Report on Bachelor / Master Thesis

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

Student:	**Tomáš Kovařík**
Advisor:	**Vilém Semerák, Ph.D.**
Title of the thesis:	**Machine learning-based approaches to forecasting international trade**

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

The author attempts to run a contest of several classes of models that could be used to forecast exports. It is not a bad idea, the author also invested admirable effort into learning many new skills, but the resulting contribution is devalued by several weaker features of the paper:

- The description of procedures used to select the final specifications is often relatively brief. The author should have tried to present more details on many of the steps and decisions he took.
- The also seems that the author uses a bit crippled version of gravity model that does not seem to fully exploit the panel nature of the data. The competition between the gravity model and machine learning approaches thus appears a bit unfair.

As such, the thesis can be considered a proof of concept and demonstration of the Mr Kovařík's ability to apply some skills if necessary, but the evidence for or against the superiority of the tested methods cannot be considered sufficiently strong. Theoretically, there could be a quite interesting practical contribution for other students – the author solved computational complexity by using external computational power hired from Google.

Methods

The methods tested by Mr Kovařík can be considered as sound and logical in general (if applied correctly), but his own implementation has some limitations:

1. When the author explains the logic of the methodology (section 3, p. 9-13), he does not try to explain the meaning of most of the variables included in the formulas (e.g. the function #4 on p. 9 and its constraints) or indeed the meaning of other terms that are clearly important for the concepts he is trying to present. Some of these terms (e.g. worst-case time complexity) can be quite familiar to readers with background in computer science, but they may be novel for a more traditional economic audience.
2. Strangely enough, even though the author is aware of the fact that he uses panel data and that there has been a rather substantial debate on the use of exporter/importer/country pair fixed effects in the literature, the regressions he reports in section 6 seemed to be just pooled OLS (or PPML) regressions. The equation which describes the specification is only provided in a general form without proper use of indexes (Equation 10, p. 16). The author thus leaves to the reader's imagination whether random effects or fixed effects or any other panel models were used. If the estimates based on the pooled sample were used, that would be a grave error. Not only there is abundant evidence that pooled OLS are typically rejected in favour of the afore-mentioned alternatives, but we have also theoretical reasons (cited by the author! in his literature review) to believe that a gravity model without the additional effects can be misspecified and the estimates biased. Finally, this approach can disadvantage the gravity model significantly in subsequent comparisons with machine learning based methods. It is true that there may be situations where pooled estimates can be preferred, but the author remains silent on his motives for this decision.

To add something positive about the regressions - the author correctly attributes the seemingly very nice results of his regressions to the large sample he was using (p. 17).

Report on Bachelor / Master Thesis

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

Student:	**Tomáš Kovařík**
Advisor:	**Vilém Semerák, Ph.D.**
Title of the thesis:	**Machine learning-based approaches to forecasting international trade**

As far as random forests, SVMs and neural networks are used, the author provides only a relatively general description of his implementation (p. 19-21). For instance, while he describes the sizes of the training and validation sets, it is not quite clear how they were selected (chronologically? randomly?). It might have been useful to add additional details (code, detailed results) to the appendix. It might have been interesting to learn more about the experiments with Keras (p. 21), tuning of hyperparameters, and on the reasons which convinced him to choose the resulting specification.
On the positive side, when the author was struggling with increasing complexity of the calculations, he opted for a less conventional solution – he outsourced the calculations to Google servers.

The comparison of the models is based only on the RMSE and MAPE, i.e. the author is taking into account neither the fact that there were 19 parameters in the gravity model but 6443 learnable parameters in the neural network described in section 6.5, nor the extreme differences in the time needed for the computations. That can be a bit misleading, as we can easily design a gravity model with additional elements (products, polynomials) which would further improve the RMSE too.
In other words, it would be desirable to find a more versatile criterium that would also penalize for complexity, something similar to e.g. AIC or BIC in traditional econometrics. It also might have been useful to provide e.g. graphical comparison of the differences in forecasts for selected economies.

Literature
The literature review looks seemingly entirely adequate for the undergraduate level. It uses proper style, refers to relevant literature. However, it is rather unbalanced (fairly good treatment of gravity, but rather shallow discussion of machine learning). Especially its second part which discusses neural networks and other machine learning based approaches on trade forecasting seems to have been finished in a hurry - as if the author was willing to cite the first few papers with the right title without making a proper search and analysis of published literature.
The author cites three papers (from 2010, 2014, and 2016), but only two of them are really focused on machine-learning-based forecasts. And both analyze data from a rather specific market (wood and wooden products)[1]. The formulations used in the literature review suggest that these are somehow the first or pathbreaking papers, but that does not seem to be true. There were many older attempts to use machine learning methods in forecasting of international trade data and some of them were published in more relevant and important journals. Indeed, Yildirim et. al. (2014) might have been "the first to use artificial neural networks to forecast non-wood forest products" (p. 7) – but how about other attempts to predict all other trade data? Attempts to use machine learning to predict macroeconomic variables are at least two decades older than the mentioned references. This treatment of neural networks in the literature review (relative to gravity models) is even more disappointing because the author clearly sees the application of machine learning as the main part of his contribution and because he definitely spent quite a lot of time reading about them.

Manuscript form
The text is logically structured, thanks to the use of latex it is also formatted in a way typical for more advanced text. There are relatively few typos and language issues, the thesis also includes quite interesting examples of graphical outputs (e.g. figure 4).
There are some strange quirks though:

1. The author completely omitted page numbering.

[1] The section also mentions an older paper from 1998, but only in relation to VAR and VECM models.

Report on **Bachelor / Master Thesis**

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

Student:	**Tomáš Kovařík**
Advisor:	**Vilém Semerák, Ph.D.**
Title of the thesis:	**Machine learning-based approaches to forecasting international trade**

2. It was also interesting to see that the author reported quartiles as well as the mean and standard deviation even of the variable which captured time in years (Figure 1, p. 15).

Summary and suggested questions for the discussion during the defense
It is not easy to provide an overall evaluation of the thesis. On the one hand, the author clearly mastered quite a few skills which most other students at undergraduate levels do not have (Python, important concepts from machine learning methods, ability to obtain large volumes of data via API access and work with them efficiently, or the use of remote computing capacity at Google). On the other hand, the thesis was clearly finished in a hurry and the author often just vaguely outlines rather than describes or explains what he actually did. Many sections would deserve some additional effort, the implementation of gravity is less than perfect as well.

If I were to evaluate just the author‘s enthusiasm for machine learning and his ingenuity and willingness to master new methods, the grade would be A. If I were to evaluate just the sections of the paper based on gravity and traditional econometrics, the final grade would be E or worse. Had the author decided to postpone the submission of his thesis and to spend some extra time improving its form, the final grade would have been much better.

Questions for the defence:
1. When you were comparing the RMSE and MAPE of the OLS and PPML respectively, did you calculate the indicators from the dependent variable in logs (as used in the OLS) or from the original values without logs (as used in the PPML)? Why can this issue be relevant for your comparisons?
2. Have you found any explanation for the differences between your results and results published by Nummelin and Haenninen?

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

CATEGORY		**POINTS**
Contribution	*(max. 30 points)*	22
Methods	*(max. 30 points)*	26
Literature	*(max. 20 points)*	11
Manuscript Form	*(max. 20 points)*	16
TOTAL POINTS	*(max. 100 points)*	**75**
GRADE	**(A – B – C – D – E – F)**	**C**

NAME OF THE REFEREE: ***Vilém Semerák***

DATE OF EVALUATION: ***January 21st, 2019***

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.*

Strong	*Average*	*Weak*
30	*15*	*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*

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*

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