Conference papers

School of Computer Sciences

2019

BREXIT: A granger causality of twitter political polarisation on the FTSE 100 Index and the Pound

James Usher

Lucia Morales

Pierpaolo Dondio

Follow this and additional works at: https://arrow.tudublin.ie/scschcomcon

Part of the Computer Sciences Commons

BREXIT: A Granger Causality of Twitter Political Polarisation on the FTSE 100 Index and the Pound

James Usher, Lucía Morales, Pierpaolo Dondio

Technology University Dublin
Kevin Street
Dublin 2, Ireland
James.usher@student.dit.ie

Abstract—BREXIT is the single biggest geopolitical event in British history since WWII. Whilst the political fallout has become a tragicomedy, the political ramifications has had a profound impact on the Pound and the FTSE 100 index. This paper examines Twitter political discourse surrounding the BREXIT withdrawal agreement. In particular we focus on the discussions around four different exit strategies known as *"Norway", "Article 50"*, the *"Backstop"* and *"No Deal"* and their effect on the pound and FTSE 100 index from the period of rumblings of the cancellation of the Meaning Vote on December 10th 2018 inclusive of second defeat on the Prime Minister's BREXIT exit strategy on February 14th to February 24th 2019. Our approach focuses on using a Naive Bayes classification algorithm to assess political party and public Twitter sentiment. A Granger causality analysis is then introduced to investigate the hypothesis that BREXIT political and public sentiment, as measured by the twitter sentiment time series, is indicative of changes in the GBP/EUR Fx and FTSE 100 Index. Our results indicate that the accuracy of the *"Article 50"* scenario had the single biggest effect on short run dynamics on the FTSE 100 index, additionally the *"Norway"* BREXIT strategy has a marginal effect on the FTSE 100 index whilst there was no significant causation to the GBP/EUR Fx.

Keywords—Web intelligence, Data mining, Twitter sentiment, Stock Market

I. INTRODUCTION

It's no secret that the modern day society encompasses political discourse via Twitter and other social media platforms. Twitter is one of the most powerful tools in modern communication. The microblogging platform allows its 300 million monthly active users to relay their personal sentiment to the general public. Using data mining tools, political scientists aim to unearth, discover and determine the political and public sentiment. In a 2010 study [1], a group of political scientists found that Twitter is a platform used for political deliberation. That study also found that Twitter is also used as a mechanism to spread political opinion. Analysis of such data is deemed important with today's news broadcasters such as the BBC, Sky News and RTE conducting regular Twitter polls to gauge the current political and public sentiment. Politicians and stock market analysts are groups that has benefitted from Twitter and relied upon it for networking [2]. Barack Obama's 2008 presidential campaign amplified Twitter's effectiveness as a tool in American politics, proving how it could be used effectively to communicate with voters [1]. Today, Twitter remains a forum for lively political debates and when the sentiment analysis is compared to stock market performance, it can create a useful barometer of how sentiment influences the capital markets [3].

This paper examines Twitter political discourse surrounding the BREXIT withdrawal and examines the effect of the key strategies such as "*Norway", "Article 50", "Backstop"* and *"No Deal"* had on the GBP/EUR Fx and FTSE 100 Index

from the period of December 10th 2018 to February 24th 2019. The selection rational behind these strategies was based on data mining the repository database for tweets that mentioned any of the predetermined BREXIT governmental strategies. Part II looks at the related works. Part III presents our data collection and architecture. Party IV takes into account the econometric methodology deployed, followed by the results in Part V, and conclusions in Part VI.

II. RELATED WORK

Today, analysis of this data is important to political parties; not only does it help them tailor their political strategies, it also allows parties define and underscore their groups and areas they need to influence. Similarly, investors can gain meaningful insights and tailor their investment strategies accordingly based on political and public sentiment. Following the success of President Barack Obama's 2008 Presidential campaign, the world saw the "crowning of the Internet as the king of all political media" [4]. Online activity indicators such as number of fans on Facebook, followers on Twitter, and likes on YouTube have been seen as indicators of a galvanized base, which ultimately contributed to Obama's victory [5]. Research was undertaken to examine Twitter messages regarding the 2009 German federal election and findings indicated that the mere number of messages reflects the election result and even comes close to traditional election polls, concluding that Twitter can be considered a valid indicator of political opinion [1]. Our study is primarily based on the ground-breaking paper [6] whereas we use public political sentiment versus each governmental BREXIT exit strategy and look for causation on the FTSE and the GBP/EUR Fx. [6] study was concerned with predicting stock market trends by measuring the mood of people on Twitter from their mood-based time series. The authors considered the tweet data of all twitter users in 2008 and used the OpinionFinder and Google Profile of Mood States (GPOMS) algorithm to classify public sentiment into six categories, namely, Calm, Alert, Sure, Vital, Kind and Happy. They cross validated the resulting mood time series by comparing its ability to detect the public's response to the Presidential elections and Thanksgiving day in 2008. They performed a causality analysis to investigate the hypothesis that public mood states, as measured by the OpinionFinder and GPOMS mood time series, are predictive of changes in DJIA closing values. [6] further used a Self-Organizing Fuzzy Neural Networks to predict DJIA values using previous values. Their results show an extraordinary accuracy of nearly 87% in predicting market movements in the Dow Jones Industrial Index closing values. [7] researched Public anxiety from Live Journal posts and established that these variations can predict S&P500 values. The study asks whether the Anxiety Index provides useful information for projecting future market prices not already contained in the market itself. [8]

51

established that pessimism using a Naive Bayes algorithm expressed in high-profile *Wall Street Journal* columns had a causal tendencies to Dow returns from 1984 to 1987. [9] looked at investor sentiment in regard to strategic allocations. [10] demonstrated how different financial language in financial news affect stock prices such that a fraction of negative words in firm-specific news stories forecasts low firm earnings.

III. DATA COLLECTION AND ARCHITECTURE

A. Political Background

We have considered the period from December 10th 2018 to February 24th 2019 surrounding the EU Withdrawal Agreement as context for extracting the political sentiment from Twitter which required the full support of Parliament in the House of Commons. This is an important period in the BREXIT negotiations as the Withdrawal agreement required the full support of the House of Parliament to agree on an exit strategy. Numerous strategies were debated such as the *"Norway", "Article 50", "Backstop" and "No Deal"* exit strategies. Numerous attempts to secure agreement failed. What ensued was complete disharmony across the political spectrum resulting in political divides and market volatility. This provided an opportunity to understand the importance of testing for a potential short-term dynamic between FTSE 100 index, GBP/EUR Fx market movements and tweets over the period to help understand if these BREXIT strategy tweets did actually impact the FTSE 100 index and the GBP/EUR Fx.

B. Twitter API

Fig. 1, illustrates the high-level system architecture diagram of our system compromising of five system components. The system components consists of the Twitter API, the python Tweepy library, the python Naïve Bayes textblob library, the data depository i.e. the SQL LITE database and the application of the Granger Analysis using the *Eviews* application. Each component works sequentially. The most common way to access Twitter data is through the Twitter REST API. Using the secure tokens obtained via the Oauth process, this provides authentication and thus allows the user to receive the requested Twitter data.

C. Python – Tweepy Library

The data is received based upon the data requirements received from the user. We utilise the "Tweepy" python library to accept the Twitter data by creating a tweepy.py file [11] Twitter offers several streaming endpoints, each customized to certain use cases. These streams are categorised as follows:

- Public Streams: Streams of the public data flowing through Twitter. Suitable for following specific users or topic or data mining.

- User Streams: Single-user streams, containing roughly all of the data corresponding with a single user's view of Twitter.

- Site Streams: The multi-user version of user streams. Site streams are intended for servers which must connect to Twitter on behalf of many users

For this study we are concerned with BREXIT public streams. We customise our *StreamListener* API from the *Tweepy* library to capture the incoming tweets from the BREXIT #Hashtags contained in Table I.

TABLE I. #HASHTAGS AND SAMPLE TWEETS

#Hashtags	Sample Tweets
#Brexit	RT @IsolatedBrit: To avoid a fascist revolution. That's why we're leaving the EU? #Brexit
#BrexitChaos	No-deal Brexit could put public at risk, warns Met chief #Brexit #BrexitChaos #BrexitCrisis
#BrexitShambles	New Labour Leader desperately needed. Preference would be Yvette Cooper, David Lammy or Chuka #BrexitShambles

We have extracted approximately 3.6 million tweets from the BREXIT public Twitter streams. For each tweet these records provide a tweet identifier, the date-time-seconds of the submission (GMT), location, verified indictor, the text content and a sentiment score derived from the Native Bayes machine learning algorithm which ranges from -1 = negative. 0 = Neutral and 1= positive for each collected tweet. Using an additional python library, we have also extracted the associated capital markets data from Yahoo! Finance for the same period. This includes GBP/EUR Fx rates and the FTSE 100 index. The time series frequency is measured in terms of seconds. Whist the data from Twitter, the FTSE index 100 and the GBP/EUR Fx was collected in terms of seconds, the data for the FTSE 100 index and the pound remained static at the weekend as a result of the markets being closed. To extract our data, we have considered the three key BREXIT hashtags mentioned in Table I. They are as follows: #BREXIT, #BREXITCHAOS and #BREXITSHAMBLES. Both the public and the politicians were tweeting to these #Hashtags. We then data mined the *#Hashtags* collectively using SQL LITE queries to determine the strategies which had the most weight in terms of discussion. The following key BREXIT exit government strategies, "*Norway", "Article 50", "Backstop" and "No Deal"* were identified as being most relevant in terms of count. Other exit strategies did exist such as the Labour party's *"Single Market"* and *"Customs Union"* exit strategy, however, these were not considered as their average weighting was significantly low at the time of this study.

D. Python – Naive Bayes using Textblob

Naive Bayes is a straightforward model for classification. It has been proven to be effective in text categorisation. *Text blob* employs a multinomial Naive Bayes classifier, where the assumption is that each feature is conditional independent to other features given the class. Bayes theorem is illustrated in equation (1)

$$P\left(\frac{C}{T}\right) = \frac{P(C)\ P(T/C)}{P(T)} \tag{1}$$

where *c* is a specific class, in our context either *positive* sentiment or *negative* sentiment, and *t* is a tweet text we want to classify. *P(c)* and *P(t)* is the prior probabilities of a sentiment class *c* and a text *t*. *P(t |c)* is the probability the text appears given this class. The goal is choosing value of *c* to maximize *P(c|t)*. Using a *bag of words* approach, each text *t* can be represented as a vector features $\{w_1, w_2, .., w_n\}$ m, where w_i represent the occurrency of each word w_i in the text *t*, usually weighted. Therefore $P(w_i|c)$ is the probability of the ith feature in text *t* appears given class *c*. In the naïve Bayes approach, each feature w_i is independent from each other. Therefore $P(w_i|c)$ is the probability of the ith feature

in text t appears given class c. In order to classify a text t, we need to compute the maximum likelihood estimation of each one. When making prediction for a new text t, we calculate the log likelihood $\log P(c) + \sum_i \log P(w_i|c)$ of different classes, and take the class with highest log likelihood as prediction

E. Data Depository

The transmission method via the API is then accepted to our remote sever which houses a SQL LITE database to store the retrieved data. The remote server environment consists of an 8 Core Intel R Xeon (R) CPU E5, a 2630 v4 2.20 GHz Intel processor with 16 GB RAM, 200GB memory and a 64-bit Windows 10 Operating System. A number of SQL queries were written to data mine the BREXIT data. To optimise and increase performance data retrieval we applied Index statements to index the critical columns. [12]

F. Granger Analysis

The Granger Analysis was carried out using the EViews 10 application.

G. Architecture
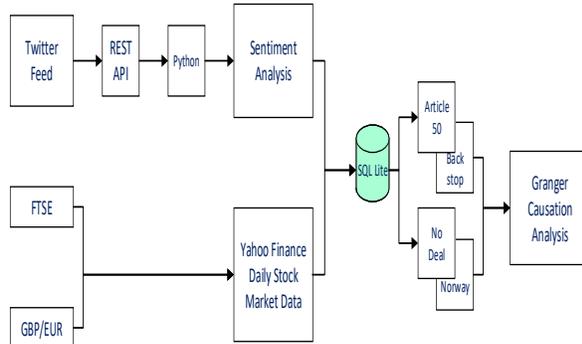
Fig. 1. High level architecture

IV. ECONOMETRIC METHOD

We have established the political BREXIT exit *strategies ("Norway", "Article 50", "Backstop" and "No Deal")* which lead to the disagreement in the House of Parliament concerning the Meaning Vote. We then wanted to understand if the Twitter sentiment from these BREXIT strategies had causal effects with the changes in the GBP/EUR Fx and FTSE 100 Index. We now introduce and apply the econometric metric known as the Granger Causality analysis to the time series from the Twitter BREXIT #*Hashtags* in conjunction with the GBP/EUR Fx and FTSE 100 Index variations from Yahoo! Finance. [13] proposed a time-series data based approach in order to determine causality. In the Granger-sense, χ is a cause of γ if it is useful in forecasting γ[1] In this framework "useful" means that χ is able to increase the accuracy of the prediction of γ with respect to a forecast, considering only past values of γ.

The FTSE time series is denoted as $ftse_t$ this essentially defines the returns in the FTSE 100 index, essentially this is the delta between $ftse_t$ - $ftse_{t-1}$. Similarly we use gbp/eur_t - gbp/eur_{t-1} for sterling/ Euro Fx rate. To facilitate our test as to whether our BREXIT Twitter public and political sentiment analysis for *Norway", "Article 50"," Backstop" and "No Deal"* time series has causal effect on changes in the FTSE 100 Index and GBP/EUR Fx markets. We firstly checked for non-stationary variables using the log difference

illustrated in equation (2) and converted the two financial variables accordingly.

$$\Delta Y_t = \frac{\ln(Y_t) - \ln(Y_{t-1})}{\ln(Y_{t-1})} \qquad (2)$$

We then computed the FTSE 100 index and the GBP/EUR Fx Granger model illustrated in equation (3) and (4) with lag values of *9 and 52* for both Twitter and the FTSE 100 index from the period 10th December 2018 to February 24th 2019 split out per BREXIT strategy. Where X_1, X_2, X_3 and X_4 represent the Twitter sentiment from the respective BREXIT strategies *Norway", "Article 50", "Backstop" and "No Deal"*. This included all tweets pertaining to each BREXIT strategy and the associated FTSE 100 index and GBP/EUR Fx price movements in a time frequency of seconds.

$$ftse_t = \alpha + \sum_{i=1}^{n} \beta_i ftse_{t-p} + \sum_{i=0}^{n} \gamma_i X_{t-q} + \epsilon_t \qquad (3)$$

$$gbp/eur_t = \alpha + \sum_{i=1}^{n} \beta_i gbp/eur_{t-p} + \sum_{i=0}^{n} \gamma_i X_{t-q} + \epsilon_t \quad (4)$$

V. RESULTS

Table II show the results of the Granger Causality tests. ADF tests were implemented for stationarity, the variables checked for cointegration and a VAR done to identify the optimal number of lags according to basic steps in time series analysis. The series were identified as I(1) processes. We can reject the null hypothesis that Twitter political sentiment time series does not have a causal effect on FTSE returns such that β_1 ,β_2.. $\eta \neq 0$ Evidence presented herein confirms a short run relationship relates to sentiment impacting the FTSE 100 index changes in the case of the tweets for X_1, X_2, where X_2 *("Article 50")* exhibits the highest Granger Causality with the FTSE 100 index for lags ranging from 9 to 52 seconds (***p-value < 0.01) and X_1, *("Norway")* exhibits a Granger Causality with the FTSE 100 index for lags ranging from 9 to 52 seconds (* p-value < 0.10) X_3 and X_4 (*"Backstop" and "No Deal"*) do not have any significant causations with the FTSE 100 Index. We fail to reject the null hypothesis that the Twitter political time series does not cause GBP/EUR Fx rate changes such that β_1 ,β_2, β_3, β_4 ,... $\eta \neq 0$. Our evidence shows that X_1, X_2, X_3, and X_4 " do not have any significant causations with GBP/EUR Fx changes.

TABLE II. GRANGER CAUSALITY STATISTICAL SIGNFINANCE (*=90% CL, ***=99% CL)

VARIABLES	NORWAY	ARTICLE 50	BACKSTOP	NO DEAL
FTSE 100 Index	1.6277 (0.092) *	1.7826 (0.0169) ***	0.5090 (0.9836)	0.6078 (0.9835)
GBP EUR Fx	0.90678 (0.5181)	0.4811 (0.9745)	0.4025 (0.9748)	1.1635 (0.1901)

VI. Conclusion and Future Work

In this paper, we conducted a number of tests to establish if Twitter public and political sentiment associated with four BREXIT exit strategies was indicative of FTSE 100 index changes and of GBP/EUR Fx rate changes. The results confirmed that sentiment tracked using a Naive Bayes machine learning algorithm indicative of public and political sentiment for *"Article 50" and Norway"* BREXIT exit strategies for the period 10th December 2018 to the 24th February 2019 was indicative of short term dynamics on the FTSE 100 Index, whilst the *"Backstop" and "No Deal"* exit strategies had no causal effects on the FTSE 100 Index changes. We further established that Twitter public and political sentiment for each of the four BREXIT strategies concerning GBP/EUR Fx changes had no causal relationship. At the time of writing this paper, a new BREXIT strategy was emerging in the form of the Labour party's move for a *"Second Referendum"*. Whilst the data mining picked up on this trend, it did not warrant enough attention for inclusion in this study. However, the Labour Party's leader Jeremy Corbin made a formal announcement on February 28th 2019 in the House of Commons that his party would support a second BREXIT referendum [14]. This would form an important variable for future research if this motion passes a House of Commons vote considering the geopolitical events and strategies of BREXIT may have far wider consequences to other markets, especially the US markets. Additional important factors to consider include the variables that were not used as part thereof this study but part thereof the original dataset. These variables include the GBP/USD Fx, the Dow Jones Industrial Average, the S&P 500, the NASDAQ, Gold and Crude Oil. Causality between BREXIT strategies and US markets could potentially identify short and long term dynamics that have the potential to influence investment decisions. The fact these BREXIT strategies are the source of geopolitical instability in the euro zone thus providing market volatility, the sentiment could also be applied to the commodity markets. With ongoing operations by a number of European, Middle East and Asian countries to move away from the dominate reserve currencies such as the US$, GBP and EUR in an effort to cut their exposure [15]. The sentiment and Granger causation analysis could provide important insights for investment decision making into the rise of safe haven commodities such as gold as a move away from traditional currency reserves in times of geopolitical risk and euro zone uncertainty. These insights could signal the decline of standard reserve currencies and propel investment decisions globally and at a government level. We are currently working on a paper examining the breakdown of the Twitter sentiment per political party with the use of Granger causation analysis to establish if British political parties with their political decision making alone have causal power within the capital markets over BREXIT political tweets from the general public. Further to this the dataset, the data contains the sentiment analysis and capital market activity from the House of Commons politicians captured during the important BREXIT debating sessions. A study could be undertaken to understand if the House of Commons speech and sentiment alone has causal power over public political sentiment. Future research could also seek to explore tweets from individuals in positions of power such as Michel Barnier the European Chief BREXIT Negotiator, Jean-Claude Juncker the President of the European Commission, Donald Tusk President of the European Council and differentiate them from the general public sentiment and the British political parties to see if there is any change in terms of the causation outcome to the capital markets. Additionally, a Dynamic Causality using a frequency domain approach would identify tweets that measure the impact of news on the capital markets. This would prove extremely useful in establishing which piece of news, or indeed if politicians' tweets, moved the markets.

References

[1] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. Association for the Advancement of Artificial Intelligence Conference (AAAI), 2010

[2] Ausserhofer, J., & Maireder, A. (2013). National Politics on Twitter. Information, Communication & Society, 16(3), 291-314.

[3] V. S. Pagolu, K. N. Reddy, G. Panda and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, Paralakhemundi, 2016, pp. 1345-1350

[4] M. Wagner. Obama election ushering in first internet presidency. http://www.informationweek.com/news/government/212000815, 2008

[5] E. Mustafaraj, S. Finn, C. Whitlock, and P. T. Metaxas. Vocal minority versus silent majority: Discovering the opinions of the long tail. International Conference on Social Computing, 2011.

[6] Bollen J, Mao H, Pepe A. Modeling public mood and sentiment: Twitter sentiment and socio-economic phenomenal/In Proceedings of the 5th International AA.Al Conference on Web logs and Social Media (ICWSM 2011), Barcelona, Spain: AA.Al Press. 2011:4S0-453.

[7] Gilbert, E & Karahalios, K. (2010) Widespread worry and the stock market

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

[9] Kenneth L. Fisher & Meir Statman (2000) Investor Sentiment and Stock Returns, Financial Analysts Journal, 56:2, 16-23, DOI: 10.2469/faj.v56.n2.2340

[10] Tetlock, P.C., Saar-Tsechansky, M., and Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. Journal of Finance, 63, pp.1437-1467. http://dx.doi.org/10.1111/j.1540-6261.2008.0136

[11] Tweepy Documentation, http://docs.tweepy.orglen/v3.S.0/

[12] http://www.sqlitetutorial.net/sqlite-index/

[13] Granger C. J., Investigating Causal Relationships by Econometrics Models and Cross Spectral Methods, Econometrica, Vol. 37, 1969, pp. 425-435

[14] https://www.bbc.co.uk/news/uk-politics-47392018

[15] James Rickards 2015 The Death of Money: The Coming Collapse of the International Monetary System, Chapter 2, p55