Report on **Master Thesis**

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

Student:	**Alex Zayat**
Advisor:	**Prof. PhDr. Ladislav Krištoufek, Ph. D**
Title of the thesis:	**Machine Learning in Macroeconomic Nowcasting**

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

Short summary

The thesis addresses the topic of nowcasting GDP, with a particular focus on the case of Argentina. Given the advent of machine learning in economics, the utilisation of these techniques in this specific context, with the incorporation of more comprehensive datasets, is a logical and well-suited for investigation. In the thesis, the author employs a variety of machine learning techniques and compares their performance to that of the benchmark AR(1) model. Additionally, the author is aware of the interpretability challenges associated with 'black-box' modelling and provides results to intepret GDP's explanatory variables. Overall, the thesis presents a range of effective tools for nowcasting, albeit in a rather superficial manner.

Contribution

The primary contribution is the nowcasting of Argentina's GDP and a comparison of several machine learning approaches. The author presents correct arguments in favour of nowcasting and its importance. The author also presents methods that are not fully part of the IES curriculum, thus representing an additional contribution to the author's own knowledge. However, when reading the text, one might also have problems connecting the problem with the approaches and data as they are presented.

Methods

Overall, the author explains many methods that are adequate for nowcasting exercise presented in the thesis. The author tests the significance of results with Diebold-Mariano test.

Unfortunately, thesis lacks a clear definition of the problem, as well as an explicit statement of the nowcasting/forecasting equation and the information set on which it is based. Moreover, the text does not provide any indication of the specific models that were employed in the analysis. It is unclear how the reader is to ascertain which models are being compared. A more detailed account of the data in use and the complete configuration would be beneficial. The fundamental characteristics and graphical representation of the nowcasted variable, GDP, are absent. The most crucial aspect of nowcasting is the inputs and the manner of their incorporation into the process, which is not presented in detail.

Many methods are presented but do not pull a reader towards better understanding of the stated problem. One example for all can be „Surrogate Models" on page 50. Statements such *"unintuitive coefficients"* or *"Furthermore, surrogate models are limited in their applicability and can only be employed under specific conditions."* make one ask why is it presented in the text.

Additionally, I have not found why the author presents the chapter 9 about revisions, since it does not bring much information to the reported estimates.

Literature

The thesis demonstrates a deficiency in the command of the literature. There are several references that are not included in the bibliography. For example, D'Amato et al. (2017) is the source of data for

Report on **Master Thesis**

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

Student:	**Alex Zayat**
Advisor:	**Prof. PhDr. Ladislav Krištoufek, Ph. D**
Title of the thesis:	**Machine Learning in Macroeconomic Nowcasting**

the study, Cybenko, G. (1989) is a reference to neural networks, and also there is missing reference to the Diebold-Mariano test.

Moreover, a considerable amount of text is presented without any reference (Chapter 3) or indication of the source from which the information or argument was derived, particularly in the description of methods.

The list of resources is somewhat limited in light of the significance of nowcasting in economics. Additionally, the list of references contains errors and is incomplete.

Manuscript form

The manuscript appears to be well structured and is formatted in the LaTeX, which gives a good foundation for the thesis. Unfortunately, there are many typos and errors, which diminsh the quality.

- The use of quatation marks and direct citations is incorrect, see first line of the introduction.
- The Czech abstract is in a poor form. Although, I understand it might be hard to do the translation from English or force correct encoding in LaTeX, unfortunately, it is wrong this way.
- The equation on page 12 contains a character that is not valid and affects the meaning of the equation.
- The text contains typos or inconsistencies in emphasizing words or phrases, e.g., *cement* on page 52.
- On page 30, at point d., the equation of MAPE does not contain the absolute value.
- On page 31, *"Results for this exercise are shown in Tables 7.2 to 7.2.!"* there a typo when labelling the table.

Overall evaluation and suggested questions for the discussion during the defense

The thesis addresses the topic of nowcasting the GDP of Argentina in an intriguing manner. However, the presentation lacks rigour and appears to deviate from the core question it seeks to address. Furthermore, the presentation of methods is somewhat general and could be more focused on the specific challenges of nowcasting.

Q1: The first question is directly related to the ML methodologies and the preprocessing techniques. In the event that your models utilise the "full data test", could you kindly elaborate on the regularisation employed during the modelling process? It would be beneficial to understand whether any techniques are employed to prevent overfitting. Is this comparison an appropriate and valid one?

Q2: Please explain the rationale behind the selection of Argentina's GDP. What insights can be gained from this particular case study?

Q3: What are the reasons for selecting a monthly rather than a quarterly GDP forecast? It would be beneficial for the author to provide a detailed explanation of the manner in which the data are incorporated into the models. Please describe the information set (predictors) for nowcasting time t and for forecasting the $t+1$ period. I have a follow-up question regarding the distinction between points *a.* and *b.* on page 30.

Q4: What is a *slack* variable in support vector machine regressions?

Report on **Master Thesis**

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

Student:	**Alex Zayat**
Advisor:	**Prof. PhDr. Ladislav Krištoufek, Ph. D**
Title of the thesis:	**Machine Learning in Macroeconomic Nowcasting**

Overall, the thesis meets the standards required for a master's thesis at IES, Faculty of Social Sciences, Charles University. Therefore, it is recommended for the defence, and a grade of **E** is suggested.

The results of the Turnitin analysis do not indicate (16%) significant text similarity with other available sources.

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

CATEGORY	POINTS
Contribution (max. 30 points)	18
Methods (max. 30 points)	18
Literature (max. 20 points)	7
Manuscript Form (max. 20 points)	12
TOTAL POINTS (max. 100 points)	**55**
GRADE (A – B – C – D – E – F)	**E**

NAME OF THE REFEREE: Luboš Hanus

DATE OF EVALUATION: September 9, 2024

Digitally signed (9/9/2024)
Luboš Hanus

_____
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**