Meta-Analysis in a Nutshell: Techniques and General Findings

Martin Paldam

Abstract

The purpose of this article is to introduce the technique and main findings of meta-analysis to the reader, who is unfamiliar with the field and has the usual objections. A meta-analysis is a quantitative survey of a literature reporting estimates of the same parameter. The funnel showing the distribution of the estimates is normally amazingly wide given their *t*-ratios. Little of the variation can be explained by the quality of the journal (as measured by its impact factor) or by the estimator used. The funnel has often asymmetries consistent with the most likely priors of the researchers, giving a publication bias.

(Published in Special Issue Meta-Analysis in Theory and Practice)

Authors

Martin Paldam, ✉ Department of Economics and Business, Aarhus University, Denmark, mpaldam@econ.au.dk

1   Introduction: Analyzing a *β*-literature

A quantitative survey of an empirical literature on one parameter – say *β* – is termed a meta-analysis. It demands that the studies covered are so similar that their differences can be coded. This is possible in many cases because meta-studies disregard theoretical models and consider results from **estimation models**. Theories may change and develop to become much more complex, but in the end they have to be reduced to a model that can be estimated on available data. Such models tend to be formally rather similar.

The analysis asks three questions to the coded estimates:

Q1:   Do the estimates converge to a meta-average that might be seen as the true one? This is, of course, the key question if *β* is used for policy-making.

Q2:   Can the main innovations of relevance for this convergence be identified?

Q3:   Do the estimates suffer from biases that should be corrected for?

Meta-studies have two levels. **Level one** is discussed in Section 2. It consists of three steps: (i) A search for the *β*-literature; (ii) the coding of this literature; (iii) a set of basic calculations that estimate a **meta-average**, which in many cases differs from the **mean**. These steps allow few choices, so the results are robust. **Level two** is discussed in Section 3. It tries to explain the variation in the results and asks questions to the literature. Here the results are less robust.

Meta-analysis came to economics from medicine around 1990.[1] In medicine an experiment is an expensive clinical trial, while it is a cheap regression in economics.[2] This has strong effects on the number of experiments done and the fraction reported. Hence, meta-analysis required a development of new tools to be useful in economics. When they became available (Stanley 2008), it caused a wave

---

[1] The first studies were Jarrell and Stanley (1990), Doucouliagos (1995) and Card and Krueger (1995).

[2] The subjects of this note are thoroughly covered by the textbook Stanley and Doucouliagos (2012). The benefit transfer literature will soon be covered in the textbook Johnson et al (2015). The standard textbook for meta-analysis in medicine is Hunter and Schmidt (1990, 2004). The tools presented may also apply to experimental economics and field experiments.

of studies. At present about 750 meta-studies have been made in economics (broadly defined), and about 40,000 papers have been coded.

The new tools have been analyzed in half a dozen simulation studies,[3] where the true value of β is known. This has built trust in the tools, and it allows the analysts to claims that the meta-average is much closer to the true value than is the mean. Consequently, the difference between the two averages is an estimate of the ***publication bias***. It is defined as a systematic difference between the published results and the true value.

The literature shows that such biases are common in most fields.[4] The research and publication process involves choices, which require judgment that may be affected by the results desired, leading to exaggeration in the direction wanted. The fact that most experiments remain unreported gives a considerable scope for exaggeration. This will be further discussed below, for now a simple rule of thumb is to expect that the true value is half the published one in the average paper.

One of the strongest beliefs in economics is that humans react to priors and incentives, and all economists know of many studies that support this belief. At the same time many economists seem to believe that they themselves are 'above' such reactions and engaged in pure truth seeking. Meta-analysis takes the view that economists are humans. This should not be controversial, but I know that it is. It is something many economists do not want to know.

Meta-analysts are human too. Hence, it is important that level one of the analysis is robust, in the sense of containing few choices requiring judgment. Two independent studies of the same literature should reach much the same result. At

_____

[3] The simulations generate β-literatures in different ways and show that while the mean is biased, the meta-average gets close to the true value; see Stanley (2008), Callot and Paldam (2011), Stanley and Doucouliagos (2014), Paldam (2013, 2015), Reed et al. (2015).

[4] Google Scholar gave 2.99 million hits on 'publication bias' on the 2/3–15. The more narrow term 'sponsor interests' gave 0.21 million hits. Many of these hits refer to studies applying various tests that show that a literature suffers from such bias. Most of the tests have been done in medicine.

level two choices have to be made. They often require judgment that may be influenced by priors and incentives.[5]

Sections 2 and 3 introduce the two levels of meta-analysis. Section 4 reports my impressions of the main findings from many meta-studies. It tells some rather persistent stories about empirical economics. Even when most of these stories are unsurprising, they are still painful to face for the average economist. Section 5 concludes.

2 Introducing the tools at level one: The funnel and the FAT-PET MRA

The coded β-literature is a set of N estimates (b_i, s_i), where b_i is the (standardized) estimate, and s_i is its standard error. From each (b_i, s_i) follows the estimates of the precision $p_i = 1/s_i$ and the *t*-ratio $t_i = b_i/s_i$. Overlined variables are (unweighted arithmetic) means over the N-set: $\overline{b}$, $\overline{s}$, etc. In meta-estimates the N-set is often referred to as the *primary* estimates. The presentation from now uses the simplifying assumptions made in Table 1.

The analyst always starts by a look at the distribution of the N-set. The most telling version of the distribution is ***the funnel***, which is the (p_i, b_i)-scatter plot. It should be narrow at the high precision top and broad at the low precision bottom. Funnels should be symmetrical, as the expected estimate of b_i should be independent of p_i.

Figure 1 shows an example of a typical funnel. The literature that has generated Figure 1 is based on data-samples with up to 1,000 observations, so the primary estimates have high precisions.

Some funnels have more than one top indicating that the literature on β is heterogeneous. In this case the analysis must start by identifying the tops. They may come from different data sets, or they may point to a strong, partly omitted

---

[5] This is illustrated by Doucouliagos and Paldam (2008, 2013) and the critical study by Mekasha and Tarp (2013). The results at level one are the same, while the results at level two differ as suggested by the incentives of the authors, notably the critics.

Table 1. Four assumptions for ease of presentation

(1)	The parameter of interest β is the effect of x on y, $\beta = \partial y / \partial x$.
(2)	Most researchers believe that $\beta > 0$, and it is actually true.[a]
(3)	The sign is not enough for the decision makers in the field.
(4)	This has caused the β-literature, with N estimates b of β.

Note a): If the profession has got the sign on β wrong that should lead to a two-topped funnel, where the second (with the true sign) starts in a few iconoclastic papers that are difficult to publish. But then, as the second top becomes clearer, it will show up in the funnel.

Figure 1. A funnel, covering a literature reporting $N = 366$ estimates
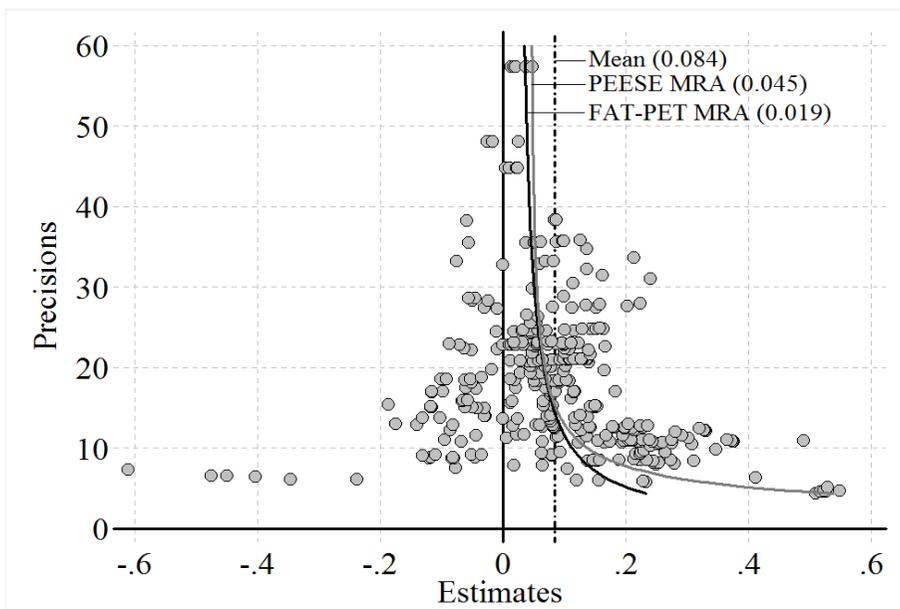
Note: The estimates are made comparable by a conversion into partial correlations. Thus, they are scaled from -1 to $+1$. The funnel is Figure 2b in Doucouliagos and Paldam (2015). The FAT-PET and PEESE are discussed below. They give the two curves $p_i = \beta_F / (b_i - \beta_M)$ and $p_i = \sqrt{\beta_F / (b_i - \beta_P)}$ respectively.

variable. Economists trying to estimate *β* like to believe that they look for a 'deep' parameter, but estimates surely differ between samples. To account for sample heterogeneity, estimating models include *ceteris paribus* controls. Such controls are not perfect, so they result in (additional) noise around the true estimate.

We would like to believe that this noise is white, so that the funnel has one top only, and that this top is in the middle of the distribution, so that the funnel is symmetric. Most *b*s in the typical *N*-set are statistically significant. Fanelli (2010) studies a large sample of typical papers in different sciences. In economics 87% of the papers confirm the thesis presented at the 5% level of significance.[6] The high *t*-ratios suggest that the width of the funnel, as measured by the standard deviation of the *N*-set is small. Consequently, ideal funnels should be one-topped, symmetric and lean.

Most funnels actually have one top only, as do Figure 1. This is the case considered from now on. However, empirical funnels are often asymmetrical and always amazingly wide (relative to the t-ratios of the estimates):

The most analyzed asymmetry is due to censoring of results that are 'wrong' by economic theory, moral/political beliefs, or sponsor interests. Authors, referees and editors all dislike 'wrong' results. And they are actually rarer than they should be if the funnels were symmetrical. Under the assumptions of Table 1 the funnel will miss most of the negative tail that should occur by symmetry. Thus, published estimates have a ***censoring bias***, making the mean result too big. However, the problems may be bigger than that.

Once the data are in the computer, it costs next to nothing to run regressions. Hence, most researcher cannot help running a great deal more than they can possibly publish. The rational researcher will surely selects the 'best' estimate based on its fit and size as modeled by economic theory. Paldam (2015) considers such selections for researchers with different preferences. They give a rather robust ***rationality bias***.

---

[6] More refined results are given in Brodeur et al. (2013). It covers all reported estimates in a large sample of papers in top journals. It finds that results cluster just above significance, while few results are just below.

To handle this situation, the meta-analyst runs MRAs,[7] meta-regression analyses, which are regressions on estimated regression coefficients. The main MRA is the FAT-PET:[8]

(1a) $b_i = \beta_F\, s_i + \beta_M + u_i$ or   (1b) $b_i = \beta_F\, /p_i + \beta_M + u_i$

$ \rightarrow \beta_M$ for $p \rightarrow \infty$

A division by s_i gives   (1c) $t_i = \beta_F + \beta_M\, p_i + v_i$

β_F is the FAT, *funnel asymmetry test*. If $\beta_F \neq 0$, the funnel is asymmetric. β_M is the PET, *precision estimate test*, which most practitioners term the **PET meta-average**. The noise terms are u_i and v_i. When the funnel is asymmetric, the FAT is non-zero, and the PET differs from the mean.[9]

The logic of the FAT-PET is that the low precision estimates scatter most, so they are more likely to be censored. The variables of the funnel are used in (1b), which is a curve that may be far from β at small precision but converges to $\beta_M \approx \beta$ as p rises. The path of convergence is hyperbolic in (1b). This appears reasonable, but somewhat arbitrary. Stanley and Doucouliagos (2014) experiment with a squared version termed the PEESE MRA: [10]

(2) $b_i = \beta_F\, s_i^2 + \beta_P + u_i$ It can be written in the same three versions as (1).

———————————————————

[7] The paper by Reed et al. (2015) discusses alternatives to the two MRAs. The most simple is to use WLS-regression with the precisions as the weights. It often works surprisingly well.

[8] From Stanley (2008). He terms $\beta_0 = \beta_M$ and $\beta_1 = \beta_F$. I like terms making the variables easier to remember.

[9] Formulation (1c) is used for estimation. Estimates within the same paper tend to cluster, so that clustered standard errors should be used. They are the same as the non-clustered standard errors if there is no paper-effect. In the typical case the clustered standard errors are 20–30% larger.

[10] PEESE is an acronym for Precision-Effect Estimate with Standard Errors. The PEESE is made to handle censoring, and Stanley and Doucouliagos (2014) shows that the PEESE is actually a bit better in that case. However, Paldam (2015) shows that the PET is substantially better in handling the biases due to the rational behavior of economists. As the meta-analyst does not know how asymmetry is generated the FAT-PET is likely to be better.

The PET and the PEESE are normally close, and they typically reduce the bias by more than 90%. Each of them is only exactly equal to β in rather special cases. In other cases they over- or undershoots a little, but we do not know if it is the one or the other. The two MRAs are shown on Figure 1.

Another analysis that should be done with the *N*-set is a study of the path over time, τ:

(3) $b_i = b_i(\tau)$, where τ is 'time' measured as the order of publication

Equation (3) allows the analyst to analyze if the primary estimates have trends and structural breaks. Most papers are announced as an improvement over previous ones, and some really are. They ought to give structural breaks in the $b_i(\tau)$-series. Such breaks should be controlled for in the final assessment of the best (current) estimate of β.

The research process for any paper demands many choices such as: Should control z be included in the estimated equation? Which year should the data sample start? Should the TSIV-estimator be used? We like to believe that all such choices are based on objective criteria, but an element of 'judgment' inevitably enters into the choice. This is precisely where priors and incentives are at play. If (1) or (2) shows publication bias, it means that judgment is affected by a ***choice bias*** that works systematically in one direction.

3   Introducing the tools at level two: Adding the moderator matrix

At level two of the meta-analysis the data of the *N*-set is supplemented by an (*M x N*) moderator matrix, *Q*. It tells how the primary estimates are reached and gives relevant information about the author and the publication outlet of the paper. Each b_i gives a row, q_i, of the *Q*-matrix with *M* coded moderators, and each moderator gives a column, q_j, of the *Q* matrix. Four typical examples show how column q in the *Q*-matrix is coded.

(Ex1)   Is control variable z included in the estimate of b_i?

(Ex2)   Does the estimator for b_i adjust for simultaneity?

(Ex3)   Does the research for b_i have a sponsor who is interested in the size of the estimate?

(Ex4)   What is the impact factor of the journal, where b_i is published?

Here (E1) to (E3) are coded as a qualitative binary variable, where $q_{ji} = 1$ if the answer is 'yes' and otherwise 0, while (E4) is a quantitative variable. Often M is as large as 50, so the coding of the moderator matrix is a major undertaking. Once done, it allows the researcher to ask many interesting questions to the literature. This is done by *augmenting* equation (1) with the relevant q-column transposed as a regressor. The augmented FAT-PET MRA is:

(4) $b_i = \beta_F\, s_i + \beta_M + \alpha\, q_i + u_i$

The estimate of α in (4) is unbiased in three cases: (i) No publication bias was found at level one. (ii) z is exogenous to the research process. (iii) The estimate of α must also be unbiased if the inclusion of q_i does not change the estimate of β_F (and thus β_M). However, if it does – and especially if it reduces β_F – it suggests that the selective inclusion of q_i is one of the factors generating the publication bias.

Some examples of (ii) are: (a) Regional dummies – it is no choice of the author if a country is Latin American. So (4) can be used to see if β is different in Latin America. (b) Some field has sponsors who are interested in certain results. If they undertake the sponsorship before the research starts the contact is exogenous. Dummies for such sponsors can be used to see if they produce consistently different results.

As mentioned it appears that priors and incentives often lead to bias; see Doucouliagos and Stanley (2012) and Paldam (2013, 2015). When publication bias is found, it follows that the literature has a dominating choice bias, so that choices involving judgment are influenced by their effect on β. Thus, the estimate of α, β_F and β_M in (4) are biased.

Think of a variable z that is included in some of the estimating models. A publication bias – as detected by (1) – implies that z is more likely to be included when it influences β in the 'right' (positive) direction. This will bias the estimates

of (4) so that β_F goes down and β_M goes up, and the estimate of α will be too large.[11]

The meta-analyst often estimates a version of that augments (1) with all or most of the M coded moderators:

(5) $b_i = \beta_F \, s_i + \beta_M + [\alpha_1 \, q_{1i} + ... + \alpha_M \, q_{Mi}] + u_i$

When (1) shows a bias, many of the estimated coefficients from (5) are biased. It can be showed that β_F becomes zero and β_M becomes the mean for a full augmentation. However, it is difficult to know how the biases are distributed across the estimated αs. Nevertheless, the estimate (5) still provides qualitative suggestions about the αs: It helps us to point to more or less important variables in the model, but thanks to the biases these suggestions needs further research.

It is a major discussion if (4) and (5) can be amended to give unbiased estimates of the αs, β_F and β_M even when (1) detect a bias. This will not be discussed at present.

4   Common findings in meta-studies: A few observations

The following observations are based on my impressions from reading and listening to the presentation of many meta-studies.[12] It has already been mentioned that all studies find excess width of funnels and most find asymmetries that often can be explained as publication bias.

One of the key subjects analyzed is 'progress'. Most of the primary papers in the β-literature present an *innovation* in the model or the estimator. It then proceeds to show that the innovation is empirically 'better'. Thus, the paper claims that it pushes the frontiers of research in the field making the 'old' literature

_____

[11] Also, it is possible that z is included only when it is significant. It typically influences the estimate of β more when it is significant. When (4) is estimated, it will show the result as if the effect of z is always what it is, when it is significant. This will bias the estimates of α, β_F, and β_M.

[12] Till now few attempt has been made to systematically summarize the findings of the many meta-studies in economics; see Doucouliagos and Stanley (2012). See also Nelson and Kennedy (2010).

obsolete.[13] After some time the innovation has been used in enough papers, so that it can be tested if it does make a significant difference in the results. This is done by eq. (4) with a q-dummy for papers using the innovation. The estimate of α shows if the innovation is significant. Often it is not. This means that the paper introducing the innovation exaggerated its importance. Researchers should work at the frontline, so insignificant innovations are a problem.

We all believe that the quality of papers is crucial and that top journals publish papers of a higher quality. Therefore the impact factor has often been used as the q-variable in (4), but I have yet to see a meta-study where this variable turns significant. My interpretation is that papers in top journals contain more innovation, while papers in other journals contain more replication.[14] Thus, top journals should report results that are more variable but not necessarily larger. Several analysts have reported that they have found signs supporting the variability idea, but till now such reports have been oral only.

Most economists also regard the right choice of estimator as very important, and spent a lot of time on mastering and applying state-of-the arts estimators. Models should be estimated just right, and researchers should demonstrate high technical skills to publish well. Many meta-studies have included estimator dummies. They normally get small coefficients which are often insignificant. Thus, these studies show that little of the big variation between studies is explained by the choice of estimators.[15] This suggests that the benefit-costs ratio from getting models and data right are greater than from getting estimators right. This points to some misallocation of talent in our field!

_____

[13] Here the complex phenomenon called 'fashion' also matters.

[14] As meta-analysis looks at replicability of results, it is crucial that it includes all published results.

[15] My own experience is that when I spent considerable efforts on estimators (or had an econometrician as co-author) it did increase the publication chance, and it gave a nice feeling in the belly to know that everything had been done, but the results did not really change!

5 Conclusion on the conventional mold of papers

It is a convention in economics to cast empirical papers in a mold as if they were done in three stages: (s1) An intuition that leads to a theory. (s2) The theory is operationalized into an estimation model for a certain data set. (s3) The model is estimated, and it is shown to confirm the theory. This is a convention, but the mold implies a research strategy.

It is well-known that the conventional strategy invites moral hazard, and it has often been criticized.[16] A number of remedies have been proposed such as robustness tests, out-of-sample predictions, etc. These remedies have often been used, but nothing prevents authors from including a dozen robustness tests and an out-of-sample data-set in the search and research efforts.

This paper mold has withstood the critique remarkably well as it has a great advantage. It is doable and leads to publications. Even when it is well-known that it is a strategy of make-believe, it has proven difficult to find an equally 'useful' alternative. Thus we have to live with the conventional mold in empirical economics.

In all sciences, results need replication to be credible, but due to the problems mentioned results in economics need a considerable amount of replication and this is precisely where meta-analysis is needed. In addition, it has another advantage: From the distribution of the results in a literature it can, in many cases, estimate a meta-average that is much closer to the true value than the average result.

Acknowledgements This note looks at research in economics done by regression analysis only. Thus, it neglects the important literature on benefit transfers. I am very grateful for the fine comments by the discussants/referees to my introduction. I have corrected the errors detected and included the most important of the comments, but some will have to wait for subsequent work.

---

[16] A classic paper in this respect is Leamer (1983), see also Summers (1991) and Paldam (2013). The problems are many: It is rare indeed that one and only one estimation model follows from the theory. This inevitably requires choices as already mentioned: Control variables and instruments in two-stage estimations have to be chosen. So have data samples, and only a fraction of the estimates made can be published. Loops from (s3) back to (s2) and (s1) are common, etc.

References

Brodeur, A., Lé, M. Sangnier, and Y. Zylberberg (2013). Star wars: The empirics strike back. IZA discussion paper, No. 7268 http://ftp.iza.org/dp7268.pdf.

Callot, L., and M. Paldam (2011). Natural funnel asymmetries. A simulation analysis of the three basic tools of meta-analysis. *Research Synthesis Methods* 2 (2): 84–102.

Card, D., and A.B. Krueger (1995). Time-series minimum-wage studies: A meta-analysis. *American Economic Review* 85 (2): 238–243. http://www.jstor.org/stable/2117925?seq=1#page_scan_tab_contents

Doucouliagos, C. (H.) (1995). Worker Participation and Productivity in Labor-Managed and Participatory Capitalist Firms: A Meta-Analysis. *Industrial and Labor Relations Review* 49 (1): 58–75. http://www.jstor.org/stable/2524912?seq=1#page_scan_tab_contents

Doucouliagos, H., and M. Paldam (2008). Aid effectiveness on growth. A meta study. *European Journal of Political Economy* 24 (1): 1–24. http://www.sciencedirect.com/science/article/pii/S0176268007000614

Doucouliagos, H., and M. Paldam (2013). The robust result in meta-analysis of aid effectiveness: A response to Mekasha and Tarp. *Journal of Development Studies* 49 (4): 584–87. http://www.tandfonline.com/doi/abs/10.1080/00220388.2013.764595#.VSZV_MkrFnU

Doucouliagos, H., and M. Paldam (2015). Finally a breakthrough? The recent rise in the size of the estimates of aid effectiveness. For Mak Arvin, ed., *Handbook on the Economics of Foreign Aid*. Edward Elgar

Doucouliagos, H., and T.D. Stanley (2012). Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys* 27 (2): 316–339. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-6419.2011.00706.x/abstract

Fanelli, D. (2010). ''Positive'' results increase down the hierarchy of the sciences. *Plos ONE* 5 (4): 1–10. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0010068

Hunter, J.E., and F.L. Schmidt (1990, 2004). *Methods of Meta-Analysis: Correcting error and bias in research findings*. New York: Sage.

Jarrell, S.B., and T.D. Stanley (1990). A meta-analysis of the union-nonunion wage gap. *Industrial and Labor Relations Review* 44 (1): 54–67. http://www.jstor.org/stable/2523429?seq=1#page_scan_tab_contents

Johnson, R., J. Rolfe, R. Rosenberger, and R. Brouver, eds. (2015). *Benefit transfer in environmental and resource values: A handbook for researchers and practitioners.* Dordrecht, Springer.

Leamer, E.E. (1983). Let's take the con out of econometrics. *American Economic Review* 73 (1): 31–43. http://www.jstor.org/stable/1803924?seq=1#page_scan_tab_contents

Mekasha, T.J., and F. Tarp (2013). Aid and growth. What meta-analysis reveals. *Journal of Development Studies*, 49 (4): 564–583. http://www.tandfonline.com/doi/abs/10.1080/00220388.2012.709621#.VSZZ1skrFnU

Nelson, J., and P. Kennedy (2010). The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environmental and Resource Economics* 42: 345–377. http://link.springer.com/article/10.1007%2Fs10640-008-9253-5

Paldam, M. (2013). Regression costs fall, mining ratio rise, publication bias looms, and techniques get fancier. Reflections on some trends in empirical macroeconomics. *Econ Journal Watch* 10 (2): 136–156. http://econjwatch.org/articles/reflections-on-some-trends-in-empirical-macroeconomics

Paldam, M. (2014). How do partly omitted control variables influence the averages used in meta-analysis in economics? *Empirical Economics*, forthcoming, p.t. at: http://www.martin.paldam.dk/Papers/Meta-method/POCtheory.pdf.

Paldam, M. (2015). Simulating an empirical paper by the rational economist. Aarhus. http://www.martin.paldam.dk/Papers/Meta-method/Simulating-rationality.pdf.

Reed, W.R., R.J.G.M Florax, and J.A. Poot (2015). A Monte Carlo analysis of alternative meta-analysis in the presence of publication bias. *Economics: The Open-Access, Open-Assessment E-Journal,* forthcoming, p.t. at: http://www.economics-ejournal.org/economics/discussionpapers/2015-9

Stanley, T.D. (2008). Meta-regression methods for detecting and estimating empirical effect in the presence of publication bias. *Oxford Bulletin of Economics and Statistics* 70(1): 103–127.

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

Stanley, T.D., and H. Doucouliagos (2014). Meta-regression approximation to reduce publication selection bias. *Research Synthesis Methods* 5 (1): 60–78. http://onlinelibrary.wiley.com/doi/10.1002/jrsm.1095/pdf

Summers, L.H. (1991). The scientific illusion in empirical macroeconomics. *Scandinavian Journal of Economics* 92 (2): 129–148. http://www.jstor.org/stable/3440321?seq=1#page_scan_tab_contents

Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

The Editor