Report on Bachelor Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague

Student:	**Mária Partelová**
Advisor:	**PhDr. Boril Šopov, MSc., LL.M.**
Title of the thesis:	**Presidential rhetoric, sentiment and their relation to stock markets**

OVERALL ASSESSMENT *(provided in English, Czech, or Slovak):*

The thesis investigates impact of the US president's Twitter comments on three major stock indices, i.e. S&P 500, NASDAQ Composite and DJIA. The author focuses on Barrack Obama's and Donald Trump's Twitter feeds from which she extracts sentiment in the form of a numerical value using Natural Language Processing approach. The applied model contains a dictionary that assigns either a positive, negative or neutral sentiment to a tweet depending on the message's lexical content. Next, bivariate VAR models are estimated for sentiment time series and stock indices. The author finds a significant impact of Barrack Obama's tweets on DJIA and S&P 500 stock index movements. No such impact did the author detect for Donald Trump's tweets. The work deals with an interesting topic, uses up-to-date approaches for sentiment analysis and is well structured. Nevertheless, several points for discussion arise:

1. In chapter 2 the author presents different approaches to sentiment analysis, i.e. machine-learning vs. lexicon-based methods, and presents her model of choice – Valence Aware Dictionary for sEntiment Reasoning (VADER) that should be more suitable for analyzing micro-blogging content. However, does this model take into account the tweet's context in addition to analyzing individual words? If not, cannot e.g. sarcasm cause that the model assigns an incorrect sentiment to a tweet?

2. In order to perform the analysis the sentiment data from tweets were aggregated within time brackets in a way to make it possible to capture their impact on closing values of stock indices. This leads me to a question about time zone differences; if a tweet is sent from a different time zone, does it contain time information from that time zone or is the time information from tweets worldwide adjusted to some common time? Tweet distributions presented in Figure 3.3 in chapter 3 might reflect some time zone adjustments since Twitter activity of both presidents seems to be pronounced in very late evening hours and in the case of Donald Trump also in very early morning hours. All in all, is it possible that a discrepancy with respect to time zone differences occurs between Twitter data and stock market closing values for a trading day?

3. To estimate the impact of presidential Twitter feeds on stock market developments Granger causality using bivariate VAR was applied. As observed in the thesis, Granger causality does not provide insights into contemporaneous relationship between the two observed variables; sentiment and stock indices. Therefore, no potential endogeneity between the two variables was investigated nor taken into account in the analysis. One variable that could have a large impact on the results of the analysis is economic news or the news effect as opposed to Twitter effect which also should be investigated. Economic news propagated by the media online or by the means of TV or newspapers have impact on stock markets which could in a way pre-empt the Twitter effect. From several presidential tweet examples presented in the thesis US presidents often comment on domestic and world events. However, these have likely already been covered by media and therefore should be already reflected in stock movements. The sentiment from Twitter could thus only be a further affirmation of good/bad news by the US president. Furthermore, other approaches could be used to investigate Twitter effect on stock markets, e.g. IV model or GMM, to account for the potential endogeneity between the variables.

Report on Bachelor Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague

Student:	**Mária Partelová**
Advisor:	**PhDr. Boril Šopov, MSc., LL.M.**
Title of the thesis:	**Presidential rhetoric, sentiment and their relation to stock markets**

4. Next, sentiment time series over periods of different length for both presidents were used. The series for Barrack Obama cover period of almost 5 years while the sentiments extracted for Donald Trump stretch across only 5 months though those tweets are more numerous. Differences in period length could contribute to the significant outcome for Barrack Obama as opposed to Donald Trump. In addition, the differences in vocabulary or parts of speech favored by the two presidents could also translate into the president's credibility and affect the results. As reported in the thesis Barrack Obama uses more sophisticated vocabulary and his comments contain more nouns while those of Donald Trump's are more simplistic and contain more adjectives. Potentially, the lexicon used in tweets by Barrack Obama might appear more credible and linked to action whereas the tweets by Donald Trump might be perceived as more emotional but passive which does not trigger a stock market reaction.

Overall, the thesis presents an analysis of presidential Twitter feed impact on stock market indices. It extends the current literature by focusing on tweets of influential individuals, i.e. the two US presidents, to study their stock market impact. The author uses Natural Language Processing for sentiment mining from tweets and proceeds by using bivariate VARs to investigate the relationship between the two variables. The econometric part of the analysis could have controlled for other variables affecting the stock market, such as economic news. An alternative model accounting for endogeneity could have been presented as well as a more thorough discussion of the results. Nevertheless, the thesis deals with current and interesting topic, it is logically structured, easy to follow and uses good English. Therefore, I recommend this thesis for defense with the suggested grade "excellent".

***SUMMARY OF POINTS AWARDED** (for details, see below):*

CATEGORY		**POINTS**
Literature	*(max. 20 points)*	17
Methods	*(max. 30 points)*	27
Contribution	*(max. 30 points)*	26
Manuscript Form	*(max. 20 points)*	20
TOTAL POINTS	*(max. 100 points)*	**90**
GRADE	**(1 – 2 – 3 – 4)**	**1**

NAME OF THE REFEREE: Mgr. Diana Žigraiová

DATE OF EVALUATION: 2.6.2017

Referee Signature