DS 501: CASE STUDY 4

TWITTER AND THE STOCK MARKET

GROUP 6:
ELIZABETH KARPINSKI
JINGNAN XU
QIAN WANG
XIAOSHUAI LI

Introduction:

Advanced technologies provide the Internet and social media with a range of opportunities to possess incredible influential power. Social media is the easiest and fastest way to transmit and receive information. As a real-time network that connects users to the latest information interesting to them, Twitter is one of the most explosive social media platforms. With millions of users tweeting about their opinions and attitudes towards varies subjects, the aggregate of tweets could be seen as an indicator of collective mood. Based on behavioral economics, people are not rational consumers and individual behaviors are significantly affected by emotions. Therefore, data scientists have been made several attempts to examine Twitter's predictive potential of consumer purchasing by observing users' mood.

One of the most famous data science papers published so far was "Twitter mood predicts the stock market" by Bollen, Mao, and Zeng. Their paper, published in 2011, used sentiment analysis to track the public mood throughout the day, and used those values in order to predict the closing value of the Dow Jones Industrial Average. The premise of this experiment rested on the importance of behavioral economics, a subset in the field of economics, which suggests that human emotion's affect on individual behavior corresponds to shifts in group behavior. This in turn suggests that microcosms of emotional display on Twitter could signify changes in the macrocosm of the mood and climate of the economic world at large. Several other studies had been done similarly to these which showed connections between chat activity, Google searches, and other online activities, and real-world economic events. Bollen, Mao, and Zeng's paper confirmed this trend by establishing a connection between the large-scale Twitter feed contents (through basic textual and sentiment analysis) and the closing value of the Dow Jones Industrial values, at a delay of three to four days. This is very interesting because it raises the question of whether this result could replicated using other measures, and how the time delay would translate across these other metrics. In order to properly address this question, our group decided to analyze Twitter data through sentiment analysis and compare it with instead with the SP500 and Nasdaq index. In order to best evaluate these, we ran four different tests with the data we collected, as described below.

Data Collection:

To build up a sample of data to conduct analysis on, we collected data from both Twitter and the various stock market measures every ten minutes during the stock market hours which are 9:30 a.m. to 4:00 p.m. Eastern Time. The dates and contents of each tweet for this research were obtained for the time period of December 1, 2015 to December 3, 2015.

The data we actually used in our experiment to analyze the Twitter mood was the ratio of positive versus negative words, and the difference between the number of positive and negative words. This was done as a form of sentiment analysis in order to evaluate the sentiment, tweet by tweet, based on the lexical content of the tweet.

In order to evaluate changes in the stock market, we used the Nasdaq and SP500 indices. Thus for the purposes of our data analysis, we had four distinct data sets, falling into two overarching categories: Twitter mood (created from the previous process) and the stock market indices.

Statistical Methods :

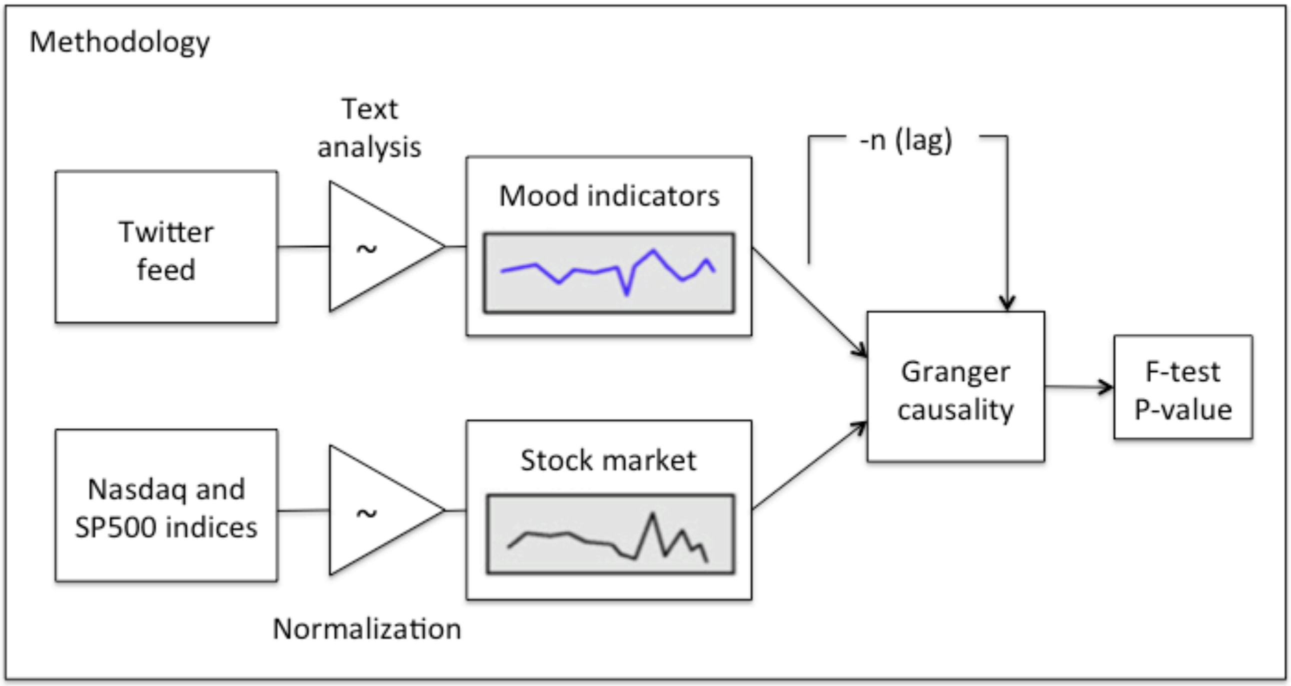

Fig 1: Diagram outlining 3 phases of methodology and corresponding datasets

Fig 1 shows the 3 phases methodology of our study. In order to make linear predictions based on the data we collected, we elected to use Granger Causality testing. The Granger Causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another.

As shown in fig 2, when times series X Granger-causes time series Y, the patterns in X are approximately repeated in Y after some time lags (two examples are indicated with arrows). Thus, past vales of X can be used for the prediction of future values of Y. A time series X is said to Granger-cause Y if it can be shown, usually through a series of F-tests on lagged values of X, that those X values provide statically significant information about future values of Y.

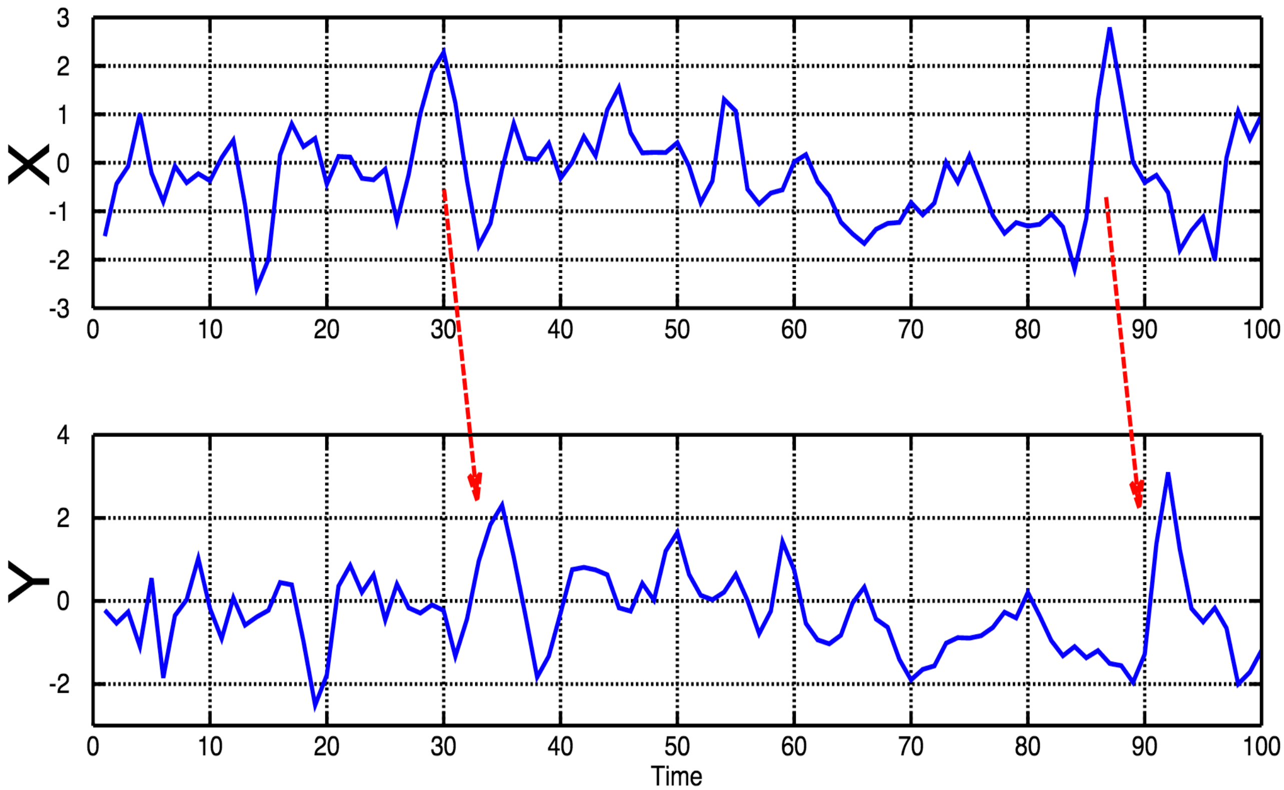

Fig 2: Granger Causality

In this study, our SP500 and Nasdaq times series, denoted Y_t, is defined to reflect real-time changes in stock market. To examine whether Twitter mood time series predicts changes in stock market, we compared the variance explained by two linear models as shown in Eqs. (1) and (2). The first model use only n lagged values of Y_t to predict the future of stock market. While the second model uses the n lagged values of both Y_t and the Twitter mood time series denoted X_t.

$$Y_t = \alpha + \sum_{i=1}^{n} \beta_i Y_{t-i} + \varepsilon_t \qquad (1)$$

$$Y_t = \alpha + \sum_{i=1}^{n} \beta_i Y_{t-i} + \sum_{i=1}^{n} \gamma_i X_{t-i} + \varepsilon_t \qquad (2)$$

We started with a one period lag instead of setting i = 0, because we are not including instantaneous causality in the model. If $\gamma_i = 0$ (for i = 1,2,…,n) then X fails to Granger-cause Y. To decide this, an F-test must be carried out to examine the null hypothesis of non-causality, $H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_n = 0$. From our research, ssr-based F-test is the standard Granger causality test and therefore we used it for our study. The p-value of F-test stands for the possibility of null hypothesis occurrence. The null hypothesis can be rejected if p-value is less than a critical value 0.05. In order to implement these, we used the Python package statsmodels to conduct F-tests.

Data Analysis & Results:

We looked at the collected data through two lenses. For the first part of the study, where we looked at data of December 2, 2015, we conducted F-testing for four data sets, which is positive and negative words ratio as Twitter mood vs. SP500 index, positive and negative words ratio as Twitter mood vs. Nasdaq index, the difference between positive and negative words as Twitter mood vs. SP500 index, and the difference between positive and negative words as Twitter mood vs. Nasdaq index.

(Positives/Negatives) vs. SP500 index					
Number of lags	1	2	3	4	5
F-statistic	3.7735	2.1788	1.2376	1.3980	0.8668
P-value	0.0602	0.1297	0.3140	0.2623	0.5182
(Positives/Negatives) vs. Nasdaq index					
Number of lags	1	2	3	4	5
F-statistic	8.1114	4.4031	1.4935	2.8027	1.7895
P-value	**0.0073****	**0.0205***	0.2371	**0.0465***	0.1548
(Positives - Negatives) vs. SP500 index					
Number of lags	1	2	3	4	5
F-statistic	7.0728	4.6334	2.1526	1.9519	1.3555
P-value	**0.0117***	**0.0171***	0.1152	0.1318	0.2772

(Positives - Negatives) vs. Nasdaq index					
Number of lags	1	2	3	4	5
F-statistic	10.1884	7.6876	2.8832	3.1745	2.4234
P-value	**0.0030****	**0.0019****	0.0527	**0.0299***	0.0664

* P-value < 0.05

** P-value < 0.01

Table 1: F-testing for December 2, 2015

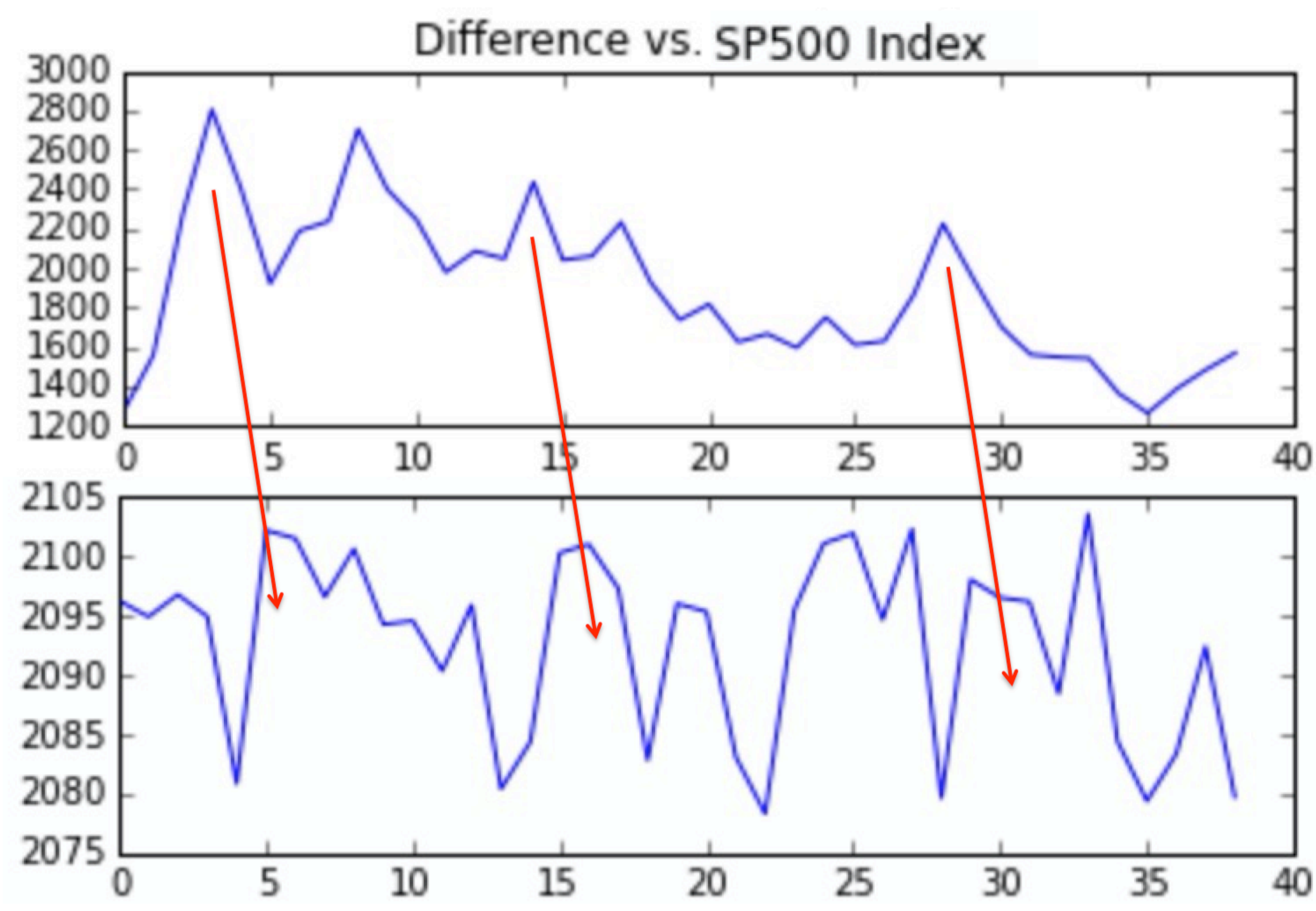

Fig 3: Twitter mood and SP500 index of December 2, 2015

Based on the results of our Granger causality (shown in Table 1), we can reject the null hypothesis that the Twitter mood time series do not predict stock price with a high level of confidence. We observed that the difference between positive and negative words as the indicator of Twitter mood has the highest Granger causality relation with SP500 index and Nasdaq index for lags ranging from 10 to 20 minutes (p-value < 0.05). The ratio as Twitter mood does not have significant causal relations with SP500 index, but does shows Granger causality with changes in Nasdaq index.

To visualize the correlation between Twitter mood difference and the stock market indices in more detail, we plotted both time series. As can be seen in Fig 3, changes in past value of Twitter mood (t-2) predicts a similar rise in SP500 values (t-0). The Twitter mood thus has predictive value with regards to the stock market indices.

Finally, we combined all of the three days' data into one dataset, which is one series is the three days Twitter moods and the other is the three days stock market indices. Using these two series data to do the Granger Causality test and the results is shown in Table 2.

(Positives/Negatives) vs. SP500 index					
Number of lags	1	2	3	4	5
F-statistic	4.1715	1.9049	1.2195	0.4457	0.5761
P-value	**0.0435***	0.1538	0.3064	0.7753	0.7182
(Positives/Negatives) vs. Nasdaq index					
Number of lags	1	2	3	4	5
F-statistic	5.2212	2.4186	1.4228	0.7359	0.7431
P-value	**0.0242***	0.0939	0.2403	0.5696	0.5931
(Positives - Negatives) vs. SP500 index					
Number of lags	1	2	3	4	5
F-statistic	12.9210	6.7096	3.6875	1.6029	1.2570
P-value	**0.0005****	**0.0018****	**0.0143***	0.1793	0.2887

(Positives - Negatives) vs. Nasdaq index					
Number of lags	1	2	3	4	5
F-statistic	11.4039	6.1914	3.6221	2.0130	1.5756
P-value	**0.0010****	**0.0028****	**0.0155***	0.0982	0.1740

* P-value < 0.05

** P-value < 0.01

Table 2: F-testing for three days period

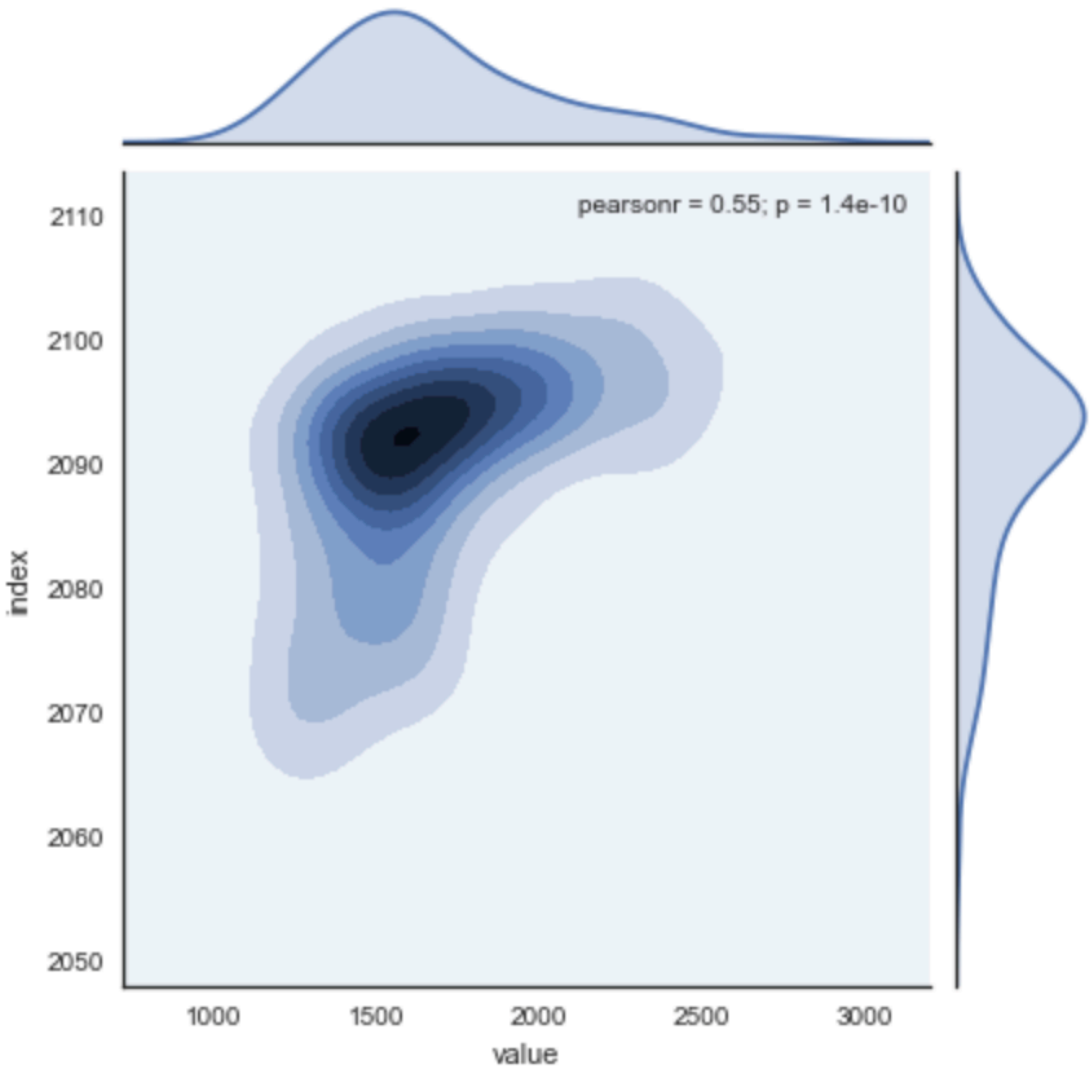

Fig 4: Twitter mood and SP500 index

Based on our results, the difference between positive and negative words as the indicator of Twitter mood has the highest Granger causality relation with SP500 index and Nasdaq index for lags ranging from 10 to 30 minutes (p-value < 0.05). Take Twitter mood difference and Nasdaq index as an example, when the number of lags is 3, the p-value of F-test is only 1.55%, which means in 95% significant level, we can reject the null hypothesis of the two series data do not have causality.

Conclusion:

In this study, we investigated whether public mood as measured from large-scale collection of tweets posted on Twitter is correlated and predictive of stock market indices. Our results show that the changes in public mood collected from Twitter can be tracked from the content of tweets by means of text processing technique. Between two Twitter mood indicators, the difference between positive and negative words is Granger causative of the stock market indices, because changes of this mood indicator match shifts in the SP500 index and Nasdaq index that occur 10-30 minutes later.

References

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

Granger-causality testing within the context of the bivariate analysis of stationary macroeconomic time series. Retrieved from:

http://www.slideshare.net/ThomasReader/granger-causality-testing

Granger causality. Retrieved from:

https://en.wikipedia.org/wiki/Granger_causality