Report on **Bachelor Thesis**

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

Student:	**Vladimír Zobal**
Advisor:	**doc. PhDr. Ladislav Krištoufek, Ph.D.**
Title of the thesis:	**Sentiment analysis of social media and its relation to stock market**

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

Please provide your assessment of each of the following four categories. The minimum length of the report is 300 words.

The thesis focuses on a rather challenging topic of using Twitter-based sentiment connected to specific stocks to predict price changes in the given stock. As it turns out, this is much more challenging and problematic that it might seem at the first glance.

Contribution

From my perspective, the biggest contribution of the thesis is very nicely described procedure of sentiment specification. This step-by-step guide of text filtering and connected procedures can help others when trying to perform similar analysis. As for the empirical part of the thesis, the contribution is rather weak. This is mainly due to the fact that the Twitter policy with respect to Tweets availability has become much stricter and Tweets can be downloaded only several days into past. Nevertheless, this is not something completely new and one can easily bypass this by downloading the Tweets on the run but one would need to start downloading much earlier that few weeks before the thesis submission.

Methods

The utilized methods are on the one hand simple – the OLS regression for the final stage of analysis – but on the other quite advanced for the text mining and sentiment setting. For the latter, it seems the author has performed everything accordingly. But for the former, the results reporting is quite poor. Taking e.g. equation (unnumbered) in section 6.1, there are not hats even though these are apparently fitted values, there are no t-statistics, or standard errors, or p-values reported. The author only says these are not significant and then resigns to perform further tests with respect to this. However, if variables are insignificant separately, they can still be significant jointly, therefore it could have paid off to actually run the F-tests, Granger-causality and DM test.

Literature

Work with the literature is probably the strongest part of the thesis as it works with relevant sources and presents them intelligibly. No issues here.

Manuscript form

The structure of the text is fine, there are typos now and there. Mainly tables could have been given more work.

*Overall, the thesis is in a way a missed opportunity as the author clearly is able to measure sentiment based on Tweets but underestimating the Twitter download policies has had its effect. **I believe the thesis can be defended and I suggest grade B (with a total of 65 points).***

Report on **Bachelor Thesis**

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

Student:	Vladimír Zobal
Advisor:	doc. PhDr. Ladislav Krištoufek, Ph.D.
Title of the thesis:	Sentiment analysis of social media and its relation to stock market

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

CATEGORY		POINTS
Contribution	*(max. 30 points)*	15
Methods	*(max. 30 points)*	15
Literature	*(max. 20 points)*	20
Manuscript Form	*(max. 20 points)*	15
TOTAL POINTS	*(max. 100 points)*	**65**
GRADE	**(1 – 2 – 3 – 4)**	**2**

NAME OF THE REFEREE: doc. PhDr. Ladislav Krištoufek, Ph.D.

DATE OF EVALUATION: 5.6.2017

Digitally signed by Ladislav Krištoufek
DN: cn=Ladislav Krištoufek, o=Faculty of
Social Sciences, Charles University in
Prague, ou=Institute of Economic Studies,
email=ladislav.kristoufek@fsv.cuni.cz, c=CZ
Date: 2017.09.07 09:55:36 +02'00'

_____

Referee Signature