

Report on Bachelor / Master Thesis

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

Student:	Petr Rýgr
Advisor:	PhDr. Ladislav Křišťoufek, Ph.D.
Title of the thesis:	Applicability of online sentiment analysis for stock market prediction

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

The author set himself a goal to investigate the sentiment analysis and its applicability in financial trading.

The text is composed from the introduction to the tools used in the sentiment analysis, history of research, an example of possible analysis and trading strategy, and online sentiment analysis.

Notes to the text itself:

Major comments:

- Goal of the thesis (or any research work for that matter) is not to introduce anybody into a subject, rather to tackle a question. This may indeed be only structural issue of the thesis, but signals a rather poor understanding of the purpose of what should be done.
- In the strategy for trading, transaction fees are not low! Approximately five dollars per transaction with closing all up after one year gives you something like \$525 in transaction costs. To cover this up you would have to trade high amounts of stocks. Moreover, you trade based on knowledge of the coefficients from future of your data. The evaluation is misleading at least. (Good self-check is to imagine you are betting high amount of your own money.)
- Total profit of 174% over last ten years is not impressive. It is about 8% annualized. S&P 500 made 6% annually. The difference is probably caused by knowledge of the future through coefficients.
- There are no hypotheses in the thesis, therefore it is easy to say that we have been introduced to the problematics itself. I find this kind of hedging unacceptable.

Minor comments:

- Often the meaning of terms in equations is missing.
- S&P is not a name of index, either S&P 500 or some small-cap indices, etc.
- Checking multicollinearity with no formal test (yet of limited validity) but statement: „Correlation among regressors is however not too high to endanger consistency of OLS estimators.“ is horrendous. Include VIF or some other methods.
- ADF test could have been used, results for other words should be included, but delayed to appendix.
- „Figure 5.3: Lag distribution with confidence intervals“ zero is never rejected, it is therefore hard to make claims about the fact that it is not incorporated into price. Rather the news are not important to the price itself.

More concisely to individual points of the evaluation:

Literature review: Being no specialist on the topic, a quick Google Scholar search reveals a highly relevant papers (more than 1000 citations) that have not been discussed:

- Twitter mood predicts the stock market (Bollen, Mao, Zeng)
- Investor sentiment in the stock market (Baker, Wurgler)
- Giving content to investor sentiment: The role of media in the stock market (Tetlock)

This might be author's intention not to include the discussion of these, however I fail to see how could that be the case if the thesis serves as an introduction to problematics. **I invite the author to explain the choice to the committee.** The bibliography provided at the end of the document is in non-standard format and often unreadable shape. (Look at the famous Fama paper from The Journal of Finance.) Therefore it is hard to evaluate the sources and their relevance.

Methodology: The estimation methodology is done well. However, as mentioned in the comments I have major problems with the trading strategy evaluation. Apart from not using some more advanced

Report on Bachelor / Master Thesis

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

Student:	Petr Rýgr
Advisor:	PhDr. Ladislav Křišťoufek, Ph.D.
Title of the thesis:	Applicability of online sentiment analysis for stock market prediction

matrices (Sharpe ratio, etc.) author should be using cross-validation schemes not plain *ex post* analysis. The exclusion of trading costs is also a bit short-sighted and is fine in theoretical financial work but not defensible when evaluating a trading strategy.

Contribution: The author does not try to provide contribution in a way of asking testable questions at the beginning of the analysis. The thesis might serve as a modest introduction into the problem of sentiment analysis. Despite that I believe that the thesis contributed highly to the skill development of the author and there is considerable work behind the thesis in which I see the reasonable value added.

Manuscript form: The thesis is well-structured. More care could have been paid to presentation of the regression results in their standard table forms. (Including screenshots of tables instead of reproducing the tables usually signifies one of the two following: carelessness, or unimportance of the table itself.)

Unfortunately, I find the thesis to be average, while with little more care and planning the thesis as it stands could be improved significantly. (Providing reasonable bibliography, more readable tables, and shaping the thesis in terms of hypotheses.) Therefore, **I suggest the thesis for defense with the grad 3, i.e. satisfactory.**

I suggest the following questions for defense:

- Discuss how the analysis of trading scheme should be improved to be valid.
- Comment on whether the mentioned papers are relevant to your thesis and why you had not included them in your thesis.

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

CATEGORY	POINTS
<i>Literature</i> (max. 20 points)	5
<i>Methods</i> (max. 30 points)	15
<i>Contribution</i> (max. 30 points)	15
<i>Manuscript Form</i> (max. 20 points)	10
TOTAL POINTS (max. 100 points)	45
GRADE (1 – 2 – 3 – 4)	3

NAME OF THE REFEREE: *Mgr. Tomáš Křehlík, M.A.*

DATE OF EVALUATION: 28. 5. 2015



Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

LITERATURE REVIEW: *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

Strong Average Weak
20 10 0

METHODS: *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

Strong Average Weak
30 15 0

CONTRIBUTION: *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

Strong Average Weak
30 15 0

MANUSCRIPT FORM: *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

Strong Average Weak
20 10 0

Overall grading:

TOTAL POINTS	GRADE		
81 – 100	1	= excellent	= výborně
61 – 80	2	= good	= velmi dobře
41 – 60	3	= satisfactory	= dobře
0 – 40	4	= fail	= nedoporučuji k obhajobě