Comments on "A Monte Carlo analysis of alternative meta-analysis estimators in the presence of publication bias"

T.D. Stanley, Hendrix College

This paper concerns an interesting and important subject. I have long suggested that we need more simulations of meta-analysis estimators. That need remains. Unfortunately, Reed, Florax and Poot (2015) does not show what it claims. Mean squared errors (MSE) and type I errors are miscalculated, invalidating any generalization that one might wish to make from this study.

Identifying the calculation error is relatively easy, once we understand how Reed, Florax and Poot (2015) define the population from which they are sampling. In several places, Reed, Florax and Poot (2015) identify their population clearly. First, they describe how their "population of M studies" (p.4) is taken from OLS estimates of a known true effect (a regression coefficient), α, after data has been generated from a specified DPG. Note the typo under equation (1); a is not the "true" effect. On page 5, they give further details about how their population is generated, "we repeat the above process M times, so that there is a sample of M studies . . .*This sample constitutes the population of all studies*. . . . M=300" (emphasis added). And, they confirm that these 300 studies is their population, "the population of 300 studies" (p. 9). This establishes their population, the mean of which is the population mean. These population means will be differ by random sampling error as a new set of 300 estimates is drawn and estimated from the DPG described on pp. 4-5. To these population, Reed, Florax and Poot (2015) apply multiple filters to sample selectively, pp.5-6. Lastly, alternative meta-analyses are performed upon these selected samples from their population of 300 estimates.

The statistical issue here is that the "true" population mean will differ in all cases from the "true" effect, α, listed in the left column of their tables (e.g., Table 2, 3, 5, 6). This difference is the result of the way Reed, Florax and Poot (2015) choose to generate their populations by random sampling 300 estimates from their estimation model and DPG (p.4). The mean for each population of 300 is the "true" population effect that all meta-analysis methods seek to estimate, not α. In statistics, sample estimates are only meant to represent the population from which they are drawn

(the population mean of these 300 estimates) and can only be judged relative to this population mean. Comparing them to some remote 'true' effect, α, is not a valid way to calculate MSE or type I errors. This is not an arbitrary judgment subject to a difference of opinion. The formulas for calculating MSE and type I errors depend on the population mean (of the estimated regression coefficient in these simulations). However, Reed, Florax and Poot (2015) do not use what they, themselves, define as their population mean for these calculations. Thus, all of their MSEs and type I errors are miscalculated. They should not be used to make any generalization, not even to the "particular specifications of the data-generation process used in our Monte Carlo simulations" (p.3).

These random population means of 300 will differ, by definition, from their true effect, α, randomly; thus, the individual biases introduced by using the wrong population means are likely to cancel out across 1,000 replications. This cancellation of biases is clearly evidenced in the tables of simulation results reported by Reed, Florax and Poot (2015). Note how biases are quite small for these meta-analysis estimates (see Tables 2, 3, 5, and 6), even in those extreme cases where nearly all negative estimates are omitted from the estimation samples. Contrary to the claims made by Reed, Florax and Poot (2015), their reported results (see Tables 2, 3, 5, and 6) show that our PET-PEESE corrections for publication bias dominate other meta-analysis methods when one looks at bias. Reducing bias is the purpose of these methods (Stanley, 2008; Stanley and Doucouliagos, 2014). Rather than providing new information, these simulations confirm a series of past simulation studies (Stanley, 2008; Moreno, et al., 2009; Stanley, Jarrell and Doucouliagos, 2010; Stanley and Doucouliagos, 2014, Stanley and Doucouliagos, 2015) or are invalid. Thus, "A Monte Carlo analysis of alternative meta-analysis estimators in the presence of publication bias" should be ignored altogether.

The reader might ask, if Reed, Florax and Poot's (2015) reported *biases are valid*, why are *their MSE and type I calculations invalid*? Simple, the errors of calculation by subtracting the wrong population mean in MSE and or in using the wrong population mean for type I errors do not cancel out as they do for bias. Consider the empirically calculated MSE for PEESE across these 1,000 simulation replications. As Reed, Florax and Poot (2015) calculate, $MSE_{RFP} = \Sigma(PEESE_i-\alpha)^2/1000$, whereas the correct calculation is: $MSE = \Sigma(PEESE_i-\alpha_i)^2/1000$ for α_i = the population mean calculated across the 300 estimates that are included in their population. Because α_i is the average of 300 draws from their DPG and estimation model, it will vary from α by

conventional random sampling error, ε_i. Substituting $\alpha_i = \alpha + \varepsilon_i$, into their formula for MSE_{RFP} gives, $\Sigma[(PEESE_i + \varepsilon_i) - \alpha]^2/1000$. This shows clearly that their calculation of MSE_{RFP} adds an inappropriate estimation error the meta-estimate. Because these errors are squared before they are summed, they do not cancel but instead will make MSEs larger. Similarly, type I errors do not cancel. In contrast, calculations of bias are the same as these MSE calculations, except the differences are not squared. Because ε_i is random sampling error, its sum will be close to zero, making Reed, Florax and Poot's (2015) bias calculations nearly correct.

What is especially unfortunate is that these mistakes are necessary. All that is needed is to sample once from the DPG and estimation model. Publication selection can be induced by the way each estimate is selected from those generated by their DPG and estimation model. The general lesson it that is not appropriate to have two different populations as the basis of a Monte Carlo study and then select that population parameter which best serves your purposes.

To anticipate a possible counter-argument, suppose that someone wishes to claim that a 2-step process is a realistic description of how reported estimates in economics are generated. That is, suppose that the first step is to generate some 'feasible' finite population of estimates by sampling from an infinite theoretical population. Then, on the second step, a given study is represented by sampling and selecting from this 'feasible' population of estimates. Such a 2-step process might be a way to account for the fact that only a finite number of data and model specification choices are likely to be available to a given researcher at any point in time. This might be a sensible approach to approximate the research process; however, it has been widely known for decades that the typical choices economists make literally result in millions of possible estimates— "I just ran two million regressions" (Sala-i-Martin, 1997). Had Reed, Florax and Poot (2015) generated millions of estimates for their populations, rather than 300, the above criticism would not have been necessary. Then, the artificial sampling error added by their first step would be so small as to be practically negligible.

The above observations are sufficient to dismiss the contribution of this simulation study. However, even if they had calculated these summary statistics correctly, it is doubtful that the idiosyncratic way that these simulation are designed would allow any generalization beyond their specific and idiosyncratic design. In particular three choices made by Reed, Florax and Poot (2015) shape and constrain what the simulations can show.

- All primary studies are assumed to be estimated using the exact same sample size, 100.

- The possible values of the standard deviations are truncated at the lower end, forcing them to be 0.2 or larger for the fixed-effects case and .5 or larger for the random-effects.

- Selection for statistical significance is assumed to be bi-directional.

Sample sizes used to estimate regression coefficients for the typical area of economics research vary widely, usually by several orders of magnitude. Likewise, many dozens of meta-analyses in economics find at least a few studies have very small standard errors. Together, these assumptions limit the range of standard errors of reported estimates, and hence the precision of these estimates. It is precisely the contrast among reported standard errors (or precisions) that give the PET-PEESE publication methods their power. It takes only one, or a few, highly precise estimate to anchor these publication correction methods, accurately. In economics, precision is often much larger than those simulated here.

Lastly, the particular selection mechanisms used by Reed, Florax and Poot (2015) are also unique. The PET-PEESE methods were deigned to correct directional selection, because economic theory usually makes a qualitative directional prediction and because directional selection presents the greatest threat (and bias) to reported economics findings. If Reed, Florax and Poot (2015) think differently, it would have been easy to adjust PET-PEESE for bidirectional statistical selection had they chose to do so (Bom and Ligthart, 2014). Reed, Florax and Poot (2015) also make a number of other assumptions and choices that makes it difficult to assess the value or applicability of their reported simulation results.

References

Bom, P. R., and J. E. Ligthart. "What have we learned from three decades of research on the productivity of public capital? *Journal of Economic Surveys*, 28 (2014):889-916.

Moreno, S.G., Sutton, A.J., Ades, A.E. Stanley, T.D., Abrams, K.R., Peters, J.L., and Cooper, N.J. "Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study." *BMC Medical Research Methodology* (2009) 9:2, http://www.biomedcentral.com/1471-2288/9/2

Reed, W. R., R.J.G.M. Florax, and J. Poot. *"*A Monte Carlo analysis of alternative meta-analysis estimators in the presence of publication bias," *Economic*s, Discussion Paper No. 2015-9 http://www.economics-ejournal.org/economics/discussionpapers/2015-9

Sala-i-Martin, X. "I just ran two million regressions," *American Economic Review*, 87 (1997):178-83.

Stanley, T.D. "Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection," *Oxford Bulletin of Economics and Statistics*, 70(2008): 103-127.

Stanley, T.D., and Doucouliagos, H. *Meta-regression analysis in economics and business*. London: Routledge, 2012.

Stanley, T.D., and Doucouliagos, H. "Better than random: Weighted least squares meta-regression analysis". Working paper, Deakin University, School of Accounting, Economics, and Finance, Economics Series, SWP 2013/2, updated February 2014.

Stanley, T. D and Doucouliagos, C. "Meta-regression approximations to reduce publication selection bias," *Research Synthesis Methods 5* (2014), 60-78

Stanley, T.D., and Doucouliagos, H. "Neither fixed nor random: Weighted least squares meta-analysis," *Statistics in Medicine* (2015), DOI: 10.1002/sim.6481

Stanley, T.D., Jarrell, S.B. and Doucouliagos, C. "Could it be better to discard 90% of the data? A statistical paradox." *The American Statistician*, 64 (2010): 70-77.