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:	**Petr Polák**
Advisor:	Doc. PhDr. Tomáš Havránek Ph.D.
Title of the Thesis:	Three Essays on Meta-Analysis
Type of Defense:	**DEFENSE**
Date of Pre-Defense	April 22, 2020
Opponent**:**	Geoff Pugh Ph.D.

Address the following questions in your report, please:

a) Can you recognize an original contribution of the author?
b) Is the thesis based on relevant references?
c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?
d) Do the results of the thesis allow their publication in a respected economic journal?
e) Are there any additional major comments on what should be improved?
f) What is your overall assessment of the thesis? (a) I recommend the thesis for defense without substantial changes, (b) the thesis can be defended after revision indicated in my comments, (c) not-defendable in this form.

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

My initial report was quite lengthy. In contrast, this one will be short. I have read the candidate's replies to my various questions and points of criticism and find them to be thoughtful and well argued. These points are then appropriately covered in the revised Introduction. In some cases, we could continue the discussion, but that would reflect only my desire to learn more. I am completely satisfied with the candidate's responses. I have no doubt that these conclude a body of work easily satisfying the criteria for the award of PhD; namely: a) contribution to knowledge; b) mastery over a substantive area of economics (demonstrated and supported by thorough and up-to-date referencing); c) international comparability of the standard of the research; and d) potential for publication in journals of international standing. Consequently, **I recommend the thesis for defense without substantial changes**.

Date:	27-11-2020
Opponent's Signature:	
Opponent's Affiliation:	Geoff Pugh Ph.D. Staffordshire University

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:	**Petr Polák**
Advisor:	Doc. PhDr. Tomáš Havránek Ph.D.
Title of the Thesis:	Three Essays on Meta-Analysis
Type of Defense:	**PRE-DEFENSE**
Opponent**:**	Geoff Pugh Ph.D.

Address the following questions in your report, please:

a) Can you recognize an original contribution of the author?
b) Is the thesis based on relevant references?
c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?
d) Do the results of the thesis allow their publication in a respected economic journal?
e) Are there any additional major comments on what should be improved?
f) What is your overall assessment of the thesis? (a) I recommend the thesis for defense without substantial changes, (b) the thesis can be defended after revision indicated in my comments, (c) not-defendable in this form.

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

Overall, the candidate has demonstrated that he fulfils the traditional criteria for the award of PhD: mastery over at least one substantive field of economic enquiry; and a significant contribution to knowledge. Accordingly, this report does not lavish praise on each and every feature of the research that I like, but raises questions and criticisms that may encourage reflection and subsequent improvements.

There is a notable – and pleasing – progression from Chapter 1 to Chapters 2 and 3. Not only is the quality of the meta-regression analysis higher, but the candidate appears to be more sure footed in the field of international macroeconomics than in the field of productivity analysis. If I were reviewing for a journal, I would recommend major revisions for Chapter 1 but only minor revisions (at most) for Chapters 2 and 3. Chapters 2 and 3 are excellent work and I have benefitted from reading them. I will certainly recommend them to my colleagues and graduate students.

Finally, I would change the title. These are not "three essays on Meta-Analysis": this title leads the reader to expect a focus on critique of existing practice and suggestions for methodological innovation. Rather these are three applications of meta-regression analysis, which – as I have commented – are together worthy of a PhD in economics.

Before offering detailed comments, I will answer the above questions explicitly.

a) Can you recognize an original contribution of the author? Yes
b) Is the thesis based on relevant references? Yes
c) Is the thesis defendable 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? Yes
f) What is your overall assessment of the thesis? (a) I recommend the thesis for defense without substantial changes, (b) the thesis can be defended after revision indicated in my comments, (c) not-defendable in this form.

   Here, I am between (a) and (b). I recommend changes to the Introduction to reflect – or refute – the comments below. However, given that Chapters 1 and 2 have been published, I would not recommend that they be changed. Any issues here can be addressed in a revised Introduction. As for Chapter 3, I leave to the candidate's discretion as to which issues – if any – to address to increase chances of favorable reviewing upon submission to a good journal.

My detailed comments follow.

Introduction

p.4

These two statements seem to be inconsistent.

- Even for model averaging, one can specify priors (ex-ante limits and restrictions), which affect the outcome.
- Model averaging ensures that the final model specification is not handpicked.

"Specify" and "Handpicked" seem, at least to me, to be similar if not synonymous.

This inconsistency informs my general impression that the decision to use BMA was never seriously in doubt. However, I would expect greater effort to be devoted to justifying this choice. I would expect both advantages and disadvantaged to be considered. For example, as far as I know, BMA practice does not incorporate diagnostic testing – e.g. the reporting and assessment of the Ramsey test for unmodelled non-linearity. In this case, averaging takes place over models that include an unknown number that are misspecified and whose results are to an unknown extent biased.

I am not hostile to the use of BMA. Indeed, under the influence of the "Czech" School of MRA, I have used it myself. However, no good purpose is served by using it uncritically.

p.6

"… the third paper does not use weighting by standard error as baseline methodology following the up-to-date works like Havranek et al. (2017)." Having told us the approach you do not use, please mention the "up-to-date" approach that you do use.

There are certain inconsistencies between the chapters (noted below) that should be acknowledged and discussed in the Introduction. I would be happy with an explanation in terms of "learning". As we learn more, we change our ideas and corresponding practices. To inform the reader about the candidate's intellectual journey, especially the reasoning that led to different approaches in successive chapters, should be part of the function of the Introduction.

Chapter 1: The Productivity Paradox

p.18

Like BMA, I have nothing against multi-level modelling as a way of handling within-study heterogeneity. However, I would like some recognition of the disadvantages. In particular, like random effects estimation more generally, multi-level modelling rests on very strong statistical assumptions. For example, do you know the distribution of the "multiple random effects" in your model? I would like to know what attempts have been made to test for whether or not these assumptions hold in the data. And, if such diagnostic testing has not been undertaken, why not? Maybe because nobody else does? That may be an excuse, but it is scarcely good econometric practice.

In addition, I would like to see some assessment of the costs and benefits of multi-level modelling in comparison with the simpler approach – also widely used in MRA – of addressing within-study heterogeneity by reporting cluster-robust standard errors. I am concerned that no recognition is displayed that more complex models typically rest on more demanding assumptions with, correspondingly, more possibilities for model fragility (i.e. lack of robustness). I would have liked some discussion of this issue in relation to the results reported in Tables 2.3 and 2.4. Particularly striking is the apparent precision of the multilevel estimates in comparison to the Clustered OLS estimates. I would be more convinced by this comparison if these results had been preceded by some (brief) discussion of (i) the assumptions underlying the computation of the SEs in each case and (ii) the extent to which these assumptions are supported by diagnostic testing.

If the case for multi-level modelling is as overwhelming as suggested in Chapter 1, why was it not used in Chapter 2? Here (p.68), we read, without further justification, that "a clustering procedure is undertaken to adjust the standard errors for intra-study correlation". See also Chapter 3, p.116.

In general, I am recommending a more balanced approach to justifying the proposed methodology. Rather than treating BMA and multi-level modelling as almost self-evident choices, let us see the terms of debate and some acknowledgement that these methods may have disadvantages as well as advantages. It is always good practice to acknowledge the limitations as well as the benefits of a particular approach to empirical analysis.

p.18

Referring to Eq.2.8, we read that: "The explained variable is the t-statistic and not the estimate of the effect size." Are you sure? The structure of your model develops Equations 2.5 and 2.6 (WLS). In Eq.2.6, the coefficients have the same interpretation as in Eq.2.5: i.e. alpha zero continues to measure publication bias (if any); and beta continues to measure the authentic empirical effect (if any) "beyond" publication bias. In effect, the coefficients switch places but preserve their initial interpretation.

In Chapter 2, this point is acknowledged (p.68): "… the interpretation of the coefficients in equation 3.4 is the same …" So there appears to be some inconsistency between the chapters.

On the same issue, on p.30 we read:

> The results of the explanatory meta-regression analysis (model 2.8) are reported in table 2.8, both showing the effects on productivity and profitability. The interpretation of the results is not straightforward because the dependent variable is the t-statistic.

In Model 2.8, the estimated beta is no longer a direct measure of the authentic empirical effect, because the SE it is interacted with each of the moderator variables. However, an estimate of the authentic empirical effect can be recovered by summing beta – the estimated precision effect – and each of the coefficients on the moderator variables with which it is interacted (where the value of each moderator variable is set at its sample mean). (In effect, this procedure sets the value of the constant and the "K" moderators to zero, thereby yielding an estimate of the authentic effect beyond – or net of – publication bias.) Alternatively, the authentic effect size for studies with particular characteristics – say, a "best practice" specification – by setting the relevant moderators to one (and irrelevant ones to zero). Stanley and Doucouliagos (2012) discuss this in Ch.5; see, in particular, pp. 96-99.

If so (if my reasoning is correct), then it is not the case – as asserted – that "we can only interpret the sign and significance".

Stanley, T. and Doucouliagos, H. (2012). *Meta-Regression Analysis in Economics and Business Research*. London: Routledge.

I note that a variant of the above procedure is adopted in both Chapter 2 (p.78) and in Chapter 3 (p.124 and 127). Why not in Chapter 1?

p.20

In the production function (Equations 2.1 etc), Q is output in physical terms. Yet your variable of interest is "the value of the ICT capital" (p.20). Hence, without any discussion or even acknowledgement, the production function has been recast into value terms. Of course, this is the typical procedure in productivity studies, because data on physical output is rarely obtainable and never widely comparable. However, researchers do need to explain precisely what is being estimated and why. In particular, when the theoretical and empirical forms of models are markedly different, readers should be informed of this and the implications assessed.

Your "important coefficient" presumably measures the "value of the ***output*** elasticity of ICT capital"; not "the elasticity of ICT capital".

p.31, Table 2.5; also p.109

Why do you include both "precision" and (the number of) "Observations" in your models? Both are generally treated as alternative measures/controls in MRA. Did you explore the likely collinearity between these two variables?

pp. 32/33.

"Kohli & Devaraj (2003) concludes that a larger sample size leads to a higher ICT payoff …" Why? It is at least plausible to argue the contrary: i.e. larger sample size should – ceteris paribus – increase the precision of estimates, thereby reducing the incentives that lead to positive section bias and thus lower estimated ICT payoff.

p.35

The headline conclusion is stated as follows: "the underlying effect is identified to be around 0.003 …" This I find genuinely frustrating. What does this mean? Is this the representative output elasticity of ICT expenditure? Or what? Who knows? Be aware that some readers will read only the introduction and the conclusion of your article.

p.36

"Another explanation could be the investment into ICT, which lowers the resulting effect." Do you mean "diminishing returns"?

Chapter 2: The Euro's trade effect

p.66

"… following recent trends in meta-analysis." OK, but your supporting references seem to indicate a local rather than a global phenomenon. The rest of us are catching on slowly!

The explanation of BMA is useful. Also p.120.

p.73; also p.113

"The preferred methodology for meta-regression analysis is the BMA …" This is pure assertion. You have not established this. If true, why did you not use it in Chapter 1? Again, there is some unexplained inconsistency.

"… and to work with a large number of regressors at the same time." In my experience, the jump from days to months of computing time is made rather quickly.

The point I am trying to make here is not that I object to BMA in MRA studies. (I have found it useful.) Rather, I am suggesting a rather more critical stance regarding your chosen methods, particularly as these have changed over time. The Introduction would be the place for such reflection.

The comparison with the established WLS approach is most useful.

p.74

"… and a large number of observations in the dataset result in a smaller euro effect." Surprised? See my comment on pp. 32/33.

p.75

I do not find the procedure outlined in the paragraph on Publication bias to be at all clear.

Chapter 3: How bad are trade wars?

p.104

"Primary studies, from which we collect the estimates, use different methods to estimate trade elasticity with respect to trade costs or exporter 'competitiveness', such as wage, exchange rate, prices and productivity." Does this give a coherent effect size? Or do we have an "apples and oranges" problem? Some discussion of this point would be useful.

Table 4.1

A nice feature of this paper is that the variables coded reflect the large literature on gravity modelling, for example with respect to the treatment of multilateral resistance and the treatment of zero flows in the trade matrix. However, I do have a couple of points to make in this regard.

1. The use of Poisson regression was not introduced primarily to deal with zero flows in the trade matrix. It is, according to its proponents, the preferred approach to estimating a gravity model *in principle*, given the nature of the error term.

2. I was surprised to see no mention of dynamic gravity specifications that capture the role of historical influences on trade patterns. Do none of the studies in this literature specify dynamic models? If so, that might be worth a comment. For references to dynamic gravity specifications more generally in the trade literature, see, for example: Gashi, P, Hisarciklilar, M & Pugh, G 2017, 'Kosovo–EU trade relations: a dynamic panel poisson approach' *Applied Economics*, vol. 49, no. 27, pp. 2642-2654. https://dx.doi.org/10.1080/00036846.2016.1245836

p.116

"The results of the test presented in Table 4.2 come from several specifications of Equation 4.7, and all are linear." How do you know that your linear specification is supported by the data? Standard econometric practice would be to report and interpret the Ramsey test.

p.118

"The IV approach is widely used in meta-analysis as a robustness check since the standard error is estimated jointly with the effect size in the primary studies. An intuitive instrument is based on the number of observations since a greater number of observations should lead to more precise estimates." Elsewhere, you include the number of observations as a moderator variable. This is not consistent with its use as an instrument. Moreover, if this is proposed as an IV estimate, why are readers not informed about the usual tests for (i) instrument exogeneity and (ii) weak instruments? On the evidence of the reported correlation coefficient, it would appear that the instrument is weak. Indeed, this is explicitly acknowledged at the bottom of p.118.

p.128.

You conclude: "Due to the data heterogeneity, conditional estimates (based on the preferred combination of explanatory variables) of trade elasticity have such large confidence intervals that the results are not significantly different from zero. Therefore, we cannot make any conclusions based on these data aggregates."

As I indicate above, I think you have done a good job of capturing the heterogeneity in this literature. As such, therefore, heterogeneity should not be an obstacle to capturing an authentic empirical effect – beyond both publication bias and heterogeneity – should such an effect be present in the data. In this case, your conclusion should be that meta regression of the literature provides no evidence of a substantial trade effect from the trade costs investigated. You seem to shrink from the implications of your own analysis. Better would be to trust your own analysis, point to any limitations in your study, and sketch a research agenda designed to check the robustness of your findings.

Some of the effects that we economists take for granted just do not seem to be there in the data (or are present to only a rather minor extent). It is part of the mission of MRA to uncover such uncomfortable conclusions.

Date:	31-03-2020
Opponent's Signature:	
Opponent's Affiliation:	Geoff Pugh Ph.D. Staffordshire University