Report on Bachelor / Master Thesis

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

Student:	**Bc. Tereza Hudcová**
Advisor:	**Mgr. Petr Polák, MSc.**
Title of the thesis:	**Does Language Drive the Crowd? Case of Czech Reward-Based Crowdfunding**

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

Please provide your assessment of each of the following four categories, summary and suggested questions for the discussion. The minimum length of the report is 300 words.

Contribution

In her thesis, the author uses Bayesian Model Averaging and logit regression to find out which factors, and to what degree, influence the probability of reaching the funding goal of crowdfunding campaigns on a Czech platform HitHit. She puts an emphasis on the influence of the project description contents and its characteristics, which turn out not to be significantly correlated with the campaign success. The author herself states that her contribution lies in using Czech data, as opposed to similar already existing studies on US, UK, or Western Europe data, and describing the Czech crowdfunding environment.

Other than growing fast, I don't think there is anything special about the Czech case. No gap in the literature is identified in the literature review part, nor does the author emphasize any reasons why Czechia should be any different than other countries. Therefore, I had hard time finding any valuable contribution really worthy of a master thesis.

Methods

The author uses methods adequate to the research question. I am not familiar with the textual analysis but I was surprised by the low number of words/terms identified by the used method. Also, I would expect to find there some catchy phrases, modern expressions, etc. Do the results of a textual analysis in a similar context in English look the same, or is Czech somehow special in this regard?

The logit regression uses a specification based on variables with high posterior inclusion probabilities from the BMA. I was surprised to find out that a number of coefficients in the logit regression are not statistically significant, even though all of the used variables had very high posterior inclusion probabilities. The author only mentions that it may be due to methodological differences between BMA and logit. It may be so, but I would expect that when facing low robustness, the author would try to provide more robustness checks to make sure the reader can believe the results. But there are none in the thesis. For example, the author could have, at least, used different priors for the BMA analysis.

There is also a number of factors, which are, in my opinion, nowadays much more important for a successful campaign. For example, the attached video – but not as a binary variable „present/not present", but its content, form, length, music background, language used IN the video, etc. Or, maybe even more importantly, the promotion of the campaign on the social networks. From my own experience, social networks are the place, where the chase for campaign backers really happens. Such a task could be even maybe automated to scrape the mentions of the product from Facebook/Twitter and use textual analysis on the posts/tweets/comments there. If not included into the analysis, it could have been at least extensively discussed.

Also, as the author finds that the textual variables do not help to explain the success of the campaign, why not try explaining the percentage funded, for instance? In the current form, the thesis feels too rushed and not really providing enough original content. Instead, the author unnecessarily derives the logit model on 3 pages (pp. 42-44).

Report on Bachelor / Master Thesis

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

Student:	**Bc. Tereza Hudcová**
Advisor:	**Mgr. Petr Polák, MSc.**
Title of the thesis:	**Does Language Drive the Crowd? Case of Czech Reward-Based Crowdfunding**

Literature

The author provides a nice literature review which only has one problem: It doesn't really identify any important gaps in the literature.

Manuscript form

The thesis is very pleasant to read, I have only a couple of minor remarks: First, the section 2.2 on natural language processing seems to be out of place. It breaks the flow of the text. Second, as I already mentioned, I find the parts deriving BMA and logit model unnecessarily detailed. Third, there is a paragraph at the beginning of chapter 6 (p. 47) on „correlation vs. causality". But then, in the rest of that chapter, the text sometimes implies causality. Such as „Results imply, that inclusion of a video has positive and significant effect on the probability of success." (p. 47). I suggest the author next time either provides more theory on why, e.g., including video should increase the probability of success, so that it can be interpreted causally, or is more careful in the interpretation of the results. Such a theoretical context (from the field of psychology and/or behavioral economics) would also nicely complement the analysis and add more interesting content. Fourth, table 5.5 (p. 46) reads „percentage correctly predicted" but doesn't report percentages.

Summary and suggested questions for the discussion during the defense

The thesis is a nice piece of work, but in my opinion rather shallow for a master thesis. On the other hand, what the author does, she does (as far as I can judge) correctly. The thesis fulfills the requirements for a master thesis, I recommend it for defense and suggest grade "C".

- Can you emphasize, what the original contribution of your thesis is?
- Do the results of a textual analysis in a similar context in English look the same, or is Czech somehow special in this regard?
- How are outliers treated in the analysis?
- Why not use also percentage funded as the dependent variable?

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

CATEGORY		POINTS
Contribution	*(max. 30 points)*	15
Methods	*(max. 30 points)*	25
Literature	*(max. 20 points)*	20
Manuscript Form	*(max. 20 points)*	17
TOTAL POINTS	*(max. 100 points)*	**77**
GRADE	**(A – B – C – D – E – F)**	**C**

NAME OF THE REFEREE: PhDr. Jiří Schwarz Ph.D.

DATE OF EVALUATION: 21.5.2020

_____
Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

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._

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._

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

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._

Overall grading:

TOTAL	GRADE
91 – 100	A
81 - 90	B
71 - 80	C
61 – 70	D
51 – 60	E
0 – 50	F