# Does Reddit know better?

## The impact of emotions on Presidential approval rates of public opinion polls

Marlijn Kruitwagen (434212)
M.Sc. Behavioural Economics
February 7, 2017
*Master thesis*

Supervisor: Vitalie Spinu
Co-reader: Chen Li

---

**Abstract**

This study investigates whether public's emotional state relates to public opinion polls, in particular the approval rate of Barack Obama during his eight years of presidency of the United States. Given the positive penetration rates of Internet users and the recently released substantial database of Reddit comments, the main objective of this study is to determine to what extent public emotions as measured from a large-scale collection of Reddit comments can be used to track public opinion polls. Despite the frequent use of polls, one must acknowledge three flaws, i.e. inaccurate data, requirement of large amount of respondents and time consuming to accomplish. Therefore, polls are often expensive to conduct. As a reply to these shortcomings, sentiment analysis uses Natural Language Processing techniques (NLP) to determine opinions from online sources. Sentiment analysis tools that automatically extract opinions from textual comments are quick, inexpensive and applicable on a large-scale. Previous research in sentiment analysis mainly focuses on a two-dimensional scale (positive versus negative). Polls are often limited to two-dimensional division (approve versus disapprove). NLP tools can provide a multidimensional view on public opinion by means of emotion analysis. A linear regression model with AR(I)MA errors is fitted to investigate the main question. The answer to the main question is twofold; on one hand, no significant results were found and therefore the study emphasizes that the approach was not sufficient to confirm the ability of public emotion to track presidential approval rates. On the other hand, emotional shifts aligned to remarkable events were captured. As a consequence of this conclusion, I still believe that expanding social media tracking tools with emotional analysis can be beneficial over relative costly and time consuming surveys and public opinion polls. Further research should move beyond the idea of supplementing public opinion polls and instead focus on substituting the polls.

---

**Keywords:** Reddit, emotion analysis, natural language processing, public opinion polling, big data, dynamic regression model, lexicon.

ERASMUS UNIVERSITEIT ROTTERDAM

**Contents**

**List of tables**



**List of figures**

## 1 Introduction

Online surveys and polls are often recognized by companies and governments as sufficient tools to gain feedback on products, services and policies all over the world. Despite the frequent use of polls in divergent fields, one must acknowledge three flaws of the method (Cody et al., 2016). First of all, people are often not very motivated to complete a poll; hence the accuracy of the data might be questionable. Secondly, the validity of the analysis strongly depends on both the sample size and the proportion of the total population. To develop an accurate analysis, a substantial amount of respondents are required which makes the technique both expensive and time consuming. The third flaw therefore accentuates the limitations of polls due to both time and costs as the restrictions in number of topics and participants.

As a reply to these shortcomings, innovative companies such as Meltwater develop sentiment analysis software that automatically extracts sentiment and opinions from online sources such as Facebook and Twitter (Meltwater, 2016). Sentiment analysis and opinion mining is widely known as the use of Natural Language Processing techniques (NLP) to determine the opinions and private states (beliefs, feelings, and speculations) of people towards entities (Liu, 2012). The analysis is a two-dimensional approach which classifies the opinion as either positive or negative (Mohammad & Turney, 2013). Despite the numerous amount of research on the architect of sentiment analysis, there is little literature to be found on the emotion analysis of text. This is at least remarkable considering the decisive impact of emotions on opinions. Nevertheless, Facebook recently acknowledged the importance of emotion classification by extending the famous “like” button. Formerly, Facebook users could only express their feelings by either liking the content or not liking the content by doing nothing. Nowadays, users can express their opinions in a more extensive way by selecting the “love, haha, wow, sad, angry or like” button. This re-classification provides the social media giant detailed information on the opinions of users and is therefore able to conduct several types of comprehensive analyses.

As stated, sentiment analysis utilizes NLP techniques to detect and extract subjective information of texts (Martin & Jurafsky, 2000; Mohammad & Turney; 2013). In addition, the aim of NLP is to create tools with high accuracy rates of public opinion towards entities (Wired, 2016). There are currently many technological developments in the field of NLP. NLP techniques are already utilized to gain detailed insights in customer feedback and other marketing purposes (Meltwater, 2016). The aim of this study is to examine if emotion analysis can also be utilized to gain a better understanding of the public opinion in politics.

Public opinion polls are often used to extract public views regarding specific topics or events. Companies such as Gallup and Rasmussen Reports mainly focus on presidential elections, presidential approvals and consumer confidence (Gallup, 2016; Rasmus, 2016a). However, these polls often fail to accurately predict the outcome of presidential elections and other general approvals. For instance, almost all polls wrongly predicated a win for Hillary Clinton in the last presidential election of the United States of America (RCP, 2016c). Hence, it is both interesting and progressive to examine if emotions captured by NLP techniques can be related to public opinion polls.

Earlier research shows some positive results of sentiment analysis regarding the prediction of financial markets (Bollen et al., 2011a; Zhang et al., 2012; Rao & Srivastava, 2012; Sprenger et al., 2014; Sul et al., 2014) and public opinion polls (O'Connor et al., 2010; Tumasjan et al., 2010; Cody et al., 2016). Contrary to these findings, numerous studies have emphasized the techniques are not sufficient enough to capture the large complexity of text (Jürgens et al., 2011; Gayo-Avello, 2012). This study would like contribute to current research and move beyond the two-dimensional scale of sentiment to a multidimensional scale of emotions.

To examine if emotion analysis is a proper technique to determine the public opinion, I utilize data from the user-generated content website www.reddit.com, which will be further referred to as 'Reddit'. To my best knowledge, no earlier study has used Reddit comments for sentiment or emotion analysis which makes this research unique and progressive. Reddit has 234 million unique users and 8 billion monthly page views, creating the 9th most-visited website in the United States (US) (Reddit, 2016). This increase means that more people in the US use Reddit and therefore it could be interesting to examine whether Reddit is a powerful tool for public opinion mining simply due to its wide area of influence.

Given these positive penetration rates of Reddit users and the volume of comments on Reddit and the negative accuracy rates of public opinion polls, the relevance of a multidimensional analysis of online texts cannot be neglected. Inspired by the expansion of lexicon-based emotion words by Saif Mohammad (Mohammad & Turney, 2013) and the rich database of public opinions on Reddit, the objective of this study is to examine the usability of online text as an opinion tool. As will be elucidated in the theoretical framework of this thesis later on, the model will be considered accurate when it is capable of capturing the public opinion as extracted from public opinion polls. Adopting a multidimensional approach, which enriches the view on public opinions, can help to attain this goal. Based on a consolidation of the above, the following research question is formulated:

> **RQ:** To what extent can *public emotion* measured from a large-scale collection of *Reddit comments* be used to track *public opinion polls*?

The use of multidimensional insights in combination with a lexicon-based NLP technique to extract public opinion from Reddit is still in its infancy. Therefore, I conduct multiple analyses to see whether this lexicon-based technique can capture the public view. The final part of this report focuses on the results regarding the usability of Reddit as a public opinion tool guided by the behavioural economic principles of emotion classification.

The paper is structured as follows: in section 2, I provide a theoretical framework concerning sentiment and emotion analysis, natural language processing and lexicon-based techniques, public opinion polling systems and the user-generated content website Reddit. Besides, the hypotheses are presented. Section 3 gives an explanation of the different data sources. In section 4, I include an extensive explanation of the models and its aligned assumptions. In section 5, the analysis of the data gathered is conducted. Lastly, I draw conclusion from the analysis, which will be extensively discussed. Based on this discussion, I provide recommendations for further research.

## 2 Theoretical Framework

### *2.1 Sentiment and emotions*

So far, the body of research in emotion analysis seems scare. This is due to the long hold believe in psychology, economics and politics that the decision making process is assumed to be a rational choice theory. More recent research argues that emotions play a significant role in human decision making (Bollen et al., 2011b). Once appeared, emotions might affect how individuals perceive and take action in subsequent events even when normatively irrelevant (Renshon & Lerner, 2012). Below, an overview is provided of the implication of emotions in decision making models in divergent fields.

#### *2.1.1 Psychology of emotions*

Until recently, the rational choice framework was the predominated theory in the field of human decision making research. The human decision making process was assumed to be a cognitive process where decision makers behave rationally and make decision based on which consequences yield the highest utility (Loewenstein, 2000). In the late 1960s, behavioural decision theories emerged that criticized the traditional view. Behavioural decision theories emphasize the influence of cognitive errors on evaluating the likelihood of future consequences and the simplifying heuristics that people use to cope with the complexity of decision making (Loewenstein, 2000). At the time, behavioural decision theories in psychology studies undervalued the influence of emotions in the process. In the 1990s, research found many examples of the impact of emotions on judgment and choice (Lerner & Keltner, 2000; Forgas 1995). Moreover, research highlighted the increase in explanatory power of decision making models including emotions (Lopes & Oden, 1998; Mellers et al., 1997), where emotions are defined as the subjective feelings and thoughts (Liu, 2012).

Emotions can be incorporated in the decision making process in several ways (Loewenstein, 2000; Renshon & Lerner, 2012). The first category, integral emotions, indicate emotions that are clearly related to the decision (at time). Loewenstein (2000) distinguishes two ways of integral emotions entering the decision making process. Expected emotions include the predictions of emotional consequences of certain decisions, such as the regret an investor might face when investing in a stock that declines in market value a short time after. The possibility of regret in the future influences the decision making today. Immediate emotions reflect the emotions experienced at the time of decision making.

Incidental emotions compose the second category of emotions entering the decision making process. Incidental emotions are often assumed to be normatively irrelevant to the judgments and decisions at hand, however can still influence decision making in an unpredictable way (Renshon & Lerner, 2012). To conclude, psychological research argues that emotions, in addition to information, play a significant role in human decision making (Bollen et al., 2011b).

In order to analyze emotions to examine the relevance for decision making models, psychologists have proposed a number of theories that classify human emotions into categories (Mohammed & Turney, 2013). Emotion classification addresses how one emotion can be distinguished from another emotion. Frijda (2008) proposes three main models by means of emotion classification: the basic-emotion, the multi-componential and the hierarchical models. Basic-emotion models are categorical models, which emphasize that basic emotions are categorized based on facial expression and processes within the human body. Basic emotions are therefore seen as discrete categories, where the emotion can be expressed in various degrees. An example of a basic-emotion model is proposed by Ekman et al. (1972). Earlier research posits that every emotion could be categorized into ‘pleasure’ or 'pain'. Ekman et al. (1972) elaborate and argue that every emotion could be categorized into six basic emotions: anger, disgust, fear, happiness, sadness and surprise. In the following decades, nineteen basic emotions are added to the model (Ekman & Cordaro, 2011).

The second and third model are process models characterized by multiple dimensions in groupings. The multi-componential models consists of a bottom-up approach and the hierarchical models of a top-down approach. Multi-componential models define emotions as a feeling or thought that is expressed and experienced in differenced ways, but provides a coherent response. For example, the emotion 'anxiety' can be expressed in nervousness (affective), worries about failing (cognitive), increased heart beat (physiological) and anxious facial expressions (expressive). A third category of models is introduced by Frijda (2008), which tends to outperform the preceding two models: hierarchy models. These models attempt to explain human emotions by means of multiple dimensions. The dimension are connected by psychology and neurobiology systems. This contradicts the basic-emotion model where emotions are assumed to be triggered by separate neural systems (Posner, Russell & Peterson, 2005). One of the most prominent hierarchical models is the wheel of emotions introduced by Plutchik in 1980 (Figure 1). Plutchik describes the relations among eight primary emotions, which are aligned to the colors in the wheel. The vertical dimension elaborates on the intensity of the emotion and the circle resembles the similarity among emotions.

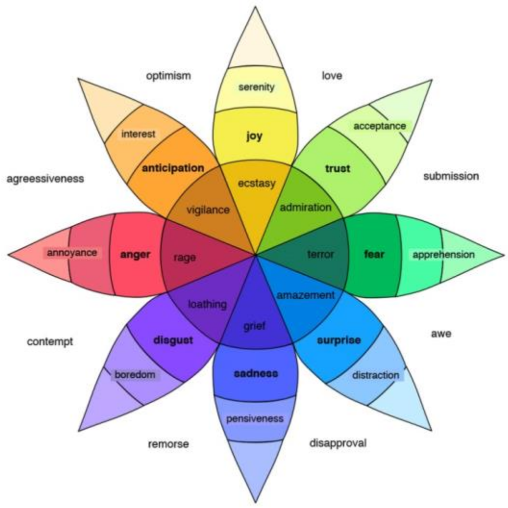



Figure 1. The circumplex model (Plutchik, 1980)

The eight primary emotions are arranged in four opposite pairs; joy versus sadness, trust versus disgust, anticipation versus surprise and fear versus anger. Plutchik (1980) arranges the four pairs of opposite primary emotions into a three-dimensional model. For example, higher intensity of the primary emotion anger is rage and lower intensity of the primary emotion anger is annoyance. Furthermore, anger is most similar to the primary emotions disgust and anticipation. The emotions in the blank spaces are primary dyads emotions and highlight the combination between primary emotions, aggressiveness for example is a mixture of anger and anticipation, contempt a mixture of anger and disgust.

To conclude, disparity among scholars about emotion classification is still present. The application of emotion classification is of great importance to determine whether a basic or process model might be relevant.

### *2.1.2 Emotions and economic theory*

Some of the earliest utility models in economics viewed utility as the net sum of positive over negative emotions (Bentham 1789). Neoclassical economists shifted the focus away from emotions to preferences, in accordance with the rational choice theory. At the time, emotions were only emphasized to express certain issues rather than adopted into economic models (Elster, 1998). In psychology, the last couple of decades are characterized by an incremental focus on behavioural decision theories accentuating the influence of emotions in decision making. In economics, this shift of focus is driven by the field of behavioural economics.

In order to increase the explanatory power of economic models, behavioural economics includes the effect of psychological, social, cognitive, and emotional factors on people's observable economic behaviour (Wilkinson & Klaes, 2012). The largest body of research in this field focusses on the cognitive side of psychology. This is due to larger contribution of cognitive psychologists, such as Kahneman & Tversky (Elster, 1998). When emotions are addressed in behavioural economics, they are often still classified as positive or negative. This is remarkable since emotions are argued to be inevitable and inseparable parts of human existence that cannot be limited to negative and positive classification (Elster, 1998). Besides, adding emotions as determinants can improve the predictive power of economic decision making models (Wilkonson & Klaas, 2012). This is illustrated in the field of behavioural finance where the following example shows that it is reasonable to assume that public emotions and sentiment can drive stock market values as much as news (Bollen et al. 2011b).

The traditional view of the financial paradigm implies several assumptions, such as that an investor is rational and uses laws of probability to form beliefs. When an investor does not behave rationally, it behaves randomly. This means that financial markets follow a random walk, where variations in financial markets are in a way unpredictable (Sewell, 2011). Due to these assumptions of the efficient market hypothesis (EMH), emotions were often neglected in stock price analysis. More and more studies discovered curious patterns in asset returns that could not be explained using the traditional view (Hansen & Shiller, 2013). An alternative view is developed, predominantly known as behavioural finance, which tries to explain the curious patterns (Barberis & Thaler, 2003). One of the core assumption of the behavioural finance perspective is that individuals are often subject to biases of judgment and errors of preference, which are the main causes of imbalances between estimated and actual market results. Sometimes, this yields the most optimal situation. More often however it leads to consequential biases in financial markets. These biases often occur after using judgmental heuristics, which can lead to systematic errors. These emotional biases are in a way predictable, to the extent that they rely on seasonal patterns, daylight and special events (Baltussen, 2009).

To conclude, the field of behavioural finance shows that identifying emotional biases contribute to the prediction of patterns in financial markets.

### *2.1.3 Emotions and politics*

The field of politics always acknowledged the relevance of emotions in political decisions and the deviations from rational choice, but appeared to be mainly dominated by the assumption that political subjects were essentially rational actors maximizing strategic interest. Normative political theorists have long considered emotion integral to understanding politics, however often with the purpose to constrain and minimize its potential harm on the deliberation process. Positive political theorists often neglected the importance of emotions. Fortunately, a growing number of scholars recognize the relevance of emotions within political studies (Clarke, Hogget & Thompson, 2006; Groenendyk, 2011; Marcus, 2000).

Based on the adopted view of positive political theorists, voting decisions are often seen by whom as a decision making process based on utility calculation of the consequences of the decision. In the beginning of the 20$^{th}$ century, scholars started to recognize deficiencies in the amount of information voters could process and the unbounded capacity of time and attention. By the end of the 20$^{th}$ century, scholars began to recognize the use of heuristics and aligned biases in voters’ behaviour (Groenendyk, 2011). Winter (2014) provides illustrative examples of the occurrence of irrationality within voting behaviour. For example, the decision to actually go out and vote can be described as an irrational decision. If the decision is interpreted as rational, the benefits of the considerable effort of voting should outweigh the potential costs. However, the individual influence of the outcome of an election is virtually zero, since the outcome of a national election cannot be determined by one single vote. Political scientists refer to this as the ‘voting paradox’. Furthermore, in a world of rationality, political debate would be minimized since exposure to the same facts should result in the same conclusion. Fact is that bounded attention causes people to focus of evidence that confirms their political orientation Winter (2014).

These deviations from rational choice in political behaviour substantiate the claim of the behavioural choice theory (Clarke, Hogget & Thompson, 2006; Groenendyk, 2011; Marcus, 2000). Therefore, it seems reasonable to assume that people engage in politics for more reasons than the utility of the consequences. An example of such reasons are emotions.

Extracting and identifying emotions to add as determinants in models can increase the explanatory power of economic decision making models. Since the same deviations from rational choice are highlighted in economics, emotions can perhaps also provide increasing explanatory power for decision making models within politics.

To conclude, psychological research argues that emotions, in addition to information, play a significant role in human decision making and the application of emotion classification is of great importance to determine whether a basic or process model might be relevant (Bollen et al., 2011b; Plutchik, 1980). The field of behavioural finance shows that identifying emotional biases contribute to the prediction of patterns in financial markets. Previous research (Bollen et al., 2011a; Zhang et al., 2012; Rao & Srivastava, 2012; Sprenger et al., 2014; Sul et al., 2014) merely focuses on the application of sentiment and emotion analysis in financial markets. This study explores the application of emotion analysis within political behaviour, since the deviations from rational choice in political behaviour substantiate the claim of the behavioural choice theory. The emotional analysis is performed using the classification model of Plutchik (1980). Most previous models of emotion extraction focus on facial expressions, gestures and postures. This study extracts emotions in language through words with the use of NLP techniques.

### *2.2 Emotion classification in Natural Language Processing*

Detecting and extracting subjective information in text documents is a very interesting research domain, since the amount of online documents and online opinion sharing is growing rapidly. North America is the continent with the highest Internet penetration rates, 89%, among all continents (World stat, 2016). An appropriate analysis of these texts can provide a rich database and have monetary benefits over surveys and polling systems. An appropriate analysis, known as sentiment analysis, has increased attention within the academic field. Sentiment analysis usage has also expanded into a variety of domains, from retail, financial services and healthcare to social event and political elections (Liu, 2012).

The field of NLP sentiment analysis is often explicated as the automatic extraction of quantitative opinions from subjective text (Nasukawa & Yi, 2003; Chowdhury, 2003; Martin & Jurafsky, 2000). NLP is a very active research area, however it took the field until 2000 to acknowledge the relevance of sentiment analysis. The relevance was instigated by the expansion of social media and the availability of many new sources of individual opinions and experiences (e.g. Facebook, Twitter, micro-blogs, comments, reviews, forums) (Liu, 2012). At the same time, increasing knowledge and developments occurred within NLP techniques. To extract sentiment and emotion from text, NLP makes a distinction between sentiment and emotion classification.

Sentiment classification is often regarded as a classification into two categories: positive and negative, where positive often expresses some desired state of quality while negative sentiment words are used to express some undesired state or qualities (Liu, 2012). Emotion classification adopts a more multidimensional view on desired states and focuses on the intensity of emotions and the mixture between motions (Plutchik, 1980; Frijda, 2008). Multiple sets of classification techniques have been developed within the field of NLP.

The first technique relates to unsupervised learning (Martin & Jurafsky 2000). Unsupervised learning techniques approach a problem without a precise idea or validation option of the results. The goal is to describe the structure or distribution from unlabeled data to gain knowledge about the data. Examples of approaches to unsupervised learning include clustering and neural networks. The second technique called supervised learning is applied when a certain dataset is given and there is a perception on what the results should look like (Liu, 2012). Thus, the task is to approximate a mapping function from labeled training data in such a way that new input data can also predict the output variables. Examples of supervised learning techniques are naïve Bayes classification and support vector machines (SVM). A third method to automatically extract sentiment and emotion from text is known as the lexicon-based approach. This approach involves calculating orientation for a document from the semantic orientation of words or phrases in the document. This is known as sentiment or affective lexicon (Martin & Jurafsky, 2000). Examples of such lexicons are EmoLex (Mohammed & Turney, 2013), WordNet Affect Lexicon (Strapparava & Valitutti, 2004), the General Inquirer (Stone et al., 1966) and MPQA Subjectivity lexicon (Wilson et al., 2005).

### *2.3 Public opinion polling*

Public opinion polling is a method to extract the public views regarding specific topics or events. The definition used by one of the largest performance-management consulting companies named Gallup is: ”A scientific, nonbiased public opinion poll is a type of survey or inquiry designed to measure the public's views regarding a particular topic or series of topics” (Gallup, 2016). For over a century, public opinion polls have been widely known as an instrument for public opinion mining.

In the early $19^{th}$ century the method was also known as ’straw polling’. Although, polling was not a very reliable measure due to its nonrandom sample, it was widely used for political parties to check the trend and sentiment of the population regarding political candidates (Gallup, 1948). One of the motivators for this development is the application of a poll as propaganda instrument, also known as the ‘band-wagon effect’. For example, a poll publishes very favorable numbers for a certain presidential candidate. This outcome can motivate other (floating or swing) voters to also favor this specific candidate, since other voters also adopted this view (Katz & Cantril, 1937; Gallup, 1948). The development of public opinion polls, encouraged by George Gallup, Archibald Crossley and Elmo Roper, created a widely used measure for research on attitudes and opinion and behaviour that continues today. However, whether the measures are reliable is very dependent on the source. For example, radio or TV shows often ask listeners to call and explain their opinion about a certain topic. These results cannot be reflected as scientific since the sample is not random (Gallup, 2016). Widely known public opinion polling companies in the United States are Gallup, Rasmussen Reports, Ipsos, Nielsen and Pew Research Center. They often work in collaboration with newspapers and news channels such as LA Times, CBS News, Fox News and ABC News and universities or research centers (RCP, 2016a). The most popular polls regard presidential elections, approval ratings and consumer confidence indices (RCP, 2016a). Below, an overview of previous research regarding public opinion polls and sentiment analysis is provided.

The main body of previous research regarding social media and public opinion polls is conducted using Twitter messages. One of the earliest papers discussing the feasibility of Twitter data to predict traditional polls is O’Connor et al. (2010). Although some significant results are gathered, O’Connor et al. (2010) only consider sentiment word frequencies and no correlation is found between electoral polls and Twitter sentiment data. Tumasjan et al. (2010) try to predict the German 2009 Federal election and state that a mere count of tweets mentioning a party or candidate accurately reflects the election results. However, Jürgens et al., (2011) comment that Tumasjan et al. (2010) make very arbitrary choices such as only using a subset of parties in their analysis. Sang & Bos (2012) and Skoric et al. (2012) also argue that applying sentiment analysis can improve performance and there consists a certain amount of correlation between Twitter and votes. Saleiro et al. (2015) design and implement the POPmine system, which is able to collect texts from web-based media and social media, and translate them into polarity measures. It is one of the first open source platforms that integrates data collection, information extraction, opinion mining and visualizes the extracted data. However, this model only considers positive versus negative measures and does not necessarily predict polls. Cody et al. (2016) conduct a sentiment analysis using Twitter messages and find that Twitter sentiment predicts Obama job approval three months in advance and correlates with surveyed consumer sentiment. However, Cody et al. (2016) only perform an emotion analysis on ambient happiness.

To conclude, some research has already been conducted in the prediction of public opinion polls using Twitter data. Many results have their limitations regarding predictions, but multiple correlations were found.

### *2.4 Reddit*

As discussed earlier, many online sources are analyzed to predict public sentiment and emotion. The recently released large dataset of Reddit posts and comments creates a very interesting new online source for analysis. Reddit describes itself as “the front page of the internet” and the website as “a source of what is new and popular on the web” (Reddit, 2016). Reddit is home to 234 million unique users and 8 billion monthly page views, creating the 9$^{th}$ largest website in the United States.

The Reddit website founded in June 2005 is a platform including user-generated content. Every registered member can upload content, such as text posts or direct links. Reddit has an original way of organizing content with the use of approximately 1 million subreddits. Subreddits are individual forums for similar topics or interests, which can serve many different purposes (e.g. asking and answering thought-provoking questions, deposit humorous content, news and discussion about U.S. politics). The largest Reddit subreddits based on the amount of subscribers are ‘AskReddit’, ‘funny’ and ‘todayilearned’. More than 15 million Reddit users are subscribed to these subreddits. The subscription to subreddits creates a personalized profile for every user (Reddit, 2016). The subreddits provide a large benefit for the plausibility of the data, since the data of interest is easily extracted. The creation of communities and subreddits also contributes to less noise in the data, due to the sense of belonging and feeling of validation within the community (Anderson, 2015).

Reddit does not provide exact figures of demographics, however a study from 2013 indicates that 6% of the online adults are Redditors with male ages 18-29 representing the largest group (Duggan & Smith, 2013). Also, Statista (2016) indicates that males form a proportionally larger group of the Reddit users. The average unique visit is about 13 minutes and users generate 25 million votes per day (Reddit, 2016). Since Reddit accounts for such a large amount of users it provides a very useful open source for a variety of opinions and ideas of individuals and communities. The amount of Reddits unique visitors per month increased from 40 million visitor in July 2012 to 234 million unique users in April 2016 (Reddit, 2016). It could be interesting to examine whether Reddit is a powerful tool for public opinion mining simply due to its usage penetration. Despite the high usage penetration rates, I need to be aware that the subreddit community may not be representative of the total population due to the characteristics of Reddit users compared to US population with voting rights.

To conclude, while most recent studies use the microblogging website Twitter to perform sentiment analysis, Reddit recently released a substantial amount of data containing posts and comments of the user-generated social news website. Due to the volume of comments and high penetration rates, the potential of the data cannot be neglected. Most recent studies perform a two-dimensional sentiment analysis of positive versus negative views. While it is difficult to determine how NLP will evolve in the future, it seems reasonable to assume that sentiment analysis needs to move beyond the two-dimensional scale. Where politics cannot be classified to a simple left or right scale, the same accounts for sentiment. Behavioural economists and political scientists should embrace the importance of multidimensional emotion analysis, since emotions can profoundly affect individual behaviour and decision making. For that reason, this study expands the study of O'Connor et. al. (2010) and applies a multidimensional emotion analysis on the large database of Reddit comments within the field of politics. This study especially focuses on the public opinion poll known as 'Obama job approval'. Therefore, the following main hypothesis is formulated:

$H_{main}$: Emotions are able to track the presidential approval rate of public opinion polls.

### *2.5 Multidimensional exploration of emotions*

Previous research has emphasized the role of sentiment and emotion in the decision making process (Lerner & Keltner, 2000). Emotions are therefore assumed to have profound impact on political decision making (Clarke, Hogget & Thompson, 2006; Groenendyk, 2011; Marcus, 2000). To study the interaction between emotions and the effect of emotions on public opinion polls, the eight-dimensional emotion framework by Plutchik (1980) is considered (Figure 1). Each emotion is individually explored and sub-hypotheses are created to address the expectations towards the presidential approval rate of Obama.

Anger tends to instigate discussion among many scholars. Previous research often perceives anger as a negative emotion, which is closely related to fear. More recent research classifies anger as an "approach" emotion, which consequences are more related to enthusiasm than fear (Brader & Marcus, 2013). Anger induces feelings of empowerment, confidence and optimism (Plutchik, 1980). In political decision making, anger motivates people to confirm prior convictions and reduces sensitivity to opposing points of view (Brader & Marcus, 2013). Therefore, I hypothesize:

$H_{anger}$: Anger leads to an increase in the presidential approval rate of Obama.

Anticipation is seen as the act of preparing for an expected or future event (Plutchik, 1980). Anticipation often enhances other emotions which provides difficulties in examining the individual effect on decision making. Anticipation is therefore seen as an indirect emotion that is more connected to other emotions than involved in individual decision making (Plutchik, 1980). The following hypothesis is formulated:

$H_{anticipation}$: Anticipation has no direct effect on the presidential approval rate of Obama.

Disgust is an emotion that is often highly connected to anger. Many studies use disgust as an indicator term for anger. The main difference is that disgust induces boredom and averse behaviour (Brader & Marcus, 2013). Only recently scholars acknowledged the existence of disgust, beyond the physical state, in moral judgment (Schnall et al., 2008). The body of research examining the relation between disgust and decision making is still limited. Scholars confirm the connection between disgust and averse behaviour, which is also shown in fear. Therefore, I hypothesize that:

$H_{disgust}$: Disgust leads to a decrease in the presidential approval rate of Obama.

The largest body of research in emotion analysis examines the influence of fear. Fear is often related to anxiety and classified as a “defensive” emotion (Plutchik, 1980). Fear might be triggered by feelings of danger or threat. For example in politics, concerns about political leaders, news about a possible loss of favorable candidate or possibilities of terrorist attacks (Brader & Marcus, 2013). As a consequence, fear motivates people to engage in averse behaviour associated with negative future consequences. The hypothesis defined is:

$H_{fear}$: Fear leads to a decrease in the presidential approval rate.

Joy is often related to serenity, ecstasy and refers most directly to feelings of pleasure. In the broader sense, joy induces a general satisfaction with life (Brader & Marcus, 2013). Joy is less specific than enthusiasm, but the two are closely related in the field of positive emotions and appear to be very similar in terms of responses. Enthusiasm in the political field can result in an increase in interests in political processes and motivation of political action. Enthusiasm can be triggered by news stories of desired politics or the positive favorable numbers in polls (Brader & Marcus, 2013). Therefore, I hypothesize that:

$H_{joy}$: Joy leads to an increase in the presidential approval rate.

Sadness and disappointment are often related to failure and loss and therefore classified as negative emotions. Plutchik (1980) relates sadness to grief and pensiveness. It seems reasonable to assume that sadness and disappointment influence political behaviour, but previous research mostly focuses on high-arousal emotion such as fear and anger. Since sadness is classified as a negative emotion, the following hypothesis is defined:

$H_{sadness}$: Sadness leads to a decrease in the presidential approval rate.

Surprise is one of briefest state of emotion and relates to amazement and distraction. Surprise often co-occurs with anticipation. Certain events can occur without anticipation and induce surprise. Depending on the event, surprise can either lead to a positive or negative mindset (Plutchik, 1980). Many scholars assume that surprise mainly influences other emotions and therefore received little attention among scholars in political science (Brader & Marcus, 2013). The following hypothesis is stated:

$H_{surprise}$: Surprise has no direct effect the presidential approval rate of Obama.

Trust is related to acceptance and admiration and evoked through familiarity (Plutchik, 1980). Trust is an important element in political science since security and cooperation among people and groups are induced. Despite, trust is seen as an indispensable positive element to create a stable and effective democracy (Brader & Marcus, 2013). Therefore, I hypothesize:

$H_{trust}$: Trust leads to an increase in the presidential approval rate of Obama.

This study is also interested in the interactions between emotion. Ekman & Cordaro (2011) state that fear often triggers anger. Anger emerges in situations where people feel threatened or encounter obstacles towards a goal. Therefore, anger and fear tend to co-occur (Brader & Marcus, 2013). The following interaction effect is formulated:

> $H_{fear\text{-}anger}$: The interaction between anger and fear has an negative effect on the presidential approval rate of Obama.

Plutchik explains primary dyads in the wheel of emotion (1980). Primary dyads are emotions that are mixtures of two primary emotions. For example, an interaction between anticipation and joy induces optimism and an interaction between sadness and surprise, disapproval. The following interaction effects are stated:

> $H_{anticipation\text{-}joy}$: The interaction between anticipation and joy has an positive effect on the presidential approval rate of Obama.

> $H_{surprise\text{-}sadness}$: The interaction between surprise and sadness has an negative effect on presidential approval rate of Obama.

## 3 Data

This study uses three different data sources: public opinion polls, Reddit comments and lexicons.

### *3.1 Polls*

This study uses three different public opinion polling sources: Real Clear Politics (RCP), Rasmussen reports and Gallup. RCP is the main poll and Rasmussen reports and Gallup are integrated for robustness tests. I extract the presidential approval rates, which are percentages between 0 and 100, from every poll. The percentages indicate the proportion of Americans who approve or disapprove the job Obama is doing as a president.

#### *3.1.1 Real Clear Politics*

RCP defines itself as the “trusted, non-partisan convener in a content-rich media environment” (RCP, 2016b). It provides public opinion polls on presidential approval ratings and elections on state and senate level. The RCP Obama job approval polling data is constructed combining 45 different polling companies, such as Reuters, Gallup, Rasmussen, the Economist and CBS as shown in appendix 1.

The daily data is extracted from the graph on the RCP website, requested with JavaScript Object Notation (JSON) and unfolded in R programming language. RCP consists of 2,828 observations in the period from 27$^{th}$ of January 2009 until 24$^{th}$ of October 2016. This period consists of 2,828 days, so the observations cover all days within this period. The average approval rate in this period is 47.56 percent (*SD*=4.58, *range*=40-65.50). The RCP approval time series are shown in Figure 2.

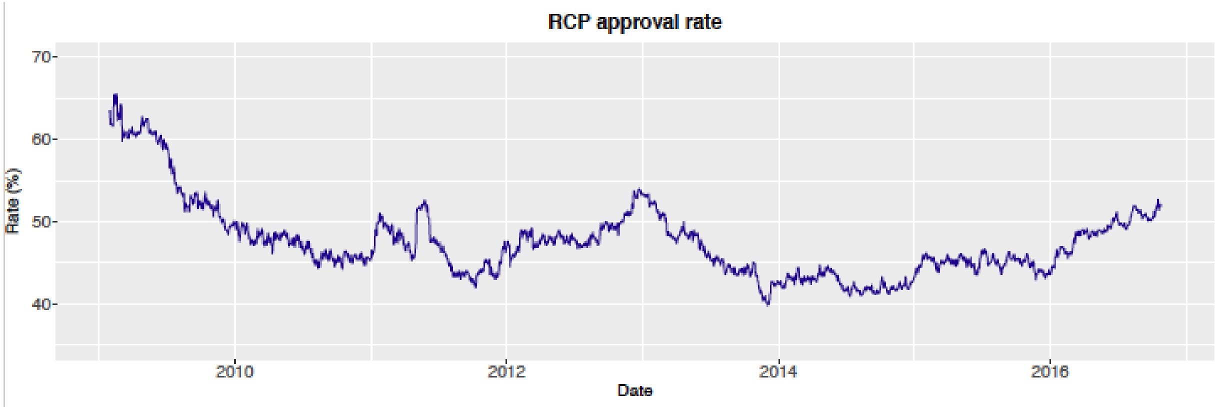



Figure 2. RCP presidential approval rate of Obama

### *3.1.2 Rasmussen Reports*

Rasmussen Reports is an American polling company, founded in 2003. Rasmussen Reports provides a wide variety of financial and econometric data, such as consumer index, investor index, employment index and banking and inflation trends. They also provide political data at national and state levels of elections, politics and presidential job approval ratings (Rasmus, 2016a). The sample used by Rasmussen to generate the data is based on likely voters (RCP, 2016a). The daily data is extracted from a table provided by the RCP website. The table contains the complete rating history and is called 'Obama Approval Index History' (Rasmus, 2016b). Rasmussen reports consists of 2,644 observations in the period from 6$^{th}$ of November 2008 until 26$^{th}$ of October 2016. This period consists of 2,911 days, so the observations cover 90.83% of the total days within this period. The average approval rate is 48.66 percent (*SD*=4.17, *range*=41-69). The Rasmussen Reports time series of the approval rate is shown in Figure 3.

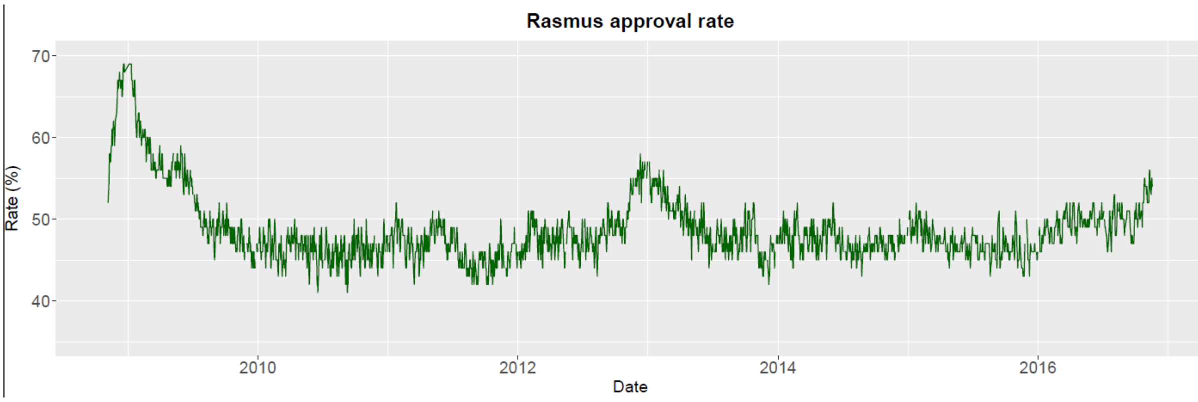



Figure 3. Rasmussen Reports presidential approval rate of Obama

### *3.1.3 Gallup*

Gallup is an American polling company existing for over 80 years. Gallup is a consulting company specialized in providing analytics and advice in a variety of industries. Besides the consulting activities, Gallup became known for conducting public opinion polls in several countries. Gallup performs thousands of daily telephone calls and face-to-face interviews to create daily, weekly and monthly tracking polls such as presidential approval, economic confidence and work engagement (Gallup, 2016). The sample used by Gallup to generate the data is based on adults (RCP, 2016a).

The daily data is extracted from the graph on the Gallup website, requested with JavaScript Object Notation (JSON) and unfolded in R. The data consists of 3-day rolling averages, therefore the last day is taken as reference point for the approval value. Gallup consists of 2,728 observations in the period from 24th of January 2009 until 9$^{th}$ of November 2016. This period consists of 2,846 days, so the observations have a coverage of 95.85% of the total days within this period. The average approval rate is 48.68 percent (*SD*=5.35,*range*=41-69). The time series approval rate of Gallup is as shown in Figure 4.

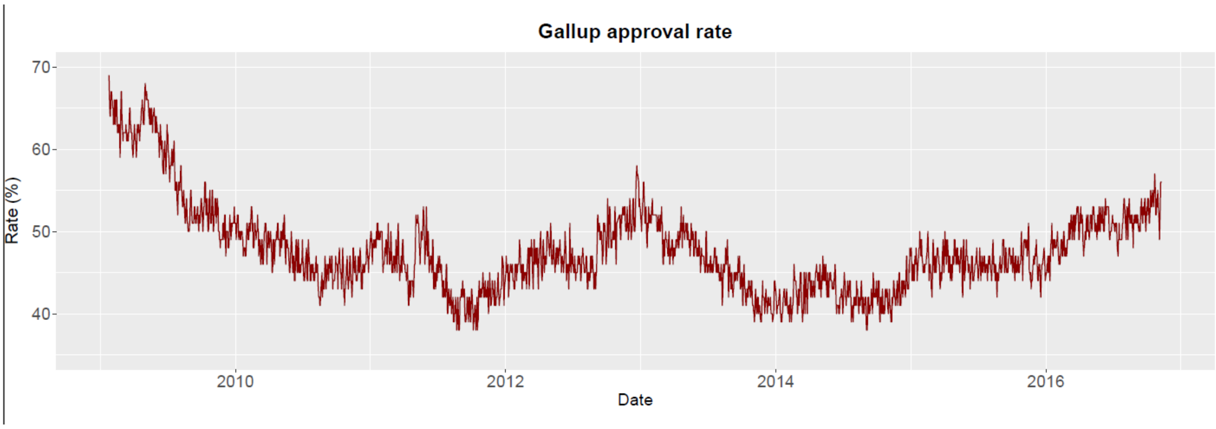



Figure 4. Gallup presidential approval rate of Obama

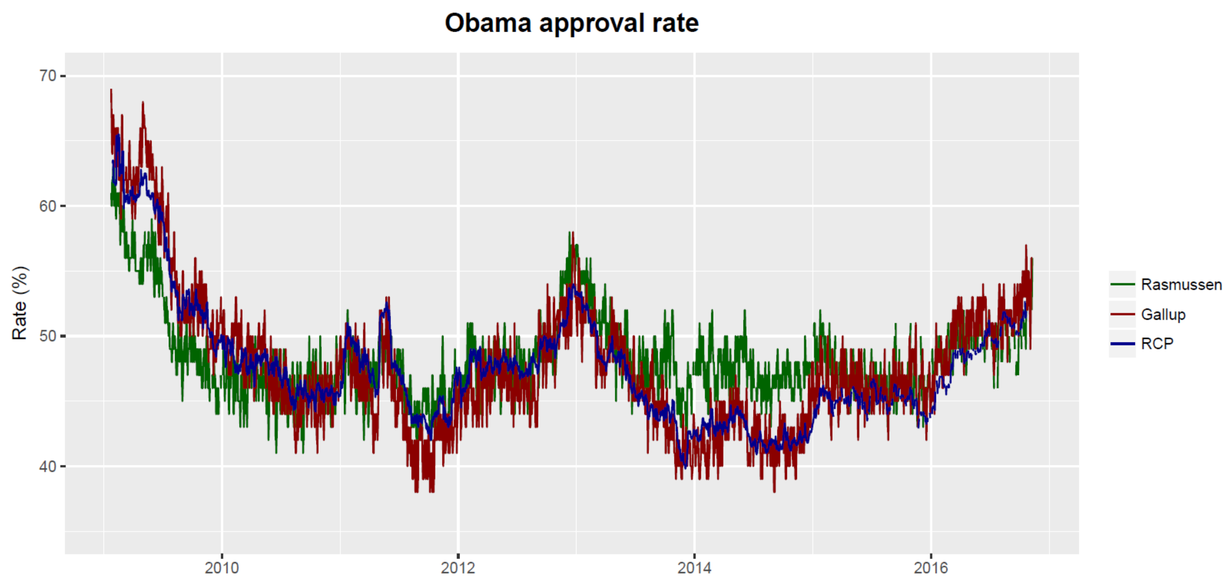
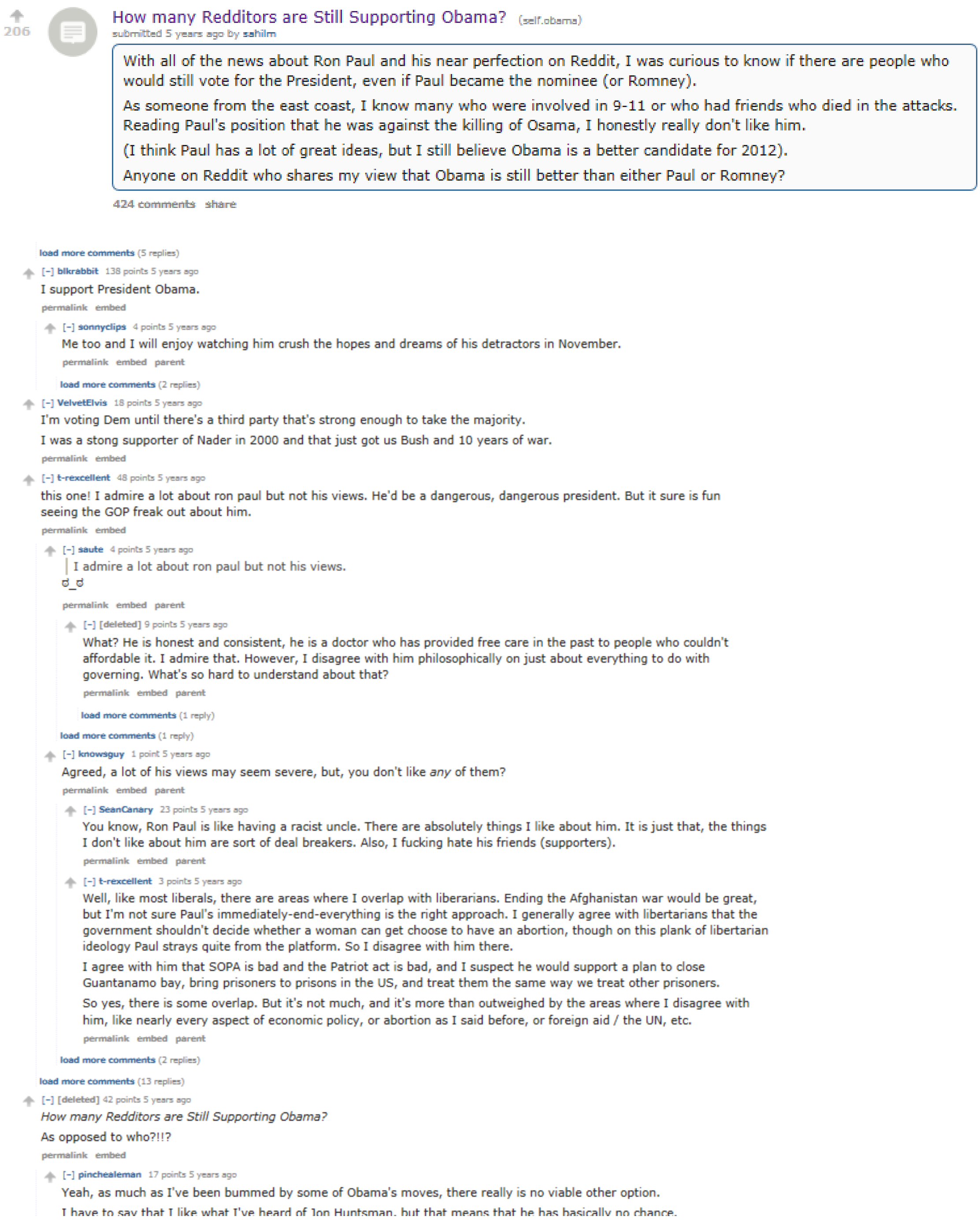



Figure 5. Presidential approval rates of Obama

Figure 5 provides an overview of the approval rate by three public opinion polls. Obama was officially elected as the president of the United States at 20th of January 2009. In his first year of presidency the approval rate dropped from around 65 percent in February 2009 to 47 percent in February 2010 for Gallup and RCP. The approval rate of Rasmussen declines more rapidly during this period. In the preceding year high positive approval peaks are displayed in January 2011 and May 2011, killing of Osama bin Laden. Throughout the year 2012 a steady increase is shown including a high positive approval peak in December 2012, the period after the re-election of Obama in November 2012 (Business Insider, 2012). The approval rate declines steadily throughout 2013 with the lowest negative approval in the beginning of December 2013. This seems to be due to the afterglow post-election period and increase in negative views about Obama. For example, due to new gun control policies after the Newtown school shootings in January, federal government's budget sequestration in March and October, the revelation of NSA spy programs by Snowden in June and problems with the healthcare insurance in October and November (Gallup, 2013). This low trend appears to maintain in 2014 with an average approval rate of 42.5 percent and a very low peak in November due to the midterm election and disapproval of the way the conflict of the Islamic State of Iraq is handled (Politico, 2014). In this period, Gallup and RCP seem to follow the same trend. Remarkably, Rasmussen displays a way higher average during the years 2013 and 2014. In 2015, the approval rate slightly increases to an average of 45 percent.

*3.2 Reddit*

The Reddit data is extracted from the Reddit API, which is a controlled way of accessing Reddit data. It answers to Hypertext Transfer Protocol (HTTP) requests with JavaScript Object Notation (JSON). This study will use comments from the subreddit '/r/politics'. The politics subreddit has over 3,2 million subscribers and the topic is ranked as the 55th largest subreddit.


206
How many Redditors are Still Supporting Obama? (self.obama)
submitted 5 years ago by sahilm

With all of the news about Ron Paul and his near perfection on Reddit, I was curious to know if there are people who would still vote for the President, even if Paul became the nominee (or Romney).

As someone from the east coast, I know many who were involved in 9-11 or who had friends who died in the attacks. Reading Paul's position that he was against the killing of Osama, I honestly really don't like him.

(I think Paul has a lot of great ideas, but I still believe Obama is a better candidate for 2012).

Anyone on Reddit who shares my view that Obama is still better than either Paul or Romney?

424 comments share

load more comments (5 replies)

[–] blkrabbit 138 points 5 years ago
I support President Obama.
permalink embed

[–] sonnyclips 4 points 5 years ago
Me too and I will enjoy watching him crush the hopes and dreams of his detractors in November.
permalink embed parent

load more comments (2 replies)

[–] VelvetElvis 18 points 5 years ago
I'm voting Dem until there's a third party that's strong enough to take the majority.
I was a stong supporter of Nader in 2000 and that just got us Bush and 10 years of war.
permalink embed

[–] t-rexcellent 48 points 5 years ago
this one! I admire a lot about ron paul but not his views. He'd be a dangerous, dangerous president. But it sure is fun seeing the GOP freak out about him.
permalink embed

[–] saute 4 points 5 years ago
> I admire a lot about ron paul but not his views.

ಠ_ಠ
permalink embed parent

[–] [deleted] 9 points 5 years ago
What? He is honest and consistent, he is a doctor who has provided free care in the past to people who couldn't affordable it. I admire that. However, I disagree with him philosophically on just about everything to do with governing. What's so hard to understand about that?
permalink embed parent

load more comments (1 reply)

load more comments (1 reply)

[–] knowsguy 1 point 5 years ago
Agreed, a lot of his views may seem severe, but, you don't like *any* of them?
permalink embed parent

[–] SeanCanary 23 points 5 years ago
You know, Ron Paul is like having a racist uncle. There are absolutely things I like about him. It is just that, the things I don't like about him are sort of deal breakers. Also, I fucking hate his friends (supporters).
permalink embed parent

[–] t-rexcellent 3 points 5 years ago
Well, like most liberals, there are areas where I overlap with liberarians. Ending the Afghanistan war would be great, but I'm not sure Paul's immediately-end-everything is the right approach. I generally agree with libertarians that the government shouldn't decide whether a woman can get choose to have an abortion, though on this plank of libertarian ideology Paul strays quite from the platform. So I disagree with him there.

I agree with him that SOPA is bad and the Patriot act is bad, and I suspect he would support a plan to close Guantanamo bay, bring prisoners to prisons in the US, and treat them the same way we treat other prisoners.

So yes, there is some overlap. But it's not much, and it's more than outweighed by the areas where I disagree with him, like nearly every aspect of economic policy, or abortion as I said before, or foreign aid / the UN, etc.
permalink embed parent

load more comments (2 replies)

load more comments (13 replies)

[–] [deleted] 42 points 5 years ago
*How many Redditors are Still Supporting Obama?*

As opposed to who?!!?
permalink embed

[–] pinchealeman 17 points 5 years ago
Yeah, as much as I've been bummed by some of Obama's moves, there really is no viable other option.

I have to say that I like what I've heard of Jon Huntsman, but that means that he has basically no chance.

Figure 6. Regular discussion on Reddit.com/r/politics

Figure 6 shows a regular discussion on the politics subreddit and the aligned comments. The title of the post is ‘How many Redditors are Still Supporting Obama?’ and in this case contains a question from the submitter. Reddit users can up or down vote the post. These up and down votes are also known as ‘Karma points’ (Anderson, 2015) and create the submission score of 206 (displayed next to the thumbnail in Figure 6). In this way, Reddit users determine what will be on the front page of the website. The post contains 424 comments, either directly aimed to the post or indirectly as a reply to another comment. Comments with the highest up votes will appear at the top and can be awarded with Reddit gold. Rubbish comments will get a lot of down votes and members of the communities will respond to spam and self-promotion (Anderson, 2015). When a comment is reported and identified as spam, the comments are distinguished by moderators or administrators.

The final dataset contains of comments regarding the topic Obama from 15$^{th}$ of October 2007 until 29$^{th}$ of February 2016. Each day is one unit of time, resulting in a timeline of 3,060 data points for time series. Per comment, the following information is provided (Github, 2016):

- *post_id:* the item identifier of the post
- *comment_id:* the item identifier of the comment
- *parent_id:* ID of the thing this comment is a reply to, either the link or a comment in it
- *author*: the account name of the poster. Often contains the value [deleted]
- *ups:* the number of upvotes per comment
- *score:* the net-score of the comment compiled by the number of upvotes minus the number of downvotes
- *gilded:* the number of times this comment received Reddit gold. Reddit provides the option to award a comment with Reddit gold, which means the comment has been well received. For example, due to the high level of humor or quality; this process is known as “gilding”
- *created_utc:* the time of creation in UTC epoch-second format
- *retrieved_on*: the time of extracting the data from Reddit API in UTC epoch-second format
- *controversiality*: a dummy variable where 1= equal amount of up- and downvotes and 0 = non equal amount of up- and downvotes

- *distinguished*: to allow determining whether the commenters have been distinguished by moderators or administrators. Where null = not distinguished, moderator = distinguished by moderator and admin = distinguished by admin.
- *url:* the link of the comment.
- *body*: the main text of the comment.

After removing subjects marked as 'moderator' and 'admin' in variable *distinguished*, we extract all subjects that include the word 'Obama' in their *body*. A total of 1,141,802 comments remain for pre-processing. These comments are distributed over time as shown in Figure 7.

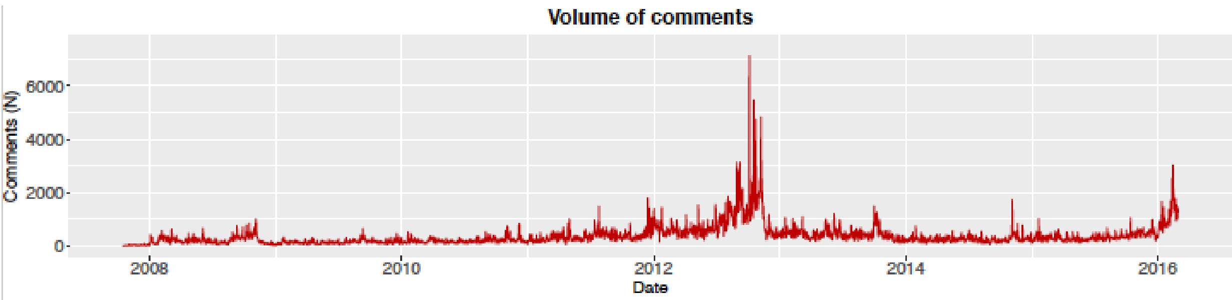



Figure 7. Volume of Reddit comments

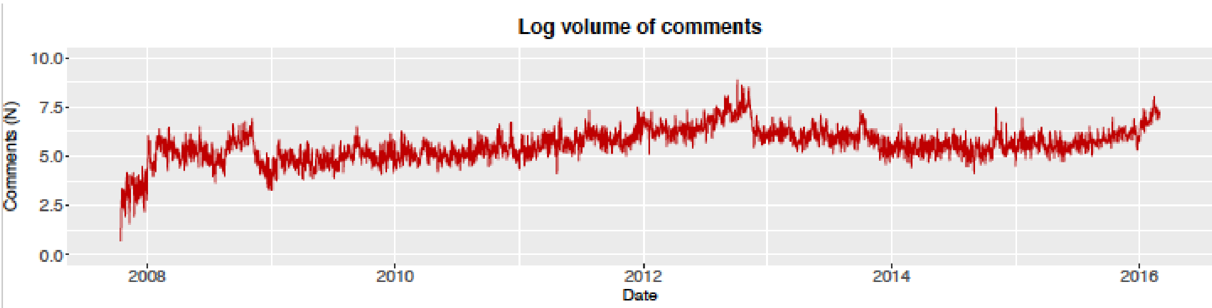



Figure 8. Log volume of Reddit comments

The large peak is centered around October and November, 2012 due to the United States presidential election. The same accounts for the increasing volume of comments in 2016. Reddit gained popularity over the years which could explain why larger average amounts of comments are captured in more recent years compared to earlier years. For example, in 2009 the average amount of comments covering the topic Obama was 151 per day. In 2015, this amount increased to 340 per day.

### *3.3 Lexicon*

This research uses dictionaries of words annotated with the words emotion orientation. The dictionary is called NRC Word-Emotion Association Lexicon, also known as EmoLex (Mohammad & Turney, 2013).

Mohammad & Turney (2013) designed the lexicon through a principle called the ‘wisdom of the crowd’ and classified the emotions according to the wheel of emotions by Plutchik (1980). Therefore, EmoLex focusses on joy, sadness, anger, fear, trust, disgust, surprise and anticipation. EmoLex provides many advantages over other lexicons and models, such as the possibility to perform both sentiment and emotion analysis and the inclusion of words that often occur on social media. The lexicon is created through asking annotators examples of words associated with different emotions and rely on intuition what the emotions mean and how text is used to express the emotion. These annotators are approached through the crowdsourcing platform Mechanical Turk, which is especially suited for tasks done over the internet or mobile phones. Low costs, less organizational overhead and quick turnaround of time provide advantages of the platform (Mohammad & Turney, 2013).

Table 1. Example of NRC Word-Emotion Association Lexicon

| **Word** | **Anger** | **Anticipation** | **Disgust** | **Fear** | **Joy** | **Sadness** | **Surprise** | **Trust** |
|---|---|---|---|---|---|---|---|---|
| abandon | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| accused | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| admiration | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| aggressive | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| alive | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| allure | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| ambitious | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*Note.* The table provides 7 examples of words in the NRC Word-Emotion Association Lexicon, where 0 denotes no association with emotion and 1 denotes association with emotion.

The annotated scores were analyzed and every word was given an individual score, as shown in table 1. For example, the word ‘abandon’ is associated with the emotions fear and sadness. The emotion lexicon can be very helpful in identifying and extracting emotions, but caution should be taken that sentiment is not simply an aggregation of lexicon-scores (Mohammad & Turney, 2013). The final list is extracted from the website of Saif Mohammad and consists of 14,182 words based on a general domain (NRC, 2016). In contrast to the NRC Hashtag Emotion Lexicon, which is based on tweets with emotion word hashtag from Twitter

(Mohammad & Turney, 2013). Since the analysis is performed on Reddit data, I expect that EmoLex provides a better fit.

## 4 Methodology

In order to investigate the relation of emotions extracted from Reddit comments with public opinion polls, the lexicon and Reddit data needs to be pre-processed. The goal is to derive an eight-dimensional daily vector of emotions and investigate its ability to track public opinion polls by means of a linear regression model with AR(I)MA errors.

### *4.1 Pre-processing lexicon*

First of all, the terms from the NRC Word-Emotion lexicon are converted to lower-case and stemmed. Stemming is a process where words are reduced to the basic form (Liu, 2012). Stemming reduces the size of words needed and captures as many words as possible for analysis. The lexicon is reduced from 14,182 to 10,943 terms. The downside of stemming is that words may have different connotation which is not captured by stemmed words. Stemming is necessary since EmoLex does not include every word individually, for example plural and singular forms. Table 2 shows how the original lexicon word is reduced to the stemmed form and the annotated scores are provided before and after stemming.

Table 2. Example of the change in annotated scores after stemming

| **Orig. Word** | **Stemmed Word** | **Anger** | **Anticipation** | **Disgust** | **Fear** | **Joy** | **Sadness** | **Surprise** | **Trust** |
|---|---|---|---|---|---|---|---|---|---|
| abandon | - | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| abandoned | - | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| abandonment | - | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| abandon | abandon | 0.6666667 | 0 | 0 | 1 | 0 | 1 | 0.333333 | 0 |
| abandoned | abandon | 0.6666667 | 0 | 0 | 1 | 0 | 1 | 0.333333 | 0 |
| abandonment | abandon | 0.6666667 | 0 | 0 | 1 | 0 | 1 | 0.333333 | 0 |

*Note.* The table provides an example of three original words, which are reduced to the same stem. The first three lines provide the annotated scores of the original words from the lexicon. The last three lines provide the annotated scores for the stemmed word. The scores of the stemmed word are averages of the aligned scores of original words.

### *4.2 Pre-processing Reddit comments*

The comments extracted from the Reddit API are preprocessed using the following steps:

- removal of all punctuations, such as brackets, hyphen, commas and dashes

- removal of all numbers and stripping the whitespace
- conversion to lower-case of all remaining characters
- removal of stopwords excluding negations
- separation of individual terms on white-space boundaries
- stemming of individual terms

The individual terms extracted from the Reddit data are filtered to contain only the terms from the NRC lexicon and matched to the eight-dimensional annotated scores of the NRC lexicon. Thus, the resulting Reddit dataset only contains stemmed terms of the lexicon in addition to negations. The total list consists of 20,586,195 terms in 922,480 unique comments. The time series of the daily number of unique terms is shown in Figure 9. The logarithmic time series are also provided in Figure 10.

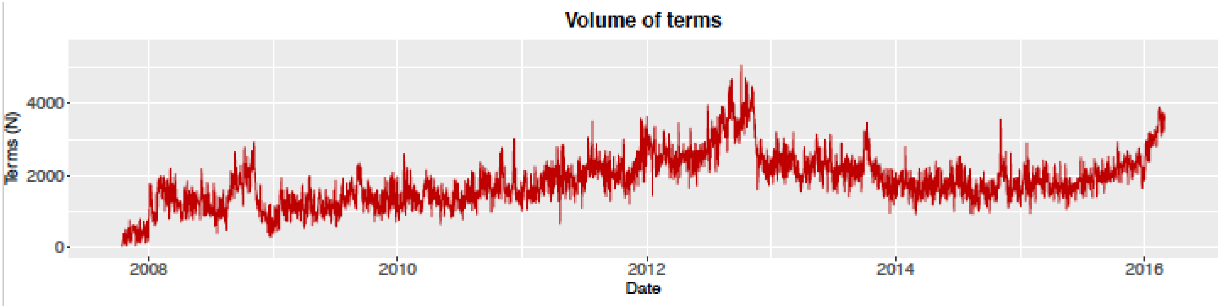



Figure 9. Volume of terms on Reddit

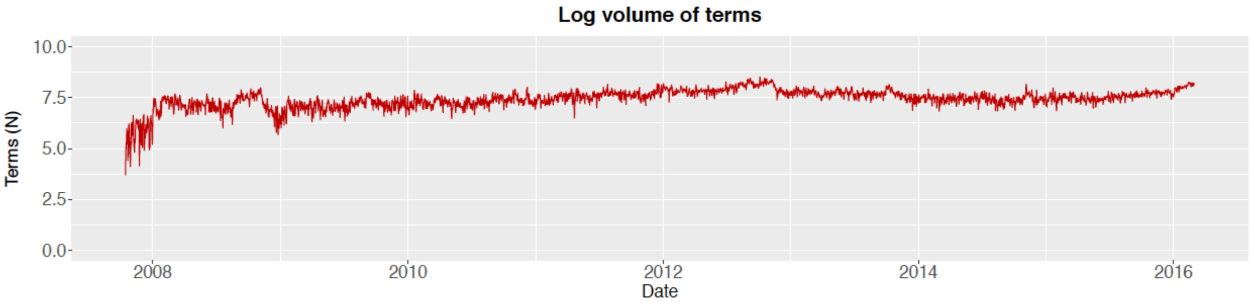



Figure 10. Log volume of terms on Reddit

The large peak in Figure 9 is centered around October and November, 2012 due to the United States presidential election. The peak is less extreme than the peak in the same period in the volume of comments (Figure 7). The increasing volume over time is also less visible in the unique terms than the comments. This is caused by the length of comments and volume of words. The comment can contain less lexicon-words in the comments. The logarithmic volumes in Figure 10 slightly increase over time.

Before aggregating the scores per day, transformation steps are needed. Since Reddit does not limit the length of the comments, the emotion scores need to be corrected for the amount of terms in the comment and the amount of times the term occurs over all comments. In order to do so, the method of term frequency-inverse document frequency (TF-IDF) is applied (Rajaraman & Ullman, 2011). One of the main goals of TF-IDF is to determine which words are useful indicators. This is valuable since the more frequent a word occurs does not necessarily indicate that this word is most significant. For example, the most common words are likely to be "the" or "and", which do not characterize emotions towards the topic Obama. Although these words are removed during stop-word removal, it is likely that the lexicon terms also contain words which tend to occur more often but do not characterize the opinion towards the topic Obama best (Rajaraman & Ullman, 2011). TF-IDF is computed as follows:

$$\mathrm{TF}_{\mathrm{tc}} = \frac{\mathrm{n}_{\mathrm{tc}}}{\mathrm{max}_{\mathrm{k}}\mathrm{f}_{\mathrm{kc}}} \tag{1}$$

Where $n_{tc}$ is the number of occurrences of term $t$ in *comment c*. $f_{tc}$ is normalized through dividing the frequency by $\mathrm{max}_k\mathrm{f}_{kc}$, the maximum number of occurrences of any term within comment $c$. Thus, the higher TF the more frequently a term occurs in comment $c$. IDF is computed to measure how much information a word provides, that is, whether the term is common or rare across all comments:

$$\mathrm{IDF}_{\mathrm{t}} = \log\frac{\mathrm{N}}{\mathrm{n}_{\mathrm{t}}} \tag{2}$$

Where $N$ represents the total number of comments in the corpus. $N$ is divided by the number of comments $n_t$ where term $t$ appears and is taken as a logarithm. Finally, TF-IDF can be calculated:

$$\mathrm{TFIDF}(\mathrm{t},\mathrm{c}) = \mathrm{TF}_{\mathrm{tc}} \times \mathrm{IDF}_{\mathrm{t}} \tag{3}$$

The higher the score of TF-IDF, the better the term characterizes the topic of the comment. After correcting the annotated scores of the terms for TF-IDF, the scores are corrected for negations. Down-weighting factors are added to reduce the impact of negations:

$$For\ all\ L < 0 \qquad e_t(L) = \begin{cases} 0.1e_t & L = -1 \\ 0.5e_t & L = -2 \\ e_t & L \leq -3 \end{cases} \tag{4}$$

where $e_t$ denotes the annotated score per term adjusted for TF-IDF. When a term lags one distance (*L=-1)* from a negation word, e.g. negation precedes the word, $e_t$ receives the down-weighting factor of 0.1. When a term lags two distances *(L=-2)* from negation, $e_t$ receives the down-weighting factor of 0.5. When a term lags more distances from the negation words ($L \leq -3$) $e_t$ does not receive a down-weighting factor. The aggregated emotion vector per day ($e_d$) is produced for the set of terms submitted on a particular date $d$ (denoted $T_d \subset T$) by averaging the emotion vector of the term submitted that day.:

$$e_d = \frac{\sum \forall t \in T_d e_t}{||T_d||} \tag{5}$$

*4.2.1 Intraday normalization*

Furthermore, I performed intraday normalization on the aggregated emotion vectors per day. Normalization removes the impact of factors that drive all emotions simultaneously. For example, the impact of trends and major events. Normalization can also contribute to the reduction of multicollinearity. Multicollinearity occurs when the emotion vectors are highly correlated. The idea behind intraday normalization is that the scores of the eight aggregated emotion vectors per day add up to 1:

$$e_d = 1 = anger_d + anticipation_d + disgust_d + fear_d + joy_d + sadness_d + surprise_d + trust_d \tag{6}$$

Where the eight transformed variables of individual emotions denote a proportion of the total emotion score of 1 per day.

*4.3 Linear regression model with AR(I)MA errors*

The goal of this study is to examine whether the aggregated emotions vectors of $anger_t$, $anticipation_t$, $disgust_t$, $fear_t$, $joy_t$, $sadness_t$, $surprise_t$ and $trust_t$, are related to $Y_t$, which represents the job approval rate of Obama. To investigate this relationship, the linear regression model with AR(I)MA errors is estimated. The model is adopted from Ruppert (2015), Hyndman & Athanasopoulos (2014) and Tsay (2005) and adjusted to this study.

The assumptions of linear regression models are linearity of the conditional expectation, independent noise, constant variance and Gaussain noise. Gaussain noise indicates normally distributed error terms, which is also known as the Gaussian white noise process[1]. When residual analysis of the linear regression models shows that the residuals are correlated, then one of the key assumptions of the linear model does not hold, and the results of the analysis are incorrect (Ruppert, 2015). The assumption of uncorrelated or independent errors for time series data is often not appropriate in business and economics because the errors commonly exhibit some serial correlation (Ruppert, 2015). A time series model, where autoregressive and moving average terms are included, can overcome this problem but only allows for inclusion of information from past observations of a series, and not for inclusion of other relevant information[2].

A solution to these problems is defined by Ruppert (2015): " replace the assumption of independent noise in the linear regression model by the weaker assumption that the noise process is stationary but possibly correlated". This can be achieved by assuming that the noise is an AR(I)MA process. The linear regression model with AR(I)MA errors combines the linear regression model and the AR(I)MA model for noise. The following linear regression model with AR(I)MA errors is estimated:

$$Y_t = \beta_0 + \beta_1 anger_{1,t} + \beta_2 disgust_{1,t} + \beta_3 fear_{1,t} + \beta_4 joy_{1,t} + \beta_5 sadness_{1,t} + +\beta_6 trust_{1,t} + \beta_7 surprise_{1,t} * sadness_{1,t} + \beta_8 joy_{1,t} * anticipation_{1,t} + \beta_9 anger * fear_{1,t} + n_t \quad (7)$$

where,

$$(1 - \emptyset B - \cdots - \emptyset B^p)n_t = \left(1 + \theta_1 B + \cdots + \theta_q B^q\right)\epsilon_t) \quad (8)$$

Where $Y_t$ is the presidential approval rate of Obama. The model has two error terms in place, where $n_t$ is the error from the regression model and $\epsilon_t$ the error from the AR(I)MA model. Only the AR(I)MA model errors are assumed to be white noise. The errors from the regression model are assumed to be stationary. $B$ is the backwards operator, $p$ determines the amount of autoregressive terms and $q$ determines the amount of moving average terms.

---

[1] $Y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \epsilon_t$ , where $\epsilon_i$ is called the noise, disturbances, or errors and it is assumed i.i.d. white so that: $\epsilon_1, \ldots, \epsilon_n$ $are\ i.i.d. with\ mean\ o\ and\ variance\ \sigma_\epsilon^2$ (Ruppert, 2015)

[2] $(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\, \epsilon_{t-1} + \cdots + \theta_q\, \epsilon_{t-q}$

The first part of the expression refers to AR and the second part refers to MA. It shows how $Y_t$ depends on lagged values of itself and lagged values of the white noise process (Ruppert, 2015).

An augmented Dickey-Fuller test is performed to check that $Y_t$ and all predictors appear to be stationary. One exception can be made when the non-stationary predictor variable is co-integrated with the stationary response variable (Hyndman & Athanasopoulos, 2014). Co-integration means whether there exist a linear combination between $Y_t$ and emotion variables. This is tested with the Johansen test for co-integration between time series. When the time series are not co-integrated, differences and transformations must be considered. Differencing might transform a non-stationary time series into a stationary time series.

When estimating the model, the sum of squared $\epsilon_t$ values need to be minimized and not those of $n_t$. The sum of squared $\epsilon_t$ is minimized in order to avoid suboptimal estimates of coefficients, incorrect statistical tests and 'spurious regression'. 'Spurious regression' might occur when the $p$-values of associated coefficients are very small. (Hyndman & Athanasopoulos, 2014).

In order to find the most parsimonious model to test the main hypothesis, I fit the proxy model with 0 moving averages (MA) and 0 autoregressive (AR) terms. Once the proxy model is estimated, the regression coefficients and the preliminary values of $n_t$, are calculated to select a more appropriate ARMA model for $n_t$, before re-estimating the entire model (Hyndman & Athanasopoulos, 2014). The autocorrelation function (ACF) and partial autocorrelation (PACF) plots provide more information about the amount of AR and MA terms that need to be added to the model. Section 5 elaborates on the use of (P)ACF and provides specific example for the Reddit dataset. AIC and BIC values present an indication of which model provides the best fit. The lower the values, the better the fit of the model. This study chooses AIC and BIC measurements over the F-test, because it is more appropriate to use for model selection since the statistical behaviour of F-test is not always known (Tsay, 2005). For linear regression models, AIC can be denoted as:

$$AIC = n\log(\hat{\sigma}^2) + 2(1+p) \qquad (9)$$

where 1+ $p$ is the number of parameters in a model with $p$ predictor variables; the intercept provides the final parameter. BIC replaces 2(1+$p$) in AIC by log($n$)(1+$p$) and therefore punishes for the addition of more predictors (Tsay, 2005). This is visible in section 5, where the BIC values become higher when models add more terms. Although the measure provides a good overview of the fit of the model, this study adopts a more parsimonious model based on more criteria or with a better economic rationale. To conclude, I check whether the $\epsilon_t$ series looks like white noise with the use of (P)ACF plots and the Ljung-Box test. The Ljung-

Box test is performed to test if the several autocorrelations are zero at various lags, where $H_0$ is rejected if the p-value is less than or equal to $\alpha$. If the p-value is larger than 0.05 the joint test confirms there is no serial correlation. More detail can be found in section 5, where the test is performed on the linear regression model, to confirm the presence of correlated errors and the regression model including AR(I)MA, to confirm the absence of correlated errors. I perform the Ljung-Box test instead of the Durbin-Watson test, because the Durbin-Watson test cannot be applied when capturing autoregressive effects in a model (Ruppert, 2015).

*4.3.1 Note on ARMAX models*

An important remark to linear regression modelling with AR(I)MA errors is the comparison with ARMAX models. Hyndman (2010) provides an overview of the difference in the 'ARIMAX model muddle'.

ARMAX models add the covariate $\beta X_t$ to the right hand side of the AR(I)MA equation. This is shown in the following function:

$$(Y_t) = \beta X_t + \phi_1(Y_{t-1}) + \cdots + \phi_p\left(Y_{t-p}\right) - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q} + z_t \qquad (10)$$

The presence of lagged values of the response variable in equation 10 causes that the coefficients are hard to interpret. The coefficient $\beta$ can only be interpreted conditional on the value of the previous values of the response variable. When rewriting the model using backshift operators, the AR coefficients cannot be isolated due to the covariates and the error term[3].

If a regression model with AR(I)MA errors is considered, the covariate $\beta X_t$ is embedded in the linear regression model equation. This is shown in the following function:

$$y_t = \beta X_t + n_t \qquad (11)$$

Where

$$n_t = \phi_1 n_{t-1} + \cdots + \phi_p n_{t-p} - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q} + z_t \qquad (12)$$

In this matter, the value of $\beta$ can be easily interpreted as the effect on $y_t$ when $x_t$ is increased by one since the lagged values of the response variable are not included. When rewriting the

[3] $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ . The AR coefficients get mixed up with both the covariates and the error term

equation using backshift operators the AR coefficient does not get mixed up with the covariates and error term[4].

Ruppert (2015) defines the linear regression model with AR(I)MA errors as an ARMAX model. Hyndman & Anthanasopoulos (2014) make a clear distinction and define the linear model with AR(I)MA errors as a dynamic regression model, which is different from the ARMAX model. The reason why this study prefers linear models over ARMAX models is that the coefficients are easier to interpret (Hyndman, 2010). Therefore, this study prefers the dynamic regression model of Hyndman & Anthanasopoulos (2014).

[4] $y_t = \beta x_t + \frac{\theta(B)}{\emptyset(B)} z_t$

Table 3. Descriptive statistics of main dataset

| **Variables** | Start | End | N | M | SD | Min | Median | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 Rcp - approval (DV1) | 2009-01-27 | 2016-10-24 | 2,828 | 47.56 | 4.580 | 40.00 | 46.90 | 65.50 | 1.431 | 5.474 |
| 2 Gallup - approval (RC1) | 2009-01-24 | 2016-11-09 | 2,728 | 47.68 | 5.350 | 38.00 | 47.00 | 69.00 | 1.248 | 5.034 |
| 3 Rasmus – approval (RC1) | 2008-11-06 | 2016-11-18 | 2,743 | 48.70 | 4.170 | 41.00 | 48.00 | 69.00 | 1.948 | 8.242 |
| 4 Anger (IV1) | 2007-10-15 | 2016-02-29 | 3,060 | .0123 | .0023 | .0007 | .0120 | .0303 | 1.119 | 7.375 |
| 5 Anticipation (IV2) | 2007-10-15 | 2016-02-29 | 3,060 | .0117 | .0023 | .0052 | .0113 | .0333 | 1.893 | 11.12 |
| 6 Disgust (IV3) | 2007-10-15 | 2016-02-29 | 3,060 | .0084 | .0017 | .0003 | .0082 | .0189 | .9299 | 5.409 |
| 7 Fear (IV4) | 2007-10-15 | 2016-02-29 | 3,060 | .0126 | .0024 | .0046 | .0122 | .0346 | 1.590 | 10.08 |
| 8 Joy (IV5) | 2007-10-15 | 2016-02-29 | 3,060 | .0101 | .0024 | .0036 | .0097 | .0482 | 2.930 | 30.13 |
| 9 Sadness (IV6) | 2007-10-15 | 2016-02-29 | 3,060 | .0115 | .0021 | .0019 | .0113 | .0451 | 2.626 | 31.38 |
| 10 Surprise (IV7) | 2007-10-15 | 2016-02-29 | 3,060 | .0066 | .0017 | .0030 | .0062 | .0275 | 2.563 | 21.96 |
| 11 Trust (IV8) | 2007-10-15 | 2016-02-29 | 3,060 | .0205 | .0034 | .0084 | .0200 | .0577 | 2.036 | 15.97 |

*Note.* DV = Dependent Variable. IV = Independent Variable. RC = Robustness check.
Emotions(4-11) are raw values adjusted for TF-IDF and negations.

## 5 Results

### *5.1 Descriptive statistics*

#### *5.1.1 Emotions*

First, I elaborate on the emotion time series of scores adjusted for TF-IDF and negations, but prior to intraday normalization. The emotion time series consist of 3,060 observations in the period from 15th of October 2007 until 29th of February 2016. The average scores per day for every emotion are displayed in table 3. For example, the average score for anger is .0123 (*SD*=.0023, *range*=.0007-.0303). The value indicates the amount of anger towards president Obama, which is expressed by Reddit users in comments. The other emotions are constructed in the same way and are therefore comparable. The other emotions can be summarized as follow; for anticipation the average amount is .0117 (*SD*=.0023, *range*=.0052-.0333), for disgust is .0084 (*SD*=.0017, *range*=.0003-.0189), for fear is .0126 (*SD*=.0024, *range*=.0046-.0346), for joy .0101 (*SD*=.0024, *range*=.0036-.0482), for sadness is .0115 (*SD*=.0021, *range*=.0019-.0451), for surprise is .0066 (*SD*=.0017, *range*=.0030-.0275), for trust is .0205 (*SD*=.0034, *range*=.0084-.0577). Concluding that, on average, the highest expressed emotion on Reddit regarding the topic ‘Obama’ is trust, followed by fear and anger. The least expressed emotions are disgust and surprise.

The density plot, Jarque-Bera and Shapiro-Wilk tests provide information about the normal distribution of the emotion time series. All the time series do not pass the Jarque-Bera and Shapiro-Wilk test of normality. The examined density plots in Figure 11 and the skewness and kurtosis values from table 1 show that the time series appear to be non-normal with heavy right-hand tails.

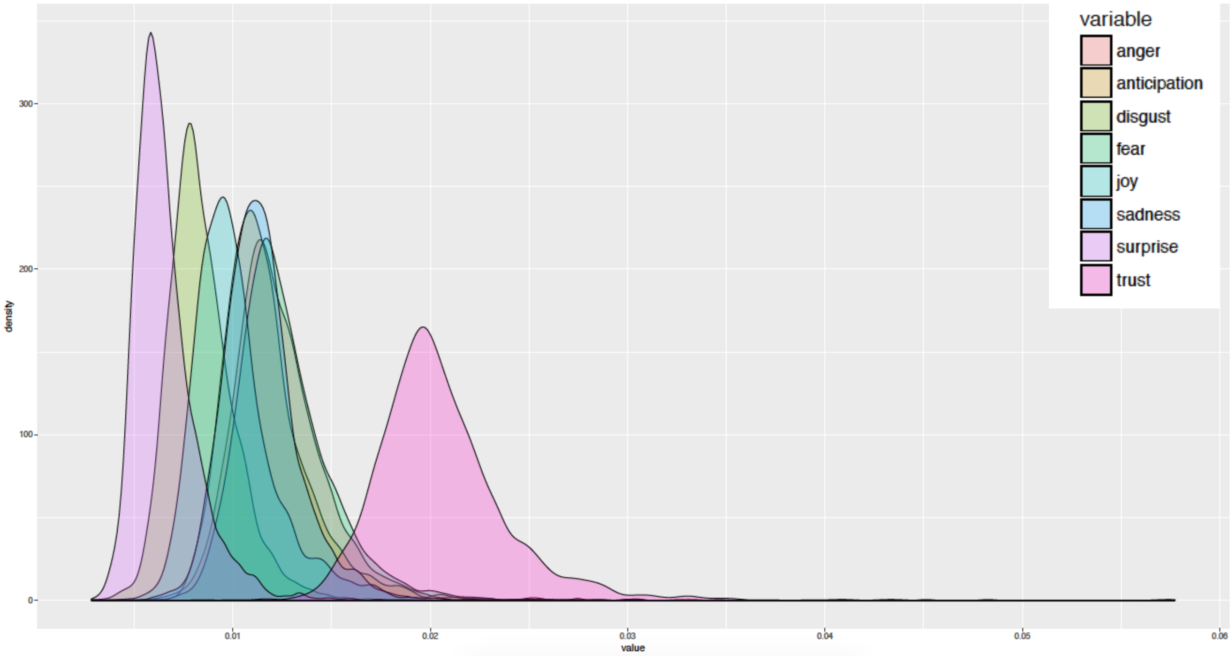



Figure 11. Density plots of emotion time series

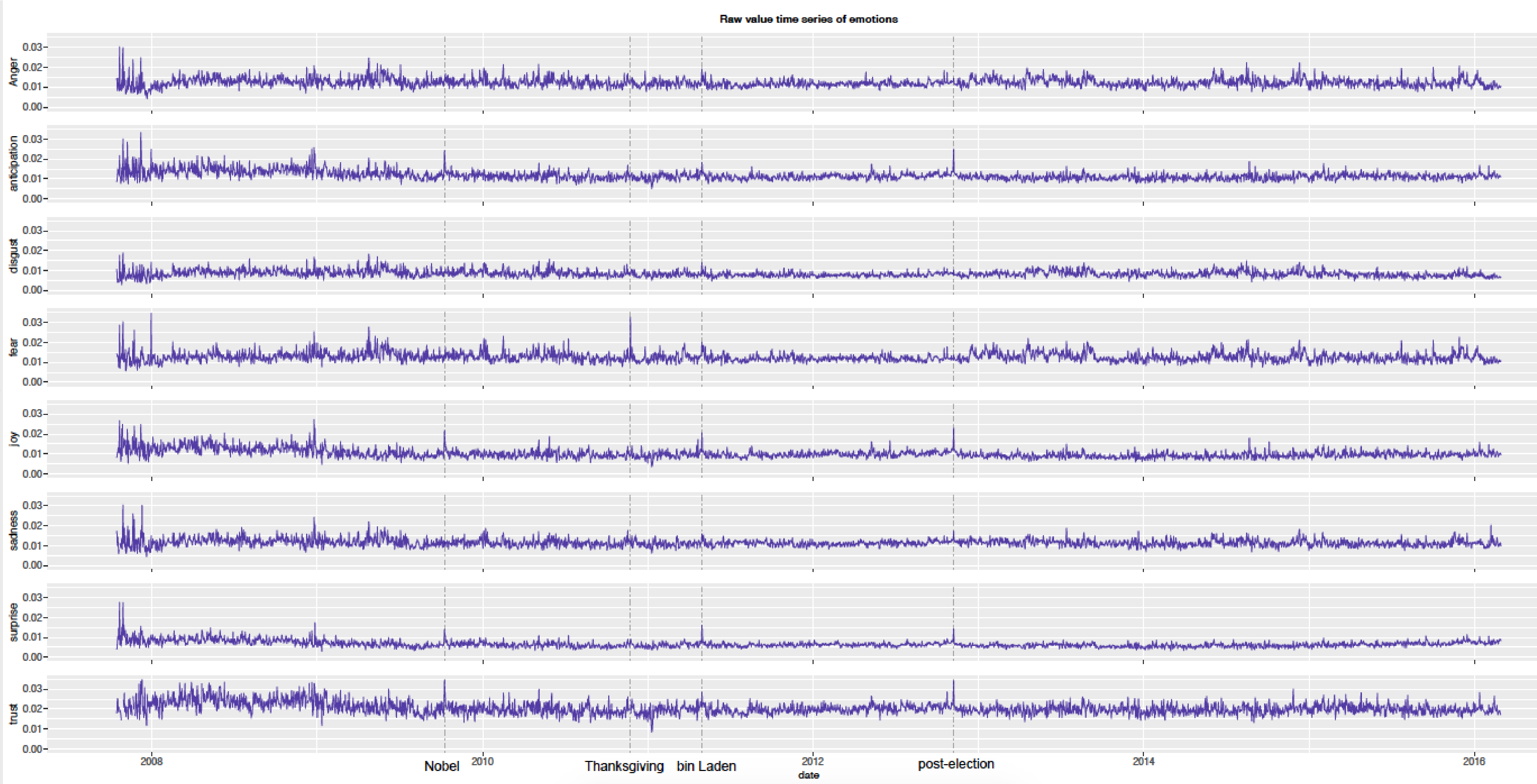



Figure 12. Emotion time series with marked events.

The time series plots of emotion are shown in Figure 12. Obama was officially elected as the president of the United States at 20th of January 2009. Therefore, the Reddit comments in the time series plot also cover a period before Obama was officially elected as president. Four specific events are highlighted in the time series in order to examine whether the emotions react to certain events. The first event relates to the 9th of October 2009 where Barack Obama wins the Nobel Peace Prize. High levels of trust, joy and anticipation are displayed and low levels of sadness, disgust, anger and fear. The second event is associated with Thanksgiving in the year 2011. In this year, Thanksgiving was celebrated at the 24th of November 2010. High levels of trust are shown, but remarkably also very high levels of fear. The high peaks of surprise and joy in May 2011 are probably due to the killing of Osama bin Laden. The high peaks of surprise and joy occur in alignment with peaks in anger and fear. The fourth event relates to the post-election period of Obama, which was a period remarked by high volumes of comments. Here, high peaks of anticipation, joy, trust and surprise are shown and low levels of anger, disgust and sadness.

To conclude, in the beginning of the time series high levels of all emotions are shown and certain events are aligned to an increase of multiple opposite emotions. This is due to the dependency of emotions on the volume of terms per day. Intraday normalization corrects for this shortcoming.

### *5.1.2 Intraday normalized emotions*

The intraday normalized emotions time series also consist of 3,060 observations in the period from 15th of October 2007 until 29th of February 2016. The average scores per day for every emotion are displayed in table 4. For example, the average score for anger is .1314 or 13.14% (*SD*=.0123, *range*=.0448-.2008). This value is adjusted for TF-IDF and negations and intraday normalized. The value indicates the intraday weight of anger on the total sum of 1 emotion per day towards president Obama expressed in Reddit comments. The other emotions are constructed in the same way and are therefore comparable. The other emotions can be summarized as follow; for anticipation the average is .1244 (*SD*=.0102, *range*=.0741-.2039), for disgust is .0899 (*SD*=.0103, *range*=.020-.1350), for fear is .1347 (*SD*=.0142, *range*=.0705-.2485), for joy .1079 (*SD*=.0118, *range*=.0566-.1874), for sadness is .1225 (*SD*=.0106, *range*=.0685-.1953), for surprise is .0697 (*SD*=.0087, *range*=.0295-.1722), for trust is .2195 (*SD*=.0158, *range*=.1461-.3087). Concluding that, on average, the highest expressed emotion on the sum of total emotion per day is trust, followed by fear and anger. The least expressed emotions are disgust and surprise. This is comparable to the values prior to intraday normalization.

Table 4. Descriptive statistics of intraday normalized values main dataset

| **Variables** | Start | End | N | M | SD | Min | Median | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 Rcp - approval (DV1) | 2009-01-27 | 2016-10-24 | 2,828 | 47.56 | 4.580 | 40.00 | 46.90 | 65.50 | 1.431 | 5.474 |
| 2 Gallup - approval (RC1) | 2009-01-24 | 2016-11-09 | 2,728 | 47.68 | 5.350 | 38.00 | 47.00 | 69.00 | 1.248 | 5.034 |
| 3 Rasmus – approval (RC1) | 2008-11-06 | 2016-11-18 | 2,743 | 48.70 | 4.170 | 41.00 | 48.00 | 69.00 | 1.948 | 8.242 |
| 4 Anger_n (IV1) | 2007-10-15 | 2016-02-29 | 3,060 | .1314 | .0123 | .0448 | .1312 | .2008 | -.1383 | 4.914 |
| 5 Anticipation_n (IV2) | 2007-10-15 | 2016-02-29 | 3,060 | .1244 | .0102 | .0741 | .1240 | .2039 | .5191 | 5.033 |
| 6 Disgust_n (IV3) | 2007-10-15 | 2016-02-29 | 3,060 | .0899 | .0103 | .0200 | .0896 | .1350 | -.0344 | 4.222 |
| 7 Fear_n (IV4) | 2007-10-15 | 2016-02-29 | 3,060 | .1347 | .0142 | .0705 | .1338 | .2485 | .4869 | 5.235 |
| 8 Joy_n (IV5) | 2007-10-15 | 2016-02-29 | 3,060 | .1079 | .0118 | .0566 | .1075 | .1874 | .4402 | 4.312 |
| 9 Sadness_n (IV6) | 2007-10-15 | 2016-02-29 | 3,060 | .1225 | .0106 | .0685 | .1224 | .1953 | .3426 | 5.928 |
| 10 Surprise _n (IV7) | 2007-10-15 | 2016-02-29 | 3,060 | .0697 | .0087 | .0295 | .0682 | .1722 | 1.293 | 9.103 |
| 11 Trust_n (IV8) | 2007-10-15 | 2016-02-29 | 3,060 | .2195 | .0158 | .1461 | .2200 | .3087 | -.0142 | 4.569 |

*Note.* DV = Dependent Variable. IV = Independent Variable. RC = Robustness check.
_n denotes the intraday normalized series.

*5.2 Correlation analysis*

When aligning the datasets of the RCP Obama job approval and the emotions extracted from Reddit, a dataset of 2,589 observations in the period from 28$^{th}$ of January 2009 until 29$^{th}$ of February 2016 remains for the joint analysis and regression modeling.

The correlation between emotions is explored prior to (Figure 13) and after intraday normalization (Figure 14) by means of correlograms. The approval rates of RCP are also displayed in the correlogram. A correlogram is an image of correlation statistics. The significance level in the correlogram is 5% and the crosses denote insignificance. Unexpectedly, all the correlations between values in Figure 13 are positive and significant. The correlogram of intraday normalized variables shows highly positive correlations between anticipation & joy and anger & fear. Highly negative correlation can be found between joy & fear and joy & anger. The negative and positive correlation outcomes of intraday normalized emotions confirm the expectations towards the interaction between emotions. This affirms the expectation that intraday normalized variables are better capable of capturing the interaction between emotions, since the values are not dependent on the volume of terms per day.

The correlations between the approval rate and emotions are also depicted in Figure 13 and 14. The correlations between approval rate and emotions in Figure 13 appear to be very low, but significant. The highest correlation of 25.25% is between approval and anticipation (table 5). All correlations between approval and the eight intraday normalized emotions are unexpectedly very low. The correlation between approval & anger, approval & joy and approval & surprise are also insignificant.

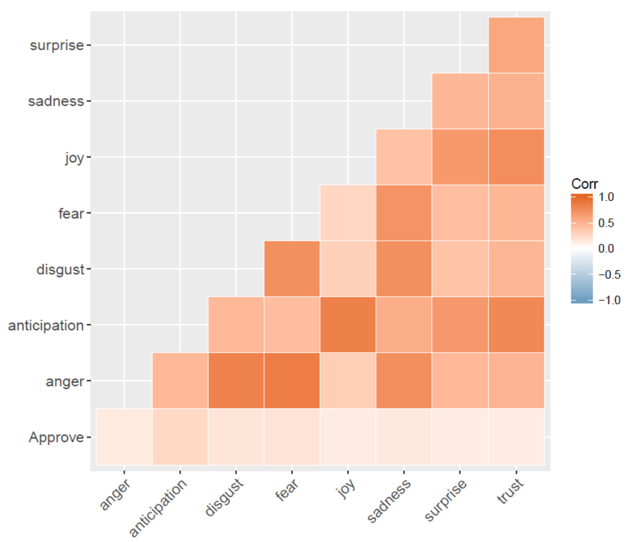



Figure 13. Correlogram of emotions

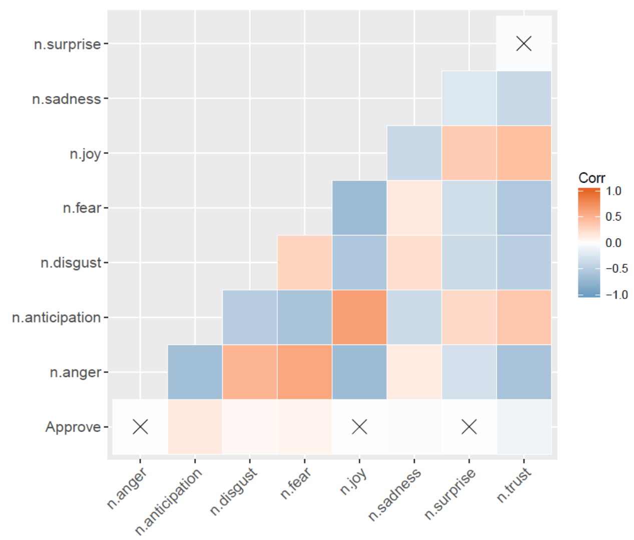



Figure 14. Correlogram of intraday normalized emotions

Table 5. Correlation matrix of main dataset

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 Rcp - approval (DV1) | 1 | .1368*** | .2525*** | .1679*** | .1818*** | .1342*** | .1458*** | .1299*** | .1275*** |
| 2 Anger (IV1) | | 1 | .4689*** | .8256*** | .8591*** | .3201*** | .7397*** | .4667*** | .4934*** |
| 3 Anticipation (IV2) | | | 1 | .4703*** | .4359*** | .8363*** | .5330*** | .6802*** | .7805*** |
| 4 Disgust (IV3) | | | | 1 | .7363*** | .3127*** | .7330*** | .3990*** | .4832*** |
| 5 Fear (IV4) | | | | | 1 | .2666*** | .7148*** | .4260*** | .4785*** |
| 6 Joy (IV5) | | | | | | 1 | .4041*** | .6772*** | .7494*** |
| 7 Sadness (IV6) | | | | | | | 1 | .4750*** | .5122*** |
| 8 Surprise (IV7) | | | | | | | | 1 | .5831*** |
| 9 Trust (IV8) | | | | | | | | | 1 |

*Note*. N = 2589 . *** $p$ < .01. ** p <.05. *p<0.10
DV = Dependent Variable. IV = Independent Variable.

*5.3 Linear regression model with AR(I)MA errors*

*5.3.1 Stationarity and co-integration*

An important consideration in estimating a linear regression model with AR(I)MA errors is that the response and predictors variables are stationary in order to estimate the correct coefficients. One exception can be made when the non-stationary predictor variable is co-integrated with the stationary response variable (Hyndman & Athanasopoulos, 2014). Before, the linear regression with AR(I)MA errors is estimated, a Dickey-Fuller and co-integration test are performed to indicate whether the series need to be differenced.

First, I perform a Dickey-Fuller test on the approval and emotion time series to test whether there is a unit root in the data. The test is performed on the raw series of approval and TF-IDF and negations adjusted values of emotion. The results are shown in table 6. The approval data shows signs of non-stationarity (DF=-3.2528, lag=1, p=.079). The time series of emotions appear to be stationary, since all p-values are smaller than 0.05.

Table 6. Augmented Dickey-Fuller test

| **Variable** | DF | p-value |
|---|---|---|
| 1 Rcp - approval (DV1) | -3.253 | .0788 |
| 2 Anger (IV1) | -35.53 | .0100 |
| 3 Anticipation (IV2) | -38.75 | .0100 |
| 4 Disgust (IV3) | -36.19 | .0100 |
| 5 Fear (IV4) | -33.65 | .0100 |
| 6 Joy (IV5) | -39.41 | .0100 |
| 7 Sadness (IV6) | -38.63 | .0100 |
| 8 Surprise (IV7) | -33.90 | .0100 |
| 9 Trust (IV8) | -40.38 | .0100 |

*Note:* when p-value is .0100, p-value might be smaller than printed value. Alternative = stationary

Next, I perform the Johansen co-integration test to examine whether the non-stationary approval rate is co-integrated with the stationary emotion variables. The rank r>0 implies a co-integrated relationship between two or possibly more time series. Table 7 shows the results of co-integration relationship between the time series adopted from the main equation (7). R ≤

9 provides sufficient evidence for rejecting the null hypothesis. Thus the best estimate of the rank of the matrix is r=10, which confirms the possibility to form linear combination (stationarity) of ten time series of equation (7).

Table 7. Values of test-statistics and critical values of Johansen test

| **Variable** | Test-statistic | 5 pct | 1 pct |
|---|---|---|---|
| r <= 9 | 11.68 | 8.18 | 11.65 |
| r <= 8 | 396.31 | 17.95 | 23.52 |
| r <= 7 | 859.64 | 31.52 | 37.22 |
| r <= 6 | 1486.80 | 48.28 | 55.43 |
| r <= 5 | 2128.93 | 70.60 | 78.87 |
| r <= 4 | 2848.11 | 90.39 | 104.20 |
| r <= 3 | 3588.71 | 124.25 | 136.06 |
| r <= 2 | 4393.50 | 157.11 | 168.92 |
| r <= 1 | 5267.62 | 192.84 | 204.79 |
| r = 0 | 6288.75 | 232.49 | 246.27 |

*Note:* when test-statistic is larger than critical value of test. The null hypothesis can be rejected.

To conclude, the linear regression model with AR(I)MA error terms is estimated with the raw values of the RCP approval rate and the TF-IDF and negation adjusted emotion variables. Despite the fact that intraday normalized variables tend to capture the dynamics of emotions better, implementing the variables in a regression model does not obtain optimal interpretation results for this study. This is elaborated in the discussion section.

### *5.3.2 Main regression results*

To validate the implementation of a linear model with AR(I)MA errors, a linear regression model was fitted to main equation (7). The results show that the errors terms are indeed correlated according to the Durbin-Watson test (AC= .9151, D-W=.1659, p=.000) and the Ljung-Box test, where the test is failed with a p-value of .0000, lag=12, fitdf=3. This provides validation for the implementation of a linear regression model with AR(I)MA errors, since the assumption of uncorrelated error terms needs to be replaced by the weaker assumption of stationary error terms.

The proxy model is estimated with the order (0,0,0). The order of this function *(p,d,q)* determines the values for AR (*p*), MA(*q*) terms and whether the series are differenced (*d*). For almost all models, *d* is set to 0 since the variables are not differenced. The time series and (P)ACF plots of the residuals of model (0,0,0) are shown in Figure 15. The plots provide information about the amount of MA and AR terms that need to be added to the model in order to correct for most of the autocorrelation. The ACF plot can be used to provide guidance for the MA-part, the PACF can be used to provide guidance for the AR-part. The plots are merely a chart of the coefficients of correlation between a time series and the lags of itself. The PACF plot is a plot of the partial correlation coefficients between the series and the lags of itself. When the spikes exceed the blue dotted significance bands, autocorrelation exists. The time series plots can be used to explain certain behaviour. For example, a strong decreasing trend in the approval rate of Obama in the first year can be explained by the increase of negative emotions but also by strong autocorrelation. When strong autocorrelation exists it might be impossible to separate the effect of individual emotions on the approval rate. Therefore, the autocorrelation need to be reduced by adding more AR and MA terms. The residuals in Figure 15 show very high autocorrelation at all lags in the ACF plots. The autocorrelation at lag 5 and above might be due to autocorrelation in previous lags. This is confirmed by the PACF plot where high autocorrelation is shown at lag 1, 2, 3 and 4. The PACF plot indicates that 4 AR terms need to be added to the model.

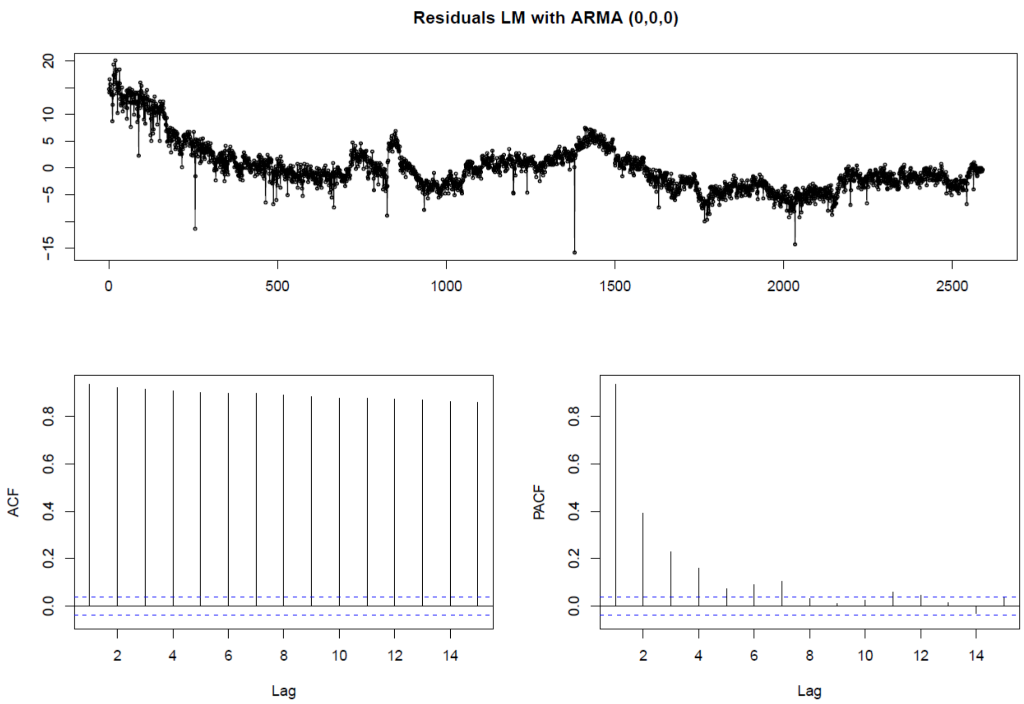



Figure 15. (P)ACF and time series plots of proxy model (0,0,0)

Table 8. Regression results linear regression models with ARMA errors

| **Model** | **1** (1,0,0) | **2** (0,0,1) | **3** (2,0,0) | **4** (0,0,2) | **5** (1,0,1) | **6** (2,0,1) | **7** (1,0,2) | **8** (2,0,2) | **9** (3,0,0) | **10** (3,0,1) | **11** (3,0,2) | **12** (3,0,3) | **13** (1,0,3) | **14** (2,0,3) | **15** (4,0,0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Anger | 7.428 | -25.37 | 7.717 | -17.00 | 7.633 | 7.472 | 7.264 | 9.333 | 7.028 | 9.094 | 9.387 | 11.24 | 10.42 | 10.71 | 11.12 |
| Disgust | 9.015 | 46.33** | 8.839 | 20.11 | 8.833 | 9.020 | 8.827 | 8.873 | 8.804 | 8.867 | 8.929 | 8.934 | 8.946 | 8.878 | 8.984 |
| Fear | 13.89 | 11.04 | 14.31 | -7.862 | 14.29 | 13.86 | 14.17 | 14.49 | 14.13 | 14.58 | 14.58 | 15.37 | 14.55 | 14.87 | 15.23 |
| Joy | -.3623 | 6.452 | .9841 | -27.93 | 1.086 | -.1792 | 1.127 | .9489 | 1.430 | .6318 | 2.311 | 2.620 | 2.782 | 2.745 | 2.595 |
| Sadness | 6.536 | -22.30 | 5.592 | 13.28 | 5.490 | 6.618 | 5.270 | 6.926 | 4.908 | 6.987 | 5.200 | 5.238 | 4.923 | 5.109 | 5.548 |
| Trust | -1.718 | -7.613 | -1.720 | -1.248 | -1.709 | -1.663 | -1.616 | -2.138 | -1.552 | -2.078 | -1.998 | -2.392 | -2.308 | -2.327 | -2.476 |
| Surprise* Sadness | -234.9 | 858.2 | -201.8 | -1,793 | -197 | -232.5 | -184.2 | -307.2 | -168.6 | -300.5 | -224.2 | -254.0 | -232.2 | -236.1 | -258.4 |
| Joy*Anticipation | -52.46 | 1,323 | -104.6 | -1,876** | -110.1 | -65.28 | -120.8 | -60.69 | -139.1 | -57.23 | -133.0 | -117.8 | -123.7 | -122.6 | -107.9 |
| Anger*Fear | -961.5 | 260.0 | -970.47 | 377.1 | -965.9 | -964.5 | -947.68 | -1,048 | -933.7 | -1,044 | -1,003 | -1,081 | -1,034 | -1,053 | -1,086 |
| AR(1) | .9986*** | | 1.054*** | | .9986*** | .0101 | .9987*** | .2373*** | 1.063*** | .2860*** | .6839*** | 1.206*** | .9992*** | 1.111*** | 1.052*** |
| AR(2) | | | | | | .9878*** | | .7601*** | -.0901** | .7866*** | -.2760 | -.3602 | | -.1111 | -.0769 *** |
| AR(3) | | | | | | | | | .0326 | -.0755 *** | .5901 *** | .1532 | | | -.1128 *** |
| AR(4) | | | | | | | | | | | | | | | .1371*** |
| MA(1) | | .9501*** | | 1.539 | .0579*** | .9893*** | .0528** | .8305*** | | .7682*** | .3997* | -.1541 | .0499*** | -.0595 | |
| MA(2) | | | | .8633 | | | -.0154 | .0930*** | | | .6291*** | .1180 | -.0233 | -.028 | |
| MA(3) | | | | | | | | | | | | -.1178 *** | -.1338 *** | -.1321 *** | |
| AIC | 1,989 | 12,089 | 1,984 | 9,514 | | 1,994 | 1,985 | 1,977 | 1,983 | 1,981 | 1,968 | 1,942 | 1,939 | 1,941 | 1,936 |
| BIC | 2,060 | 12,160 | 2,060 | 9,590 | | 2,076 | 2,067 | 2,065 | 2,065 | 2,068 | 2,062 | 2,042 | 2,027 | 2,034 | 2,024 |

*Note*. N = 2589 . *** $p$ < .01. ** p <.05. *p<0.10

Model 1 to 15 denote linear regression models with ARMA errors with AIC and BIC values for the estimated set of ARMA order (p,d,q) for errors.

The (P)ACF plot of the proxy model in Figure 15 recommends to add 4 AR terms to the model. Besides (P)ACF plots of the proxy model, the significance of error terms and AIC and BIC values of other models are examined to estimate the most parsimonious model. The measures can help to avoid overfitting of the model. Table 8 provides a summary of the coefficients and AIC and BIC values of the estimated models. Model 12, 13, 14 and 15 have the best (lowest) AIC values. Model 15 and model 13 have better BIC values than model 12 and 14, because BIC punishes for adding too many terms. All error terms are significant in model 15 and most error terms in model 13. Based on the significance of error term and AIC and BIC values, model 15 (4,0,0) is chosen as the most parsimonious model.

To validate model 15 as the most parsimonious model to analyze the hypotheses, (P)ACF plots of residuals are examined and the Ljung-Box test is performed. Model (4,0,0) has small correlation at short lags for residual ACF (Figure 16), which is an indication that the model fits well. The residuals show signs of white noise, because the model seems to capture the pattern in data quite well. Although there is some autocorrelation left, suggesting that the model can be slightly improved, it is unlikely that will make much difference in examining the hypotheses. To test jointly if the several autocorrelations of the residuals are zero, I perform a Ljung-Box test. The test is passed when the p-value exceeds .05 which means no autocorrelation exists. Model (4,0,0) passes the Ljung-Box tests for various choices of lags below lag 12. For example, the Ljung-Box test is passed with a p-value of .195 at lag=12 with fitdf=3. When the amount of lags is increases above lag 12, model (4,0,0) fails the test.

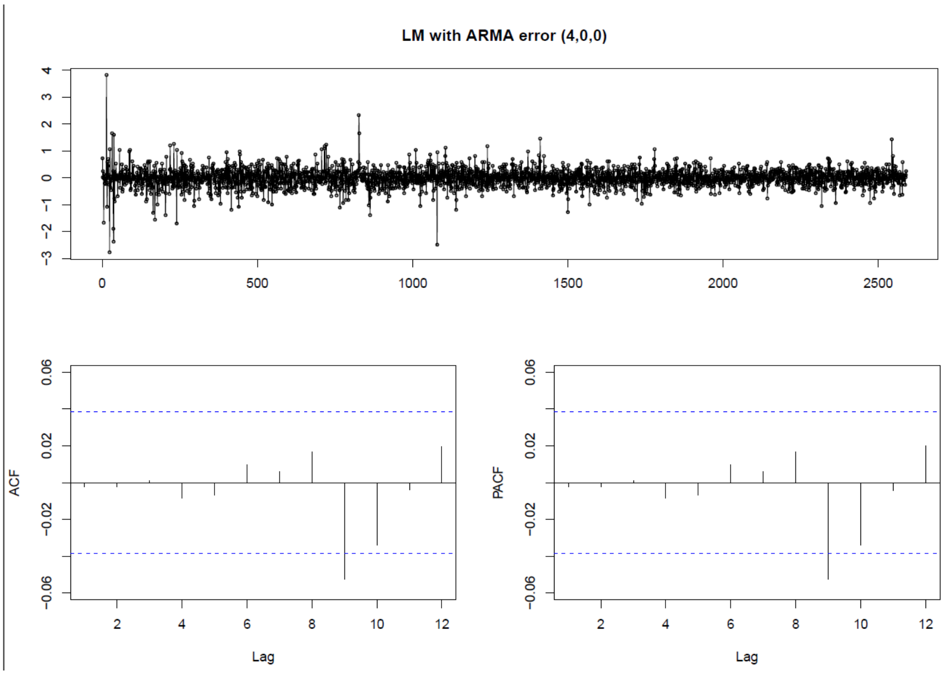



Figure 16. (P)ACF and time series plots of most parsimonious model (4,0,0)

To conclude, based on ACF plots, amount of parameters, AIC and BIC, significance error terms and Ljung-Box I consider model 15 (4,0,0) as the most parsimonious model. Thus, the fitted model from Equation (7) is:

$$Y_t \;\; = 45.68 + 11.12 anger_{1,t} + 8.984 disgust_{1,t} + 15.23 fear_{1,t} + 2.596 joy_{1,t} + 5.548 sadness_{1,t} - 2.476 trust_{1,t} - 258.4 surprise_{1,t} * sadness_{1,t} - \; 107.9 joy_{1,t} * anticipation_{1,t} - 1{,}086 anger * fear_{1,t} + n_t \qquad (7)$$

Where

$$n'_t = \; \emptyset_1 n'_{t-1} - \; \emptyset_2 n'_{t-2} - \; \emptyset_3 n'_{t-3} - \emptyset_4 n'_{t-4} + \; e_t$$
$$= \; 1.052 * n'_{t-1} - \; .0769 * n'_{t-2} - \; .1128 * n'_{t-3} + \; .1371 * n'_{t-4} + \; \epsilon_t$$

$H_{main}$ regards the ability of emotions to track the presidential approval rate of public opinion polls. The expectations regarding individual emotions are stated in $H_{anger}$, $H_{anticipation,}$ $H_{disgust,}$ $H_{fear}$, $H_{joy,}$ $H_{sadness,}$ $H_{surprise,}$ $H_{trust}$. The results can be found in table 8. Unfortunately, the coefficients of anger, anticipation, disgust, fear, joy, sadness, trust and surprise appear to be non-significant, neither at 5% nor at 10% significance levels. Therefore, the hypotheses regarding the relation of individual emotions with presidential approval rates cannot be confirmed.

Furthermore, interaction effects between emotions are hypothesized in $H_{surprise\text{-}sadness,}$ $H_{joy\text{-}anticipation}$ and $H_{anger\text{-}fear}$. The hypotheses are based on the assumption that, for example, experiencing two negative emotions at the same time increases the chance of a decreasing approval rate compared to experiencing only one negative emotion. All interaction effect coefficients are large and negative, but also appear to be non-significant, neither at 5% nor at 10% significance levels. The linear dependency of the interaction between emotions on approval cannot be examined. Still, it seems reasonable to assume the interactions exist based on correlation measures and theory. The non-significance might occur due to multicollinearity. Multicollinearity occurs when variables are highly correlation. In means of this study, if the emotion predictor variables interact a lot it is difficult to estimate the individual effects on the approval rate. This however, does not affect the fit of the model as a whole.

For error terms it is not always clear how to interpret the structure and coefficients. The equation above shows that every day includes information from the previous day, the day before, the day before that and the day before that..

Table 9. Regression results of robustness checks compared to main model 15

| **Model** | **15** (4,0,0) | Gallup | Rasmus | Winsor |
|---|---|---|---|---|
| Anger | 11.12 | 22.89 | -59.33* | -9.638 |
| Disgust | 8.984 | -20.49 | 50.71*** | 9.778 |
| Fear | 15.23 | 11.67 | -9.518 | -4.620 |
| Joy | 2.595 | 14.62 | -24.04 | 11.87 |
| Sadness | 5.548 | -30.31 | -22.55 | 9.664 |
| Trust | -2.476 | .2081 | 4.618 | -2.390 |
| Surprise* Sadness | -258.4 | 1,480 | 165.9 | -779.9 |
| Joy*Anticipation | -107.9 | 42.91 | 594.8 | -551.6 |
| Anger* fear | -1,086 | -785.8 | 2,434 | 584.7 |
| AR(1) | 1.052*** | .9529 | .9803*** | 1.052*** |
| AR(2) | -.0769 *** | .0246 | -.0191 | -.0786 *** |
| AR(3) | -.1128 *** | -.4636 *** | -.4174 *** | -.1114 *** |
| AR(4) | .1371 | .4733 *** | .4326 *** | .1367 *** |
| AIC | 1,936 | 7,923 | 7,449 | 1,935 |
| BIC | 2,024 | 8,010 | 7,533 | 2,023 |

*Note*. *** $p < .01$. ** $p < .05$. *$p<0.10$
15 (4,0,0): N= 2589, Gallup: N = 2482, Rasmu; N= 2474, Winsor; N= 2589
Linear regression models with ARMA errors with AIC and BIC values for the estimated set of ARMA order (p,d,q) for errors.

### *5.4 Robustness checks*

The structural validity of the main results are examined by several robustness checks. The dependent variable RCP approval is replaced by approval rates of other polling sources, since different polling systems use different samples of the population (e.g. adults versus likely voters). The certainty of the main findings can thus be investigated since the regression specification is modified to a similar response variable. To explore the influence of outliers, a model with winsorized values of the emotion variables is considered.

### *5.4.1 Gallup*

The main differences between Gallup and RCP occur in the AIC and BIC values. Remarkably, the AIC and BIC values for Gallup are a lot higher indicating a worse fitting model than the main model for RCP. Possible explanations could be that the approval values of Gallup round up to whole numbers creating a less precise time series than the RCP approval series. Also, the Gallup approval values are averaged over a 3-day window. Another explanation could be that the (4,0,0) model is not the best fitted model for Gallup and a large amount of autocorrelation remains (appendix 3). This is confirmed by the Ljung-Box test, which is failed with a p-value of .0000, lag=12, fitdf=3. The RCP could be better in capturing the overall public opinion since it is an average of a tremendous amount of polling systems. Besides, the differences are likely to occur since RCP and Gallup are not perfectly correlated and both polling systems use different techniques.

### *5.4.2 Rasmussen*

The AIC and BIC values for Rasmussen reports are also a lot higher than the main model for RCP. Remarkably, the values of disgust and anger show some significance. The results should be taken with care due to multicollinearity and therefore I cannot isolate the influence of disgust and anger on the predictor variable Rasmussen. Besides, the model fails the Ljung-Box test with a p-value of .0000, lag 12, fitdf=3 and the (P)ACF plots still show large amounts of autocorrelation in the residuals (appendix 3).

The predictability of the model depends on the poll to be predicted and therefore the outcomes of the robustness test indicate that caution should be taken when generalizing the results from the most parsimonious model applied to RCP approval data.

### *5.4.3 Winsorized values*

The most parsimonious regression model with AR(4) errors is also examined using winsorized values of the main dataset. Winsorizing is a transformation of the values to correct for outliers or extreme values and essentially make the data look more normal. Since the distribution can be heavily influenced by outliers, I am interested in constructing a model using the winsorized values to examine the difference with the main model.

Firstly, I examine the density plots to see whether the winsorized values encounter large differences in normal distributions with the emotions of the main dataset in Figure 11. The winsorized values are created using a fraction of alpha .05. Figure 17 presents the density plots and indeed some winsorized values for emotions appear to be more concentrated in the middle, but heavy right-hand tails are also observed.

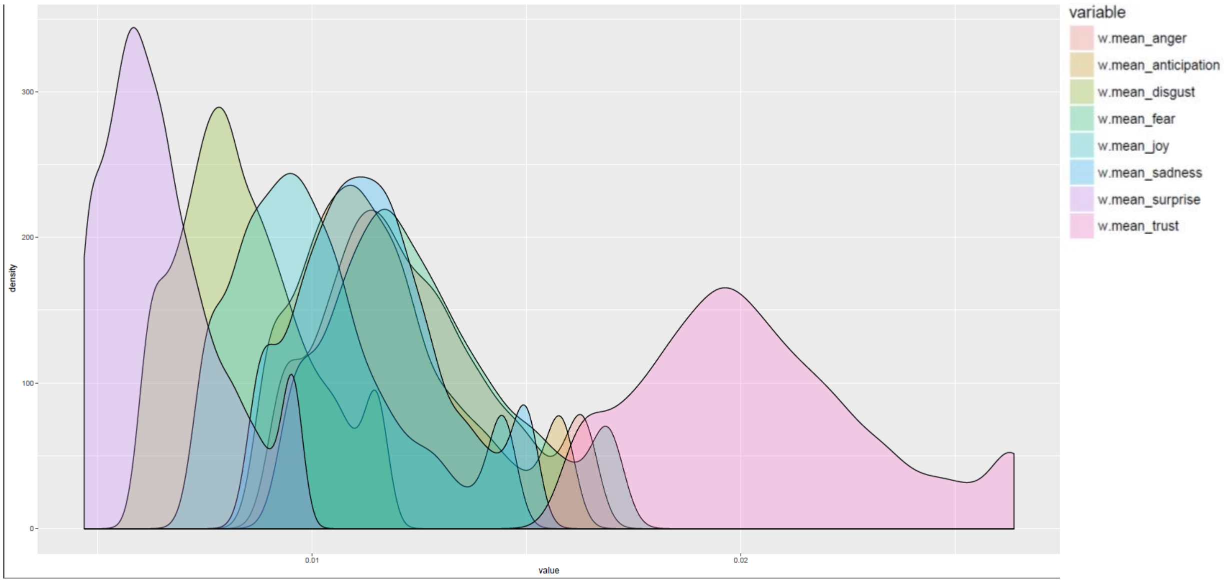



Figure 17. Density plot of winsorized emotions

Table 7 shows the results of the regression model with winsorized values. There are no large differences captured between the winsorized and main model. A small improvement in AIC and BIC values is observed, which could be an indication that the winsorized values construct a better fitted model. The model passes the Ljung-Box test for uncorrelated residuals with a p-value of .16486, lag=12, fitdf=3.

## 6 Semantic features in target identification

### *6.1 Model*

As an attempt to improve the fit of estimated coefficients in the main model, I increase the weighting of words surrounding the word Obama. The aim of this approach is to identify the target and increase the accuracy of emotion detection towards this target.

In order to find the most parsimonious model, this study uses different techniques, such as TF-IDF, negation weighting and intraday normalization. Another technique which might be appropriate to define the most optimal model is the identification of a target. Identification of a target is related to stance detection. In emotion analysis, a piece of text is examined on expressing a certain emotion. In stance detection, the opinion towards a given target is examined (Taboada et al.,2011). This research examines the emotions of the person towards the target ‘Obama’. For example, consider the following sentence from a Reddit comment:

> Sentence: *Barack Obama* did an awesome job in contrast to Donald Trump who already fails miserably.

The word ‘job’ would have been extracted by the lexicon as a positive annotated score and the words ‘fails’ and ‘miserably’ as negative scores. The load of negative sentiment would exceed the load of positive sentiment regarding Obama, which is intuitively wrong when reading the sentence. Increasing the annotated scores of words close to Obama, in this case ‘job’, increases the proportion of positive sentiment compared to negative sentiment and therefore increases the accuracy of sentiment towards the target Obama.

Earlier studies introduce approaches to automatically extract sentiment from a target (Chen et al., 2012). Chen et al. (2012) asses target-dependent polarity of each sentiment expression. The polarity is determined by the nature of its target. Chen et al. (2012) confirm that emphasizing the target entity can indeed improve the public opinion tool. Sobhani et al. (2016) develop a simple stance detection system with sentiment features. Sobhani et al. (2016) show that the features can be very helpful, but alone are not sufficient and challenges are still present. Taboada et al. (2011) introduce semantic orientation which refers to the polarity and strength of words, phrases, or texts towards a target. Earlier research shows that identifying a target entity is relevant and can indeed improve the public opinion tool. Therefore, I hypothesize:

> $H_{semantic}$: Adding semantic features improves the ability of emotion to track the presidential approval rate in public opinion polls.

This study does not aim to explore the task of stance detection in Reddit comments, since Sobhani et al. (2016) emphasize the complexity of the task. This study applies the same reasoning from stance detection, but uses rather simple techniques. Therefore, I am interested in the effect of implementing up-weighting factors to words surrounding Obama as semantic features. The semantic features are added to the annotated scores per term after the adjustments for TF-IDF and corrections for negations and before the daily aggregated vector is created. The up-weighting factors applied are:

$$For\ all\ L \neq 0 \quad e_t(L) = \begin{cases} 2.0e_t & L = -1 \mid L = 1 \\ 1.8e_t & L = -2 \mid L = 2 \\ 1.6e_t & L = -3 \mid L = 3 \\ 1.4e_t & L = -4 \mid L = 4 \\ 1.2e_t & L = -5 \mid L = 5 \\ e_t & L \leq -6 \mid L \geq 6 \end{cases} \tag{9}$$

$e_t$ denotes the annotated score per term adjusted for TF-IDF and negations. The up-weighting factors can be interpreted as follow: for example when a term lags one distance ($L$=-1) or leads one distance ($L$=1) from the word 'Obama', e.g. Obama precedes or succeeds the word, $e_t$ receives the up-weighting factor of 2.0.

Table 10. Regression results of semantic features in target identification compared to main hypothesis

| **Model** | **Main** (4,0,0) | **Semantic** (4,0,0) |
|---|---|---|
| Anger | 11.12 | 5.330 |
| Disgust | 8.984 | 6.202 |
| Fear | 15.23 | 9.045 |
| Joy | 2.595 | 1.937 |
| Sadness | 5.548 | 5.836 |
| Trust | -2.476 | -1.811 |
| Surprise* Sadness | -258.4 | -103.2 |
| Joy*Anticipation | -107.9 | -133.0 |
| Anger* fear | -1,086 | -554.4 |
| AR(1) | 1.052*** | 1.052*** |
| AR(2) | -.0769 *** | -.0772 *** |
| AR(3) | -.1128 *** | -.1114 *** |
| AR(4) | .1371*** | .1363 *** |
| AIC | 1,936 | 1,937 |
| BIC | 2,024 | 2,025 |

*Note*. N = 2589 . *** $p$ < .01. ** p <.05. *p<0.10
Linear regression models with ARMA errors with AIC and BIC values for the estimated set of ARMA order (p,d,q) for errors.

### *6.2 Results*

The results of the most parsimonious model (4,0,0) on the renewed dataset with semantic features are shown in table 10. None of the coefficients are statistically significant at all significance levels and cannot be interpreted. The measurements of model fit are comparable to the main model. Furthermore, the (P)ACF plots in Figure 18 show similar behaviour to the (P)ACF plots in Figure 16. The plots have small correlation at short lags for residual ACF, which is an indication that the model fits well. The residuals shows signs of white noise, because the model seems to capture the pattern in data quite well. The Ljung-Box test is also passed for various choices of lags below lag 12. For example, the Ljung-Box test is passed with a p-value of .256 at lag=12 with fitdf=3. When the amount of lags increases above lag 12, model (4,0,0) fails the test. Since the results are comparable at all levels to the main results, the hypothesis $H_{semantic}$ cannot be confirmed. To conclude, I cannot substantiate the claim that adding semantic features improves the ability of emotion to track the presidential approval rate in public opinion polls.

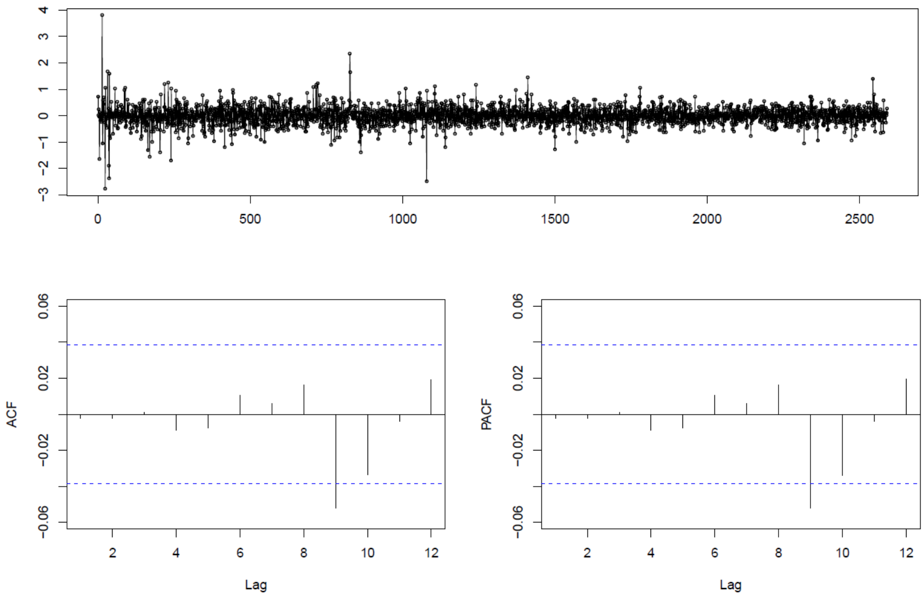



Figure 18. (P)ACF plots and time series of model (4,0,0) semantic features in target identification

# 7 Conclusion

## *7.1 Summary of study results*

The main objective of this paper is to examine whether public emotions as measured from comments posted on Reddit relate to the presidential approval rates of public opinion polls. This study is motivated by the three flaws of surveys and public opinion polls, i.e. inaccurate data, requirement of a large amount of respondents and time consuming to conduct. Hence, public opinion polls often fail to capture the high temporal resolution of public opinion. With the use of a substantial amount of data from Reddit, it is possible to capture the opinion of a large proportion of the population. To examine this inquiry the following research question was formulated:

> **RQ:** To what extent can *public emotion* measured from a large-scale collection of *Reddit comments* be used to track *public opinion polls*?

The results partly show that changes in the public's emotional state can indeed be tracked from the content of Reddit comments by means of text processing techniques. The results show examples where the emotional response to certain events, such as the killing of Osama bin Laden and the pre- and post-elections, is captured. Therefore, it partly confirms the study of Bollen et al. (2011a), Bollen et al. (2011b) and O'Connor et al. (2010). When addressing the ability of independent emotional states to track the approval rate, the data show non-significant results for the six sub-hypotheses. Also unexpectedly, non-significant results are found for the interaction between emotions.

Second expectations address the impact of semantic features on target identification. In the main model, techniques such as TF-IDF and down-weighting factors for negations are used to determine which words are useful indicators. The second model tests whether adding up-weighting factors of words surrounding Obama improves the identification of the target. Chen et al. (2012) and Sobhani et al. (2016) believe that identification of the target positively influences the accuracy of capturing emotions. Unfortunately, non-significant results are found and measurements of model fit are comparable to the main model. Therefore, overall results show that emotions extracted from Reddit do not necessarily provide high predictive power of the changes in the approval rate of Obama. This partly confirms the opinion of Jürgens et al. (2011) and Gayo-Avello (2012) who believe that tracking tools are not predictive at all for public opinion, since most of the analysis is done post hoc.

While this study is unable to find significant evidence to accept the hypotheses, there are aspects in the methodology that can be used in further studies. The study shows that adding AR(I)MA errors can overcome the problem of correlated errors in linear regression modeling and provide improved values of model fit. Also, the study shows that Reddit can indeed capture some shifts in emotions with simple lexicon-based techniques, although the individual effect of an emotion could not be examined. This could be due to the finding of this research that a lot of emotions are highly correlated. The study does not discover opposite relations between emotions, since the volume of comments drives the scores. The study confirms that the relation between emotions is complex. Further research should consider whether these shifts can be separated and the individual effect of an emotion can be explored. Proportional independent emotion variables might be a proper solution. The study also addresses the importance of the difference between public opinion polls and the consequence for finding optimal models. A lot of polling companies use different sampling techniques and therefore a model could be predictive for one poll and confirming it captures the public opinion well, but could have contradicting results for another poll.

To conclude, the answer to the formulated research question is twofold. On the one hand, the current research emphasizes that the approach was not sufficient to confirm the ability of public emotion to track presidential approval rates as no significant results are found. On the other hand, adjusting the utilized method and making the technique more suitable, emotional shifts aligned to remarkable events are captured. Therefore, I believe that expanding the sentiment-tracking tool with emotional analysis can be beneficial over relative expensive and time consuming surveys and public opinion polls. It would be interesting to automatically extract emotions from textual comments; since tracking systems for social media are fast, cheap and can be applied on a large-scale. Emotional tracking tools provide a more multidimensional view on public opinion than the two-dimensional division of polls. Economics can gain a lot from this transfer of knowledge, as certain behavioural irregularities, which could not be explained by standard economic model, can be understood when emotions are taken into account. For example, predicting market irregularities within behavioural finance. For further research, it is debatable whether scholars should pursue a model that is predictive of public opinion polls or pursue a model that is predictive of other public opinion sources. Polls do have their flaws, which could have been the reason why this study encountered non-significant results. Despite, there is still a lot of work to do in the field of NLP and the automatically extraction of public opinion. An elaboration on recommendations for further research will be provided after the discussion.

*7.2 Discussion*

Since the study encountered results that did not support the hypotheses, a discussion of the results might be appropriate. Evaluating the methodology and findings of this paper enables the recommendation for further research. For example: why do certain emotions not capture public opinion as well as expected? Is this mainly due to flaws in the methodology or are other explanations more suitable?

A discussion regarding methodological flaws in the model and analysis as an explanation for the unsupported hypotheses is possible. The extremely complicated structuring of text may not be simply grasped by a lexicon-based approach. Even when implementing different weighting schemes, the study did not find results to confirm the hypotheses. As a matter of fact, when significant results would have been found, the problem of multicollinearity could have been encountered. I am confident that a lot of emotions are highly correlated, therefore it is difficult to examine their separate effects on the approval rate. Extensive analyses would be necessary to make sure whether hard conclusions can be drawn about the influence of an individual emotion on presidential approval rates. While the EmoLex database is one of the most comprehensive databases available, it might not be able to capture the broad and complex picture. It might capture emotions with a lexicon-based approach, but there should probably be an extension of weighting schemes and large annotated datasets of lexicon-words.

The emotions were transformed using TF-IDF, down- and up-weighting techniques. Despite, it should be acknowledged that the dataset might be influenced by small events, which drive a high volume of comments and therefore a peak in emotions' score. Comments reflect a broad spectrum of conversations and reactions to major events, including difficulties such as bots and sarcasm. The stochastic nature of Reddit time series makes it difficult to compare low-resolution polls to high-resolution Reddit comments. Although the study performed many weighting techniques and controls in order to make sure it did not simply use an aggregated score of emotion lexicon-words, I am unable to conclude which method is a more accurate indicator of public opinion. Advanced proportional techniques should overcome this deficiency in the future.

The non-significant results can also be explained by a phenomenon known as the vocal minority versus the silent majority. Discovering the opinions of the long tail is discussed by Gayo-Avello (2011). Gayo-Avello (2011) discusses the behaviour of two different groups on social media. On the one hand there is a minority of users producing most of the content (vocal minority) and on the other hand there is a majority of users that hardly produces any content (silent majority). With the use of TF-IDF, I control for this phenomenon when users comment on Reddit. However, I cannot possibly control for the opinion of the silent majority that does not comment on Reddit. Therefore, Gayo-Avello (2011) suggests that extreme caution should be taken when building predictive models based on social media. Even when performing several robustness checks, there is no knowledge of the "ground truth" neither for emotion states nor in fact for the particular subsample of Reddit users.

A discussion regarding the methodology of the public opinion polls and its accuracy in capturing the public opinion is also appropriate. Polls often fail to accurately predict the outcome of presidential elections and other general approvals. For instance, the case explained in the introduction where almost all polls wrongly predicated a win for Hillary Clinton in the last presidential election of the United States of America (RCP, 2016c). Therefore, it could be possible that the model can indeed capture public emotion but is not predictive of public opinion polls. This does not necessarily mean that the model provides a negative fit for capturing public emotion. Since polls often fail to accurately predict the outcome of presidential elections and other general approval, further research can shift the focus from supplementing polls to substituting polls.

### *7.3 Further research*

The analysis of this study does not acknowledge a number of important factors that could form the basis of further studies. Earlier, the need to extent weighting schemes and large annotated datasets of lexicon-words and the shift of focus from supplementing to substituting polls was already discussed. Other options regarding further research are also considered.

Emotion analysis is an interesting approach to gain multidimensional insights in the public opinion. The goal of emotion analysis is to examine a piece of text on expressing a certain emotion. In 2016, Mohammad et al. (2016) presented for the first time a shared task on detecting stance from tweets where NLP systems must determine whether the tweeter is in favor of the given target. To clarify stance detection is related to, but different from sentiment and emotion analysis. In stance detection, the opinion towards a given target is examined. This study uses a couple of simple semantic features to increase the importance of a target entity. Scholars strongly believe that identifying the target entity is important, especially in stance detection (Chen et al., 2012; Sobhani et al., 2016; Taboada et al., 2011). Stance detection is still its infancy, but it could be a very interesting development for further research to apply these techniques and examine whether stance detection models add relevance in capturing the public opinion on presidential approval.

The number of lexicon-terms per comment drives the results in this research. Sensitivity to volume results in positive correlations at all levels for the TF-IDF scores. A possible solution for further research is the implementation of intraday normalization. This research shows confirmative correlations when implementing the intraday normalized emotions. However, the study lacks advanced econometrical techniques to interpret the variables in a model. Further research might benefit from developing proper techniques to interpret a set of independent variables that are proportions that sum to one in linear regression models.

The Reddit corpus is still an ambiguous source for research due to its recent release. However, it is one of the most comprehensive databases of human interactions out there. The database grows rapidly and the subreddits provide a rare classification of topics. Further research on the Reddit database might provide essential insights in political and behavioural decision making models.

## References


Anderson, K. E. (2015). Ask me anything: what is Reddit?, *Library Hi Tech News, 32(5),* 8-11.

Baltussen, G. (2009). Behavioral Finance: an introduction, working paper SSRN. Accessed November,2,2016.Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1488110.

Barberis, N. and Thaler, R.H. (2003), A survey of behavioral finance, *In Handbook of the Economics of Finance*, G. Constantinides, M. Harris and R. Stulz (Eds.), North- Holland, Amsterdam.

Bentham, J. (1789). A utilitarian view. *Animal rights and human obligations*, 25-26.

Bollen, J., Mao, H., & Pepe, A. (2011a). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *ICWSM*, *11*, 450-453.

Bollen, J., Mao, H., & Zeng, X. (2011b). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

Brader, T., & Marcus, G. E. (2013). Emotion and political psychology.

Business Insider (2012). *Business Insider Obama Approval.* Accessed January 4, 2017. Available at: http://www.businessinsider.com/obama-approval-rating-polls-rasmussen-50-percent-gallup-2012-11?international=true&r=US&IR=T

Chen, L., Wang, W., Nagarajan, M., Wang, S., & Sheth, A. P. (2012). Extracting diverse sentiment expressions with target-dependent polarity from twitter.

Chowdhury, G. G. (2003). Natural language processing. *Annual review of information science and technology*, *37*(1), 51-89.

Clarke, S., Hoggett, P., & Thompson, S. (Eds.). (2006). *Emotion, politics and society*. Springer.

Cody, E. M., Reagan, A. J., Dodds, P. S., & Danforth, C. M. (2016). Public Opinion Polling with Twitter. *arXiv preprint arXiv:1608.02024*.

Duggan, M., & Smith, A. (2013). 6% of Online Adults are reddit Users. Pew Internet & American Life Project, 3. Accessed November, 1, 2016. Available at: http://www.pewinternet.org/2013/07/03/6-of-onlineadults- are-reddit-users/ (accessed 10 March 2015).

Ekman, P., Liebert, R. M., Friesen, W. V., Harrison, R., Zlatchin, C., Malmstrom, E. J., & Baron, R. A. (1972). Facial expressions of emotion while watching televised violence as predictors of subsequent aggression. *Television and social behavior, 5,* 22-58.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review,* *3*(4), 364-370

Elster, J. (1998). Emotions and economic theory. *Journal of economic literature, 36(1),* 47- 74.

Forgas, J. P. (1995). Mood and judgment: the affect infusion model (AIM). *Psychological bulletin*, *117*(1), 39.

Fox News (2011). *Fox news 2011.* Accessed January 4, 2017. Available at: http://www.foxnews.com/politics/2011/08/14/obamas-approval-rating-drops-to- lowest-ever-according-to-gallup.html

Frijda, N. H. (2008). The psychologist's point of view. In *Handbook of emotions* (3rd ed., pp. 68–87). New York: Guilford Press.

Gallup, G. H. (1948). A guide to public opinion polls (p. 73). Princeton, NJ: Princeton University Press.

Gallup Jr, G. (2001). The Gallup Poll: Public Opinion, 2000. Rowman & Littlefield.

Gallup (2013). *Gallup poll 2013.* Accessed January 4, 2017. Available at: http://www.gallup.com/poll/166727/obama-job-approval-declined-steadily-throughout-2013.aspx .

Gallup (2016). *Gallup facts.* Accessed November 5, 2016. Available at: http://media.gallup.com/muslimwestfacts/pdf/pollingandhowtouseitr1dreveng.pdf

Gayo-Avello, D., Metaxas, P. T., & Mustafaraj, E. (2011). Limits of electoral predictions using twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence.

Gayo-Avello, D. (2012). " I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper"--A Balanced Survey on Election Prediction using Twitter Data. *arXiv preprint arXiv:1204.6441*.

Github (2016). *Github Reddit.* Accessed November 18, 2016. Available at: https://github.com/reddit/reddit/wiki/JSON

González-Bailón, S., Banchs, R. E., & Kaltenbrunner, A. (2012). Emotions, Public Opinion, and US Presidential Approval Rates: A 5-Year Analysis of Online Political Discussions. *Human Communication Research*, *38*(2), 121-143.

Groenendyk, E. (2011). Current emotion research in political science: How emotions help democracy overcome its collective action problem. *Emotion Review*, *3*(4), 455-463.

Hansen, L. P. and Shiller, R. J. (2013). *Understanding asset pricing.* Accessed November, 5, 2016. Available at: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/advanced-economicsciences2013.pdf.

Huberman, G., and T. Regev. (2001). Contagious speculation and a cure for cancer: a non-event that made stock prices soar, *Journal of Finance 56*, 387-396.

Hyndman, R.J. (2014). The Armax Muddle. Accessed January 30, 2017. Available at: http://robjhyndman.com/hyndsight/arimax/

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts.

Jürgens, P., Jungherr, A., & Schoen, H. (2011). Small worlds with a difference: New gatekeepers and the filtering of political information on Twitter. In *Proceedings of the 3rd International Web Science Conference* (p. 21). ACM.

Katz, D., & Cantril, H. (1937). Public opinion polls. *Sociometry, 1(1/2)*, 155-179.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & Emotion*, *14*(4), 473-493.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review, 90(2)*, 426-432.

Lopes, L. L., & Oden, G. C. (1998, December). The role of aspiration level in risky choice. In *Journal of Mathematical Psychology* (Vol. 42, No. 4, pp. 478-478). 525 B ST, STE 1900, SAN DIEGO, CA 92101-4495 USA: ACADEMIC PRESS INC.

Martin, J. H., & Jurafsky, D. (2000). Speech and language processing. *International Edition*, *710*.

Marcus, G. E. (2000). Emotions in politics. *Annual Review of Political Science*, *3*(1), 221-250.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science, 8*(6), 423-429.

Meltwater (2016). *Meltwater sentiment analysis.* Accessed January 10, 2017. Available at: https://www.meltwater.com/za/blog/social-media-sentiment-analysis/)

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, *29(3*), 436-465.

Mohammad, S. M., Kiritchenko, S., Sobhani, P., Zhu, X., & Cherry, C. (2016). Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the International Workshop on Semantic Evaluation*, *SemEval* (Vol. 16).

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture* (pp. 70-77). ACM.

NRC (2016). *NRC Word-Emotion Association Lexicon.* Accessed July 15, 2016. Available at: http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. ICWSM, 11(122-129), 1-2.

Parrott, W. G. (2001). *Emotions in social psychology: Essential readings.* Key Readings in Social Psychology. Psychology Press.

Pixley, J. (2002). Emotions and economics. *The Sociological Review*, *50*(S2), 69-89.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. Theories of emotion, 1, 3-31.

Politico (2014). *Politico approval 2014.* Accessed January 4, 2017. Available at: http://www.politico.com/story/2014/10/poll-obama-approval-rating-111902

Posner, Jonathan; Russell, J.A.; Peterson, B. S. (2005). *"The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology"*. Developmental and Psychopathology. **17**: 715–734. *doi*:*10.1017/s0954579405050340*. *PMC 2367156*. *PMID 16262989*

Rajaraman, A., Ullman, J. D. (2011). "Data Mining". *Mining of Massive Datasets*, pp. 1–17. *doi*:*10.1017/CBO9781139058452.002*. *ISBN 9781139058452*.

Rao, T. & Srivastava, S. (2012). Analyzing stock market movements using Twitter sentiment analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)* (pp. 119– 123). IEEE Computer Society.

Rasmus (2016a). *Rasmus.* Accessed November 10, 2016. Available at: http://www.rasmussenreports.com/public_content/rasmussen_reports_proprietary_econometric_data

Rasmus (2016b). *Rasmus data.* Accessed November 10 , 2016. Available at: http://www.rasmussenreports.com/public_content/politics/obama_administration/obama_approval_index_history

Real Clear Politics (2016a). *RCP.* Accessed November 1, 2016. Available at: http://www.realclearpolitics.com/epolls/other/president_obama_job_approval-1044.html

Real Clear Politics (2016b). *RCP about.* Accessed November 18, 2016. Available at: http://www.realclearpolitics.com/about.html

Real Clear Politics (2016c). *RCP elections.* Accessed January 15, 2017. Available at: http://www.realclearpolitics.com/epolls/2016/president/us/general_election_trump_vs_clinton-5491.html

Reddit (2016). *Reddit Statistics.* Accessed November 2, 2016. Available at: https://about.reddit.com/advertise/

Renshon, J., & Lerner, J. S. (2012). The role of emotions in foreign policy decision making. *Encyclopedia of Peace Psychology*, 313-317.

Ruppert, D. (2015). *Statistics and data analysis for financial engineering* (pp. 183-216). New York: Springer.

Saleiro, P., Amir, S., Silva, M., & Soares, C. (2015). Popmine: Tracking political opinion on the web. In Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on (pp. 1521-1526). IEEE.

Sang, E. T. K., & Bos, J. (2012). Predicting the 2011 dutch senate election results with twitter. In *Proceedings of the Workshop on Semantic Analysis in Social Media* (pp. 53-60). Association for Computational Linguistics.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and social psychology bulletin*, *34*(8), 1096-1109.

Sewell, M. (2011). History of the efficient market hypothesis. *RN*, *11*(04), 04.

Skoric, M., Poor, N., Achananuparp, P., Lim, E-P., and Jiang, J. (2012). In Proceedings of the 45th

Hawaii International Conference on System Sciences.
Sobhani, P., Mohammad, S. M., & Kiritchenko, S. (2016). Detecting Stance in Tweets And Analyzing its Interaction with Sentiment. *cc2016 The* SEM 2016 Organizing Committee. All papers cc2016 their respective authors. This proceedings volume and all papers therein are licensed under a Creative Commons Attribution 4.0 International License. License details: http://creativecommons. org/licenses/by/4.0*, 159.
Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20*(5), 926–957. doi:10.1111/j.1468-036X.2013.12007.x
Statista (2016). *Statista Reddit.* Accessed 3 November, 2016, available at: https://www.statista.com/statistics/517155/reddit-user-distribution-usa-gender/
Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The General Inquirer: A Computer Approach to Content Analysis.
Strapparava, C., & Valitutti, A. (2004). WordNet Affect: an Affective Extension of WordNet. In *LREC* (Vol. 4, pp. 1083-1086).
Sul, H., Dennis, A. R., & Yuan, L. I. (2014). Trading on Twitter: The financial information content of emotion in social media. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on System Sciences* (pp. 806–815). IEEE.
Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.
Tsay, R. S. (2005). *Analysis of financial time series* (Vol. 543). John Wiley & Sons.
Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*, *10*, 178-185.
Wilkinson, N., & Klaes, M. (2012). *An introduction to behavioral economics*. Palgrave Macmillan.
Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., & Patwardhan, S. (2005). OpinionFinder: A system for subjectivity analysis. In *Proceedings of hlt/emnlp on interactive demonstrations* (pp. 34-35). Association for Computational Linguistics.
Winter, E. (2014). *Feeling smart: Why our emotions are more rational than we think*. PublicAffairs.
Wired (2016). *Wired end of code.* Accessed January 10, 2017. Available at: https://www.wired.com/2016/05/the-end-of-code/
World stat. (2016). *World stat.* Accessed November 4, 2016. Available at: http://www.internetworldstats.com/stats.htm
Zhang, X., Fuehres, H., & Gloor, P. A. (2012). Predicting asset value through Twitter buzz. In *Advances in collective intelligence 2011* (pp. 23–34). Springer.

## Appendix

*Appendix 1. RCP public opinion polling sources*

| | |
|---|---|
| 1 | ABC News/Wash Post |
| 2 | Associated Press-GfK |
| 3 | Bloomberg |
| 4 | CBS News |
| 5 | CBS News/NY Times |
| 6 | CNN/Opinion Research |
| 7 | CNN/Time |
| 8 | Cook/RT Strategies |
| 9 | Democracy Corps (D) |
| 10 | Diageo/Hotline |
| 1 | Economist/YouGov |
| 12 | FOX News |
| 13 | Gallup |
| 14 | GWU/Battleground |
| 15 | Hotline/FD |
| 16 | IBD/TIPP |
| 17 | Ipsos/McClatchy |
| 18 | Marist |
| 19 | McClatchy/Marist |
| 20 | Monmouth |
| 21 | National Journal |
| 22 | National Journal/FD |
| 23 | NBC News |
| 24 | NBC News/Wall St. Jrnl |
| 25 | Newsweek |
| 26 | NPR |
| 27 | NPR /POS/GQR |
| 28 | Pew Research |
| 29 | Politico/GWU/Battleground |
| 30 | POS (R) |
| 31 | PPP (D) |
| 32 | Quinnipiac |
| 33 | Rasmussen Reports |
| 34 | Reason-Rupe/PSRAI |
| 35 | Resurgent Republic (R) |
| 36 | Reuters/Ipsos |
| 37 | The Economist/YouGov |
| 38 | Time |
| 39 | USA Today/Gallup |
| 40 | USA Today/Pew Research |
| 41 | USA Today/PSRAI |
| 42 | USA Today/Suffolk |
| 43 | Washington Post |
| 44 | Zogby |

*Appendix 2.* Correlation matrix of intraday normalized values

The values in the correlation matrix with corresponding significance provide the input for the correlogram in Figure 14

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 Rcp - approval (DV1) | 1 | -.0219 | .1377*** | .0454*** | .0731*** | -.0198 | -.0406*** | -.0048 | -.1227*** |
| 2 Anger_n (IV1) | | 1 | -.6540*** | .4863*** | .5875*** | -.6797*** | .1282*** | -.3082*** | -.6056*** |
| 3 Anticipation_n (IV2) | | | 1 | -.4988*** | -.6035*** | .6317*** | -.3616*** | .2472*** | .3616*** |
| 4 Disgust_n (IV3) | | | | 1 | .2830*** | -.5667*** | .2107*** | -.3620*** | -.4793*** |
| 5 Fear_n (IV4) | | | | | 1 | -.6872*** | .1451*** | -.3290*** | -.5510*** |
| 6 Joy_n (IV5) | | | | | | 1 | -.3770*** | .3391*** | .4238*** |
| 7 Sadness_n (IV6) | | | | | | | 1 | -.2435*** | -.3869*** |
| 8 Surprise _n (IV7) | | | | | | | | 1 | -.0280 |
| 9 Trust_n (IV8) | | | | | | | | | 1 |

*Note*. N = 2589 . *** *p* < .01. ** p <.05. *p<0.10
DV = Dependent Variable. IV = Independent Variable.

*Appendix 3.* Time series and (P)ACF plots for Robustness checks

The ACF plot can be used to estimate the MA-part, the PACF can be used to estimate the AR-part. The plots are merely a chart of the coefficients of correlation between a time series and the lags of itself. The PACF plot is a plot of the partial correlation coefficients between the series and the lags of itself. When the spikes exceed the blue dotted significance bands, autocorrelation exists. When strong autocorrelation exists it might be impossible to separate the effect of individual emotions on the approval rate. The residuals do not look like white noise, since autocorrelation exist at various lags for Gallup and Rasmus.