**CHARLES UNIVERSITY**

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

**Mária Partelová**

# Presidential rhetoric, sentiment and their relation to stock markets

*Bachelor Thesis*

Prague 2017

**Author**: Mária Partelová

**Supervisor**: PhDr. Boril Šopov, MSc., LL.M.

**Academic Year**: 2016/2017

## Bibliographic note

PARTELOVÁ, Mária. *Presidential rhetoric, sentiment and their relation to stock markets.* Prague 2017. 74 pp. Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor: PhDr. Boril Šopov, MSc., LL.M.

## Abstract

This thesis intends to uncover the linkages between the emotions contained within remarks of the president of the United States expressed on Twitter and movements of the stock market indices. The daily comments of the two consecutive presidents, Barack Obama and Donald Trump are annotated with sentiment intensity values using the lexicon-based model called VADER. Our analysis further focuses on testing for Granger causality using the bivariate vector autoregression. Overall, three major stock market indices are employed in testing, namely DJIA, S&P 500 and NASDAQ. The results yield a statistically significant Granger causal relationship in the case of the first differences of DJIA and S&P 500 logarithms with time series of Barack Obama's sentiment values.

| | |
|---|---|
| **Keywords** | sentiment analysis, stock markets, Granger causality, presidential rhetoric, natural language processing |
| | |
|---|---|
| **Author's e-mail** | `mpartelova@gmail.com` |
| **Supervisor's e-mail** | `boril.sopov@gmail.com` |

## Bibliografický záznam

PARTELOVÁ, Mária. *Presidential rhetoric, sentiment and their relation to stock markets.* Praha 2017. 74 stran. Bakalářská práce (Bc.) Univerzita Karlova, Fakulta sociálních věd, Institut ekonomických studií. Vedoucí práce: PhDr. Boril Šopov, MSc., LL.M.

## Abstrakt

Tato práce je zaměřená na odkrytí vztahů mezi emocemi obsaženými v příspěvcích prezidenta Spojených států amerických na Twitteru a pohyby indexů akciových trhů. Komentářům na denní bázi dvou po sobě jdoucích amerických prezidentů, Baracka Obamy a Donalda Trumpa, jsou přiřazeny hodnoty intenzity jejich sentimentu použitím lexicon-based modelu VADER. Tato analýza se dále zabývá testováním Grangerovy kauzality pomocí vektorové autoregrese s dvěmi proměnnými. Tři hlavní akciové indexy jsou použity v této práci, DJIA, S&P 500 and NASDAQ. Výsledky přináší statisticky významnou Grangerovu kauzalitu v případě vztahu prvních diferencí logaritmů DJIA and S&P 500 a časovou řadou sentimentu Baracka Obamy.

**Klíčová slova** analýza sentimentu, akciové trhy, Grangerova kauzalita, prezidentská rétorika, zpracování přirozeného jazyka

**E-mail autora** `mpartelova@gmail.com`

**E-mail vedoucího práce** `boril.sopov@gmail.com`

## Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and that this work has not been used to gain any other title.

The author grants the Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, 19 May 2017

Signature

## Acknowledgement

I would like to express my sincere gratitude to my supervisor, PhDr. Boril Šopov, MSc., LL.M., for his remarks and valuable guidance, without which this thesis would not be possible.

# Bachelor Thesis Proposal

| | |
|---|---|
| **Author** | Mária Partelová |
| **Supervisor** | PhDr. Boril Šopov, MSc., LL.M. |
| **Proposed topic** | Presidential rhetoric, sentiment and their relation to stock markets |

**Topic characteristics:**

A vast number of researchers have been concerned with the question regarding the factors that have predictive power on stock markets. Numerous studies emerged over the last decades trying to relate the stock market‘s movement to sentiment extracted from the big data. The focus of their research have been aimed mainly at the public mood values hidden within the textual content posted on social networks, such as microblogging website Twitter. Moreover, the presidential rhetoric on the social networks has received a lot of the public attention recently, and especially the Twitter feed of the incumbent President of the United States, Donald Trump, has appeared uncountable times in the headlines of the newspapers all over the world. With the degree of power that is so characteristic for every President of the United States, it might be worth studying if there is a relation of his remarks posted online and the stock indexes‘ movements. Various studies have examined the rhetoric of the President of the US and its impact on economy, but none of them has been concerned with his Twitter feed and its relation to stock markets so far. Thus, the goal of this thesis is to fill this gap and examine the Twitter time lines of two consecutive US Presidents, Barack Obama and Donald Trump. Not only is the study concerned with the textual analysis and comparison of the characteristics of their online rhetoric using Natural Language Processing techniques, but also tries to relate the emotional states extracted from their tweets to the selected stock indexes.

**Methodology:**

Firstly, the VADER, lexicon – based model for sentiment analysis would be employed on the Twitter data sets over the course of their presidential mandate, for which the data is available of two consecutive US presidents, Barack Obama (@BarackObama) and Donald Trump (@realDonaldTrump). The Granger causality analysis of the compound sentiment time series aggregated into trading days in the data sets and three stock indexes, namely DJIA, S&P 500, and NASDAQ would be carried out consequently.

**Hypotheses:**

1. The sentiment extracted from tweets posted by Donald Trump Granger causes the time series of selected stock indexes.
2. The sentiment extracted from tweets posted by Barack Obama Granger causes the time series of selected stock indexes.

**Outline:**

1. Introduction
2. Literature review
3. Methodology
4. Results and interpretation
5. Conclusion

**Core bibliography:**


1. Bird, Steven, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* O'Reilly Media, Inc., 2009.

2. HUTTO, Clayton J., and Eric GILBERT. *Vader: A parsimonious rule-based model for sentiment analysis of social media text.* Eighth international AAAI conference on weblogs and social media. 2014.
3. BOLLEN, Johan, Huina MAO, and Xiaojun ZENG. *Twitter mood predicts the stock market.* Journal of computational science 2.1 (2011): 1-8.
4. MITTAL, Anshul, and ARPIT Goel. *Stock prediction using twitter sentiment analysis.* Stanford University, CS229 (`http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf`) 15 (2012).
5. WOOLDRIDGE, Jeffrey M. *Introductory econometrics: A modern approach.* Nelson Education, 2015.

# Contents

# List of Tables

# List of Figures

# Acronyms

| | |
|---|---|
| **DJIA** | Dow Jones Industrial Average |
| **S&P** | Standard & Poor's |
| **NASDAQ** | National Association of Securities Dealers Automated Quotations |
| **VADER** | Valence Aware Dictionary for sEntiment Reasoning |
| **VAR** | vector autoregression |
| **BIC** | Bayesian information criterion |
| **AIC** | Akaike information criterion |
| **ML** | machine learning |
| **OM** | opinion mining |
| **SA** | sentiment analysis |
| **NLP** | natural language processing |
| **NLTK** | Natural Language Toolkit |
| **SA** | sentiment analysis |
| **NN** | neural network |
| **OF** | Opinion Finder |
| **GI** | General Inquirer |
| **LIWC** | Linguistic Inquiry and Word Count |
| **ANEW** | Affective Norms for English Words |
| **POMS** | Profile of Mood States |
| **GPOMS** | Google Profile of Mood States |
| **EMH** | Efficient market hypothesis |
| **SVM** | support vector machines |

# Introduction

The question regarding the factors that have predictive power over the stock markets is as old as the stock markets themselves. Rather surprising events have been hypothesized to have an impact on the stock prices recently, for instance the “Hathaway” effect, which suggests that every time a popular actress Anne Hathaway makes headlines, the stock of Berkshire Hathaway Inc. rises. Furthermore, a lot of discussion has been devoted to the topic concerning the influence of the incumbent president of the United States over the stock prices primarily through the remarks posted on his popular communication channel, a microblogging website Twitter.

Moreover, a vast amount of studies arose over the course of the last years in an attempt to unveil a relationship between the sentiment retained from the textual comments of people on the social networks and the movements of the stock returns. The prevailing stream of research is aimed to detect the relationship between public opinion and stock indexes employing sentiment analysis on a random dataset of tweets downloaded from Twitter. Another widely applied approach is a topic-based study conducted on the tweets' sentiment of people mentioning a particular company on Twitter and its impact on that company's stock.

This thesis diverts from the main course of research in such a way that it attempts to assess the linkages between the comments of two consecutive presidents of the United States posted on their Twitter accounts and the three major stock indexes. More specifically, those are DJIA, S&P 500, NASDAQ and the studied presidents are Donald Trump and Barack Obama. The tweets are assessed with values produced by extensively used natural language processing technique, lexicon-based sentiment analysis and the output is further used in the bivariate VAR model alongside the first differences of the stock indexes logarithms applying the Granger causality testing. To our best knowledge, there has been no such research conducted yet.

The study is built on the premise that the enormous degree of power and attention concerning the president of the United States is an adequate reason for deeper analysis of his textual statements and their impact. Furthermore, the nowadays' popular means of communication, such as Twitter feeds, provide an effective way of obtaining the desired large data sets.

The thesis is organized in a following way. Chapter 1 is devoted to the literature review concerning the factors that cause the stock market movements, sentiment analysis and its applications, and the summary of up-to-date studies on presidential rhetoric. In Chapter 2, the methodology applied in the thesis is explained with a focus on the data sets, VADER, the lexicon-based model employed for the sentiment analysis and the Granger causality approach. The data sets' characteristics, the sentiment analysis time series and the results of Granger causality tests are described in Chapter 3. Finally, Chapter 4 summarises the main conclusions emerging from our analysis and proposes further extension of our research.

# Chapter 1

# Literature Review

## 1.1 Forces that move stock prices

### 1.1.1 Early work

#### Random walk vs. chartist theories

Researchers proposed the question regarding the factors that move the stock prices countless times throughout the history of the stock markets. Firstly, two main categories of theories based on different assumptions about the past behavior of the stock prices were developed. The random walk theory suggests that the "future path of the price level of a security is no more predictable than the path of a series of cumulated random numbers" (Fama, 1965). The changes in stock price are i.i.d. random variables and therefore, the analysis of the past shifts in the stock prices with the purpose of future values' prediction is pointless. Oppositely, the chartist theories are built on the premise that the history repeats itself in such a way that the hidden "patterns" in historic prices of security occur over and over (Fama, 1965). Hence, the future stock price can be predicted by applying precise analysis of its past values.

#### Efficient Market Hypothesis

Efficient Market Hypothesis (EMH) was postulated by Nobel-prize awarded economist Eugene Fama in 1970 and states that the security prices always "fully reflect" all available information in the market. It is defined as follows:

$$E(\tilde{p}_{j,t+1}|\Phi_t) = [1 + E(\tilde{r}_{j,t+1}|\Phi_t)]p_{jt},$$

where $\tilde{p}_{j,t+1}$ is the price of a security $j$ at time $t+1$, $\tilde{r}_{j,t+1}$ is the one-period percentage return, $\Phi_t$ is the symbol for information set to be "fully reflected" in the price at time $t$, $E$ is the sign for expected value (Fama, 1970). He further differentiates between three forms of market efficiency. The weak form assumes that stock prices reflect only the historical prices, the semi-strong form proposes that the prices also respond to changes in publicly available information (e.g. earnings announcements etc.) and the strong form adds that the prices reflect the insider information as well.

### 1.1.2 General classification of prediction approaches

#### Technical analysis

Technical or "chartist" analysis aims to predict the stock price by employing various pieces of the related past information, such as purchasing and selling price or volume traded (Nazário et al., 2017). Thus, the basic assumption of this technique is the dependence of consecutive stock price shifts and the endogenous character of stock prices. Most professionals in the field of stock market prediction consider the pure chartist methods suspicious since they do not take any other possible factors into account than the price or volume itself (Fama, 1965). The study on economic forces and stock markets by N.-F. Chen et al., 1986 states: "The co-movements of asset prices suggest the presence of underlying exogenous influences, but we have not yet determined which economic variables, if any, are responsible".

#### Fundamental analysis

Fundamental analysis of the stock price of a particular company is focused on quantitative and qualitative study of that company's financial statements, its overall functioning, and the market in which it operates. It is defined as "aimed at determining the value of corporate securities by a careful examination of key value-drivers, such as earnings, risk, growth, and competitive position" (Lev & Thiagarajan, 1993). The goal is to determine the fair or intrinsic value of a stock since it is based on the assumption that the stock is either overpriced or undervalued, and that it will converge to the intrinsic value over time (Suresh, 2013).

#### Analysis based on data mining and machine learning

Data mining can be described as a search for valuable information in large volumes of data (Weiss & Indurkhya, 1998), which may be used as the input in the stock market prediction and include Google Trends (Preis et al., 2013), Wikipedia searches count (Moat et al., 2013), data generated by social networks, such as Facebook (Karabulut, 2013) or Twitter (Bollen et al., 2011) among others. Machine learning provides computers "the ability to learn without being explicitly programmed" (Samuel, 1960). The related techniques have developed their own branch of applied artificial intelligence since the 1960s (Liao et al., 2012). The majority of recent research in the field employs Machine Learning models such as Artificial Neural Networks, Support Vector Machines(Madge, n.d.), k-Nearest Neighbour classifiers (Alkhatib et al., 2013) or Decision Tree Learning (Qian & Rasheed, 2007).

### 1.1.3 Market sentiment in the spotlight

Theories of economic choice are mostly built on the assumption that the individual's decisions are derived from rational consideration of future outcomes and cost-benefit analysis (Loewenstein et al., 2001). However, the fields of neurology and behavioural economics suggest that the behaviour of individuals is not completely rational. The somatic marker hypothesis postulates that decision-making is a process impacted by marker signals, including the ones that arise from emotions, and that this process can be either conscious or non-conscious (Bechara & Damasio, 2005).

A number of studies have emerged over the last years trying to predict the stock market movement based on the aggregated market sentiment or entity-related sentiment. Papers relating emotions to stock returns using the Twitter feeds' data include Bollen et al., 2011; Mittal & Goel, 2012; Pagolu et al., 2016; Zhang et al., 2011; Ranco et al., 2015; Kim et al., 2014 or Kordonis et al., 2016. Furthermore, researchers predicting stock market movements detained sentiment data also from Google Trends (Brochado et al., 2016), Yahoo Message Board (Nguyen & Shirai, 2015; Das & M. Y. Chen, 2007), Facebook (Karabulut, 2013; Siganos et al., 2014) or daily news articles (Wong & Ko, 2016; Schumaker et al., 2012).

## 1.2 Natural language processing with a focus on sentiment analysis

### 1.2.1 General description and models

Natural Language Processing (NLP) in a broad sense means managing the language people use in daily communication by a computer in any way one finds useful. It has expanded heavily with its applications, such as word frequency analysis, translations, emotion detection or comprehension of a language and attempts to response accordingly (Bird et al., 2009). Sentiment analysis or Opinion Mining refer to the assessment of feelings from a piece of text in a form of a numerical value.

#### Lexicon-based models

The lexicon-based approach relies on a predefined dictionary of the words that express some kind of emotion and are labelled accordingly. These are called *opinion lexicons* and via a function of opinion words in a text, these models provide either simple polarity classification (positive vs. negative) or multidimensional emotional classification. Furthermore, some of them are also able to extract the intensity of emotions hidden in a given text.

#### Machine Learning-based models

The machine learning based approach is applied in order to train a sentiment classifier using various features. The frequency of n-grams related to specific emotion belongs to the most applied features and these are extracted with the Bag-of-words model. N-grams are widely used in the field of sentiment analysis and can be described as all the possible combinations of n consecutive words from a text, i.e. the unigrams represent individual words, bigrams consist of pairs of words standing next to each other in a text, etc. The Bag-of-words model is aimed to create a set or "bag" of all n-grams in a text and then, to count the individual frequency of each one in separate subtexts. The mostly employed techniques for the classification itself include Naive Bayes, Support Vector Machines and Maximum Entropy (Khan et al., 2015). This approach requires extensive manual labelling of the training data – words, sentences or pieces of text, which is costly and time-consuming.

**Hybrid models**

The lexicon-based approach can be combined with machine learning so that labelling of the text for the training set of a classifier is done rather by the opinion lexicon than manually.

### 1.2.2 Sentiment analysis for stock market prediction

A number of papers have emerged over a last decade trying to relate the public sentiment to stock markets movements. In the influential study by Bollen et al., 2011 10 million public tweets by approximately 2,7 million users over 10-month period were collected and then processed using two sentiment tools, OpinionFinder (OF) for polarity classification and Google-Profile of Mood States (GPOMS), for 6-dimensional classification within "Calm", "Alert", "Sure", "Vital", "Kind" and "Happy". The bivariate Granger causality analysis was carried out based on the equations:

$$L_1 : D_t = \alpha + \sum_{i=1}^{n} \beta_i D_{t-i} + \epsilon_t,$$

$$L_2 : D_t = \alpha + \sum_{i=1}^{n} \beta_i D_{t-i} + \sum_{i=1}^{n} \gamma_i X_{t-i} + \epsilon_t,$$

where $D_t$ is the DJIA value at time $t$, $X_{t-i}$ are the values of each of the mood time series at the time $t-i$ and $n$ is the number of lags applied. The null that the mood time series does not predict the DJIA values was rejected when applied to GPOMS mood dimension "Calm" time series and a 3-day lag. Finally, a five-layer hybrid self-organizing fuzzy neural network (SOFNN) with the input including combination of DJIA values and "Calm" mood values of the past 3 days yielded 86% direction accuracy.

Goel and Mittal used only 4-dimensional classification, whereas four models were employed in order to predict the stock prices, namely linear regression, logistic regression, SVMs and SOFNN. Similarly to Bollen et al., 2011, they found that SOFNN performs the best, but the highest accuracy of 76,56% was observed with the combination of mood time series "Calm" and "Happy".

In the study by Zhang et al., 2011 all the tweets containing emotional words, such as "fear", "worry", "hope" among others on each day were summed up and a collective measure of positive and negative emotions was extracted. The correlation analysis of these groups with the use of total number of tweets as a baseline against DJIA, NASDAQ and S&P 500 was conducted and yielded strong results supporting the hypothesis that as people express the negative emotions the indexes go down and they rise when the positive emotions and less negative emotions are present on Twitter, respectively.

Other stream of studies focused on sentiment extraction from the tweets mentioning a specific company and analysed impact of this sentiment on the stock prices of that particular company. Pagolu et al., 2016 derived their own sentiment classifier based on the human-annotated training set of tweets and predicted the rise or fall of stock prices with 69,01% accuracy when classifier was trained using logistic regression algorithm and 71,82% accuracy, when trained with SVM. A strong relationship among the public emotions regarding a company and a rise or fall of stock prices of that company was concluded. Ranco et al., 2015 studied relation between stock returns and sentiment retained from the tweets about 30 DJIA companies. Granger causality test was carried out after the correlation of the sentiment time series and price return time series was calculated. However, the correlation values were low and majority of the companies did not pass the Granger causality test. The "event study" was employed consequently and a statistically significant relationship between the stock returns and sentiment extracted from Twitter was found.

### 1.2.3 Other usage

Many meaningful studies surged in the field of the NLP apart from stock market analysis, such as box-office revenues forecast based on movie reviews (Asur & Huberman, 2010), election prediction (Tumasjan et al., 2010), measuring the emotions towards popular brands (Mostafa, 2013), prediction of the number of Influenza-like-Illness cases with the use of Twitter (Achrekar et al., 2011), examining sentiment towards tobacco products (Myslín et al., 2013), personal characteristics detection (Golbeck et al., 2011), predicting the spread of depression (De Choudhury et al., 2013b) or suicidal behavior (Brochado et al., 2016), analysis of emotional states of mothers that gave birth (De Choudhury et al., 2013a), measuring national happiness through the social media content (Durahim & Coşkun, 2015) and an overwhelming number and variety of others.

## 1.3 Presidential rhetoric and its impact

### 1.3.1 Presidential power in the United States

The fact that the political system in the United States is a federal presidential constitutional republic indicates that the President of the United States has a substantial degree of executive power at his disposal. Not only names he the members of the Cabinet and has the right to veto bills, but he also is the commander-in-chief of the military. Moreover, "none can compete effectively with the president in terms of prestige, status, media access, public attention and interest" (Cohen, 1995) and the public seems to "rely psychologically" on the president (Greenstein, 1974). The above-stated implies he is a dominant public figure and the content of his remarks might have an impact on the public sentiment or the economy and the market itself.

### 1.3.2 Studies on the presidential rhetorics

Early work was concerned with the State of the Union presidential speeches and their impact on the citizens. Cohen, 1995 found empirical evidence in favor of a significant influence, but Young & Perkins, 2005 concluded that the presidential rhetoric has been losing its power with the "End of the Golden Age of Television" and president has to find new channels to be heard. The study by Wood et al., 2005 examined optimism within the content of all public statements of the U.S. President, and how it might possibly alter public sentiment using vector autoregression model. It was found that the presidential comments heavily influence public views and confidence about the economic conditions. The relationship between public, president and media was also studied with the vector autoregression by Lee, 2014, where the statistically significant connection between media and both the president and the public was observed. Oppositely, the president and the public were found not to interact directly.

A recent study (Chin et al., 2016) analysed emotions of the public towards presidential candidates in the United States presidential elections 2016 based on the related tweets and employing machine learning classifiers and provided a great set of visualizations. However, to my best knowledge, there has been no paper written on the sentiment contained in presidential remarks posted via social networks and their relation to stock market.

# Chapter 2

# Methodology

## 2.1 Data

### 2.1.1 Twitter data sets

The data were collected from Twitter, popular microblogging website founded on 21st of March 2006 with approximately 313 million monthly active users as of June 2016 (*Twitter usage: Company facts*, 2016), who can express their opinions in a form of written updates called tweets, follow other users or repost others' statuses. The tweets can include a photo or a link to website and can not exceed the length of 140 characters. The study on the purposes of the tweets (Java et al., 2007) depicts four major ones, i.e. posts about current actions, public conversations, news commentary and sharing photos or links with additional comments. All of these purposes can transmit user's subjective feelings and mood.

The tweets were obtained using the Twitter API in JSON format containing various pieces of information, such as date, time, core text or number of retweets among many others. The full decomposition of a tweet with an example is provided in the Table 5 in the Appendix.

The posts from Twitter feeds of private accounts of two successive American presidents, Barack Obama and Donald Trump were used in this study. The accounts' user names are @BarackObama and @realDonaldTrump. Twitter limits downloading of the feeds based on criteria related to allowed time span and number of tweets. Therefore, the maximum dataset for Barack Obama contained 2598 tweets over the period from the 16th of December 2014 until the 23rd of March 2017 and for Donald Trump it included 3242 tweets over the period from the 3rd of April 2016 until the 6th of April 2017. These sets were employed within the frequency analysis and sentiment analysis. However, in the case of Granger Causality analysis, the data sets were shortened in such a way that they correspond to the period relevant to the power possession of each of the presidents and thus, potential influence of the president's comments on public.

### 2.1.2 Stock indexes' time series

Three American stock market indices were used in this study, specifically the S&P 500, the DJIA and the NASDAQ Composite. The S&P 500 consists of 500 large companies, whose stocks are listed on NASDAQ or NYSE and it provides a great reflection of the market due to this variety. The DJIA includes 30 major US firms, whereas the NASDAQ Composite Index contains around 3000 equities and is known for being weighted towards technology securities. The time series of the closing prices over the time spans corresponding to the Twitter data sets were downloaded from the Yahoo! Finance (finance.yahoo.com).

## 2.2 Getting to know the data and the frequency analysis

For the purposes of the better understanding of data characteristics and their visualization, various simple statistical measures were computed (see Table 2.1) and the distributions based on aspects such as the posts' time of day, week day or the daily tweet count were plotted for both studied presidents.

**Table 2.1:** List of computed statistical measures

| **Computed measure** |
|---|
| Average of the daily tweets |
| Mode of the daily tweets |
| Median of the daily tweets |
| Maximum of the daily tweets |
| Standard deviation of the daily tweets |
| Sample variance of the daily tweets |
| Skewness of the daily tweets' distribution |
| Average of the daily retweets |
| Median of the daily retweets |
| Maximum of the daily retweets |

*Source: Author.*

The frequency analysis was carried out consequently, which means that the distribution of the words used most often was plotted for both data sets. In order to compile a list of meaningful words and number of times they were used, the data had to undergo certain preprocessing procedures with the help of the Natural Language Toolkit (NLTK), extensive Python library created for Natural Language Processing tasks (Bird et al., 2009):

1. **Removal of several Twitter-related features**
   Apart from words and punctuation, a tweet may contain link to an article or web page, the name of a user account, to which the message is addressed or XML special values. Firstly, the patterns were found, based on which these features were detected within the text and then removed. For instance, the name of other user's account always starts with: "@", URL begins with "http" and a XML special value starts with "&" and ends with ";". Furthermore, the punctuation was eliminated as well.

2. **Lower-case conversion**

   The text was converted to lower-case characters consequently, so that our frequency distribution is not distorted by counting the same word as two distinct words, just because one of them would contain an upper-case character.

3. **Tokenization**

   Every tweet's text was split into parts, the so-called tokens, based on the location of spaces in the text. The resulting tokens are either words or key phrases written in a popular conjuncted manner known as hashtags, i.e. #MakeAmericaGreatAgain.

4. **Stop-words removal**

   As a last step, the words with no special meaning and value added were removed from the data sets. This was done using the list of so-called Stop-words covered by NLTK and this full list is provided in the Table 4 in the Appendix .

Conducting these procedures, we are left with individual words that have a semantic meaning. The frequency distribution of these words was plotted for each studied person and analysed.

## 2.3 Sentiment analysis

### 2.3.1 Problem definition and assumptions

The initial task of our study is to perform the sentiment analysis in order to extract an accurate emotion from each of downloaded tweets for both data sets. Much of the existing work is concerned with a withdrawal of a collective sentiment out of large data sets. However, this thesis is aimed to study a single-person generated textual data. The key fact, which differentiates the studied person from the others, is that as the President of the United States he holds one of the most influential positions worldwide. He has a substantial degree of power, that concerns not only the national economy and lives of the US citizens, but also the global stability since he has the right to declare a war against another country among other competencies. He stands in the spotlight of media and public interest. Therefore, there might be meaningful and valuable insights contained in his public statements.

The issue encountered at this point is to determine if Twitter as a communication channel is relevant for the study. At least in the case of the incumbent president, Donald Trump, his Twitter account has been in the centre of public attention, whereby his posts have appeared in the headlines of various newspapers and magazines numerous times. Some of the publishers have even created a separate section for the news about his tweets on the newspapers' web pages. Oppositely, in case of Obama, there had not been such a media interest in his Twitter feed during the time he was the president. However, the number of followers of Obama's account is 87,6 million as of April 2017, what could be taken as a valid point for evaluation of his textual statuses.

#### Opinion definition

Opinion can be defined as a quadruple,

$$(g_i, s_{ijk}, h_j, t_k), \quad \text{where} \quad i, j, k \in \mathbb{N},$$

$g_i$ is the opinion target, $s_{ijk}$ is the sentiment regarding the target, $h_j$ is the holder of the opinion and $t_k$ is the time when the opinion was expressed (Liu, 2012).

In our case $g_i$ is the subject of the opinion in each tweet. This is assumed to be an issue related to economy, politics, diplomacy or anything what the US President tweets about that might consequently have an impact on investors‘ confidence, mood and decisions. The main drawback of this assumption is that the president sometimes posts objective statements or comments on his daily life, which might not have a significant effect on others.

The value of $s_{ijk}$ is the output of our model in a form of numerical value representing mood intensity extracted for each tweet. The holder of the opinion is represented by the US president over the time span corresponding to intersection of his presidency and the data available.

The microblogging nature of the data implies that the analysis makes sense only for the incumbent and for the previous US President, Donald Trump and Barack Obama, respectively. This is due to the fact that Twitter was founded in 2006 and there are only the two above-stated US presidents with active Twitter accounts during their presidency. Thus, $h_j$ represents only two values, i.e. $j = 1, 2$, where $h_1$ is Barack Obama and $h_2$ is Donald Trump.

The time $t_k$ is the exact date and time, when a tweet was posted. Our sentiment analysis is applied on the data set of 3242 tweets of Donald Trump over the period 3/4/2016 - 6/4/2017 and 2598 tweets of Barack Obama over the time span 16/12/2014 - 23/3/2017. However, for the purposes of the Granger Causality analysis, the data sets were reduced according to certain logic explained in the next section.

For a better comprehension, examples of specific values or information within the quadruples are provided:

Example 1: *"Getting ready to meet President al-Sisi of Egypt. On behalf of the United States, I look forward to a long and wonderful relationship."*
In this example, $g$ is the US relationship with Egypt, $h$ or the holder of the opinion is Donald Trump, the date and time $t$ of the tweet is the 3rd of April in 2017, 16:00:33 and the sentiment $s$ classified by our model is highly positive.

Example 2: *"122 vicious prisoners, released by the Obama Administration from Gitmo, have returned to the battlefield. Just another terrible decision!"*
In the Example 2, $g$ represents the act of releasing the prisoners, the opinion holder is again $h$, Donald Trump, the date and time $t$ of a tweet is the 7th of March in 2017, 12:04:13 and the extracted emotion $s$ is significantly negative.

Example 3: *"Watch @JudgeJeanine on @FoxNews tonight at 9:00 P.M."* The

Example 3 illustrates a tweet with no opinion or emotion. It was posted by Donald Trump on 25th of March in 2017, 14:41:14 and our model classified it as neutral with neither positive nor negative emotion. How this study deals with the objective tweets will be described in more detail in the next section.

### 2.3.2 Machine learning vs. Lexicon-based method

When selecting the model for the sentiment analysis, two main categories have to be considered, i.e. the machine learning approach and the lexicon-based sentiment extraction. Machine learning requires a training set labelled with the emotional orientation. It employs specific features for the classification, such as terms and their frequencies. Firstly, it creates a bag of words from the training set and then assigns a vector with the term frequencies to each of the texts. Other extensively used feature is Part-of-speech (POS) that assigns each word to a grammatical category (e.g. adverbs, adjectives, particles, etc.) based on its properties. In Pak & Paroubek, 2010 the usefulness of this method is shown with the help of the distributions of POS, which were plotted both in the subjective vs. objective texts and in the positive vs. negative sets. The features that can be also applied are rules of opinions (e.g. punctuation, such as "???", "!!!", etc.) or sentiment shifters (e.g. negation).

When the features are specified, the classifier is applied on the training set. The Bayesian probability and the premise that the feature probabilities are independent is incorporated in the Naive Bayes classifier, whereas Support Vector Machines rely on separation of data points by hyperplanes. The Maximum Entropy classifier applies multinomial logistic regression and makes no assumptions on independence among features (Liu, 2012). The classifier then evaluates the test set based on what it has learned from the training set.

These approaches can perform very precisely, but require manual labelling of the training data set and extensive data collection (Hutto & Gilbert, 2014). The number of downloaded tweets is 2598 for Barack Obama and 3242 for Donald Trump. Therefore, the supervised machine learning methods would not be appropriate, since there is a lack of the training data what would in turn cause poor performance of the classifier.

The Lexicon-based methods rely on the opinion dictionary that contains words and phrases associated with specific emotional orientation. There are three methods (Liu, 2012), how these lexicons are built:

1. **Dictionary-based approach** uses a short list of seed words labeled manually and the algorithm adds another words to this set using a common dictionary with the synonyms and antonyms related to each seed word.

2. **Corpus-based approach** employs a list of seed words with a certain sentiment orientation and the algorithm explores a given corpus, i.e. the list of words in the studied set of texts, and collects other sentiment words with their emotional orientations. The main drawback of this approach is the requirement of a large and diverse corpus.

3. **Manual approach** requires a lot of time and effort, but yields valuable results in return.

### 2.3.3 Lexicon selection

A number of lexicons has been constructed already and could be downloaded for research purposes. It is crucial to decide which one is the most appropriate for this study. Among the most broadly used pertain General Inquirer (GI) (Stone et al., 1966), Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001), SentiWordNet (Esuli & Sebastiani, 2007), Sentiment lexicon (Hu & Liu, 2004), Affective norms for English words (ANEW) (Bradley & Lang, 1999) or Emotion lexicon (Mohammad & Turney, 2010).

The GI is one of the first lexicons with over 11 000 words classified into various categories, whereby a list of 4500 categorized words is employed by the LIWC. Both of them have been proved valid by years of research but the main disadvantage is that neither of them can detect the sentiment intensity. In the Sentiment lexicon constructed by Hu & Liu, 2004 the same problem as in the previous two cases arises and that is the possibility of binary classification only. Oppositely, SentiWordNet belongs to the group of dictionaries with the emotional intensity-detecting ability. It contains rankings of positivity, negativity and neutrality for each word. However, the scores have been computed by various algorithms and not annotated by humans as in all of the other stated dictionaries. The ANEW also includes words with the annotated strength of the emotions, with the rating between 1, the most negative, up to 9, the most positive. The Emotion Lexicon detects types of mood within the words, such as anger, joy, fear or sadness. The undesirable aspect of all the described lexicons is that they are not adjusted to the microblogging nature of text.

Twitter has its own characteristics that should be taken into account when deciding which lexicon to use. Therefore, the Valence Aware Dictionary for sEntiment Reasoning (VADER) (Hutto & Gilbert, 2014), constructed with the aim to perform especially well on the microblogging content, was chosen for this study. It captures intensity of emotions not only within the words, but also within the emoticons, sentiment-related abbreviations and slang. Furthermore, the model applies additional rules focused on detecting even more precise sentiment intensity based on the Twitter-related aspects. Finally, as employed in Hutto & Gilbert, 2014, it yielded classification accuracy of 0,96 outperforming individual human raters.

### 2.3.4 Applied Model: the VADER

#### Lexicon construction

The VADER incorporates a gold standard dictionary which means it was compiled manually. Firstly, the authors collected a list of words from the validated lexicons, for instance GI, LIWC or ANEW and then added Twitter-associated features to this list, such as slang words (e.g., "nah"), emoticons (e.g., ":)", ":(") or abbreviations (e.g., "OMG", "LOL"). The intensity of each expression was then ranked using Amazon Mechanical Turk by 10 human raters within the range starting at -4, the most negative, up to 4, the most positive and 0 for neutral. The raters' English and rating skills were checked beforehand and only the ones that scored well were able to proceed to the real tasks. An example of words and their ratings from the lexicon is provided in the Appendix table 6.

The documentation and the Python code of the VADER model has been made available by its creators on https://github.com/cjhutto/vaderSentiment. It can be also imported as a function of the NLTK and that is exactly, how the analysis was done in this thesis.

**Rules and model description**

At first, the function, which computes the compound score, attempts to find words from the lexicon in the text and sums up their emotional values. After that, it aims to identify the presence of 5 groups of features which affect the intensity of a given text as determined by the authors in Hutto & Gilbert, 2014:

1. The punctuation, such as the exclamation marks intensifies the emotion within the text.
2. The word written in capital letters increases magnitude of sentiment that is related to it.
3. The authors compiled a list of intensifiers, so that when they are located before a given word the sentiment within a given word increases and the list of downtoners, which decrease the sentiment value of a word that follows.
4. The tri-grams before a sentiment-linked expression were studied, in order to indentify, where the negation twisted the sentiment of the sentence.
5. Conjuction „but" outlines the flip of the sentiment orientation and the rise of emotional intensity covered in the second part of the sentence.

A question arises, how these features are dealt with in quantitative terms. The authors created up to 10 variations of each of the 30 selected tweets and collected their intensity values from 30 raters. Then, the mean differences between each distribution were added to the model. For instance, when the sentences ended with ".", the mean difference of sentiment intensity was 0,291, when compared to the sentences finished with "!". Therefore, 0,291 is added to the emotional score of the text. For the sake of clarity, other examples of the rules application are stated in the Table 2.2.

**Table 2.2:** Description of rules' application

| **Rule** | Example | **Quantitative application** |
|---|---|---|
| 1. | presence of "!" | adds 0.291 to the overall score |
| 2. | some of the words are ALL CAPS | adds 0.733 to the intensity of the word |
| 3. | "considerably", "intensely" | adds 0,293 to the intensity of following word |
| | "marginally", "barely" | decreases the intensity of a following word by -0,293 |
| 4. | negation between emotional word | flips the sentiment polarity |
| 5. | "but" in a sentence | emotional intensity before "but" decreases |
| | | by 50%, after "but" increases by 50% |

The compound score calculated applying all of the stated rules and other similar ones in the same fashion, is then normalised in equation (2.1) between -1 (the most negative score) and +1 (the most positive score) with 0 being the neutral one:

$$c_i = \frac{m_i}{\sqrt{m_i^2 + \alpha}}, \tag{2.1}$$

where $c_i$ is normalized compound score of text $i$, $m_i$ is the score of a text $i$ as the output of the functions employing the specified rules and $\alpha$ is the approximation of maximal expected value set to 15.

**Our application**

The tweets were not preprocessed in any way, because any procedure that would modify tweets in similar fashion as it was done in the frequency analysis would deteriorate the performance of the rules applied by specific functions of VADER. Hence, each raw text was taken from the tweet and then evaluated by the above-described model computing the output in a form of the sentiment values time series for positivity, negativity, neutrality and the compound score for all of the tweets. The compound scores calculated based on the words intensity and rules included in VADER were taken along with dates, times and number of retweets of the recorded tweet's sentiment for our analysis.

After that, these data sets were modified in such a way that they reflected the aggregated sentiment intensity of tweets contained in the time window starting at 16:00 UTC -5 on a previous trading day (e.g. end of trading hours of the NYSE, NASDAQ) terminating at 16:00 UTC -5 on a next trading day. This logic was applied in order to separate the effect of tweets on individual closing values properly. The calculation of the aggregated sentiment was further performed in a following way:

$$S_i = \sum_{j}^{n_i} s_{ij} r_{ij}, \tag{2.2}$$

where $S$ is the aggregated sentiment, $i$ is a time window ($i \in 1, ..., 478$ for Obama and $i \in 1, ..., 102$ for Trump), $s$ is the sentiment of tweet $j \in 1, ..., n$, $r$ is the number of retweets of tweet $j \in 1, ..., n$ and $n_i$ is the number of tweets in the window $i$.

The multiplication of the sentiment intensity by the number of retweets was decided to be employed in order to catch the effect of a specific tweet on public. The retweets serve as a measure of the users, which reposted the tweet, so there is a reasonable chance that more users have read it or even shared its content among others as well. As already stated, VADER has the ability to tell whether a tweet is neutral and evaluates its sentiment as equal to 0, if so. Thus, the objective posts would not enter our model. The time series resulting from calculated values for all of the time windows were plotted in Figures 3.7 and 3.8 and employed in the Granger causality analysis.

## 2.4 Granger Causality analysis

### 2.4.1 Data

When the data sets were studied, it was found that Barack Obama had given up on his Twitter account on the day after the presidential election 2016 and tweeted only a few times since then. Therefore, his data set for Granger causality analysis was reduced as to account for this discrepancy.

Regarding Donald Trump, the thesis is built on the assumption that the newly-elected President of the United States takes over the substantial influence over public opinion on the day after the election. It follows from the reasoning that there is a significant certainty that the most recently elected president will substitute the previous one in a short time and proceed with his policies. Hence, his data set starts on the US Presidential elections' day dated to 8/11/2016 instead of his inauguration. The summary of the data collections are provided in the Table 2.3.

In terms of frequency of the data sets, it is based on the trading days in the United States as explained in the previous section. As for the sentiment time series, it corresponds to the time span between two trading days' closing hours and the index time series consist of closing values for each trading day.

**Table 2.3:** Time series overview

| | **Period** | **Description** | Sample size |
|---|---|---|---|
| i | 9/11/2016 - 6/4/2017 | Donald Trump's sentiment | 102 |
| ii | 9/11/2016 - 6/4/2017 | DJIA closing prices | 102 |
| iii | 9/11/2016 - 6/4/2017 | NASDAQ closing prices | 102 |
| iv | 9/11/2016 - 6/4/2017 | S&P 500 closing prices | 102 |
| v | 16/12/2014 - 7/11/2016 | Barack Obama's sentiment | 478 |
| vi | 16/12/2014 - 7/11/2016 | DJIA closing prices | 478 |
| vii | 16/12/2014 - 7/11/2016 | NASDAQ closing prices | 478 |
| viii | 16/12/2014 - 7/11/2016 | S&P 500 closing prices | 478 |

*Source: Author.*

### 2.4.2 Definition and underlying approach

Granger Causality analysis is based on the Hume's definition of causality and thus, relies on the premise that if variable X causes Y then changes in X will occur at the point in time before the change in Y (Granger, 1969). The analysis is initiated with the definition of the equations:

$$y_t = \alpha_1 + \sum_{i=1}^{n} \beta_{1,i} y_{t-i} + e_{1,t}, \tag{2.3}$$

$$y_t = \alpha_2 + \sum_{i=1}^{n} \beta_{2,i} y_{t-i} + \sum_{i=1}^{n} \gamma_{2,i} x_{t-i} + e_{2,t}, \tag{2.4}$$

where $\alpha_1$ and $\alpha_2$ are constant terms, $\beta_{1,i}, \beta_{2,i}$ and $\gamma_{2,i}$ for $i \in \mathbb{N}$ are coefficients and $e_{1,t}$ and $e_{2,t}$ are i.i.d. error terms. The sum of squared residuals of equations (2.3) and (2.4) are computed as follows:

$$SSR_1 = \sum_{t=1}^{N} \hat{e}_{1,t}^2,$$

$$SSR_2 = \sum_{t=1}^{N} \hat{e}_{2,t}^2.$$

The test statistic is calculated:

$$T = \frac{\frac{SSR_1 - SSR_2}{n}}{\frac{SSR_2}{N-2n-1}},$$

where $n$ is equal to number of lags included in the model and $N$ is the sample size and this test statistic follows asymptotically $F(n, N-2n-1)$ distribution. In case the null:

$$H_0 : \gamma_{2,i} = 0, \quad \text{for} \quad i \in \mathbb{N}$$

is rejected, it can be stated that the variable $x_t$ Granger-causes the $y_t$.

### 2.4.3 Unit root testing

Before the Granger causality analysis is carried out, one has to make sure that the variables in the regression are stationary, i.e., examine, if they contain a unit root or not. Stationarity means that the time series stays around a constant mean value and has a constant variance over time. If the time series has a unit root, it is non-stationary and its presence in a model can falsely imply existence of economic relationship, unless it combines with other non-stationary variable, whereby creating a stationary process (Harris & Sollis, 2003).

The testing for a unit root is done by following the consecutive steps of the Perron's sequential testing procedure (Harris & Sollis, 2003) until the null of a unit root is rejected or until all of the steps are carried out:

1. The most general autoregressive model with time trend and drift is tested for unit root with the Augumented Dickey Fuller test (ADF), Dickey Fuller critical values, and the number of lags included is chosen based on the general-to-specific strategy applying t-testing. The maximum lag length studied within the general-to-specific approach is computed with the help of the Schwert's formula (Harris & Sollis, 2003):

$$p_{max} = integer\left[12 \times \sqrt[4]{\frac{N}{100}}\right],$$

   where $N$ is the sample size. For the sake of clarification, the autoregression of a variable on its the maximum lag length based on Schwert's formula is run and if the highest lag's p-value does not imply reasonable significance level, the process is repeated with the number of lags decreased by one, until the highest lag in the model is found to be significant. The highest lag's number is then employed as the appropriate lag length in the model that undergoes ADF test.

2. If the null hypothesis is not rejected, the process continues to the joint F-test of a unit root and the inclusion of time trend in the model employing the Dickey and Fuller critical values. The rejection of this hypothesis leads again to the unit root test of the model with time trend and drift, but applying the t-test statistic with standard normal critical values.

3. If the joint hypothesis in F-test or the null in t-test is not rejected, then the time trend is dropped and a model with drift and the lag length chosen within another general-to-specific procedure is tested for a unit root with Dickey and Fuller critical values.

4. If the null hypothesis is not rejected, the joint F-test for significance of drift and unit root is employed with the Dickey and Fuller critical values. If it is rejected, the t-test for the presence of unit root is carried out with standard normal critical values.

5. Finally, similarly as in step 3., if the joint hypothesis in F-test or the null in t-test is not rejected, then the model without constant term and time trend is tested with ADF for the presence of a unit root implementing another set of Dickey and Fuller critical values and the lag length selected within another general-to-specific procedure.

If one fails to reject the null hypothesis, then the time series contains at least one unit root. For the sake of clarity, the individual steps are summarised in the Table 2.4 (Harris & Sollis, 2003).

**Table 2.4:** Perron's (1988) sequential procedure using ADF test

| **Step** | **Model** | **Null** | **Test statistic** | **Critical values** |
|---|---|---|---|---|
| 1. | $\Delta y_t = \mu_c + \gamma_c t + (\rho_c - 1)y_{t-1} + u_t$ | $(\rho_c - 1) = 0$ | $\tau_{tau}$ | Fuller |
| 2a. | $\Delta y_t = \mu_c + \gamma_c t + (\rho_c - 1)y_{t-1} + u_t$ | $(\rho_c - 1) = \gamma_c = 0$ | $\Phi_3$ | Dickey-Fuller |
| 2b. | $\Delta y_t = \mu_c + \gamma_c t + (\rho_c - 1)y_{t-1} + u_t$ | $(\rho_c - 1) = 0$ | $t$ | Standard Normal |
| 3. | $\Delta y_t = \mu_b + (\rho_b - 1)y_{t-1} + u_t$ | $(\rho_b - 1) = 0$ | $\tau_\mu$ | Fuller |
| 4a. | $\Delta y_t = \mu_b + (\rho_b - 1)y_{t-1} + u_t$ | $(\rho_b - 1) = 0$ | $\Phi_1$ | Dickey-Fuller |
| 4b. | $\Delta y_t = \mu_b + (\rho_b - 1)y_{t-1} + u_t$ | $(\rho_b - 1) = \mu_b = 0$ | $t$ | Standard Normal |
| 5. | $\Delta y_t = (\rho_a - 1)y_{t-1} + u_t$ | $(\rho_a - 1) = 0$ | $\tau$ | Fuller |

*Source:(Harris & Sollis, 2003).*

The widely employed procedure that deals with non-stationarity caused either by a time trend or a unit root is the differencing of the series (Verbeek, 2008):

$$\Delta y_t = y_t - y_{t-1}.$$

The process containing a unit root that becomes stationary after first differencing is said to be integrated of order 1 or I(1). Therefore, if any of our time series of stock indices' returns or sentiment values is found to contain a unit root, the differencing has to be conducted and the resulting time series has to be tested for the presence of unit root employing the same procedure.

### 2.4.4 Cointegration

Firstly, if two different time series X and Y are integrated of the same order, i.e. contain equal number of unit roots $n \in 1, 2$, and there exists a constant $\beta$, such that the order of integration of $Y - \beta X$ is lower than the order of integration of those two series themselves, then they are cointegrated (Wooldridge, 2015). It means that these time series have an underlying relationship. Thus, if the unit root testing reveals that two of our time series within any of the models are integrated of order 1 or 2, the cointegration has to be tested. This can be done by the Engle - Granger procedure (Engle & Granger, 1987). Furthermore, if the cointegration within two series is found, the Vector Error Correction Model has to be employed.

### 2.4.5 Natural logarithms of stock returns

In an extensive amount of applied work, the variables in a functional form of natural logarithms are encountered and this follows from the fact that they allow to incorporate nonlinearities in regressions. They often appear in the time series regressions with constant percentage effect (Wooldridge, 2015). Thus, the the first differences of natural logarithms of index values are utilized in this study.

### 2.4.6 VAR($l$) model

Bivariate Vector Autoregression model or VAR($l$), where $l$ is the number of lags, will be employed in this thesis for the purposes of the Granger Causality analysis. It is a popular choice among researchers (Wooldridge, 2015) and it can be written as follows:

$$y_t = \alpha_1 + \sum_{i=1}^{l} \beta_{1,i} y_{t-i} + \sum_{i=1}^{l} \gamma_{1,i} x_{t-i} + e_{1,t},$$

$$x_t = \alpha_2 + \sum_{i=1}^{l} \beta_{2,i} x_{t-i} + \sum_{i=1}^{l} \gamma_{2,i} y_{t-i} + e_{2,t},$$

where $\alpha_1$ and $\alpha_2$ are constant terms, $\beta_{1,i}$ and $\beta_{2,i}$ for $i \in \mathbb{N}$ are coefficients and $e_{1,t}$ and $e_{2,t}$ are i.i.d. error terms. The $y_t$ would be in our case the first differences of natural logarithms of stock indexes and $x_t$ time series of Trump's and Obama's sentiment. On the whole, 6 final pairs of baseline models are studied and the variables, which enter the models are stated in Table 2.5 for the purposes of clarification.

**Table 2.5:** Models

| | $y_t$ | $x_t$ |
|---|---|---|
| i | First differences of natural log of DJIA values | Trump's sentiment |
| ii | First differences of natural log of NASDAQ values | Trump's sentiment |
| iii | First differences of natural log of S&P 500 values | Trump's sentiment |
| iv | First differences of natural log of DJIA values | Obama's sentiment |
| v | First differences of natural log of NASDAQ values | Obama's sentiment |
| vi | First differences of natural log of S&P 500 values | Obama's sentiment |

*Source: Author.*

### 2.4.7 Lag length selection using AIC and BIC

The well known criterion for the selection of number of lags to be included in the model are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (Verbeek, 2008) and both of them are applied in this thesis.

The AIC is computed as:

$$AIC = -2logL + 2p,$$

and the BIC as:

$$BIC = -2logL + plogN,$$

where L is maximized likelihood function, $p$ is the number of parameters in the model and $N$ is the sample size.

For the purposes of clarification, one has to construct the models with a range of lags to be studied, estimate them and then choose the model with the value of AIC or BIC that is the lowest among all. The maximum of the lags applied for our model selection is based on the study conducted in similar fashion by Bollen et al., 2011 and equals 7.

# Chapter 3

# Results

## 3.1 Getting to know the data and the frequency analysis

With regards to the elementary statistical measures computed for both studied presidents (Table 3.1), it appeared that Trump is using Twitter more often than Obama. Moreover, it had to be concluded that the frequency of Donald Trump's daily tweets' count had recorded substantially higher sample variance than Obama's. The daily tweet count followed the left tailed distribution in both cases, which could be implied not only from the graphical analysis in Figures 3.1 and 3.2, but also from the skewness values. Furthermore, the average and the median number of daily retweets favoured the notion that Trump's daily activity on Twitter tends to be more striking than Obama's.

**Table 3.1:** Computed statistical measures

| **Computed measure** | **@BarackObama** | **@realDonaldTrump** |
|---|---:|---:|
| Average of the daily tweets | 4 | 8 |
| Mode of the daily tweets | 4 | 3 |
| Median of the daily tweets | 4 | 7 |
| Maximum number of daily tweets | 60 | 87 |
| Standard deviation of the daily tweets | 4,91 | 8,08 |
| Sample variance of the daily tweets | 24,20 | 65,27 |
| Skewness of the daily tweets' distribution | 3,54 | 2,22 |
| Average number of the daily retweets | 2632 | 11613 |
| Median of the daily retweets | 877 | 8445 |
| Maximum number of daily retweets | 446895 | 756920 |

*Source: Author.*

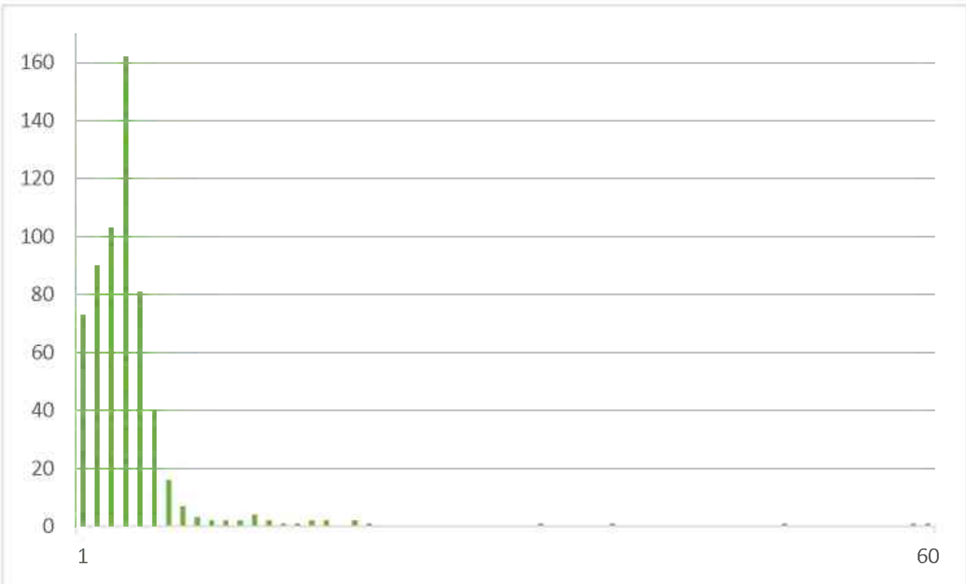



**Figure 3.1:** The daily tweets' count distribution of @BarackObama

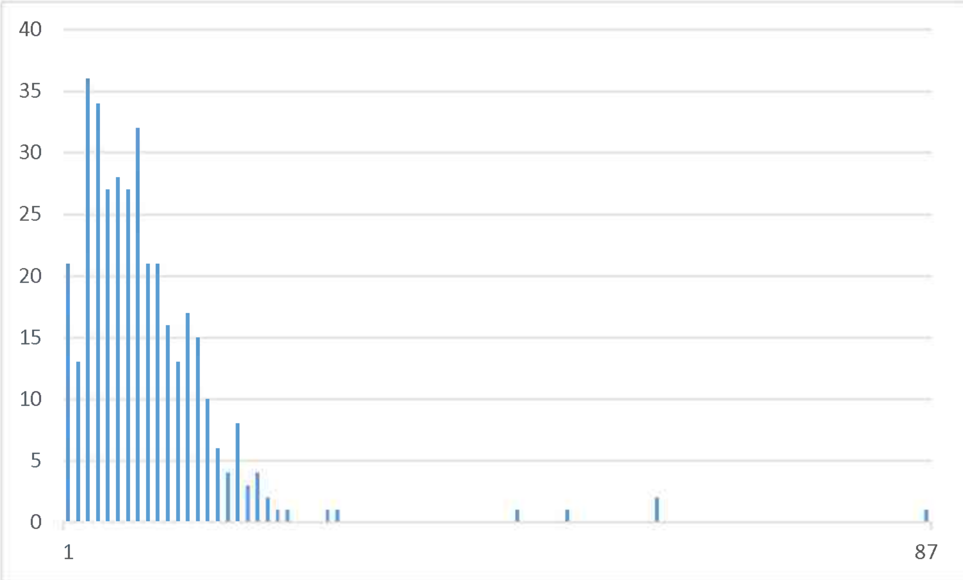
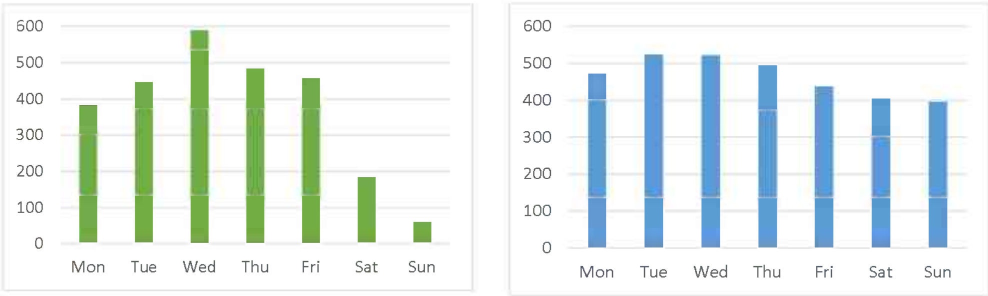



**Figure 3.2:** The daily tweets' count distribution of @realDonaldTrump

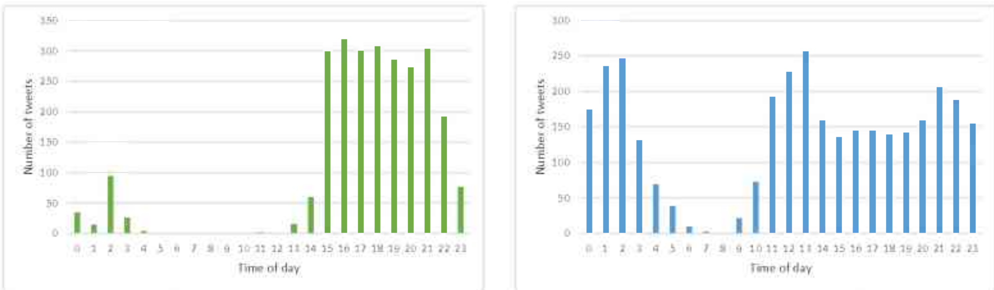
As for the distribution of tweets' count on the specific weekdays plotted in Figures 3.3, Obama's activity peaked on Wednesdays and ceased during the weekends, whereas Trump's activity decreased only slightly at the end of the week. Since the weekends are not trading days, this uneven distribution should not interfere with the results of our analysis in any way. Moreover, the tweets posted during the weekends or national holidays in the United States were added to the aggregate sentiment score of the next closest trading day.

In terms of the frequency of tweeting during the different times of day, the Figures 3.4 depicted that Obama almost never posted during the night or early morning hours, while the Trump's distribution unveiled the opposite result.




(a) (b)

**Figure 3.3:** Distribution of tweets of @BarackObama (a) and @realDonaldTrump (b) during the week




(a) (b)

**Figure 3.4:** Distribution of tweets of @BarackObama (a) and @realDonaldTrump (b) during the day

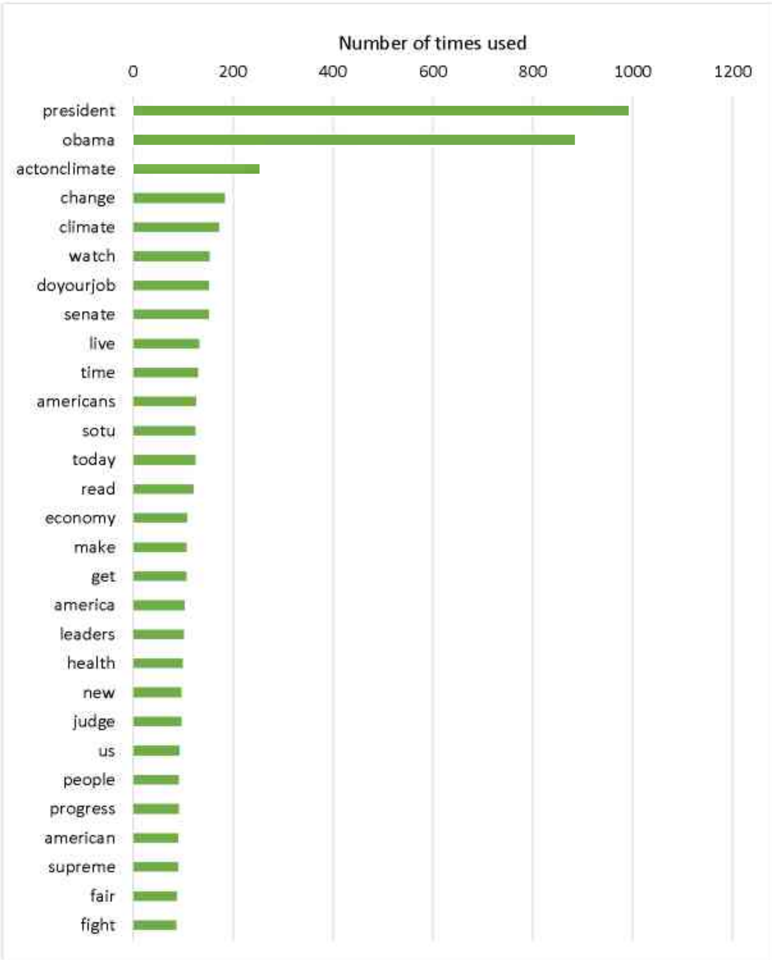



**Figure 3.5:** Top 30 most frequently used words with semantic meaning of @BarackObama

Firstly, the word frequency analysis revealed the topics that both presidents had been focused on. In the data set of the user account @BarackObama (Figure 3.5), it could be noticed that he had often mentioned climate change, which actually corresponded to his policy priorities. The phrase "#DoYourJob" in a form of hashtag was initiated by him and illustrated a national effort to make Congress consider a nominee to the Supreme Court suggested by Obama in March 2016. The senate, economy and the health care seemed to be regularly commented topics by him as well.

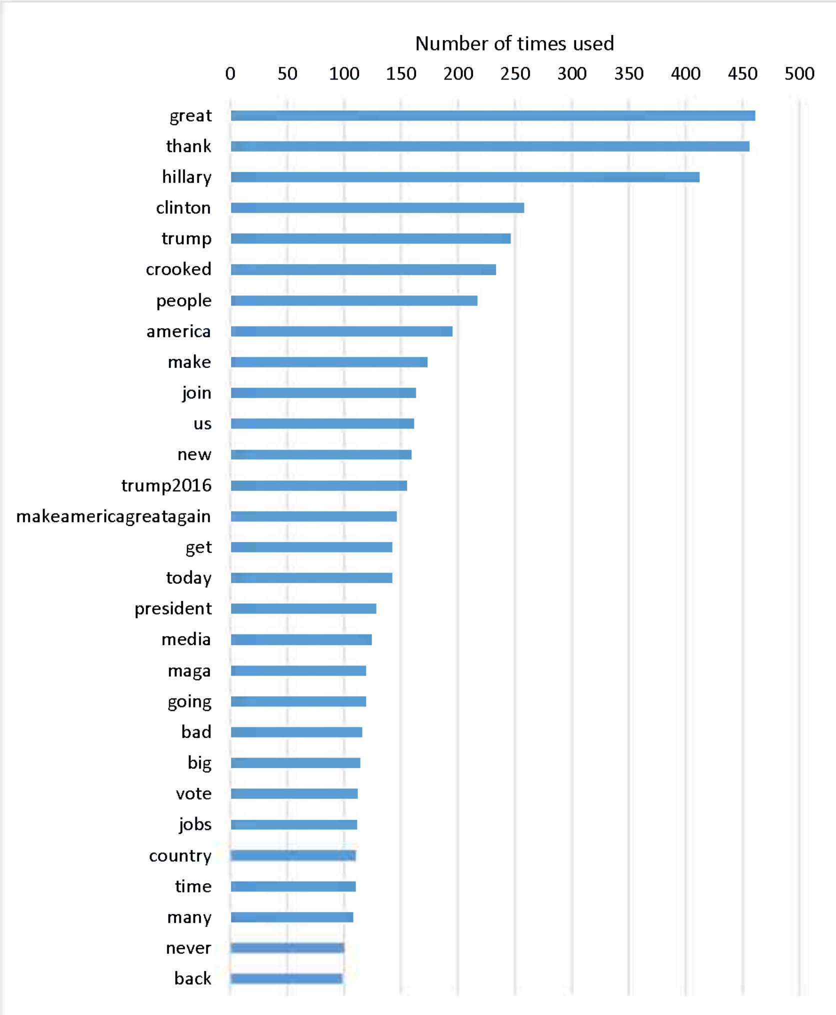



**Figure 3.6:** Top 30 most frequently used words with semantic meaning of @realDonaldTrump

Based on the list of words that were extracted from the user account @realDonaldTrump (Figure 3.6) it could be implied that Trump had made remarks on Hillary Clinton in his tweets many times. His hashtagged campaign motto "#MakeAmericaGreatAgain" or "#MAGA" also appeared a number of times since the data set used for the frequency analysis covered also months before the elections. Apart from this, he commented on media and jobs quite often. This fact was in accordance with his known policy priorities which aim to create more jobs and beat unemployment in the U.S. at the same time.

Taking a closer look at the similarities of the data sets, the same words in both of the top 30 were: “people”, “president”, “America/ Americans”, “today”, and verbs, such as “make” and “get”. Oppositely, the most appealing difference was that there were several commonly used simple adjectives in Trump’s top 30, such as “great”, “bad” or “big”, which did not make it to Obama’s top 30. Instead of those, Obama appeared to use nouns more, such as “senate”, “leaders”, “progress” or “fight”. Finally, the analysis possibly indicated a more pronounced use of a simplistic lexicon in case of Donald Trump and a more sophisticated lexicon, when it came to Barack Obama.

## 3.2 Sentiment analysis

The resulting time series of sentiment analysis in a form of aggregated sentiment computed based on the equation (2.2) for each of the presidents were plotted in Figures 3.7 and 3.8. A few outliers could be spotted in the time series of Obama’s sentiment. They corresponded to the events such as State of the Union speech on 21/1/2015, Supreme Court rule that the same-sex couples have a right to marry on 26/6/2015, Ahmed Mohamed clock incident, when Obama stood up for a little boy on 17/9/2016 causing massive reaction or the tweet supporting protests over gun control on 26/6/2016 after Orlando shootings and a sequence of tweets regarding health care progress, jobs creation, Paris agreement about climate protection and economic strength of the US on 4/11/2016.

With relation to Trump’s sentiment time series, several outlying movements were identified as well. The spikes on 9/11/2016 and 23/1/2017 corresponded to happiness regarding the election’s result and inauguration, respectively. A substantial outlier could be found on 30/1/2017, when Trump’s Immigration ban was enforced or on 27/3/2017, when he made positive comment on the Secretary of the National Security activities with regards to immigrants.

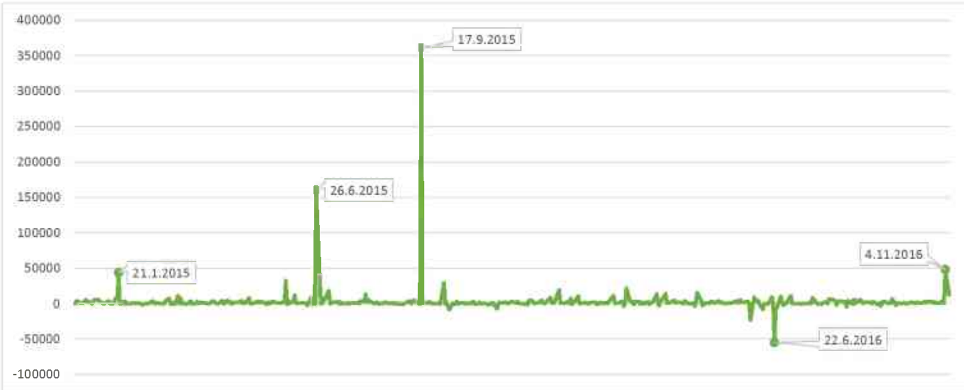



**Figure 3.7:** The aggregated compound sentiment of @BarackObama over the course of trading days

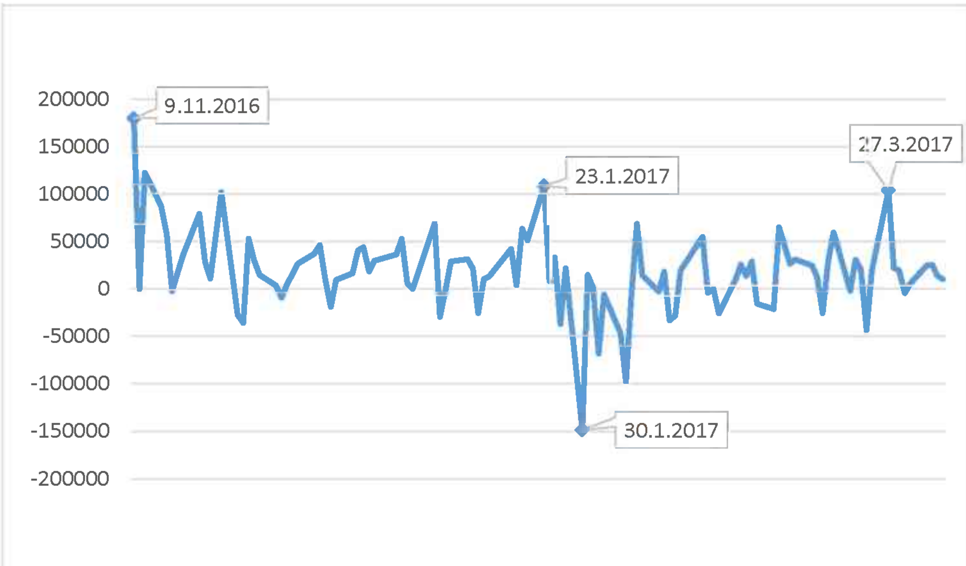



**Figure 3.8:** The aggregated compound sentiment of @realDonaldTrump over the course of trading days

## 3.3 Granger causality analysis

### 3.3.1 Unit root testing

The results of unit root testing are summarised in the Table 3.2, employing the general-to-specific strategy in order to select appropriate lag length for the models in the Perron's sequential procedure and then testing for the presence of the unit root. In all of the stock indices' time series, the null hypothesis that there is a unit root was not rejected at any stage of Perron's testing sequence. Therefore, first differences were constructed and the exact same procedure was carried out on them. The null was rejected at the first step in all of the cases. Thus, the original series have been transformed to first difference stationary processes and could be claimed as integrated of order 1, containing one unit root. As for the sentiment time series, both of them were found to be stationary processes integrated of order 0. These findings are consistent with the nature of the stock prices' movements, with an up-to-date research and an educated guess of this result could be taken ex ante from the plotted line graphs as well.

It must be further stressed out, that only stationary processes can be employed within the VAR models and because of that, the first differences of natural logarithms of stock indices' closing prices entered our model. Those were additionally tested for stationarity and the null of a unit root was rejected at the first step of Perron's procedure with a high level of significance. The sentiment time series were applied without differencing as they proved to be stationary.

**Table 3.2:** Unit root testing overview

| | **Period** | **Description** | **Result** |
|---|---|---|---|
| i | 9/11/2016 - 6/4/2017 | Trump's sentiment | No unit root, stationary I(0) |
| ii | 9/11/2016 - 6/4/2017 | DJIA closing prices | One unit root, first difference stationary I(1) |
| iii | 9/11/2016 - 6/4/2017 | NASDAQ closing prices | One unit root, first difference stationary I(1) |
| iv | 9/11/2016 - 6/4/2017 | S&P 500 closing prices | One unit root, first difference stationary I(1) |
| v | 16/12/2015 - 7/11/2016 | Obama's sentiment | No unit root, stationary I(0) |
| vi | 16/12/2015 - 7/11/2016 | DJIA closing prices | One unit root, first difference stationary I(1) |
| vii | 16/12/2015 - 7/11/2016 | NASDAQ closing prices | One unit root, first difference stationary I(1) |
| viii | 16/12/2015 - 7/11/2016 | S&P 500 closing prices | One unit root, first difference stationary I(1) |

*Source: Author.*

A more detailed version of the results including the chosen lag lengths, test statistics, critical values and significance levels extracted employing ADF test throughout the Perron's steps is provided in the Appendix tables 7 - 20.

### 3.3.2 Cointegration

The unit root testing revealed that stock indices are I(1) and the sentiment time series are I(0), and none of the distinct indices' time series entered any of the models alongside each other. This fact implied that the constructed VAR models contained only time series, which are integrated of different order and thus, the cointegration can not be present. Therefore, there was no necessity to apply Engle-Granger procedure and Vector Error Correction model.

### 3.3.3 Granger Causality

Firstly, the models with 1 up to 7 lags were estimated and the best fitting one was selected based on the lowest value of AIC and BIC, and it was either VAR(1) or VAR(2). In the cases, when AIC and BIC yielded conflicting options, both of the chosen models were employed and tested accordingly. The precise results of the model selection, computed test statistics and p-values are summarised in Table 3.3.

The p-values calculated for Barack Obama's sentiment outlined Granger causal impact on the first differences of logarithms of DJIA and S&P 500 were surprisingly significant in all three cases in the model VAR(1) and for DJIA for both VAR(1) and VAR(2). It has to be noted, that the model of Obama's sentiment and first differences of NASDAQ logarithms recorded a p-value that was only 0,001 higher than the one within the 10% of significance. Oppositely, in case of Donald Trump's sentiment, the p-values were rather high and not significant. The retained values imply that Obama's sentiment as extracted from his Twitter account Granger causes the percentage changes in DJIA and S&P 500 in the model with 1 lag and for DJIA also in the model with 2 lags employed. The Trump's sentiment did not yield any significant relationship in terms of Granger causality as for the models with 1 lag as well as 2 lags.

However, these results have to be looked at with reasonable amount of caution. The terms "causes" and "Granger causes" contain a distinct meaning. As noted by Wooldridge, 2015, the Granger causality incorporates no information about "contemporaneous" causality between two variables and thus, it doesn't help us to identify the exogeneity or endogeneity of the variable that is the source of the Granger causality. There might still be an underlying third variable that guides the ones, on which the study is conducted.

| **Trump's sentiment** ($S$) | | | | | | |
|---|---|---|---|---|---|---|
| $X$ | AIC | BIC | test statistic $X \to S$ | p-value $X \to S$ | test statistic $S \to X$ | p-value $S \to X$ |
| DJIA | VAR(2) | VAR(1) | 1,984; 1,3726 | 0,371; 0,241 | 0,48646; 0,05204 | 0,784; 0,820 |
| S&P 500 | VAR(2) | VAR(1) | 0,45589; 0,34208 | 0,796; 0,559 | 0,67474; 0,22143 | 0,714; 0,638 |
| NASDAQ | VAR(2) | VAR(1) | 0,73548; 0,32779 | 0,692; 0,567 | 0,65553; 1,5533 | 0,721; 0,213 |
| **Obama's sentiment** ($S$) | | | | | | |
| $X$ | AIC | BIC | test statistic $X \to S$ | p-value $X \to S$ | test statistic $S \to X$ | p-value $S \to X$ |
| DJIA | VAR(2) | VAR(1) | 5,994; 4,7452 | **0,050**; 0,029**** | 0,96113; 0,29002 | 0,618; 0,590 |
| S&P 500 | VAR(1) | VAR(1) | 4,0787 | **0,043**** | 0,35242 | 0,553 |
| NASDAQ | VAR(1) | VAR(1) | 2,6848 | 0,101 | 0,10333 | 0,748 |

**Table 3.3:** Model selection and Granger causality tests' results: the test statistics and p-values are stated as for the model selected by AIC and BIC, respectively. The p-values significant on 5% level are marked with **

# Conclusion and further discussion

Conclusion

The primary goal of this thesis was to detect the potential existence of causal relationship of the numeric emotional values as extracted from the Twitter accounts of two consecutive presidents of the United States on the three major stock market indices' returns. More specifically, the daily textual remarks of Barack Obama and Donald Trump were examined employing the natural language processing technique called the sentiment analysis, where its output was further analysed and applied within the bivariate VAR models estimation and testing for Granger causality.

While an extensive stream of research has emerged over the last years attempting to relate the public sentiment to the stock indices' movements or the individual stock prices' shifts, to our best knowledge none of the studies has been concerned with the analysis of emotional intensity of the comments of the president of the United States and their relation to stock markets so far. The fact that the studied person is in a possession of a substantial degree of power and thus influence over the citizens and their lives, was one of the key motivations this thesis was constructed on.

Regarding the results of the Granger causality analysis, a relationship was found between the sentiment time series of ex-president Barack Obama, which Granger causes the first differences of logarithmic DJIA values as well as the first differences of logarithmic S&P 500 time series on a 5% significance level. Oppositely, no such output was present in the case of Donald Trump.

This is rather surprising since there has been a lot of “buzz” about incumbent president’s tweeting manners and their impact recently.

Furthermore, the interpretation of the promising result was carried out in a rather cautious and conservative manner, whereby it was stressed out that the Granger causality approach has its drawbacks. A true underlying causal relationship could not be concluded based solely on this result as the variable acting as a source of Granger causality could be exogenous as well as endogenous.

Future research could be extended to an application of more sophisticated models aimed at proving true causal relationship or the models that account for nonlinear linkages and predictions, such as neural networks.

# References

ACHREKAR, Harshavardhan; GANDHE, Avinash; LAZARUS, Ross; YU, Ssu-Hsin; LIU, Benyuan, 2011. Predicting flu trends using twitter data. In: *Predicting flu trends using twitter data. Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, pp. 702–707.

ALKHATIB, Khalid; NAJADAT, Hassan; HMEIDI, Ismail; SHATNAWI, Mohammed K Ali, 2013. Stock price prediction using k-nearest neighbor (kNN) algorithm. *International Journal of Business, Humanities and Technology.* Vol. 3, no. 3, pp. 32–44.

ASUR, Sitaram; HUBERMAN, Bernardo A, 2010. Predicting the future with social media. In: *Predicting the future with social media. Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on.* Vol. 1, pp. 492–499.

BECHARA, Antoine; DAMASIO, Antonio R, 2005. The somatic marker hypothesis: A neural theory of economic decision. *Games and economic behavior.* Vol. 52, no. 2, pp. 336–372.

BIRD, Steven; KLEIN, Ewan; LOPER, Edward, 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc."

BOLLEN, Johan; MAO, Huina; ZENG, Xiaojun, 2011. Twitter mood predicts the stock market. *Journal of computational science.* Vol. 2, no. 1, pp. 1–8.

BRADLEY, Margaret M; LANG, Peter J, 1999. *Affective norms for English words (ANEW): Instruction manual and affective ratings.* Technical report. Citeseer.

BROCHADO, Ana et al., 2016. *Retail Investor Sentiment: Can We Google It?* Technical report. EcoMod.

CHEN, Nai-Fu; ROLL, Richard; ROSS, Stephen A, 1986. Economic forces and the stock market. *Journal of business*, pp. 383–403.

CHIN, Delenn; ZAPPONE, Anna; ZHAO, Jessica, 2016. *Analyzing Twitter Sentiment of the 2016 Presidential Candidates.* STANFORD (ed.)

COHEN, Jeffrey E, 1995. Presidential rhetoric and the public agenda. *American Journal of Political Science*, pp. 87–107.

DAS, Sanjiv R; CHEN, Mike Y, 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management science*. Vol. 53, no. 9, pp. 1375–1388.

DE CHOUDHURY, Munmun; COUNTS, Scott; HORVITZ, Eric, 2013a. Major life changes and behavioral markers in social media: case of childbirth. In: *Major life changes and behavioral markers in social media: case of childbirth. Proceedings of the 2013 conference on Computer supported cooperative work*, pp. 1431–1442.

DE CHOUDHURY, Munmun; GAMON, Michael; COUNTS, Scott; HORVITZ, Eric, 2013b. Predicting Depression via Social Media. In: *Predicting Depression via Social Media. ICWSM*, p. 2.

DURAHIM, Ahmet Onur; COŞKUN, Mustafa, 2015. # iamhappybecause: Gross National Happiness through Twitter analysis and big data. *Technological Forecasting and Social Change*. Vol. 99, pp. 92–105.

ENGLE, Robert F; GRANGER, Clive WJ, 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pp. 251–276.

ESULI, Andrea; SEBASTIANI, Fabrizio, 2007. SENTIWORDNET: A high-coverage lexical resource for opinion mining. *Evaluation*, pp. 1–26.

FAMA, Eugene F, 1965. The behavior of stock-market prices. *The journal of Business*. Vol. 38, no. 1, pp. 34–105.

FAMA, Eugene F, 1970. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*. Vol. 25, no. 2, pp. 383–417.

GOLBECK, Jennifer; ROBLES, Cristina; EDMONDSON, Michon; TURNER, Karen, 2011. Predicting personality from twitter. In: *Predicting personality from twitter. Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pp. 149–156.

GRANGER, Clive WJ, 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pp. 424–438.

GREENSTEIN, Fred I, 1974. What the President Means to Americans: Presidential'Choice'Between Elections. *Choosing the president*, pp. 121–147.

HARRIS, Richard; SOLLIS, Robert, 2003. *Applied time series modelling and forecasting.* Wiley.

HU, Minqing; LIU, Bing, 2004. Mining and summarizing customer reviews. In: *Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177.

HUTTO, Clayton J; GILBERT, Eric, 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Vader: A parsimonious rule-based model for sentiment analysis of social media text. Eighth International AAAI Conference on Weblogs and Social Media.*

JAVA, Akshay; SONG, Xiaodan; FININ, Tim; TSENG, Belle, 2007. Why we twitter: understanding microblogging usage and communities. In: *Why we twitter: understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pp. 56–65.

KARABULUT, Yigitcan, 2013. Can Facebook predict stock market activity?

KHAN, Aamera ZH; ATIQUE, Mohammad; THAKARE, VM, 2015. Combining lexicon-based and learning-based methods for Twitter sentiment analysis. *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJEC-SCSE)*, pp. 89.

KIM, Yoosin; JEONG, Seung Ryul; GHANI, Imran, 2014. Text opinion mining to analyze news for stock market prediction. *Int. J. Advance. Soft Comput. Appl.* Vol. 6, no. 1.

KORDONIS, John; SYMEONIDIS, Symeon; ARAMPATZIS, Avi, 2016. Stock Price Forecasting via Sentiment Analysis on Twitter. In: *Stock Price Forecasting via Sentiment Analysis on Twitter. Proceedings of the 20th Pan-Hellenic Conference on Informatics*, p. 36.

LEE, Han Soo, 2014. Analyzing the Multidirectional Relationships Between the President, News Media, and the Public: Who Affects Whom? *Political Communication.* Vol. 31, no. 2, pp. 259–281.

LEV, Baruch; THIAGARAJAN, S Ramu, 1993. Fundamental information analysis. *Journal of Accounting research*, pp. 190–215.

LIAO, Shu-Hsien; CHU, Pei-Hui; HSIAO, Pei-Yuan, 2012. Data mining techniques and applications–A decade review from 2000 to 2011. *Expert systems with applications.* Vol. 39, no. 12, pp. 11303–11311.

LIU, Bing, 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies.* Vol. 5, no. 1, pp. 1–167.

LOEWENSTEIN, George F; WEBER, Elke U; HSEE, Christopher K; WELCH, Ned, 2001. Risk as feelings. *Psychological bulletin.* Vol. 127, no. 2, pp. 267.

MADGE, Saahil. Predicting Stock Price Direction using Support Vector Machines.

MITTAL, Anshul; GOEL, Arpit, 2012. Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*. Vol. 15.

MOAT, Helen Susannah; CURME, Chester; AVAKIAN, Adam; KENETT, Dror Y; STANLEY, H Eugene; PREIS, Tobias, 2013. Quantifying Wikipedia usage patterns before stock market moves. *Scientific reports*. Vol. 3.

MOHAMMAD, Saif M; TURNEY, Peter D, 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In: *Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pp. 26–34.

MOSTAFA, Mohamed M, 2013. More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*. Vol. 40, no. 10, pp. 4241–4251.

MYSLÍN, Mark; ZHU, Shu-Hong; CHAPMAN, Wendy; CONWAY, Mike, 2013. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of medical Internet research*. Vol. 15, no. 8, pp. e174.

NAZÁRIO, Rodolfo Toríbio Farias; SILVA, Jéssica Lima e; SOBREIRO, Vinicius Amorim; KIMURA, Herbert, 2017. A Literature Review Of Technical Analysis On Stock Markets. *The Quarterly Review of Economics and Finance*.

NGUYEN, Thien Hai; SHIRAI, Kiyoaki, 2015. Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction. In: *Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction. ACL (1)*, pp. 1354–1364.

PAGOLU, Venkata Sasank; CHALLA, Kamal Nayan Reddy; PANDA, Ganapati; MAJHI, Babita, 2016. Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *arXiv preprint arXiv:1610.09225*.

PAK, Alexander; PAROUBEK, Patrick, 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: *Twitter as a Corpus for Sentiment Analysis and Opinion Mining. LREc*. Vol. 10. No. 2010.

PENNEBAKER, James W; FRANCIS, Martha E; BOOTH, Roger J, 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*. Vol. 71, no. 2001, pp. 2001.

PREIS, Tobias; MOAT, Helen Susannah; STANLEY, H Eugene, 2013. Quantifying trading behavior in financial markets using Google Trends.

QIAN, Bo; RASHEED, Khaled, 2007. Stock market prediction with multiple classifiers. *Applied Intelligence*. Vol. 26, no. 1, pp. 25–33.

RANCO, Gabriele; ALEKSOVSKI, Darko; CALDARELLI, Guido; GRČAR, Miha; MOZETIČ, Igor, 2015. The effects of Twitter sentiment on stock price returns. *PloS one*. Vol. 10, no. 9, pp. e0138441.

SAMUEL, Arthur L, 1960. Programming computers to play games. *Advances in computers*. Vol. 1, pp. 165–192.

SCHUMAKER, Robert P; ZHANG, Yulei; HUANG, Chun-Neng; CHEN, Hsinchun, 2012. Evaluating sentiment in financial news articles. *Decision Support Systems*. Vol. 53, no. 3, pp. 458–464.

SIGANOS, Antonios; VAGENAS-NANOS, Evangelos; VERWIJMEREN, Patrick, 2014. Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*. Vol. 107, pp. 730–743.

STONE, Philip J; DUNPHY, Dexter C; SMITH, Marshall S, 1966. The general inquirer: A computer approach to content analysis.

SURESH, AS, 2013. A study on fundamental and technical analysis. *International Journal of Marketing, Financial Services & Management Research*. Vol. 2, no. 5, pp. 44–59.

TUMASJAN, Andranik; SPRENGER, Timm Oliver; SANDNER, Philipp G; WELPE, Isabell M, 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*. Vol. 10, no. 1, pp. 178–185.

*Twitter usage: Company facts*, 2016 [`https://about.twitter.com/company`]. Accessed: 2016-06-30.

VERBEEK, Marno, 2008. *A guide to modern econometrics*. John Wiley & Sons.

WEISS, Sholom M; INDURKHYA, Nitin, 1998. *Predictive data mining: a practical guide*. Morgan Kaufmann.

WONG, Chayanin; KO, In-Young, 2016. Predictive Power of Public Emotions as Extracted from Daily News Articles on the Movements of Stock Market Indices. In: *Predictive Power of Public Emotions as Extracted from Daily News Articles on the Movements of Stock Market Indices. Web Intelligence (WI), 2016 IEEE/WIC/ACM International Conference on*, pp. 705–708.

WOOD, B Dan; OWENS, Chris T; DURHAM, Brandy M, 2005. Presidential rhetoric and the economy. *Journal of Politics*. Vol. 67, no. 3, pp. 627–645.

WOOLDRIDGE, Jeffrey M, 2015. *Introductory econometrics: A modern approach*. Nelson Education.

YOUNG, Garry; PERKINS, William B, 2005. Presidential rhetoric, the public agenda, and the end of presidential television's "golden age". *Journal of Politics*. Vol. 67, no. 4, pp. 1190–1205.

ZHANG, Xue; FUEHRES, Hauke; GLOOR, Peter A, 2011. Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences*. Vol. 26, pp. 55–62.

# Appendix

## A Apendix A

**Table 4:** List of stop-words

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| a | about | above | after | again | against | ain | all |
| am | an | and | any | are | aren | as | at |
| be | because | been | before | being | below | between | both |
| but | by | can | couldn | d | did | didn | do |
| does | doesn | doing | don | down | during | each | few |
| for | from | further | had | hadn | has | hasn | have |
| haven | having | he | her | here | hers | herself | him |
| himself | his | how | i | if | in | into | is |
| isn | it | its | itself | just | ll | m | ma |
| me | mightn | more | most | mustn | my | myself | needn |
| no | nor | not | now | o | of | off | on |
| once | only | or | other | our | ours | ourselves | out |
| over | own | re | s | same | shan | she | should |
| shouldn | so | some | such | t | than | that | the |
| their | theirs | them | themselves | then | there | these | they |
| this | those | through | to | too | under | until | up |
| ve | very | was | wasn | we | were | weren | what |
| when | where | which | while | who | whom | why | will |
| with | won | wouldn | y | you | your | yours | yourself |
| yourselves | | | | | | | |

*Source:* .

Table 5: Tweet's decomposition

| **Attribute** | **Description** | Example |
|---|---|---|
| created_at | Date, time and weekday of tweet's postage | Thu Mar 23 13:00:20 +0000 2017 |
| favorite_count | Number of users, which "liked" the tweet | 803757 |
| hashtags | List of keywords, which were marked with # in the tweet | [] |
| Id | Numerical Id of the user | 844896595179180034 |
| Id_str | Id in the string format | 844896595179180034 |
| Lang | Language of the tweet | en |
| Retweet_count | Number of users, which re-posted the tweet | 144566 |
| source | Determines operational system of the user, Android/iPhone/etc. | source: "textlessa href=ḧttp://twitter.com/download/iphoneërel=n̈ofollowẗextgreaterTwitter for iPhone</a>", |
| Text | The core text of the tweet | My heart goes out to the victims and their families in London. No act of terror can shake the strength and resilience of our British ally. |
| urls | The links to the articles/websites included in the tweet | [] |
| User – created_at | Date, time and weekday of user account creation | Mon Mar 05 22:08:25 +0000 2007 |
| User – description | Own text written by the user aimed to describe himself/herself | Dad, husband, President, citizen. |
| User – favourites_count | Sum of the tweets he/she liked | 10 |
| User – followers_count | Sum of the users, which follow the feed of the user | 86608691 |
| User – friends_count | Sum of the users, which are followed by the user | 630521 |
| User – lang | Language of the user account | en |
| User – listed_count | The number of lists the user is a member of | 222690 |
| User - location | Pre-defined location of the user | Washington, DC |
| User - name | Stated name of the user | Barack Obama |
| User – profile_background_color | Self-explanatory information related to account's design | 77B0DC |
| User – profile_background_image_url | Self-explanatory information related to account's design | http://pbs.twimg.com/profile_background_images/451819093436268544/kLbRvwBg.png |
| User – profile_image_url | Self-explanatory information related to account's design | https://pbs.twimg.com/profile_banners/813286/1484945688 |
| User – profile_banner_url | Self-explanatory information related to account's design | http://pbs.twimg.com/profile_images/822547732376207360/5g0FC8XX_normal.jpg |
| User – profile_link_color | Self-explanatory information related to account's design | 2574AD |
| User - profile_sidebar_fill_color | Self-explanatory information related to account's design | C2E0F6 |
| User - profile_text_color | Self-explanatory information related to account's design | 333333 |
| User – screen_name | Twitter name of the user as it appears in the feed | BarackObama |
| User - statuses_count | Sum of all of the user's posts | 15441 |
| User – time_zone | Pre-defined time-zone of the user | Eastern Time (US & Canada) |
| User - url | The link to user's personal webpage | https://t.co/93Y27HEnnX |
| User – utc_offset | The difference in seconds from Coordinated Universal Time | -14400 |
| User – verified | Determines, if the account of public interest is authentic. appears as the white tick mark in blue circle next to the user's name. | True |
| User_mentions | List of other Twitter users mentioned in the tweet | [] |

*Source: Author.*

**Table 6:** Example of evaluation of words in VADER

| **word** | Average score | Vector of individual scores |
|---|---|---|
| awards | 2,0 | [2, 2, 2, 2, 1, 2, 2, 3, 2, 2] |
| awesome | 3,1 | [3, 4, 2, 3, 2, 2, 4, 4, 4, 3] |
| awful | -2 | [-2, -2, -3, -3, -2, -3, 4, -3, -3, -3] |
| awkward | -0,6 | [-2, -1, -1, -1, -1, -1, -1, -1, 4, -1] |
| awkwardly | -1,3 | [-1, -1, -2, -1, -1, -1, -1, -2, -2, -1] |
| awkwardness | -0,7 | [-1, -2, -2, -1, -2, 2, -1, -1, 2, -1] |
| axe | -0,4 | [-2, 0, 0, 0, 0, 0, -1, -1, -1, 1] |
| axed | -1,3 | [-1, -2, 0, -3, -1, -2, -1, -1, -1, -1] |
| backed | 0,3 | [0, 0, 1, 0, 0, 0, 0, 0, 0, 0] |
| backing | 0,1 | [1, 1, -1, 0, 0, 0, -1, 1, -1, 1] |
| backs | -0,2 | [0, 0, -1, 0, 0, 0, 0, -1, 0, 0] |
| bad | -2,5 | [-3, -2, -4, -3, -2, -2, -3, -2, -2, -2] |

*Source: Author.*

**Table 7:** Perron's sequential procedure output using DF test for unit root in Trump's sentiment

| Trump's sentiment time series | | | |
|---|---|---|---|
| | 1. step | 2a. step | 3. step |
| n lags employed | 6 | 6 | 6 |
| t-test statistic for n lags | 2,05 | 2,05 | 2,130 |
| p-value for n lags | 0,044 | 0,044 | 0,036 |
| Test statistic ADF test | -2,687 | 3,74 | -2,742 |
| critical value | -3,45 | 1,12 | -2,58 |
| significance level | 5% | 5% | 10% |
| rejected | no | no | yes |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 8:** Perron's sequential procedure output using DF test for unit root in NASDAQ

| NASDAQ over the time span relevant for Trump | | | | | |
|---|---|---|---|---|---|
| | 1. step | 2a. step | 3. step | 4a. step | 5. step |
| n lags employed | 3 | 3 | 3 | 3 | 3 |
| t-test statistic for n lags | -2,890 | -2,890 | -2,030 | -2,030 | -2,000 |
| p-value for n lags | 0,005 | 0,005 | 0,045 | 0,045 | 0,048 |
| test statistic: ADF test | -3,025 | 4,990 | -1,489 | 4,120 | 1,941 |
| critical DF value | -3,450 | 1,120 | -2,890 | 0,500 | -1,95 |
| significance level | 5% | 5% | 5% | 5% | 5% |
| rejected | No | No | No | No | No |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 9:** Perron's sequential procedure output using DF test for unit root in first differences of NASDAQ

| First differences of NASDAQ over the time span relevant for Trump | |
|---|---|
| | 1. step |
| n lags employed | 4 |
| t-test statistic for n lags | -1,970 |
| p-value for n lags | 0,050 |
| test statistic: ADF test | -5,386 |
| critical DF value | -3,450 |
| significance level | 5% |
| rejected | Yes |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 10:** Perron's sequential procedure output using DF test for unit root in DJIA

| DJIA over the time span relevant for Trump | | | | | |
|---|---|---|---|---|---|
| | 1. step | 2a. step | 3. step | 4. step | 5. step |
| n lags employed | 3 | 3 | 3 | 3 | 2 |
| t-test statistic for n lags | -2,230 | -2,230 | -2,020 | -2,020 | 2,630 |
| p-value for n lags | 0,028 | 0,028 | 0,046 | 0,046 | 0,010 |
| test statistic: ADF test | -1,680 | 2,240 | -1,801 | 3,100 | 1,589 |
| critical DF value | -3,450 | 1,120 | -2,890 | 0,500 | -1,950 |
| significance level | 5% | 5% | 5% | 5% | 5% |
| rejected | No | No | No | No | No |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 11:** Perron's sequential procedure output using DF test for unit root in first differences of DJIA

| First differences of DJIA over the time span relevant for Trump | |
|---|---|
| | 1. step |
| n lags employed | 2 |
| t-test statistic for n lags | 1,950 |
| p-value for n lags | 0,054 |
| test statistic: ADF test | -5,098 |
| critical DF value | -3,450 |
| significance level | 5% |
| rejected | Yes |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 12:** Perron's sequential procedure output using DF test for unit root in S&P 500

| S&P 500 over the time span relevant for Trump | | | | | |
|---|---|---|---|---|---|
| | 1. step | 2a. step | 3. step | 4a. step | 5. step |
| n lags employed | 3 | 3 | 3 | 2 | 2 |
| t-test statistic for n lags | -2,670 | -2,670 | -2,500 | -2,500 | 2,910 |
| p-value for n lags | 0,009 | 0,009 | 0,014 | 0,014 | 0,004 |
| test statistic: ADF test | -2,401 | 3,650 | -1,787 | 2,790 | 1,665 |
| critical DF value | -3,450 | 1,120 | -2,890 | 0,500 | -1,950 |
| significance level | 5% | 5% | 5% | 5% | 5% |
| rejected | No | No | No | No | No |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 13:** Perron's sequential procedure output using DF test for unit root in first differences of S&P 500

| First differences of S&P over the time span relevant for Trump | |
|---|---|
| | 1. step |
| n lags employed | 8 |
| t-test statistic for n lags | -2,20 |
| p-value for n lags | 0,031 |
| test statistic: ADF test | -3,531 |
| critical DF value | -3,450 |
| significance level | 5% |
| rejected | Yes |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 14:** Perron's sequential procedure output using DF test for unit root in Obama's sentiment

| Obama's sentiment time series | |
|---|---|
| 1. step | |
| n lags employed | 11 |
| t-test statistic for n lags | -2,87 |
| p-value for n lags | 0,004 |
| test statistic: ADF test | -6,097 |
| critical DF value | -3,421 |
| significance level | 5% |
| rejected | Yes |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 15:** Perron's sequential procedure output using DF test for unit root in NASDAQ

| NASDAQ over the time span relevant for Obama | | | | | |
|---|---|---|---|---|---|
| | 1. step | 2a.step | 3.step | 4a.step | 5.step |
| n lags employed | 0 | 0 | 0 | 0 | 0 |
| t-test statistic for n lags | N/A | N/A | N/A | N/A | N/A |
| p-value for 1 lag | N/A | N/A | N/A | N/A | N/A |
| test statistic: ADF test | -2,814 | 3,990 | -2,772 | 4,00 | 0,436 |
| critical DF value | -3,421 | 1,130 | -2,870 | 0,510 | -1,950 |
| significance level | 5% | 5% | 5% | 5% | 5% |
| rejected | No | No | No | No | No |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 16:** Perron's sequential procedure output using DF test for unit root in first differences of NASDAQ

| First differences of NASDAQ over the time span relevant for Obama | |
|---|---|
| | 1. step |
| n lags employed | 0 |
| t-test statistic for n lags | N/A |
| p-value for n lags | N/A |
| test statistic: ADF test | -20,866 |
| critical DF value | -3,421 |
| significance level | 5% |
| rejected | Yes |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 17:** Perron's sequential procedure output using DF test for unit root in DJIA

| DJIA over the time span relevant for Obama | | | | | |
|---|---|---|---|---|---|
| | 1.step | 2a.step | 3.step | 4a.step | 5.step |
| n lags employed | 0 | 0 | 0 | 0 | 0 |
| t-test statistic for n lags | N/A | N/A | N/A | N/A | N/A |
| p-value for n lag | N/A | N/A | N/A | N/A | N/A |
| test statistic: ADF test | -2,643 | 3?490 | -2,614 | -3,48 | 0,25 |
| critical DF value | -3,421 | 1,130 | -2,87 | 0,51 | -1,95 |
| significance level | 5% | 5% | 5% | 5% | 5% |
| rejected | No | No | No | No | No |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 18:** Perron's sequential procedure output using DF test for unit root in first differences of DJIA

| First differences of DJIA over the time span relevant for Obama | |
|---|---|
| | 1. step |
| n lags employed | 0 |
| t-test statistic for n lags | N/A |
| p-value for n lags | N/A |
| test statistic: ADF test | -22,378 |
| critical DF value | -3,421 |
| significance level | 5% |
| rejected | Yes |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 19:** Perron's sequential procedure output using DF test for unit root in S&P 500

| S&P 500 over the timespan relevant for Obama | | | | | |
|---|---|---|---|---|---|
| | 1.step | 2a.step | 3.step | 4a.step | 5.step |
| number of lags employed | 0 | 0 | 0 | 0 | 0 |
| t-test statistic for 1/5 lags | N/A | N/A | N/A | N/A | N/A |
| p-value for 1/5 lags | N/A | N/A | N/A | N/A | N/A |
| test statistic: ADF test | -2,841 | 4,080 | -2,851 | 4,14 | 0,285 |
| critical DF value | -3,421 | 1,130 | -2,87 | 0,51 | -1,95 |
| significance level | 5% | 5% | 5% | 5% | 5% |
| rejected | No | No | No | No | No |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*

**Table 20:** Perron's sequential procedure output using DF test for unit root in first differences of S&P 500

| First differences of S&P 500 over the time span relevant for Obama | |
|---|---|
| 1. step | |
| n lags employed | 0 |
| t-test statistic for n lags | N/A |
| p-value for n lags | N/A |
| test statistic: ADF test | -21,626 |
| critical DF value | -3,421 |
| significance level | 5% |
| rejected | Yes |

*Source: Author, Stata 12 output and (Harris & Sollis, 2003).*