RESPONSE TO REFEREES: DOES GROWTH ATTRACT FDI?

We will like to thank the three referees for taking the time to read, review and assess our paper. We have tried to respond to their comments, though in some cases we disagree.

Referee No. 1

The authors use the typical inverse-variance-based weights in their analysis. I believe that weights based on the inverse of the number of estimates reported in each study would be more appropriate in this case, and have five major reasons for this claim. First, although multiple Monte Carlo simulations show that inverse-variance weights bring good results in meta-analysis, these simulations do not consider the case when each study reports several estimates of the effect in question, and moreover if the number of estimates per study varies. When weights are not constant across panels, the interpretation of the weighted results with panel data is unclear, which is why some statistical packages (for example, Stata) do not allow the use of such weights.

Our response: As the reviewer notes, we use the standard inverse-variance weights. Our approach is to control for within-study dependence using cluster-robust. Yet, there is some merit to the reviewer's criticism. It is true that the issue of multiple estimates from the same study has not (yet) been addressed by Monte Carlo simulations. However, it is also true that the number of estimates as weights has also not been addressed by Monte Carlo simulations. It is not obvious that these weights will be more appropriate. Is the inverse number of estimates a 'better' weight than inverse variance? We doubt that this is so. However, this is an issue that is unresolved in the literature and will be unresolved for some time to come. The broader issue of unequal number of observations in a cluster is a major issue in econometrics. We took the view that with a large number of clusters, the unequal number of observations within clusters becomes less of an issue. This would be a particular pressing issue if we had a small number of clusters (studies).

Moreover, we have three good reasons for using inverse variance weights. First, if we do not do so, we know that our standard errors will be biased. Using the inverse of the number of estimates is likely to produce biased standard errors. Second, we can justify inverse variance weights on the basis of statistical power alone. Many studies have very low power. One can argue that they probably should be omitted from a review. However, the approach in meta-analysis is to combine estimates and doing so increasing statistical power. The benefit of inverse variance weights is that they automatically down weight these studies appropriately. Third, there is a nice link between the use of standard errors in testing for publication bias and also for WLS.

We can add a discussion on alternate weights in a revised version of the paper.

Second, some method variables do not vary within studies (or their within-study variance is very limited). With multiple estimates reported per study the introduction of inverse-variance weighting brings artificial variation to the study-level variables, because they suddenly vary within-studies. Again, it is not clear how to interpret such results, and there have been no Monte Carlo simulations that would help us with inference.

Our response: It is unclear to us why estimates will vary within studies if methods do not vary within studies. Estimates vary because specification is altered, or functional form changes, sub-samples are used, or different estimators are employed. If these methods change, estimates change and their precision may also change. The weights we employ accommodate this.

Third, in meta-analysis the reported standard errors are likely to be endogenous to the reported point estimates. Certain method choices (for example, simple OLS versus instrumental variables) influence both the standard errors and the point estimates (see Havranek, 2015). If the influence of the method on the two statistics goes in the same direction, a large coefficient in the funnel asymmetry test may simply reflect this endogeneity instead of any publication or small-sample bias, and vice versa (moreover, as meta-analysis becomes better known in economics, standard errors themselves might become the target of publication selection in order for researchers to increase the weight of their results in meta-analyses). One solution is to use the inverse of the square root of the number of observations as an instrument for the standard error, because this instrument is proportional to the standard error, but not likely to be correlated with method choices. It is unclear how to interpret results of a specification where the employed weights are potentially endogenous to both the response and explanatory variable.

Our response: We agree that this is potentially an issue. We will note this in the revised version of the paper. However, we find that endogeneity does not alter the results: OLS or IV seems to produce similar results in this literature. So, we do not feel that this is an issue for this literature. Indeed, we have employed the inverse square root of the number of observations and find that the results are very similar.

Fourth, inverse-variance weights are highly sensitive to outliers in precision. In most meta-analyses there are a couple of studies that report very small standard errors for no obvious reasons other than idiosyncratic methodology, and very often they also report small point estimates (this issue is connected to the endogeneity problem). The meta-analyst can either omit these studies, which is difficult to justify, winsorize these observations, or include them as they are. The differences between these three approaches increase dramatically when inverse-variance weights are used.

Our response: We agree that this can be an issue. Indeed, we have seen meta-studies with such small standard errors and these do raise eyebrows. We followed Bollen and Jackman (1990) in identifying any observation to be influential if |DFBETA| >1. This procedure did not identify any observation as influential. This can also be seen from the funnel plot. So, we do not see this as an issue for our data.

Fifth, the weights based on the inverse of the number of observations reported in a study give each study the same importance, which in my opinion is more intuitive than to give each study a weight based on the number of estimates it reports (which is what happens when we use other weights). It would strengthen the results of the paper if the authors could produce a robustness check using the weights I suggest (such as in Havranek et al., 2015), and discuss the limitations of inverse-variance-weighting with unbalanced panel data. I would also like to see within estimates (regressions including study dummies).

Our response: In the revised paper we can add this discussion on inverse-variance weights and we can report results using weights suggested by Havranek et al. However, we do believe that we should stick with inverse variance weights as these are more conventional and we do not believe that the case for the alternate weights has been made. We are rather sceptical about the use of study dummies in meta-analysis, as it is not obvious to us that the real issue is differences in estimates within studies. Rather, the more interesting issue is the between study heterogeneity - that is where the real action takes place.

I am also skeptical about the ``general-to-specific'' approach, which involves sequential t-tests, and would prefer the use of Bayesian model averaging (BMA). Sequential t-tests are not statistically valid, because each subsequent test does not take into account that the result is conditional on the previous one. BMA, in contrast, can be thought of as an extension of the typical frequentist practice in which different specifications with various control variables are estimated to evaluate the robustness of results.

Our response: It is unlikely that the Bayesian approach will not have some real issues. All methods do. The advantage of the G-to-S is that it is probably the least objectionable way to deal with the issue of model reduction. In our view, the G-to-S is just as statistically valid as is the Bayesian approach.

I don't fully understand the statement on page 15: ``...relatively little is known about the properties of the Bayesian approach for the MRA of economics data.'' Properties of the BMA estimator are well known (Koop, 2003), and I can't think of a reason why they should be different for meta-data (actually, data in meta-analysis are likely to have more favorable properties than other data sources used in econometrics). In contrast, because most economics meta-analysis have a large number of explanatory variables, BMA is an attractive method for this field, because it helps tackle the obvious model and parameter uncertainty.

Our response: We mean by this that there have not been sufficient number of meta-studies that use this approach from which we can learn from the experience of alternate estimators. And, to our knowledge, there is no study that compares the performance of WLS to Bayesian methods when applied to meta-analysis data.

A common objection to BMA is the claim that the method is atheoretical, throwing in many potential explanatory variables and using statistical techniques to find the most important ones. The problem is that in meta-analysis we always have a large number of explanatory variables that might (or might not) potentially influence the reported point estimates. For some of them our economic intuition is stronger, for some of them weaker; nevertheless, we want to control for all the major aspects of data, methodology, and publication characteristics. The economic theory rarely helps us decide which of the variables we should omit, and the choice between BMA and OLS with sequential t-tests is not connected to this issue. But I understand that many authors might have strong priors against Bayesian methods, and in my experience the practical differences between BMA and sequential t-testing are small (although that doesn't have to hold in general). An alternative is to exclude a number of insignificant variables jointly using an F-test, and avoid any sequential testing. For example, test whether all variables with p-vales above 0.2 are jointly insignificant, and, if so, exclude them from the model. See Havranek and Irsova (2011) for an application of this approach.

Our response: The G-t-S approach results in an MRA with 10 variables (reported in the original submission). Adopting an F-test approach produces essentially the same final specific model. For example, an F-test confirms that we can safely remove any variable that has a p-value of 0.3. Estimating the MRA with this threshold gives us the same 10 variables as the G-t-S strategy reported in the paper plus 6 other variables only one of which is statistically significant (with a t-statistic of 1.94). If we employ sample size instead of variance weights all the original 10 variables remain statistically significant while the new 6 variables are not. So, we are confident that the G-t-S strategy produces a robust and defensible model.

My next comment concerns the control for study quality. The authors do not control for study quality and claim that an objective measure of quality is precision. I disagree: precision is noisy and, when endogeneity is present, precision is often negatively correlated with quality. In this case, FDI is obviously endogenous to economic growth, so the authors of primary studies must use a method that accounts for this problem by employing TSLS or GMM. The use of instruments increases variance of the estimated coefficients, thus decreasing precision, but it is the correct approach nevertheless. One of the things that puzzles me about this paper is that the authors find no systematic differences between studies that ignore endogeneity and those that account for it.

Our response: We naturally respect the reviewer's view on this matter, but we will have to respectfully disagree. It is true that precision is noisy. But, all empirical evidence is noisy. In our view, precision is not noisier than other measures of study quality. It should be noted that ours is not the first meta-analysis to find no substantive difference between say OLS and GMM/IV estimates. The most likely reason for this is that the instruments used are weak and once other research design issues are controlled, there is no real difference between these estimators. Indeed, this is what we would expect. Much of the heterogeneity in reported empirical economics is entirely artificial.

I would like to see whether the results hold when different weights are used, when study fixed effects are included, and when sequential t-tests are replaced by a statistically valid approach.

Our response: We are happy to provide a table with such results in a revised version of the paper.

I would argue for including study-level characteristics that are likely to be correlated with quality aspects not captured by methods choices: the number of Google Scholar citations and RePEc discounted recursive impact factors (available almost for all journals and are discounted, so that it doesn't matter much how long each journal has been listed on RePEc).

Our response: Actually, we did include impact factors as a moderator variable and found this to be statistically insignificant (see Appendix B). This is also not an uncommon finding in the meta-analysis (though some meta-studies do find that this makes a difference). We have not considered citations for the reason we note on footnote 6: "the use of the number of citations might bias meta-averages against newer studies in favor of older ones."

Finally, I believe the authors should discuss more the issue of FDI spillovers, because it is a crucial part of their motivation. There are at least two other meta-analyses of FDI spillovers: Meyer and Sinani (2009) and Havranek and Irsova (2012). My impression from the literature is that there is strong evidence for vertical spillovers, weak evidence for horizontal spillovers, and that most meta-analyses suggest that the benefits from FDI depend on the characteristics of the FDI projects and host and home countries.

Our response: We are happy to do this in a revised version of the paper.

Also, I think that the findings of this meta-analysis are consistent with a positive, but quite small effect of growth on FDI. The sentence ``We show that there is a robust positive correlation between growth and FDI'' in the abstract does not reflect well the findings of the paper (from the funnel plots and Table 1 I would argue that the mean effect is essentially zero).

Our response: We can revise. However, note that there is a clear effect for developing countries and from single country studies.

Referee No. 2

"I have some concerns regarding the contribution of this paper from a policy perspective. … This paper analyses the inverse relationship economic growth-FDI and does not aims to contribute to any actual policy debates.

Our response: That is correct. Ours is not a policy orientated paper.

"At the beginning of the paper, authors should discuss the motivation of this assessment."

Our response: We feel that it is clear what the motivation of our paper is. We state in the introduction: "researchers and policy makers seek to identify the factors that make a host country attractive to foreign investors. One such factor is the host country's economic performance, specifically economic growth. The focus of this paper is to identify and quantify the importance of a host country's economic growth on FDI inflows. Does economic growth attract FDI?" We are not sure how we can express this differently.

Referee No. 3

On the weakness side the paper is blurred in the identification of the strong link between the growth-FDI literature and the so-called "market seeking motivation FDI" (vis-à-vis the other three, efficiency seeking, resource seeking and strategic seeking). The reasons why growth might attract FDI are in fact rooted in the strategy of multinational corporations in looking for expanding markets (e.g. in developing countries) that are growing relatively more than other on the global scale.

Our response: We do actually make this point in the paper. In the sub-section *Economic growth as an FDI attractor*, we present several reasons why foreign investors might prefer to invest in faster growing countries, including cost efficiency of production, the realization of economies of scale and scope in production, and greater opportunities for profits and incentive to invest. We didn't specifically refer to *strategy of multinational corporations*. We are happy to add this line of argument in the revised paper.

"when discussing the growth-FDI relationship results, the paper still cites papers of the FDI-Growth nexus by rendering the punch line much blurred and by weakening the strength of the overall argument. It would be useful the mention the FDI-growth literature at the beginning but then stick to the growth-FDI when the empirical analysis starts."

Our response: Actually, we thought that that is what we did do. We mention the FDI-growth nexus and then proceed to focus on growth-FDI.

The exclusion of sector level studies and unpublished papers is not really well justified, but it should not change the overall message of the empirical analysis.

Our response: On page 9 we wrote: "Second, the study had to focus on macroeconomic relationships. Hence, studies of FDI at the firm level or a specific sectors were excluded." In our view the micro studies are not comparable to the macro studies. We agree that we should discuss why we do not include unpublished papers. We can add this explanation to a revised version of the paper. We also agree with the reviewer that this will not make much difference to the final results.

"there is an enormous "endogeneity" concerns in all this literature (see also the second point on methodology and analysis) and the paper is not using convincing arguments to rule of the possibility that the effect is "purely" driven by endogeneity, as many economists suspect. There is much more work to be done here."

Our response: Potentially this is so. It might be the case that there is endogeneity. But it might also be the case that there is no endogeneity. We do look at this issue directly in the paper. We acknowledge the issue. We test for it specifically and find that controlling for endogeneity does not alter the results. We note on page 23 that: "The statistical insignificance of *Endogeneity* might reflect poor instrumentation strategies, so that endogeneity is not adequately controlled in the primary studies." Beyond the discussion in the text, the tests provided in the paper and this qualification, there is really nothing more that we can say about this matter. Perhaps there is no endogeneity.

Overall the analysis is correct (main strength), but it is highly misleading and not transparent, i.e. potentially driven by statistical artifact (main weakness).

Our response: We don't understand how the analysis can be both correct and misleading. If our methodology and inferences are correct, then the paper is not misleading. The results of

primary studies (the individual studies that serve as the data for our meta-analysis) are vulnerable to statistical artifact. That is why we apply meta-analysis to correct publication bias (if it is present) and to identify what drives differences in results. We do not understand in what way our work is misleading and not transparent. Please clarify this so that we can respond accordingly.

The last row of table 1 should be presented as a separate table with all the summary stats (number of obs., of studies etc.) and should not be confined to a single line in whole table. For example the FAT-PET test is not significant in columns two and four and this casts doubt on the validity of the overall conclusion that there is a positive (and significant) relationship between Growth and FDI when controlling for publication bias. Even if the PEESE (columns 3 and 6) is marginally significant, there is an indication (Stanley and Doucouliagos 2013, Research Synthesis) to rely on the latter only if the FAT-PET does reject the null (B=0), which is not the case here.

Our response: We can revise and present as a separate table and include sample size etc. However, this will of course not alter the results. We presented only the basic results for the sake of brevity and also because such results are less meaningful than the multiple MRA results that are presented in Table 3. So, we don't think that a separate table with these results will add anything to the paper. We present the PEESE results in Panel B purely for the sake of robustness. We can remove these of course, but that makes the case for a separate table even weaker.

Furthermore, the fact that there are not enough "single countries studies" controlling for endogeneity (bottom of page 16) is even more problematic, because it could indicate that the results are actually driven by very badly estimated single countries studies empirical analysis. In other words, it would be useful to show the results of cross-countries studies only and especially cross countries studies controlling for endogeneity. Table 1 is mixing two very much different type of regressions.

Our response: Table 1 is presenting an analysis of the entire research base. This is essential. One can then proceed to looking at sub-samples, but it is critical in meta-analysis to report the basic results for all estimates included in the database. So, we really strongly feel that it is critical to report table 1. We can add a new table for cross-country studies in a revised paper, though we do not see the need for this.

Results in table 2 cannot lead to the conclusion (end page 18) that there is a much larger role for economic growth in attracting FDI because time series analysis are fundamentally biased by endogeneity problems. These single studies paper cannot be mixed with the cross-countries analyses to start with (table 1 should have covered only the latter) because they use fundamentally different econometric specification that are completely unreliable on the endogeneity side (single countries studies can use granger causality tests but they are prone to major critiques anyway). Table 3 tries to partially address this point by including the singelcontry moderator variable, but this is not anyway enough to convince the reader that cross-countries and single countries should not me mixed together in the first place. For example in single countries studies FE should also be included.

Our response: All studies are potentially biased by endogeneity, though the evidence from this literature suggests that this is not the case. Again, we disagree that Table 1 should exclude single country studies. Actually, in our view, the single country studies might actually be *more* informative. Pooling several countries together might disguise the underlying heterogeneity. Single country studies are free of between country heterogeneity. Arguably, Table 2 is the more informative table. In the ideal world, we would have numerous estimates for each *individual* country that we can then apply meta-analysis. We don't understand how it will be possible to include FE in a single country study.

It would be useful to split the sample also for not developing and show a separate table to see the results separately. Alternatively, it is also possible to add a dummy in the whole sample and see whether it is significantly different from the constant (implicitly the estimate of the not-developing)

Our response: We can report such results in a revised paper. However, we are not sure that this will be informative.

Overall this is a very interesting and well-structured analysis. Unfortunately, the results are partially convincing, due to a lack in transparency when explaining the endogeneity as well as singlecontry estimates role within the overall effect. The latter can be much more muted than emphasized in the paper.

Our response: We feel that there is an important distinction to be made between single and cross-country studies and that this is one important finding of the paper. Hence, we would rather not mute this distinction.