

Opponent's Report on Dissertation Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University
Opletalova 26, 110 00 Praha 1, Czech Republic
Phone: +420 222 112 330, Fax: +420 222 112 304

Author:	Nino Buliskeria
Advisor:	PhDr. Jaromír Baxa Ph.D. (IES)
Title of the Thesis:	Essays on Research Reproducibility in Economics
Type of Defense:	DEFENSE
Date of Pre-Defense	April 17, 2024
Opponent:	Prof. Robert Reed Ph.D. (University of Canterbury)

Address the following questions in your report, please:

- a) Can you recognize an original contribution of the author?
YES
- b) Is the thesis based on relevant references?
YES
- c) Is the thesis defensible at your home institution or another respected institution where you gave lectures?
YES
- d) Do the results of the thesis allow their publication in a respected economic journal?
YES
- e) Are there any additional major comments on what should be improved?
NO
- f) What is your overall assessment of the thesis?
(a) I recommend the thesis for defense without substantial changes

(Note: The report should be at least 2 pages long.)

Chapter 1

1) There are no tests for weak instruments. Should the candidate have found that the instruments were weak, it would have been useful to provide Anderson-Rubin confidence sets to re-evaluate the inference claims of both B&D and the results from this chapter.

Our analysis confirmed that the instruments used are indeed strong. Consequently, we decided not to discuss F-tests. However, the F-tests for our main regressions are presented in Table 1.A1 [now 2A.1]. Additional F-tests related to trend breaks can be found in Figure 1.1 (b) [now 2.1] and Table 1.B1 [now 2B.2]. I have also added further diagnostic tests to the Appendix of the Chapter 2. These were initially shown in Figure 8 (here Figure 2B1.1) of the working paper version of our study. Unfortunately, due to the journal's page limitation, we could only include the main results in the published

version. However, thanks to your comment, I have now included F-tests for varying cut years as well as multiple cut-year exercises. Please refer to Figure 2B1.1 (Figure 8 in the working paper version) and Tables 2B1.1 – 2B1.4 and 2B2.1 – 2B2.3.

REED: I am satisfied with these changes.

2) It would have been helpful if Buliskeria had included their code in the Appendix to this chapter. This would have helped the reader to figure things out when the text wasn't sufficiently clear.

I am afraid the code is too long to be provided as part of the dissertation; however, I have provided the link to the code and data in the footnote on the title page of Chapter 1 [now 2], and also a printout of the main code in Appendix B, at the end of this document.

REED: I am satisfied with these changes.

3) In the interests of open science, Buliskeria might think of making the data and code publicly available so that the results of the chapter are push-button replicable.

The data and code are publicly available at *Journal of Applied Econometrics Data Archive* [<http://qed.econ.queensu.ca/jae/datasets/buliskeria001/>] I also provide the link to the replication package in the footnote on the title page of Chapter 1 [now 2].

REED: This satisfies my comment.

Chapter 2:

Our team discussed the potential application of uncertainty within an IV framework several times, but we have not yet reached any meaningful specification. The main problem is that it is hard to argue that one indicator of uncertainty is exogenous while the others are endogenous because all of those indicators seem to be, to some extent, causing shifts in economic activity and being affected by economic activity at the same time. One of the possibilities would be to define some events that could be considered as exogenous shocks to uncertainty, unrelated to developments of economic activity and unexpected, and use a binary variable set at one of those events as an instrument in the spirit of narrative identification. However, the candidates for such shocks are relatively scarce, perhaps the Brexit referendum and few elections with unexpected outcomes. In addition, shock identification could help with a more precise estimation of the effects of uncertainty shocks on economic activity, but we have not found any meaningful way for the IV approach to compare alternative indices of uncertainty. Therefore, any idea on this matter is more than welcome.

REED: I am satisfied with this answer, however, my IV suggestion was not addressed to the endogeneity issues you mention above. Rather, it came from the perspective that each of the indicators of uncertainty are characterized by measurement error. If both are unbiased measures with measurement error, one can be used as an instrument of the other.

1) *I know Buliskeria is simply replicating the work of Baker et al. (2016), but I didn't understand why the word "today" was used in the keyword searching. Some explanation of this would have been helpful.*

Baker et al. (2016) chose the word "today" as the word that has no relation to economics, politics & uncertainty. Therefore it would serve as a reasonable proxy for all articles. Please also see footnote 6, chapter 3, on page 63: "Factiva's search engine does not allow "blank" searches to obtain all articles published in a given month. Baker et al. (2016) address this issue by searching for the word "today" instead; however, it appears that when we search for the article "the" ("la" in Italian and French), the count of retrieved articles differs (is larger) from the search results using "today." Given the sensitivity of the series to the "generic" word, we find the use of "today" to be problematic and agree with your concern. "

REED: This satisfies my comment.

2) *Figure 2.3 should either have a note or a legend that identifies the black line as the EPU and the red line as the WUI.*

Thank you for noticing. I have added the legend to Figure 2.3 [now Figure 3.3]

REED: This satisfies my comment.

3) *It also would have been interesting to pursue the use of AI (such as ChatGPT) to produce key words. On a lark, I asked ChatGPT for keywords and it gave me the list below.*



ChatGPT

For researching articles on economic uncertainty, consider using a combination of the following keywords:

1. "Economic uncertainty" - The primary topic of interest.
2. "Macroeconomic indicators" - To find discussions on how economic uncertainty affects or is affected by broader economic conditions.
3. "Business cycle volatility" - For insights into how economic uncertainty fluctuates with economic cycles.
4. "Policy uncertainty" - To explore how government policies contribute to economic uncertainty.
5. "Market uncertainty" - For a focus on how financial markets respond to or are impacted by economic uncertainty.
6. "Risk assessment" - Useful for finding articles on the evaluation and management of risks associated with economic uncertainty.
7. "Investor sentiment" - To understand how economic uncertainty influences investor behavior.
8. "Global economic outlook" - For articles that contextualize economic uncertainty within global economic trends.

Combining these keywords with specific countries or regions, time periods, or particular aspects of the economy (like "unemployment," "inflation," or "foreign investment") can help refine your search results.

I agree that this could indeed be an interesting pursuit and thank you for the suggestion. Scraping databases was a very time-consuming endeavor therefore, while we consider his suggestion interesting, I am not sure if it is feasible to dive into collecting data for all countries again. The problem is that FACTIVA does not allow quick text mining analysis, so we had to perform all the searching for most of the newspapers manually.

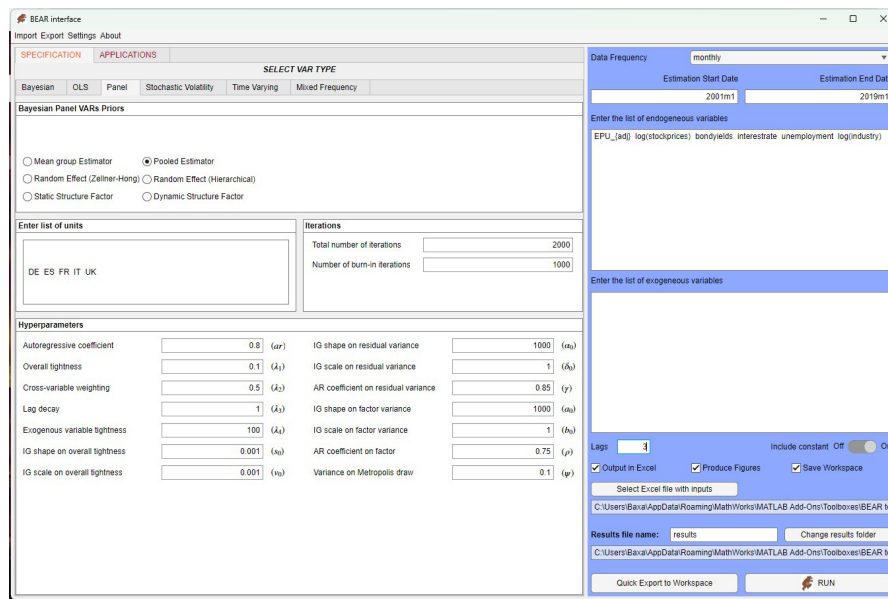
REED: This satisfies my comment.

4)It would have been helpful to provide a table(s) in the appendix reporting some of the key VAR results, along with the code used to produce those results (and the associated impulse response functions).

We have used the Matlab package BEAR with a neat Graphical User Interface that, as can be seen in the figure below, does not require any coding at all:

We have provided following specifications to the bear toolbox:

- Panel VAR: pooled estimator;
- structural decomposition: choleski factorisation;



- units: DE; ES; FR; IT; UK;
 - endogenous variables: EPU_{adj} ; $\log(\text{stockprices})$; bondyields ; interest rate ; unemployment ; $\log(\text{industry})$;
 - exogenous variables: constant;
 - estimation sample: 2001m1-2019m12;
 - sample size (omitting initial conditions): 225;
 - number of lags included in regression: 3;
- hyperparameters:
- autoregressive coefficient (ar): 0.8;
 - overall tightness (lambda1): 0.1;
 - lag decay (lambda3): 1;
 - exogenous variable tightness (lambda4): 100;

However, we will prepare the workspace file for publication in the future to allow for the replicability of our research.

REED: This satisfies my comment.

5) In the interests of open science, Buliskeria might think of making the data and code publicly available so that the results of the chapter are push-button replicable.

Absolutely! Currently, the data set is available on my webpage ninobuliske-ria.com; the replication package is also being prepared for the publication stage and will be available soon.

REED: This satisfies my comment.

CHAPTER THREE:

1) I'm not sure whether this affects the results of the analysis, and I am not 100% certain this is correct, but I think the MAIVE estimator produces a biased estimate of the SE. This arises because it relates average sample size to average SE values. But average SE values are biased downwards because researchers always "cheat" in one direction. They never increase the size of their SEs. They always decrease them. Thus the first stage regression of the MAIVE will underestimate the true value of the SE.

Thank you for your comment. I believe that it can indeed bias the fitted SE values. This concern is valid since a downward measurement error in SE would decrease the strength of the relation between the instrument and SE, thus threatening the validity of the procedure.

I have discussed this point with Tomas Havranek, who with his coauthors of the MAIVE paper plans to develop the extension that can count for the downward bias in the first stage of estimation. On the other hand, I also want to point to the rigorous simulations that show the efficiency of the MAIVE technique in the case of the existence of p-hacking. In their simulations, the authors compare the MAIVE version of conventional methods to their baseline models (simple average, FE/SLW, PET-PEESE, EK, WAAP, Andre & Kasy, p-uniform) in the case of p-hacking. Irsova et al. (2024) show three stylized facts in their simulations. First, spurious precision can plausibly arise in observational research. Second, a small portion of spuriousness creates serious problems for current meta-analysis models. Third, and importantly, the new meta-analysis instrumental variable estimator (MAIVE) substantially limits the resulting bias in meta-analysis. Therefore, while I believe that your concern is very important and should be addressed in the future, MAIVE still performs better than baseline models in case of p-hacking. Although perfecting the MAIVE approach is beyond the scope of this paper, rather relates more to the original paper by Irsova et al. (2024), I greatly appreciate your comments and I plan to think more on this issue in the next revisions and mention this point in the final version of the paper. In my next major revision, I intend to extend the paper by including methods by Andrews & Kasy (2019), Mathur (2022), RoBMA and others that will allow further analysis of the robustness of the results.

REED: This satisfies my comment.

*2) It looks like the β_{FE} and β_{BE} coefficients are estimated in separate regressions. I don't know whether you can do this with IV estimation, but without IV it is possible to estimate both in the same equation use REWB (Random Effects Within Between) estimation. See: Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & Quantity*, 53, 1051-1074. This is what*

Wooldridge calls "correlated random effects" and I believe it also called the Mundlak model.

Thank you for pointing me to this study and method; I will explore it more during my next major revision, where I also plan to approach the research question using other methodologies, including WAAP, Andre & Kasy, p- uniform, and Mathur (2024).

REED: This satisfies my comment.

- 3) Figure 3.11. Please label axes. Thank you for noticing, corrected.
- 4) Page 128, 7 lines from the bottom. "I imply median regression" should be "I apply median regression". Thank you for noticing, corrected.
- 5) Page 108, sentence immediately above Equation (3.3): It should be footnote "7", not "27". Thank you for noticing, corrected.
- 6) In the interests of open science, Buliskeria might think of making the data and code publicly available so that the results of the chapter are push-button replicable.

I intend to build a replication package and make it available in the future.

REED: Consistent with open science principles, I would recommend that the replication package be made available as a condition for completion of the thesis.

Date:	May 17, 2024
Opponent's Signature:	
Opponent's Affiliation:	Prof. Robert Reed Ph.D. (University of Canterbury)