

# Report on Bachelor Thesis

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

<b>Student:</b>	<b>Bc. Vojtěch Nedvěď</b>
<b>Advisor:</b>	<b>doc. PhDr. Jozef Baruník, Ph.D.</b>
<b>Title of the thesis:</b>	<b>Careless Society: Drivers of (Un)Secure Passwords</b>

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

### **Short summary**

This is a high-quality thesis, and I enjoyed reading it. The topic is exciting and original. The author takes advantage of an enormous (in fact publicly available) dataset of historically leaked account names and passwords and provides an exceedingly sophisticated analysis of the two generally observed phenomena of a „poor“ password management endangering the individual-user level cybersecurity: similarity of a username and a respective password and so-called „password reuse,“ i.e., the use of a same/similar password across multiple accounts.

The author considers various potential drivers to explain this behavior: besides a microeconomic variable (gender), variables of the „macroeconomic nature“ (e.g., democracy level or literacy rate of the user’s county of origin, among others) as well as of the „technical nature“ (e.g., password length or the estimated thesaurus sentiment of the given password) are utilized.

### **Contribution**

I can clearly detect several areas of a significant contribution of the work. First, the thesis is well embedded in related scientific literature, and it clearly shifts further the knowledge related to human password management behavior. Second, it is exceptionally elaborated regarding technical details (it has 284 pdf pages from which more than 100 is an Appendix with many additional details and sketches of code) of the analysis. It suggests some innovative statistical/technical solutions to various issues that other researchers can take advantage of in the future (not only wrt password-related research topics but generally in the natural language processing field.)

Third, while delivering econometrically valuable results, which are finally well summarized, the author also suggests various potential ways how his results can be used for practical (policy) implementation. One example for all: as considerable differences are found for different language groups, the author would recommend providers in different countries (e.g., google.cz, google.de, and google.com) specific customization of their default password requirements policy to enhance its security level.

The only issue I value negatively is that the general economic contribution and the most important findings are sometimes overlaid by and hidden in long and repeated technical descriptions and discussions.

### **Methods**

Definitely the strongest part of the thesis. The author combines an advanced toolkit that, to my mind, could potentially „fill up“ several individual high-quality diploma theses. Again one example for all: I believe that section 3.4 (pg. 49-74), which actually deals with only a partial methodology for the creation of one of a set of explanatory variables, the thesauric sentiment of passwords base on Twitter data, could accommodate an individual diploma thesis topic on the sentiment analysis.

Everything is well explained in rich detail, starting with data processing, econometric model construction, specification of all variables and their proxies, the sentiment analysis, the similarity measures for the dependent variable, sampling strategy, and the correlation analysis that leads to a PCA to sort out high correlation among the macroeconomic variables.

Generalized ordered logistic regression is used with 30 cutoffs allowing for different coefficients in each interval. This offers a rich (and sometimes complicated to generalize) structure of estimates for interpretation but allows for the detection of interesting non-monotonous impacts across the scheme of

# Report on Bachelor Thesis

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

<b>Student:</b>	<b>Bc. Vojtěch Nedvěď</b>
<b>Advisor:</b>	<b>doc. PhDr. Jozef Baruník, Ph.D.</b>
<b>Title of the thesis:</b>	<b>Careless Society: Drivers of (Un)Secure Passwords</b>

cutoffs. Finally, after interpreting results, a robustness check is performed via analysis of potential subsampling bias and via BMA to design the most probable models (which provides somewhat poor results compared to „standard“ BMA outputs, see especially Fig 5.6).

## Literature

The literature section is large, rich, and offers a comprehensive selection of related research. It is logically structured to several topical subsections. However, it is designed in a paper-by-paragraph fashion in some sections, which is neither appealing nor effective. I also lack a clear link between the commented findings and the structure of the thesis. In other words, how specifically the presented literature shaped the thesis structure, the specific details of the analysis, and the tools selected. For instance, how specifically is the selection of the explanatory variables (pg. 23) linked to the preceding subsections, and why it differs for the two phenomena. Finally, citations are done properly using a standard style for economic papers.

## Manuscript form

The thesis is written in decent English and typeset in LaTeX. The bibliography section seems complete and well-formatted. Referencing tables and figures is done correctly in the text. The tables are reasonably labeled (with some exceptions: 4.6, 4.9 where an explanation, although seemingly obvious, is missing), and the figures are reasonably self-contained. The graphics are often taken from other sources but correctly cited.

On the other hand, some sections are overfilled by too many technical details with questionable benefits to the actual elaboration of the work (e.g., section 2.6). The section with results is demanding to follow, but the author does his best to aggregate and summarize the most important findings in the Conclusion, which I appreciate.

## Overall evaluation and suggested questions for the discussion during the defense

In my opinion, the assessed thesis undoubtedly surpasses the IES, Faculty of Social Sciences, Charles University master level standards. Thus I can confidently recommend it for the defense and suggest the grade A (excellent).

The results of the Urkund analysis do not indicate significant text similarity with other available sources.

Potential topics for discussion:

- The country of origin of the users is derived from the domain suffixes of the accounts (.cz for the Czech Republic, etc.). Based on this, the macroeconomic variables are associated with each observation, and perhaps this also affects the language analysis. But many people use, e.g., facebook.com or google.com in English from their home countries. Are then these observations coded, in fact, incorrectly in the dataset as users from the U.S.? If yes, what complications might this bring?
- Password reuse is defined here as the use of the same/similar password across multiple accounts. Still, another closely related phenomenon that is repeatedly mentioned in the literature review is the reuse within one account during periodical changes of the password (often forced by the provider). Can this phenomenon also be somehow studied using the dataset in hand?

# Report on Bachelor Thesis

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

Student:	Bc. Vojtěch Nedvěď
Advisor:	doc. PhDr. Jozef Baruník, Ph.D.
Title of the thesis:	Careless Society: Drivers of (Un)Secure Passwords

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

CATEGORY	POINTS
Contribution (max. 30 points)	29
Methods (max. 30 points)	30
Literature (max. 20 points)	18
Manuscript Form (max. 20 points)	16
<b>TOTAL POINTS</b> (max. 100 points)	<b>93</b>
<b>GRADE</b> (A – B – C – D – E – F)	<b>A</b>

**NAME OF THE REFEREE:** Jiří Kukačka

**DATE OF EVALUATION:** 8. 6. 2021

---

*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