reason for business cycles. Unifying the two approaches may be timely as the economy presently moves into a decade that has been marked so far by a systemic shock (COVID-19) and largely global events such as war and investor uncertainty. To do so, it is a timely endeavour to find empirical evidence of how this interplay works since even policy makers agree on how credit spreads affect the market and should be interwoven in policy. The difficulty arises in acquiring historical data that can be used in economic analysis for periods that cover more than just the recent two decades.

This analysis makes usage of the Survey of Professional Forecasters to first motivate the systematic biases narrative from 2010 through 2022, and then uses the data to train a machine learning model that uses textual analysis on Wall Street Journal title pages, such as in Figure 6, to derive textual factors that I use to calculate errors in credit spread expectations. The resulting series, which I think of as a stand in for sentiment, is then used in predictive regressions to forecast macroeconomic indicators on a quarterly basis. I find that increases in the errors for credit spread expectations, associated with over optimism in the credit market, generally predict downturns in economic activity.

These findings, further confirming predictable patterns in credit spread expectation errors, suggest real implications for policy makers who may try to abate economic activity declines by incorporating credit market analysis in their considerations. Additionally, the proxy for sentiment here is derived from a popular news source for market participants but speaks little to the drivers of sentiment and how it manifests into actual market behavior. To do so would require more work than the current analysis but proves useful guidance if we can hope to provide quantitative advice to policy makers.

References

Andoni, A., Indyk, P., Laarhoven, T., Razenshteyn, I., and Schmidt, L. (2015). Practical and optimal lsh for angular distance. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Arellano, C., Bai, Y., and Kehoe, P. J. (2019). Financial Frictions and Fluctuations in Volatility. *Journal of Political Economy*, 127(5):2049–2103.

Baron, M. and Xiong, W. (2017). Credit Expansion and Neglected Crash Risk. *The Quarterly Journal of Economics*, 132(2):713–764.

Bernanke, B. and Gertler, M. (1989). Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review*, 79(1):14–31.

Bholat, D., Hans, S., Santos, P., and Schonhardt-Bailey, C. (2015). *Text mining for central banks*. Number 33 in Handbooks. Centre for Central Banking Studies, Bank of England.

Bianchi, J. (2011). Overborrowing and Systemic Externalities in the Business Cycle. *American Economic Review*, 101(7):3400–3426.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020a). Overreaction in Macroeconomic Expectations. *American Economic Review*, 110(9):2748–2782.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020b). Overreaction in Macroeconomic Expectations. *American Economic Review*, 110(9):2748–2782.

Bordalo, P., Gennaioli, N., Porta, R. L., and Shleifer, A. (2019). Diagnostic Expectations and Stock Returns. *Journal of Finance*, 74(6):2839–2874.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2018). Diagnostic Expectations and Credit Cycles. *Journal of Finance*, 73(1):199–227.

Caldara, D. and Iacoviello, M. (2022). Measuring Geopolitical Risk. *American Economic Review*, 112(4):1194–1225.

Cascitti, J., Niebler, S., Müller, A., and Schmidt, B. (2022). Rnacache: A scalable approach to rapid transcriptomic read mapping using locality sensitive hashing. *Journal of Computational Science*, 60:101572.

Cheng, B., Ling, X., and Qiang, Y. (2022). Deep representation learning of text-based economics and other knowledges. *China Journal of Econometrics*, 2(1):58.

Cong, L. W., Liang, T., Yang, B., and Zhang, X. (2020). Analyzing Textual Information at Scale. In *Information for Efficient Decision Making Big Data, Blockchain and Relevance*, World Scientific Book Chapters, chapter 10, pages 239–271. World Scientific Publishing Co. Pte. Ltd.

Doh, T., Song, D., and Yang, S.-K. X. (2020). Deciphering Federal Reserve Communication via Text Analysis of Alternative FOMC Statements. Research Working Paper RWP 20-14, Federal Reserve Bank of Kansas City.

Dubovik, A., Fiedler, C., and Parakhonyak, A. (2022). Temporal Patterns in Economics Research. CPB Discussion Paper 440, CPB Netherlands Bureau for Economic Policy Analysis.

Eggertsson, G. B. and Krugman, P. (2012). Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach. *The Quarterly Journal of Economics*, 127(3):1469–1513.

Fano, S. and Toschi, G. (2022). COVID-19 pandemic and the economy: sentiment analysis on Twitter data. *International Journal of Computational Economics and Econometrics*, 12(4):429–444.

Gennaioli, N., Ma, Y., and Shleifer, A. (2016). Expectations and investment. BIS Working Papers 562, Bank for International Settlements.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, 57(3):535–574.

Giavazzi, F., Iglhaut, F., Lemoli, G., and Rubera, G. (2020). Terrorist Attacks, Cultural Incidents and the Vote for Radical Parties: Analyzing Text from Twitter. Working Papers 659, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Gilchrist, S. and Zakrajsek, E. (2012). Credit Spreads and Business Cycle Fluctuations. *American Economic Review*, 102(4):1692–1720.

Gorodnichenko, Y., Pham, T., and Talavera, O. (2021). The Voice of Monetary Policy. NBER Working Papers 28592, National Bureau of Economic Research, Inc.

Greenwood, R. and Hanson, S. G. (2013). Issuer Quality and Corporate Bond Returns. *Review of Financial Studies*, 26(6):1483–1525.

Hansen, S., McMahon, M., and Prat, A. (2018). Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach. *The Quarterly Journal of Economics*, 133(2):801–870.

Jordá, O., Schularick, M., and Taylor, A. M. (2013). When Credit Bites Back. *Journal of Money, Credit and Banking*, 45(s2):3–28.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2019). Predicting Returns With Text Data. NBER Working Papers 26186, National Bureau of Economic Research, Inc.

Kelly, B., Manela, A., and Moreira, A. (2021). Text Selection. *Journal of Business & Economic Statistics*, 39(4):859–879.

Kubin, I., Zörner, T. O., Gardini, L., and Commendatore, P. (2019). A credit cycle model with market sentiments. *Structural Change and Economic Dynamics*, 50(C):159–174.

Leiva-León, D., Perez-Quiros, G., Sapriza, H., and Zakrajšek, E. (2022). Introducing the Credit Market Sentiment Index. *Richmond Fed Economic Brief*, 22(33).

López-Salido, D., Stein, J. C., and Zakrajšek, E. (2017). Credit-Market Sentiment and the Business Cycle. *The Quarterly Journal of Economics*, 132(3):1373–1426.

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

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

Minsky, H. P. (1977). The Financial Instability Hypothesis: An Interpretation of Keynes and an Alternative to"Standard" Theory. *Challenge*, 20(1):20–27.

Newey, W. K. and West, K. D. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, 61(4):631–653.

Schularick, M. and Taylor, A. M. (2012). Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008. *American Economic Review*, 102(2):1029–1061.

Shapiro, A. H. and Wilson, D. J. (2019). The Evolution of the FOMC's Explicit Inflation Target. *FRBSF Economic Letter*.

Stark, T. (2013). SPF panelists' forecasting methods: A note on the aggregate results of a November 2009 special survey. Research working paper, Federal Reserve Bank of Philadelphia.

Stein, J. C. (2014). Incorporating Financial Stability Considerations into a Monetary Policy Framework : a speech at the International Research Forum on Monetary Policy, Washington, D.C., March 21, 2014. Speech 796, Board of Governors of the Federal Reserve System (U.S.).

Tobback, E., Nardelli, S., and Martens, D. (2017). Between hawks and doves: measuring central bank communication. Working Paper Series 2085, European Central Bank.

Yue, T. and Jing, M. (2022). Sentiment Analysis Based on Bert and Transformer. In Yang, H., Qiu, R., and Chen, W., editors, *AI and Analytics for Public Health*, Springer Proceedings in Business and Economics, pages 317–328. Springer.

Table 1. [Motivation] Predictability Tests on Credit Spreads (Actual, Forecast, and Error)

	Actual Spread (1)	Forecast Spread (2)	Error (Actual - Forecast) (3)
Current Spread	0.32*	0.51***	-0.31***
	(0.19)	(0.11)	(0.08)
Constant	1.16**	1.23**	1.82***
	(0.49)	(0.51)	(0.37)
Total Observations	51	51	51
R^2	0.64	0.54	0.26

Notes: Quarterly time series regressions following Table I in Bordalo et al. (2018). In columns (1) - (3), the independent variable is the actual credit spread averaged over quarters $t-4$ to $t-1$ prior to the forecast given in quarter t. Then, (1) is the actual credit spread averaged between the Q1 and Q4 forecasts ahead of survey date (Actual average between $t+1$ to $t+4$), (2) is the forecasts of credit spreads averaged between Q1 and Q4 forecasts ahead of survey date (Forecast average between $t+1$ to $t+4$), and (3) is the forecast error (actual minus forecast) of credit spreads. Credit spread forecasts are the consensus forecasts computed from the Survey of Professional Forecasters spanning 2010Q1 to 2022Q3. Newest-West standard errors, with the automatic bandwidth selection from Newey and West (1994), are in parentheses.
* p < 0.10, ** p < 0.05, *** p < 0.001

Table 2. [Results] Predictive Regressions for Log-Difference of Real GDP (Growth)

	Real GDP Growth (h quarters ahead)				
	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
	(1)	(2)	(3)	(4)	(5)
Δ Predicted Expectation Error	-0.010	-0.018	-0.019	-0.016	-0.029***
	(0.006)	(0.013)	(0.012)	(0.011)	(0.010)
Most Recent Credit Spread	0.006	0.012	0.008	0.011	0.014**
	(0.004)	(0.009)	(0.012)	(0.007)	(0.006)
Most Recent log Δ GDP	0.459***	0.516***	0.613***	0.514***	0.556***
	(0.084)	(0.080)	(0.091)	(0.063)	(0.079)
Most Recent CPI	0.071	0.089	0.084	0.101	0.076
	(0.051)	(0.068)	(0.083)	(0.85)	(0.127)
Δ 3-Month Treasury Yield	-0.169	-0.097	0.107	-0.018	0.004
	(0.141)	(0.157)	(0.154)	(0.149)	(0.163)
Δ 10-Year Treasury Yield	-0.478	-0.461	-0.459	-0.513*	-0.610*
	(0.398)	(0.349)	(0.363)	(0.267)	(0.340)
R^2	0.419	0.261	0.213	0.224	0.233

Notes: Quarterly time series regressions with sample period 1948Q1 through 2022Q3. All specifications include a constant (not reported). Δ Predicted Expectation Error is the change of the predicted credit spread expectation error from quarter $t-1$ to t, Most Recent Credit Spread is the credit spread in quarter $t-1$, Most Recent log Δ GDP is the log-difference of GDT from quarter $t-2$ to $t-1$, Most Recent CPI is the CPI inflation rate in quarter $t-1$, Δ 3-Month Treasury Yield is the change from quarter $t-2$ to $t-1$, and Δ 10-Year Treasury Yield is the change from quarter $t-2$ to $t-1$. Heteroskedasticity- and autocorrelation-consistent asymptotic Newey-West standard errors are reported in parentheses and use the automatic lag selection method of Newey and West (1994).

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

Table 3. [Results] Predictive Regressions for Changes in Unemployment Rate

	Change in Unemployment(h quarters ahead)				
	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
	(1)	(2)	(3)	(4)	(5)
Δ Predicted Expectation Error	1.871**	1.654**	1.665**	1.598**	1.601**
	(0.903)	(0.815)	(0.820)	(0.781)	(0.766)
Most Recent Credit Spread	-0.471	-0.416	-0.443	-0.439	-0.509**
	(0.291)	(0.281)	(0.274)	(0.279)	(0.219)
Most Recent Δ Unemployment	0.761***	0.707***	0.699***	0.774***	0.797***
	(0.292)	(0.264)	(0.248)	(0.282)	(0.287)
Most Recent CPI	0.026	0.039	0.040	0.033	0.031
	(0.018)	(0.026)	(0.027)	(0.023)	(0.019)
Δ 3-Month Treasury Yield	-0.004	0.001	0.000	-0.001	-0.003
	(0.004)	(0.003)	(0.001)	(0.002)	(0.004)
Δ 10-Year Treasury Yield	-0.011	-0.009	-0.010	-0.011	-0.007
	(0.009)	(0.010)	(0.012)	(0.016)	(0.008)
R^2	0.319	0.147	0.189	0.109	0.192

Notes: Quarterly time series regressions with sample period 1948Q1 through 2022Q3. All specifications include a constant (not reported). Δ Predicted Expectation Error is the change of the predicted credit spread expectation error from quarter $t-1$ to t, Most Recent Credit Spread is the credit spread in quarter $t-1$, Most Recent Δ Unemployment is the change in unemployment rate from quarter $t-2$ to $t-1$, Most Recent CPI is the CPI inflation rate in quarter $t-1$, Δ 3-Month Treasury Yield is the change from quarter $t-2$ to $t-1$, and Δ 10-Year Treasury Yield is the change from quarter $t-2$ to $t-1$. Heteroskedasticity- and autocorrelation-consistent asymptotic Newey-West standard errors are reported in parentheses and use the automatic lag selection method of Newey and West (1994).

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

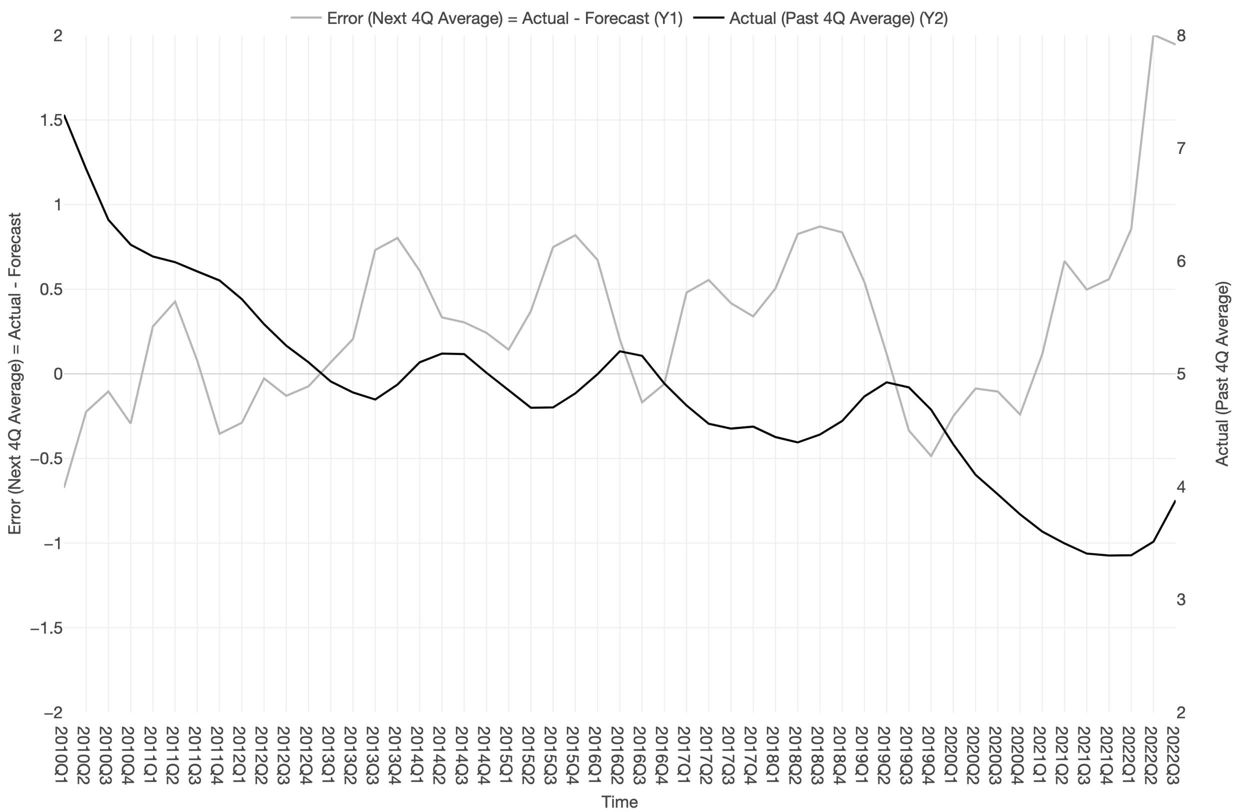
Table 4. [Results] Predictive Regressions for Log-Difference of Domestic Investment (Growth)

	Domestic Investment Growth (h quarters ahead)				
	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
	(1)	(2)	(3)	(4)	(5)
Δ Predicted Expectation Error	-0.019	-0.026	-0.021	-0.024	-0.029***
	(0.014)	(0.019)	(0.016)	(0.015)	(0.011)
Most Recent Credit Spread	-0.014**	-0.016**	-0.011**	-0.018**	-0.020**
	(0.007)	(0.007)	(0.005)	(0.008)	(0.010)
Most Recent log Δ Investment	0.185**	0.197*	0.251*	0.213*	0.194*
	(0.082)	(0.116)	(0.137)	(0.112)	(0.104)
Most Recent CPI	0.017	0.009	0.007	0.016	0.004
	(0.010)	(0.006)	(0.009)	(0.012)	(0.008)
Δ 3-Month Treasury Yield	0.008	0.006	0.003	0.001	0.005
	(0.012)	(0.005)	(0.009)	(0.011)	(0.012)
Δ 10-Year Treasury Yield	0.0012	0.002	0.005	0.013	0.007
	(0.013)	(0.011)	(0.015)	(0.012)	(0.017)
R^2	0.213	0.189	0.181	0.193	0.204

Notes: Quarterly time series regressions with sample period 1948Q1 through 2022Q3. All specifications include a constant (not reported). Δ Predicted Expectation Error is the change of the predicted credit spread expectation error from quarter $t-1$ to t, Most Recent Credit Spread is the credit spread in quarter $t-1$, Most Recent log Δ Investment is the log-difference of private domestic investment from quarter $t-2$ to $t-1$, Most Recent CPI is the CPI inflation rate in quarter $t-1$, Δ 3-Month Treasury Yield is the change from quarter $t-2$ to $t-1$, and Δ 10-Year Treasury Yield is the change from quarter $t-2$ to $t-1$. Heteroskedasticity- and autocorrelation-consistent asymptotic Newey-West standard errors are reported in parentheses and use the automatic lag selection method of Newey and West (1994).

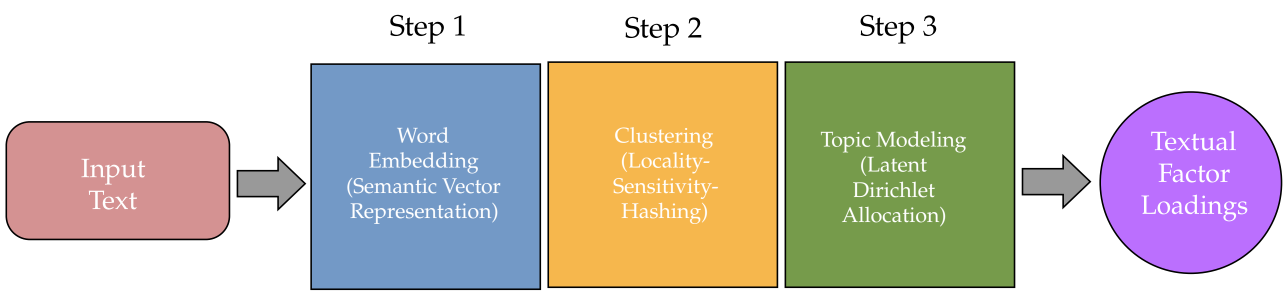
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Predictable errors in forecasts of credit spreads.



Notes: Quarterly time series plot following Figure 1 in Bordalo et al. (2018). In each quarter t, the gray line shows credit spread expectation errors (actual minus forecast) averaged over quarters $t+1$ to $t+4$ (left scale), and the black line shows the actual credit spread average over quarters $t-4$ to $t-1$, where $t-1$ is the latest quarterly credit spread prior to the forecast (right scale). Credit spread forecasts are the consensus forecasts computed from the Survey of Professional Forecasters spanning 2010Q1 to 2022Q3.

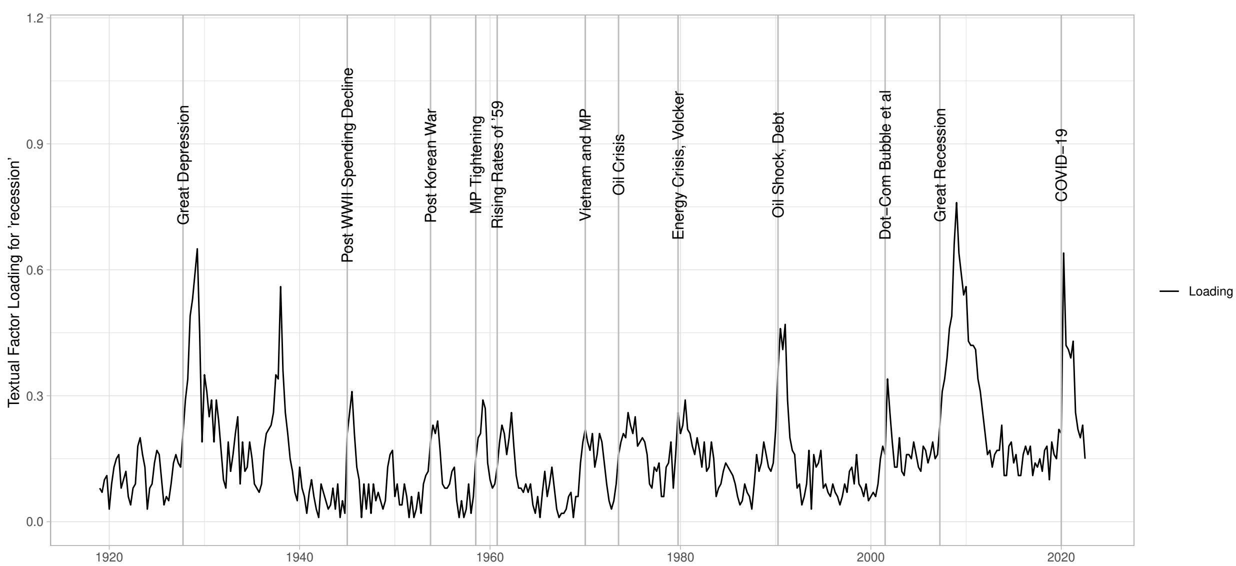
Figure 2: Method Diagram to obtain Textual Factor Loadings



Notes: Diagram depicting the process of obtaining the Textual Factor Loadings that are used in Equation 1. Input Text is coming from the digital versions of newspapers obtained through the OCR process. Step 1 transforms the text into vector representations (text embeddings) using the *word2vec* function. Step 2 clusters these vector forms of words in a way that reduces dimensionality. Step 3 creates topics from the clustered word vectors. The result, the textual factors, allow me to create loadings on various topics that are fed into predictive regressions in Section 4.

Figure 3: Accuracy check on Constructed Variable Data following the Textual Factors Model

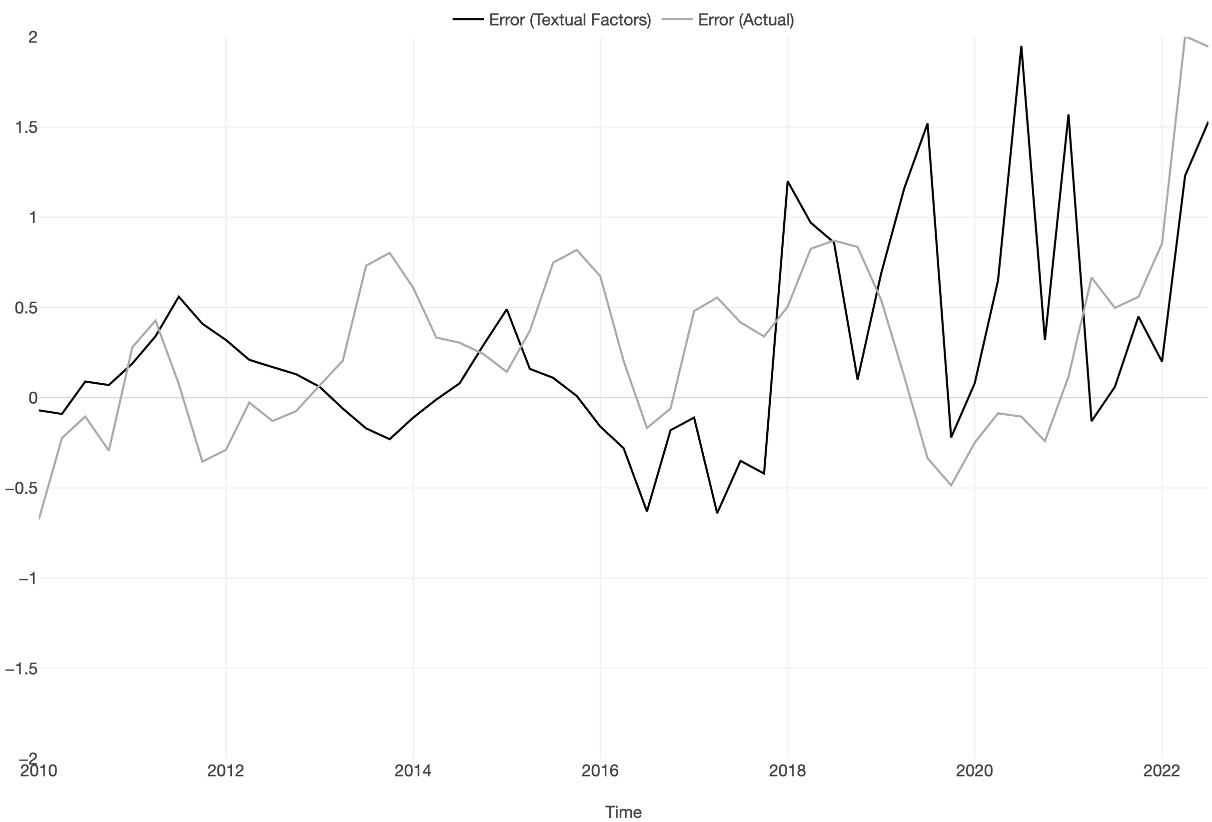
(a)



(b)

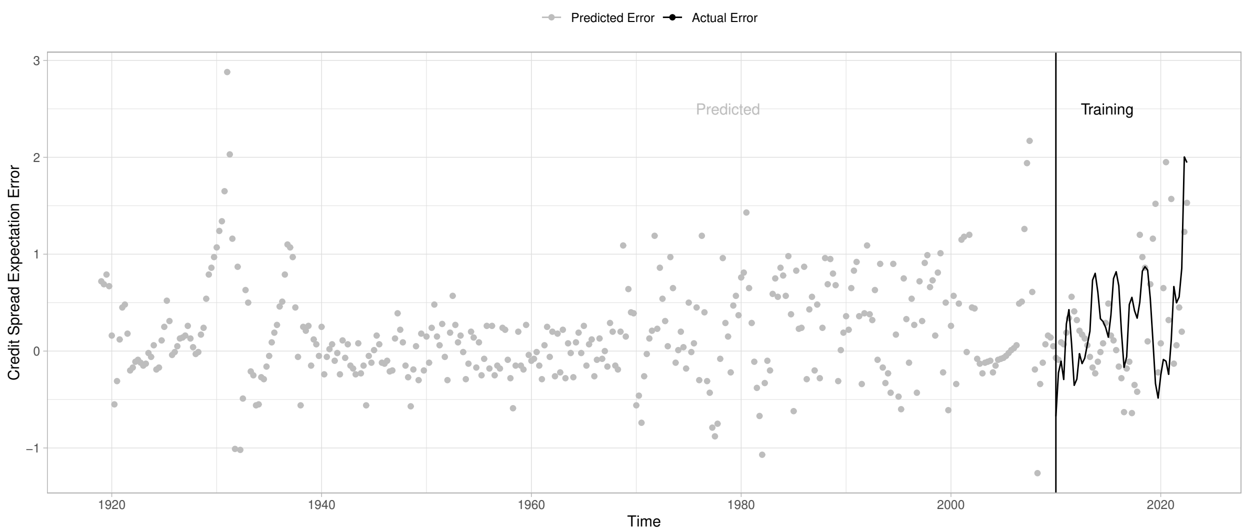
Notes: (a) Sample cluster based on the computational linguistics approach in Section 3 for the word "recession". (b) Loadings on the textual factors for the word "recession" through time using the Wall Street Journal front page data as the input source data for the algorithm. Data processing follows the filtering methodology of Manela and Moreira (2017) to leave out common phrases and stop-words including, but not limited to, "an", "the", "it".

Figure 4: Comparison between Actual Credit Spread Expectation Error and Predicted Errors using Textual Factors, 2010Q1 - 2022Q3.



Notes: Quarterly time series plot. The gray line shows credit spread expectation errors (actual minus forecast) from the consensus forecasts computed from the Survey of Professional Forecasters. The black line shows the predicted credit spread expectation errors calculated using the textual factors.

Figure 5: Predicted Credit Spread Expectation Error, 1919Q1 - 2022Q3.



Notes: Solid line is the realized expectation error between SPF consensus forecasts . Dots are credit spread expectations error as derived from the penalized Lasso regressions in Equation 2 that derive the expected BAA credit spread expectations from textual factor loadings and are used to calculate the error relative to historical 10-Year Treasury Yields. The Training sub-sample, 2010Q1 to 2022Q3, is used to estimate the dependency between the textual factors and implied expectation error. The Predicted sub-sample includes all earlier observations for which BAA Credit Spread Expectations and, hence, Credit Spread Expectations Error, are not available.

Figure 6: Sample document of the Wall Street Journal Front Page; November 9, 1988.

Notes: Throughout its publication, the WSJ has had a number of fixed entities on the front page. For example, the "What's News—" section has either been in the middle-left or left-most column of the front page and dedicates brief snippets to news topics. Over time, graphics became more prominent on the front page, cutting the amount of text as well.