

# Report on Master Thesis

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

<b>Student:</b>	<b>Bc. Jan Kutman</b>
<b>Advisor:</b>	<b>prof. PhDr. Tomáš Havránek, Ph.D.</b>
<b>Title of the thesis:</b>	<b>GDPNow for the Czech Republic</b>

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

### **Short summary**

This is an excellent master thesis, and I enjoyed reading it. The topic is notably original and of considerable practical macroeconomic importance. The thesis aims to develop a 'running' nowcasting model of the quarterly GDP growth for the Czech Republic. That is a model that would estimate the current GDP growth at any time, based on the most actual available data. I have to appreciate that, although the elaboration of the thesis must have been technically 'dense' and the analysis full of details, the author managed to summarize, present, and communicate the result excellently, with sufficient foresight and yet without neglecting essential information.

### **Contribution**

First, such a real-time tool is missing for the Czech Republic. Thus, it can markedly support pertinent policy-making as official GDP data are typically announced quarterly with a significant delay and often subsequently revised. Second, the author further plans to automate the model, which will be updated daily and publicly available online. After completion, this will be an original undertaking as only a few nowcasting models have been open to the public worldwide. Third, the author presents a well-elaborated analysis that contrasts and compares 3 standard nowcasting models, 3 machine-learning-based models, some trivial benchmark models, and finally deals with model averaging to tackle model uncertainty and get rid of various imperfections of individual models. The author provides clear evidence that his approach beats the standard CNB forecast in normal times. As the last exercise, while the model is for good reasons developed based on data until the Covid pandemic, the entire analysis is repeated with the most current data, which brings some nontrivial new findings.

### **Methods**

The author combines three standard approaches (Dynamic Factor Model, Mixed Data Sampling Model, and a combination of the two approaches) with three popular machine-learning algorithms (Gradient Boosting, Random Forest, and Support Vector Machine). The first three models are well defined in detail while the three ML approaches are well explained briefly as to present them in detail is not contributive for the reader. Also, the cross-validation for the ML models seems appropriately done. Finally, a combination of model averaging and GDP nowcasting is itself scarce in the literature and evidence of its practical usefulness, which this thesis, at least to some extent, also delivers.

The only part that remained not completely clear to me (although the author did a good job in the description and defended his decision based on the current literature) is the actual data manipulation in time. If I understand correctly, the author has decided to follow the stream of literature that ignores multiple data revisions leading to so-called pseudo-real-time data. Still, I could not understand how the technical aspects of the analysis would be affected if the other approach were selected. And what is then the importance of the section '4.1.2 Impact of new data releases'. Also, I am not sure how the online tool that should be automatically updated daily goes together with the monthly dataset (pg. 17) and three vintage datasets constructed for each month (pg. 19).

What is missing is some technical specification of the implementation (what language, what packages, etc.). Although this can be detected in the attachment, I believe it should have been briefly mentioned in the main text. Especially for the still-developing field of machine learning, it can be important what package was used to implement the ML algorithms. Also, the author suggests that some analysis was avoided as it is extremely computationally expensive, but no benchmark of the computational budget is provided. What might be completely infeasible using a PC might be well possible using high-speed computational clusters or online services such as AWS by Amazon. Finally, the analysis of the impact of news related to Fig. 5.7. provides only a single example. A more aggregate analysis of this phenomenon would be appreciated.

# Report on Master Thesis

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

<b>Student:</b>	<b>Bc. Jan Kutman</b>
<b>Advisor:</b>	<b>prof. PhDr. Tomáš Havránek, Ph.D.</b>
<b>Title of the thesis:</b>	<b>GDPNow for the Czech Republic</b>

## Literature

The literature section is rich and offers a comprehensive selection of related research divided into four chapters covering nowcasting in general, related empirical research, the situation in the Czech Republic, and a discussion on model averaging in the nowcasting context. It is well elaborated analytically: various studies are compared and contrasted, advantages and disadvantages of particular methods and approaches are highlighted, and I can see a clear link between the commented findings and the thesis structure. The expected contribution of the model combining exercise is well explained. The discussion is also based on the most recent working papers from 2021.

## Manuscript form

The thesis is written in decent English and typeset in LaTeX. The bibliography section is complete, but there is some formatting issue as info for several working papers seems incomplete (e.g., Adam et al. (2021) should be 'CNB Working Paper Series,' I believe). A statement at the bottom of pg. 2 would require citation. Referencing tables and figures is done correctly in the text. The tables are reasonably labeled, but some of them are not entirely self-contained. In Fig. 5.1. the vertical axis is not marked (especially, where is the zero?). In Fig. 5.2, should factor loadings also be inverted to fit Fig 5.1 multiplied by -1? It is also not clear why RW benchmarks produce such a trivial pattern in RMSE graphs, and I think this is not explained in the text. I could only detect one example of non-academic English ("it's" on pg. 27) and only a handful of typos.

## 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 author might want to re-explain the unclear points related to data manipulation and frequencies (the second paragraph of 'Methods' above).
- At what stage is the preparation of the online tool. What differences will it contain in comparison with this work?

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

CATEGORY	POINTS
<i>Contribution</i> (max. 30 points)	30
<i>Methods</i> (max. 30 points)	28
<i>Literature</i> (max. 20 points)	20
<i>Manuscript Form</i> (max. 20 points)	18
<b>TOTAL POINTS</b> (max. 100 points)	<b>96</b>
<b>GRADE</b> (A – B – C – D – E – F)	<b>A</b>

**NAME OF THE REFEREE: Jiří Kukačka**

**DATE OF EVALUATION: 19. 1. 2022**

---

**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