UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ADMINISTRAÇÃO, CONTABILIDADE E ECONOMIA
DEPARTAMENTO DE ECONOMIA

Textual Analysis in Finance

Camila Cardoso Pereira

Brasília
2019

Camila Cardoso Pereira

Textual Analysis in Finance

Tese de doutorado submetida ao Departamento de Economia da Universidade de Brasília como requisito parcial para a obtenção do grau de Doutor em Economia.

Orientador: Daniel Oliveira Cajueiro

Brasília
2019

Camila Cardoso Pereira

Textual Analysis in Finance

Tese de doutorado submetida ao Departamento de Economia da Universidade de Brasília como requisito parcial para a obtenção do grau de Doutor em Economia.

Trabalho aprovado. Brasília, 10 de junho de 2019:

Banca examinadora:

Prof. Daniel Oliveira Cajueiro, PhD
Universidade de Brasília (UnB)
Orientador

Profª. Marina Delmondes de Carvalho Rossi, PhD
Universidade de Brasília (UnB)

Prof. Herbert Kimura, PhD
Universidade de Brasília (UnB)

Prof. Regis Augusto Ely, PhD
Universidade Federal de Pelotas (UFPel)

Prof. José Guilherme de Lara Resende, PhD
Universidade de Brasília (UnB)

Resumo

Esta tese é composta por três estudos que têm como objetivo estudar o impacto da mídia escrita no mercado acionário. No primeiro estudo, fazemos uma pesquisa acerca dos trabalhos que utilizam análise textual para quantificar variáveis econômicas e resumimos os principais resultados dos estudos que investigam seu impacto no mercado acionário. Como o uso de textos como dados em pesquisas científicas é um campo que está em crescimento, este estudo tem como objetivo sintetizar os principais resultados para delinear onde está a fronteira do conhecimento na literatura de finanças. Os dois estudos restantes investigam a relação entre duas variáveis estimadas a partir de notícias e o mercado acionário brasileiro. Assim, no segundo estudo que compõe esta tese estudamos o impacto da incerteza econômica nos retornos acionários semanais. Neste estudo, propomos um novo método para estimar incerteza econômica a partir de notícias usando vetores de palavras para representar o vocabulário. Encontramos um efeito significativo da nossa medida de incerteza econômica na precificação das ações e mostramos que medidas de incerteza propostas na literatura mensuradas a partir de notícias geram um efeito similar. No terceiro estudo, estimamos corrupção a partir de notícias e analisamos sua relação com o desempenho de ações de duas empresas que estiveram envolvidas em escândalos de corrupção nos últimos anos. Este estudo tem como objetivo quantificar o custo da corrupção para essas empresas. O impacto da corrupção abordada nas notícias nos retornos acionários divergem entre as empresas. No caso em que a empresa possui controle privado, a corrupção nas notícias impactam negativamente os retornos acionários. Para o caso em que a empresa possui controle estatal, o efeito é insignificante. Encontramos, ainda, um efeito de longo prazo dos escândalos de corrupção nos preços das ações.

Palavras-chave: análise textual; incerteza econômica; retorno acionário; retorno anormal; corrupção; controle sintético.

Abstract

This thesis is composed of three studies that aim to investigate the impact of written media on stock performance. In the first study, we make a survey of the literature that uses textual analysis to quantify economic variables and review the main results of the studies that examine its effect on the stock market. Since the use of texts as data in scientific research is a growing field, this study aims to summarize the main findings to draw where the frontier knowledge in finance literature is. The remaining two studies investigate the relation between two variables estimated from news stories and the Brazilian stock market. Thereby, in the second study, we investigate the impact of economic uncertainty on weekly stock returns. We propose a new method to estimate economic uncertainty from news stories using word vectors for word representation. We find that there is a significant effect of our economic uncertainty measure on pricing individual stocks and provide similar evidence with uncertainty measures from news stories proposed in the literature. In the third study, we estimate corruption from news stories and investigate its relation to the stock performance of two firms that were involved in corruption scandals in the latest years with the primary goal of estimating the cost of corruption for the firms. The impact of the corruption reported in the news stories on the stock returns diverges between companies. In the case the company has private ownership, corruption in news negatively impacts stock returns. For the state-owned company, the effect is insignificant. We also find a long-term effect of the corruption scandals in the stock prices.

Keywords: textual analysis; economic uncertainty; stock returns; abnormal returns; corruption; synthetic control.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Individuals use text to get all type of information for different purposes. In the finance context, we can mention, for example, corporate filings, analysts reports, social networks and newspapers, all relevant data for the decision-making process. The use of natural language processes applied to finance is still a growing literature, less explored than quantitative variables. Undoubtedly, the use of texts is not as simple as the use of numbers as data because we do not have, until now, an algorithm that accurately simulates the human interpretation of a document. It is a hard task even if we consider a median reader, since people may interpret a text differently depending on their cognitive processes. Furthermore, documents may be compounded by ambiguous sentences, terms that have a different meaning in multiple contexts and proper names mimicking common names. Thus, it is still a challenge to find an algorithm that gets close to human reading and thereby quantify written content automatically.

In light of that, the primary goal of this thesis is to estimate variables with economic consequences from written public information using textual analysis, which is a methodology to interpret the content of documents. The methods to process texts is an emerging area in the finance literature and became popular in the last years with the increase in computing power that allows us to process a higher amount of data. Also, written materials are substantial fonts of information, which give us the opportunity to study valuable unexplored data. This study focuses on searching for new estimations for variables that investors consider in their investment decision and which is widely disseminated information in media. Therefore, we may capture in our estimations a media component which the quantitative variables do not reflect.

This thesis is divided into two empirical studies, each one with a specific purpose involving measuring variables that do not appear in the literature that uses texts to obtain it or proposing a new algorithm to estimate it. Additionally, we develop a survey of the relevant studies in finance involving textual analysis with the objective to delineate the issues that are still not investigated in the literature.

Chapter two presents the survey of the primary methods and findings in the finance literature that uses texts in the quantitative analyses. The purpose of this chapter is to resume what topics appear in the literature and what are the main questions that lack answers.

Chapter three reports the first empirical study in which we propose a new measure for economic uncertainty estimated from public information. We aim to establish the relation between the tone of economic uncertainty in the news stories with individual stock returns. We also show how our measure is related to alternative uncertainty measures established in the literature to examine the differences and similarities between them. In sum, we show in this chapter whether we can measure economic uncertainty from media studying whether there is an effect on investor behavior and whether there is a component in economic uncertainty the media reports that known risk factors do not reflect.

Chapter four reports the second empirical study that comprises this thesis. In this chapter, we investigate how the tone in the news stories of corruption involving specific firms is related to their valuation. The purpose of this study is to establish the cost for a firm as the result for the decision to get in corruption schemes in the case the schemes become public. In sum, we investigate investor behavior when news stories report a corruption scheme and estimate the effect in the long-term of those corruption scandals on firm value using an econometric method to estimate the value the firm would have in the absence of the scandal.

Chapter 2

Textual Analysis in Finance: An Introduction

2.1. Introduction

In this chapter, we analyze the primary tools in textual analysis the finance literature explore and identify the main topics addressed in these studies. Textual analysis is an approach to analyze the content of natural language in texts with the purpose to obtain structured data. Without computational support, texts can be useful fonts of information if we have a small sample of documents to analyze. Nevertheless, when the amount of information begins to grow in a level that human cognitive capability does not assimilate it, algorithms that interpret automatically these texts and return simplified information are required.

The natural language processing still has many difficulties we have to handle. Even for a human reader sometimes doubts emerge about the meaning of a slice of a text. The most advanced algorithms do not completely solve this problem until now. Therefore the existing methods can be efficient in generalizing through data and return effective output easy to read and manage. Since it is an emerging topic in several areas, the main goal of this chapter is to offer a brief explanation of the most used methods in the finance literature.

We stress that we focus in this survey on the methods that appear in the finance literature and do not consider the advances in other areas. Study the complexity of a text[1], the similarity between documents[2], the sentiment in public informations[3] and the

[1]See, for example, Li (2008); Miller (2010); Lawrence (2013); Lundholm, Rogo, and Zhang (2014); Lo, Ramos, and Rogo (2017).

[2]See, for example, Hoberg and Phillips (2010); Brown and Tucker (2011); Ibriyamova, Kogan, Salganik-Shoshan, and Stolin (2016); Hoberg and Phillips (2016, 2017); Box (2017).

[3]See, for example, Antweiler and Frank (2004); Tetlock (2007); Tetlock, Saar-Tsechansky, and Mac-

tone in documents related to a specific topic[4] are among the main features the literature explores to investigate their relation to financial variables. The following sections describe how the studies define those variables from texts using textual analysis tools.

2.2. Readability

When we use algorithms to interpret texts to study their impact on some economic variable, the first issue we assume is that people have full access to the information and consider it in the decision-making process. An implicit fact of this assumption is that the reader fully assimilates what the text intends to communicate. Although, this is not always true. There are measures that reflect the complexity of a text and hence classify the documents as more or less readable. The Fog index, for example, measures how readable a text is taking into account the number of words with more than two syllables and the average length of a sentence.

Measuring the complexity of a text, Miller (2010) finds that more complex annual reports are associated with a lower trading volume of small investors and Lundholm et al. (2014) evidence that foreign firms make effort to produce more readable reports to compensate the geographic distance. Also, Lawrence (2013) find that more complexity in the financial disclosures is associated with lower individual investments less financially-literate.

2.3. Similarity between Documents

When analyzing the content of a sample of documents, we can compare each one with each other and measure the similarity between them. The most popular way to measure it is to represent each document as a vector of word frequencies and then define the similarity as the cosine between them. More specifically, most authors create vectors of n coordinates, where each coordinate i represents the word i in the vocabulary of the entire sample. Therefore, the value of the coordinate i for a document is the frequency the word i appears in it.

To give examples of applications of this method, we can cite Brown and Tucker

skassy (2008); Li (2010); Loughran and McDonald (2011); García (2013); Jegadeesh and Wu (2013); Huang, Zang, and Zheng (2014); Liu and McConnell (2013); Ferguson, Philip, Lam, and Guo (2015); Agarwal, Chen, and Zhang (2016); Tsai, Lu, and Hung (2016); Bajo and Raimondo (2017); Fraiberger, Lee, Puy, and Rancire (2018).

[4]See, for example, Li (2006); Balvers, Gaski, and McDonald (2016); Audi, Loughran, and McDonald (2016); Karapandza (2016).

(2011), in which the authors show that firms with more economic changes do more modifications in their Management Discussion and Analysis disclosures (comparison of documents from the same firm through time) and the investors react to those modifications. Another example is from Ibriyamova et al. (2016). The authors apply the so-called semantic fingerprinting method and show that higher correlation between stock returns is associated with a higher similarity between the firm descriptions (comparison of same documents from different firms).

2.4. Bag of Words Hyphotesis

The bag of words hypothesis is the assumption of independence between words. This hypothesis simplify the analyses and facilitate the estimations since it ignores the order of the words in a document. Although it is a strong and naive hypothesis, there are interesting results in the finance literature, which we mention below. The disadvantage of assuming this hypothesis is that it considers that the position of a word in a document is not important and hence we do not have any information about the context the word is inserted.

2.5. Word Lists

Using the bag of words hypothesis, many authors quantify data from texts using word lists when the interest is in quantifying the texts according to a characteristic. Also called dictionaries, word lists are lists containing terms related to an attribute. Therefore, word lists are used to select words previously classified as having an intrinsically meaning to consider their occurrence in a text. Essentially, if a text contains a higher frequency of terms related to an attribute relative to other documents, we assign a higher score for its tone related to the attribute.

Two well-known lists are the Harvard-IV-4 psychosocial dictionary[5] and the Loughran-McDonald master dictionary[6]. The first is a dictionary that classifies words according to their meaning in the psychology area. For example, Tetlock (2007) and Tetlock et al. (2008) use the positive and negative categories to measure the bearish and bullish of the market. The second word list is proposed by Loughran and McDonald (2011). The authors propose a dictionary built with financial terms for financial texts analyses, since a word in psychology or in another area may have a different meaning in the financial

[5] Available in `http://www.wjh.harvard.edu/$\sim$inquirer/homecat.htm`.
[6] Available in `https://sraf.nd.edu/textual-analysis/resources/`.

context. They show that considering simply word frequencies, their dictionary outperformance the Harvard-IV-4 psychosocial dictionary. Since its release, the finance literature widely explores their dictionary (García, 2013; Liu and McConnell, 2013; Ferguson et al., 2015; Agarwal et al., 2016; Tsai et al., 2016; Bajo and Raimondo, 2017; Fraiberger et al., 2018).

Some authors define word lists with terms related to a specific topic they are exploring to analyze their occurrence in a document. For example, examining corporate annual reports, Audi et al. (2016) use a list of trust words to consider their occurrence and Karapandza (2016) considers the frequency of the verbs in the future tense to study its relation with stock returns. More examples are Li (2006) that count the frequency of "risk" and "uncertainty" words and their variants to measure the risk sentiment in corporate annual reports and Balvers et al. (2016) that consider the frequency of terms related to "customer satisfaction" in corporate annual reports to study its relation with American Customer Satisfaction Index score.

2.5.1. Term Weighting

A more sophisticated process using word lists consists of weighing words differently depending on its importance in the text instead of considering simply the frequency of some specific words. These methods are based on the idea that some words may be more relevant than others depending on the document collection they appear. A popular method to weight words differently to measure the tone of a text is proposed by Loughran and McDonald (2011) that use a weighting scheme known as tf-idf (term frequency-inverse document frequency). This method take into account besides the frequency the word appears in a document relative to the document length, the frequency the word appears in all collection and the document collection length, which assign the weight for a word in a specific document considering the importance of the word for the entire sample. The authors show that using this weighing scheme the performance of their financial dictionary is equivalent to the Harvard-IV-4 psychosocial dictionary. Alternatively, Fraiberger et al. (2018) offer a weighting scheme which also considers the occurrence of the words in a document related to its occurrence in the entire corpus to weigh words beside to consider weight equal 1 for each word to build a country sentiment index. We can mention also Jegadeesh and Wu (2013) that develop a weighting scheme based on the market reaction to corporate annual reports to define weights based on its importance for the corpus. These methods help to eliminate the bias of considering irrelevant words from a list from the point of view of the investors.

2.6. Naive Bayes Classifier

The machine learning methods became popular with advances in the computational area. These methods provide tools to automatically extract patterns from data. Goodfellow, Bengio, and Courville (2016) point out that "the introduction of machine learning allowed computers to tackle problems involving knowledge of the real world and make decisions that appear subjective." Thereby, machine learning can be practical when we aim to automatically extract pieces of information from texts and organize them as structured data.

The Naive Bayes classifier is a supervised machine learning method, which means that the learning process uses previously correctly classified documents to calculate the most likely category for the remaining documents. The Naive Bayes classifier is an application of the Bayes' Theorem and it gives as output the probabilities for each document to belong to each category in a set of categories we choose in advance. Although the method is more sophisticated than word lists usage, it still assumes the bag of words hypothesis and hence it ignores the context around each word since it considers individual words for the probabilities estimation.

Examples of the application of the Naive Bayes approach in the finance literature comprise to predict the pessimism and optimism of internet messages (Antweiler and Frank, 2004), the tone in analyst reports (Huang et al., 2014) and corporate filings (Li, 2010), among other applications (Antweiler and Frank, 2006; Das and Chen, 2007; Agarwal et al., 2016; Buehlmaier and Whited, 2018a).

2.7. Getting Into Context

The challenge in the textual analysis literature is at constructing a method that considers the context of a document. Most of the studies that explore the context of a text instead of considering the meaning of the isolated words uses deep learning models. Deep learning models are part of the machine learning methods and incorporate concepts inspired by the biological brain. Surely, still the scientists do not understand all connections a brain makes to process information, but the advances in this direction are the base for such models.

To incorporate the context in textual analysis, Mai, Tian, Lee, and Ma (2019) construct deep learning models, such as averaging embedding and convolutional neural network, to find patterns in textual disclosures to predict bankruptcy. For the same purpose, Barboza, Kimura, and Altman (2017) compare alternative machine learning methods in the estimation, also incorporating the context in the analyses. Mamaysky and Calomiris (2018) is another example of a study that estimates sentiment in news stories without assuming the bag of words hypothesis. The authors use a 4-grams, which is a contiguous sequence of four words, in the sentiment measure.

Other examples of studies that examine the role of sentiment in the news stories use the output from the neural network Thomson Reuters News Analytics provides to examine its impact on stock market movements (Smales, 2015; Hendershott, Livdan, and Schrhoff, 2015; Heston and Sinha, 2017; Sun, Najand, and Shen, 2016; Araújo, Eleutério, and Louçã, 2018). The output of this neural network is the sentiment relative to how positive or negative a news story is for each new story in the database. The database classifies a text at a sentence level rather than word level as the use of the word lists.

2.8. Conclusion

We present here the main methods the finance literature explore to study the relation between public information containing natural language and financial variables. The frontier in the finance literature that uses texts to extract quantitative data from them is at building an algorithm that analyzes the content of a document and incorporates the context in the output. The most advanced methods try to reproduce human interpretation of a text, but there is still a lot to evolve. Nonetheless, the increase in the complexity of these methods is remarkable in the latest years.

Chapter 3

Measuring Economic Uncertainty from Textual Analysis

3.1. Introduction

We examine the impact of economic uncertainty covered in the news in individual stock returns. We suggest a new estimation using word vectors for word representation to quantify economic uncertainty. We then compare our measure with two other well accepted measures in the literature. Our paper analyzes the role of economic uncertainty in predicting stock returns and examines if there is a component related to the media coverage that contributes to explain returns that is not associated with known risk factors.

Firms are affected by the economy they are inserted in, and hence expectancy about economic fundamentals must be a relevant determinant in asset prices. Economic uncertainty generates an unpredictable business scenario that companies are about to face and therefore affects companies decisions about investments in the present, which changes the current demand for their stocks (Merton, 1973). In this paper, we examine the role of economic uncertainty in pricing stocks and explore some alternatives to measure it from news stories using textual analysis. We argue that when we quantify economic uncertainty from news stories we may capture something related to investors behavior that others proxies based on fundamentals does not consider. As a consequence, the media may represent an important component in pricing stock returns in the short term.

Investors react to public information about issues that impact firms traded in the stock market, which makes media coverage essential for disseminating information. We are interested in finding a measure to quantify public documents content related to a specific topic, economic uncertainty, that may be not captured by quantitative information. Irrational investors can lead movements in the stock market, and since media can cause a sentiment in these investors, we hypothesize that quantifying news stories can give us subjective elements that are not considered in quantitative variables involved in the pricing of individual stocks. Some authors support our hypothesis showing that investor sentiment plays an important role in pricing stocks (Lee, Jiang, and Indro, 2002; Schmeling, 2009; Stambaugh, Yu, and Yuan, 2012). Furthermore, media frequently covers economic and political issues and thus, if our hypothesis is valid, the use of the media to quantify economic uncertainty can measure a sentiment component that may be not associated to economic fundamentals.

If we find a proper measure from news stories, it could capture either the new information incorporated by investors or the sentiment of the market involved in the news, both involved in the stock pricing. The use of the media as a source to build a quantitative variable is based on two possible situations. First, a news story can reflect the real situation of the economy, and even if the investors are only reacting to economic fundamentals or using the information to investment decisions, an appropriated measure of the media reflects what we are interested in estimating. In this case, we need to find a measure that fully capture the information reported. One advantage in using this source of information is that even if the media is not reporting all relevant information to pricing stocks, we can have attention focused to the information reported, which is evidenced to influence investors behavior (Huberman and Regev, 2001; Barber and Odean, 2008; Lou, 2014). Second, if the media is speculating in some news stories, investors behavior can change, either for a sentiment caused by the news or a misperception of the fundamentals. Since the two possibilities can drive movements in the stock market, we argue that the media is a relevant component to be explored in pricing stocks.

To investigate the relation between economic uncertainty and stock returns, we construct an economic uncertainty measure from news stories using word vectors for word representation. The methods to generate word vectors consist in mapping words in the vector space where similar words tend to be grouped nearby based on the context of a sample of documents. This method can be useful when one has documents in a specifical context and want to define the word vectors that represent the meaning of the words for that context. We collect over 300.000 news stories from Valor Econômico and Folha de São Paulo online, two popular sources of information in Brazil. We then select the sections related to economic and political issues to build a sample with the specific context we are interested in studying.

Our paper mainly contributes to the literature proposing a new method to estimate economic variables from news stories. Estimating media is practical if the variable we aim to estimate has more importance when it is known by investors than an estimate from economic fundamentals information poorly disseminating. In sum, we are more interested in capture the interpretation of investors regard to a specific situation than an interpretation of quantitative data of the same situation. Our method for estimating economic uncertainty provides evidence consistent with the literature, and it can be applied in other contexts where one is interested in quantifying a qualitative characteristic of a variable.

Alternatively, we build two measures used in the literature to estimate uncertainty using textual analysis. The first measure is the economic policy uncertainty of Baker, Bloom, and Davis (2016), which consider the appearance of words related to economic, uncertainty and policy makers to classify articles in economic policy uncertainty articles. The index is based on counting these articles in the newspapers. The second measure is the word count with a weighting scheme method as in Loughran and McDonald (2011) using their uncertainty dictionary. The authors propose to use weights for words in which higher weights are assigned to the more relevant words giving the documents in the sample.

For studying the effect of events related to economic uncertainty that impact the market as a whole, the Brazilian case had favorable conditions in the last years. We had significant variance in expectations generated by the political environment and financial distress. At the end of the period we analyze, these difficulties remain, which give us a considerable time window with substantial variation in uncertainty about the future. The impeachment of President Rousseff, the decision to freeze the government spending and changes in labor legislation are examples of critical events that directly impacted the stock market. These events were accompanied by mass protests and debates in Congress over months, which draw the attention of the market day after day. The attempts of Minister Joaquim Levy to stabilize the public finances is another example of a relevant situation in the period we analyze that lasts 11 months and, during this time, media daily covered each event or announcements. Situations like that make investors react to new informations, which are highly covered by media. These events make the period immediately before the event date more uncertain than the day of the decision per se, which make the discussion in the news stories in the previous days valuable information to study investors reaction.

Our main empirical results are as follows. We show that the economic uncertainty index we propose helps to predict contemporaneous stock returns and abnormal returns in a weekly analysis. We provide evidence that an increase in economic uncertainty measure makes stock prices to drop. Moreover, an increase in economic uncertainty predicts negative abnormal returns. This result is robust after controlling for firm specific characteristics, media coverage effect for stocks, public firms, and illiquidity. The effect is not related to small-cap stocks effect and remains along the sample. Also, the effect is not related to the political crisis or the recession period. The performance of the word count with a weighting scheme uncertainty measure generates similar results. Also, the weekly economic policy uncertainty measure of Baker et al. (2016) perform well in predicting stock returns and abnormal stock returns similar to other measures and its effect is concentrated in stocks mentioned in the news during the week.

3.2. Literature Review

Our paper is related to the literature that studies the impact of the uncertainty in the stock market. Several studies investigate the relation between uncertainty and stock market movements. Theoretical studies demonstrate a negative relation between uncertainty and stock returns in an equilibrium model (Pástor and Veronesi, 2012; Armstrong, Banerjee, and Corona, 2013). Empirical studies like Ozoguz (2009) confirm these results. Other empirical studies show different aspects of how uncertainty is related to stock returns that evidences the importance of uncertainty in market efficiency. For example, Bali, Brown, and Tang (2017) show that investors demand a higher premium to hold stocks with a negative effect due to uncertainty, Starks and Sun (2016) argue that in times with higher uncertainty flow-performance sensitivity decreases in mutual funds and Zhang (2006) evidences that the greater information uncertainty about firms, the greater reaction of the market following new informations.

Our paper is also related to the literature that aims to quantify economic uncertainty. Some authors uses quantitative variables as a proxy for uncertainty (Carrière-Swallow and Céspedes, 2013; Segal, Shaliastovich, and Yaron, 2015) or events that cause uncertainty shocks (Liu, Shu, and Wei, 2017). As in Baker et al. (2016) we get our measure from news stories. Media is an important element in disseminating information, and we have empirical evidence that it can influence investors decisions (Dougal, Engelberg, García, and Parsons, 2012; Gurun and Butler, 2012; Solomon, 2012; Peress, 2014).

Individuals dedicate limited intellectual resources to collect and process information optimizing their behavior (Sims, 1998; Peng and Xiong, 2006; Maćkowiak and Wiederholt, 2015). In this context, we may think that the media is a relevant source of information the investors use to make investment decisions. Since information is only valid to change investors behavior if it is disseminated, the media is not only important to measure economic uncertainty, but also to filter the most important information in impacting stocks performance. Empirical evidence that shows the investor inattention influences the market reaction to new information also supports this hypothesis (Cohen and Frazzini, 2008; Dellavigna and Pollet, 2009).

Our paper is also related to the growing literature of textual analysis. Several studies try to quantify qualitative data from news stories. Earlier studies quantify media tone based on word counting using dictionaries of negative and positive words (Tetlock, 2007; Tetlock et al., 2008; Davis, Piger, and Sedor, 2012; García, 2013; Liu and McConnell, 2013; Agarwal et al., 2016; Tsai et al., 2016) or a specific list of words related to the issue that it is tried to be quantified (Li, 2006; Balvers et al., 2016; Audi et al., 2016; Karapandza, 2016). Some authors use dictionaries with different weights for each word (Loughran and McDonald, 2011; Jegadeesh and Wu, 2013). Other papers, such as Antweiler and Frank (2004), Antweiler and Frank (2006), Li (2010), Huang et al. (2014) and Buehlmaier and Whited (2018b), use the Naive Bayes algorithm to classify texts. All these methods assume the bag-of-words hypothesis, which consider independence among words.

More sophisticated methods of textual analysis are being explored to try to get the meaning of words taking into account the context they are inserted in examining. Reuters database provides a sentiment for each news story constructed with a neural network, and some authors use it for different purposes. For example, to study the relation between sentiment in the news and stock returns (Heston and Sinha, 2017), gold futures market (Smales, 2015) and institutional trading (Hendershott et al., 2015). We propose a new approach to develop qualitative measures from texts incorporating context features, which consists in training vectors to represent words.

3.3. Data and Method

3.3.1. Sample and data sources

The stock sample includes all stocks listed on the BM&FBovespa. We select only liquid stocks for the analyses since illiquidity can lead to mispricing. We consider a stock to be liquid at day t if it presents at least one trade per day in 95% of trading days in the previous year. We still exclude stocks that reach a minimum closing price of 4 reais[7] in the analyzed period. After filtering the sample, our sample remains with 174 stocks. We collect financial data from Economática database, and the risk factors for Brazilian market come from data provided by São Paulo university[8]. For the analyses, we estimate the results in a weekly frequency, as in Heston and Sinha (2017), where the authors show that weekly news predict returns in a larger window than daily news. Since an emerging market as the Brazilian market can present more inefficiencies than a developed market, for example, a slower reaction to new information, we consider a weekly frequency analysis more appropriate.

Our news sample comes from two popular newspapers in Brazil, Valor Econômico, which is one of the most relevant font for investments, and Folha de São Paulo Online, a well known font of information in the country. We collect news stories from sections related to finance, business and politics of Valor Econômico[9] and sections related to international news, politics, finance, economy and investments of Folha de São Paulo Online[10], both from January 2012 through June 2018. In the analyses, we suppose a news story has an impact on the day it is published. So, if a story becomes known to investors in a day with the market closed or in a day after half an hour before market closure, we consider this story is published in the next trading day.

3.3.2. Parsing text and word representation

We consider each news story as a unique document. Before filtering texts, we create a dictionary with synonyms for each company in the sample. We then convert all letters to lowercase and take off accents, except those proper names in the dictionary that become ambiguous with the process when we look at the isolated word. Before parsing the text, we substitute each word related to a company in the dictionary by a unique word that represents the company. We then filter punctuation, links, and numbers, except percentage numbers, which we replace by $+[\%]$ ($-[\%]$) if it is a positive (negative) number. After replacements, we remove terms that occur less than five times in documents vocabulary, except if it is documented in the dictionary. This process is needed to remove very infrequent terms, which meaning is hard to detect.

To build an uncertainty measure from news stories, we use vectors for word repre-

[7] Real is the official currency of Brazil. One US dollar equaled 1,87 reais and 3,86 reais at the beginning and the end of the period, respectively. The Brazilian Real depreciation evidences the economically troubled period we are considering.

[8] The data are available in http://nefin.com.br/.

[9] *Finanças*, *Empresas* and *Política* sections in `https://www.valor.com.br/`

[10] *Mundo*, *Poder* and *Mercado* sections in `https://www.folha.uol.com.br/`

sentation. In this representation, similar words are mapped together in the vector space. Because it uses the entire sample to define the vectors, they have no time dependence. In our context, we have a set of uncertainty words that we want to relate with terms associated with the economy. After the training process, some uncertainty words will be closer to economic terms than others, which means that in our sample, those words are more used in economic terms context than the others. Hence, we assume that the closer uncertainty word vectors are to economy word vectors, more uncertainty is associated with economic scenario.

We use the algorithm GloVe for obtaining vectors. The algorithm defines vectors based on global co-occurrence counts for words appearing in the same context from a set of documents (Pennington, Socher, and Manning, 2014). We define a 10 word symmetric context window to train 300-dimensional vectors. In unreported analysis, we test others values for parameters and test the analyses after build the index defined in the next subsection. Changes in vectors dimension to 100 and context window to 5 or 15 generate similar results. It is worth emphasizing that we train the vectors only with our documents, selected from specific sections of news that are directly related to the stock market. So, we assume the trained vectors fully represent the context we are working with, but they are probably not useful in other contexts.

3.3.3. Uncertainty measure

To build the economic uncertainty index, we consider two groups of words: (a) the uncertainty category of Loughran-McDonald master dictionary developed in Loughran and McDonald (2011) translated to Portuguese, which we call V^U; and (b) a group containing the words {economia, econômico}, which is the Portuguese translation for {economy, economic} and we call it V^E. We define the economic uncertainty index for document j as the Euclidean distance between the mean of the vectors of the economy words set and the uncertainty words dictionary. Specifically, the economic uncertainty EU for document j is as follows:

$$EU_j = \left\| \frac{1}{\|V_j^E\|} \sum_{e_{ij} \in V_j^E} e_{ij} - \frac{1}{\|V_j^U\|} \sum_{e_{ij} \in V_j^U} e_{ij} \right\|$$

in which V_j^S, $S \in \{E, U\}$, is the intersubsection of V^S and the set of words contained in document j, and e_{ij} is the vector representation for word i in document j.

Then, for a weekly index, we take the mean of documents indexes, EU_j, for each week. Let D_t be the set of all documents at week t. So we have the economic uncertainty EU for week t:

$$EU_t = \frac{1}{\|D_t\|} \sum_{j \in D_t} EU_j$$

In sum, our economic uncertainty index represents a distance between economic and uncertainty terms in the news stories. We interpret a value of lower magnitude as representing a higher level of economic uncertainty. To change this direction to make interpretation easier, we take the inverse of the index:

$$EU_t^i = (EU_t)^{-1}$$

and we call it EU Index from now on.

3.3.3.1 Alternative uncertainty measures

We also build two alternative uncertainty indexes:

- The economic policy uncertainty index (EPU Index) developed as in Baker et al. (2016) in weekly frequency.
- An index based on word count with term weighting (Fin-Unc Index) as in Loughran and McDonald (2011);

The EPU Index from Baker et al. (2016), widely used in the literature (Brogaard and Detzel, 2015; Bekiros, Gupta, and Majumdar, 2016; Starks and Sun, 2016; Gu, Sun, Wu, and Xu, 2018; Phan, Sharma, and Tran, 2018; Xiong, Bian, and Shen, 2018), is a normalized index defined as the total number of newspaper articles that are discussing economic policy uncertainty scaled by the total number of articles in the newspaper. An article is identified as an economic policy uncertainty article if it contains at least one word of each group: (1) uncertainty terms, (2) economy terms and (3) policy-relevant terms. The authors build a monthly index for the Brazilian market, which is available in their site[11]. The translated and adapted words for the Brazilian market to build our weekly index are taken from their method.

Loughran and McDonald (2011) develops a dictionary of different categories for financial purposes, including the *uncertainty* category. The words are selected from 10-K

[11] `http://www.policyuncertainty.com/brazil_monthly.html`.

reports and are also widely used in finance literature (García, 2013; Liu and McConnell, 2013; Ferguson et al., 2015; Bajo and Raimondo, 2017). The authors use a weighting scheme to define weights for each word to assign more importance to more important words. We use the same method as theirs to define weights, known as tf-idf (term frequency-inverse document frequency), which is defined as follows:

$$w_{i,j} = \begin{cases} \frac{(1+log(tf_{i,j}))}{(1+log(a_j))} log\frac{N}{df_i} & \text{if } tf_{i,j} \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

where $tf_{i,j}$ is the raw count of the word i in the document j, a_j is the average word count in the document j, N is the total number of documents in the sample and df_i is the number of documents containing the word i.

Before calculating these indexes, we remove stop words. In these indexes, the context is not considered, so eliminating words that have no meaning without context, we keep only relevant words. The analyses below are made with these two uncertainty indexes besides the EU Index. Also, to diminish differences in interpretation between them, we select only documents that contain one of the words {economia, econômico}, since the EU Index select only these news stories by definition.

3.3.4. Descriptive statistics

For the following analyses, to facilitate the interpretation in comparing the results we normalize the uncertainty indexes between 0 and 1 following the transformation for each value x:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where $x_{normalized}$ is the value x after transformation and, x_{min} and x_{max} are the minimum and maximum values, respectively, of the index we are normalizing.

To ensure our method reflects the fluctuation in uncertainty in our sample, we calculate the average values for the uncertainty measures we define in section 3.3.3 for each year and report the results in Table **1**. As we mention above, we had very unusual events with economic consequences in the latest years. These critical events begin in 2014 with the presidential elections. The fiscal adjustment unsuccessful attempts mainly characterize the year of 2015. We had the impeachment o President Rousseff in 2016, and an unsettled new government in 2017, which did continue until the end of our sample in 2018. The increase in the average values along the years for our economic uncertainty measure gives us evidence that it is reflecting the uncertainty we perceive to increase over the years, a pattern the EPU Index do not replicate and Fin-Unc Index replicate only since 2015.

Table 1: Uncertainty indexes average values

This table presents the annual average values for the three uncertainty measures we define in section 3.3.3.

	2012	2013	2014	2015	2016	2017	2018
EU Index	0.313	0.482	0.514	0.574	0.653	0.697	0.689
EPU Index	0.167	0.229	0.211	0.277	0.312	0.361	0.206
Fin-Unc Index	0.31	0.35	0.323	0.419	0.475	0.509	0.561

We plot the normalized uncertainty indexes in Figure **1**. Panels A, B, and C represent the EU Index, EPU Index, and the Fin-Unc Index, respectively, in weekly frequency. Table **2** reports Pearson correlation between the uncertainty indexes defined above. The indexes have a high correlation with each other and have the same signs. Panel A of Table **3** reports some descriptive statistics of the news sample we use to calculate the indexes and the uncertainty indexes. Panel B of Table **3** shows the Pearson correlation between uncertainty indexes and stock returns. We compute a negative correlation with weekly returns (Return) and weekly abnormal returns (AR), which is defined below in model 2. In the next section we investigate the impact of the uncertainty indexes on stock returns controlling for variables used in the literature for similar analyses.

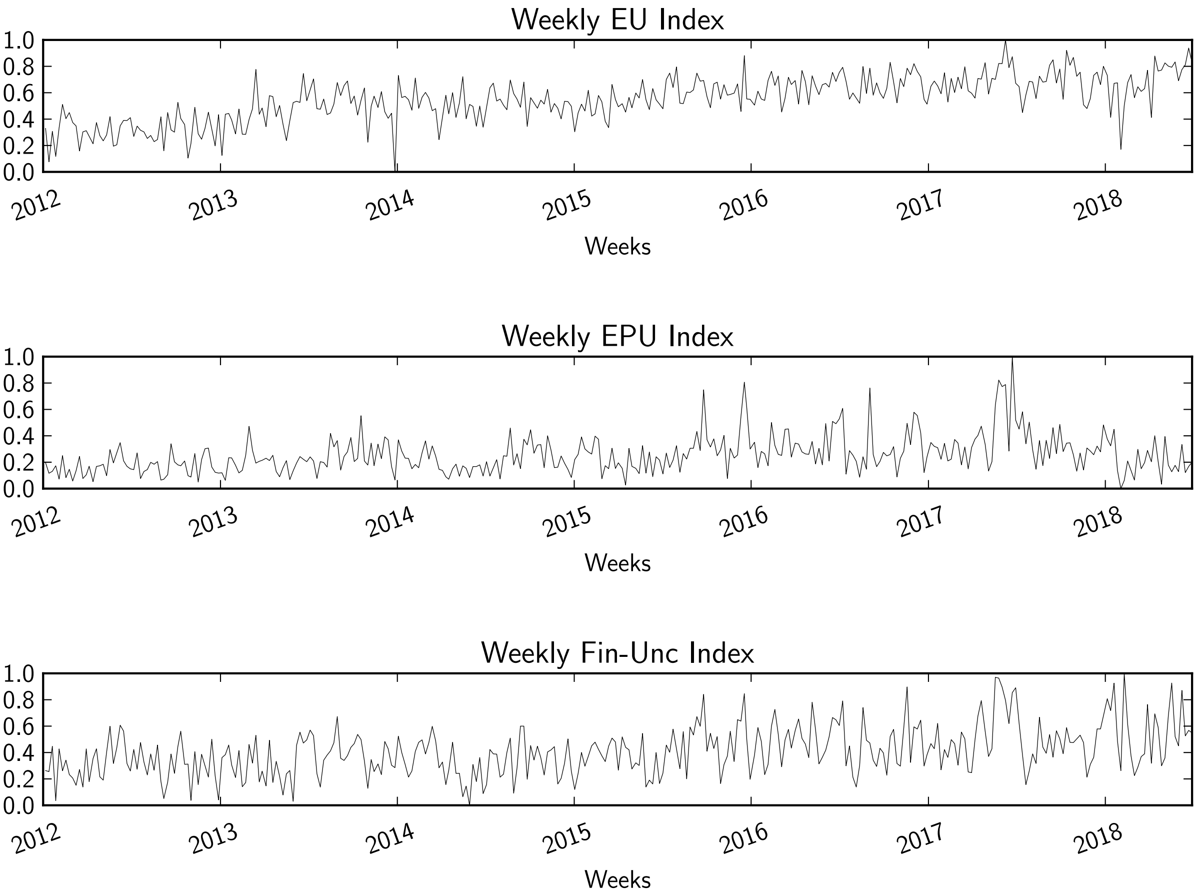
Table 2: Uncertainty indexes correlation

This table presents the Pearson correlation between indexes. ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	EU Index	EPU Index	Fin-Unc Index
EU Index	1.000***		
EPU Index	0.469***	1.000***	
Fin-Unc Index	0.543***	0.557***	1.000***

A concern about the use of the GloVe algorithm is the appearance of ambiguity in the word vectors meaning, for example, antonyms mapped nearby. Since the method uses co-occurrence of words, antonyms could be closer than similar words of different word classes because they appear in the same context more often depending on the set of documents. If the set of documents brings the same sentence structure for words with

Fig. 1. Weekly indexes
This figure represents the uncertainty indexes defined in Section 3.3.3 built in weekly frequency.



opposite meanings, our index could reflect no valid measure or, worse, could reflect something in the opposite direction. For example, *good* can be mapped closer to *bad* than to *goodness*. We perform a few tests with the trained vectors to eliminate concerns about antonyms. We find some cases that indicate this concern can be ignored. For example, the distance between *undefined* and *defined* is larger than *undefined* and *definition*, and *defined* and *define*[12]. Moreover, several uncertainty words in the Loughran-McDonald master dictionary have no clear antonym as *speculate*, *risk*, *probability*, *predict*, *rumors*, *caution* and *assumption*. In this context, it is worth mention that ambiguity is a general difficulty in natural language processing and the other methods we use to estimate

[12]The terms mentioned are a free translation from Portuguese. The original words are in the Appendix. We do the same with the terms involved in the indexes construction cited in this article from now on.

Table 3: Descriptive statistics

This table presents descriptive statistics about the news and the uncertainty indexes. Panel B presents the Pearson correlation with stock returns and abnormal returns. ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: News and indexes								
	N	mean	s.d.	1%	25%	50%	75%	99%
News per week	73,330	216.313	56.76	93.38	177.0	213.0	252.0	356.1
Stocks in news per week	167	48.988	9.411	26.38	43.0	49.0	55.0	71.0
EU Index		0.55	0.173	0.12	0.448	0.552	0.681	0.88
EPU Index		0.255	0.144	0.051	0.158	0.23	0.321	0.782
Fin-Unc Index		0.41	0.185	0.039	0.274	0.4	0.52	0.926

Panel B: Correlations		
	Return	AR
EU Index	-0.012**	-0.016***
EPU Index	-0.017***	-0.018***
Fin-Unc Index	-0.038***	-0.037***

uncertainty are also subject to it. For instance, for the Fin-Unc Index we consider the appearance of an uncertainty term in the text as raising the level of uncertainty in the economy and discard any negation term that could take the meaning of the uncertainty term to the opposite direction since the index assumes independence between words. We also discard the possibility of the uncertainty terms to be related to other objects rather than the economy, such as companies or politics. EPU Index is also subject to misclassification. The appearance of economy, uncertainty and policy makers terms do not guarantee the text assign uncertainty to the economic policies because we discard the context the terms are cited.

To illustrate our trained vectors, we select the relevant words we consider in the indexes we construct. In sum, we have three groups of words: (1) economy terms, which is used in EU Index and EPU Index; (2) policy-relevant terms, relevant for EPU Index; and (3) uncertainty terms, which is the Loughran-McDonald master dictionary, which contains the uncertainty terms proposed by Baker et al. (2016). We then filter the 40 most similar vectors to the vector representing the term *economy* and use t-SNE (t-Distributed Stochastic Neighbor Embedding) technique (Maaten and Hinton, 2008) to reduce the vectors to two-dimensions and plot them in Figure **2**. The sequence of numbers in the graph represents the vectors sorted by similarity to *economy*. There are similarities among words grouped together in the vector space, which indicates the algorithm works properly in our case. For example, *possibility* (noun), *may* (verb) and *might* (verb), and *depending* (verb), *dependent* (adjective) and *depend* (verb), are closely mapped in the graph. Similar nouns as *law*, *legislation* and *regulation*, and *budget* and *deficit*, are also closely mapped.

Moreover, we argue that the strong correlation between the index we propose and the indexes suggested in the literature reported in Table **2** indicates that the index we define using vector representations for words reflects economic uncertainty covered in the news. Also, we have the expected signal in the correlation with stock returns reported in Table **3** suggesting that our index works appropriately.

3.4. Empirical Results

In this section, we present the main results of this paper. We expect the uncertainty indexes generate a negative reaction in stock returns. We first run an OLS regression of weekly stock returns and the uncertainty indexes depicted in section 3.3.3. We control for firm characteristics including book-to-market ratio (B/M), leverage (Leverage), which is defined as the ratio of total assets to the market value of a firm, log of market value in millions of reais (Size), turnover ratio (Turnover), idiosyncratic volatility (IVol), defined as the log of the standard deviation of the daily residuals in a month from the Fama-French three-factor model. To examine whether economic uncertainty measures impact individual stock returns we estimate the following regression:

$$R_{it} = \alpha + \beta_1 Index_t + \beta_2 B/M_{it} + \beta_3 Leverage_{it} + \beta_4 Size_{it} + \beta_5 Turnover_{it} + \beta_6 IVol_{it} + \epsilon_{it} \quad (1)$$

where R_{it} is the log return of stock i at week t and $Index_t$ is the uncertainty index we want to test the predictive power at week t.

In addition, we run the model 1 with weekly stock abnormal returns as dependent variable. We define abnormal return for stock i at week t (AR_{it}) as the residual of the Fama-French three-factor model:

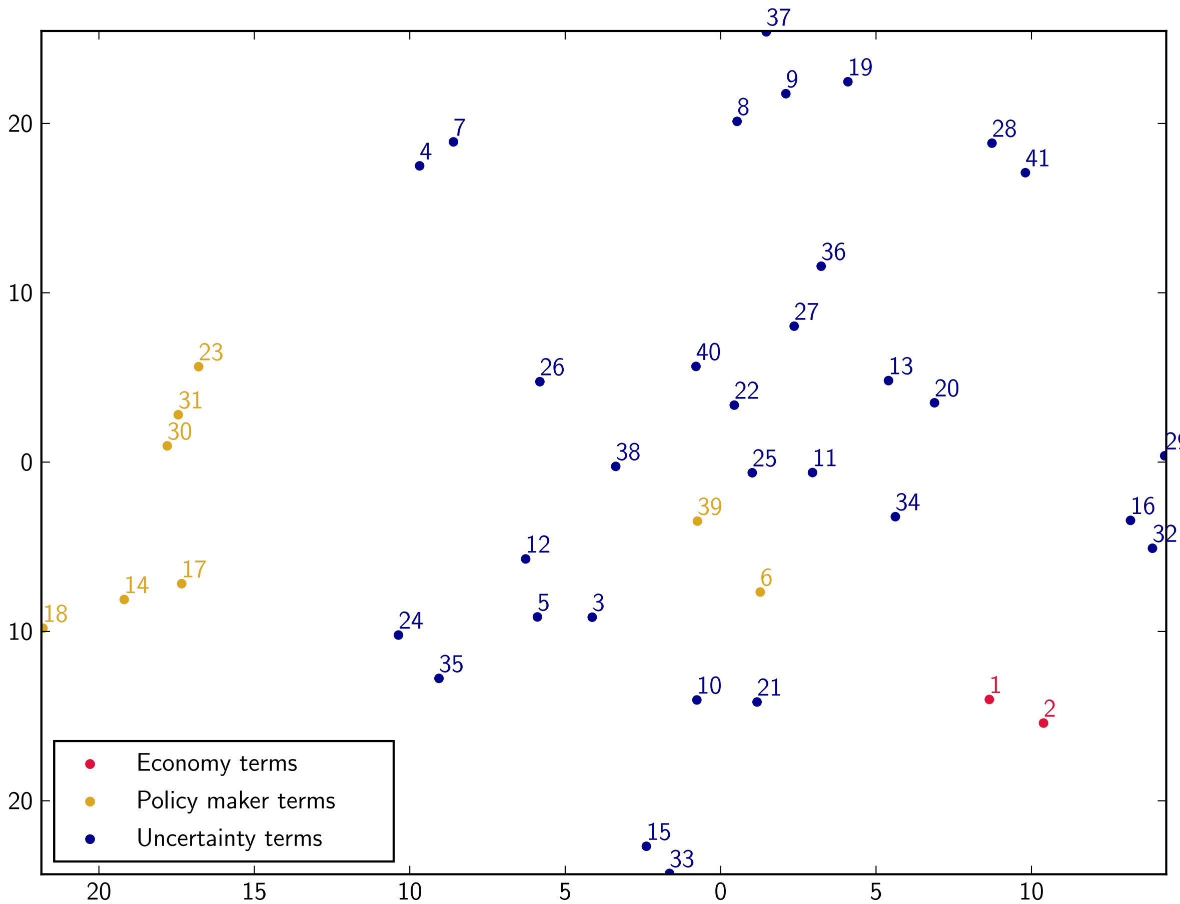
$$AR_{it} = R_{it} - \hat{\alpha} - \hat{\beta_{1i}} MKT_t - \hat{\beta_{2i}} SMB_t - \hat{\beta_{3i}} HML_t \quad (2)$$

where R_{it} is the natural logarithm of return of the stock i at week t and MKT, SMB and HML are the excess market return, firm size factor and book-to-market equity factor, respectively, from Fama-French three factors model (Fama and French, 1992, 1993).

The results of model 1 using stock returns and abnormal returns as dependent variables are reported in Table **4**. The first three columns of Table **4** uses stock returns as

Fig. 2. Word vectors around 'economy'

This figure represents the word vectors most similar to the vector representing 'economy'. The numbers are sorted by similarity to 'economy' before reduction to two-dimension. Each vector in the graph represent a term as follows. 1: economy, 2: economic, 3: may, 4: risk, 5: might, 6: central bank, 7: risks, 8: uncertainties, 9: uncertainty, 10: somewhat, 11: probably, 12: possibility, 13: believe, 14: budget, 15: almost, 16: depend, 17: congress, 18: deficit, 19: instability, 20: believe, 21: different, 22: suggesting, 23: regulation, 24: may, 25: possibly, 26: predicted, 27: predict, 28: cautiousness, 29: dependent, 30: law, 31: legislation, 32: depending, 33: approximately, 34: normally, 35: assume, 36: uncertain, 37: predictability, 38: revise, 39: tax, 40: predicting, 41: cautious. The terms are a free translation from Portuguese. The original words are reported in the Appendix.



dependent variable regressed on each uncertainty index. Columns 4, 5 and 6 make the same analysis with abnormal stock returns used as dependent variable. Columns 1 and 4 shows model 1 with $Index$ equal to $EU\ Index$, in columns 2 and 5 we substitute $Index$ by $EPU\ Index$ and columns 3 and 6 show results with $Fin-Unc\ Index$ as uncertainty index. We find that the uncertainty indexes are good predictors of current stock returns and abnormal returns. The uncertainty measures $EU\ Index$ and $EPU\ Index$ similarly explain stock returns and abnormal returns, and the impact in the abnormal stock returns has a greater magnitude comparing to the effect on stock returns significant at 1% level. The effect of $Fin - Unc\ Index$ on stock returns and abnormal stock returns are similar, and when comparing to the other uncertainty measures the effect is larger.

In line with the literature, the indexes are negatively related with individual stock returns. Also, we have a negative relation between uncertainty indexes and abnormal returns statistically significant at 1% level, which indicates that the impact of the uncertainty indexes estimated from news stories are not related to risk factors. In sum, our estimates of economic uncertainty from news stories make stock prices to drop at the same week and generate negative returns that are not priced by the risk factors we are controlling.

Table 4: OLS regression of stock returns and uncertainty indexes

This table reports the effect of the uncertainty indexes defined in section 3.3.3 on the individual stock returns. The dependent variable of the first and the last three columns are individual stock returns (Return_{it}) and individual stock abnormal returns (AR_{it}), respectively, both in percentage. AR_{it} is defined as the residuals of the Fama-French three-factor model, B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	Return_{it}(%)	Return_{it}(%)	Return_{it}(%)	AR_{it}(%)	AR_{it}(%)	AR_{it}(%)
	(1)	(2)	(3)	(4)	(5)	(6)
EU Index	-0.340**			-0.494***		
	(0.162)			(0.173)		
EPU Index		-0.570***			-0.670***	
		(0.211)			(0.226)	
Fin-Unc Index			-1.184***			-1.143***
			(0.164)			(0.171)
B/M	-0.093***	-0.093***	-0.094***	-0.083***	-0.083***	-0.084***
	(0.027)	(0.027)	(0.027)	(0.029)	(0.029)	(0.029)
IVol	0.303**	0.311**	0.334**	0.551***	0.559***	0.578***
	(0.138)	(0.138)	(0.138)	(0.194)	(0.195)	(0.193)
Leverage	-0.038***	-0.038***	-0.038***	-0.035***	-0.035***	-0.036***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Size	0.175***	0.175***	0.178***	0.177***	0.177***	0.179***
	(0.024)	(0.024)	(0.024)	(0.026)	(0.026)	(0.025)
Turnover	0.126*	0.125*	0.127*	0.276**	0.276**	0.277**
	(0.076)	(0.076)	(0.075)	(0.137)	(0.137)	(0.137)
Intercept	-0.072	-0.085	0.321	0.673	0.602	0.955
	(0.587)	(0.591)	(0.579)	(0.810)	(0.816)	(0.784)

3.4.1. Panel Data Regressions

In the regressions above, we ignore any effect of particular characteristics of companies that we are not controlling for, which could lead to a bias in the parameters of the uncertainty indexes reported in Table **4**. Stocks react differently to economic uncertainty because specific characteristics related to sectors or firm fundamentals can make some stock prices more sensitive to changes in the economy than others. Since our economic uncertainty indexes are based on the appearance of uncertainty words in the news stories, we could measure from some texts about specific firms uncertainty related to firms and not economy. We reduce that type of misguided estimation selecting only news stories where at least one word of economy terms have at least one occurrence, but we can not affirm that we do not have that kind of mistakes in some news stories.

To check that possibility, we run a panel data with fixed effects of stock returns and the uncertainty indexes controlling for the explanatory variables in model 1. The results are presented in Table **5** for each uncertainty measure. Comparing with the OLS regressions, the parameters of the uncertainty indexes do not have a relevant difference in their interpretation about explaining stock returns or abnormal stock returns. Still, the magnitude of these parameters are slightly higher, which eliminates concerns about firms characteristics leading the results. Moreover, we have 73,330 news stories that contain at least one word from economy terms, which is a condition to include the news story in the sample before estimating the uncertainty measures, and only 20,336 from those that contains some keyword for a company or stock in our sample.

When investors read a news story, they interpret the text and keep an impression of the actual context, even if it is not entirely reflecting the real situation. It is more likely that the impressions impact investments instead of the real situation revealed by a deep analysis. When we look at the results that considers known risk factors in Tables **4** and **5**, the result that the economic uncertainty indexes help to explain stock returns remains. Since we intentionally select the same news stories to estimate each index, we quantify precisely the same information using three methods considerably different. This may indicate that besides the impact of the economic uncertainty component reflected in each index related to the actual situation of the economy, media may be an additional determinant in the stock pricing.

The results in the next section explore some known components in the literature that impact investment decisions, which could be driving the impact we find reported in Tables **4** and **5**. We test if the uncertainty indexes we estimate reflect or are related

Table 5: Panel data regression of stock returns and uncertainty indexes

This table reports results from panel data regressions with fixed effects for stocks. Stock returns are regressed on the uncertainty indexes defined in section 3.3.3. The dependent variable of the first and the last three columns are individual stock returns (Return_{it}) and individual stock abnormal returns (AR_{it}), respectively, both in percentage. AR_{it} is defined as the residuals of the Fama-French three-factor model, B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	Return_{it}(%)	Return_{it}(%)	Return_{it}(%)	AR_{it}(%)	AR_{it}(%)	AR_{it}(%)
	(1)	(2)	(3)	(4)	(5)	(6)
EU Index	-0.521**			-0.641***		
	(0.202)			(0.232)		
EPU Index		-0.623***			-0.666***	
		(0.212)			(0.227)	
Fin-Unc Index			-1.340***			-1.281***
			(0.145)			(0.154)
B/M	-0.124***	-0.124***	-0.125***	-0.108***	-0.108***	-0.110***
	(0.019)	(0.020)	(0.019)	(0.015)	(0.015)	(0.016)
IVol	0.533***	0.542***	0.576***	0.725***	0.732***	0.763***
	(0.172)	(0.172)	(0.174)	(0.229)	(0.229)	(0.229)
Leverage	-0.008	-0.009	-0.008	-0.005	-0.006	-0.005
	(0.008)	(0.008)	(0.007)	(0.005)	(0.006)	(0.005)
Size	1.044***	1.030***	1.048***	1.083***	1.066***	1.084***
	(0.135)	(0.132)	(0.139)	(0.123)	(0.120)	(0.125)
Turnover	0.212***	0.211***	0.214***	0.446***	0.445***	0.448***
	(0.082)	(0.082)	(0.081)	(0.164)	(0.164)	(0.163)
Intercept	-6.646***	-6.612***	-6.252***	-6.597***	-6.599***	-6.284***
	(1.042)	(1.045)	(1.057)	(1.192)	(1.207)	(1.197)

with these variables.

3.5. Robustness Tests

3.5.1. Media effect

The primary concern we have since we are dealing with media is the existence of a news or no news effect, as in Fang and Peress (2009), where the authors show that there is a higher return for stock with no media coverage, or in Solomon, Soltes, and Sosyura (2014), where the authors demonstrate that stocks with high past returns attract greater investment when media cover them. Because stocks mentioned in the news are in the spotlight, and hence get more attention from the investors, we expect they are more susceptible to be affected by the tone of economic uncertainty in media than stocks with no media coverage. For the $Fin - Unc\ Index$, the concern is even higher. $Fin - Unc\ Index$ assumes the bag-of-words hypothesis, which means that the order of words in a text is not relevant, and then we can not imply if the uncertainty words are related to the firms mentioned in the document or to the economy words.

To examine if there is a media effect, we include a dummy variable in model 1 that indicates if a stock appears in the news in the current week (News) and combine it with the uncertainty indexes. Table **6** present these results. The first three columns present results with stock returns as dependent variable and the last columns are results from regressions using abnormal stock returns as dependent variable. Each column is the result for each uncertainty index and dependent variable.

For $EU\ Index$, we have the same interpretation as we have in the results reported in Table **4**. Even after controlling for media coverage, economic uncertainty still helps to explain stock returns. The effect of $EPU\ Index$ in stock returns and abnormal returns is concentrated in the combination $News \times EPU\ Index$, which indicates that its effect is concentrated in those stocks mentioned in the news stories. We can still affirm that $EPU\ Index$ helps to predict stock returns and abnormal returns, but only for those stocks highlighted by the media. For the other indexes, the effect of economic uncertainty remains in the stock market as a whole. The positive coefficient on $News \times Fin - Unc\ Index$ suggests that companies mentioned in the news present positive abnormal returns in the presence of the uncertainty measure $Fin - Unc\ Index$.

Table 6: OLS regressions of stock returns and uncertainty indexes with news effect

This table reports the effect of the uncertainty indexes defined in section 3.3.3 on the individual stock returns controlling for media effect. The dependent variable of the first and the last three columns are individual stock returns (Return_{it}) and individual stock abnormal returns (AR_{it}), respectively, both in percentage. AR_{it} is defined as the residuals of the Fama-French three-factor model, News is a dummy variable indicating if the stock was mentioned in the news at the current day, B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio, and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_{it}$(%) (1)	$Return_{it}$(%) (2)	$Return_{it}$(%) (3)	AR_{it}(%) (4)	AR_{it}(%) (5)	AR_{it}(%) (6)
EU Index	-0.508**			-0.704***		
	(0.219)			(0.230)		
EPU Index		-0.222			-0.275	
		(0.290)			(0.300)	
Fin-Unc Index			-1.474***			-1.507***
			(0.230)			(0.254)
News × EU Index	0.258			0.310		
	(0.321)			(0.343)		
News × EPU Index		-0.704*			-0.811*	
		(0.416)			(0.430)	
News × Fin-Unc Index			0.501			0.620*
			(0.332)			(0.360)
B/M	-0.091***	-0.091***	-0.092***	-0.080***	-0.080***	-0.081***
	(0.026)	(0.026)	(0.026)	(0.029)	(0.029)	(0.029)
IVol	0.330**	0.336**	0.362***	0.590***	0.596***	0.618***
	(0.136)	(0.137)	(0.136)	(0.191)	(0.192)	(0.190)
Leverage	-0.036***	-0.036***	-0.036***	-0.033***	-0.033***	-0.033***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
News	-0.354*	-0.033	-0.435***	-0.478**	-0.098	-0.574**
	(0.198)	(0.136)	(0.168)	(0.242)	(0.170)	(0.224)
Size	0.206***	0.205***	0.211***	0.222***	0.221***	0.226***
	(0.027)	(0.027)	(0.027)	(0.033)	(0.033)	(0.033)
Turnover	0.134*	0.132*	0.135*	0.288**	0.286**	0.289**
	(0.077)	(0.077)	(0.077)	(0.141)	(0.141)	(0.140)
Intercept	-0.043	-0.236	0.371	0.697	0.412	1.009
	(0.593)	(0.595)	(0.582)	(0.815)	(0.824)	(0.792)

3.5.2. Public firms

We also test if public or private firms drive the effect above. As discussed before, within the period we analyze, the market is subjected to high political uncertainty. In this context, we have two hypotheses. First, public firms may react more easily to changes in the political scenario with economic consequences. In that scenario the indexes could reflect in part uncertainty related to political issues associated with public firms, for example, different parties disputing elections defending different ideas about public firms management. Second, economic uncertainty may mainly affect private companies since it is expected that public firms are more likely to get financial assistance from government. To test these possibilities and check if the effect is present in both companies, we include a dummy variable indicating if the stock belongs to a public firm (Public) combining it with uncertainty indexes.

We build the indexes to reflect the uncertainty in the economy as a whole, hence it should reflect some systematic risk related to economic uncertainty and not only firms with specific characteristics as to be public or private. As expected, the results in Table **7** indicate that the uncertainty indexes help to forecast returns over the entire sample, without distinction between public or private companies.

We find that for $EPU\ Index$, public companies do not have a different impact from private companies. The coefficients on $EPU\ Index$ in the regressions with stock returns and abnormal returns controlling for public firms are -0.627 and -0.709, respectively, significant at the 1% level, similar to the initial results in Table **4**.

When considering $EU\ Index$ as economic uncertainty measure in a regression with stock returns as dependent variable, the positive coefficient on $Public \times EU\ Index$ in column 1 of Table **7** indicates that in the presence of economic uncertainty, public firms perform better than private firms. When controlling for public firms, the absolute value of the coefficient on $EU\ Index$ increases about 0.1 in both regressions, with stock returns and abnormal returns as dependent variable (columns 1 and 4 of Table **7**). When we regress stock returns and abnormal stock returns on the $Fin - Unc\ Index$, the results are similar to those reported in Table **4**, which indicates that public and private firms similarly perform in the presence of economic uncertainty.

3.5.3. Small caps effect

There are evidences that sentiment predicts returns on small stocks (Lemmon and Portniaguina, 2006; Baker and Wurgler, 2006) since small stocks are held mostly by

Table 7: OLS regressions of stock returns and uncertainty indexes with public firms effect

This table reports the effect of the uncertainty indexes defined in section 3.3.3 on the individual stock returns controlling for public firms effect. The dependent variable of the first and the last three columns are individual stock returns (Return_{it}) and individual stock abnormal returns (AR_{it}), respectively, both in percentage. AR_{it} is defined as the residuals of the Fama-French three-factor model, Public is a dummy variable indicating if the stock belongs to a public firm, B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio, and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_{it}$(%)	$Return_{it}$(%)	$Return_{it}$(%)	AR_{it}(%)	AR_{it}(%)	AR_{it}(%)
	(1)	(2)	(3)	(4)	(5)	(6)
EU Index	-0.442***			-0.591***		
	(0.169)			(0.181)		
EPU Index		-0.627***			-0.709***	
		(0.218)			(0.231)	
Fin-Unc Index			-1.261***			-1.242***
			(0.170)			(0.179)
Public × EU Index	0.981*			0.933		
	(0.578)			(0.588)		
Public × EPU Index		0.541			0.377	
		(0.780)			(0.856)	
Public × Fin-Unc Index			0.735			0.939
			(0.611)			(0.689)
B/M	-0.094***	-0.094***	-0.095***	-0.084***	-0.084***	-0.085***
	(0.028)	(0.028)	(0.028)	(0.032)	(0.032)	(0.032)
IVol	0.300**	0.307**	0.332**	0.547***	0.555***	0.576***
	(0.139)	(0.140)	(0.139)	(0.196)	(0.197)	(0.195)
Leverage	-0.038***	-0.038***	-0.038***	-0.036***	-0.036***	-0.036***
	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)
Public	-0.500	-0.101	-0.266	-0.465	-0.051	-0.341
	(0.332)	(0.240)	(0.263)	(0.339)	(0.290)	(0.269)
Size	0.176***	0.174***	0.178***	0.177***	0.176***	0.179***
	(0.024)	(0.024)	(0.024)	(0.026)	(0.025)	(0.026)
Turnover	0.126*	0.125*	0.127*	0.276**	0.276**	0.277**
	(0.076)	(0.076)	(0.075)	(0.137)	(0.137)	(0.137)
Intercept	-0.042	-0.080	0.344	0.702	0.605	0.984
	(0.588)	(0.591)	(0.580)	(0.814)	(0.817)	(0.787)

individual investors (Lee, Shleifer, and Thaler, 1991). We check if small caps do not drive our results since these stocks are known to be riskier. We filter the sample keeping only the lowest tercile sorted by size, which is computed at the beginning of each month.

Comparing with results reported in Table **4** we find in Table **8** that the effect is higher in stocks with lower market values when we consider $EPU\ Index$ or $Fin - Unc\ Index$ as uncertainty measure. For the $EU\ Index$, the effect on stock returns disappear, but it remains in explaining abnormal returns. $EU\ Index$ and $EPU\ Index$ lose significance compared with results from the whole sample. This can be explained by the decrease in the number of observations. Nevertheless, we want to test if there is an effect associated with this group of stocks that is not present in the rest of the sample. To ensure that the effect does not disappear in the rest of the sample, we run the same model with the excluded stocks.

Table **9** present results for the remaining stocks not considered in the results reported in Table **8**. The effect on stock returns and abnormal returns remains in those stocks indicating that the economic uncertainty indexes help to explain returns in the entire sample. In line with the literature, the effect is smaller in the remaining stocks, but it is not vanished. Comparing with the first results in Table **4**, the smallest decrease in the absolute value of the coefficients is for $EU\ Index$, about 0.06 in both regressions, with stock returns and abnormal stock returns as dependent variable. That difference when we consider $EPU\ Index$ as uncertainty measure is higher, 0.17 in the impact on stock returns and 0.23 for abnormal stock returns. We find that the effect of $Fin - Unc\ Index$ in explaining stock returns and abnormal stock returns reported in Table **4** drops, but it remains large. This evidence indicates that the economic uncertainty index we propose measures risk in the stock market more related to a systematic risk than a risk associated with market capitalization. Nevertheless, the effect we find for the three uncertainty indexes is not entirely driven by small cap stocks.

3.5.4. Illiquidity

We also check if the effect of economic uncertainty we measure from news stories is related to the illiquidity of the stocks. Following the illiquidity measure of Amihud (2002), we define illiquidity (ILLIQ) for stock i at day t as the average daily ratio of absolute stock return to its volume within a month m:

$$ILLIQ_{it} = 10^6 \frac{1}{D_{im}} \sum_{t \in m} \frac{\|R_{it}\|}{Vol_{it}} \tag{3}$$

Table 8: OLS regression of stock returns and uncertainty indexes in small-cap stocks

This table reports the effect of the uncertainty indexes defined in section 3.3.3 on the individual stock returns of small-cap stocks. We consider a stock to be small-cap if it belongs to the lowest tercile of the sample sorted by size. The dependent variable of the first and the last three columns are individual stock returns (Return_{it}) and individual stock abnormal returns (AR_{it}), respectively, both in percentage. AR_{it} is defined as the residuals of the Fama-French three-factor model, B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio, and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_{it}$(%) (1)	$Return_{it}$(%) (2)	$Return_{it}$(%) (3)	AR_{it}(%) (4)	AR_{it}(%) (5)	AR_{it}(%) (6)
EU Index	-0.469			-0.633*		
	(0.332)			(0.374)		
EPU Index		-0.858**			-1.021**	
		(0.432)			(0.479)	
Fin-Unc Index			-1.529***			-1.452***
			(0.331)			(0.361)
B/M	-0.108***	-0.107***	-0.108***	-0.097**	-0.096**	-0.097**
	(0.036)	(0.036)	(0.036)	(0.040)	(0.040)	(0.040)
IVol	0.591**	0.600**	0.620**	1.058**	1.069**	1.084**
	(0.293)	(0.294)	(0.292)	(0.434)	(0.437)	(0.432)
Leverage	-0.039***	-0.039***	-0.040***	-0.039***	-0.039***	-0.040***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Size	0.430***	0.430***	0.435***	0.433***	0.433***	0.437***
	(0.075)	(0.075)	(0.075)	(0.094)	(0.094)	(0.094)
Turnover	0.170*	0.169*	0.170*	0.364**	0.362**	0.364**
	(0.100)	(0.100)	(0.100)	(0.180)	(0.180)	(0.180)
Intercept	-0.679	-0.683	-0.232	0.854	0.807	1.172
	(1.030)	(1.037)	(1.007)	(1.446)	(1.456)	(1.379)

Table 9: OLS regression of stock returns and uncertainty indexes in mid-cap and large-cap stocks

This table reports the effect of the uncertainty indexes defined in section 3.3.3 on the individual stock returns of mid-cap and large-cap stocks. We consider a stock to be large-cap if it belongs to the highest tercile of the sample sorted by size. A stock that do not belongs to the lowest or the highest tercile of the sample sorted by size, we consider to be a mid-cap stock. The dependent variable of the first and the last three columns are individual stock returns (Return$_{it}$) and individual stock abnormal returns (AR$_{it}$), respectively, both in percentage. AR$_{it}$ is defined as the residuals of the Fama-French three-factor model, B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio, and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_{it}(\%)$ (1)	$Return_{it}(\%)$ (2)	$Return_{it}(\%)$ (3)	$AR_{it}(\%)$ (4)	$AR_{it}(\%)$ (5)	$AR_{it}(\%)$ (6)
EU Index	-0.283*			-0.430**		
	(0.172)			(0.171)		
EPU Index		-0.403*			-0.445**	
		(0.223)			(0.224)	
Fin-Unc Index			-0.977***			-0.927***
			(0.175)			(0.173)
B/M	-0.065	-0.064	-0.073	-0.049	-0.046	-0.054
	(0.065)	(0.065)	(0.065)	(0.063)	(0.063)	(0.063)
IVol	0.174*	0.180*	0.204*	0.233**	0.238**	0.259**
	(0.105)	(0.105)	(0.105)	(0.105)	(0.105)	(0.105)
Leverage	-0.029	-0.030	-0.027	-0.027	-0.029	-0.026
	(0.024)	(0.024)	(0.024)	(0.023)	(0.023)	(0.023)
Size	0.080**	0.079**	0.084***	0.091***	0.089***	0.093***
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Turnover	-0.015	-0.015	-0.012	0.004	0.004	0.006
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
Intercept	0.388	0.371	0.716	0.461	0.376	0.681
	(0.503)	(0.502)	(0.504)	(0.500)	(0.498)	(0.500)

where D_{im} is the number of days with available data for stock i at month m, R_{it} is the return for stock i at day t and Vol_{it} is the volume in reais for stock i at day t.

Table **10** present results with ILLIQ added as an explanatory variable and its combination with the uncertainty indexes in each regression. We find no effect on the illiquidity measure on stock returns or abnormal returns. The coefficients on the economic uncertainty measures are quite similar to the initial results reported in Table **4**. We conclude that illiquidity is not a concern that could invalid our uncertainty measures. That is not an unexpected result since we filter our sample by a strong restriction for liquidity before the analyses. Still, liquidity is a concern when we are dealing with emerging markets, which is the case of the Brazilian market.

3.5.5. Political crisis

In our sample, we have two events of great political uncertainty that lead to great uncertainty about the economic environment in the future. One of the events is the election in October 2014. In that scenario, we had two candidates with a close dispute and any new information about something that favored one, or another candidate was enough to induce a market reaction. We had the same sensitivity to the news in the impeachment process of President Rousseff from December 2015 to August 2016. When the result of the process was not clear, the market reacted easily to new informations and facts that pointed to a favored or unfavored result to President. The results of these events were decisive to economic decisions in the future because decisions would depend on the incumbent President. Surely, economic uncertainty is strongly correlated with political crisis and we can not separate the several concerns about different issues involved in those moments. Nevertheless, we select only news stories with words related to the economy, and even if the uncertainty involved in a story is about other political concern, we argue that it has economic consequences. Therefore, we examine how our results are related to those events and if the effect we find is concentrated in that period because we want an economic uncertainty measure that it is efficient independently of whether we have extreme uncertainty or stable economic times. With that purpose, we define a dummy variable that indicates if the week belongs to one of these events period (Crisis) and run model 1 with $Crisis$ added as an explanatory variable and its combination with economic uncertainty indexes. Table **11** reports these results.

The effect of $EPU\ Index$ on stock returns is concentrated in periods of high political uncertainties. The combination with political crisis dummy $Crisis \times EPU\ Index$ have a coefficient of -2.532 significant at 1% level, and the index alone does not explain stock

Table 10: OLS regression of stock returns and uncertainty indexes with illiquidity effect

This table reports the effect of the uncertainty indexes defined in section 3.3.3 on the individual stock returns controlling for illiquidity effect. The dependent variable of the first and the last three columns are individual stock returns (Return_{it}) and individual stock abnormal returns (AR_{it}), respectively, both in percentage. AR_{it} is defined as the residuals of the Fama-French three-factor model, ILLIQ is the illiquidity measure from Amihud (2002), B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio, and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_{it}(\%)$ (1)	$Return_{it}(\%)$ (2)	$Return_{it}(\%)$ (3)	$AR_{it}(\%)$ (4)	$AR_{it}(\%)$ (5)	$AR_{it}(\%)$ (6)
EU Index	-0.325**			-0.482***		
	(0.162)			(0.173)		
EPU Index		-0.582***			-0.680***	
		(0.211)			(0.224)	
Fin-Unc Index			-1.175***			-1.141***
			(0.163)			(0.171)
ILLIQ × EU Index	-0.100			-0.071		
	(0.077)			(0.077)		
ILLIQ × EPU Index		0.055			0.069	
		(0.173)			(0.173)	
ILLIQ × Fin-Unc Index			-0.078			-0.019
			(0.078)			(0.100)
B/M	-0.095***	-0.094***	-0.096***	-0.084***	-0.083***	-0.084***
	(0.027)	(0.027)	(0.027)	(0.029)	(0.029)	(0.029)
ILLIQ	0.043	-0.020	0.012	0.035	-0.015	0.003
	(0.041)	(0.026)	(0.021)	(0.040)	(0.029)	(0.025)
IVol	0.314**	0.319**	0.347**	0.554***	0.559***	0.582***
	(0.139)	(0.140)	(0.138)	(0.195)	(0.197)	(0.194)
Leverage	-0.038***	-0.038***	-0.039***	-0.035***	-0.035***	-0.036***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Size	0.172***	0.173***	0.172***	0.176***	0.177***	0.178***
	(0.024)	(0.024)	(0.024)	(0.026)	(0.026)	(0.027)
Turnover	0.125	0.124	0.125	0.276**	0.276**	0.277**
	(0.076)	(0.076)	(0.076)	(0.138)	(0.138)	(0.138)
Intercept	-0.004	-0.023	0.427	0.684	0.604	0.981
	(0.599)	(0.606)	(0.592)	(0.834)	(0.844)	(0.812)

Table 11: OLS regression of stock returns and uncertainty indexes with political crisis effect

This table reports the effect of the uncertainty indexes defined in section 3.3.3 on the individual stock returns controlling for political crisis effect. The dependent variable of the first and the last three columns are individual stock returns (Return_{it}) and individual stock abnormal returns (AR_{it}), respectively, both in percentage. AR_{it} is defined as the residuals of the Fama-French three-factor model, Crisis is a dummy variable indicating if the day belongs to a period of high political uncertainty, B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio, and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_{it}(\%)$	$Return_{it}(\%)$	$Return_{it}(\%)$	$AR_{it}(\%)$	$AR_{it}(\%)$	$AR_{it}(\%)$
	(1)	(2)	(3)	(4)	(5)	(6)
EU Index	-0.528***			-0.674***		
	(0.165)			(0.177)		
EPU Index		-0.367			-0.579**	
		(0.224)			(0.235)	
Fin-Unc Index			-1.121***			-1.219***
			(0.174)			(0.186)
Crisis × EU Index	1.487			4.130***		
	(1.018)			(0.992)		
Crisis × EPU Index		-2.532***			-0.544	
		(0.682)			(0.789)	
Crisis × Fin-Unc Index			-1.549**			0.623
			(0.611)			(0.685)
B/M	-0.092***	-0.091***	-0.093***	-0.083***	-0.083***	-0.084***
	(0.027)	(0.027)	(0.027)	(0.029)	(0.029)	(0.029)
Crisis	-0.576	1.173***	1.147***	-2.600***	0.152	-0.275
	(0.656)	(0.255)	(0.307)	(0.632)	(0.315)	(0.319)
IVol	0.258*	0.259*	0.279**	0.553***	0.560***	0.577***
	(0.139)	(0.139)	(0.139)	(0.195)	(0.195)	(0.194)
Leverage	-0.038***	-0.038***	-0.038***	-0.035***	-0.035***	-0.036***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Size	0.174***	0.173***	0.176***	0.177***	0.177***	0.179***
	(0.024)	(0.024)	(0.024)	(0.025)	(0.025)	(0.025)
Turnover	0.129*	0.128*	0.130*	0.277**	0.275**	0.277**
	(0.076)	(0.076)	(0.075)	(0.138)	(0.138)	(0.137)
Intercept	-0.197	-0.378	0.032	0.775	0.589	0.975
	(0.593)	(0.596)	(0.587)	(0.820)	(0.822)	(0.799)

returns. However, $EPU\ Index$ still helps to explain stock abnormal returns. The effect of $EU\ Index$ and $Fin-Unc\ Index$ on stock returns and abnormal stock returns remains.

Column 4 of Table **11** shows that in periods of high political uncertainty we have positive abnormal returns associated with variation in $EU\ Index$. This evidence confirms Pástor and Veronesi (2013) findings, where the authors argue that risk premia for political uncertainty is stronger in a weaker economy, in which the current policies are considered harmful, which is the case of Brazil. Since these periods had several uncertainties about policies to be adopted in the future, including economic issues, these results are reasonable. Column 3 of Table **11** reveals that stock returns are negatively correlated with $Fin-Unc\ Index$, and in period of high political uncertainty this effect is more expressed. Although the effect is more pronounced in periods with high economic uncertainty possibly because of political crisis, the uncertainty measures still help to explain returns through different economic scenarios, especially the $EU\ Index$ and the $Fin - Unc\ Index$.

3.5.6. Recession

One particular characteristic of the period we analyze is a considerable time window that the Brazilian economy finds itself in a recession. From the second quarter of 2014 to fourth quarter of 2016 we have a negative real GDP. García (2013) finds that media helps to predict stock returns and the effect is more concentrated in recessions than in expansions in daily frequency analysis. To examine how the economic uncertainty indexes behave in periods with poor economic activities, we construct a dummy variable indicating if the week belongs to the recession period and include it in model 1 with its combination with each uncertainty index. We consider a month belongs to a recession period from the first negative real GDP reported in a quarter to the immediately previous month the first positive result reported in a quarter. The data are reported three months later the result, and then we consider the recession period from September 2014 to February 2017. The results are reported in Table **12**.

We have the same positive effect associated with higher uncertainty levels in recessions as in political uncertainty periods. The difference here is that the effect is present for all uncertainty measures and the effect appears in the regressions with stock returns and abnormal returns used as dependent variable. One explanation for this effect is the higher payment investors require to hold stocks in recessions with higher levels of economic uncertainty. Moreover, the effect on economic uncertainty indexes alone remains, which indicates that the measures we construct from news stories reflect some component not related to the recession period either.

Table 12: OLS regression of stock returns and uncertainty indexes with recession effect
This table reports the effect of the uncertainty indexes defined in section 3.3.3 on the individual stock returns controlling for recession effect. The dependent variable of the first and the last three columns are individual stock returns (Return_{it}) and individual stock abnormal returns (AR_{it}), respectively, both in percentage. AR_{it} is defined as the residuals of the Fama-French three-factor model, Recession is a dummy variable indicating if the day belongs to a recession period, B/M is the book-to-market ratio, Leverage is the ratio of total assets to the market value of a firm, Size is the log of market value in millions of reais, Turnover is the turnover ratio, and IVol is the idiosyncratic volatility. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_{it}$(%) (1)	$Return_{it}$(%) (2)	$Return_{it}$(%) (3)	AR_{it}(%) (4)	AR_{it}(%) (5)	AR_{it}(%) (6)
EU Index	-1.014***			-1.021***		
	(0.178)			(0.182)		
EPU Index		-0.920***			-0.982***	
		(0.251)			(0.257)	
Fin-Unc Index			-1.675***			-1.740***
			(0.193)			(0.204)
Recession × EU Index	3.600***			3.474***		
	(0.516)			(0.613)		
Recession × EPU Index		0.774*			0.965**	
		(0.442)			(0.478)	
Recession × Fin-Unc Index			1.283***			1.744***
			(0.362)			(0.384)
B/M	-0.091***	-0.093***	-0.094***	-0.081***	-0.082***	-0.083***
	(0.027)	(0.027)	(0.027)	(0.030)	(0.029)	(0.030)
IVol	0.313**	0.308**	0.330**	0.590***	0.585***	0.606***
	(0.139)	(0.139)	(0.139)	(0.195)	(0.196)	(0.194)
Leverage	-0.037***	-0.038***	-0.038***	-0.034***	-0.035***	-0.035***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Recession	-2.123***	-0.212	-0.516***	-2.199***	-0.416***	-0.868***
	(0.318)	(0.140)	(0.164)	(0.381)	(0.158)	(0.171)
Size	0.175***	0.175***	0.178***	0.177***	0.176***	0.180***
	(0.024)	(0.024)	(0.024)	(0.026)	(0.026)	(0.025)
Turnover	0.125*	0.125*	0.127*	0.274**	0.273**	0.275**
	(0.076)	(0.076)	(0.076)	(0.138)	(0.138)	(0.138)
Intercept	0.307	-0.011	0.479	1.155	0.845	1.345*
	(0.591)	(0.597)	(0.592)	(0.813)	(0.823)	(0.806)

3.6. Conclusion

This paper investigates the relation between economic uncertainty tone in media coverage and individual stock returns. We introduce a new approach to estimate economic uncertainty from news stories based on word vectors for word representation. We find that economic uncertainty that we estimate helps to predict returns and abnormal returns in the stock market. We also estimate an index based on word count with a term weighting scheme as in Loughran and McDonald (2011) using the uncertainty list from Loughran-McDonald master dictionary and find that it performs quite similar to our measure in predicting returns. We also use the economic policy uncertainty from Baker et al. (2016) as an economic uncertainty measure and find that it affects stock prices, and the effect is concentrated in stocks mentioned in the news.

The advantage of using vector representations for words instead of dictionaries is that the first take into account the context of the documents sample, while applications of dictionaries usually assume the bag-of-words hypothesis, which ignores the position in the document. The method we use to quantify economic uncertainty from news stories can be adapted to other contexts to measure a particular variable.

Appendix: Translated words

The list below reports the translation to Portuguese of the terms involved in the uncertainty measures cited in this paper.

Almost: Quase.

Approximately: Aproximadamente.

Assume: Assumir.

Believe: Acreditam, acreditar.

Budget: Orçamento.

Cautious: Cauteloso.

Cautiousness: Cautela.

Central bank: Banco central.

Congress: Congresso.

Deficit: Déficit.

Define: Definir.

Defined: Definido.

Definition: Definição.

Depend: Depender.

Dependent: Dependente.

Depending: Dependendo.

Different: Diferente.

Economic: Econômico.

Economy: Economia.

Instability: Instabilidade.

Law: Lei.

Legislation: Legislação.

May: Pode, poder.

Might: Poderia.

Normally: Normalmente.

Possibility: Possibilidade.

Possibly: Possivelmente.

Predict: Prever.

Predictability: Previsibilidade.

Predicted: Previsto.

Predicting: Prevendo.

Probably: Provavelmente.

Regulation: Regulação.

Revise: Rever.

Risk: Risco.

Risks: Riscos.

Somewhat: Algo.

Suggesting: Sugerindo.

Tax: Imposto.

Uncertain: Incerto.

Uncertainties: Incertezas.

Uncertainty: Incerteza.

Undefined: Indefinido.

Chapter 4

Measuring Corruption: Evidence from Brazil Scandals

4.1. Introduction

This paper studies the performance of the individual stock prices of two Brazilian companies involved in corruption scandals after the information about the corruption schemes becomes public. First, we examine how the amount of corruption covered by the media mentioning a company affects their stock returns in the short-term. We find evidence that media helps to explain stock returns when reporting corruption about a specific firm, but only in the case the firm is not state-owned. We also examine if there is a long-term effect by constructing a synthetic unit for both companies involved in corruption scandals. The results show that the level of stock prices of both cases after the scandal is below the level of synthetic companies stock prices.

Corruption hampers economic growth (Shleifer and Vishny, 1993), but for the firms politically connected, the scheme can be beneficial (Faccio, 2006; Bunkanwanicha and Wiwattanakantang, 2009; Ovtchinnikov and Pantaleoni, 2012). Nethertheless, even if a particular firm takes advantage from keeping political connections, when the relationship ends the firm value may decrease (Pan and Tian, 2017; Wang, Xu, Zhang, and Shu, 2018; Xu, 2018). Investors expect the firms politically connected outperform. For example, there is evidence that stock returns of politically connected firms are negatively more impacted than less politically connected firms when rumors appears about a government change (Fisman, 2001). Also, the public information about the ending of a connection negatively impacts stock returns (Acemoglu, Johnson, Kermani, Kwak, and Mitton, 2016). Our study contributes to this literature focusing on corruption scandal events, when the information about corruption schemes comes out.

Brazil is known for having corruption scandals, especially in the latest years when several corruption schemes came to light. In 2018, Brazil ranked 105th out of 180 countries on the CPI (Corruption Perception Index) from Transparency International[13]. Therefore, the country presents a favorable scenario to study the impact of not only government corruption, but corruption in specific firms with political connections, especially after the beginning of the operation "Car Wash" (lava-jato), which expose several corruption schemes including public and private companies. The operation "Car Wash" is an ongoing investigation carried out by the Brazilian Federal Police. In March 2014 Federal Police launched the operation to investigate corruption schemes that soon would become the largest anti-corruption probe in Brazil's history.

We have evidence that there is a positive effect in companies that keep political connections (Fisman, 2001; Amore and Bennedsen, 2013; Wang, 2015), and hence companies have incentives to pay bribes for receiving favors. In Brazil case, Claessens, Feijen, and Laeven (2008) find that companies which donate for political campaigns of elected politicians have greater access to bank financing relative to other firms and present higher stock returns around the election. Despite the evidence that investors expect corrupt firms to have a better performance, we aim to examine whether the market continues to believe the firms outperform when the corruption schemes are suddenly confirmed. In sum, we propose to answer the following question: how do investors react if the information about the corruption practices become public?

To examine the cost of corruption after the scandals emerge, we consider two companies involved in corruption scandals occurred in the latest years: Petrobras, a state-owned oil company, and JBS, a global food private industry. Petrobras scandal, emerged in the operation "Car Wash", brought out the information that the company was used to receive bribes for signing contracts with private firms. JBS scandal was part of the "Weak Flesh" (carne fraca) scandal, an investigation into corruption schemes that consisted of paying bribes to public agents in exchange for failing inspection.

To analyze the role of a corruption scandal, we first construct a corruption measure from media coverage and examine its relation with individual stock returns. We find that there is an effect of our corruption measure on stock returns for JBS, but no effect for Petrobras. We argue that a possible explanation for this result is that the investors expect the state-owned companies as Petrobras receive financial assistance from the government, while a private company, as the case of JBS, is exposed to higher risk and possible bankruptcy.

Second, to study how stock prices behave in the long-term, we apply the synthetic

[13] Data are available at `https://www.transparency.org/cpi2018`

control method developed by Abadie and Gardeazabal (2003) for each company to define a comparable evolution of the stock prices in the absence of the corruption scandal. Both cases evidence that after the scandal, the stock prices level is below their respective synthetic control stock prices. We conclude that there is a negative effect of a corruption scandal on firm value, but in the short term corruption news stories impact more significantly private firms relative to public firms.

4.2. Brazilian Corruption Scandals

4.2.1. Petrobras Scandal

Petrobras is a Brazilian state-owned oil company that was highlighted in the media after corruption scandals discovered in 2014. Investigations into corruption schemes in the company are part of the operation "Car Wash", a large operation carried out by the Brazilian Federal Police against corruption started in March 2014.

Petrobras scandal emerges with a plea bargaining signed by the former director of Petrobras Paulo Roberto Costa whereby Costa agreed to explain the corruption scheme and reveal the names of the beneficiaries involved in the scandal in exchange for a lighter sentence. Federal Police found that Petrobras executives were receiving bribes from companies to sign contracts. Companies were paying to ensure they get the contract, which was a very lucrative business. Later, many people were discovered to be involved in the scheme, including executives from private companies and politicians.

After the information of the corruption scheme became public, Petrobras preferred stocks were traded at a minimum price of 4.37 reais on January 18, 2016, while at the beginning of the scandal, on August 25, 2014, the price reached 23.15 reais[14]. The price drops evidence the importance of the public information about corruption schemes in driving stock prices.

4.2.2. JBS Scandal

In March 2017, Brazil receives the information that public agents, including inspectors, were taking bribes to allow firms to sell products with irregularities. These pieces of information are part of the operation so-called "Weak Flesh". Carried out by the Brazilian Federal Police, the operation is a criminal investigation that began to investigate irregularities in the Federal Inspection System (Sistema de Inspeção Federal - SIF).

[14] Real is the official currency of Brazil. One US dollar equaled 4.04 reais on January 18, 2016 and 2.28 reais on August 25 2014.

Companies paid for failing inspection and falsifying export documents. JBS, a global food industry, is one of these companies.

After the "Weak Flesh" scandal, more corruption schemes involving the company came out. JBS owners paid bribes for contracts with the Brazilian Development Bank (Banco Nacional de Desenvolvimento Econômico e Social - BNDES), which one of its owners, Joesley Batista, declared after the scandal emerged. The company owners had political connections that allowed them to receive benefits from government. The company received loans from the Brazilian Development Bank as a development policy of the government and made donates to elect candidates of different parties, which guarantted influence in National Congress. In May 2017, Brazilian Federal Police began the operation so-called "Bullish" to investigate frauds and irregularities in BNDES loans to JBS.

4.3. Literature Review

Part of the literature that examines the impact of corruption on investor behavior evidences that stock returns increase in the presence of corruption. For example, stock returns increase when managers have a higher propensity to corrupt (Mironov, 2015), the firm has a political privileged geography (Kim, Pantzalis, and Park, 2012; Pantzalis and Park, 2014) and when the firm contributes to candidates (Claessens et al., 2008; Cooper, Gulen, and Ovtchinnikov, 2010). In line with these results, some authors show that when the political connection ends, stock returns decrease (Wang et al., 2018) and firm value decreases (Xu, 2018). Furthermore, adverse information for politically connected firms that threatens this relation makes stock returns to drop (Fisman, 2001).

We make our analyses considering two companies involved in corruption scandal: one under private ownership and one under public ownership. Therefore, our paper is also related to the literature that separately analyses these two kinds of company. The literature indicates a different effect of political connections in a company depending on if it is private or public. Furthermore, these different effects differ among studies. For example, Pan and Tian (2017) find that investment in private firms decreases more and becomes less efficient relative to public firms after the political connection ends. Chen, Li, Luo, and Zhang (2017) shows that political connection has a negative effect on public firms. For private firms, the impact is positive, but only at lower levels of political connection. Also, Wang (2015) evidence that political connection increases firm value in private firms and decreases the value in public firms. About corruption in the business environment, Nguyen and van Dijk (2012) find that there is a negative relation between perceived corruption and private firms growth, and on public firms growth, the impact of perceived corruption is insignificant.

In a non-adverse economic and political condition in Brazil, political connection benefits the firm and increase their stock returns (Claessens et al., 2008). On the other hand, the political crisis in Brazil generates a stronger negative effect on stock returns for politically connected firms relative to non-connected firms (Hillier and Loncan, 2019). In line with these results, Padula and Albuquerque (2018) evidence, also for the Brazil case, a devaluation of state-owned companies due to corruption scandals. That study does not consider the impact on the specific company involved in the scandals as ours, but the impact of scandals on public companies as a whole.

4.4. Data and method

Our approach is divided into two steps. We first build a corruption measure based on news stories. Our goal is to examine how corruption reported in the news affects stock returns. In a second moment, we aim to examine the impact of corruption in the long-term. For this purpose, we build a synthetic Petrobras and a synthetic JBS to determine the stock prices trajectory that each company would have in the absence of the corruption scandal. The subsequent analyses present results when we select the preferred stocks for Petrobras (PETR4) to represent the company, but in unreported analyses, we find similar results with the common stocks (PETR3). JBS only trade common stocks (JBSS3).

We collect financial data from January 2012 through June 2018 of stocks traded on the BM&FBovespa from Economática database, and risk factors for Brazilian market from data provided by São Paulo University[15] (Universidade de São Paulo). For evaluating the effect of corruption covered by the media on the stock returns, we run weekly and monthly regressions for each company. To estimate the synthetic control, we consider the data in monthly frequency.

4.4.1. Corruption measure

To estimate corruption reported in the news, we first collect news stories from two popular newspapers in Brazil, Valor Econômico and Folha de São Paulo Online. We collect news stories from sections related to finance, business and politics of Valor

[15]The data are available in http://nefin.com.br/.

Econômico[16] and sections related to international news, politics, finance, economy and investments of Folha de São Paulo Online[17], both from January 2012 through June 2018. In the analyses, we suppose a news story has an impact on the day it is published. So, if a story becomes known to investors in a day with the market closed or in a day after half an hour before market closure, we consider this story is published in the next trading day.

For filtering the news stories, we convert all letters to lowercase and take off accents, except those proper names in the dictionary that become ambiguous with the process when we look at the isolated word. Before parsing the text, we create a dictionary with synonyms for each company in the sample and substitute each word related to a company in the dictionary by a unique word that represents the company. We then filter punctuation, links, and numbers, except percentage numbers, which we replace by +[%] (−[%]) if it is a positive (negative) number. After replacements, we remove terms that occur less than five times in documents vocabulary, except if it is documented in the dictionary. This process is needed to remove very infrequent terms, which meaning is hard to detect.

We build a set of corruption words and before constructing the corruption index for a company, we select only news stories that mention the company. We consider the following words as terms related to corruption: *corruption*, *bribe*, *scandal*, *scheme* and *slush fund*[18]. We define a corruption index for each news story j (Corruption$_j$) as the weighted count of all corruption words that occur in the text. As in Loughran and McDonald (2011) we use the method known as tf-idf (term frequency-inverse document frequency) to define weights, which is defined as follows:

$$w_{i,j} = \begin{cases} \frac{(1+log(tf_{i,j}))}{(1+log(a_j))} log\frac{N}{df_i} & \text{if } tf_{i,j} \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

where $tf_{i,j}$ is the raw count of the word i in the document j, a_j is the average word count in the document j, N is the total number of documents in the sample and df_i is the number of documents containing the word i. If there is no corruption term in the document, we define Corruption$_j = 0$. Then the corruption of the company c (Corruption$_c$) in a period

[16] *Finanças*, *Empresas* and *Política* sections in https://www.valor.com.br/
[17] *Mundo*, *Poder* and *Mercado* sections in https://www.folha.uol.com.br/
[18] The corruption words are a free translation from Portuguese. The original words are *corrupção*, *propina*, *escândalo*, *esquema* and *caixa 2*, respectively.

is defined as the average corruption$_j$ for all news story j in the period.

Table **13** reports descriptive statistics about the news stories and corruption indexes. We consider the period before scandal for Petrobras, all periods until August 22, 2014, when occurred the event we assume as the scandal beginning. The periods before scandal for JBS are the period before March 17, 2017, when the "Weak Flesh" scandal emerged. In Panel A, we present descriptive statistics for Petrobras and in Panel B, for JBS. Figures **3** and **4** present the quantity of news stories per month that mention the company and contains at least one corruption word. For both companies, we have an increase in the number of news stories about corruption after the scandal becomes public, which demonstrate that media widely cover such information and hence plays an important role in disseminating the scandals.

Table 13: Descriptive statistics

This table presents descriptive statistics about the news and the corruption indexes. Panel A and B presents descriptive statistics for Petrobras and JBS, respectively.

Panel A: Petrobras							
	mean	s.d.	1%	25%	50%	75%	99%
Corruption news per week before scandal	4.949	9.135	0.0	0.0	1.0	3.0	44.0
Corruption news per week after scandal	61.713	47.025	8.01	27.0	45.5	86.75	202.87
Corruption news per month before scandal	20.645	37.791	1.3	4.0	6.0	14.5	154.4
Corruption news per month after scandal	266.043	176.394	59.92	124.0	215.0	378.0	688.22
Weekly corruption	0.735	0.987	0.0	0.01	0.345	1.134	4.237
Monthly corruption	0.623	1.046	0.0	0.0	0.191	0.807	4.09
Panel B: JBS							
	mean	s.d.	1%	25%	50%	75%	99%
Corruption news per week before scandal	0.716	1.924	0.0	0.0	0.0	1.0	5.6
Corruption news per week after scandal	27.353	35.762	0.67	5.0	14.0	36.25	142.97
Corruption news per month before scandal	3.113	6.008	0.0	0.0	1.0	4.0	25.53
Corruption news per month after scandal	116.312	122.171	11.2	26.25	63.0	175.75	389.05
Weekly corruption	0.319	1.132	0.0	0.0	0.0	0.038	5.174
Monthly corruption	0.269	1.033	0.0	0.0	0.0	0.0	4.454

Figures **5** and **6** show the plot in monthly frequency of our corruption measures for Petrobras and JBS, respectively. We highlight some main events of each scandal that we hypothesize impact the stock returns of each company. As we expect, our corruption measure reflects those critical events. Higher levels of corruption occur while investigations about the corruption schemes proceed.

Fig. 3. Amount of Petrobras corruption news stories per month

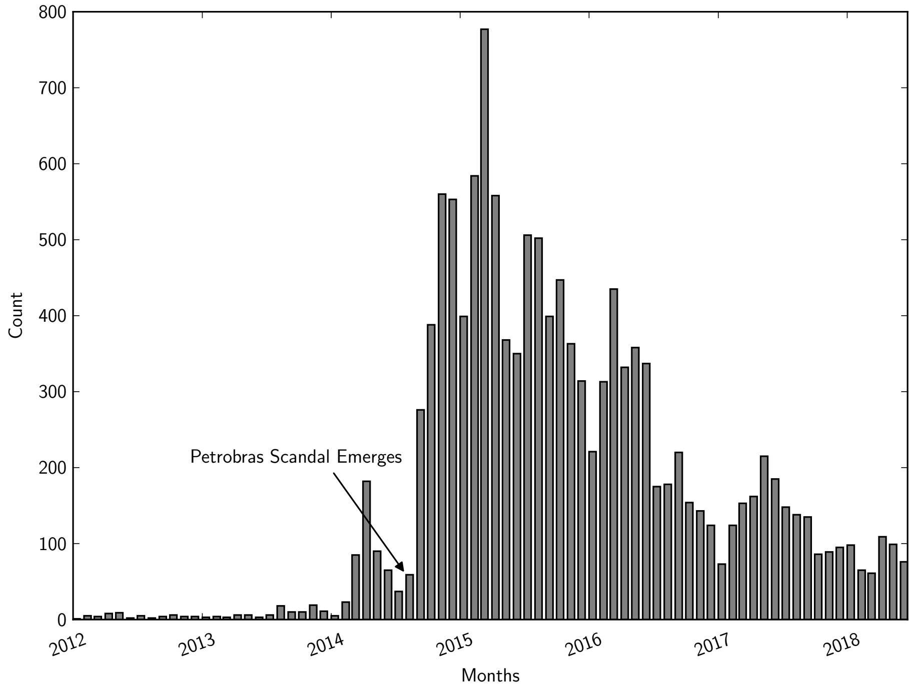


Fig. 4. Amount of JBS corruption news stories per month

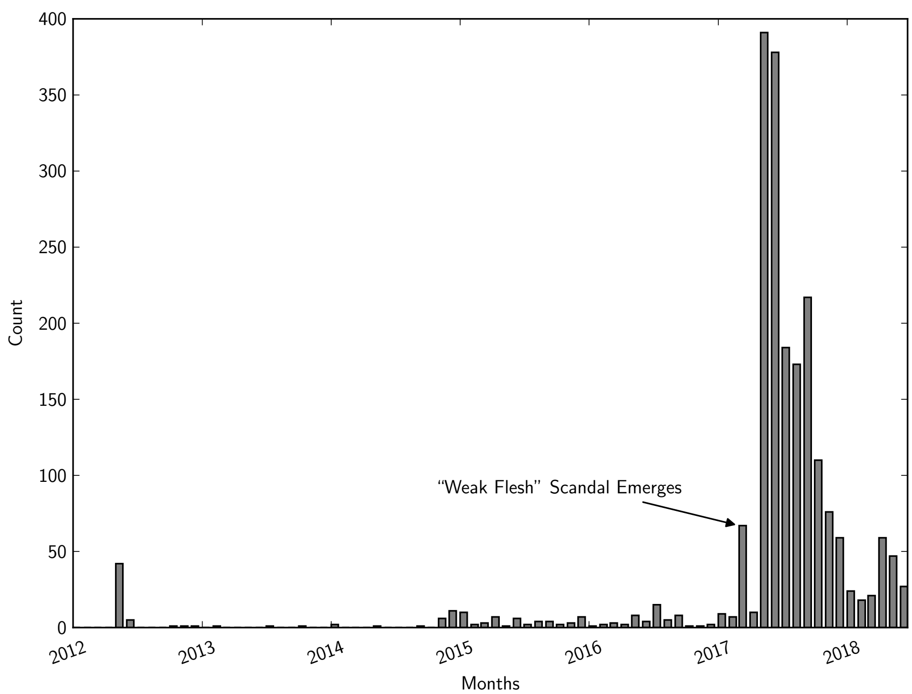


Fig. 5. Petrobras monthly corruption index

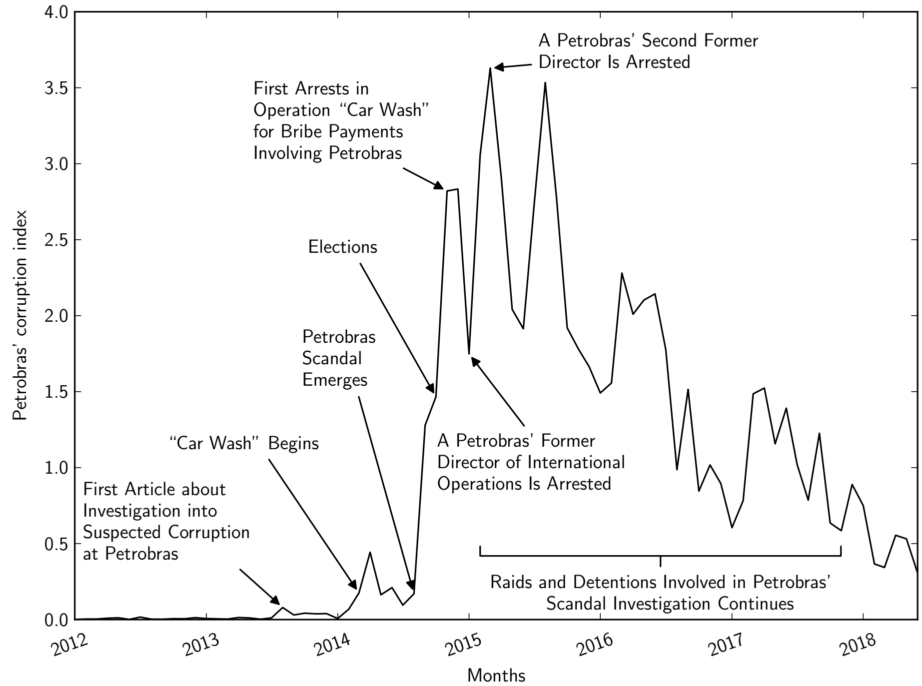
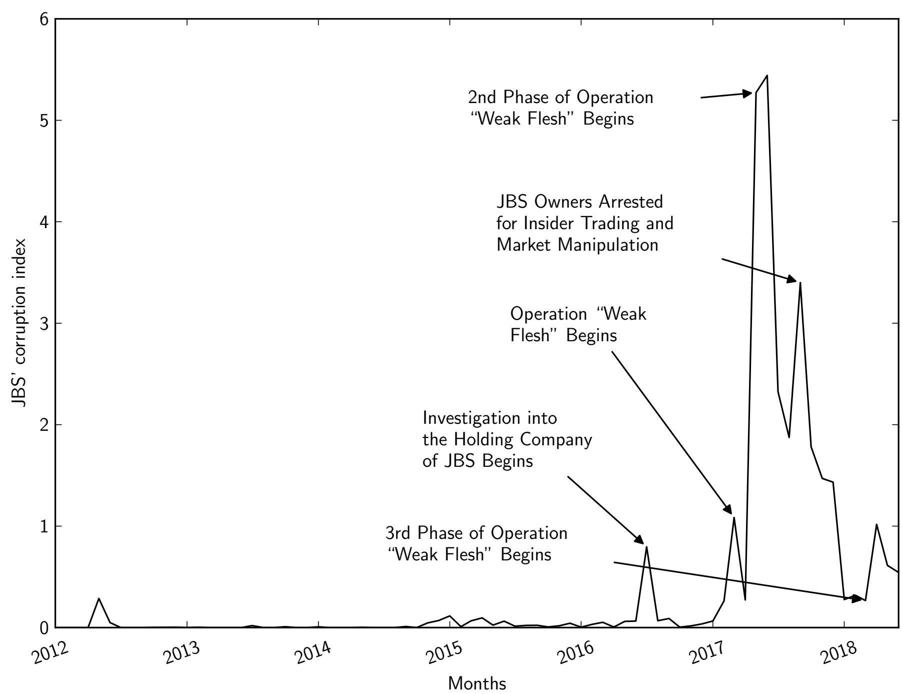


Fig. 6. JBS monthly corruption index



4.4.2. Synthetic control

There is no firm in the stock market with similar characteristics of Petrobras or JBS that we could use as a control to imply the effect on stock prices of the corruption scandal by comparing with it. We then use the synthetic control method developed by Abadie and Gardeazabal (2003) to build a synthetic company. The synthetic control method consists in constructing a control unit by generating weights to potential controls that result in a trajectory with similar values for the variable we want to evaluate the impact before the event we suppose to change the trajectory, in our case, the corruption scandal. Therefore, after the event, the trajectory of the control unit would be the result in the absence of the event.

To select potential controls for Petrobras and JBS, we select only stocks that are traded all over the analysis period and has no missing data for any predictor. We also select one stock per company to represent the company and exclude those companies with suspect of corruption. In constructing the synthetic Petrobras we consider only public firms, and for the synthetic JBS, we select only private firms that compose the main performance indicator of the stocks traded in Brazilian Stock Exchange (IBovespa). We evaluate the impact on the closing price and consider the following predictors to obtain the synthetic company: profit per share in reais, turnover, sharpe ratio, earnings-to-price (E/P), return over equity (ROE), and market value in reais. We choose the pre-treatment period as the period that precedes each corruption scandal. For Petrobras, the pre-treatment period goes until July 2014, one month before the scandal, and for JBS, we estimate weights for the synthetic firm until February 2017, one month before the "Weak Flesh" operation beginning.

4.5. Empirical results

In this section, we provide the main results of the paper that examines the impact of public information about corruption in Petrobras and JBS on their stocks price.

4.5.1. Media and stock returns

We first run the following regression model for each company to estimate the impact of the corruption measure depicted in Section 4.4.1 on the stock returns:

$$Return_t = \alpha + Corruption_t + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t \tag{4}$$

where $Return_t$ and $Corruption_t$ is the log return and the corruption measure of the stock we are evaluating the effect at week t, and MKT, SMB and HML are the excess market return, firm size factor and book-to-market equity factor, respectively, from Fama-French three factors model (Fama and French, 1992, 1993).

Table **14** presents results of model 4 in weekly frequency for both companies. Corruption$_{jbs}$ and Corruption$_{petr}$ are the corruption measures for JBS and Petrobras, respectively. We also run the model with year fixed effect (Columns 3 and 4) and with a dummy variable indicating if the week belongs to a recession period (Columns 5 and 6) since recession can potentialize the media effect, as in García (2013). From the second quarter of 2014 to the fourth quarter of 2016 we have a negative real GDP. We define the recession period as the period from the first negative real GDP reported in a quarter to the immediately previous month the first positive result reported in a quarter, in our case, from September 2014 to February 2017.

Results in Table **14** show that our corruption measure helps to explain stock returns for JBS and the result is robust when we add year fixed effect and consider the recession in the analysis. When we estimate the JBS scandal from news stories, we find a significant impact in their stock returns, which evidences that the market expects the company has a poor performance. On the other hand, the Petrobras corruption measure does not explain the stock returns of the company, except when we add a year fixed effect. Nevertheless, the effect is not robust for further analyses. Table **15** reports results for the same analyses as in Table **14**, but in monthly frequency. We find similar results for both analyses. Corruption scandals impact stock returns in a company, but only in the case it is under private ownership.

We explain the different impact between the companies as having support in the characteristic of having public or private ownership. In the Petrobras case, a public company, investors presume that the government will offer financial help, and hence they do not await the company goes belly up, which makes the demand for their stocks continue to exist. The same does not occur with private companies. In the JBS case, the government does not have a commitment with eventually financial loss that the company will face. Therefore, private companies hold a riskier situation in a possible corruption scandal relative to public companies.

There is evidence that firms with political connections receive benefits from the government in Brazil and the investors expect the company has that distinguished treatment (Claessens et al., 2008). Our results do not oppose that evidence, but they indicate that in the case the corruption scheme becomes public, investors expect it has a negative

Table 14: Impact of weekly corruption on stock returns

This table reports the effect of the weekly corruption measure defined in Section 4.4.1 on the stock returns. Columns 1 and 2 present results from model 4 for JBS and Petrobras, respectively. In Columns 3 and 4 we add a year fixed effect, and in columns 5 and 6, a dummy indicating the recession periods. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_t$(%) (1)	$Return_t$(%) (2)	$Return_t$(%) (3)	$Return_t$(%) (4)	$Return_t$(%) (5)	$Return_t$(%) (6)
$Corruption_{jbs}$		-0.861***		-1.141***		-0.904***
		(0.326)		(0.320)		(0.320)
$Corruption_{petr}$	-0.247		-0.784***		-0.305	
	(0.236)		(0.302)		(0.284)	
MKT	1.883***	0.762***	1.899***	0.760***	1.884***	0.758***
	(0.123)	(0.161)	(0.120)	(0.160)	(0.123)	(0.161)
SMB	0.196	-0.375**	0.252	-0.383**	0.196	-0.384**
	(0.172)	(0.175)	(0.186)	(0.191)	(0.173)	(0.179)
HML	0.140	0.244	0.092	0.254	0.138	0.257
	(0.181)	(0.215)	(0.187)	(0.221)	(0.183)	(0.220)
Intercept	0.339	0.432	-0.389	-0.002	0.324	0.633**
	(0.254)	(0.284)	(0.413)	(0.752)	(0.261)	(0.321)
Year fixed effects	No	No	Yes	Yes	No	No
Recession	No	No	No	No	Yes	Yes

Table 15: Impact of monthly corruption on stock returns

This table reports the effect of the monthly corruption measure defined in Section 4.4.1 on the stock returns. Columns 1 and 2 present results from model 4 for JBS and Petrobras, respectively. In Columns 3 and 4 we add a year fixed effect, and in columns 5 and 6, a dummy indicating the recession periods. The standard errors of the parameters are reported in parentheses. The standard errors of the parameters are reported in parentheses. Standard errors are heteroscedasticity and autocorrelation robust (HAC). ***, ** and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

	$Return_t$(%) (1)	$Return_t$(%) (2)	$Return_t$(%) (3)	$Return_t$(%) (4)	$Return_t$(%) (5)	$Return_t$(%) (6)
$Corruption_{jbs}$		-2.603**		-3.462**		-2.659**
		(1.067)		(1.382)		(1.039)
$Corruption_{petr}$	-0.388		-2.297		-0.121	
	(1.093)		(1.624)		(1.771)	
MKT	2.158***	0.302	2.247***	0.290	2.157***	0.297
	(0.219)	(0.282)	(0.196)	(0.278)	(0.219)	(0.288)
SMB	0.350	-0.176	0.719***	-0.013	0.351	-0.184
	(0.221)	(0.280)	(0.274)	(0.354)	(0.221)	(0.281)
HML	-0.193	0.498	-0.504**	0.487	-0.187	0.511
	(0.258)	(0.308)	(0.238)	(0.360)	(0.262)	(0.323)
Intercept	0.803	1.327	-2.999	-2.004	0.817	1.561
	(1.137)	(1.249)	(1.876)	(3.087)	(1.119)	(1.386)
Year fixed effects	No	No	Yes	Yes	No	No
Recession	No	No	No	No	Yes	Yes

effect in the company even if before the scandal there is an indication of a corruption scheme that benefits the company and increases the firm value.

An alternative explanation for the negative abnormal return we evidence for JBS may be a result of the valuation the investors attribute for the company due to the corruption before it becomes public, as the effect Acemoglu et al. (2016) evidence. Thus, the negative effect we find can reflect the expectation about the end of the schemes and hence, after the scandal, there is a reversal movement. For the Petrobras case, following the same logic, investors may not value corruption schemes because the company is under public ownership, which explains the absence of an effect. In this case, our result is in line with Chen et al. (2017) and Wang (2015) that evidence differences between the value of political connection for private firms and public firms.

4.5.2. Synthetic control

To examine the effect of the corruption scandal on the stock prices in the long-term, we apply the synthetic control method. For each company, we construct a synthetic unit and examine the divergence between the stock prices trajectories. If the corruption scandal has an impact on the stock prices, we expect the difference between the stock prices of the company and the synthetic company diverge after the scandal.

4.5.2.1 Petrobras case

We notice that after the information about corruption schemes in Petrobras became public, the company stock prices decreased. Although, the period we analyze is economic and politically troubled, which could impact the stock prices also, especially for the public stocks, as Padula and Albuquerque (2018) evidence. Therefore, building a synthetic Petrobras allow us to estimate the evolution of the stock prices in the absence of the corruption scandal.

To estimate the synthetic control for Petrobras, we select only public companies as potential controls, since the effect in the stock prices could come from political issues or the economic situation of the country that impact state-owned companies differently relative to private companies. The set of potential controls for Petrobras is compounded by *Banco do Brasil* (*Bank of Brazil*); the electric power generation companies *Celesc*, *CEMIG*, *CESP*, *Copel*; and the water and waste management companies *COPASA*, *SANEPAR* and *SABESP*. Table **16** reports the weights of each control firm in the synthetic Petrobras. Prior to the corruption news, the Petrobras stock prices is best reproduced by a combination of *Banco do Brasil* (66.8%) and *Celesc* (33.2%).

Table 16: Firm weights in the synthetic Petrobras

Company	Stock	Weight
Banco do Brasil	BBAS3	0.675
Celesc	CLSC4	0.325
CEMIG	CMIG4	0
CESP	CESP6	0
COPASA	CSMG3	0
Copel	CPLE6	0
SANEPAR	SAPR4	0
SABESP	SBSP3	0

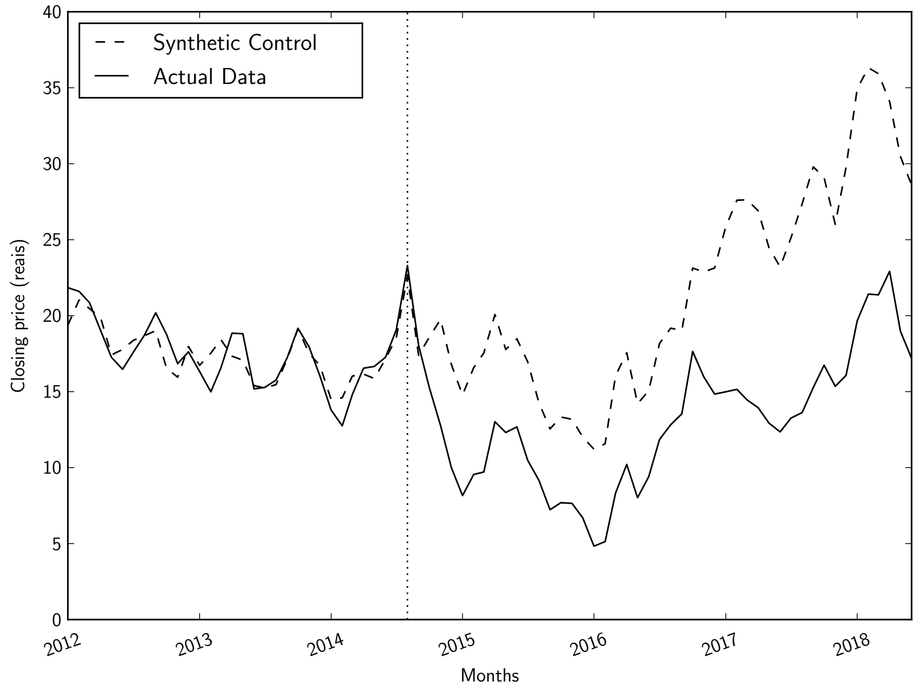
Figure **7** presents the evolution of the stock prices of Petrobras and synthetic Petrobras. After the scandal emerges, the two lines begin to diverge. The Petrobras stock price falls while the synthetic Petrobras stock price moves up, and until the end of the period we analyze, we notice no sign of recovery. When we look at the Petrobras stock prices only, we can imagine the price drops as a consequence of the scandal and then recover at the price level before the scandal. Nevertheless, when we compare with the evolution of the synthetic control, we notice a considerable difference in prices, which suggests that the recovery is associated with some systematic component rather than the end of the effect.

Even though we do not have a short-term effect media on the stock returns, as we evidence in Tables **14** and **15**, we evidence a long-term effect in prices following the corruption scandal. These results suggest that investors do not react to corruption news stories, but the scandal decreases the firm value.

Recall that we construct the synthetic Petrobras as a weighted average of potential controls including only public companies, which removes any natural effect of Petrobras being a state-owned company due to the context we are considering. Public companies stocks had a bad performance in the period after the operation "Car Wash" emerges (Padula and Albuquerque, 2018). Therefore, the cost of the corruption scandal we measure as the difference between the two lines in Figure **7** is only the cost of corruption in Petrobras. Our method does not capture the cost of corruption in the government.

The Petrobras scandal emerges in August 2014 with the information that Costa would give specific information to prosecutors about a corruption scheme in the company. Besides, our news database has a first news story about suspicious negotiations in August

Fig. 7. Petrobras Synthetic Control

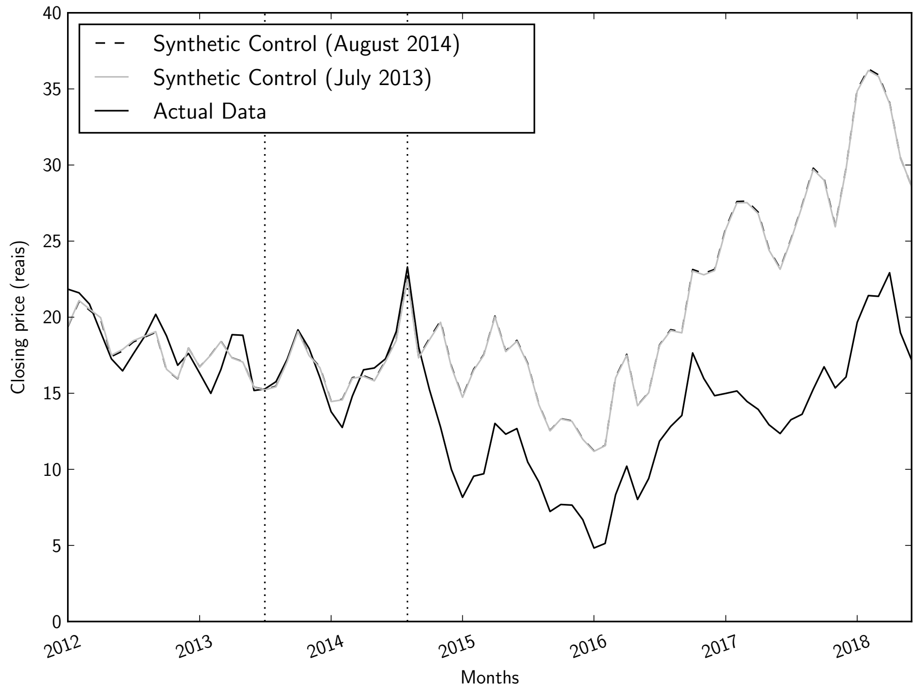


2013[19]. Since it could cause an anticipated reaction in the market, we plot in Figure **8** the synthetic control we obtain using as the event date that divides our sample in pre-treatment and post-treatment period, the first news story in our database about suspicion of a corruption scheme in Petrobras. The weights are similar for both cases. Although the market reacts after the schemes become public and news stories become more frequent, this test ensures that we do not have any anticipated movement before the corruption schemes become clear. In the following months, the prices of the Petrobras and the synthetic Petrobras stocks are similar until the scandal emerges, which demonstrates that the method works properly.

To ensure the decrease in the Petrobras stock price has a causal relation with the scandal, we perform a couple of tests suggested in the literature. First, to guarantee that the synthetic control reproduces the trajectory of the treated unit in the absence of the event, in our case the corruption scandal, Abadie and Gardeazabal (2003) suggest a placebo study that consists to suppose that the event happened in a unit that did not. Following Abadie, Diamond, and Hainmueller (2010), we construct a synthetic company for every company we use as a potential control assuming that the corruption scandal

[19]The news story can be accessed in `https://www.valor.com.br/empresas/3223210/compra-de-pasadena-foi-um-negocio-normal-diz-gabrielli`

Fig. 8. Petrobras Synthetic Control (Treatment in July 2013)



involved that company. If the corruption scandal did impact Petrobras, then we would notice a different effect from other companies that did not have a corruption scandal. In Figure **9** we plot the gap between the real and the synthetic evolution of the company stock price. We notice a persistent drop in the Petrobras stock price after the scandal that is not reproduced by any other unit, which indicates the gap we find in Figure **7** is due to the corruption scandal.

We also calculate the Root Square Mean Percentage Error (RMSPE) for every company we perform the synthetic control method, before and after the scandal. We plot in Figure **10** the ratio between the post-event RMSPE and the pre-event RMSPE. The RMSPE for Petrobras is the largest ratio among companies, more than three times larger than the second largest RMSPE. Since the RMSPE is an adjust measure between the two lines, this result indicates that the event generated an effect only on Petrobras stock price, which is in line with the result in Figures **7** and **8**.

We assume there is no insider trading that could drive the stock price and hence we consider that before the media reports the information about the corruption schemes, there is no effect on the stock prices. If the corruption practice benefits the company before the scandal and the firm value increase as a consequence, we might think that the

Fig. 9. Closing price gaps in Petrobras and placebo gaps in all control companies

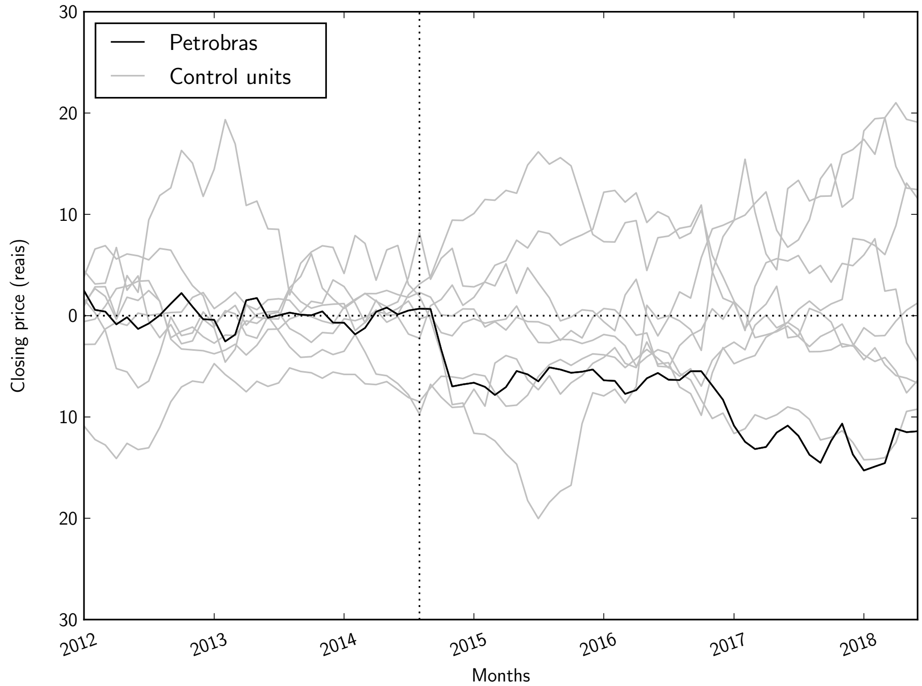
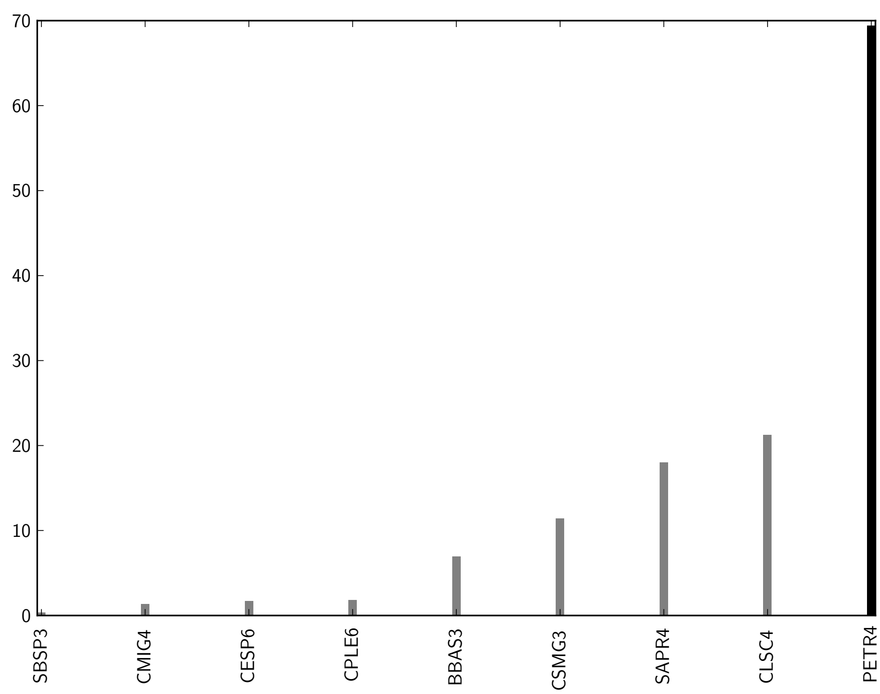


Fig. 10. post/pre-Petrobras scandal RMSPE



negative effect we estimate is actually smaller if we consider the situation of the company without any corruption scheme. Nevertheless, since we are comparing the company with companies with similar variables based on fundamentals, we assume this is not a concern.

4.5.2.2 JBS case

Table **17** presents the weights of each control firm in the synthetic JBS and the set of potential controls. Prior the scandal, the JBS stock price is best reproduced by a combination of mainly *WEG Industries* (56.3%), *Santander Brasil* (10.7%) and *MRV Engenharia* (10.3%).

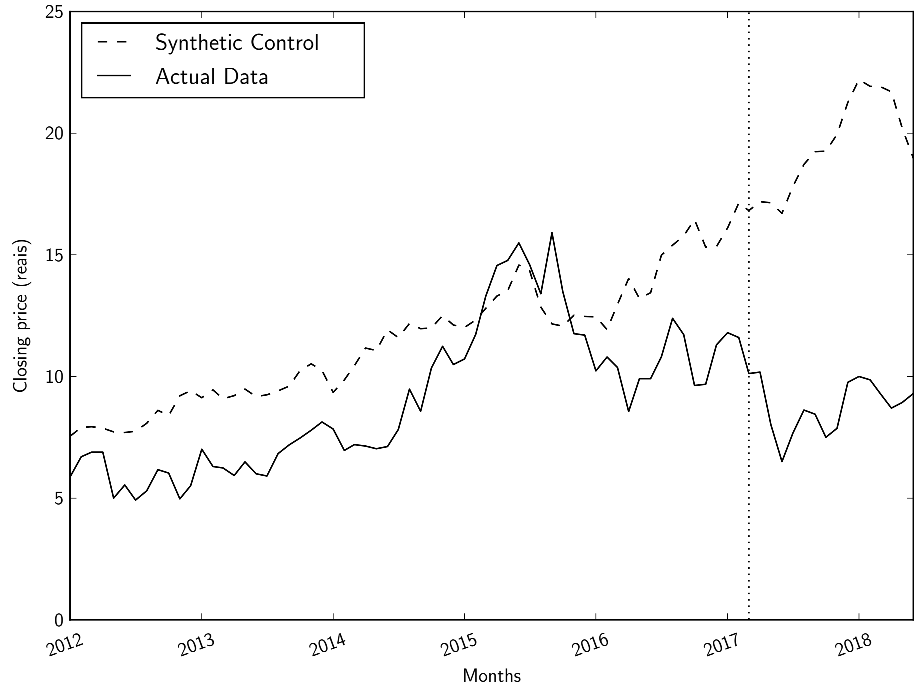
Table 17: Firm weights in the synthetic JBS

Company	Stock	Weight	Company	Stock	Weight
AmBev	ABEV3	0,005	Equatorial Energia	EQTL3	0,008
B3	B3SA3	0,006	Estácio Participações	ESTC3	0,012
BR Malls	BRML3	0,004	Fibria Celulose	FIBR3	0,004
Bradesco	BBDC3	0,008	Fleury	FLRY3	0,007
Bradespar	BRAP4	0,001	Iguatemi	IGTA3	0,005
Braskem	BRKM5	0,006	Localiza	RENT3	0,006
CCR	CCRO3	0,006	Lojas Renner	LREN3	0,005
Cielo	CIEL3	0,07	MRV Engenharia	MRVE3	0,103
Comgás	CGAS5	0,003	Multiplan	MULT3	0,005
Cosan	CSAN3	0,005	Natura & Co	NATU3	0,003
CPFL Energia	CPFE3	0,005	Santander Brasil	SANB11	0,107
Cyrela	CYRE3	0,004	CSN	CSNA3	0,004
EcoRodovias	ECOR3	0,003	Telefônica Brasil	VIVT3	0,003
Embraer	EMBR3	0,005	TIM Brasil	TIMP3	0,009
EDP Brasil	ENBR3	0,018	Vale	VALE3	0,002
ENGIE Brasil	EGIE3	0,004	WEG Industries	WEGE3	0,563

Figure **11** presents the evolution of the stock price of JBS and synthetic JBS. We notice a larger difference between the two lines after the operation "Weak Flesh" begins. We notice not only that the JBS stock price drops after the corruption scandal, but the synthetic JBS stock price increase, which evidences a real cost larger than simply taking the difference in prices before and after the scandal. This result indicates that corruption news stories not only impact the firm in the short-term, as we also evidence in Tables **14** and **15** results, but the negative impact in the long-term persists.

The "Weak Flesh" scandal emerged in March 2017. The information had obviously a negative impact on the firm value, but public information about corruption involving the

Fig. 11. JBS Synthetic Control

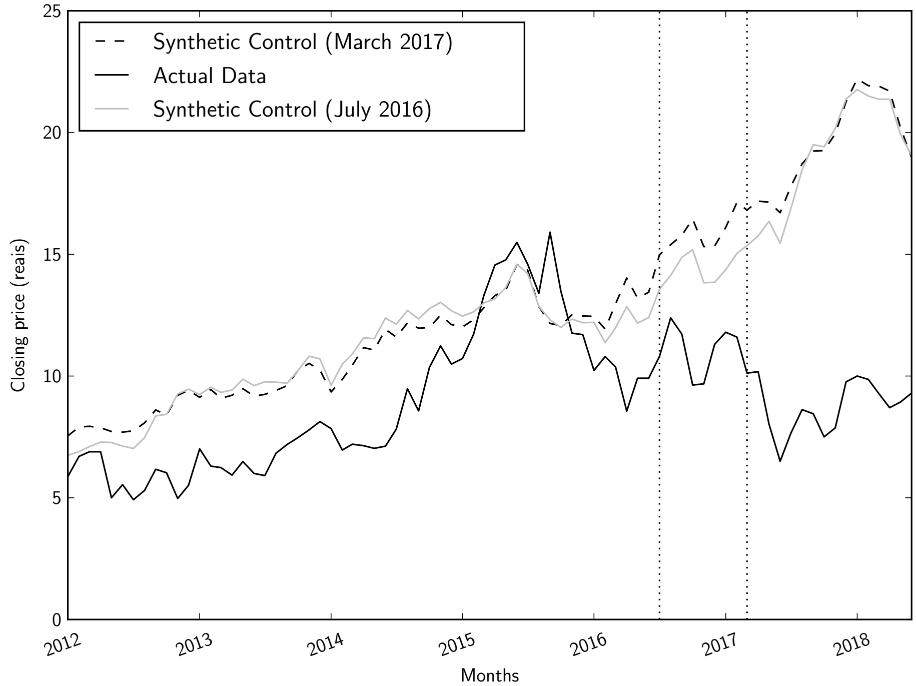


company and its owners began sooner. In July 2016 Federal police starts to investigate the holding of JBS in a new operation, and Joesley Batista, one of the JBS owners, was accused of paying bribes in exchange for receiving financial resources of a workers severance fund. To ensure that the effect we are estimating is due to the scandal, we estimate the weights for the control firms using the period before those news stories as the pre-treatment period. We plot the result in Figure **12**.

The lines begin to diverge before the "Weak Flesh" scandal emerges, which indicates that not only the scandal impacted the JBS stock prices, but also the evidence of corruption involving one of the owners and the holding that controls the company. When we consider the corruption investigations into the JBS owners and the holding that controls the company instead of the scandal date, the gap between the real evolution and synthetic unit is similar.

We apply the two tests as in Petrobras case to ensure the trajectory we are estimating in the absence of the corruption scandal is due to the corruption schemes. First, we run the synthetic control method for each control company we use to estimate the synthetic JBS as the corruption scandal happened involving that company. Figure **13** present the gap between the real evolution in stock prices and the respective synthetic control for

Fig. 12. JBS Synthetic Control (Treatment in July 2016)



each placebo test. The gap between the JBS and the synthetic JBS stock prices continues to drop after the event month with no recovery, which is a pattern not reproduced by any other placebo test, suggesting that the gap we observe after the event is not a coincidence and there is a causal effect of the corruption scandal in the JBS stock prices.

Figure **14** present the ratio between the post-event RMSPE and the pre-event RMSPE for JBS case and the placebo cases. The ratio for JBS is larger than the majotity of the control companies. Besides the sinthetic JBS weakly reproduces the JBS trajectory before the corruption scandal relative to the Petrobras case, the results we present in this paper suggest that there is a negative effect of the corruption scandal and the company stock prices that persist in the long-term.

4.6. Conclusion

Our paper investigates the effect of corruption scandals on stock prices. For this purpose, we examine two cases of Brazilian companies corruption scandals: the state-owned company Petrobras and the private company JBS. We find that there is a negative impact of the amount of corruption reported in the news stories about a company on their stock returns, but only in the case the company ownership is private. For the

Fig. 13. Closing price gaps in JBS and placebo gaps in all control companies

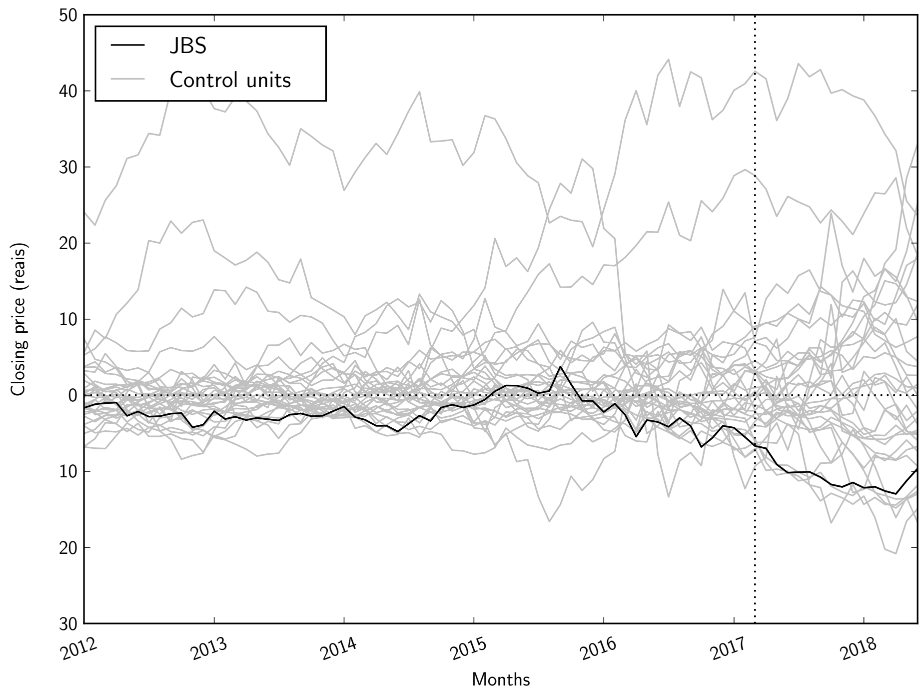
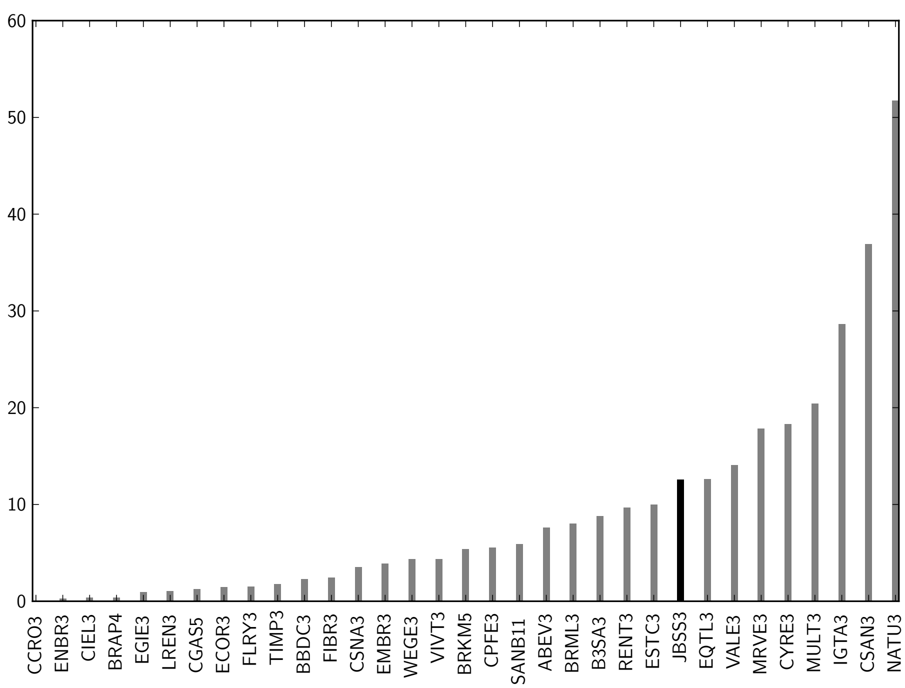


Fig. 14. post/pre-JBS scandal RMSPE



state-owned company, we do not evidence a significant effect. We also find a persistent drop in the stock prices after the corruption scandal for both cases with no recovery in the long-term.

There is a risk for the company relative to the decision to get into corruption schemes in exchange for benefits. We estimate in this study the cost for the company valuation when it has to face legal consequences. Evidence in Brazil points out that there is an incentive for the company to seek for political connections, and the investors expect the company benefits from this relation (Claessens et al., 2008). Our results do not contradict that study but instead complements it. We argue that investors presume the corruption is no longer lucrative after the scandal because the agents involved in the scheme can not continue with the practice, which makes not only the benefits from corruption ends, but the negative effects due to the scandal appear, as lower revenue and legal consequences.

References

Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of Californias Tobacco Control Program. Journal of the American Statistical Association 105, 493–505.

Abadie, A., Gardeazabal, J., 2003. The Economic Costs of Conflict: A Case Study of the Basque Country. American Economic Review 93, 113–132.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., Mitton, T., 2016. The value of connections in turbulent times: Evidence from the United States. Journal of Financial Economics 121, 368–391.

Agarwal, S., Chen, V. Y. S., Zhang, W., 2016. The Information Value of Credit Rating Action Reports: A Textual Analysis. Management Science 62, 2218–2240.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Amore, M. D., Bennedsen, M., 2013. The value of local political connections in a low-corruption environment. Journal of Financial Economics 110, 387–402.

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.

Antweiler, W., Frank, M. Z., 2006. Do US Stock Markets Typically Overreact to Corporate News Stories? SSRN Scholarly Paper ID 878091, Social Science Research Network, Rochester, NY.

Araújo, T., Eleutério, S., Louçã, F., 2018. Do sentiments influence market dynamics? A reconstruction of the Brazilian stock market and its mood. Physica A: Statistical Mechanics and its Applications 505, 1139–1149.

Armstrong, C. S., Banerjee, S., Corona, C., 2013. Factor-Loading Uncertainty and Expected Returns. The Review of Financial Studies 26, 158–207.

Audi, R., Loughran, T., McDonald, B., 2016. Trust, but verify: A language and the role of trust in corporate culture. Journal of Business Ethics pp. 551–561.

Bajo, E., Raimondo, C., 2017. Media sentiment and IPO underpricing. Journal of Corporate Finance 46, 139–153.

Baker, M., Wurgler, J., 2006. Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance 61, 1645–1680.

Baker, S. R., Bloom, N., Davis, S. J., 2016. Measuring Economic Policy Uncertainty. The Quarterly Journal of Economics 131, 1593–1636.

Bali, T. G., Brown, S. J., Tang, Y., 2017. Is economic uncertainty priced in the cross-section of stock returns? Journal of Financial Economics 126, 471–489.

Balvers, R. J., Gaski, J. F., McDonald, B., 2016. Financial Disclosure and Customer Satisfaction: Do Companies Talking the Talk Actually Walk the Walk? Journal of Business Ethics 139, 29–45.

Barber, B. M., Odean, T., 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. The Review of Financial Studies 21, 785–818.

Barboza, F., Kimura, H., Altman, E., 2017. Machine learning models and bankruptcy prediction. Expert Systems with Applications 83, 405–417.

Bekiros, S., Gupta, R., Majumdar, A., 2016. Incorporating economic policy uncertainty in US equity premium models: A nonlinear predictability analysis. Finance Research Letters 18, 291–296.

Box, T., 2017. Qualitative Similarity and Stock Price Comovement. SSRN Scholarly Paper ID 2139708, Social Science Research Network, Rochester, NY.

Brogaard, J., Detzel, A., 2015. The Asset-Pricing Implications of Government Economic Policy Uncertainty. Management Science 61, 3–18.

Brown, S. V., Tucker, J. W., 2011. Large-Sample Evidence on Firms Year-over-Year MD&A Modifications. Journal of Accounting Research 49, 309–346.

Buehlmaier, M. M. M., Whited, T. M., 2018a. Are Financial Constraints Priced? Evidence from Textual Analysis. The Review of Financial Studies 31, 2693–2728.

Buehlmaier, M. M. M., Whited, T. M., 2018b. Are Financial Constraints Priced? Evidence from Textual Analysis. The Review of Financial Studies 31, 2693–2728.

Bunkanwanicha, P., Wiwattanakantang, Y., 2009. Big Business Owners in Politics. The Review of Financial Studies 22, 2133–2168.

Carrière-Swallow, Y., Céspedes, L. F., 2013. The impact of uncertainty shocks in emerging economies. Journal of International Economics 90, 316–325.

Chen, C. R., Li, Y., Luo, D., Zhang, T., 2017. Helping hands or grabbing hands? An analysis of political connections and firm value. Journal of Banking & Finance 80, 71–89.

Claessens, S., Feijen, E., Laeven, L., 2008. Political connections and preferential access to finance: The role of campaign contributions. Journal of Financial Economics 88, 554–580.

Cohen, L., Frazzini, A., 2008. Economic Links and Predictable Returns. The Journal of Finance 63, 1977–2011.

Cooper, M. J., Gulen, H., Ovtchinnikov, A. V., 2010. Corporate Political Contributions and Stock Returns. The Journal of Finance 65, 687–724.

Das, S. R., Chen, M. Y., 2007. Yahoo! for amazon: Sentiment extraction from small talk on the web. Management Science 53, 1375–1388.

Davis, A. K., Piger, J. M., Sedor, L. M., 2012. Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. Contemporary Accounting Research 29, 845–868.

Dellavigna, S., Pollet, J. M., 2009. Investor Inattention and Friday Earnings Announcements. The Journal of Finance 64, 709–749.

Dougal, C., Engelberg, J., García, D., Parsons, C. A., 2012. Journalists and the Stock Market. The Review of Financial Studies 25, 639–679.

Faccio, M., 2006. Politically Connected Firms. American Economic Review 96, 369–386.

Fama, E. F., French, K. R., 1992. The Cross-Section of Expected Stock Returns. The Journal of Finance 47, 427–465.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fang, L., Peress, J., 2009. Media Coverage and the Cross-section of Stock Returns. The Journal of Finance 64, 2023–2052.

Ferguson, N. J., Philip, D., Lam, H. Y., Guo, J. M., 2015. Media Content and Stock Returns: The Predictive Power of Press. Multinational Finace Journal 19, 1–31.

Fisman, R., 2001. Estimating the Value of Political Connections. American Economic Review 91, 1095–1102.

Fraiberger, S. P., Lee, D., Puy, D., Rancire, R., 2018. Media Sentiment and International Asset Prices. National Bureau of Economic Research Working Paper Series .

García, D., 2013. Sentiment during Recessions. The Journal of Finance 68, 1267–1300.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. The MIT Press.

Gu, M., Sun, M., Wu, Y., Xu, W., 2018. Economic Policy Uncertainty and Momentum. SSRN Scholarly Paper ID 3133832, Social Science Research Network, Rochester, NY.

Gurun, U. G., Butler, A. W., 2012. Don't Believe the Hype: Local Media Slant, Local Advertising, and Firm Value. The Journal of Finance 67, 561–598.

Hendershott, T., Livdan, D., Schrhoff, N., 2015. Are institutions informed about news? Journal of Financial Economics 117, 249–287.

Heston, S. L., Sinha, N. R., 2017. News vs. Sentiment: Predicting Stock Returns from News Stories. Financial Analysts Journal 73, 67–83.

Hillier, D., Loncan, T., 2019. Political uncertainty and Stock returns: Evidence from the Brazilian Political Crisis. Pacific-Basin Finance Journal 54, 1–12.

Hoberg, G., Phillips, G., 2010. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. The Review of Financial Studies 23, 3773–3811.

Hoberg, G., Phillips, G., 2016. Text-Based Network Industries and Endogenous Product Differentiation. Journal of Political Economy 124, 1423–1465.

Hoberg, G., Phillips, G. M., 2017. Text-Based Industry Momentum. SSRN Scholarly Paper ID 2504738, Social Science Research Network, Rochester, NY.

Huang, A. H., Zang, A. Y., Zheng, R., 2014. Evidence on the Information Content of Text in Analyst Reports. The Accounting Review 89, 2151–2180.

Huberman, G., Regev, T., 2001. Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar. The Journal of Finance 56, 387–396.

Ibriyamova, F., Kogan, S., Salganik-Shoshan, G., Stolin, D., 2016. Using Semantic Fingerprinting in Finance. SSRN Scholarly Paper ID 2755585, Social Science Research Network, Rochester, NY.

Jegadeesh, N., Wu, D., 2013. Word power: A new approach for content analysis. Journal of Financial Economics 110, 712–729.

Karapandza, R., 2016. Stock returns and future tense language in 10-K reports. Journal of Banking & Finance 71, 50–61.

Kim, C. F., Pantzalis, C., Park, J. C., 2012. Political geography and stock returns: The value and risk implications of proximity to political power. Journal of Financial Economics 106, 196–228.

Lawrence, A., 2013. Individual investors and financial disclosure. Journal of Accounting and Economics 56, 130–147.

Lee, C. M. C., Shleifer, A., Thaler, R. H., 1991. Investor Sentiment and the Closed-End Fund Puzzle. The Journal of Finance 46, 75.

Lee, W. Y., Jiang, C. X., Indro, D. C., 2002. Stock market volatility, excess returns, and the role of investor sentiment. Journal of Banking & Finance 26, 2277–2299.

Lemmon, M., Portniaguina, E., 2006. Consumer Confidence and Asset Prices: Some Empirical Evidence. The Review of Financial Studies 19, 1499–1529.

Li, F., 2006. Do stock market investors understand the risk sentiment of corporate annual reports? SSRN Electronic Journal .

Li, F., 2008. Annual report readability, current earnings, and earnings persistence. Journal of Accounting and Economics 45, 221–247.

Li, F., 2010. The Information Content of Forward-Looking Statements in Corporate FilingsA Nave Bayesian Machine Learning Approach. Journal of Accounting Research 48, 1049–1102.

Liu, B., McConnell, J. J., 2013. The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? Journal of Financial Economics 110, 1–17.

Liu, L. X., Shu, H., Wei, K. C. J., 2017. The impacts of political uncertainty on asset prices: Evidence from the Bo scandal in China. Journal of Financial Economics 125, 286–310.

Lo, K., Ramos, F., Rogo, R., 2017. Earnings management and annual report readability. Journal of Accounting and Economics 63, 1–25.

Lou, D., 2014. Attracting Investor Attention through Advertising. The Review of Financial Studies 27, 1797–1829.

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.

Lundholm, R. J., Rogo, R., Zhang, J. L., 2014. Restoring the Tower of Babel: How Foreign Firms Communicate with U.S. Investors.

Maaten, L. v. d., Hinton, G., 2008. Visualizing Data using t-SNE. Journal of Machine Learning Research 9, 2579–2605.

Maćkowiak, B., Wiederholt, M., 2015. Business Cycle Dynamics under Rational Inattention. The Review of Economic Studies 82, 1502–1532.

Mai, F., Tian, S., Lee, C., Ma, L., 2019. Deep learning models for bankruptcy prediction using textual disclosures. European Journal of Operational Research 274, 743–758.

Mamaysky, H., Calomiris, C. W., 2018. How News and Its Context Drive Risk and Returns Around the World. Journal of Financial Economics .

Merton, R. C., 1973. An Intertemporal Capital Asset Pricing Model. Econometrica 41, 867–887.

Miller, B. P., 2010. The Effects of Reporting Complexity on Small and Large Investor Trading. The Accounting Review 85, 2107–2143.

Mironov, M., 2015. Should one hire a corrupt CEO in a corrupt country? Journal of Financial Economics 117, 29–42.

Nguyen, T. T., van Dijk, M. A., 2012. Corruption, growth, and governance: Private vs. state-owned firms in Vietnam. Journal of Banking & Finance 36, 2935–2948.

Ovtchinnikov, A. V., Pantaleoni, E., 2012. Individual political contributions and firm performance. Journal of Financial Economics 105, 367–392.

Ozoguz, A., 2009. Good Times or Bad Times? Investors' Uncertainty and Stock Returns. The Review of Financial Studies 22, 4377–4422.

Padula, A. J. A., Albuquerque, P. H. M., 2018. Government corruption on Brazilian capital markets: A study on Lava Jato (Car Wash) investigation. Revista de Administração de Empresas 58, 405–417.

Pan, X., Tian, G. G., 2017. Political connections and corporate investments: Evidence from the recent anti-corruption campaign in China. Journal of Banking & Finance .

Pantzalis, C., Park, 2014. Too close for comfort? Geographic propinquity to political power and stock returns. Journal of Banking & Finance 48, 57–78.

Pástor, L., Veronesi, P., 2012. Uncertainty about Government Policy and Stock Prices. The Journal of Finance 67, 1219–1264.

Pástor, L., Veronesi, P., 2013. Political uncertainty and risk premia. Journal of Financial Economics 110, 520–545.

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics 80, 563–602.

Pennington, J., Socher, R., Manning, C. D., 2014. Glove: Global vectors for word representation. In: *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.

Peress, J., 2014. The Media and the Diffusion of Information in Financial Markets: Evidence from Newspaper Strikes. The Journal of Finance 69, 2007–2043.

Phan, D. H. B., Sharma, S. S., Tran, V. T., 2018. Can economic policy uncertainty predict stock returns? Global evidence. Journal of International Financial Markets, Institutions and Money 55, 134–150.

Schmeling, M., 2009. Investor sentiment and stock returns: Some international evidence. Journal of Empirical Finance 16, 394–408.

Segal, G., Shaliastovich, I., Yaron, A., 2015. Good and bad uncertainty: Macroeconomic and financial market implications. Journal of Financial Economics 117, 369–397.

Shleifer, A., Vishny, R. W., 1993. Corruption. The Quarterly Journal of Economics 108, 599–617.

Sims, C. A., 1998. Stickiness. Carnegie-Rochester Conference Series on Public Policy 49, 317–356.

Smales, L. A., 2015. Asymmetric volatility response to news sentiment in gold futures. Journal of International Financial Markets, Institutions and Money 34, 161–172.

Solomon, D. H., 2012. Selective Publicity and Stock Prices. The Journal of Finance 67, 599–637.

Solomon, D. H., Soltes, E., Sosyura, D., 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. Journal of Financial Economics 113, 53–72.

Stambaugh, R. F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302.

Starks, L. T., Sun, S. Y., 2016. Economic Policy Uncertainty , Learning and Incentives : Theory and Evidence on Mutual Funds.

Sun, L., Najand, M., Shen, J., 2016. Stock return predictability and investor sentiment: A high-frequency perspective. Journal of Banking & Finance 73, 147–164.

Tetlock, P. C., 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance 62, 1139–1168.

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.

Tsai, F.-T., Lu, H.-M., Hung, M.-W., 2016. The impact of news articles and corporate disclosure on credit risk valuation. Journal of Banking & Finance 68, 100–116.

Wang, F., Xu, L., Zhang, J., Shu, W., 2018. Political connections, internal control and firm value: Evidence from China's anti-corruption campaign. Journal of Business Research 86, 53–67.

Wang, L., 2015. Protection or expropriation: Politically connected independent directors in China. Journal of Banking & Finance 55, 92–106.

Xiong, X., Bian, Y., Shen, D., 2018. The time-varying correlation between policy uncertainty and stock returns: Evidence from China. Physica A: Statistical Mechanics and its Applications 499, 413–419.

Xu, Y., 2018. Anticorruption regulation and firm value: Evidence from a shock of mandated resignation of directors in China. Journal of Banking & Finance 92, 67–80.

Zhang, X. F., 2006. Information Uncertainty and Stock Returns. Journal of Finance 61, 105–137.