Contents

Early warning systems for macroeconomic risks through news-based outlier detection

Abstract

Macroeconomic risks, such as economic downturns or financial crises, threaten businesses and their operations. Hence, firms not only closely monitor the economic climate in order to be able to adapt quickly, but they also demand early warning systems. This paper proposes an innovative system that infers macroeconomic risk from the soft content of financial news. Previous research has documented the relation between price response on the market and news sentiment (Tetlock, 2007; Tetlock et al., 2008; Feuerriegel & Prendinger, 2016). Here we follow Mamaysky & Glasserman (2016) whereby a deviation from earlier news stories coincides with a higher level of risk. Mathematically, we formalize this through an outlier detection analysis applied to the content of financial narratives: an excessive number of outliers serves as an early warning of economic downturns. As part of our empirical demonstrations, we utilize 11, 902 regulatory disclosures from January 2004 to June 2011, finding a statistically significant relationship between outliers and recessions. The benefits of this approach are obvious as it makes fairly little assumptions (e. g. with regard to macroeconomic models) and is thus straightforwardly generalizable. Thereby, our work entails direct implications for operations management, as practitioners can readily exploit our proposed approach in practice.

Keywords:
Text mining, Financial crisis, Macroeconomic risk, Outlier detection, Financial news, Anomaly analysis

1. Introduction

If there is a feature that contemporary societies share across the world, it is to be prone to economic crises. The 20th and the 21st centuries have been characterized by an increasing number of recessions, price bubbles and other type of economic crises. Recent examples are the 1991 Indian economic crisis, the 1997 Asian financial crisis, the Dot-com bubble of the early 2000's and the 2008 global financial crisis from which the American and European economies have only recently recovered (Allen & Gale, 2009). Economic crises are characterized by their far reaching and long lasting impact on central sectors of society including, but not limited to, health care provision (Cutler et al., 2002), structural unemployment (Perkins et al., 2013), income inequality (Perkins et al., 2013) and the efficiency of credit markets (Bernanke, 1983). It is for instance estimated that the American economy shrank by 50 % during the first five years of the Great Depression (Romer, 1988). Chang et al. (2013) also observed that suicide rate significantly rose during the years that followed the 2008 economic crisis in the 27 European countries as well as in most countries of the American continent.

Anticipating economic crises with appropriate fiscal and monetary policies can mitigate the impact that they have on society, increase the resilience of the economy and help accelerate the rate of recovery (Perkins et al., 2013; Christiano et al., 2002; Braggion et al., 2007). For these reasons, economic policy-makers have devoted much effort to the improvement of models forecasting economic crises. The existing literature on the prediction of financial instability primarily relies on time series analyses of macroeconomic indicators (Frankel & Saravelos, 2012). For instance, Estrella & Mishkin (1998) examine the performance of various financial variables as predictors of past U. S. recessions, Lo Duca & Peltonen (2013) find early warnings of financial instability in domestic and global macrofinancial vulnerabilities and Borio & Lowe (2002) show that widespread financial distress typically arises from the unwinding of financial imbalances. Yet, the fact that the 2008 economic crisis took almost everybody by surprise – only a handful of economists had foreseen it – shows that the existing forecasting techniques are in dire needs of improvements; Greenspan (2013) even talks of an *"existential crisis for economic forecasting"*.

In this paper we take an alternative approach to economic forecasting: instead of conducting time series analyses of macroeconomic indicators, we draw upon the theoretical framework developed by Antweiler & Frank (2004); Tetlock (2007); Tetlock et al. (2008) and base our forecast of the future state of the economy on information embedded in soft content. Their work are descendants of the intuition of John maynard Keynes who coined the term *animal spirit* 70 years ago to indicate that a significant part of the behavior of investors on the stock market appears to be unjustified by macroeconomic fundamentals. These studies show that soft content, such as news articles or social media posts, contains information about the economy that macroeconomic indicators may fail to reflect; thereby shedding light on Keynes' notion of *animal spirit*. Indeed, investors rarely directly observe firms' productive activities but usually get their information secondhand, for the most part from financial news. For this reason, it should not come as a surprise that soft content is valuable material for analyzing market activity (Tetlock, 2007). A growing literature has explored this theme: Tetlock et al. (2008) show that financial news convey information about firm earnings above and beyond stock experts forecasts and historical data analyses, Bollen et al. (2011) quantifies the general mood of Twitter's users in an effort to predict the return direction of the S&P500 and Hanley (2016) uses computational linguistics of bank's risk disclosures to predict financial instability.

In the continuity of these researches, we conduct an anomaly analysis of financial news in an effort to forecast economic crises. We conceptualize economic crises as large-scale events that offset the normal disposition of the economy and therefore approach the task of crises forecasting as an anomaly detection problem[1]. To this end, we build upon earlier studies which extract economic information from soft content (Tetlock et al., 2008; Bollen et al., 2011; Hanley, 2016) and investigate whether an anomaly analysis of financial news provides reliable forecasts for economic crises. Drawing upon the work of Mamaysky & Glasserman (2016) who find that the unusualness of financial news forecast economic volatility up to several months ahead, we conjecture that the months leading an economic crisis are characterized by an increasing number of unusual events which are reflected in financial news via the occurrence of unusual messages and can be detected by an anomaly analysis. We further compare the predictive power of an anomaly analysis conducted onsoft content, i. e. financial news, with that of an anomaly analysis conducted on macroeconomic indicators. Since the 2008 economic crisis is the only major financial downturn that has occurred since the widespread use of the Internet made financial news readily available to all stakeholders, it is the focus of the paper.

We find that soft content can be modeled to reliably predict the occurrence of the 2008 economic crisis. In particular, we show that an anomaly analysis conducted on financial news provides a forecast of economic instability up to 24 months ahead. Importantly, the anomaly analysis conducted on the term-document matrix of the corpus of financial news offers more reliable predictions for economic crises than an anomaly analysis conducted on sentiment scores, suggesting that term-document matrices are valuable material to analyze textual data in the domain of finance. Furthermore, our findings show that anomaly levels computed from macroeconomic variables have less explanatory power than those computed from financial news, confirming the claim by Tetlock (2007); Tetlock et al. (2008) that soft content contains information about the future economic outlook that is not necessarily reflected in common macroeconomic indicators and extending these results to the forecasting of financial instability.

The common wisdom in economics is that macroeconomic indicators offer reliable economic forecasts (Frankel & Saravelos, 2012). Yet the 2008 financial crisis showed that they failed to predict what has been the most devastating economic event since the Great Depression (Greenspan, 2013). A growing literature has started to explore the influence of soft content on stock market activity; yet, most researchers have only conducted a sentiment analysis of soft content (Tetlock et al., 2008; Bollen et al., 2011; Hanley, 2016). Our research distinguishes itself by extending the existing research on soft content to anomaly analysis and showing that it is a reliable predictor for the future state of the economy.

These results entail considerable implications for research and practice. First, future researches should further the study of anomaly in soft content in order to refine the relationship between changes in anomaly levels and the occurrence of economic crises. Researchers may also want to investigate additional anomaly detection algorithms: since each identifies a spe-

[1] In this paper, we refer to outliers and anomalies interchangeably.

cific type of anomaly, considering different anomaly detection techniques may reveal new insights (Chandola et al., 2009). Secondly, stakeholders of the financial sector e.g. investors, policy makers should integrate the results of this research in their forecasting models where soft content should complement macroeconomic indicators. This will help them increase the accuracy of their predictions regarding the future evolution of the economy. How soft content is to be incorporated to existing forecasting models should be the topic of future researches. Finally, social media content is gaining more importance in the study of financial phenomena. Future research should extend the anomaly analysis we conducted on financial news to social media content such as the tweets of the platform Twitter.

This paper is organized as follows. Section 2 provides an overview of related work on text mining of financial news. Next, Section 3 describes our methodology with a focus on the processing of the textual data. Section 4 presents the main findings and Section 5 discusses their implications for academia and the stakeholders of the financial sector, while Section 6 concludes.

2. Related Work on Natural Language Processing Techniques Applied to Financial News

Much research has been conducted to study the impact that the content of financial news has on the stock market. Three studies seem particularly relevant to this paper. Firstly, Tetlock (2007) and Tetlock et al. (2008) attempt to characterize the relationship between soft material such as media reports and the behavior of investors on the stock market. Tetlock (2007) shows that the content of news media content can be modeled to predict movements in indicators of stock market activity. To accomplish this, he constructs a measures of media pessimism using a principal component analysis. He finds that high levels of pessimism in the media forecast significant decreases in market prices and that unusually levels of media pessimism (either very high or very low levels) predict larger-than-usual trading volumes on the market. Similarly, Tetlock et al. (2008) quantify the language used in financial news stories to forecast stock returns. They find that stock prices respond to information embedded in media content with a delay of one day, indicating a potential source of profits with the use of high-frequency trading strategies. Together, these findings suggest that media content incorporate aspects of firms fundamentals that are otherwise hard-to-quantify and are not captured by macroeconomic indicators and which investors quickly incorporate into stock prices.

More recently, Feuerriegel & Prendinger (2016) have furthered this line of research by successfully developing a trading system that generates profit by utilizing the relationship between news-based financial data and movements on the stock market. More specifically, they conduct a sentiment analysis to extract subjective information from financial disclosures and thereby determine a measure of the tone of the texts which then provides an indicator of the expected return direction. Together with the momentum of the associated stock price, the measure of the tone of the texts provides the basis for a trading strategy which can potentially generate benefits.

Finally, more closely related to the approach taken in this paper, Mamaysky & Glasserman (2016) investigate the relation between the level of unusualness of financial news and market volatility. In particular, they study how the former can be modeled to forecast the latter. Their findings show that not only sentiment matters, but so does unusualness when predicting financial stability. Working with a 4-gram model and the measure of entropy, they find that the unusualness of financial news can be modeled to forecast economic volatility up to several months ahead.

3. Method Development

3.1. Intuition Behind Methodology

Economic policy-makers heavily rely on forecasting models to estimate the future state of the economy. Improving the accuracy of these models can help them better anticipate future recessions and thereby mitigate their effects on society (Perkins et al., 2013; Christiano et al., 2002; Braggion et al., 2007). Tetlock et al. (2008) shows that financial news contains information about the economy which, if extracted, can help improve economic forecasts. One may consequently be tempted to develop a classifier to predict economic crises; yet, we oppose this approach for the following two reasons. Firstly, there are only a handful of economic crises whose contemporary financial news have been well-preserved, offering little training data for the development of a classifier. Secondly, every crisis is inherently different from the others for the following reason: once the root causes of a crisis are identified, policies are implemented in an effort to prevent similar events from happening again. It is thus not surprising that Frankel & Saravelos (2012) find that the indicators that are relevant for predicting a particular crisis tend to be irrelevant for predicting others, making the task of developing a classifier unlikely to succeed.

Instead, we conceptualize economic crises as large-scale events that offset the normal disposition of the economy and therefore approach the task of crises forecasting as an anomaly detection problem. To this end, we build upon earlier works which extract economic information from soft content (Tetlock et al., 2008; Bollen et al., 2011; Hanley, 2016) and investigate whether an anomaly analysis of financial news provides reliable forecasts for economic crises. Drawing upon the work of Mamaysky & Glasserman (2016) who find that the unusualness of financial news forecast economic volatility up to several months ahead, we conjecture that the months leading an economic crisis are characterized by an increasing number of unusual events which are reflected in financial news via the occurrence of unusual messages and can be detected by an anomaly analysis.

We therefore propose a methodology based on the identification of unusual messages in financial news. To accomplish this, we apply the k-nearest-neighbor algorithm to the term-document matrix of a given corpus of financial news. The main advantage of the k-nearest-neighbor algorithm is its capacity to detect documents with unusual terms without requiring the researchers to determine in advance which words are to be looked for, thereby avoiding a potential source of human bias. In addition, unlike most outlier detection algorithms, it works in an unsupervised setting and therefore does not require training data. Chandola et al. (2009) offers an extensive discussion of the respective merits and limitations of a myriad of outlier detection techniques. To our knowledge no prior study examining the influence of soft content on market activity has based its analysis on the term-document matrix of a corpus of financial news; this is a major innovation of our paper. The absence of term-document matrix in this type of research may be due to the fact that the technique maps the documents to a very high-dimensional space where the sparsity of the data points makes the identification of insightful patterns difficult to realize. In order to make the data tractable, we omit the terms with a sparsity level exceeding 0.9 from the term-document matrix. Besides reducing the number of dimensions, this step also eliminates terms with no direct link to the economy, allowing the researchers to focus on the more critical content words. Furthermore, since in practice, one only has access to the financial news that have been published in the past, when we compute the outlier score of a given document, we only consider the news that have been published *before* the document in question.

To evaluate the performance of this pioneering method for forecasting economic crises, we compare it to two alternative approaches. Firstly, we conduct a more common sentiment analysis (Pang & Lee, 2008; Manning & Schuetze, 1999; Feldman, 2013). Following the method employed by Feuerriegel & Prendinger (2016), we use a dictionary-based algorithm to quantify the polarity, i. e. positiveness or negativeness of the documents, and we conduct an anomaly analysis of the obtained sequence of polarity values with the seasonal-hybrid ESD algorithm. The seasonal-hybrid ESD algorithm is an anomaly detection algorithm specifically designed for univariate temporal data (Rosner, 1983). Its main advantages are its ability to take seasonal components into account and its capacity to detect both local and global outliers – two features essential for an outlier analysis of temporal data (Chandola et al., 2009). Secondly, we conduct an anomaly analysis based on a selection of four variables: the company's stock price at the moment of the document's publication, the real return of the stock, the number of words in the text and its sentiment score. We use this additional outlier score to compare the forecasting power of an outlier measure based on soft content, i. e. on a term-document matrix, to that of an outlier measure based on a mix of economic variables, i. e. the price and the real return of the stocks, and textual characteristics, i. e. the number of words and the sentiment of the texts.

Since crises are usually defined with a monthly resolution (Allen & Gale, 2009), we aggregate the obtained anomaly scores at the month level. More precisely, for each month, we compute the mean anomaly score obtained by the k-nearest-neighbor algorithm and we count the frequency of outliers found by the seasonal-hybrid ESD algorithm.

We then link the obtained output for each month, i. e. the mean outlier score calculated with the k-nearest-neighbor algorithm and the number of news identified as anomalous identified by the seasonal-hybrid ESD algorithm, to the state of the economy with a logistic model. This model uses the anomaly scores as predictors and the occurrence of an economic crisis in the following 24 months as a response variable. We compare the respective predictive power of both predictors using the Akaike Information Criterion (AIC) (Akaike et al., 1998). Models with smaller AIC offer the best fit of the data and are thus preferred.

We finally check the robustness of our results by varying the maximum level of sparsity of the terms included in the term-document matrix and by varying the k number of neighbors consider by the k-nearest-neighbor algorithm. Namely, we consider two term-document matrices with maximum sparsity levels of 0.9 and 0.99 and we consider four values for k in the k-nearest-neighbor algorithm: 5, 10, 50 and 100.

3.2. Overview of Methodology

Our methodology is as follows (see Figure 1). We first process the corpus of financial news by computing its term-document matrix and calculating the sentiment score of each document with a dictionary-based method. This is a common approach in sentiment analysis (Pang & Lee, 2008; Manning & Schuetze, 1999; Feldman, 2013; Feuerriegel & Prendinger, 2016). We then conduct an anomaly analysis on each of these two objects: we apply the k-nearest-neighbor algorithm to the term-document matrix and the seasonal-hybrid ESD algorithm to the sentiment scores. The former yield an anomaly score for each text and the latter determines whether or not a text is an outlier. Finally, we aggregate the results at the monthly level using the mean anomaly score and the monthly number of outliers before link the obtained output to the state of the economy with a logistic model. We compare the predictive power of the obtained models with the Akaike Information Criterion.

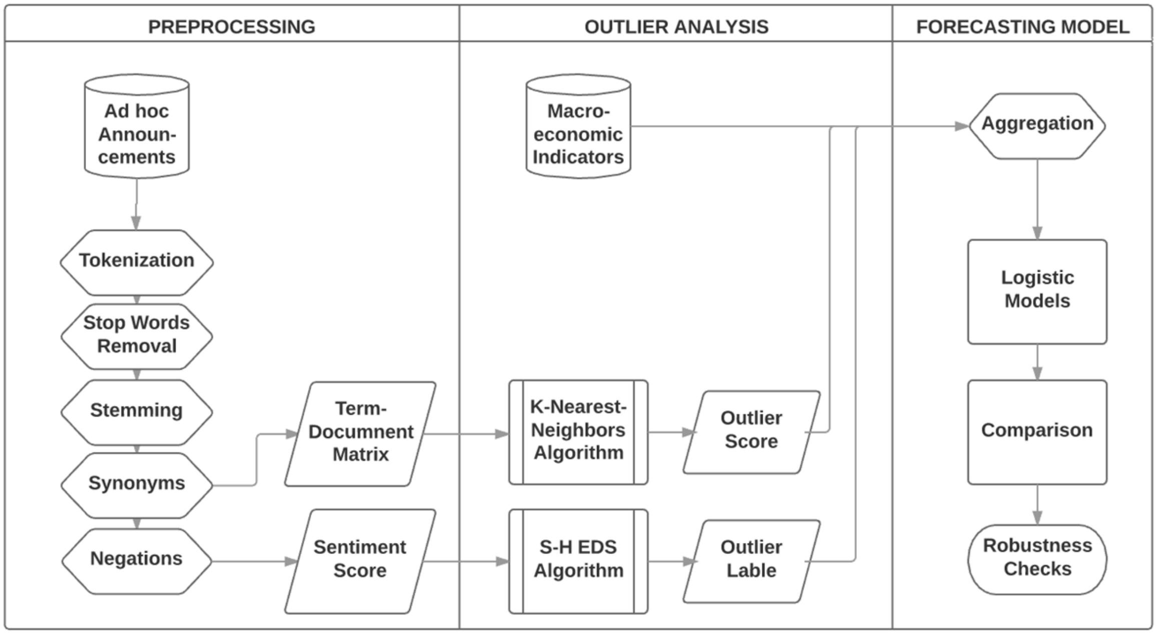

Figure 1: Flowchart

3.3. Processing of Narrative Material

We start by computing a term-document matrix of the corpus of financial news. Term-document matrices have to our knowledge never been used in prior researchers studying the influence of financial news on stock market activity; this is a major innovation of this paper. We first compute the frequency of each term present in the document and then calculate their term frequency-inverse document frequency (TF-IDF) (Salton & Buckley, 1988). The TF-IDF term-weighting system gives a larger score to terms appearing multiple times in the document being investigated while penalizing terms that are frequent in the corpus. This enables researchers to focus on the words that are discriminative for the documents. It is therefore a popular ponderation method in text mining and information retrieval (Crnic, 2011). Each column of the matrix corresponds to a term present in the corpus and each row to a text, with the entry $f_{i,j}$ of the term-document matrix indicating the TF-IDF of term j in document x_i. Since the produced term-document matrix maps the documents into a very high-dimensional space, we reduce the total number of variables by omitting the terms with a sparsity level exceeding 0.9. Besides making the data tractable, this step also eliminates terms with no direct link to the economy, allowing the researchers to focus on the more critical content words. In our empirical analysis, this step reduces the number of 12444 different terms down to the most critical 281 terms. In order to check the robustness of our findings, we also compute the term-document matrix that includes terms with a sparsity level inferior to 0.99. The latter term-document matrix therefore includes more terms (see Table 1).

Unless specified otherwise, the term-document matrices referred to in the remaining of this paper are term-document matrices with a maximum sparsity level of 0.9.

Table 1: Term-Document Matrix

Maximum Sparsity Level	**Number of Terms**
No maximum	$12,444$
0.99	$1,480$
0.9	281

To evaluate the performance of the term-document matrix in forecasting economic crises, we also conduct a sentiment analysis of the financial news and compare the two approaches. The latter technique is common in text mining and will serve as a benchmark. Sentiment analysis refers to methods dealing *"with the computational treatment of opinion, sentiment, and subjectivity in text"* (Pang & Lee, 2008, p. 4). In this paper, we quantify the polarity of financial news, that is, whether they have a globally *positive* or *negative* tone. To accomplish this, we follow common procedures in sentiment analysis and utilize a dictionary-based approach (Pang & Lee, 2008; Manning & Schuetze, 1999; Feldman, 2013). The main advantage of this approach is the possibility that it gives to researchers to use dictionaries tailored to their research topic. In our case, since we attempt to quantify the polarity of financial news, we use the Loughran-McDonald Dictionary (Loughran & McDonald, 2011). This dictionary is popular among researchers in economics since it was specifically created to analyze textual data in the field of economics and contains more than 85000 terms (Bodnaruk et al., 2015; Loughran & McDonald, 2016; Feuerriegel & Prendinger, 2016).

To quantify the polarity of financial news, we follow the method employed by Feuerriegel & Prendinger (2016). We first split the corpus entries into single terms (tokens) and remove stop words, i. e. terms without a deeper meaning such as *the*, *and* or *is* using a list of 571 stop words proposed by Lewis et al. (2004). Removing stop words is a common practice in text-mining that helps focus on significant terms (Manning & Schuetze, 1999). We then use the so-called Porter stemming algorithm to reduce words to their stem (Porter, 1980) and we utilize the finance-oriented Loughran-McDonald Dictionary to label the obtained terms as either positive or negative (Loughran & McDonald, 2011). For instance, the terms *advantages* and *damaging* are reduced to their stems *advantage* and *damage*, and respectively receive a *positive* and a *negative* label. Finally, we invert the polarity of the terms of negated sentences. The sentiment score $s(x_i)$ of a document x_i is calculated as the difference between the number of positive $p(x_i)$ and negative $n(x_i)$ stem terms as determined by the Loughran-McDonald Dictionary divided by the total number of stem words $w(x_i)$ of document x_i, i. e.

$$s(x_i) = \frac{p(x_i) - n(x_i)}{w(x_i)}. \tag{1}$$

The resulting sentiment score s ranges from -1 to $+1$ with larger values indicating a more positive tone and smaller values a more negative tone. These values later serve as input to the anomaly analysis.

3.4. Anomaly Analysis

We conduct two outlier analyses respectively on the term-document matrix and the sentiment scores. These two objects have specific characteristics requiring different methodologies (Chandola et al., 2009). Table 2 offers an overview of the different anomaly scores used in this paper. The term-document matrix maps the documents to a high-dimensional space where the data points are sparsely distributed. It therefore requires an approach for outlier detection that scales well to a large number of variables. Beyer et al. (1999) show that the k-nearest-neighbor algorithm scales well to high-dimensional spaces if the data points are relatively close to each other. The entries, i. e. the tf-idf of the terms, of the term-document matrix are inherently very small values and thereby ensure that the condition is fulfilled. Furthermore, unlike most outlier detection techniques that output a binary label indicating whether or not an observation is an outlier, the k-nearest-neighbor algorithm outputs a continuous outlier score indicating the level of anomaly of each instance with larger values indicating a larger level of unusualness. Since continuous variables always contain more information than categorical variables, having a continuous score instead of a label as output of the anomaly analysis will make our empirical model more powerful. Since in practice, one only has access to the disclosures that have been published in the past, when we compute the outlier score of a given document, we only consider the disclosures that have been published before it. That is, the k-nearest-neighbor algorithm look for the nearest neighbors of a document published at time t only among the pool of disclosures whose date of publication is prior to time t.

In contrast, the sentiment scores s_i obtained from the sentiment analysis form a univariate time series densely distributed around 0 and comprised in the interval $[-1, 1]$. It consequently requires an algorithm that accounts for seasonal effects and which is not only able to detect global outliers but also local ones, that is, data points that are outliers with regards to their most direct neighbors (Chandola et al., 2009). We therefore use the seasonal-hybrid ESD algorithm to conduct an anomaly analysis of the sentiment scores. The seasonal-hybrid ESD algorithm is an anomaly detection algorithm specifically designed for univariate temporal data. It is able to take seasonality into account and to detect both local and global outliers (Rosner, 1983). The algorithm is furthermore capable to differentiate between *positive* outliers i.e. observations with an abnormally large values and *negative* outliers i.e. observations with an abnormally small values. This latter feature will help us develop a more precise empirical model.

Table 2: Anomaly Scores

Anomaly score	**Input**	**Algorithm**
Multivariate anomaly score	4 variables (see Section 3.4.1)	k-nearest-neighbor algorithm
Term-document matrix anomaly score	Term-document matrix	k-nearest-neighbor algorithm
Positive sentiment anomaly score	Sentiment scores	Seasonal-hybrid ESD algorithm
Negative sentiment anomaly score	Sentiment scores	Seasonal-hybrid ESD algorithm

3.4.1. The k-Nearest-Neighbor Anomaly Detection Algorithm

The k-nearest-neighbor anomaly detection algorithm uses the distance between data points to compute outlier scores (Chandola et al., 2009). Using the term-document matrix, we first measure the Euclidean distance between the documents of the corpus via the formula

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (f_{i,k} - f_{j,k})^2}, \tag{2}$$

where $d(x_i, x_j)$ denotes the Euclidean distance between documents x_i and x_j, $f_{i,k}$ represents the tf-idf of term k in text x_i and n corresponds to the total number of terms present in the term-document matrix. The algorithm then identifies the k nearest neighbors of each data points using the Euclidean distance. The identification of neighboring data points is realized in an incremental way – we limit the search for nearest neighbors of an observation to the data points that occurred *before* the observation in question – in order to reproduce the way the algorithm would be used in practice. Namely, to find the k nearest neighbors of a document x_n, we search among the texts $x_1, x_2, \ldots, x_{n-1}$. The outlier score of a text then corresponds to the mean distance to its k nearest neighbors. Let $s(x_i)$ denote the anomaly score of text x_i and $d(x_i, x_j)$ be the Euclidean distance between text x_i and text x_j, with text x_j one of the k nearest neighbors of the text x_i, then

$$s(x_i) = \frac{1}{k} \sum_{j=1}^{k} d(x_i, x_j). \tag{3}$$

Texts that are very different from the rest of the data are consequently characterized by large anomaly scores and texts similar to the others receive small scores (Chandola et al., 2009).

Since the k-nearest-neighbor algorithm finds the nearest neighbors of the data points in an incremental way, the neighboring points of latter documents are found in a larger pool of documents than those of documents published earlier. Since later documents have a larger pool of previously published documents, it is easier to find close neighbors of document published late in time, than of documents published early. This means that the anomaly scores of the first documents published will tend to be inflated compared to those of later documents. In order to avoid this bias in anomaly score over time, we limit the pool of documents in which to look for neighboring data points to the 250 previously published documents. Furthermore, since the first 250 documents do not have a pool of 250 texts published before them, we burn them in, i. e. we do not compute their anomaly scores and exclude them in the empirical model.

Finally, we compute an additional outlier score using the k-nearest-neighbor algorithm that takes the following four variables as input: the company's stock price at the moment of the document's publication, the real return of the stock, the number of words in the text and its sentiment score. Throughout the remaining of the paper, the former anomaly score will be referred to as *term-document matrix anomaly score* and the latter one as *multivariate anomaly score*. The values of the two anomaly scores are extremely volatile over time. This makes the identification of relevant trends, i. e. an increase in anomaly score prior to the crisis, difficult to realize. In order to facilitate the identification of patterns, we construct indexes for the term-document matrix anomaly score and the multivariate anomaly score.

3.4.2. The Seasonal-Hybrid ESD Algorithm

The seasonal-hybrid ESD algorithm builds upon the Generalized Extreme Studentized Deviate (ESD) test for detecting anomalies in univariate variables (Rosner, 1983). This test has the advantage of only requiring that an upper bound for the suspected number of outliers be specified. In brief, given an upper bound r for the maximum number of outliers, the generalized ESD performs r Grubbs test sequentially (Grubbs, 1969): given a vector $X = (x_1, \ldots, x_n)^T$ of n real numbers, to test whether or not a value $x_i \in X$ is an outlier, the Grubbs test first computes its absolute standardized score

$$Z(x_i) = \frac{|x_i - \overline{x}|}{s}, \tag{4}$$

where $\overline{x}$ is the mean of vector X and s its standard deviation, and concludes that the value x_i is an outlier if

$$Z(x_i) > \frac{n-1}{n}\sqrt{\frac{(t_{\alpha/(2n),n-2})^2}{n-2+t_{\alpha/(2n),n-2}}}, \tag{5}$$

where $t_{\alpha/(2n),n-2}$ is the critical value of the t-distribution with $n-2$ degrees of freedom at the significance level $\alpha/(2n)$ and n the length of vector X (Grubbs, 1969).

Building upon the generalized ESD test, the seasonal-hybrid ESD algorithm furthermore takes seasonal effects and temporal correlation into account to detect outliers and is also able to differentiate between positive and negative outliers, making it a suitable algorithm for conducting anomaly analyses of financial news. Throughout the remaining of the text, the monthly frequency of documents identified as positive outliers by the seasonal-hybrid ESD algorithm will be referred to as *positive sentiment anomaly score* and the frequency of negative outliers as *negative sentiment anomaly score*.

3.5. Empirical Model

We use a logistic regression to model the outputs of the anomaly analyses conducted with the k-nearest-neighbor algorithm and the Seasonal-Hybrid ESD Algorithm in an effort to forecast the occurrence of an economic crisis in the following 24 months. In order to test the robustness of the results, we consider the following values for k in the k-nearest-neighbor algorithm: $5, 10, 50$ and 100. Since crises are usually defined with a monthly resolution (Allen & Gale, 2009), we aggregate the results of the anomaly analyses at a monthly level. More precisely, for each month, we compute the mean anomaly score computed by the k-nearest-neighbor algorithm and we count the number of positive and negative outliers found by the seasonal-hybrid ESD algorithm. Furthermore, we include a dummy variable for the month in the logistic regression in order to control for seasonal variations (Box et al., 2016). Finally, since our goal is to predict the imminence of a crisis when the economy is still growing, we exclude the observations that occurred during the crises.

The logistic regression produces a statistical model giving the probability that a crisis will occur in the next 24 months based on the anomaly score of the month in question and which month of the year it is using the equation

$$p_i = \frac{1}{1 + e^{-y_i}}, \tag{6}$$

where p_i denotes the probability of a crisis in the next 24 months following month i and y_i is a linear combination of the predictors with

$$y_i = \alpha + \beta\, x + \sum_{j=1}^{11} \gamma_j\, month_j, \tag{7}$$

where α, β and γ_j are regression coefficients estimated via the maximum-likelihood method, x is the output of one of the two anomaly analyses and $month_j$ is the dummy control variable for months (Hosmer et al., 2013). We do not allow for interaction between the predictors for reasons of parsimony.

We finally compare the respective predictive power of both predictors using the Akaike Information Criterion (AIC) in order to identify which of the created models offers the best fit of the data. Models with smaller AIC are preferred. The AIC is an estimator of the complete quality of a statistical model (Akaike et al., 1998). It takes into account not only the likelihood of the model, i. e. how well it fits the data points, but also its simplicity, i. e. how many parameters the model has. By penalizing models with large numbers of parameters, the AIC takes into consideration the trade-off between goodness-of-fit and parsimony. The main advantage of using models AIC as a selection criterion rather than conducting a Chi-square test to compare them is that, unlike the latter approach, it is not limited to models that are nested. As long as the data on which they are fitted are the same, models - no matter how drastically different they are - can be compared using the AIC. Given the large number of predictors that will be modeled on the same dataset, this characteristic of the AIC makes an ideal selection criterion.

4. Empirical Findings

4.1. Data

In order to identify unusual events which would indicate subtle disruptions in the economy, our analysis focuses on ad hoc announcements. Unlike earning announcements, ad hoc announcements are not published on a regular basis, but follow specific events such as management changes, layoffs or earning warnings. This implies that, if unusual events occur in the economy, they are likely to be reflected in ad hoc announcements. The corpus of texts we investigate comprises the 11902 regulated ad hoc announcements published by German companies in English between January 2004 and June 2011. This type of financial disclosure is an important source of information since listed companies are legally required by the Securities Trading Act of Germany to publish these disclosures in order to prevent illegal insider trading. Ad hoc announcements must also be signed by company executives, quality-checked by governmental bodies and released within a certain time limit. As such, they have shown a strong influence on financial markets and their official character has made them a popular choice in previous research (Hagenau et al., 2013; Muntermann & Guettler, 2007; Feuerriegel & Prendinger, 2016).

Since we want to link the narrative of financial disclosures to the state of the economy, we incorporate a binary variable describing whether or not the economy is in recession. We focus on the 2008 global economic crisis because it is the only major crisis that has occurred since the widespread use of the Internet made financial news readily available to all stakeholders, thereby making the 2008 economic crisis an ideal case study for our research. We draw upon data from the Center for Economic Policy Research and accordingly define the recession as starting in January 2008 and ending in Mai 2009.

4.2. Hypotheses

4.2.1. Predictive Power of Economic Indicators

The multivariate anomaly score computed with the help of the k-nearest-neighbor algorithm appears to have no predictive power with regard to the occurrence of economic crises. Table 3 presents the results of the empirical analysis conducted of this anomaly score. Note that the *naive model* corresponds to a logistic model that only contains a dummy variable for *month* as predictor and, as such, offers a baseline from which to compare the AIC of the other logistic models. We observe that the AIC of the naive model is smaller than that of the models including a multivariate anomaly score as predictor across the four values for k in the k nearest-neighbor algorithm. This means that the naive model is the model that offers the bets fit of the data; that is, adding a multivariate anomaly score as predictor does not improve the goodness of fit of the logistic model. Moreover the p-value of the multivariate anomaly score are not significant at the significance level 0.05. The data therefore suggest that the multivariate anomaly score is not related to the occurrence of economic crises.

Figure 2 presents the evolution of the index of the multivariate anomaly scores computed with four different values of k for the k nearest-neighbor algorithm over time. The plot indicates the reasons why the multivariate anomaly score lacks predictive power. The index of the multivariate anomaly scores increases between the year 2004 and 2006 and decreases between the years 2009 and 2011, indicating that the anomaly score is higher than average during the former economic peak and smaller than average during the latter. Moreover, the index is stable between 2006 and 2009, indicating that the anomaly score is close to its average during the pre-recession period of time. Taken together, these two observations, i. e. economic peaks characterized by both large and small anomaly scores and average scores during the pre-recession period of time, explain why the multivariate score fails to be modeled with a logistic regression to forecast the occurrence of economic crises: an increase in the multivariate anomaly score is not related to an increase or a decrease in the probability of being in the months leading to an economic crisis.

Table 3: Logistic Regression: Multivariate Anomaly Score

k	**p-value of anomaly score**	**AIC**
Naive Model	-	116.05
5	0.894	118.04
10	0.890	118.04
50	0.955	118.05
100	0.966	118.05

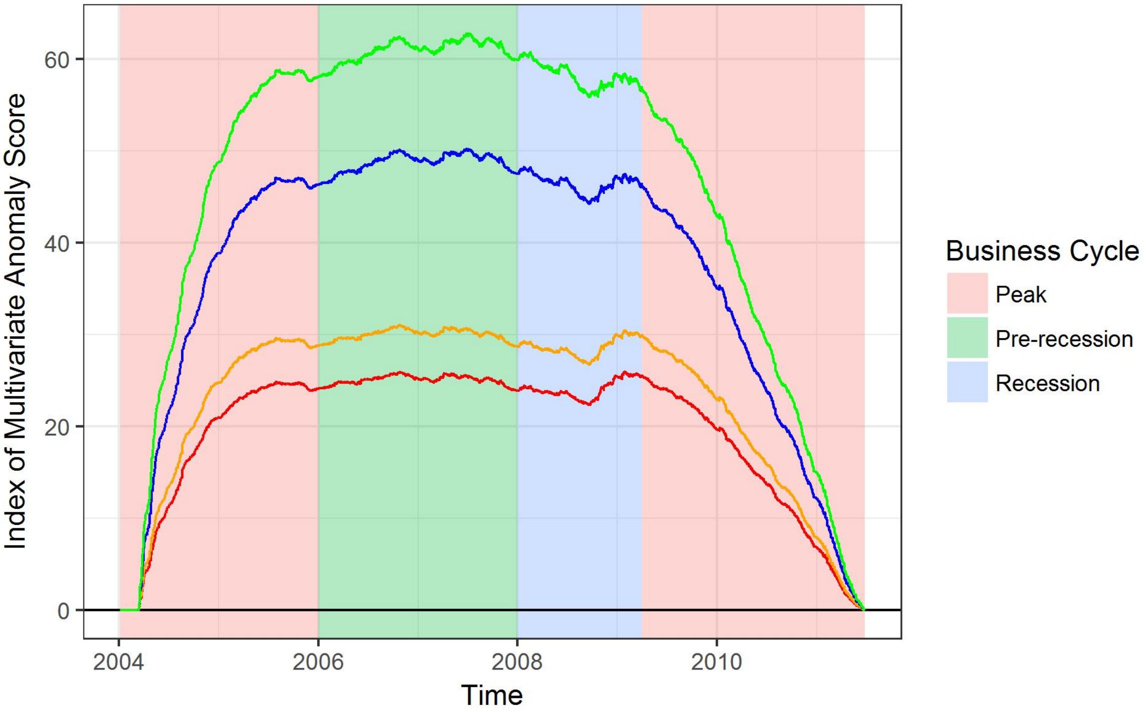

Figure 2: Index of Multivariate Anomaly Score ($k = 5$ [red], $k = 10$ [orange], $k = 50$ [blue] and $k = 100$ [green])

4.2.2. Predictive Power of Sentiment Scores

The positive sentiment anomaly score computed with the help of the seasonal-hybrid ESD algorithm appears to be a good predictor of the occurrence of economic crises, while the negative sentiment anomaly score seems to have no predictive power. Table 4 presents the results of the empirical analysis conducted on the two sentiment anomaly scores. We observe that the logistic model including the positive sentiment anomaly score as predictor is the model that offers the best fit of the data (AIC of 108.87) and that the logistic model including the negative sentiment anomaly score as predictor (AIC of 117.99) does not surpass the naive model (AIC of 116.05). Moreover, the p-value of the positive sentiment anomaly score is significant at the significance level 0.05 while that of the negative sentiment anomaly score is not significant. The data therefore shows that, unlike the negative sentiment anomaly score, the positive anomaly score can be modeled with a logistic regression model to forecast the occurrence of economic crises.

Figure 3 and Figure 4 indicate the direction of the relation between the positive sentiment anomaly score and the occurrence of economic crises. Figure 3 presents the evolution of the anomaly scores over time. We observe that the score peaks during the pre-recession period of time. This means that the 2008 economic crisis is preceded by large levels of unusualness and suggests that months with larger positive sentiment anomaly scores are more likely to be followed by an economic crisis withing the following two years. This observation is confirmed by the logistic regression (see Figure 4). The regression curve indicates that, the larger the positive sentiment anomaly score of a month, the larger the probability that the month in question occurs in the pre-recession period of time.

Table 4: Logistic Regression: Sentiment Anomaly Score

Anomaly score	**p-value of anomaly score**	**AIC**
Naive model	-	116.05
Positive sentiment anomaly score	0.00558	108.87
Negative sentiment anomaly score	0.799	117.99

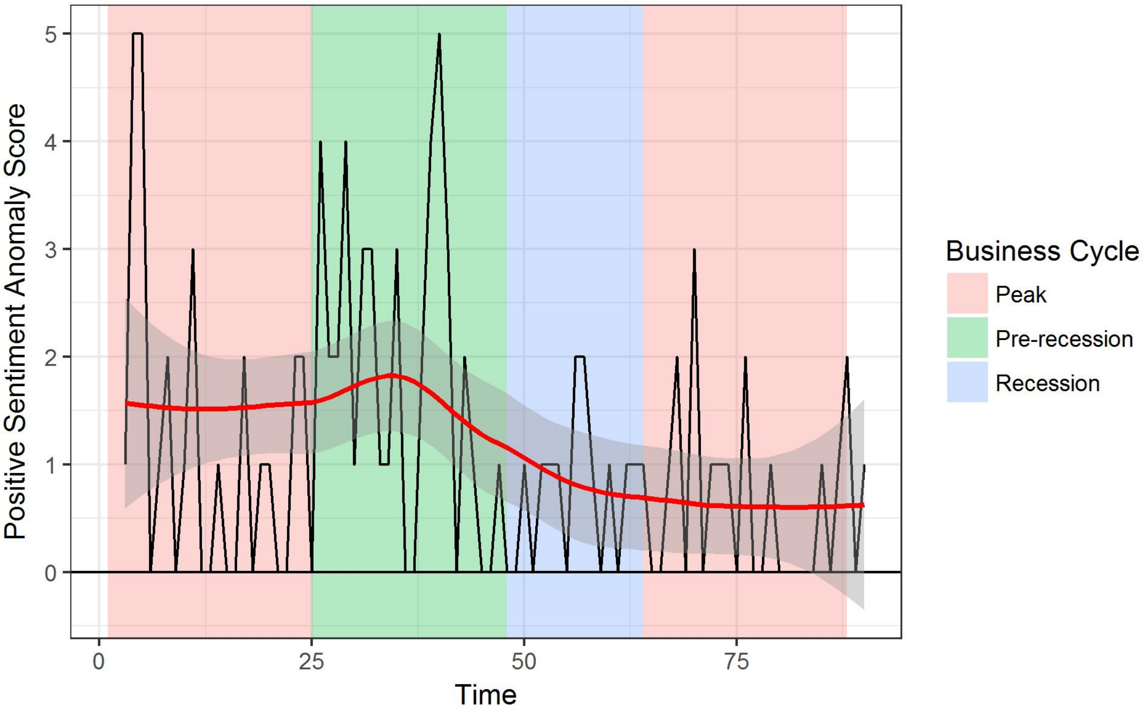

Figure 3: Evolution of the Positive Sentiment Anomaly Score over Time

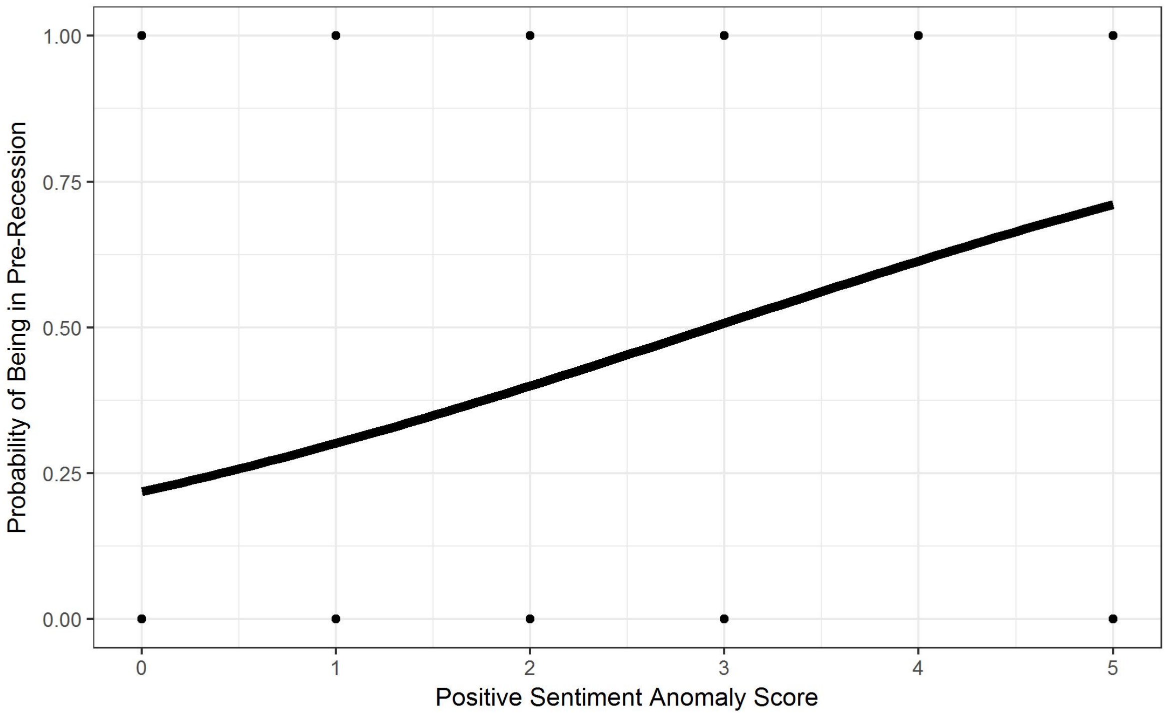

Figure 4: Logistic Regression: Positive Sentiment Anomaly Score

4.2.3. Predictive Power of the Term-Document Matrix

The term-document matrix anomaly score computed with the help of the k-nearest-neighbor algorithm appears to have a strong predictive power with regard to the occurrence of economic crises. The results of the empirical analysis indicates that the naive model is the model that offers the worst fit of the data (see Table 5). The logistic models including a term-document matrix anomaly score as predictor have a smaller AIC than that of the naive model across the fours values for k in the k nearest-neighbor algorithm. We also observe that the AIC increases as the value of k increases, suggesting that term-document matrix anomaly score computed with larger values for k have more predictive power. Moreover, the p-value of the term-document matrix anomaly scores are significant at the significance level 0.05. The data therefore shows that the term-document matrix anomaly score can be modeled with a logistic regression to forecast the occurrence of economic crises.

Figure 5 and Figure 6 indicate the direction of the relation between the term-document matrix anomaly score and the occurrence of economic crises. Figure 5 presents the evolution over time of the index of the term-document matrix anomaly scores computed with four different values of k for the k nearest-neighbor algorithm. We observe that the index increases during the two economic peaks and decreases during the pre-recession period of time. This means that economic booms are characterized by levels of unusualness that are larger than average and the months leadings to the economic crisis of 2008 by levels smaller than usual.

This pattern is similar across all four levels of k. This implies that months with larger term-document matrix anomaly scores are more likely to occur during economic booms and those with a smaller score are more likely to occur in the 2 years leading to the crisis. This observation is confirmed by the logistic regression (see Figure 6). The regression curve indicates that, the larger the term-document matrix anomaly score of a month, the smaller the probability that the month in question occurs in the pre-recession period of time.

Table 5: Logistic Regression: Term-Document Matrix Anomaly Score

k	**p-value of anomaly score**	**AIC**
Naive model	-	116.05
5	0.000187	92.973
10	9.28e-05	88.776
50	5.07e-05	83.004
100	4.34e-05	80.633

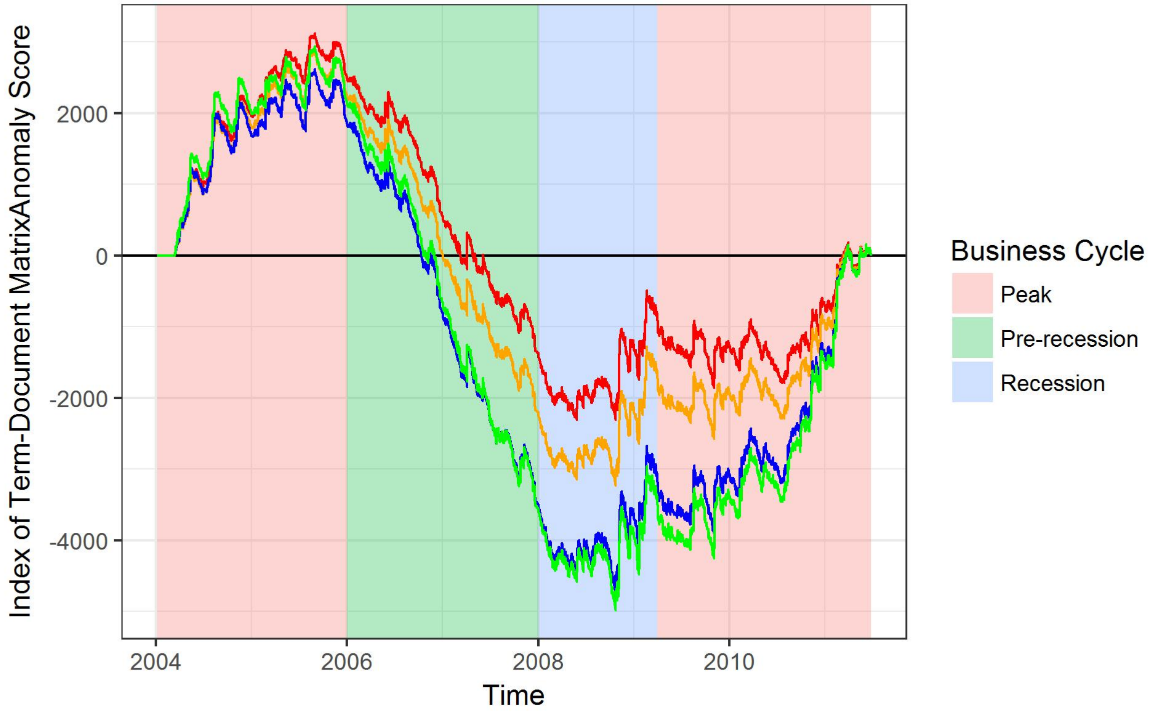

Figure 5: Index of Term-Document Matrix Anomaly Score ($k = 5$ [red], $k = 10$ [orange], $k = 50$ [blue] and $k = 100$ [green])

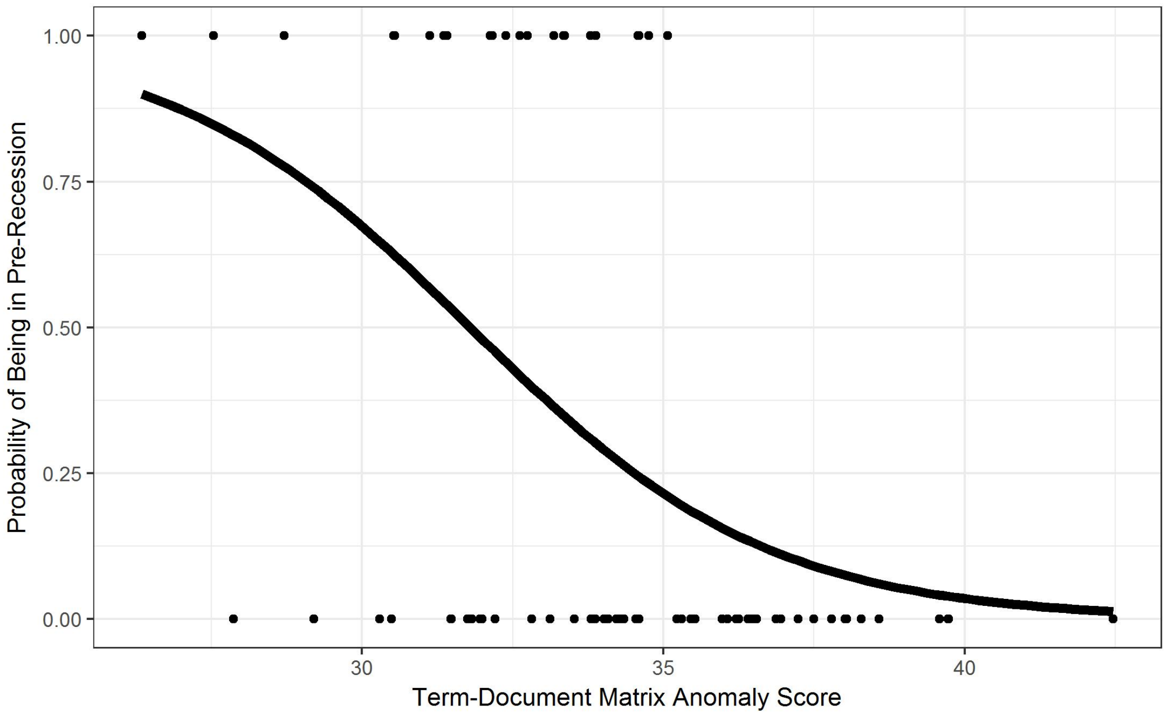

Figure 6: Logistic Regression: Term-Document Matrix Anomaly Score ($k = 100$)

4.3. Comparison

The previous empirical analysis shows that the three anomaly scores, i. e. the multivariate anomaly score, the sentiment anomaly score, and the term-document matrix anomaly score, have different levels of forecasting power. Namely, logistic models including a multivariate anomaly score as a predictor are outperformed, AIC-wise, by the naive model across the four values of k in the k nearest-neighbor algorithm. This suggests that the multivariate anomaly scores we have computed possess little to no predictive power with regard to the occurrence of economic crises. On the other hand, the term-document matrix anomaly score appears to be a good predictor of the future occurrence of economic crises. The AIC of the logistic model that includes the term-document matrix anomaly score computed with $k = 100$ in the k nearest-neighbor algorithm is 80.6338, compare to an AIC of 116.05 for the naive model. Finally, the sentiment anomaly score offers mitigated results. The positive sentiment anomaly score appears to have a good forecasting power – although less good than that of the term-document matrix anomaly score – and the negative sentiment anomaly score appears to have no predicting power.

These results indicate that anomaly levels computed from macroeconomic variables have less explanatory power than those computed from financial news, confirming the claim by Tetlock (2007); Tetlock et al. (2008) that soft content contains information about the economy that is not necessarily reflected in common macroeconomic indicators. It furthermore suggests that term-document matrices are valuable material to analyze textual data in the domain of finance and that conducting an anomaly analysis of financial news can help forecast economic crises.

4.4. Robustness Check

To check the robustness of our findings, we run the model with anomaly scores computed by the k-nearest-neighbor algorithm with various values of k. We use the following values for k: $5, 10, 50, 100$. Table 3 and Table 5 indicate that our results are consistent across all four values of k.

In addition, we also computed a term-document matrix including all terms present in the corpus with a sparsity level inferior to 0.99 (see Table 1). We apply the k nearest-neighbor algorithm to this object in order to obtain a new series of term-document matrix anomaly scores that take a larger number of terms into account. Table 6 presents the empirical findings. We observe that these term-document matrix anomaly scores behave analogously to those computed from the 0.9 term-document matrix: they are good predictors across the four values for k in the k nearest-neighbor algorithm and the AIC decreases as the value of k increases. In fact, the term-document matrix anomaly score computed with $k = 100$ on the term-document matrix with sparsity 0.99 is a slightly better predictor (AIC of 79.8) than term-document matrix anomaly score computed with $k = 100$ on the term-document matrix with sparsity 0.9 (AIC of 80.6338). We could unfortunately not consider term-document matrices including a larger number of terms due to the limited computational power available to us.

Table 6: Logistic Regression: Term-Document Matrix Anomaly Score with Maximum Sparsity Level of 0.9

k	**p-value of anomaly score**	**AIC**
Naive model	-	116.05
5	0.000219	90.635
10	0.000195	87.302
50	0.000170	82.02
100	0.000163	79.8

5. Discussion

5.1. Implications to Research

Term-documents matrices have never been used in sentiment analysis before. This makes our approach innovative, hence perfectible in many ways. This section highlights the most important limitations of our analysis and how researchers could overcome them in the future. Firstly, due to time constraints, this paper considers only two anomaly detection algorithms,

i. e. the k-nearest-neighbor anomaly detection algorithm and the seasonal-hybrid ESD algorithm. Chandola et al. (2009) explain that there exists a myriad of anomaly detection algorithms, each capable of discovering a specific type of anomalous data. As the analysis we conducted shows, different algorithms produce different results: the pre-recession period seems to be characterized by an increase in unusualness as measured by the seasonal-hybrid ESD algorithm applied to the positive sentiment anomaly score but by a decrease in unusualness as measured by the k-nearest-neighbor algorithm applied to the term-document matrix anomaly score. Since this paper only considers two algorithms, we may have missed some important patterns in the data that other anomaly detection algorithms could have detected. Future research should thus consider additional anomaly detection algorithms. This will yield a more refined appreciation of the nature of outliers in financial news in the context of economic crisis prediction and will provide new insight concerning the relation between the level of unusualness in soft content and future financial instability.

Secondly, our analysis suffers from the limited computational capacities we had at our disposition. We could not compute term-document matrices with maximum sparsity levels superior to 0.99. Such matrices would have contained a number of elements that our computers could not handle. In addition, we could not compute anomaly scores with the LOF algorithm taking more than 5 neighboring points into consideration (see Appendix Appendix A). Would we have had more powerful computers at our disposition, we could have computed anomaly scores from term-document matrices that contain more terms and we could have considered an additional anomaly detection method, i. e. the LOF algorithm, in our analysis.

Thirdly, researchers could use additional measures of financial instability. A binary variable indicating the occurrence of an economic crisis is a crude representation of financial instability that does not do justice to the complexity of risk analysis in economics. Using a continuous measure of risk such as the *Volatility Index* (VIX) would allow for a more refines analysis of the relation between the level of anomaly in financial news and economic instability. In the same line, further research may also consider additional types of textual data, for instance messages posted on social media such as tweets on Twitter.

Finally, the main limitation of this study is that it only considers the 2008 financial crisis. The fact that the empirical data we use in this paper are limited to one specific economic crisis may constraint the generalizability of our method. The 2008 economic crisis was characterized by its far reaching impact on society and its worldwide reach. Whether our findings are relevant to smaller, more regional crises should be investigated in a further study.

5.2. Implications to Management and Policy-Makers

In this paper, we attempted to develop a method whose characteristics make its generalizable to several economic sectors. The method we propose is a very general approach to measuring risk. It does not make assumptions regarding the form of outliers: what defines an outlier in this context need not be specified in advance. This makes the system flexible. This feature is particularly useful for the study of economic crises since all economic crises are different, yet all are characterized by a severe disruption of economic fundamentals; but it is also useful to the modeling of risk in other economic sectors. Kremer et al. (April 2013), for instance, propose to conduct a sentiment analysis on textual data to improve banks' credit-risk assessment, portfolio-management system, early-warning systems for SMEs and the monitoring of industry trends.

Furthermore, these findings also bear significant implication to firm management where the overall economic outlook is a crucial determinant of production and to economic policy making where financial instability can have far reaching consequences on society. Yet, given my limited knowledge of the two disciplines, I let to the professionals thereof the development of managerial strategies and economic policies to be adopted in case our system forecasts an economic crisis in the near future.

6. Conclusion

Financial instability forecasting is in dire need of improvements. Common wisdom among policy makers is to base forecasts of economic stability on time series analysis of macroeconomic indicators. This research takes an innovative approach and conducts an anomaly analysis on financials news in an effort to forecast economic instability. We find that soft content can be modeled to reliably predict the occurrence of the 2008 economic crisis. In particular, we show that an anomaly analysis conducted on financial news provides a forecast of economic instability up to 24 months ahead. Importantly, the anomaly analysis conducted on the term-document matrix of the corpus of financial news offers more reliable predictions for economic crises than an anomaly analysis conducted on sentiment scores, suggesting that term-document matrices are valuable material to analyze textual data in the domain of finance. Furthermore, our findings show that anomaly levels computed from macroeconomic variables have less explanatory power than those computed from financial news, confirming the claim by Tetlock (2007); Tetlock et al. (2008) that soft content contains information about the future economic outlook that is not necessarily reflected in common macroeconomic indicators and extending these results to the forecasting of financial instability.

References

Akaike, H., Parzen, E., Tanabe, K., & Kitagawa, G. (1998). *Selected papers of Hirotugu Akaike*. Springer series in statistics. Perspectives in statistics. New York: Springer.

Allen, F., & Gale, D. (2009). *Understanding financial crises* volume 1 of *Clarendon lectures in finance*. Oxford and New York: Oxford University Press.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, *59*, 1259–1294. doi:10.1111/j.1540-6261.2004.00662.x.

Bernanke, B. (1983). *Non-Monetary Effects of the Financial Crisis in the Propagation of the Great Depression*. Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w1054.

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is nearest neighbor meaningful? In G. Goos, J. Hartmanis, J. van Leeuwen, C. Beeri, & P. Buneman (Eds.), *Database Theory — ICDT'99* (pp. 217–235). Berlin, Heidelberg: Springer Berlin Heidelberg volume 1540 of *Lecture Notes in Computer Science*. doi:10.1007/3-540-49257-7{\textunderscore}15.

Bodnaruk, A., Loughran, T., & McDonald, B. (2015). Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis*, *50*, 623–646. doi:10.1017/S0022109015000411.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*, 1–8. doi:10.1016/j.jocs.2010.12.007.

Borio, C. E. V., & Lowe, P. W. (2002). Asset prices, financial and monetary stability: Exploring the nexus. *SSRN Electronic Journal*, . doi:10.2139/ssrn.846305.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* volume 1 of *Wiley series in probability ans statistics*. (Fifth edition ed.). Hoboken, New Jersey: Wiley.

Braggion, F., Christiano, L., & Roldos, J. (2007). *Optimal Monetary Policy in a 'Sudden Stop'*. Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w13254.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection. *ACM Computing Surveys*, *41*, 1–58. doi:10.1145/1541880.1541882.

Chang, S.-S., Stuckler, D., Yip, P., & Gunnell, D. (2013). Impact of 2008 global economic crisis on suicide: Time trend study in 54 countries. *Bmj*, *347*, f5239. doi:10.1136/bmj.f5239.

Christiano, L., Gust, C., & Roldos, J. (2002). *Monetary Policy in a Financial Crisis*. Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w9005.

Crnic, J. (2011). Introduction to modern information retrieval20112g.g. chowdhury. introduction to modern information retrieval . london: Facet publishing, isbn: 1–85604–480–7 3rd edition. *Library Management*, *32*, 373–374. doi:10.1108/01435121111132365.

Cutler, D. M., Knaul, F., Lozano, R., Méndez, O., & Zurita, B. (2002). Financial crisis, health outcomes and ageing: Mexico in the 1980s and 1990s. *Journal of Public Economics*, *84*, 279–303. doi:10.1016/S0047-2727(01)00127-X.

Estrella, A., & Mishkin, F. S. (1998). Predicting u.s. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, *80*, 45–61. doi:10.1162/003465398557320.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*, 82. doi:10.1145/2436256.2436274.

Feuerriegel, S., & Prendinger, H. (2016). News-based trading strategies. *Decision Support Systems*, *90*, 65–74. doi:10.1016/j.dss.2016.06.020.

Fodo, I. K. (). A survey of dimension reduction techniques. URL: https://e-reports-ext.llnl.

gov/pdf/240921.pdf.

Frankel, J., & Saravelos, G. (2012). Can leading indicators assess country vulnerability? evidence from the 2008–09 global financial crisis. *Journal of International Economics*, *87*, 216–231. doi:10.1016/j.jinteco.2011.12.009.

Greenspan, A. (2013). Never saw it coming. *Foreign Affairs*, *92*, 88–96.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, *11*, 1–21. doi:10.1080/00401706.1969.10490657.

Hagenau, M., Liebmann, M., & Neumann, D. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, *55*, 685–697. doi:10.1016/j.dss.2013.02.006.

Hanley, K. W. (2016). Dynamic interpretation of emerging systemic risks. *SSRN Electronic Journal*, . doi:10.2139/ssrn.2792943.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* volume 1 of *Wiley series in probability and statistics*. (Third edition ed.). Hoboken, US: J. Wiley.

Kremer, A., Malskorn, W., & Strobel, F. (April 2013). Ratings revisited: Textual analysis for better risk management. *McKinsey&Company*, .

Lewis, D. D., Yang, Y., Rose, T. G., & Li, F. (2004). Rcv1: A new benchmark collection for text categorization research. *Journal of machine learning research*, *5*, 361–397.

Lo Duca, M., & Peltonen, T. A. (2013). Assessing systemic risks and predicting systemic events. *Journal of Banking & Finance*, *37*, 2183–2195. doi:10.1016/j.jbankfin.2012.06.010.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*, 35–65. doi:10.1111/j.1540-6261.2010.01625.x.

Loughran, T. I. M., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*, 1187–1230. doi:10.1111/1475-679X.12123.

Mamaysky, H., & Glasserman, P. (2016). Does unusual news forecast market stress? *The Office of Financial Research Working Paper*, *4*, 1–75.

Manning, C. D., & Schuetze, H. (1999). *Foundations of statistical natural language processing* volume 1. Cambridge and London: MIT Press.

Muntermann, J., & Guettler, A. (2007). Intraday stock price effects of ad hoc disclosures: The german case. *Journal of International Financial Markets, Institutions and Money*, *17*, 1–24. doi:10.1016/j.intfin.2005.08.003.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundation and Trends in Information Retrieval*, *2*, 1–135. doi:10.1561/1500000011.

Perkins, D. H., Radelet, S., & Lindauer, D. L. (2013). *Economics of development* volume 1. (7th ed.). New York: Norton.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, *14*, 130–137. doi:10.1108/eb046814.

Romer, C. (1988). *The Great Crash and the Onset of the Great Depression*. Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w2639.

Rosner, B. (1983). Percentage points for a generalized esd many-outlier procedure. *Technometrics*, *25*, 165–172. doi:10.1080/00401706.1983.10487848.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*, 513–523. doi:10.1016/0306-4573(88)90021-0.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*, 1139–1168. doi:10.1111/j.1540-6261.2007.01232.x.

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*, 1437–1467. doi:10.1111/j.1540-6261.2008.01362.x.

Appendix A. Local Outlier Factor Anomaly Detection Algorithm

In this study, we applied another outlier detection algorithm to the term-document matrix: the *local outlier factor* anomaly detection algorithm. The local outlier factor algorithm is an extension of the k-nearest-neighbor algorithm that not only takes into account the nearness of the neighbors of a given data point, but also the density of the region surrounding those neighboring points (Chandola et al., 2009). We did not include this algorithm in our analysis because we could only run it with $k = 5$ due to computational constraints; we could not compute anomaly scores for larger values of k.

Table A.7 presents the results of the empirical analysis. Unlike the anomaly scores computed via the k-nearest-neighbor algorithm, those computed with the local outlier factor algorithm do not appear to be statistically significant when applied to term-document matrices. Furthermore, the AIC of their corresponding logistic models are larger than that of the logistic models including the anomaly scores computed via the k-nearest-neighbor algorithm. It is worth noting that the multivariate score computed with the local outlier factor algorithm appears to be a good predictor, unlike that computed with the k-nearest-neighbor algorithm. This confirms Chandola et al. (2009)'s claim that different types of data require different algorithms and shows that the choice of the anomaly detection technique is crucial for the analysis and should be carefully considered.

Table A.7: Logistic Regression: Local Outlier Factor Anomaly Detection Algorithm ($k = 5$)

Input of local outlier factor algorithm	**p-value of anomaly score**	**AIC**
Naive model	-	116.05
Term-document matrix (sparsity < 0.9)	0.980	118.05
Term-document matrix (sparsity < 0.99)	0.134	115.7
Set of 4 variables (see Section 3.4.1)	0.00614	108.92

Appendix B. Additional Levels of Granularity: Daily Level and Document Level

We did not only conduct the analysis at the granularity level of month, but we also conducted it at the level of days and of the individual documents themselves. The results we obtain at the level of days and documents are similar to those obtained at the month level, confirming the hypothesis that unlike multivariate anomaly scores and negative sentiment anomaly scores, which have no predictive power, the term-document matrix anomaly score can be modeled to forecast the occurrence of economic crises (see Tables B.8 to B.13).

Appendix B.1. Day Level

Table B.8: Logistic Regression: Multivariate Anomaly Score

k	**p-value of anomaly score**	**AIC**
Naive model	-	2037.3
5	0.7594	2039.2
10	0.7987	2039.2
50	0.5049	2038.8
100	0.4770	2038.8

Table B.9: Logistic Regression: Term-Document Matrix Anomaly Score

k	**p-value of anomaly score**	**AIC**
Naive model	-	2037.3
5	0.000596	2025.6
10	0.00011	2021.9
50	9.19e-06	2016.3
100	2.74e-06	2013.5

Table B.10: Logistic Regression: Local Outlier Factor Algorithm

Input of local outlier factor algorithm	**p-value of anomaly score**	**AIC**
Naive model	-	2037.3
Term-document matrix	0.9944	2039.3
Set of 4 variables	0.00348	2030.7

Appendix B.2. Document Level

Table B.11: Logistic Regression: Multivariate Anomaly Score

k	**p-value of anomaly score**	**AIC**
Naive model	-	$11,502$
5	0.8012	$11,503$
10	0.8209	$11,503$
50	0.5954	$11,503$
100	0.5987	$11,503$

Table B.12: Logistic Regression: Term-Document Matrix Anomaly Score

k	**p-value of anomaly score**	**AIC**
Naive model	-	$11,502$
5	2.04e-07	$11,473$
10	1.15e-08	$11,466$
50	2.13e-10	$11,457$
100	2.79e-11	$11,452$

Table B.13: Logistic Regression: Local Outlier Factor Algorithm

Input of LOF algorithm	**p-value of anomaly score**	**AIC**
Naive model	-	$11,502$
Term-document matrix	0.5367	$11,503$
Set of 4 variables	0.000228	$11,490$