ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Hong, Sanghyun

Article

Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work? A replication study of Alinaghi & Reed (Research Synthesis Methods, 2018)

International Journal for Re-Views in Empirical Economics (IREE)

Suggested Citation: Hong, Sanghyun (2019) : Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work? A replication study of Alinaghi & Reed (Research Synthesis Methods, 2018), International Journal for Re-Views in Empirical Economics (IREE), ISSN 2566-8269, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg, Vol. 3, Iss. 2019-4, pp. 1-22, https://doi.org/10.18718/81781.13

This Version is available at:
https://hdl.handle.net/10419/206820

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

https://creativecommons.org/licenses/by/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work?

A replication study of Alinaghi & Reed (*Research Synthesis Methods*, 2018)

Sanghyun Hong*

International Journal for Re-Views in Empirical Economics, Volume 3, 2019-4, DOI: *10.18718/81781.13*

JEL: B41, C15, C18
Keywords: Meta-Analysis, Publication Bias, Funnel Asymmetry Test (FAT), Precision Effect Estimate with Standard Error (PEESE), Monte Carlo Simulations, Replication Study

Data Availability: The original programming by Alinaghi & Reed can be downloaded at Harvard's *Dataverse* (DOI: 10.7910/DVN/4IOLOP). The R-code to reproduce the results of this replication-can be downloaded at IREE's data archive (DOI: 10.15456/iree.2018280.233725).

Please Cite As: Hong, Sanghyun (2019). Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work? A replication study of Alinaghi & Reed (Research Synthesis Methods, 2018). *International Journal for Re-Views in Empirical Economics*, Vol 3(2019-4). DOI: 10.18718/81781.13

Abstract

A meta-analysis is a tool for aggregating estimates of a similar "effect" across many studies. Publication bias is the phenomenon where literature is sample selected in favor of studies having statistically significant results and/or having estimates that satisfy pre-conceived expectations. A popular procedure used for conducting meta-analyses in the presence of publication bias is the FAT-PET-PEESE (FPP) procedure. In a recent paper published in *Research Synthesis Methods*, Alinaghi and Reed (2018), utilizing Monte Carlo simulations, report that the FPP procedure does not work well when used in "realistic" data environments where true effects differ both across and within studies. AR's findings are important because the FPP approach is dominant in the economics meta-analysis literature. I replicate their results and discover two mistakes, which I subsequently correct. The first mistake is found in a descriptive statistics table, misrepresenting the overview of simulated dataset. The second is associated with the fixed effect estimation, generating erroneous estimated effects and Type I error. Further, I extend their analysis by making their simulation environment even

*University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand. E-mail: sanghyun.hong@pg.canterbury.ac.nz
Acknowledgements: I gratefully thank Professor W. Robert Reed for his invaluable support.

Received October 10, 2018; Revised February 20, 2019; Accepted March 7, 2019; Published July 16, 2019.
©Author(s) 2019. Licensed under the Creative Common License - Attribution 4.0 International (CC BY 4.0).

more realistic. Despite producing somewhat different results, my replications generally confirm AR's conclusions about the unreliability of the FPP procedure in realistic data environments.

1 Introduction

Meta-analysis is a tool for aggregating estimates of a similar "effect" across many studies, such as the effect of an increase in the minimum wage on unemployment, or the value of a statistical life (Glass, 1976). Publication bias is the phenomenon where literature is sample selected in favor of studies having statistically significant results and/or having estimates that satisfy pre-conceived expectations (Rothstein et al., 2006). The problem of publication bias in the economics literature is widely recognized (Christensen and Miguel, 2016; Andrews and Kasy, 2017). A popular procedure used for meta-analyses in the presence of publication bias is the FAT-PET-PEESE (FPP) procedure (Stanley and Doucouliagos, 2012).

In a recent paper published in *Research Synthesis Methods*, Alinaghi and Reed (2018, henceforth AR), conduct a series of Monte Carlo simulations to investigate the performance of the FPP procedure. They find that the FPP procedure works well in a simulated "fixed effects"[1] environment where each study only has one estimate, and all estimates derive from a single, population value. However, it becomes unreliable when tested in "random effects" environments, where studies have only one estimate, but the underlying population values differ across studies; and "panel random effects" environments, where studies have multiple estimates and population values differ both between and within studies. AR's findings are important because the FPP approach is dominant in the economics meta-analysis literature.

I recently attempted to reproduce AR's results, some of which are based on Stata programming code, by re-programming their simulations in the programming language R. In doing so, I uncovered two mistakes. This study reports the results of re-running their simulations with the corrected code. I then extend their analysis by modifying their simulation design to be even more realistic. Despite producing somewhat different results, my overall findings generally confirm AR's conclusions about the unreliability of the FPP procedure in realistic data environments.

This study proceeds as follows. Section 2 identifies two mistakes in AR's simulation codes and presents the results of correcting those mistakes. In Section 3, I make their simulations more realistic in the "panel random effects" case by allowing individual studies to have differing numbers of estimates. In contrast, AR's simulations restricted studies to each have 10 estimates per study. Section 4 summarizes my main findings and concludes the study.

While not reported, I confirm that I was able to exactly reproduce AR's results using their code. Together, this, along with the replication analyses in Sections 2 and 3, respectively correspond to "Reproduction", "Repetition" and "Extension" in Reed's (2017) taxonomy of replications.[2] All the

[1] In the meta-analysis literature, the terms "fixed effects" and "random effects" have very different meanings than in the panel data econometrics literature. In this study, "fixed effects", "random effects" and "panel random effects" will always refer to their meta-analysis meaning (as described in the text), unless I specifically indicate otherwise.

[2] Reed (2017) defines a "Reproduction" as using the same measurement and/or analysis on the same dataset as the original study. A "Repetition" is similar in that it uses the same measurement and/or analysis but applied to a different sample drawn from the same population. Any differences between the original sample and the replication sample is solely

programs to reproduce my results are downloadable from the journal's data archive.[3]

2 Repetition

AR posted all their programming code at *Dataverse*, including the Stata do files examined in this study.[4] I confirm that their do-files exactly reproduced the tables in their paper. However, during the process of re-programming their Monte Carlo experiments in R, I found two coding mistakes, which I subsequently corrected.

2.1 Mistake #1

AR create artificial meta-analysis datasets and then simulate sample selection by identifying estimated effects that do not pass their selection filters. AR then calculate the descriptive statistics of the resulting simulated datasets. Due to the way STATA codes missing values, AR report inflated "Percent significant" values.[5]

I provide corrected "Percent significant" values for the affected tables in Tables 1 and 2. These replicate the bottom two panels of their Tables 2 and 3, respectively. The incorrect values that AR report are in parentheses, in red, italicized and bold-faced. The correct values are directly above them and are in black type and bold-faced. For example, in Table 1, AR report a median "Percent significant" value of 0.93 for a representative meta-analysis sample created from a "Random Effects" data generating process (DGP) with a mean overall effect size (α) of 1.0, after sample selection for insignificance. In contrast, the correct median "Percent significant" value is 0.77. Likewise, for a representative meta-analysis sample created from the same underlying DGP but where publishing discriminates against negatively signed estimates, AR report a median "Percent significant" value of 0.49. The correct value is 0.31. Similar examples of incorrect, inflated "Percent significant" values can be seen in Table 2, where the underlying DGP is "Panel Random Effects" with $\alpha = 1$.

I provide the respective programming code, both AR's incorrect and my correct code, in the Appendix. It should be noted, however, that while the results change substantially as a result of fixing this mistake, the "Percent significant" numbers do not affect any of AR's conclusions, as they only show up in sample statistics describing the simulated data.

due to sampling error. The analogy here is that mistakes in coding are like "sampling errors", resulting in different data despite drawing from the same "population". An "Extension" is defined as a replication where the sample data are drawn from a related, but different, population. The analogy here is that the population of studies having different numbers of estimates per study is related, but different, than the population of studies all having 10 estimates per study.

[3] My programs can be found here: DOI 10.15456/iree.2018280.233725.

[4] Their programs can be found here: DOI 10.7910/DVN/4IOLOP.

[5] When AR impose the selection filters, they replacing the effect t-value with a missing value, indicated by "." in the statistical software package Stata. AR then calculate the percent of estimated effects that are statistically significant by counting the number of estimated effects that have t-statistics greater than 2. This causes a problem because Stata internally assigns a (very) large number to missing values (presumably to aid in sorting). As a result, AR include the "sample selected", missing values in their count of estimated effects with t-values larger than 2. This inflates their reported "Percent significant" values.

Table 1: Replication of AR's Table 2 — Sample Characteristics of the Simulated Data (Random Effects/$\alpha = 1$)

Variable	Median	Minimum	P5%	P95%	Maximum
Sample after selection against insignificance (33.1 percent of estimates)					
Estimated effect	1.81	−7.52	−2.04	5.64	9.45
t-statistic	2.54	−13.22	−2.31	12.40	42.72
Percent significant	**0.77**	**0.70**	**0.74**	**0.81**	**0.85**
	(0.93)	(0.89)	(0.91)	(0.94)	(0.95)
I-squared	0.94	0.87	0.91	0.96	0.98
Sample after selection against negative estimates (74.6 percent of estimates)					
Estimated effect	1.55	−5.04	0.05	4.74	9.45
t-statistic	1.28	−4.91	0.04	7.29	42.72
Percent significant	**0.31**	**0.26**	**0.29**	**0.34**	**0.38**
	(0.49)	(0.44)	(0.46)	(0.51)	(0.53)
I-squared	0.82	0.58	0.74	0.88	0.92

Note: Values in the table are constructed by simulating 1000 meta-analysis studies given the respective conditions, and then averaging the results on the respective dimensions (e.g. median value, 5% quantile value, etc.). "Percent significant" identifies the average percent of estimates that are significant at the 5 percent level. The red numbers in parentheses are taken from AR's Table 2. The black, boldfaced numbers directly above them are the corrected values. "I-squared" measures the extent of effect heterogeneity (Higgins and Thompson, 2002).

Table 2: Replication of AR's Table 3 — Sample Characteristics of the Simulated Data (Panel Random Effects/$\alpha = 1$)

Variable	Median	Minimum	P5%	P95%	Maximum
Sample after selection against insignificance (21.9 percent of estimates)					
Estimated effect	2.39	−5.29	−2.97	5.97	8.89
t-statistic	3.68	−18.41	−7.63	16.55	33.47
Percent significant	**0.91**	**0.81**	**0.86**	**0.95**	**0.99**
	(0.98)	(0.95)	(0.97)	(0.99)	(1.00)
I-squared	0.98	0.80	0.94	0.99	1.00
*Sample after selection against negative estimates (**56.9** (80.5) percent of estimates)*					
Estimated effect	2.21	−5.39	−0.83	6.18	10.91
t-statistic	1.73	−2.86	−0.50	10.05	33.48
Percent significant	**0.44**	**0.28**	**0.37**	**0.52**	**0.61**
	(0.68)	(0.57)	(0.62)	(0.74)	(0.80)
I-squared	0.84	0.48	0.69	0.94	0.98

Note: Values in the table are constructed by simulating 1000 meta-analysis studies given the respective conditions, and then averaging the results on the respective dimensions (e.g. median value, 5% quantile value, etc.). "Percent significant" identifies the average percent of estimates that are significant at the 5 percent level. The red numbers in parentheses are taken from AR's Table 3. The black, boldfaced numbers directly above them are the corrected values. "I-squared" measures the extent of effect heterogeneity (Higgins and Thompson, 2002).

2.2 Mistake #2

AR commit a more serious error in their implementation of the FAT-PET-PEESE (FPP) procedure. They estimate the underlying regression model using panel regression, fixed effects/study dummy variables.[6] The use of panel fixed effects/study dummy variables in their regression distorts the interpretation of the constant term, which in turn affects results of two tests common to meta-analyses: the Funnel Asymmetry Test (FAT) for the existence of publication selection bias, and the Precision Effect Test (PET) for the existence of a non-zero, mean overall effect.[7] In particular, the constant term estimates the mean value of the true effect for the omitted study. However, in the

[6]The *Fixed Effect* model in the meta-analysis literature is very different from the fixed effects model in the panel data literature. In the panel data literature, the term 'fixed effects' refers to a unit-specific, constant effect. However, the fixed effect model in the meta-analysis literature assumes that there is one true effect size and that all the differences in observed effects are due to sampling error. Furthermore, the parameter of interest in meta-regression analysis is the intercept which captures a mean overall (weighted) effect. Therefore, the term 'common-effect' model would be a more descriptive term for the fixed effect model in the meta-analysis.

[7]See Equation (1) in AR and the associated discussion for further detail about these tests.

Panel Random Effects environment, each study has a unique mean true effect, different from the mean overall (across all studies) true effect. AR confuse the estimate of the study-specific mean effect for the mean overall true effect.

Table 3 reports FAT and PET rejection rates for different values of mean true effects (α), different types of DGPs (Fixed Effects, Random Effects, and Panel Random Effects), and different types of publication selection bias (against statistical significance and against wrongly-signed estimates). It replicates Table 4 in AR.

As above, AR's incorrect values are in parentheses, in red, italicized and bold-faced. The correct values are immediately to their left in black and bold-faced. (The highlighting of cells in red and grey is explained below). In many cases, the values are similar, but there are instances where they differ substantially. For example, when the data environment is characterized by Panel Random Effects, and publication selection bias discriminates against statistical insignificance, and $\alpha = 0$, AR report a FAT rejection rate of 0.56. The correct value is 0.63. They report a PET rejection rate of 0.31. The correct value is 0.65. The interpretation of these numbers follows.

For the FAT, rejection of the null indicates the existence of publication selection bias. In Table 3, the only scenarios where there is no publication selection bias are when $\alpha = 0$ and there is selection bias against statistical insignificance.[8] Publication bias exists everywhere else. Thus, the expected rejection rates for the FAT is 0.05 when $\alpha = 0$ and there exists publication bias against insignificance. Everywhere else, the expected rejection rate is 1.00.[9]

For the PET, rejection of the null indicates that the estimated mean effect is different from zero. Thus, the expected rejection rate for the PET is 1.00 for all $\alpha > 0$, and 0.05 when $\alpha = 0$. AR highlight scenarios where FPP performs poorly by color-coding the respective cells in the table in red and grey.[10] They conclude that the performance of the FPP procedure declines as the DGP moves away from the unrealistic case of one true population effect for all estimates (Fixed Effects/FE), and becomes more realistic, with population effects allowed to be heterogeneous across studies (Random Effects/RE), and both across and within studies (Panel Random Effects/PRE). Here is how they summarize their results:

> "... we see a consistent pattern of declining performance of the FPP procedure on the FAT and PET as we move from the unrealistic, simplistic environment of FE, to the more realistic RE and PRE data environments. This is represented by the increasing prevalence of red cells as one moves from the top panel of Table 4 down through the bottom panel. It is true for both selection bias against statistical insignificance, and selection bias against negative estimates."

In Table 3, we see the same pattern of red and grey cells with the new, corrected FAT and PET rejection rates. In fact, the pattern is somewhat stronger than in AR. The FPP procedure works well in the unrealistic data environment of Fixed Effects. Once the data environments become more

[8] See AR (page 291) for a discussion of the relationship between α and publication selection bias.

[9] Following AR, I ignore issues associated with statistical power.

[10] The notes at the bottom of Tables 3 and 4 give more detail about the different types of cell color-coding.

Table 3: Replication of AR's Table 4 — FAT and PET

	Publication Bias against Insignificance			Publication Bias against Wrong Sign		
	Fixed Effects (FE)					
α	Percent Published	FAT Rejection Rates	PET Rejection Rates	Percent Published	FAT Rejection Rates	PET Rejection Rates
0.0	14.3	**0.06** (0.07)	**0.15** (0.16)	55.1	**1.00** (1.00)	**0.08** (0.09)
0.5	23.1	**1.00** (1.00)	**1.00** (1.00)	71.8	**1.00** (1.00)	**1.00** (1.00)
1.0	31.8	**1.00** (1.00)	**1.00** (1.00)	80.6	**1.00** (1.00)	**1.00** (1.00)
1.5	40.0	**1.00** (1.00)	**1.00** (1.00)	86.5	**1.00** (1.00)	**1.00** (1.00)
2.0	47.5	**1.00** (1.00)	**1.00** (1.00)	90.6	**1.00** (1.00)	**1.00** (1.00)
2.5	54.6	**1.00** (1.00)	**1.00** (1.00)	93.5	**0.98** (0.98)	**1.00** (1.00)
3.0	61.0	**1.00** (1.00)	**1.00** (1.00)	95.5	**0.81** (0.81)	**1.00** (1.00)
3.5	67.0	**1.00** (1.00)	**1.00** (1.00)	97.0	**0.55** (0.52)	**1.00** (1.00)
4.0	72.3	**1.00** (1.00)	**1.00** (1.00)	98.0	**0.31** (0.29)	**1.00** (1.00)
	Random Effects (RE)					
α	Percent Published	FAT Rejection Rates	PET Rejection Rates	Percent Published	FAT Rejection Rates	PET Rejection Rates
0.0	27.1	**0.09** (0.09)	**0.08** (0.09)	55.0	**0.63** (0.62)	**0.89** (0.89)
0.5	28.7	**0.34** (0.33)	**0.66** (0.64)	65.3	**0.63** (0.61)	**0.99** (1.00)
1.0	33.1	**0.70** (0.68)	**0.98** (0.99)	74.7	**0.54** (0.59)	**1.00** (1.00)
1.5	39.2	**0.78** (0.80)	**1.00** (1.00)	81.9	**0.47** (0.47)	**1.00** (1.00)
2.0	45.9	**0.80** (0.78)	**1.00** (1.00)	87.5	**0.35** (0.36)	**1.00** (1.00)
2.5	52.8	**0.77** (0.77)	**1.00** (1.00)	91.3	**0.25** (0.21)	**1.00** (1.00)
3.0	59.2	**0.69** (0.67)	**1.00** (1.00)	94.0	**0.18** (0.18)	**1.00** (1.00)
3.5	65.0	**0.61** (0.62)	**1.00** (1.00)	95.9	**0.13** (0.13)	**1.00** (1.00)
4.0	70.5	**0.53** (0.53)	**1.00** (1.00)	97.2	**0.11** (0.09)	**1.00** (1.00)

Table 3 continued on next page

Table 3 continued: Replication of AR's Table 4 — FAT and PET

	Publication Bias against Insignificance			Publication Bias against Wrong Sign		
	Fixed Effects (FE)					
α	Percent Published	FAT Rejection Rates	PET Rejection Rates	Percent Published	FAT Rejection Rates	PET Rejection Rates
0.0	19.1	**0.63** (0.56)	**0.65** (0.31)	38.5	**0.64** (0.46)	**0.99** (0.77)
0.5	19.8	**0.61** (0.61)	**0.65** (0.34)	47.6	**0.65** (0.47)	**1.00** (0.85)
1.0	22.1	**0.67** (0.61)	**0.71** (0.49)	56.7	**0.63** (0.42)	**1.00** (0.89)
1.5	25.2	**0.70** (0.72)	**0.79** (0.62)	65.5	**0.62** (0.45)	**1.00** (0.91)
2.0	29.6	**0.73** (0.65)	**0.88** (0.70)	73.6	**0.60** (0.47)	**1.00** (0.92)
2.5	34.5	**0.73** (0.61)	**0.94** (0.84)	80.2	**0.59** (0.49)	**1.00** (0.96)
3.0	40.3	**0.72** (0.63)	**0.97** (0.89)	86.1	**0.61** (0.45)	**1.00** (0.97)
3.5	46.6	**0.71** (0.67)	**0.99** (0.95)	90.7	**0.58** (0.42)	**1.00** (0.98)
4.0	52.7	**0.71** (0.65)	**1.00** (0.97)	93.8	**0.60** (0.39)	**1.00** (0.99)

Note: α is the mean true effect in the simulations underlying a given experiment (see TABLE 1A). "Percent Published" represents the percentage of estimates (out of the original 1000) that survive publication selection bias and are available to the meta-analyst for study. The values in the FAT and PET columns represent the rejection rates for the respective null hypotheses ($\beta_1 = 0$ and $\beta_0 = 0$, respectively, in Equation (1) in the text). Rejection rates are expected to be 0.05 for (i) the FAT when $\alpha = 0$ and publication selection is biased against insignificant estimates; and (ii) the PET when $\alpha = 0$ under both types of publication selection bias. Everywhere else, rejection rates are expected to be 1.00.
The red and bold-faced numbers in parentheses are taken from AR's Table 4. The black, boldfaced numbers directly to their left are the corrected values. Red- and grey-colored cells indicate that the associated rejection rates represent "poor performance": Red-colored cells indicate (i) a rejection rate > 0.15 when the expected rejection rate is 0.05; or (ii.a) a rejection rate < 0.80 when the expected rejection rate is 1.00, and (ii.b) more than 10% of the estimates have been been censored due to publication selection bias. Grey-colored cells indicate a rejection rate < 0.80 when the expected rejection rate is 1.00, and less than 10% of the estimates have been been censored due to publication selection bias.

realistic, by allowing true effects to differ across and within studies, and by allowing studies to produce more than just one estimate, the performance of the FPP procedure declines markedly.

AR's error also affects their results in their Table 6, in which the FPP estimator is compared with two other, weighted least squares estimators: WLS-FE and WLS-RE.[11] These latter estimators do not make any corrections for publication selection bias. The point of this comparison is to see whether the FPP procedure provides improved performance compared to estimators that ignore publication bias. They compare the estimators on three dimensions: (i) Bias; (ii) Mean Squared Error; and (iii) Inference, where the estimated effect is tested for equality with its true value.

In the first two panels of their table, AR identify the least biased and most efficient of the three estimators by yellow-highlighting the estimator that is best on the respective dimension. In the last panel, they test whether the estimated mean effect equals the true effect, so that the expected rejection rate is 0.05. Once again, the original AR values are in parentheses, red, bold and italicized. The corrected values are in black and bold immediately to their left. Once again, there are some substantial differences between the original AR values and the corrected values. AR report that the FPP procedure is almost always less biased than the WLS-FE and WLS-RE estimators. However, it is never the most efficient estimator, consistently dominated by the two WLS estimators. Further, neither the FPP nor the other two estimators can be relied upon for inference.

The corrected values indicate that the FPP procedure does even worse than AR report. For example, when $\alpha = 2.0$ and publication bias is directed against statistical insignificance, AR report that FPP is best in terms of bias, as the estimated mean overall effect is 2.17 versus 2.34 and 3.13 for WLS-FE and WLS-RE. In contrast, the corrected values indicate that WLS-FE is best with an estimated overall mean value of 1.97, versus 2.16 and 1.93 for FPP and WLS-RE, respectively. Similarly, when $\alpha = 2.0$ and publication bias is directed against wrong-signed estimates, AR report that the least biased estimator is again FPP, whereas the corrected values indicate that WLS-RE is best.

Overall, the results in Table 4 confirm AR's finding that FPP does poorly on efficiency and inference. They also indicate that FPP is frequently more biased than the WLS estimators. Thus, as in Table 3, correcting AR's programing code produced results that marginally strengthen their main arguments.

[11] WLS-FE refers to the weighted least squares estimate of β_0, the overall mean effect, in Equation (4) of AR, where the weights are related to the precision of the estimated effects in the original study. WLS-RE refers to the weighted least squares estimate of β_0 in the same equation, except that the weights are related to the precision of the estimated effect in the original study, plus a measure of heterogeneity of the true effects across studies. The interested reader is referred to AR for more detail.

Table 4: Replication of AR's Table 6 — Comparison of FPP, WLS-FE and WLS-RE (Panel Random Effects)

	Publication Bias against Insignificance			Publication Bias against Wrong Sign		
	Mean Value of $\hat{\beta}_0$					
α	FPP	WLS-FE	WLS-RE	FPP	WLS-FE	WLS-RE
0.0	-0.01 (0.06)	0.00 (0.04)	-0.03 (-0.01)	1.75 (1.69)	1.70 (1.77)	1.67 (1.88)
0.5	0.51 (0.58)	0.46 (0.67)	0.37 (1.04)	1.94 (1.93)	1.90 (1.97)	1.84 (2.08)
1.0	0.97 (1.25)	0.86 (1.36)	0.80 (1.92)	2.14 (2.13)	2.10 (2.17)	2.07 (2.29)
1.5	1.51 (1.71)	1.38 (1.84)	1.33 (2.58)	2.34 (2.36)	2.30 (2.40)	2.29 (2.53)
2.0	2.16 (2.17)	1.97 (2.34)	1.93 (3.13)	2.65 (2.62)	2.61 (2.67)	2.58 (2.80)
2.5	2.63 (2.70)	2.41 (2.83)	2.44 (3.58)	2.98 (2.93)	2.95 (2.96)	2.90 (3.09)
3.0	3.19 (3.18)	2.97 (3.32)	2.91 (4.02)	3.33 (3.30)	3.31 (3.33)	3.26 (3.44)
3.5	3.66 (3.69)	3.46 (3.78)	3.33 (4.40)	3.75 (3.72)	3.73 (3.75)	3.67 (3.83)
4.0	4.09 (4.07)	3.92 (4.17)	3.76 (4.77)	4.13 (4.09)	4.11 (4.12)	4.11 (4.21)
	Mean Squared Error					
α	FPP	WLS-FE	WLS-RE	FPP	WLS-FE	WLS-RE
0.0	1.14 (1.69)	2.33 (0.93)	1.20 (0.45)	3.40 (3.66)	3.46 (3.47)	2.95 (3.59)
0.5	1.18 (1.63)	2.14 (0.86)	0.99 (0.66)	2.45 (2.78)	2.57 (2.48)	1.99 (2.55)
1.0	1.20 (1.59)	1.93 (0.91)	0.92 (1.15)	1.69 (1.99)	1.83 (1.67)	1.29 (1.71)
1.5	1.25 (1.60)	1.72 (0.90)	0.58 (1.37)	1.09 (1.50)	1.24 (1.13)	0.78 (1.10)
2.0	1.09 (1.47)	1.46 (0.79)	0.42 (1.43)	0.81 (1.27)	0.99 (0.79)	0.51 (0.68)
2.5	1.04 (1.19)	1.39 (0.67)	0.35 (1.28)	0.67 (0.95)	0.91 (0.54)	0.34 (0.40)
3.0	0.82 (1.25)	1.22 (0.63)	0.28 (1.15)	0.54 (0.91)	0.79 (0.44)	0.24 (0.24)
3.5	0.64 (1.14)	1.08 (0.58)	0.27 (0.88)	0.52 (1.01)	0.81 (0.45)	0.18 (0.16)
4.0	0.63 (1.02)	1.12 (0.47)	0.29 (0.66)	0.51 (0.88)	0.81 (0.38)	0.18 (0.09)

Table 4 continued on next page

Table 4 continued: Replication of AR's Table 6 — Comparison of FPP, WLS-FE and WLS-RE (Panel Random Effects)

	Publication Bias against Insignificance			Publication Bias against Wrong Sign		
	Mean Value of $\hat{\beta}_0$					
α	FPP	WLS-FE	WLS-RE	FPP	WLS-FE	WLS-RE
0.0	**0.73** (0.27)	**0.26** (0.19)	**0.52** (0.06)	**1.00** (0.77)	**0.90** (0.99)	**1.00** (1.00)
0.5	**0.77** (0.31)	**0.23** (0.18)	**0.50** (0.14)	**1.00** (0.79)	**0.74** (0.95)	**1.00** (1.00)
1.0	**0.84** (0.33)	**0.22** (0.22)	**0.52** (0.39)	**0.98** (0.62)	**0.53** (0.77)	**0.98** (1.00)
1.5	**0.81** (0.32)	**0.22** (0.25)	**0.54** (0.63)	**0.94** (0.47)	**0.39** (0.58)	**0.91** (1.00)
2.0	**0.80** (0.31)	**0.23** (0.22)	**0.54** (0.78)	**0.91** (0.40)	**0.28** (0.39)	**0.82** (0.98)
2.5	**0.84** (0.27)	**0.24** (0.21)	**0.55** (0.87)	**0.88** (0.29)	**0.26** (0.23)	**0.69** (0.85)
3.0	**0.80** (0.26)	**0.21** (0.21)	**0.57** (0.91)	**0.90** (0.26)	**0.20** (0.18)	**0.61** (0.57)
3.5	**0.82** (0.30)	**0.18** (0.21)	**0.57** (0.89)	**0.85** (0.30)	**0.18** (0.19)	**0.52** (0.37)
4.0	**0.85** (0.26)	**0.23** (0.16)	**0.59** (0.84)	**0.83** (0.26)	**0.21** (0.15)	**0.52** (0.16)

Note: α is the mean true effect in the simulations underlying a given experiment. The top panel reports the mean estimated value of β_0 in Equation (4) of AR, where β_0 represents the mean true effect. The associated estimate, $\hat{\beta}_0$, is averaged over 1000 simulated meta-analysis studies using three different methods. "FPP" reports the estimate of β_0 in Equation (1) in AR using the FPP procedure. WLS-FE refers to the weighted least squares estimate of β_0 in Equation (4) of AR, where the weights are related to the precision of the estimated effects in the original study. WLS-RE refers to the weighted least squares estimate of β_0 in the same equation, except that the weights are related to the precision of the estimated effect in the original study plus a measure of heterogeneity of the true effects across studies. The interested reader is referred to AR for more detail. The middle panel reports the average mean squared error (MSE) value for each of the three methods. The bottom panel reports rejection rates associated with the null hypothesis $\beta_0 = \alpha$. Rejection rates are expected to be 0.05 for all experiments.
The red bold-faced numbers in parentheses are taken from AR's Table 6. The black boldfaced numbers directly to their left are the corrected values. Yellow-colored cells indicate that the respective estimator is "best" of the three estimators (FPP, WLS-FE, WLS-RE) for a given experiment. "Best" means either least biased or smallest MSE. Red-colored cells indicate a rejection rate > 0.15 when the expected rejection rate is 0.05.

3 Extension: different numbers of estimates per study

AR's simulation framework for the Panel Random Effects (PRE) DGP restricts each primary study to have an equal number of estimated effects (i.e., balanced panel data). In this section, I extend their simulation framework by allowing each primary study to have different numbers of effects. In the economics literature, even though studies examine the same research question, the number of estimates per study varies widely. My extension generates meta-analysis samples that consist of unbalanced panel datasets. This more closely matches the kinds of data a typical economics meta-analyst works with.

Under this extension, the simulated dataset would not only have the same mean number of estimates per study, but also have the amount of variations observed in the real economics meta-analysis study datasets. This would allow us to examine the performance of each meta-analysis estimator when the cluster structure is asymmetric. Therefore, the performance analysis under this extension framework would better represent the true performance of the estimators in the meta-analysis studies. Having said the relevance and usefulness of this extension, I am also aware of its limitations. When a meta-analysis dataset is simulated to match the first two moments of the observed number of estimates per study, the higher moments (e.g., skewness and kurtosis) are ignored. As not a few studies report only one estimate and a few studies report more than hundred estimates, skewness and kurtosis do appear in real meta-analysis datasets. However, simulating a dataset that matches all the characteristics that we observe from real meta-analysis datasets is beyond the scope of this study, so I leave this for future work.

In order to determine a representative distribution for the number of estimated effects per study, I examine 13 Meta-Regression-Analyses (MRAs) that include a total of 964 primary studies. The 13 MRAs, along with details about the number of included primary studies and the total number of estimated effects, are listed in Table 5.

On average, an MRA study has about 75 ($\approx$ 964/13) primary studies. The total number of effects reported in the 964 primary studies is 12,020, so that the average number of effects per study is about 12.5 ($\approx$ 12020/964). The number of effects in a primary study is highly skewed. About 20% of the primary studies (187 studies) have only one estimated effect. Approximately 50% (500 studies) have 5 or less estimated effects. On the other end, several studies have more than 200 estimated effects.

Figure 1 presents a histogram and a kernel density of the number of effects per study for the 964 primary studies. The mean and standard deviations are 12.47 and 23.41, respectively. The distribution has a lower bound of one (the minimum number of estimated effects is one) and a very long tail. Visually, it resembles a log normal distribution.

Table 5: Numbers of Primary Studies and Estimated Effects in Economics and Business MRA Studies

Authors/Journal	Number of Primary Studies	Number of Estimated Effects
Dalhuisen et al. (2003) *Land Economics*	51	314
Nijkamp and Poot (2005) *Journal of Economic Surveys*	16	208
Melguizo and González-Páramo (2013) *SERIEs*	48	143
Havranek, Irsova and Janda (2012) *Energy Economics*	41	202
Haile and Pugh (2013) *Journal of International Trade & Economic Development*	89	1,255
Doucouliagos and Paldam (2013) *The Journal of Development Studies*	130	1,921
Ogundari and Abdulai (2013) *Food Policy*	46	115
Nataraj et al. (2014) *Journal of Economic Surveys*	9	220
de Linde Leonard et al. (2014) *British Journal of Industrial Relations*	16	710
Havranek and Kokes (2015) *Energy Economics*	241	2,247
Havránek (2015) *Journal of the European Economic Association*	169	2,735
Bruno and Cipollina (2018) *World Economy*	46	1,643
Havranek et al. (2018) *Land Economics*	62	307
Total	964	12,020

I calibrate the location and shape parameters of the log normal distribution to match the distribution of effects per study in Figure 1. The respective parameter values are $\log(5.5)$ and $\log(3.5)^2$. As a check, Figure 2 displays a distribution of simulated numbers drawn from a log normal distribution having location and shape parameters equal to these values.[12]

In order to incorporate the distribution of the number of effects per study into the simulation framework of AR, I first randomly draw a number n_i from the log normal distribution, $LN(\log(3.5), \log(5.5)^2)$ foreach study i. I then simulate $[n_i]$ estimated effects for each study i.[13,14] After artificially creating the associated meta-analysis datasets, I reproduce AR's performance comparison of the three different estimators (FPP, WLS-FE, and WLS-RE). The results of this analysis are reported in Table 6.

I find that average biases in Table 6, measured by the difference between the true mean value and its estimate, are not much different from the values reported in Table 5. However, I find a substantial increase in the mean squared errors (MSE). Type I error rates are also marginally higher with unbalanced panel data. Overall, the results are qualitatively similar to what AR obtain for their PRE simulations assuming a fixed number of estimates per study, as can be confirmed by comparing Table 6 with Table 4.

4 Concluding Remarks

This study replicates the recent simulation work of Alinaghi and Reed (2018). During the replication process, I found two errors in their programming code. This study reports the results of correcting those mistakes and re-simulating AR's experiments. I also extend their analysis by relaxing a restriction on the number of estimates per study. After correcting their mistakes and extending their analysis, I find that some values differ substantially, but the qualitative results remain the same. My replication confirms this summary, taken from AR's conclusion:

> "The FPP procedure is generally used for 3 purposes: (1) to test whether a sample of estimates suffers from publication bias, (2) to test whether the estimates indicate that the effect of interest is statistically different from zero, and (3) to obtain an estimate of the overall, mean effect. . . .Our findings indicate that the FPP procedure performs well in the basic but unrealistic environment of Fixed Effects, where all estimates are assumed to derive from a single, population value and sampling error is the only reason for why studies produce different estimates. However, when we study its performance in more realistic data environments, where there is heterogeneity in population effects across and within studies, the FPP procedure becomes unreliable for the first 2 purposes, and less efficient than some other estimators that do not correct for publication bias. Further, hypothesis tests about the overall mean effect often cannot be trusted."

[12] Mean and standard deviation of $LN(\log(3.5), \log(5.5)^2)$ are 12.05 and 23.51 respectively.
Mean = $e^{\log(5.5)+\log(3.5)^2/2} \approx 12.05$; Standard Deviation = $[\left(e^{\log(3.5)^2} - 1\right) * e^{2*\log(5.5)+\log(3.5)^2}]^{\frac{1}{2}} \approx 23.51$.

[13] I first draw a real number from the log distribution (e.g. $n_i = 2.6$), and then I use the associated ceiling number (e.g., ceiling = $[n_i]$ = [2.6] = 3 when creating effects for each study. I do this because I need an integer and a number greater than or equal to one.

[14] In AR, primary studies are assumed to have exactly 10 estimated effects. This is close to the mean number of effects (12.47) that I find from the sample of 964 studies in Table 5.

Figure 1: Histogram and Kernel Density of the Number of Reported Effects in 964 Primary Studies

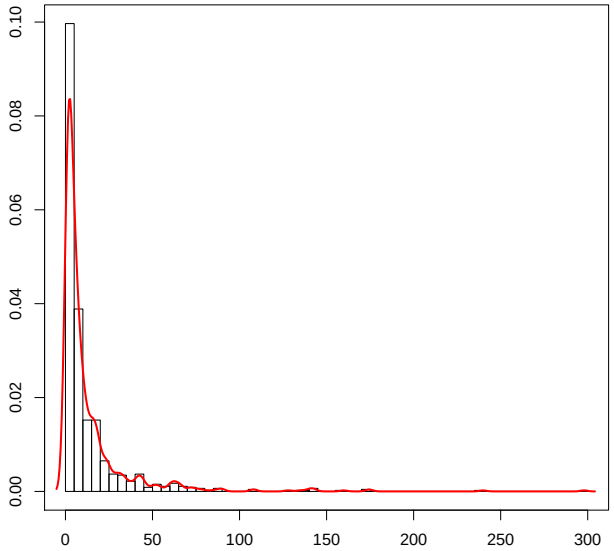

Figure 2: Kernel Density for the Simulated Number of Reported Effects

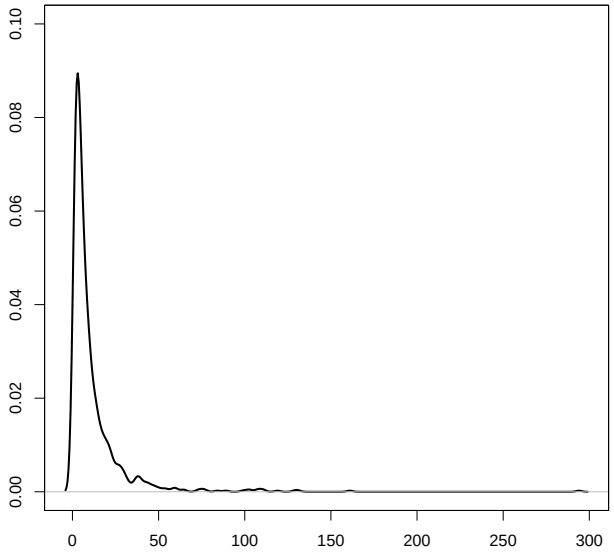

Table 6: Comparison of FPP Estimates with WLS-FE and WLS-RE: Unbalanced Panel Data Setting

	Publication Bias against Insignificance			Publication Bias against Wrong Sign		
	Mean Value of $\hat{\beta}_0$					
α	FPP	WLS-FE	WLS-RE	FPP	WLS-FE	WLS-RE
0.0	-0.05	-0.06	-0.09	1.75	1.67	1.66
0.5	0.51	0.43	0.41	1.98	1.92	1.88
1.0	1.12	0.99	0.93	2.19	2.12	2.07
1.5	1.60	1.43	1.40	2.42	2.37	2.30
2.0	2.20	1.98	1.98	2.72	2.68	2.61
2.5	2.73	2.49	2.44	3.02	2.97	2.95
3.0	3.24	2.99	2.91	3.33	3.28	3.28
3.5	3.62	3.39	3.30	3.67	3.62	3.62
4.0	4.13	3.94	3.76	4.12	4.09	4.08
	Mean Squared Error					
α	FPP	WLS-FE	WLS-RE	FPP	WLS-FE	WLS-RE
0.0	1.76	3.40	2.71	3.50	3.45	3.16
0.5	1.81	3.36	2.47	2.68	2.74	2.29
1.0	1.78	2.74	1.86	1.95	2.09	1.58
1.5	1.82	2.56	1.49	1.38	1.59	1.05
2.0	1.61	2.21	1.07	1.23	1.56	0.85
2.5	1.36	1.96	0.81	0.89	1.19	0.65
3.0	1.16	1.78	0.75	0.74	1.04	0.54
3.5	1.07	1.59	0.75	0.72	1.04	0.56
4.0	1.03	1.61	0.74	0.73	1.07	0.56

Table 6 continued on next page

Table 6 continued: Comparison of FPP Estimates with WLS-FE and WLS-RE: Unbalanced Panel Data Setting

	Publication Bias against Insignificance			Publication Bias against Wrong Sign		
	Mean Value of $\hat{\beta}_0$					
α	FPP	WLS-FE	WLS-RE	FPP	WLS-FE	WLS-RE
0.0	0.78	0.34	0.67	1.00	0.88	0.99
0.5	0.80	0.35	0.67	1.00	0.71	0.99
1.0	0.87	0.30	0.68	0.98	0.54	0.92
1.5	0.89	0.34	0.69	0.94	0.45	0.86
2.0	0.86	0.31	0.66	0.92	0.39	0.81
2.5	0.87	0.35	0.67	0.92	0.31	0.76
3.0	0.86	0.32	0.73	0.91	0.29	0.73
3.5	0.88	0.31	0.75	0.88	0.27	0.74
4.0	0.88	0.33	0.74	0.87	0.27	0.73

Note: α is the mean true effect in the simulations underlying a given experiment. The top panel reports the mean estimated value of β_0 in Equation (4) of AR, where β_0 represents the mean true effect. The associated estimate, $\hat{\beta}_0$, is averaged over 1000 simulated meta-analysis studies using three different methods. "FPP" reports the estimate of β_0 in Equation (1) in AR using the FPP procedure. WLS-FE refers to the weighted least squares estimate of β_0 in Equation (4) of AR, where the weights are related to the precision of the estimated effects in the original study. WLS-RE refers to the weighted least squares estimate of of β_0 in the same equation, except that the weights are related to the precision of the estimated effect in the original study plus a measure of heterogeneity of the true effects across studies. The interested reader is referred to AR for more detail.
The middle panel reports the average mean squared error (MSE) value for each of the three methods. The bottom panel reports rejection rates associated with the null hypothesis $\beta_0 = \alpha$. Rejection rates are expected to be 0.05 for all experiments.Yellow-colored cells indicate that the respective estimator is "best" of the three estimators (FPP, WLS-FE, WLS-RE) for a given experiment. "Best" means either least biased or smallest MSE. Red-colored cells indicate a rejection rate > 0.15 when the expected rejection rate is 0.05.

References

Alinaghi, Nazila and W. Robert Reed (2018). "Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work?" *Research Synthesis Methods* 9: 285–311. DOI: 10.1002/jrsm.1298.

Andrews, Isaiah and Maximilian Kasy (2017). "Identification of and correction for publication bias." *NBER Working Paper* No. 23298. DOI: 10.3386/w23298.

Bruno, Randolph Luca and Maria Cipollina (2018). "A meta-analysis of the indirect impact of foreign direct investment in old and new EU member states: Understanding productivity spillovers." *The World Economy* 41: 1342–1377. DOI: 10.1111/twec.12587.

Christensen, Garret S. and Edward Miguel (2016). "Transparency, reproducibility, and the credibility of economics research." *NBER Working Paper* No. 22989. DOI: 10.3386/w22989.

Dalhuisen, Jasper M., Raymond J. G. M. Florax, Henri L. F. de Groot and Peter Nijkamp (2003). "Price and income elasticities of residential water demand: A meta-analysis." *Land Economics* 79(2): 292–308.

Glass, Gene V. (1976). "Primary, secondary, and meta-analysis of research." *Educational Researcher* 5(10): 3–8. DOI: 10.2307/1174772.

Doucouliagos, Hristos and Martin Paldam (2013). "The robust result in meta-analysis of aid effectiveness: a response to Mekasha and Tarp." *The Journal of Development Studies* 49(4): 584–587. DOI: 10.1080/00220388.2013.764595.

Haile, Mekbib and Geoff T. Pugh (2013). "Does exchange rate volatility discourage international trade? A meta-regression analysis." *The Journal of International Trade & Economic Development* 22(3): 321–350.

Havránek, Tomáš (2015). "Measuring intertemporal substitution: The importance of method choices and selective reporting." *Journal of the European Economic Association* 13: 1180–1204. DOI: 10.1111/jeea.12133.

Havránek, Tomáš, Zuzana Iršová, and Karel Janda (2012). "Demand for gasoline is more price-inelastic than commonly thought." *Energy Economics* 34(1): 201–207. DOI: 10.1016/j.eneco.2011.09.003.

Havránek, Tomáš, Zuzana Iršová, and Tomas Vlach (2018). "Measuring the income elasticity of water demand: The importance of publication and endogeneity biases." *Land Economics* 94: 259–283. DOI: 10.3368/le.94.2.259.

Havránek, T. and Ondrej Kokes (2015). "Income elasticity of gasoline demand: A meta-analysis." *Energy Economics* 47: 77–86. DOI: 10.1016/j.eneco.2014.11.004.

Higgins, Julian P. T. and Simon G. Thompson (2002). "Quantifying heterogeneity in a meta-analysis." *Statistics in Medicine* 21(11): 1539–58. DOI: 10.1002/sim.1186.

de Linde Leonard, Megan, T. D. Stanley, and Hristos Doucouliagos (2014). "Does the UK minimum wage reduce employment? A meta-regression analysis." *An International Journal of Employment Relations* 52(3): 499–520. DOI: 10.1111/bjir.12031.

Melguizo, Ángel and José Manuel González-Páramo (2013). "Who bears labour taxes and social contributions? A meta-analysis approach." *SERIEs* 4(3): 247–271. DOI: 10.1007/s13209-012-0091-x.

Nataraj, Shanthi, Francisco Perez-Arce, Krishna B. Kumar, and Sinduja V. Srinivasan (2014). "The impact of labor market regulation on employment in low-income countries: A meta-analysis." *Journal of Economic Surveys* 28(3): 551–572. DOI: 10.1111/joes.12040.

Nijkamp, Peter and Jaques Poot (2005). "The last word on the wage curve?" *Journal of Economic Surveys* 19: 421–450. DOI: j.0950-0804.2005.00254.x.

Ogundari, Kolawole and Awudu Abdulai (2013). "Examining the heterogeneity in calorie–income elasticities: A meta-analysis." *Food Policy* 40: 119–128. DOI: j.foodpol.2013.03.001.

Reed, W. Robert. (2017). "Replications in labor economics." *IZA World of Labor* 2018: 413. DOI: 10.15185/izawol.413.

Rothstein, Hannah R., Alexander J. Sutton, and Michael Borenstein (Eds.) (2006). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments.* Chichester: John Wiley & Sons.

Stanley T.D. and Hristos Doucouliagos (2012). *Meta-regression analysis in economics and business.* London: Routledge.

Appendix - Original Program Codes and Corrections

A1) AR's Tables 2 and 3

TABLES 2 and 3 in AR have "Percent significant" values too low because of the following command:

```
gen sig=(absteffect>=2)
```

This assigned "1" to all observations have abs(teffect) values > 2. It unintentially also included any abs(teffect) observations with missing values, because "." counts as a large number from Stata's perspective. The correct command is:

```
gen sig=(absteffect>=2) if absteffect != .
```

A2) AR's Table 4

AR's program for their TABLE 4 is wrong because it includes study (panel) fixed effects.

```
tab ID, gen(dum)
forvalues i = 1/100 {
generate SE`i' = seeffect*dum`i'
}
regress teffect dum1-dum100 pet, vce(cluster ID)
return scalar effect_PET = _b[pet]
scalar effect_PET = _b[pet]
test _b[_cons] = 0
return scalar pvalue_FAT = r(p)
scalar pvalue_FAT = r(p)
test pet = `alpha'
return scalar pvalue_PET = r(p)
scalar pvalue_PET = r(p)
test pet = 0
return scalar pvalue_PETFPP = r(p)
scalar pvalue_PETFPP = r(p)
regress teffect SE1-SE100 pet, noc vce(cluster ID)
return scalar effect_PEESE = _b[pet]
scalar effect_PEESE = _b[pet]
test pet = `alpha'
return scalar pvalue_PEESE = r(p)
scalar pvalue_PEESE = r(p)
return scalar effect_FPP = effect_PET
return scalar pvalue_FPP = pvalue_PET
if pvalue_PETFPP < 0.05 {
return scalar effect_FPP = effect_PEESE
return scalar pvalue_FPP = pvalue_PEESE
}
return scalar N = e(N)
```

The following is a correction of AR's program:

```
regress teffect pet, vce(cluster ID)
test _b[_cons] = 0
return scalar pvalue_FAT = r(p)
test pet = 0
return scalar pvalue_PET = r(p)
scalar N = e(N)
```

A3) AR's Table 6

AR's program for their TABLE 6 is wrong for the same reason as above. Here is their program:

```
tab ID, gen(dum)
forvalues i = 1/100 {
generate SE`i' = seeffect*dum`i'
}
regress teffect dum1-dum100 pet, vce(cluster ID)
return scalar effect_PET = _b[pet]
scalar effect_PET = _b[pet]
test _b[_cons] = 0
return scalar pvalue_FAT = r(p)
scalar pvalue_FAT = r(p)
test pet = `alpha'
return scalar pvalue_PET = r(p)
scalar pvalue_PET = r(p)
test pet = 0
return scalar pvalue_PETFPP = r(p)
scalar pvalue_PETFPP = r(p)
regress teffect SE1-SE100 pet, noc vce(cluster ID)
return scalar effect_PEESE = _b[pet]
scalar effect_PEESE = _b[pet]
test pet = `alpha'
return scalar pvalue_PEESE = r(p)
scalar pvalue_PEESE = r(p)
regress feteffect fepet, noc vce(cluster ID)
return scalar effect_FE = _b[fepet]
test fepet = `alpha'
return scalar pvalue_FE = r(p)
quietly metareg effect, wsse(seeffect) mm
scalar tau2 = e(tau2)
gen revarR= seeffect^2 + tau2
gen reseR = sqrt(revarR)
gen reteffect = effect/reseR
gen repet = 1/reseR
regress reteffect repet, noc vce(cluster ID)
return scalar effect_RE = _b[repet]
test repet = `alpha'
```

```
return scalar pvalue_RE = r(p)
return scalar effect_FPP = effect_PET
return scalar pvalue_FPP = pvalue_PET
if pvalue_PETFPP < 0.05 {
return scalar effect_FPP = effect_PEESE
return scalar pvalue_FPP = pvalue_PEESE
}
```

The following is the corrected program:

```
regress teffect pet, vce(cluster ID)
scalar effect_PET = _b[pet]
test pet = `alpha'
scalar pvalue_PET = r(p)
test pet = 0
scalar pvalue_PETFPP = r(p)
regress teffect seeffect pet, noc vce(cluster ID)
scalar effect_PEESE = _b[pet]
test pet = `alpha'
scalar pvalue_PEESE = r(p)
regress feteffect fepet, noc vce(cluster ID)
return scalar effect_FE = _b[fepet]
scalar effect_FE = _b[fepet]
test fepet = `alpha'
return scalar pvalue_FE = r(p)
scalar pvalue_FE = r(p)
quietly metareg effect, wsse(seeffect) mm
scalar tau2 = e(tau2)
gen revarR= seeffect^2 + tau2
gen reseR = sqrt(revarR)
gen reteffect = effect/reseR
gen repet = 1/reseR
regress reteffect repet, noc vce(cluster ID)
return scalar effect_RE = _b[repet]
scalar effect_RE = _b[repet]
test repet = `alpha'
return scalar pvalue_RE = r(p)
scalar pvalue_RE = r(p)
scalar effect_FPP = effect_PET
scalar pvalue_FPP = pvalue_PET
if pvalue_PETFPP < 0.05 {
return scalar effect_FPP = effect_PEESE
return scalar pvalue_FPP = pvalue_PEESE
}
```