Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
No. 2015-3 | January 19, 2015 | http://www.economics-ejournal.org/economics/discussionpapers/2015-3

Meta-Analysis in a Nutshell: Techniques and General Findings

Martin Paldam

Abstract
The purpose of this note 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 results is normally amazingly wide given their *t*-ratios. Little of the variation can be explained by the quality of the journal 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)

(Submitted as Survey and Overview)

JEL B4 C2
Keywords Meta-analysis; selection of regressions; publication bias

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

Citation Martin Paldam (2015). Meta-Analysis in a Nutshell: Techniques and General Findings. Economics Discussion Papers, No 2015-3, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2015-3

Received January 14, 2015 **Accepted as Economics Discussion Paper** January 15, 2015 **Published** January 19, 2015

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 considers 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 they converge to a meta-average we might see as the truth?

Q2: Do they suffer from biases that should be corrected for?

Q3: Can we identify the main innovations of relevance for the convergence?

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 substantially from the ***mean***. These steps can, in principle, be done in one way, 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.[2] In medicine an experiment is an expensive clinical trial, while it is a cheap regression in economics. This has strong effects on the number of experiments done and the fraction reported. Hence, economics required a development of new tools. When they became available (Stanley 2008),[3] it caused a wave 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,[4] 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

2. The first studies were Jarrell and Stanley (1990), Doucouliagos (1995) and Card and Krueger (1995).
3. The textbook by Hunter and Schmidt (1990, 2004) covers the meta tools in medicine, while the textbook by Stanley and Doucouliagos (2012) covers economics.
4. 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, 2014b), Reed *et al.* (2014).

difference between the two averages is an estimate of the ***publication bias***. The literature shows that such biases are common in most fields.[5] This has been developed into the theory of the research and publication process. The 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. 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. Meta-analysis takes the view that economists are humans. This should not be controversial, but I think that anyone who has received a referee reports on a study finding such bias knows that it is – it is something we 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 level two choices have to be made. They often require judgment that may be influenced by priors and incentives.[6]

Sections 2 and 3 are written as introductions to 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 economists. 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 estimate,[7] and s_i is its standard error. From each (b_i, s_i) follow the precision $p_i = 1/s_i$ and the t-ratio $t_i = b_i/s_i$. Underlined variables are (unweighted arithmetic) means over the N-set: $\underline{b}$, $\underline{s}$ etc. The presentation from now uses the simplifying assumptions made in Table 1.

5. Google Scholar gave 3.9 million hits on 'publication bias' on the 24/10-14. The more narrow term 'sponsor interests' gave 0.23 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.
6. This is illustrated by Doucouliagos and Paldam (2008, 2013) and Mekasha and Tarp (2013). The results at level one are the same, while the results at level two differ in the way suggested by the incentives of the authors.
7. The *b*s should be standardized to the same scale, i.e., all *b*s may be elasticities or at least partial correlations.

Table 1. Four assumptions for ease of presentation

(1)	The parameter is the effect of x on y, $\beta = \partial y/\partial x$
(2)	Most researchers believe that $\beta > 0$, and it is actually true,
(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

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. As the expected estimate of b_i should be independent of p_i funnels should be symmetrical.

Some funnel has two (or more) tops 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 variable. Economists trying to estimate β like to believe that they look for a 'deep' parameter, but we also know that estimates differ between samples. To account for sample heterogeneity, estimating models include *ceteris paribus* controls. Such cp-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.

Most funnels actually have one top. This is the case considered from now on. Thus, we expect that the funnel is one-peaked and symmetric. It should look much like a Xmas tree! As $\beta > 0$, most estimates in the N-set are positive, but it is likely that some negative estimates occur especially at low levels of precisions.

Most of the bs in the N-set are statistically significant. Fanelli (2010) studies a large sample of typical papers in different sciences.[8] In economics 87% of the papers confirm the thesis presented at the 5% level of significance. The high t-ratios suggest that the coefficient of variation of the N-set is small. Consequently, ideal funnels should also be lean.[9]

One of the first observations from meta-analysis is that funnels are amazingly wide, i.e., they have ***excess width*** compared to ideal ones. Also, most funnels are ***asymmetrical*** in ways suggested by likely priors and incentives: Results that are 'wrong' according to economic theory, moral/political beliefs, or sponsor interests are rarer than they should be if the funnels were symmetrical. Under the assumptions of Table 1 the funnel typically misses most of the negative tail that should occur by symmetry. Thus, the mean result is too big.

8. 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.
9. A simple example shows the orders of magnitude. Imagine the true value $\beta = 1$ and that all bs in the N-set have a t-ratio of 2, i.e., $2 = \underline{t} = \underline{b}/\underline{s} = 1/s = \underline{p}$. Here the funnel is twice as high as it is wide!

We all know why this occurs. 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, even when the typical paper publish a dozen estimates to show the robustness of the main estimate. The rational researcher will surely selects the 'best' estimate based on its fit and size. Paldam (2014b) considers different selection rules: The two extremes are selection by fit alone and by size alone. They give almost the same bias. Consequently, it is still the same in the realistic case where the selection is by a combination of fit and size.

To handle this situation, the meta-analyst runs MRAs, meta-regression analysis, which are regressions on estimated regression coefficients. The main MRA is the FAT-PET: [10]

(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, or 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.

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.[11] Stanley and Doucouliagos (2014) experiment with a squared version termed the FAT-PEESE MRA:

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

The PET and the PEESE are normally close, and they both reduce the bias by more than 95%. Each of them is only exactly equal to β in rather special cases. The 'residual' biases in other cases are either positive or negative, but we do not normally know if it is one or the other.

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

The purpose of (3) is to allow the analyst to see if trends or structural breaks occur over time. Most papers are announced as an improvement over previous ones, and some really are. They

10. From Stanley (2008). He terms $\beta_0 = \beta_M$ and $\beta_1 = \beta_F$. I like terms making the variables easier to remember.
11. 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.

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: Should control z be included or not? Should the data set start at year A or go back to year B? Should the TSIV-estimator be used? etc. 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. Introduction 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 estimates are reached and gives relevant information about the author and the paper where the estimates are published. Each b_i gives a row, q_i, of Q with M coded moderators, and each moderator is a column, q_j, in 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 result?
(Ex4) 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 estimating equation (4) with the relevant q-column transposed as a regressor:

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

The estimate of α in (4) is unbiased in two cases: (i) No publication bias was found at level one. (ii) z is exogenous to the research process.

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. 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, 2014b). 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 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 of course the estimate of α will be too significant.[12]

The meta-analyst often estimates a version of (4) that contains all M coded:

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

When (1) shows a bias, we know that many of the estimated αs from (4) are biased, and it is difficult to know how the biases are distributed across the coefficients. In particular, the estimates of β_F and β_M in (5) make little sense. However, the estimate (5) still provides qualitative knowledge: It helps us to point to more or less important variables in the model.

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.[13] 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 papers present an *innovation* in models or estimators and show that it is empirically 'better'. Thus, the paper claims that it pushes the frontiers of research in the field making the 'old' literature obsolete.[14] 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. We all want to

12. 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, but the direction of the bias is unclear.
13. Till now only one attempt has been made to systematically summarize the findings of the many meta-studies in economics; see Doucouliagos and Stanley (2012).
14. Here the complex phenomenon called 'fashion' also matters.

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.[15] 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. 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.[16] This evidence points to some misallocation of talent in our field. It appears that the benefit-costs ratio from getting models and data right are greater than from getting estimators right!

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 is given to justify 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.[17] 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 advan-

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

16. 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!

17. 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.

tage. It is doable and leads to publications. Even when we all know that it is a strategy of make-believe, it has proven difficult to find an equally ‘useful’ alternative. We thus 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 additional, 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.

References:

Brodeur, A., Lé, M., Sangnier, M., and Y. Zylberberg (2013). *Star Wars: the Empirics Strike Back.* URL: 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): 238-43.

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.

Doucouliagos, H, and M. Paldam (2008). Aid effectiveness on growth. A meta study. *European Journal of Political Economy* 24 (1): 1-24.

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.

Doucouliagos, H., and T.D. Stanley (2012). Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity. *Journal of Economic Surveys* 27 (2): 316-39.

Fanelli, D. (2010). ‘‘Positive’’ Results Increase Down the Hierarchy of the Sciences. *Plos ONE* 5 (4): 1-10.

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.

Leamer, E.E. (1983). Let's Take the Con Out of Econometrics. *American Economic Review* 73 (1): 31-43.

Mekasha, T.J., and F. Tarp (2013). Aid and growth. What meta-analysis reveals. *Journal of Development Studies*, 49 (4): 564–83.

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): 136-56.

Paldam, M. (2014a). How do partly omitted control variables influence the averages used in meta-analysis in economics? URL: http://www.martin.paldam.dk/Papers/Meta-method/POCtheory.pdf.

Paldam, M. (2014b). Simulating an empirical paper in economics. Why publication bias is so common. URL: http://www.martin.paldam.dk/Papers/Meta-method/Simulating-pub-bias.pdf.

Reed, W.R., Florax, R.J.G.M., and J.A. Poot (2014). A Monte Carlo analysis of alternative meta-analysis in the presence of publication bias. MEAR Colloquium in Athens 2014.

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): 103-27.

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.

Summers, L.H. (1991). The Scientific Illusion in Empirical Macroeconomics. *Scandinavian Journal of Economics* 92 (2): 129-48, and Hylleberg, S., and M. Paldam, eds. *New Approaches to Empirical Macroeconomics*. Cambridge, U.K.: Blackwell.