2012 Eighth Latin American Web Congress

Opinion Dynamics of Elections in Twitter

Felipe Bravo-Marquez*, Daniel Gayo-Avello†, Marcelo Mendoza‡§ and Barbara Poblete*§
*Department of Computer Science, University of Chile, Chile
†University of Oviedo, Spain
‡Universidad Técnica Federico Santa María, Chile
§Yahoo! Research Latin America, Chile

***Abstract*—In this work we conduct an empirical study of opinion time series created from Twitter data regarding the 2008 U.S. elections. The focus of our proposal is to establish whether a time series is appropriate or not for generating a reliable predictive model. We analyze time series obtained from Twitter messages related to the 2008 U.S. elections using ARMA/ARIMA and GARCH models. The first models are used in order to assess the conditional mean of the process and the second ones to assess the conditional variance or volatility. The main argument we discuss is that opinion time series that exhibit volatility should not be used for long-term forecasting purposes. We present an in-depth analysis of the statistical properties of these time series. Our experiments show that these time series are not fit for predicting future opinion trends. Due to the fact that researchers have not provided enough evidence to support the alleged predictive power of opinion time series, we discuss how more rigorous validation of predictive models generated from time series could benefit the opinion mining field.**

***Keywords*-opinion mining; time series analysis; volatility; twitter**

I. Introduction

Public opinion is a complex collection of the beliefs of many different people about specific subjects. It is more than the simple aggregation of individual points of view. The process by which public opinion arises involves several factors such as the credibility of opinion leaders, peer to peer influence and social media pervasiveness, among many other aspects.

Leveraging public opinion in relation to different subjects and entities, is a matter which interests public and private institutions from a wide range of fields. The unbiased understanding and representation of public opinion is a problem that concerns the areas of Politics and Economics.

Commonly, the extraction of public opinion regarding a certain topic is studied through opinion polls. These are a set of questions about a certain subject which are surveyed on a small sample of the population. This approach has one major drawback which is that polls can inadvertently bias the outcome and, thus, report a wrong opinion. The problem resides in the fact that the wording used in questions, the questions themselves and the order in which they are asked, can influence how people respond (cf. [21]).

The emergence of the *social web* has allowed researchers to observe how people feel and react regarding a great variety of topics. The richness of this information is due to the fact that it is provided by the users freely and willingly. Social web platforms allow users to post and exchange (short) messages about topics which interest them, using the language that they feel more appropriate at that moment. On the social web we can find subjective, as well as objective information, including opinions and emotions. Thus, the social web becomes an ideal environment for *live* public opinion observation.

In spite of all of its advantages, opinion mining on the social web cannot replace opinion polls. The main concern with this type of analysis is that the population which uses the social web is not necessarily a representative sample of the *real world*. For instance, in December 2008, only 11% of the U.S. adult population was using Twitter[1], which is one of the most popular services. Although this amount of users is a significant fraction of the population, the usage of Twitter itself can produce a biased dataset. Therefore, the conclusions obtained from that kind of analysis will only reflect the public opinion regarding a particular fraction of the population. Nevertheless, opinions extracted from social media provide some benefits in comparison to traditional polling.

First of all, this approach allows to cheaply process greater amounts of data. Secondly, as social media opinions become available in continuous time streams, we believe that social media is more suitable for studying the temporal properties of public opinion. In this work we define *opinion dynamics* as how public opinion evolve through time.

As we show in Section II, there has been a recent surge in papers claiming impressive forecasting abilities derived from social media. Unfortunately, most of these claims are misleading because they ignore the temporal aspects of the opinions. We address this issue by applying ARMA/ARIMA and GARCH models [4], [7], [3] to opinion time series to determine if they are appropriate for making reliable forecasts. We focus on different aspects of opinion time series, especially on the notion of opinion *volatility*.

We conduct an experimental exploration of opinion time series extracted from Twitter related to the 2008 U.S. Presidential elections. We analyze these time series finding

[1]http://www.twitter.com

978-0-7695-4839-5/12 $26.00 © 2012 IEEE
DOI 10.1109/LA-WEB.2012.11

that these series present an important volatility factor, which would impede producing accurate long-term predictions.

This article is organized as follows. In Section II we discuss related work. We introduce opinion time series analysis tools in Section III. In Section IV we show how opinion time series are created from Twitter and in Section V-A the dataset is explained. Experimental results are presented in Section V-B and discussed in Section V-C. Finally, we conclude in Section VI with a brief discussion about some open questions and future work.

II. Related Work

There exists a substantial body of work regarding opinion mining and sentiment analysis. [18] provides a thorough survey on both the theoretical aspects of the field and on the different methods that can be applied to obtain opinion oriented information. Unsurprisingly, machine learning has been widely applied (see for instance [19]) although the performance of ML methods in sentiment analysis is noticeably worst than in other areas of NLP. Apart from machine learning, the most straightforward and widely applied way of finding the sentiment (or polarity) of a piece of text is by relying on the prior polarity of the words it's composed of (e.g. [23], [8]).

Most recently, there has emerged an interest in understanding temporal aspects of opinions and furthermore, in predicting future events from social media. [15] showed that sentiment dynamics in social media exhibit a certain degree of seasonality and, thus, can be predicted with reasonable accuracy. Later, a number of scenarios have been studied looking for correlation between "real-world" events and those online sentiment dynamics such as movie-box-office sales performance [1], [16] or the evolution of the stock market [2].

In this regard, politics has constituted its own subfield in the realm of prediction: studying the predictive power of sentiment analysis of social generated content with regards to the electoral outcome. The accuracy of such methods is debatable to say the least: while some researchers claim that their predictive power is *"close to traditional election polls"* [22], others argue that such power *"is greatly exaggerated, especially for political elections"* [9].

However, it must be noted that such electoral "predictions" were made *post-facto* and, unlike the works regarding box-office performance or stock market evolution most of them ignored the temporal component of the social data collected (i.e. they did not infer any time series from the social data).

Furthermore, public opinion is not static but dynamic and significant conclusions cannot be drawn by just relying on the existence of a few positive reports which could be attributed to chance [14]. That's why rigorous analyses of opinion dynamics are needed and in this sense, there have been some research on this area, albeit different from our approach.

[12] proposed the ARSA model which is an autoregressive model subsuming a Latent Semantic model for sentiment analysis; ARSA was then used to predict box office performance. [5], [20] used SVM regressors to predict time series; to that end, the SVMs were trained using both historical values of the time series and texts published at each time point. Finally, [10] applied an ARMAX model to predict the presidential approval rating by combining historical values from both the ratings and sentiment information from social media.

In reference to the above works, and to the best of our knowledge, no other research work has performed a deep statistical analysis of opinion time series created from social media. We strongly believe that the study of the volatility and other aspects of opinion time series, like seasonality and stationarity, will allow us to determine the limitations of assessing opinion dynamics from social media.

III. Time series Analysis Tools

An opinion time series can be defined as a sequence of opinion values spaced at uniform time intervals $X_1, \ldots, X_t, \ldots, X_n$. An opinion value X_t is a measure which reflects a dimension of the public opinion in a certain period regarding a specific topic or event.

In order to establish if an opinion time series should be discarded as a basis for a predictive model, we strongly recommend that at least a minimum amount of tests should be performed. We believe that without these methodological tests, favorable results of predictive models do not provide enough evidence to support their forecasting power.

We propose the application of the Box-Jenkins methodology [4] based on ARMA/ARIMA models for the expected mean, and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models [7], [3] for the volatility analysis.

An ARMA(p,q) process is defined by the following expression:

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=0}^{q} \beta_j \epsilon_{t-j} \tag{1}$$

The first summation refers to the autoregressive part of the model ($AR(p)$), the second one refers to the moving average part ($MA(q)$) and ϵ_t is a series of uncorrelated random variables with mean zero and variance σ^2. An ARIMA(p,d,q) model is a process whose d-th difference is an ARMA process.

As the Box-Jenkins methodology suggests, first of all, a model specification step must be conducted. Here, the stationarity of the time series must be checked using methods such as the Augmented Dickey-Fuller unit root test. A time series is weakly stationarity when the mean function is constant through time and the covariance function for any pair of periods be finite and identical when the lags of the periods are the same.

Due to the fact that ARMA models are defined for stationary time series, in the case of having a non-stationary time series, it should be differenced until the stationarity conditions are satisfied, where d is the number of times the time series is differenced. Moreover, the order of the autoregressive and moving average parameters (p,q) of the ARIMA(p,d,q) model can be identified from the shape of the autocorrelation and partial autocorrelation plots. Parameters α_i, β_j from Equation 1 can be estimated by one of several methods such as: the methods of moments, least squares or maximum likelihood.

It is recommended to fit a grid of plausible models varying the values of p and q. Then the best model can be chosen according to the Akaike's information criterion (AIC), the Bayesian information criterion (BIC) among others criteria. Once the correct parameters are estimated, a diagnostic check or residual analysis should be performed.

The previous procedure generates an ARMA or ARIMA predictive model, but it will not be reliable if the one-step-ahead conditional variance does not always show the same value as the noise variance. Time series which present volatility do not meet this criteria, therefore ARIMA models cannot generate reliable predictive models.

We hypothesize that volatility can be a very relevant aspect in opinion time series. Intuitively, during hectic periods, people tend to be more sensitive to information and hence opinion trends register larger fluctuations. Therefore, large opinion changes are followed by large opinion fluctuations. In this situation, ARMA/ARIMA predictive models are not reliable.

Volatility effects have been studied in price theory for many years. Mandelbrot [13] observed that large price changes were followed by large price fluctuations and small price changes were followed by small price fluctuations. The patterns of changing from quiet to volatile periods is named as *volatility clustering*. Time-sensitive volatility analysis allows the identification of hectic periods (large fluctuations) and calm periods (small fluctuations). The most suitable models that deal with volatility are the GARCH models which are discussed below.

Let $\sigma^2_{t|t-1}$ be the expected conditional variance or volatility of a zero-mean time series r_t at period t, the GARCH(q,p) process that models $\sigma^2_{t|t-1}$ is defined as follows:

$$\sigma^2_{t|t-1} = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon^2_{t-1} + \sum_{j=1}^{p} \beta_j \sigma^2_{t-j|t-j-1} \qquad (2)$$

Here, the first term α_0 together with the first summation refer to the autoregressive conditional heteroscedasticity (ARCH) part of the model and the second summation reflects how past values of $\sigma^2_{t|t-1}$ are fedback into the present value. It is clear from Equation 2 that ARCH models are included in GARCH models as a special case. For the remainder of this paper, we will use the term "GARCH" to refer to both ARCH and GARCH models. Considering that GARCH models require a zero-mean time series, a common approach used in financial time series is to model the continuously compounded return r_t of a positive time series X_t (e.g. stock prices), where the return values are expressed as: $r_t = log(\frac{X_t}{X_{t-1}})$. It is also possible to work with the residuals of a non-zero-mean time series. It is important to state, that while ARIMA models are used to model the expected mean of the time series, GARCH models are focused on modeling the past conditional variance. As stated in [6], the past conditional variance or volatility of a time series given past observations, measures the uncertainty in the deviation of the time series from its conditional mean.

We rely on the McLeod-Li test for detecting conditional heteroscedasticity in the time series. This test is equivalent as the Box-Ljung statistics applied to the squared returns[2] or residuals, and basically detects if the squared data are autocorrelated. In the case of rejecting the null hypothesis, this implies the presence of a past conditional variance. In this scenario, opinion series are volatile, being very difficult to forecast the opinion trends in the long term using ARMA or ARIMA models. Moreover, if the volatile time series is correctly specified, a GARCH model could be fitted using maximum likelihood estimators and hence future volatility values could be forecasted. A deeper discussion of the methods referenced in this section can be checked in [6].

IV. Building Twitter Opinion Time Series

Now we describe how opinion time series from temporal collections of tweets can be created. First of all, we define a temporal event-related collection of tweets $\mathcal{C}$, as a collection of timestamped tweets associated with a certain event e. Here, the hypothesis is that an event e is represented by a set of keywords and that a tweet containing these keywords will be associated with e. Then, a sentiment evaluation for each tweet belonging to $\mathcal{C}$ is performed. As was discussed in Section II, this evaluation can be achieved by several approaches. In this article we apply the lexicon-based approach proposed in [23] which is based on counting positive and negative opinion words from a given list. The main advantage of this simple method is that, unlike supervised methods, which require hand-labeled corpora for training, by relying on a lexicon we could skip the laborious –and subjective– task of labeling a corpus of tweets. Furthermore, although the lexicon method is quite simple, machine learning methods applied to sentiment analysis are still rather far away from the performance usually achieved in other NLP tasks.

[2]In GARCH models the squared returns are unbiased estimators of the unobserved conditional variance.

As tweets in the collection are timestamped, they can be aggregated by time periods such as days, weeks or months. Below, we describe opinion values which were extracted from tweets in this work: The **activity level**, is the number of tweets associated to the event in the time period. The **positiveness** for a certain event e in a time period t is defined as the average number of positive words per tweet in the period. Likewise, we define the **negativeness** as the average number of negative words in the period. The commonly used measure of **polarity** can also be expressed as the difference between positiveness and negativeness. It is important to note, that we model temporal occurrence of positive and negative words as two separated processes as in [11]. We believe that this representation facilitates a better observation of the temporal properties of the event and at least in the datasets used for this paper, negative and positive sentiment are mutually uncorrelated as we will show in the next section. Furthermore, as both positiveness and negativeness measures are always positive they can be transformed into log-return values which are more appropriated for GARCH models and, hence, for assessing the volatility.

V. Experimental Evaluation

A. Dataset Description

The dataset consists of tweets associated with the U.S. elections of 2008. A thorough description of the data gathering process for the first collection is detailed in [9]. This collection contains 250,000 tweets, published by 20,000 Twitter users from June 1, 2008 to November 11, 2008. All of the tweets are related to either the Democrat ticket (Obama-Biden) or to the Republican one (McCain-Palin); the Twitter Search API was used using one query per candidacy.

Opinion time series created from this dataset are shown in Figure 1. From top to bottom, the first plot shows polarity time series for each candidate, the second one represents activity levels in Twitter together with relevant dates related to the event, and the last ones represents positiveness and negativeness time series for each candidate. By the remainder of this work opinion time series of Obama and McCain will be referred as $(O.+)$ and $(M.+)$ for positiveness and $(O.-)$ and $(M.-)$ for negativeness respectively.

B. Experimental Results

Following the previously described methodology (Section III), we have analyzed each opinion time series described in Section V-A. We first performed an exploratory analysis of the time series.

Scatter plots between positiveness and negativeness opinion time series are shown in Figure 2, (a) Obama-Biden, and (b)McCain-Palin. Pearson correlation between positiveness and negativeness are -0.0698, and 0.0682 for Obama and McCain respectively.

Cross-correlation coefficients between different series-pairs are: $(O.+, M.+) = 0.21$, $(O.-, M.-) = -0.14$, $(O.+, M.-) = 0.17$, and $(O.-, M.+) = 0.01$. Figure 2,(c) shows a scatter plot between Twitter activity and opinion polarity, using log axes. Pearson correlation coefficients for $O.+$, $O.-$, $M.+$, and $M.-$ are 0.13, 0.08, 0.08, and 0.11, respectively. Furthermore, we performed Pearson correlation tests between all pairs mentioned above with significance level of $\alpha = 0.05$. With the exception of pair $(O.+, M.+)$, all p-values obtained are greater than 0.05. These results validate the idea of modeling positiveness and negativeness as separate time series and show us that sentiment measures have no linear relationship with the level of activity in the period.

To check the stationarity of the time series we conduct the Augmented Dickey-Fuller test whose results are shown in Table I. Obtained Augmented Dickey-Fuller (ADF) statistics and p-values allow to reject the null hypothesis for every opinion time series. Stationarity implies that each time series can be studied by fitting only one model. Thus, we can apply the Box-Jenkins methodology by fitting stationary models to each time series.

Time series	ADF test	p-value
O.+	-7.117	< 0.01
O.-	-9.567	< 0.01
M.+	-10.715	< 0.01
M.-	-6.016	< 0.01

Table I: **Augmented Dickey-Fuller statistics for trend non-stationarity testing**.

Seasonal patterns are also studied in order to find cyclical periods for the positiveness and negativeness in the data. A possible approach is to estimate multiplicative seasonality factors (e.g day of the week) for each season. As was suggested in [15], we estimated weekly seasonal coefficients for each US elections time series. For each day of the week, we calculate the ratio of actual values divided by predicted values, according to linear trend regression models applied to the series. These values should be near to 1 in the absence of seasonal patterns. As it can be seen from Table II, there are coefficients being different than one when we suppose a period of a week. Correlograms for each time series shows similar results when we analyze the 7-th lag suggesting that seasonal patterns are conditioned to the day of the week.

Opinion time series can also be smoothed in order to derive a more consistent signal as was done in [17]. Some possible smoothing approaches are moving average, moving median, exponential smoothing, among others. We evaluate the use of moving averages of seven days according to the weekly seasonal patterns described above. It is important to consider that smoothed opinion time series can cause the opinion variable to respond more slowly to recent changes [17]. Thus, we fit ARMA/ARIMA models to each time series, considering also its smoothed versions.

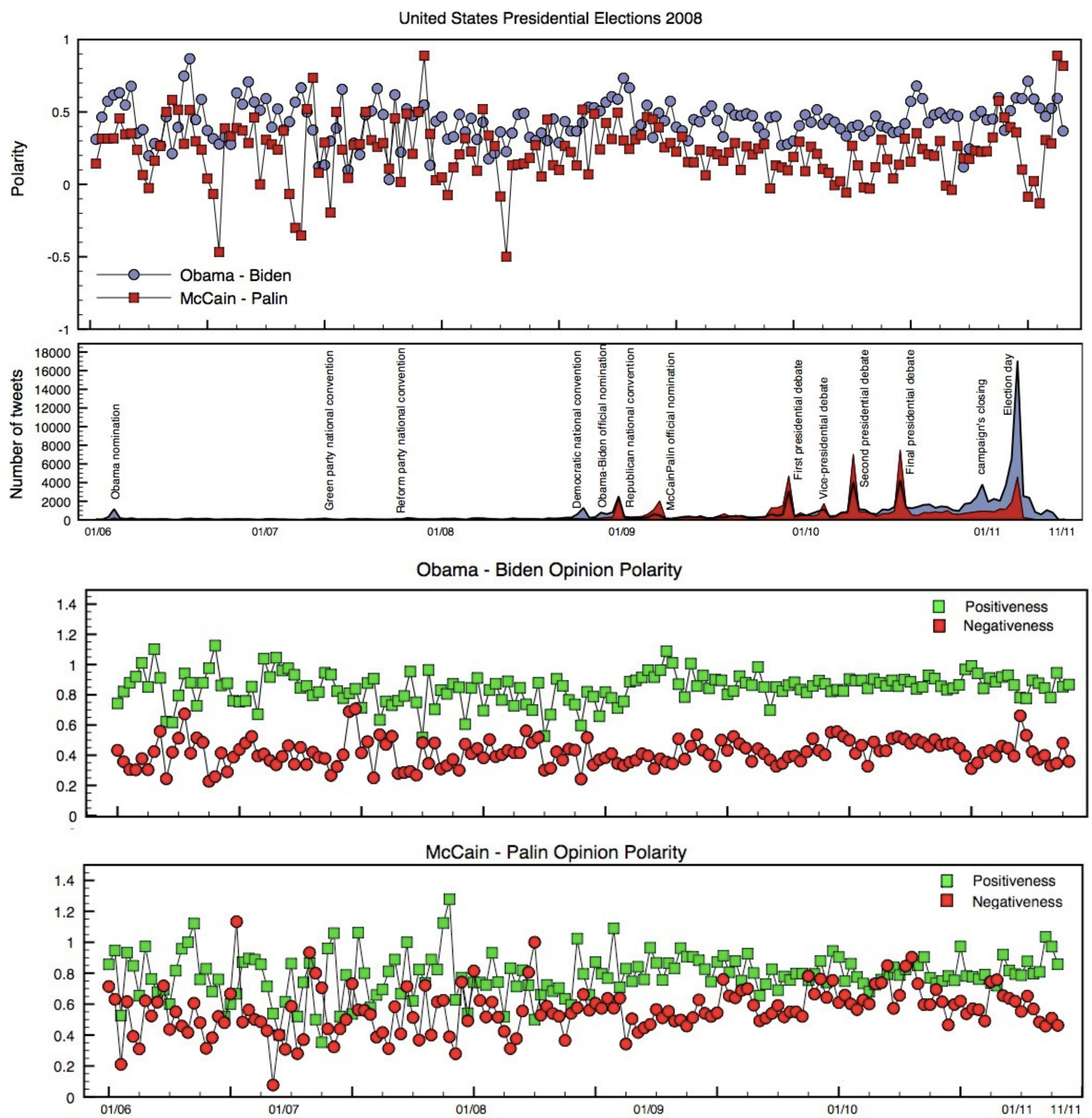
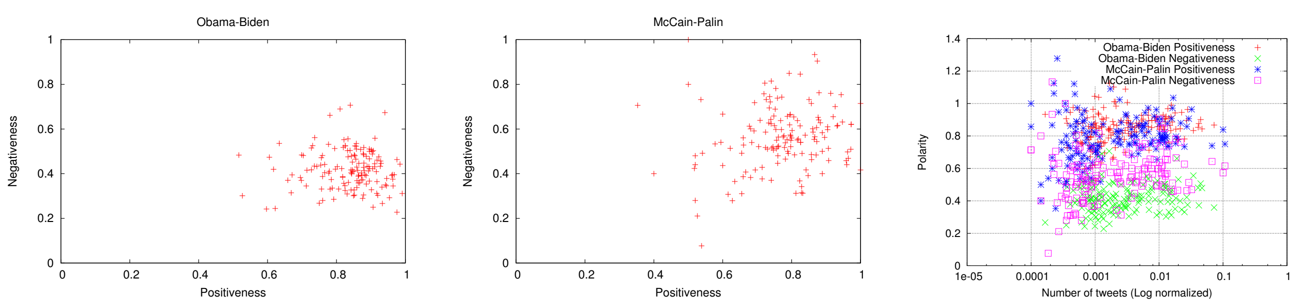


Figure 1: **Opinion time series for the US Presidential Election of 2008**.



(a) (b) (c)

Figure 2: **Scattering analysis for the polarity time series of the US Election 2008**.

Day	O.+	O.-	M.+	M.-
Sunday	1.018	0.944	1.055	1.156
Monday	0.961	1.019	0.995	0.942
Tuesday	0.971	0.986	0.975	0.930
Wednesday	1.026	1.063	0.963	0.944
Thursday	1.002	1.020	1.013	0.932
Friday	0.990	0.992	0.987	1.044
Saturday	1.030	0.975	1.007	1.046

Table II: **Trend seasonality factors for the polarity time series of the US Election 2008**.

Model selection was performed by fitting high order models to each time series. In the case of the US elections we consider an additional multiplicative seasonal ARMA model to incorporate seasonality. The use of high order models allows us the observation of the coefficient values for over-parameterized models, identifying coefficients with significant error standard measures. We avoid the use of models with poor fitting properties by conducting model cheeking with a confidence level of 95%. Model check tests were rejected under this threshold, suggesting to us the presence of ARIMA components. Then we fit ARIMA models to this subset of time series, conducting similar model selection and model evaluations steps. We separate each time series into two parts, one for model fit/test and a second part for time series forecasting. Model fitting/testing was conducted over the first three months of the US elections.

Forecasting results obtained from each time series are shown in the first two columns of Figure 3. From top to bottom, we show Obama Positiveness, and Negativeness, and McCain Positiveness, and Negativeness forecasts, respectively. The first column shows the results achieved for the original time series and the second one shows results for smoothed versions of each time series, using moving averages of seven days. Actual values are shown with black points and predicted values with blue points. Error margins for predicted values are depicted with segmented lines.

Forecasting results are far from being accurate in the original time series. Forecasts can at least model the mean of future outcomes but not the variance. For the smoothed versions of the time series, forecasting results are improved in some cases. As will be see in the following analysis, the difficulty of predicting future outcomes in these time series is due to the presence of volatility.

In order to assess the volatility we used the financial risk management convention of converting the original time series to log return values as was explained in Section III. The return time series are referred as $R_O.+$ and $R_O.-$ for the positiveness and negativeness of Obama and likewise as $R_M.+$ and $R_M.-$ for McCain respectively. The results of the volatility analysis for the transformed time series is summarized in Table III. Below we describe the tests and measures which were considered in the analysis.

We first checked if the desirable conditions for GARCH modeling were met in the transformed series. The zero-mean condition was tested through a zero-mean t-test, were in all cases we failed to reject the null hypothesis. Due to the fact that volatile time series capture a non-Gaussian fat-tailed distribution [6], we evaluated the excess of kurtosis of the returns[3], which were positive in all cases. Then, we checked the presence of volatility in the return series through the McLeod-Li test at considering different lags. For all the series the average p-values for the lags considered were less than 0.05. All these results indicate that our transformed time series are appropriate for GARCH modeling. Moreover, it is important to remark that this conditions were not satisfied in the original series. We fitted a grid of possible GARCH models to the series varying the orders of q and p from 1 to 3, where in all cases the best resulted model was a GARCH(1,1) model. The quality of the fitted models was assessed by considering the significance of the α_i and β_j coefficients through zero-mean t-tests. As it shown in Table III the p-values obtained from the t-tests applied to the coefficients of the GARCH(1,1) models were all close to zero. Thus, we have statistical evidence that GARCH(1,1) models are appropriate for modeling our transformed time series. Finally the fitted models were used to estimate the conditional variance of the transformed time series.

	$R_O.+$	$R_O.-$	$R_M.+$	$R_M.-$
Kurtosis	2.09	0.978	0.346	1.99
Zero-mean t-test p-value	0.936	0.955	0.999	0.925
McLeod-Li avg. p-value	0.000	0.023	0.002	0.000
α_1 t-test p-value	0.000	0.015	0.004	0.001
β_1 t-test p-value	0.000	0.000	0.000	0.000
Mean Volatility	0.028	0.073	0.058	0.119

Table III: **Volatility Analysis of log return time series of the US Election 2008**.

The third column of Figure 3 shows from top to bottom the fitted conditional variances of $R_O.+$, $R_O.-$, $R_M.+$ and $R_M.-$ time series respectively. Although all volatility time series exhibit calm and volatile periods, an interesting insight derived from these plots is that the volatility or conditional variance of the series tends in all cases to decrease while approaching the election day. This can be interpreted in the following way: at the beginning of the election period people could have been more open to new information and hence there was more uncertainty about the voting preferences. However, while getting closer to the election day the preferences of the voters became clearer and hence the change in the opinion pattern was reduced.

C. Discussion

The experimental results from Section V-B show us how opinion time series created from Twitter data regarding the 2008 U.S elections tend to be volatile and hence predictions

[3]Fat-tailed distributions have positive kurtosis.

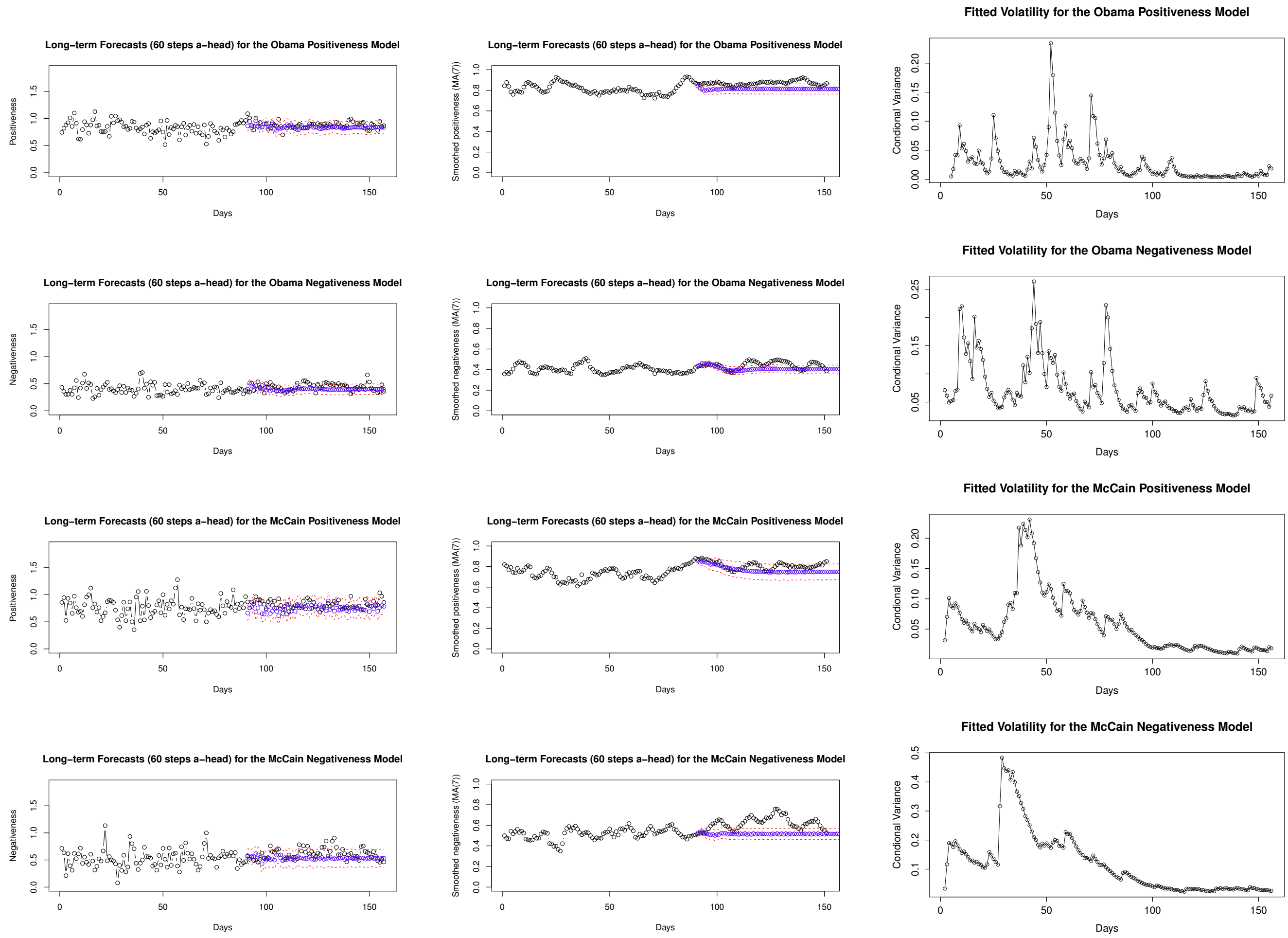


Figure 3: **Long term forecasting and volatility for opinion time series of the US Election 2008**.

of future outcomes in the long-term following the Box-Jenkins methodology are limited. A possible approach to reduce the level of uncertainty is to smooth the time series. These transformed time series are maybe more appropriate for prediction, but their outcomes will fit slower to opinion changes. The presence of volatility in our return transformed time series suggests that the past conditional variance can be modeled by estimating the parameters of the GARCH model. Thus, calm and volatile periods could be identified and predicted in the future. This means that although the conditional mean cannot be predicted using ARMA/ARIMA models, the volatility could be properly modeled and forecasted using GARCH models. Forecasting the volatility could be used as a measure of risk in public opinion as it is used in finance. As it is shown in Section II, a significant amount of work has been done regarding prediction based on opinion mining in social media. According to this and to our experimental results we address the following question: Is social media an accurate proxy for opinion dynamics? Yes? No? Depending on the scenario? If the former was the case, how could it be checked beforehand to prove that forecasting is feasible and, thus, that positive results are not a product of chance? Certainly, we cannot say a priori if the event we want to analyze will be predictable or not. The data itself can only answer that question. Furthermore, as more accurate NLP methods for assessing the sentiment in social media messages are developed, opinion time series created with those methods will reflect in a better manner the opinion dynamics of the population. The main contribution of this work, is a methodology that can be used as a framework for analyzing opinion time series. This methodology will allow to identify if the opinion time series are indeed predictable, or if the past conditional variance of the log returns of the process could be modeled due to the presence of volatility.

VI. Conclusions

In this work we conducted an empirical study to check whether Twitter extracted opinion time series related to a specific event are or not suited for generating predictive models. Mainly we have argued that time series volatility is a key aspect which needs to be studied before claiming that a model's predictions are product of more than just random chance. We have shown how to apply this methodology to opinion time series created from Twitter concluding that these time series do not generate reliable forecasting models. We show interesting properties of these time series in an in-depth exploratory analysis. In addition we discussed prior work on opinion time series and how this research can benefit from rigorous analysis of volatility estimators.

For future work we expect to analyze opinion time series from different sources, such as face-to-face opinion polls and data in other languages, such as the recent Spanish elections and the US elections 2012. In the last case, an extensive analysis of some of the most important features of the time series will be conducted prior to the election day, trying to conduct an *in vivo* analysis of this event.

VII. Acknowledgment

This work has been partially supported by FONDEF-CONICYT project D09I1185. The first author is granted by CONICYT's Master Scholarship.

References

[1] Sitaram Asur and Bernardo A. Huberman. Predicting the future with social media. In *Web Intelligence*, pages 492–499, 2010.

[2] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. Twitter mood predicts the stock market. *J. Comput. Science*, 2(1):1–8, 2011.

[3] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, April 1986.

[4] George Edward Pelham Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 3rd edition, 1994.

[5] Munmun De Choudhury, Hari Sundaram, Ajita John, and Dore D. Seligmann. Can blog communication dynamics be correlated with stock market activity? In *Hypertext*, pages 55–60. ACM, 2008.

[6] Jonathan D. Cryer and Kung-Sik Chan. *Time Series Analysis: With Applications in R (Springer Texts in Statistics)*. Springer, 2nd edition, June 2009.

[7] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):pp. 987–1007, 1982.

[8] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, pages 417–422, 2006.

[9] Daniel Gayo-Avello. Don't turn social media into another 'literary digest' poll. *Commun. ACM*, 54(10):121–128, 2011.

[10] Sandra Gonzalez-Bailon, Rafael E. Banchs, and Andreas Kaltenbrunner. Emotional reactions and the pulse of public opinion: Measuring the impact of political events on the sentiment of online discussions. *CoRR*, abs/1009.4019, 2010.

[11] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188, 2009.

[12] Yang Liu, Xiangji Huang, Aijun An, and Xiaohui Yu. Arsa: a sentiment-aware model for predicting sales performance using blogs. In *SIGIR*, pages 607–614. ACM, 2007.

[13] Benoit Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4):394–419, 1963.

[14] Panagiotis T Metaxas, Eni Mustafaraj, and Daniel Gayo-avello. How (not) to predict elections. In *Proceedings of IEEE SocialCom Conference*, October 2011.

[15] G.A. Mishne and M. de Rijke. Capturing global mood levels using blog posts. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*. AAAI Press, 2006.

[16] Gilad Mishne and Natalie Glance. Predicting movie sales from blogger sentiment. In *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pages 155–158, 2006.

[17] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2010.

[18] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[19] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86, 2002.

[20] Joshua Ritterman, Miles Osborne, and Ewan Klein. Using prediction markets and Twitter to predict a swine flu pandemic. In *Proceedings of the 1st International Workshop on Mining Social Media*, November 2009.

[21] Howard Schuman and Stanley Presser. *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Sage Publications, Inc., 1996.

[22] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *ICWSM*. The AAAI Press, 2010.

[23] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 347–354, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.