School Working Paper

ECONOMICS SERIES

SWP 2012/4

The robust result in meta-analysis of aid effectiveness:
A response to Mekasha and Tarp

Hristos Doucouliagos and Martin Paldam

www.deakin.edu.au

The difference is **Deakin University.**

The robust result in meta-analysis of aid effectiveness:

A response to Mekasha and Tarp

Hristos Doucouliagos,[1] Deakin University, Melbourne, Australia

Martin Paldam,[2] Aarhus University, Denmark

Abstract:

Our study, Doucouliagos and Paldam (2008), has recently been critically discussed by Mekasha and Tarp (2011). In this paper we show that contrary to what they state, their study validates our basic analysis: Both papers confirm that the literature has shown that aid is of little economic importance in generating growth. M&T find some random coding errors with virtually no effect on the basic results. Furthermore, we discuss some methodological disagreements and show that their choice of the random effects model is not appropriate for the problem at hand, and that the way they use multiple MRAs contradicts the robust results reached at the basic analysis.

---

1. School of Accounting, Economics and Finance, Deakin University, 221 Burwood Highway, Burwood, Victoria 3125. Australia. Mail: chris.doucouliagos@deakin.edu.au. Meta home page: http://www.deakin.edu.au/meta-analysis.
2. Department of Economics and Business, University of Aarhus, Bartholins Allé 10, DK 8000 Aarhus C, Denmark. Mail: mpaldam@econ.au.dk. Our project is at the URL: http://www.martin.paldam.dk. Click on: 'working papers', 'meta project'

1.   Introduction: A project, a critique and a response

The most important test of any empirical result is whether it survives independent replication, *i.e.,* replication by others on new data. Meta-analysis is a systematic analysis of the robustness of empirical results to replications. During the past six years we have made a major meta-study of the large Aid Effectiveness Literature (AEL); see D&P06, D&P08, D&P09, D&P10 and D&P11.

In D&P08 and D&P11 we show that *aggregate* aid has no robust effect on growth – what looks like a small positive effect is, if fact, mostly a selection bias. Aid ineffectiveness is a tragic result, and we have stressed positive and promising *disaggregate* results. In D&P08 we note that it appeared that aid had a positive effect on growth in Asia, while in D&P11 we note that it appeared that some *components* of aid had a positive effect on growth.

A recent meta-study by Mekasha and Tarp (2011, M&T) is critical of D&P08. M&T criticize our dataset, our measure of the effect of aid on growth, our use of the fixed effects estimator and our treatment of heterogeneity. They conclude that the AEL is free of selection bias and that aid is effective in generating growth. Their critique is comprehensive and largely invalid:

(1) M&T find random errors in a small proportion of the coded cells, and this has virtually no effect on the results; (2) our measure of effect size is indeed appropriate, even in the context of interaction terms; (3) the method used by M&T to reject publication bias is incorrect. We show that accommodating heterogeneity does not alter our conclusions; (4) their support for the random effects estimator is ill advised for the data at hand; and (5) M&T misinterpret their own results; they actually show that aid has little effect on growth.

The central message of our meta-studies is not publication bias. Rather, it is the absence of robust evidence indicating aid effectiveness as a cross-country experience. Also, the AEL shows *declining* aid effectiveness over time (see D&P08, 09, 11 and section 3.4 below). If M&T are correct that the AEL has no publication bias, then we are forced to accept that the aid industry has un-learning by doing. M&T fail to note this critical fact and offer no explanation of how such a process could occur.

Section 2 is a short introduction to meta-analysis. Section 3 looks at the data-issues raised in M&T. The recoding of M&T validates our results as much as anybody could wish. Section 4 discusses more substantial methodological issues. Section 5 deals with heterogeneity through multiple meta-regression analysis (MRA). We show that M&T claim more than their MRA can actually provide. Indeed, a simple re-specification of the MRA reveals only selection bias and the absence of aid effectiveness. Finally, section 6 concludes.

2. The AEL, bias and meta-regression analysis

The AEL is a subset of the growth regression literature. It is well-known to have a robustness problem: It is far too easy to come across a significant coefficient to growth by mining the many possible controls.[3] Also, aid effectiveness is an emotional issue, with large economic interests. Section 2.1 deals with the robustness problem while section 2.2 discusses the issue of priors and biases. Section 2.3 is a brief introduction to meta-regression analysis.

2.1 The AEL: Mining the control set causes a robustness problem

The AEL arises from the well-known *zero-correlation* fact that aid and growth are uncorrelated. The effect is modeled by a relation explaining growth by aid and a *set of control*s. The basic AEL relation for estimating aid effectiveness β is:

$$g_{it} = \alpha + \beta h_{it} + [\lambda_1 z_{1it} + ... + \lambda_n z_{nit}] + u_{it}, \qquad \text{(1a) analyzed in D\&P08 and here}$$

$$g_{it} = \alpha + \beta h_{it} + \gamma h_{it} z_{0it} + \delta z_{0it} + [\lambda_1 z_{1it} + ... + \lambda_n z_{nit}] + u_{it} \qquad \text{(1b) analyzed in D\&P10 see section 4.2}$$

where g is growth, h is the aid share, the square brackets contains the controls, the residuals are u and the subscripts i and t are for country and time. (1b) contains an interaction term, z_0.

The zero-correlation result means that $\beta \approx 0$ in (1a), if the []-term is deleted. Consequently, the finding of a positive β depends on the choice of the control set.[4] If a sufficiently large set of potential controls is mined, the researcher generates a wide range of estimated aid effectiveness coefficients to choose from. This produces excess variation (heterogeneity) in the reported estimates. Thus, a major part of the meta-analysis is to study this variation and to quantify the effect that specification has on estimates of β, and vice versa: How do priors about the sign and statistical significese of β influence the results reported?

The data set for aid and growth using 5-year averages has now reached about 1,100 observations and the AEL has reported more than 1,700 estimates of β. The published estimates have the iceberg property – what is reported is only a small fraction of what has been generated. It is reasonable to assume that *at least* 30,000 regressions have been run on the 1,100 observations; the AEL-research community has thoroughly mined the data. Great ingenuity has been applied to reach the 1,700 estimates: About $M = 60$ variables have been tried in many combinations for the control set,[5]

---

3. On robustness in growth regressions see *e.g.* Durlauf, Johnson and Temple (2005).
4. In the term *control set* we include data 'fine tuning' and the estimator used.
5. The reader may compare the control sets used in Hansen and Tarp (2000) and in Dalgaard, Hansen and Tarp (2004).

the variables have been averaged, they have been lagged, they have been interacted and a whole set of estimators has been applied. Meta-analysis is a way to study the *stability of the effect* of the variable of interest, which in our case is aid effectiveness.

It is important that the literature has made a thorough search of the possible models. But then, of course, one has to pause from time to time and consider what has been found. This is the main purpose of meta-regression analysis.

2.2   Two types of publication bias: Polishing and censoring

Once the literature has the iceberg property, one has to ask how the selected estimates are chosen. It appears inevitable that the selection is influenced by researchers' priors (and those of referees and editors). Two types of publication bias are commonly detected by meta-studies:

A polishing bias occurs if selection is influenced by the t-ratios. Most of us have a prior for clear results with high t-ratios. This preference is common and it has the effect of inflating any individual effect size that is calculated from t-ratios, such as the partial correlations used by D&P08 and M&T, but it may not necessarily influence the average result.

A censoring bias occurs if the selection is caused by widespread priors about the size and sign of β. In the AEL two such priors are likely: (i) Almost all of us want to find a positive effect from aid for moral and political reasons.[6] (ii) Aid had a turnover of US$ 129 billion in 2010 so it is a major 'industry'. It is a main sponsor of the AEL, and the industry naturally wants research to show that aid works.[7] As (i) and (ii) are widespread and point in the same direction, it is likely that the AEL has a strong bias: The profession is reluctant to publish negative estimates of β.

2.3   The two levels of meta-analysis

The seminal work on meta-regression analysis in economics is Stanley and Jarrell (1989). The present front-line methodology was developed by Stanley (2008).[8] The final submission date of the D&P08 to the *EJPE* was May 2007. We have used the improved techniques as they were developed, but the technical progress in meta-analysis does not alter the central conclusions, as shown in sections 3 and 4 below.

---

6. The bias can, of course, run the other way, *e.g.* libertarians trying to show that aid reduces growth by inflating the public sector and Marxists trying to show that aid contributes to capitalist dependency – this explains the negative points on Figure 2a. However, as shown in the text, the AEL has a net bias in favor of positive coefficients.
7. The finding of aid ineffectiveness is a threat to aid budgets, and hence the aid industry dislikes such results, as predicted by the theory of bureaucratic behavior (see Mueller 2003, chapters 16-17 for a lucid survey).
8. Most of the methods for economics are developed as notes to the approximately 500 meta-studies done in economics, but a few methodology papers do exist, *e.g.* Stanley and Doucouliagos (2010) and Callot and Paldam (2011). Stanley and Doucouliagos (2012) is a new textbook specific to economics.

Meta-analysis is conducted at two levels. The first level involves the literature search, data coding and the estimation of the basic FAT-PET MRA:

$$effect_{ij} = \beta_0 + \beta_1 SE_{ij} + \varepsilon_{ij} \tag{2}$$

where $effect_{ij}$ is the ith estimated effect of aid of growth from study j and SE is the estimate's standard error. Eq. 2 is estimated using weighted least squares, with precision ($1/SE$) used as weights,[9] so that more precise estimates are given greater weight. The FAT tests H0: $\beta_1 = 0$. This is a test for censoring (D&P08 and D&P09). The PET tests H0: $\beta_0 = 0$. This is a test for the existence of an effect of aid on growth corrected for selection bias; this is the meta-average of the effect of aid on growth, β from Eq. 1. The FAT-PET-MRA is an objective test. Once the literature is collected and coded there is *one and only one* FAT-PET MRA to run; Eq. 2. Section 3 shows that M&T find virtually the same result as we did.

The second level of analysis expands Eq. 2 to accommodate heterogeneity in β:

$$effect_{ij} = \beta_0 + \sum \beta_k Z_{kji} + \beta_1 SE_{ij} + \varepsilon_{ij} \tag{3}$$

where Z is a vector of controls that 'moderate' the reported estimates. Eq. 3 is a multiple MRA that can be used for two tasks: (i) identify the key factors that result in excess variation in reported estimates and (ii) adjust the estimated meta-average for partly omitted variables bias. (i) is much easier to do than (ii). The focus of the multiple meta-regression analysis in D&P08 was primarily (i). M&T use it for task (ii). However, it is not a straightforward matter to separate the sources of genuine variation from the variation created by the research process itself.

The first question to ask of a control is if it is relevant or irrelevant to the size of β. If a control is correlated with the aid share h, it changes the estimate of β if it is included. Hence it is relevant. If a control is uncorrelated with the aid share h, it is an irrelevant control. Researchers need to make an assessment as to which variables in the moderator matrix can be taken to reflect genuine differences in the effect of aid on growth and which merely reflect specification bias and differences in the research process. Once a substantial publication bias is found it is clear that the latter possibilities are important.

---

9. An *equivalent* approach is to divide *all* of Eq. 2 through by SE and estimate using OLS, so that the dependent variable becomes the t-statistic yielding $t_i = \beta_0 p_i + \beta_1 + u_i$, where β_0 is the meta-average, p is 1/SE and the FAT-test has H0: $\beta_1 = 0$. Unfortunately, many applications of the multiple MRA version of this model (see section 5 below) have misinterpreted the regression coefficients from this MRA. Hence, Stanley and Doucouliagos (2012) advocate the estimation of Eq. 2 using WLS; the interpretation of the MRA coefficients is then clear-cut.

3. M&T's critique: The non-issue of data and the FAT-PET MRA

Section 3.1 deals with the literature 'selection' and coding, where we have little disagreement with M&T.[10] We do not understand why they make this an issue. Section 3.2 examines the robustness to coding and estimators of the results from D&P08, while section 3.3 looks at the results from D&P11. Section 3.4 gives a simple graphic illustration of the issue of publication bias in the AEL.

To make studies comparable we need a common measure of the effect analyzed. Our project uses the partial correlation. This is calculated as: $r = t / \sqrt{t^2 + df}$. To help interpret the size of r we use Cohen's well-known guidelines: 0.1 is *small*, 0.3 is *medium*, and anything larger than 0.5 is *large* (see Cohen 1988, pp. 79-80). Tables 1, 2 and 3 below consistently report meta-averages of partial correlations ranging from 0.02 to 0.04.

3.1 Literature search and coding

We have tried very hard *not* to 'select' the literature base, but to be as inclusive as possible in our search for studies. Naturally, we had to use cut-off dates for the search. This was 1/1-2005 for D&P06, 08 and 09 and 1/1-2009 for D&P10 and 11. All publications in international scientific journals and all working papers from the top institutions quickly enter the searchable levels on the net. However, there are at least 5,000 institutions in the world that carry out some research in economics.[11] Unless they are cited, working papers from many less prominent institutions are difficult to find. Our bibliography has been published on the net (see Christensen *et al.* 2007 and 2010) and nobody has pointed to missing papers, though there must be some. M&T identify no missing papers, so they confirm the care of our search.

For D&P08, we coded over 100,000 cells.[12] It is undoubtedly important to do a careful coding, and we are grateful to M&T for making a second rechecking of the coding. We have now undertaken our own extensive third recoding of the dataset,[13] which includes an update of the D&P08 and D&P11 datasets. A coding/recoding is a labor intensive process and with two independent recodings it has required the use of a handful of research assistants. It is notoriously difficult to

---

10. The only disagreement has to do with the coding of aid interaction terms, where we have written a whole paper, D&P10. See section 4.2 below.
11. The 'World of Learning' covers 30,000 institutions of 'higher learning'. Economics is a fairly common field, but many of the institutions have modest research programs.
12. We have now extended the dataset – see section 3.3 below – resulting in over 600,000 cell entries. A major difference between conducting another primary study and a meta-analysis is the labor effort required to collect data.
13. In the process, we exclude five observations that were initially included in D&P08; these observations relate to food aid. Our revised dataset contains 536 observations. Below we present the results from extending the D&P08 dataset from 536 to 1,361 observations.

get everything right. M&T report coding errors in less than 0.2% of the coded cells in the D&P08 dataset. We found a similar number of errors in the third recoding. Many studies exist on the fraction of coding errors in different fields, and we believe that we have been reasonably careful.

The coding is complicated by wide differences in reporting standards between econometric studies. Most studies are straightforward (but still time consuming) to code, but some studies contain contradicting or incomplete information, especially with regard to information about the number and names of countries included in the sample. There are several cases where the reported information is doubtful, but we stuck with the original results. In some cases, we were able to go back to primary data sources and from there infer the information required. Given these factors, the coding of a small fraction of the coded coefficients is uncertain. Also, of course, there are typos and other 'clerical' mistakes in the coding.

As mentioned our comparable effect measure is the partial correlation. It can be calculated if the t-ratio or the standard error and the degrees of freedom are available. The *df*, are not always explicitly given, but can be calculated with a fair precision from the number of observations.[14] Like the elasticity it is a unitless measure, but they are not the same, see section 4.1 below.

Statistically, random errors in variables are only a serious problem in small samples. In meta-analysis, samples are normally so large that they do not make any real difference to the inferences drawn. This assessment is confirmed in the following section that shows how robust the basic results are to coding, recoding and triple coding.

3.2 *The 68 studies of D&P08: The effects of triple recoding*

Table 1 reports 2-3 estimates of the FAT-PET MRA on 5 versions for the D&P08 data set. The estimates use either robust standard errors, standard errors adjusted for data clustering or robust regression. The first two sets of estimates use weighted least squares with precision (the inverse of the partial correlation's standard error) weights. Adjusting standard errors for data clustering is recommended because of the multi-level nature of the data (several estimates reported within studies). Robust regression corrects for the effects of influential outliers.

Panel (i) reports the original D&P08 results. Panel (ii) reports M&Ts results. They show a marginally larger publication bias. But for all practical considerations the results are identical. This validates our coding – we think that M&T should have been so fair as to note this fact.

---

14. The partial correlation, *r*, is fairly robust to uncertainty about the degrees of freedom, *df*. In D&P08 the mean value of the t-statistic is 1.04 and the mean value of *df* is 162. This yields *r* = 0.08, which is not statistically significant from zero. For all values of *df* from 149 to 191, *r* remains at 0.08, so the results are robust to imprecise values the *df*.

Table 1. Estimates of the FAT-PET MRA (Eq. 2) for the D&P08 literature base

Regression/s.e.	(1) FAT Funnel asymmetry	(2) PET Meta-average	N
	(i) Original estimate from D&P08 (p.11)		
Robust s.e.	**0.73** (4.41)	0.03 (1.82)	541
Clustered s.e.	**0.73** (2.43)	0.03 (1.00)	541
Robust regression	**0.83** (4.77)	0.02 (1.32)	541
	(ii) M&T results (p.11)		
Robust s.e.	**0.79** (4.84)	0.03 (1.73)	537
Clustered s.e.	**0.79** (2.67)	0.03 (0.94)	537
	(iii) D&P08 with new revision		
Robust s.e.	**0.69** (4.18)	**0.03** (2.23)	536
Clustered s.e.	**0.69** (2.30)	0.03 (1.19)	536
Robust regression	**0.79** (4.49)	0.03 (1.73)	536
	(iv) D&P08 with published version of estimates		
Robust s.e.	**0.66** (3.76)	**0.04** (2.28)	512
Clustered s.e.	**0.66** (2.09)	0.04 (1.22)	512
Robust regression	**0.82** (4.49)	0.02 (1.43)	512
	(v) D&P08 with additional estimates		
Robust s.e.	**0.70** (4.52)	0.03 (1.90)	618
Clustered s.e.	**0.70** (2.34)	0.03 (0.99)	618
Robust regression	**0.81** (4.85)	0.02 (1.14)	618

Notes: The dependent variable is the partial correlation. The brackets hold t-ratios, using robust standard errors, clustered standard errors or robust regression, respectively. Bolded estimates are significant at the 5% level. N is the number of observations. Panel (i) reports the original D&P08 results. Panel (ii) reports M&T results. Panel (iii) uses our own revised dataset. Panel (iv) replaces the unpublished versions of the estimates used in the original D&P08 study with the published versions. Panel (v) includes additional estimates as discussed in the text. FAT measures the degree of publication bias. PET measures the effect of aid on growth corrected for publication bias.

Panel (iii) uses our own revised dataset. In order to be as inclusive as possible, DP08 used many unpublished studies, some of which are now published.[15] Panel (iv) replaces the unpublished versions of the estimates used in the original D&P08 study with the published versions when relevant. In D&P08 we excluded some estimates that were reported in appendices or other tables that were not central to the authors' main analysis. Here, we take the opportunity to expand the dataset by including these estimates as well. Panel (v) includes these additional estimates.

The results reported in Table 1 are remarkably consistent. The exercise in coding, recoding, and triple coding confirms the great robustness of the analysis in D&P08. In all cases, the FAT shows robust evidence of a significantly positive funnel asymmetry consistent with the reluctancy bias.

---

15. Many researchers claim that the inclusion of unpublished studies minimizes publication bias; restricting the sample to just published estimates is said to potentially result in the appearance of selection bias where none existed.

M&T should seriously consider the result they have replicated. M&T show that the AEL systematically selects controls to generate a positive estimate of the aid effectiveness parameter β. This is crucial knowledge for the analysis discussed in section 5 below.

The meta-average is positive, as we all want, but its size is so small as to be of no practical significance.[16] By Cohen's criteria, the size of the meta-average is negligible (0.02 to 0.04). When data dependence is accommodated by using clustered standard errors, the meta-average is not even statistically significant.[17] *The null hypothesis of no effect cannot be rejected.*

Table 2. Estimates of the FAT-PET MRA (Equation 2) for the D&P11 literature base

	(1) FAT	(2) PET	
Regression/s.e.	Funnel asymmetry test	Meta-average	N
(i) D&P11 estimates after new revision			
Robust s.e.	**0.34** (3.44)	**0.02** (3.99)	1,039
Clustered s.e.	0.34 (1.04)	0.02 (1.22)	1,039
Robust regression	**0.27** (2.63)	**0.03** (4.97)	1,039
(ii) D&P011 plus – aggregate aid			
Robust s.e.	**0.33** (3.74)	**0.03** (5.37)	1,361
Clustered s.e.	0.33 (1.20)	0.03 (1.83)	1,361
Robust regression	**0.30** (3.23)	**0.03** (5.83)	1,361
(iii) Disaggregate aid alone			
Robust s.e.	0.14 (0.42)	**0.07** (2.67)	417
Clustered s.e.	0.14 (0.24)	0.07 (1.52)	417
Robust regression	0.36 (1.34)	**0.04** (2.49)	417

Notes: See noted to Table 1.

3.3   Adding of 35 studies in the D&P11 and 30 new ones

D&P11 adds 35 studies to the 68 studies covered in D&P08. This doubles the data set to 1,039. Table 2, Panel (i) uses the same 103 studies as in DP011, but with an additional 233 estimates reported in these papers (the original DP011 uses 984 estimates). DP011 ended their search for studies in 1/1 2009. We have now conducted a new search and collected studies available as of September 2011. This produced an additional 30 studies. Panel (ii) uses this new and larger dataset

16. M&T (p.2) argue that this does not matter, stating that: 'Moreover, it has long been understood in the medical profession that it does not follow (in any simple way) from a zero meta-impact result that the medical practitioner should immediately stop 'treatment' and leave the ailing patient alone.' We think that a zero impact result is a good reason to change to another cure. As economists, we are mindful of the opportunity cost of scarce funds and seek interventions that will have the greatest cost-benefit payoffs. In Ziliak and McCloskey's (2008) nomenclature, aid lacks 'oomph'.
17. M&T also present the results of the MST test (Tables 5 and 6). However, the MST is no longer used in the field as it has been shown to have inflated type I errors; it tends to find a genuine empirical effect when there really is none. M&Ts findings of a positive aid-growth effect are probably just a type I error.

of 1,361 estimates of the effects of aggregate aid on growth from 133 studies. Panel (iii) uses 417 estimates of disaggregate measures of aid from 31 studies. This is also an update of DP011. The results are virtually the same as in D&P11.

Over time more estimates accumulate, and the distribution of the estimates should become more symmetrical, so the FAT coefficient should decrease. The number of regressions reported per study increases from about 8 in the first 68 studies to almost 20 in the last 36 studies, as the literature increasingly reports systematic model search processes. Also, the database of aid shares increases annually with about 150 observations, and longer time series are less easy to mine, since pure flukes in the data tend to even out. This explains why the FAT for aggregate aid falls to half, and its statistical significance becomes more debatable. However, the direction of the bias is still the same.

Moreover, the meta-average is the same in Tables 1 and 2, so it is still economically insignificant, and the clustered standard errors still show that the meta-average is statistically insignificant. Despite a doubling of the number of studies since D&P08 and more than doubling of the number of estimates, the literature finds the same results, and the conclusions from D&P08 still hold.

The 417 estimates for disaggregate measures of aid are the best results till now. The meta-averages are more than twice as large as the ones reported for aggregate aid. The results are still economically insignificant, but they show no sign of publication bias. However, it is a problem that these results are from a collection of diverse components of aid.[18]

By comparing the results in D&P08 and D&P11 we know that even if a few papers could be found that were overlooked in our literature base for D&P08, it is very unlikely that they will change anything. Whether we use the studies in D&P08 or more than double these in the updated dataset, the same conclusion of aid ineffectiveness remains.

3.4   Selection bias in AEL

Figure 1 presents a graphical illustration of selection bias in the AEL.[19] In the absence of publication bias, there should be no association between the aid-growth effects and their standard errors. However, the FAT-PET MRA shows that the FAT term is statistically significant. The upward slopping line plotted in Figure 1 illustrates the search intensity (and hence bias) to find 'correct coefficients'.

---

18. In D&P11 we show that some of the components of aid actually reveal economically significant effects on growth.
19. Figure 1 is basically a funnel plot that is inverted and then has its axis reversed. The upward sloping line is thus a plot of the FAT-PET regression (Eq. 2), see Stanley and Doucouliagos (2010).
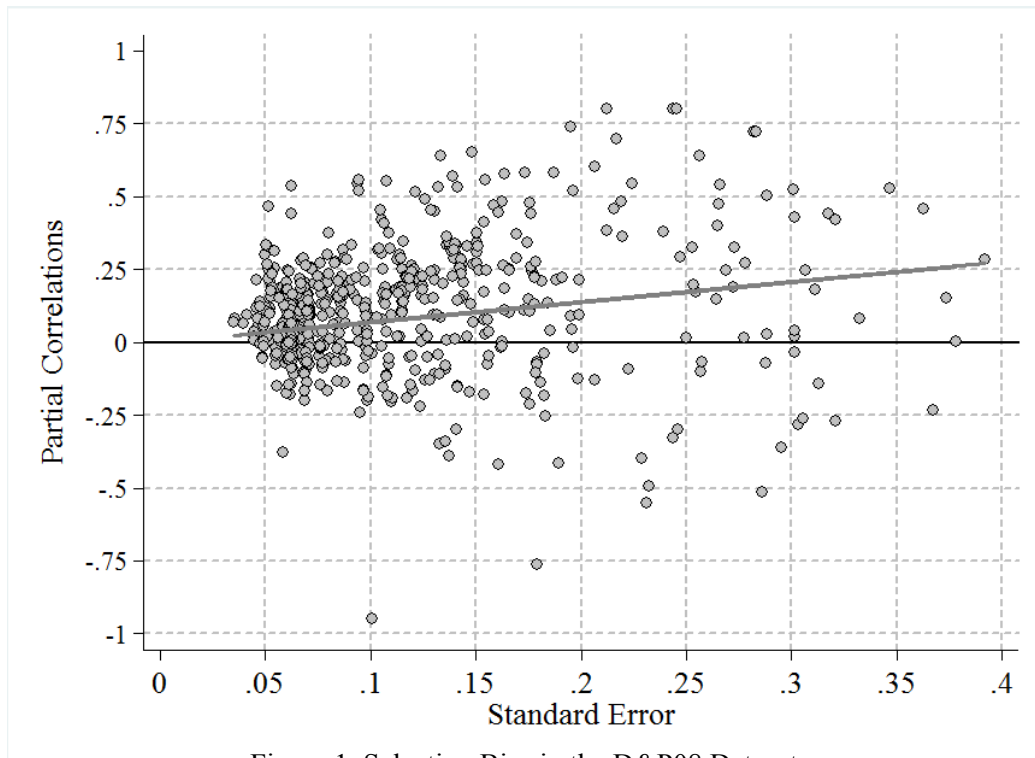
Figure 1. Selection Bias in the D&P08 Dataset

It is instructive to trace out the evolution of the AEL. We do this via funnel plots that show the relationship between the partial correlations and their estimated precision. Figure 2 illustrates the evolution in the distribution of the reported partial correlations.[20] The reader should remember that aid data did not start before the late 1960s.

Figure 2a shows the 28 estimates reached – on very limited data – before 1980. It depicts two distinct sets of estimates: One group reporting positive effects and the other reporting negative effects, with only two estimates close to zero. Here the PET converges to the largest group of estimates, which is obviously biased. This figure is a nice example of Young, Ioannidis and Al-Ubaydli's (2008) *winners' curse* idea. Early estimates show dramatic effects which do not withstand the test of time. However, these dramatic estimates are often an outcome of selection bias. Figure 2b shows that ten years later, the distribution had changed, with more results close to zero. As the literature develops, the estimates come to cluster around a zero effect. With the availability of larger datasets, precision has increased over time and the funnel part of the plot becomes more pronounced and clustered around a tiny positive effect.

---

20. The reader may note that Figure 2 uses precision and not the sample size as used in D&P08. In 2007 the relative merit of the two presentations was still discussed (*e.g.* Hunter and Schmidt (2004) and Schulze (2004) strongly recommended sample size for the meta-analysis of correlations). The funnels look much the same.

This illustrates the point made in the introduction: All three averages of aid effectiveness fall steadily throughout the decades; the AEL shows that aid becomes less and less effective in generating aid. Our assessment is that this is an artifact that follows from decreasing publication bias, due to increases in the data that gradually reduce the scope for data mining.
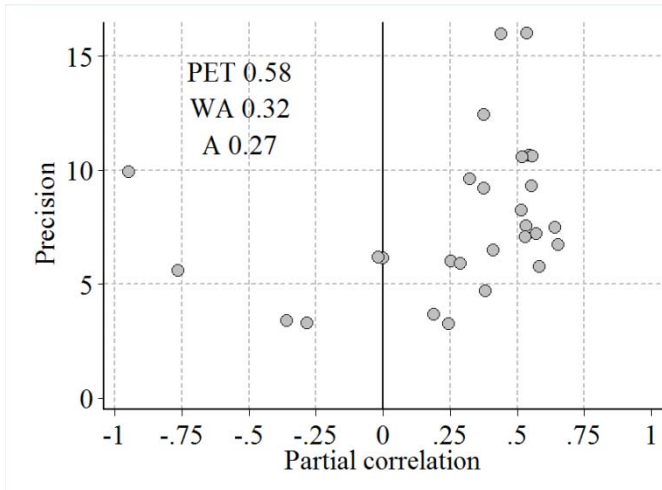

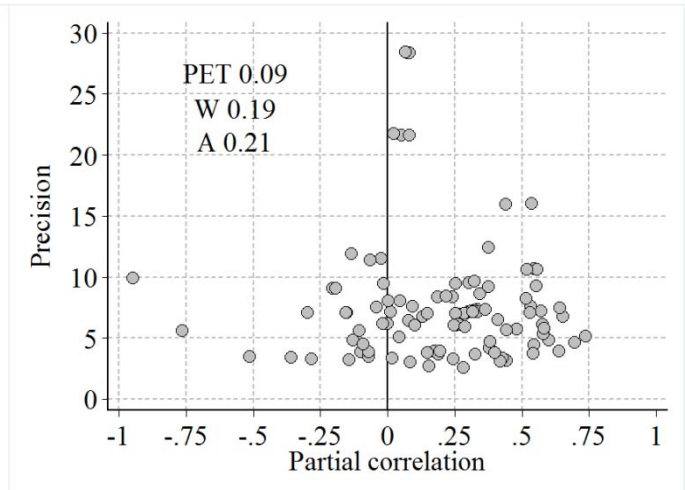Figure 2a. Pre-1980 estimates   Figure 2b. Pre-1990 estimates
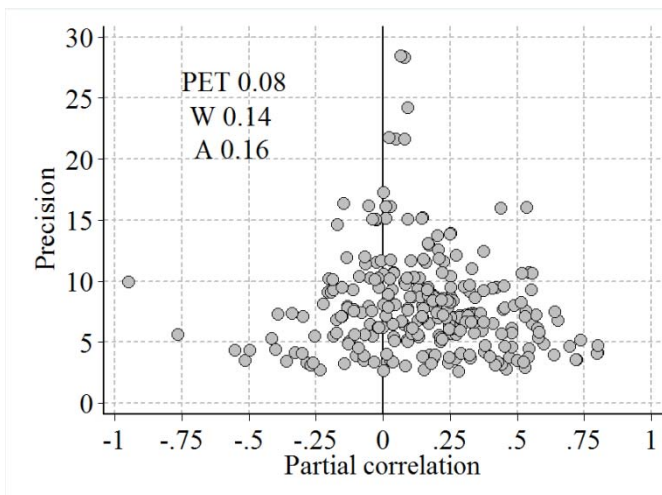
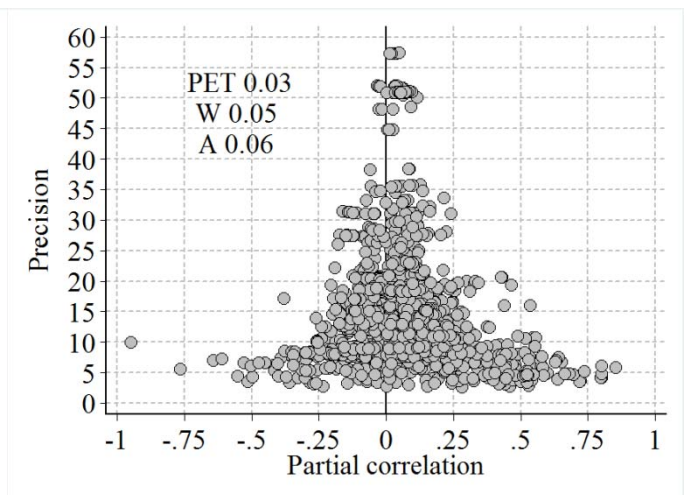Figure 2c. Pre-2000 estimates   Figure 2d. All estimates

Figure 2. Funnel plots for four different periods, all estimates of aggregate effects of aid

Note: PET is the meta average from Eq. 2, W is the fixed effects weighted average, and A is the arithmetic average.

The FAT-PET MRA from both D&P08 and M&T confirms the absence of a genuine empirical effect. Hence, taking zero as the centre of the funnel plots, we can see that they become more symmetrical over time. However, the imbalance continues to exist. If the true effect is indeed

practically zero, then we should find a more equal distribution of positive and negative estimates reported. Pre-1980, only 18% of the estimates were negative. This rose slightly to 22% in the pre-1990 period, 24% in the pre-2000 period and 37% for all time periods. This asymmetry reflects selection bias. On the other hand, 70% of *all* reported estimates have 95% confidence intervals that include a zero partial correlation. That is, even though the majority of the reported point estimates are positive, they are estimated with poor precision such that most estimates cannot reject the null of a zero growth effect.

4.   Some methodological issues

Section 4.1 shows that M&T misinterpret the partial correlation. Section 4.2 discusses the appropriate treatment of interactive terms. Section 4.3 considers the issue of fixed versus random effects in meta-analysis. Section 4.4 discusses the treatment of moderator variables in D&P08 and M&T.

4.1   Interpretation of partial correlations: They are not elasticities
The partial correlation coefficient is a statistical measure of the directional strength of an association between two variables, holding other variables constant. However, M&T incorrectly interpret the partial correlation as if it was an elasticity. For example, M&T find a partial correlation between aid and growth of 0.17. They then conclude that: 'This implies that a one percentage point increase in aid is associated with a 0.17 percentage point rise in GDP growth. This result is again close to the Arndt et al. (2010) estimate of 0.13' (M&T, p. 2).

We show that the meta-average of the partial correlation is much smaller than 0.17. However, taking their figure at face value, Cohen's guidelines (see section 3.1) suggest that this is a relatively small effect of little practical significance to growth.

Further, polishing distorts our measure of effect size. It is easy to mine economic data and authors have strong incentives to polish results, so the real level of statistical significance is surely inflated (see Caudill and Holcolmbe, 1999). Hence, our and M&T's estimates of the effect of aid on growth as a genuine effect are likely to be inflated and are best viewed as upper bound estimates.

4.2   The issue of interaction terms
M&T criticize the way we treat interaction terms (*e.g.*, aid interacted with policy, aid interacted with itself and aid interacted with institutions). They claim that our method leads to a biased estimate of the growth effect of aid. We know that it does not:

D&P10 is a separate meta-regression analysis of the two main classes of hypothesized interactions[21] – aid interacted with policy (*Aid·Policy*) and aid interacted with itself (*Aid·Aid*). We found that these interactions have failed in independent replications. Consequently, it would be wrong to include the interaction terms in the calculation of the partial correlation between aid and growth. M&T entirely overlook the findings in D&P10 where the null of no conditionality cannot be rejected.

Due to these results we handled interaction terms differently in D&P08 by including binary variables in the meta-regression analysis to capture specification differences between studies. The effects of the inclusion of interaction terms in a growth model can be quantified by the associated coefficient in the MRA. We found a negative coefficient on *Aid·Institutions* and a positive coefficient for studies that include the *Aid·Aid* term. M&T interpret these findings in D&P08 as lending support for their contention that our measure of effect is biased.

Our interpretation is entirely different. The use of interaction terms can be viewed as an attempt to impose a structure on the data that ultimately leads to a misspecification bias. The MRA coefficients merely quantify that bias. Both the *Aid·Aid* and *Aid·Institutions* terms are strictly variables that capture specification differences, rather than representing a genuine empirical effect.

4.3 *Fixed versus random effects*

In D&P08 we report estimates from both fixed effects and random effects estimators, but rely upon the results of the fixed effects model for statistical inference. A major part of M&T's critique is that they strongly advocate the use of the random effects model. Random effects can be applied to weighted meta-averages or to meta-regressions in the form of a random effects multilevel unbalanced panel analysis.[22]

Stanley (2008) and Stanley and Doucouliagos (2012) show that while both fixed effects and random effects weighted averages are biased in the presence of publication selection, fixed effects averages are less biased. M&T find significantly larger meta-averages with the random effects weighted average. They note (p. 7) that the D&P08 fixed effects meta-average: 'does not fall in our 95 per cent confidence interval which indicates that we can reject their 0.08 estimate at 5 per cent level of significance.' However, this is *exactly* what publication bias does. Their results in Table 1 do not prove that the aid on growth meta-average is larger than our fixed effects average. Rather, their Table 1 merely proves that their meta-average is artificially inflated by selection bias. M&Ts

---

21. D&P08 contains adequate cross references to the working paper version of D&P10.
22. The two types of analysis are different. One involves a weighted average while the other involves random effects in a meta-regression, similar to random effects panel econometrics models. Both approaches are fundamentally problematic when applied to economic data.

Table 2 suffers from exactly the same problem. Hence, Tables 1 and 2 in M&T merely corroborate the robust publication bias found by the FAT-PET.

The random effects model is extensively used in the meta-analysis of medical research, where estimates are drawn from controlled clinical trials. The nature of medical research means that it can be assumed that research variation is entirely random. Meta-analysis of econometric studies (such as those of the aid on growth literature) use estimates derived from mining *observational* data from readily available datasets. It is trivially easy to manufacture excess variation in applied econometrics. *This variation cannot be assumed to be random.*

M&T's preferred random effects model assumes that heterogeneity is random and independent of all of other moderator variables included in the MRA. Their model is valid only if there is no correlation between the MRA explanatory variables and the study-level effects. Stanley and Doucouliagos (2012, p. 82) argue that this assumption is unlikely to be true: ' … the random effects will be routinely correlated with the standard error when there is publication selection. Random effects are then, in part, the result of these greater efforts to select and report desired estimates.' Random effects MRA models in economics research will result in larger biases than fixed effects MRA models. For this reason we continue to prefer the fixed effects estimator (be it for calculating weighted averages or for meta-regression).

4.4   Partly omitted control variables: Genuine effect heterogeneity or omitted variable bias?

Consider four facts: (i) If all estimates were derived from models that included all controls there would be no funnel asymmetry.[23] (ii) There will still be study heterogeneity, *e.g.* due to differences in the countries and time periods covered. (iii) All controls are partly omitted variables, as they are used in some studies and not in others. (iv) The FAT-PET MRA shows that the selection of the controls is non-random, but influenced by their effect on the estimate of *β*, the aid effectiveness parameter.

Consider two partly omitted control variables: *Income* and *Africa*. They contribute to study heterogeneity, and hence to the width of the funnel, but they are different in an important respect: *Income* (*e.g.* initial income) is a variable that should be included in growth regressions according to the literature on convergence. Hence, if it is not included, it might be interpreted as a missing variable leading to an *omitted variable* bias. Whether this bias matters, is an empirical question. *Africa* asks if aid is less (or more) effective in African countries than in other countries. Thus, it

---

23. This argument assumes that controls for data selection and the estimator are included.

asks about country *homogeneity*. *Africa* is a variable that can be included only in studies that cover both African and other countries.[24]

Multiple MRAs are estimated to study both sources of study heterogeneity. But the result of the study should be treated differently. Country heterogeneity is no bias, but a fact to be noted in its own right. Omitted variable bias is a problem that should be adjusted for. The standard adjustment (the augmented FAT-PET MRA) makes the adjustment by assuming that the effect would have been the same in the estimates where the control is omitted as it is in the ones where it is included. Due to the above fact (iv) this is unlikely. The FAT-PET proves that.

Consequently, it is not an issue that it is important to use the multiple MRA to study result heterogeneity, but it is a complex issue if it makes sense to adjust the FAT-PET MRA for omitted variable biases, when the FAT-PET detects selection bias.

Both D&P08 and M&T use the same set of moderator variables to run multiple MRAs. D&P08 use the moderator variables to study the observed variation in the reported estimates. M&T use them for an augmentation process that increases the estimate of the aid effectiveness effect and decreases the funnel asymmetry. This approach disregards the results of the basic analysis.

5.   Multiple meta-regression analysis

In this section we reconsider the issue of heterogeneity using multiple MRA. The results are presented in Table 3. All estimates use WLS (weighted least squares) with precision as weights and standard errors adjusted for data clustering. In addition, we control for selection bias. In contrast, the multiple MRA in D&P08 focused only on explaining the variation in reported estimates; we did not attempt to report multiple MRA estimates corrected for selection bias. Also, the specification differs from D&P08 in three ways: (i) The variable *Expectations*, is excluded as it is a too subjective variable to code.[25] (ii) A new binary time dummy – *2000s* – is included. (iii) *1960s* and *Africa* are omitted so that the constant in the multiple MRA models represents studies that include African data from the 1950s and 1960s.

Most variable names are probably self explanatory (the reader is referred to D&P08).

---

24. However, *Africa* may give rise to biases if the included African countries are selected selectively, see section 5.2.
25. *Expectations* is meant to capture whether the empirical results met the author's expectations of the effect of aid on growth (negative, positive or neutral). However, there was much disagreement among our coding checkers regarding this variable and hence we decided to omit it from the multiple MRA.

5.1 The six estimates

Column 1 reports estimates of Eq. 3 using the same data as D&P08, but with coding errors corrected. *Column 2* reports results derived from applying a general-to-specific modeling strategy to the results reported in column 1. *Column 3* uses the fixed effects multilevel (FEML) estimator that includes dummy variables for individual authors:

$$effect_{ij} = \beta_0 + \sum \beta_k Z_{kij} + \beta_1 SE_{ij} + v_j + \varepsilon_{ij} \tag{4}$$

Where v captures any unobserved quality differences specific to studies.[26] Stanley and Doucouliagos (2012) show that fixed-effects unbalanced panel methods provide consistent and unbiased estimate of the MRA coefficients. The MRA model of Eq. 4 informs on the within study differences and naturally requires within study variation.

Column 4 tests our contention that M&T have misinterpreted the MRA model. Perhaps all of the observed variation in aid-growth effects is driven entirely by selection effects, so that the observed heterogeneity is simply an outcome of research design. We model this by extending Eq. 2 to allow heterogeneity in selection bias, rather than in genuine effects:

$$effect_{ij} = \beta_0 + \beta_1 SE_{ij} + \sum \alpha_j SE_{ij} K_{ji} + \varepsilon_{ij} \tag{5}$$

In this model, the constant (β_0) represents the *only* genuine empirical effect, and *all* variation around this is driven either by random errors or the systematic search for 'correct' effects. Hence, while in columns 1 and 2 selection bias is captured only by *Standard error* and the other variables probably represent only specification differences, column 3 allows a more complex form of selection bias. Selection bias now occurs through both the SE term, as well as SE interacted with other study characteristics. Eq. 5 explicitly models research design and the choice of models, data and estimators to get a 'desired' result. The null of no publication bias now requires a test of:

$$\beta_1 = \sum \alpha_j = 0 \tag{6}$$

Finally, *columns 5 and 6* use the new and updated dataset, for Eqs. 3 and 4, respectively. The MRA results are fairly robust across the models and confirm most of the inferences made in D&P08.

---

26 An alternative is to include individual *study* effects. However, since some papers are generated by the same author (or groups of authors), we prefer to include individual author effects. This is an extension of D&P08 where we included variables for two groups of authors, World Bank and Danida.

Table 3. Multiple FAT-PET Models

Variable	D&P08 dataset				Extended dataset	
	(1) General (Eq. 3)	(2) Specific (Eq. 3)	(3) FEML (Eq. 4)	(4) Only bias (Eq. 5)	(5) General (Eq. 3)	(6) FEML (Eq. 4)
Publication bias						
Standard error	0.755 (1.63)	**0.859** (2.62)	**1.109** (2.98)	-0.313 (0.27)	0.662 (1.81)	**0.753** (2.21)
Journals and Authors						
Unpublished	**-0.095** (3.23)	**-0.044** (3.13)	**-0.140** (4.23)	**-1.094** (2.98)	0.022 (1.02)	-0.039 (1.78)
Cato	**-0.148** (2.11)	-0.086 (1.68)	**-0.302** (4.20)	-0.840 (1.36)	**-0.136** (2.06)	-0.090 (1.43)
EDCC	**-0.200** (2.27)	**-0.226** (3.76)	**-0.700** (7.50)	**-2.167** (3.10)	**-0.159** (2.60)	**-0.862** (7.39)
JDS	-0.041 (0.90)	-	**-0.174** (4.89)	-0.282 (0.50)	0.036 (0.92)	-0.051 (0.93)
AER	-0.040 (0.99)	-	**0.214** (3.19)	-0.190 (0.31)	0.011 (0.28)	0.020 (0.43)
AppliedEconomics	-0.088 (1.63)	-0.087 (1.79)	0.085 (0.45)	-0.453 (0.87)	-0.007 (0.15)	**0.115** (2.27)
Gender	-0.082 (1.98)	-0.057 (1.77)	**-0.171** (3.78)	**-0.905** (2.03)	-0.011 (0.54)	0.017 (0.33)
Theory	**0.070** (2.41)	0.026 (1.79)	0.016 (0.31)	**0.736** (2.11)	0.001 (0.02)	0.027 (1.01)
Influence	**0.107** (4.14)	**0.105** (6.01)	**0.289** (4.95)	**1.459** (4.48)	0.024 (1.01)	**0.093** (2.02)
Data heterogeneity						
Panel	**0.142** (2.26)	0.080 (1.76)	**0.212** (3.76)	0.964 (1.54)	0.041 (0.58)	0.037 (0.91)
No. Countries	0.001 (1.33)	**0.001** (2.75)	0.002 (1.98)	0.010 (1.19)	-0.001 (1.66)	**-0.001** (2.43)
No. Years	-0.003 (1.13)	-	**-0.011** (3.97)	-0.031 (1.20)	-0.001 (0.87)	0.001 (0.82)
Sub-sample	-0.017 (0.61)	-	-0.024 (0.74)	-0.342 (1.07)	**-0.035** (2.03)	-0.015 (1.39)
Single	0.061 (0.38)	-	**-0.427** (4.61)	0.562 (0.42)	-0.134 (1.00)	-0.228 (1.68)
Specification						
Aid·Institutions	**-0.098** (2.36)	**-0.147** (3.15)	-0.023 (0.47)	**-1.543** (3.20)	-0.040 (1.14)	-0.029 (0.72)
Aid·Policy	0.008 (0.32)	-	0.010 (0.46)	-0.117 (0.36)	-0.038 (1.84)	-0.001 (0.57)
Aid·Aid	**0.043** (2.03)	**0.041** (2.07)	**0.069** (4.14)	**0.659** (2.27)	0.016 (0.94)	0.003 (0.23)
Lagged Aid	**0.058** (2.01)	**0.060** (2.19)	-0.010 (0.35)	0.690 (1.84)	0.029 (0.85)	0.045 (0.99)
SystemAid	**-0.103** (3.23)	-0.065 (1.90)	**-0.080** (2.59)	**-0.985** (2.45)	-0.061 (1.93)	0.032 (0.88)
GapModel	0.002 (0.03)	-	0.007 (0.07)	-0.250 (0.57)	0.094 (1.67)	**0.206** (2.16)
Capital	**0.137** (2.40)	**0.130** (3.58)	0.031 (0.67)	**1.207** (2.01)	-0.018 (0.69)	-0.042 (1.67)
Policy	-0.033 (1.24)	**-0.047** (3.42)	-0.005 (0.20)	-0.449 (1.31)	-0.012 (0.45)	0.012 (0.45)
Fiscal	0.021 (0.64)	-	**0.047** (2.16)	0.453 (1.05)	0.002 (0.07)	0.021 (0.75)
Sizegovt	**0.062** (2.40)	**0.083** (3.67)	0.046 (1.46)	**0.710** (2.57)	**0.075** (3.22)	0.042 (1.65)
RegionDummy	-0.028 (1.39)	-	-0.017 (0.85)	-0.469 (1.92)	-0.007 (0.57)	**-0.025** (2.02)
Ethno	-0.042 (0.91)	-	-0.059 (1.05)	-0.418 (0.64)	-0.039 (1.99)	0.006 (0.30)
Population	0.033 (1.28)	-	**0.056** (2.02)	0.338 (1.07)	-0.008 (0.43)	0.036 (1.55)
Income	-0.007 (0.10)	-	-0.112 (0.86)	0.151 (0.27)	-0.024 (1.87)	-0.024 (1.55)
Time and region						
1970s	-0.108 (1.78)	**-0.115** (2.77)	-0.059 (0.93)	-0.736 (1.56)	0.016 (0.58)	-0.047 (0.88)
1980s	-0.003 (0.04)	-	**0.107** (2.85)	-0.089 (0.22)	-0.052 (1.52)	-0.041 (0.98)
1990s	0.076 (1.82)	-	0.092 (0.94)	0.555 (1.26)	-0.013 (0.61)	0.026 (1.28)
2000s	-0.032 (1.05)	**-0.036** (3.58)	**0.079** (2.58)	**-0.894** (2.25)	0.003 (0.19)	-0.038 (1.78)
Asia	**0.176** (3.33)	**0.167** (3.85)	**0.145** (2.35)	**0.729** (2.44)	-0.021 (0.50)	0.010 (0.16)
Latin	**-0.133** (3.16)	**-0.120** (3.11)	-0.113 (1.85)	**-0.656** (2.10)	0.037 (0.89)	0.038 (0.73)
Constant (Africa)	-0.129 (0.68)	-0.081 (0.72)	-0.078 (1.54)	0.034 (0.67)	0.099 (0.83)	0.055 (1.23)
Standard error	0.115	0.115	0.100	1.426	0.093	0.080
Adjusted R^2	0.44	0.43	0.67	0.39	0.23	0.50
N	520	520	520	520	1,344	1,344

Notes: See Table 1. Estimation using WLS, with precision weighs. Columns 3 and 6 include author fixed effects (not reported). Adjusted R^2 is not strictly comparable across the different models. The variables: *JID, WorldBank, Danida, Outliers, FDI, FinDevelopment, LowIncome, EDA, HumCapital, OLS, Inflation, Stability, SystemCap* are included but never reach t-ratios of 1.7, so their coefficients are omitted to make the table easier to read.

5.2   Selection bias

Table 3 confirms that selection bias is a real phenomenon in the AEL. The coefficient on *Standard error* is always positive and in most cases it is statistically significant. That is, the evidence is rather robust that there is selection bias in this literature and that it is in favor of finding positive growth effects. Doucouliagos and Stanley (2012) establish guidelines for the practical significance of the size of selection bias. They argue that a FAT coefficient that is less than 1 does not reflect a large degree of selection bias, *on average*. Following this guideline, we can conclude that the selection bias in the AEL is of a moderate size.[27] However, the results show that there is indeed selection bias in this literature, consistent with results from the simple FAT-PET MRA. It confirms our conjecture of reluctance to report aid ineffectiveness. Interestingly, once we adopt the multiple MRA framework we do not find the decline in FAT reported in Table 2.

Standard error is not statistically significant in column 4. However, this column reports estimates of Equation 4 where selection bias is more complex than just a single term. Indeed, 15 variables that reflect selection bias are statistically significant in column 4.[28] Specifically, MRA suggests that selection for aid effectiveness works through *Theory, Influence, Aid·Aid, Lagged Aid, Capital, SizeGovt* and *Asia*. In contrast, we find that there is less selection among unpublished studies, those that explore conditionality with institutions, those that model aid as a system of equations, those that include regional dummies, and those that include data from the new century and from Latin America.

Consistent with this selection effect is the coefficient on *Influence*. This variable is almost always statistically significant and positive indicating that studies that receive feedback from friends report stronger aid effectiveness. Column 4 suggests that this is linked to selection (*Influence·SE*); on average, researchers who receive feedback from other authors who have published in AEL are more likely to selectively report aid effectiveness; on average, it inflates t-statistics by 1.46.

A simple illustration of the publication bias in the AEL is to see what happens to the number of bolded (significant) coefficients when the same multiple MRA is calculated for the first 520 estimates and for all 1,344 estimates. When the number of observations rises any stable structure in the data should become clearer, and hence the number of bolded coefficients should go up. As seen from Table 4 the reverse happens.

---

27. For example, in the beta-convergence literature, the size of the FAT coefficient (*Standard error*) is 4.31 which is a 'severe' form of selection (see Doucouliagos and Stanley, 2012).
28. A Wald test rejects the null of the joint statistical insignificance of all the selection bias terms (Eq. 6); p-value is 0.000.

Table 4. Significance in comparable columns in Table 3 for 520 and 1,344 estimates

| Number of | Eq.3 General | | | Eq. 4 FEML | | |
Estimates	Column	Bold	R^2	Column	Bold	R^2
520	(1)	14	0.44	(2)	18	0.67
1,344	(5)	5	0.23	(6)	7	0.50

5.3   *Genuine growth effects*

While selection bias is interesting, the essential issue is whether there exist genuine empirical effects. The only variables in the MRA that can (potentially) be interpreted as estimating heterogeneity in the size of the genuine empirical effect, are the region dummies (*Asia*, *Latin* and *Africa*) and the decade dummies. Here too, however, it is possible that these variables are not actually picking up genuine effects but rather selection. Column 4 of Table 3 confirms that this indeed the case; authors vary samples to try to get the 'right' results.[29] Nevertheless, we will assume for the sake of the argument that these variables do reflect heterogeneity in the effect of aid on growth.

In columns 1 to 3 in Table 3, *Asia* has a positive and statistically significant coefficient. This finding reproduces the results reported in D&P08 (p.17), where we noted that: 'While there does not appear to be a direct aid-growth effect on average across nations, it appears that Asian countries experience higher aid-growth effects.' However, this result is not robust and appears to break down when we use the larger dataset (columns 5 and 6, Table 3). In contrast to *Asia*, *Latin* has a negative coefficient, suggesting that aid to this region had an adverse effect on growth (assuming that is that the MRA coefficients actually can be interpreted as picking up genuine effects rather than selection). The base of the MRA models, *Africa*, is not statistically significant. The MRA coefficients in column 1 show an aid on growth effect of 0.176 for Asia, -0.133 for Latin America and zero for Africa. Thus, holding time period constant, the MRA estimate of the effect of aid on growth for studies that include data from all three regions is +0.043. That is, allowing for region heterogeneity at best increases the aid-growth effect from +0.03 (Tables 1 and 2) to +0.04 (Table 3). This shrinks to zero when we add more observations (columns 5 and 6).

The MRA results for the decade dummies also do not support aid effectiveness. The coefficients for the decade dummies are not robust, but they do indicate that reported estimates of aid effectiveness vary depending on the time period used. Aid effectiveness is weaker if data from the 1970s is added to a sample. It is also weaker, in some cases, when data from the new century is added (2000s). Taken literally as estimates of the genuine empirical effect, the MRA predictions from columns 1, 2 and 3 assuming studies include data from all regions and all time periods are -

---

29. Note that in column 4, the PET coefficient is 0.034, with a t-statistic of 0.67. This is similar to the size of the PET reported in Tables 1 and 2.

0.15, -0.19 and +0.17, respectively; the between study estimates show a negative effect on growth whereas the within study estimates find a positive effect on growth.

It is instructive to conduct an additional thought experiment assuming that all the MRA coefficients could be used to construct the genuine effect. The issue then is which of the rival growth model specifications should be adopted as 'best practice'? One possibility is the Burnside and Dollar (2000) specification. This would mean setting the following MRA variables to 1: *Panel*; *Income*; *Ethno*, *Stability*, *RegionDummy*, *Aid·Policy* and *Policy*. We assume that data is used for all regions and all time periods and we use the mean number of countries and years. This gives an estimated partial correlation of -0.07, -0.11 and +0.13, for columns 1, 2 and 3, respectively. The predictions using the extended dataset are -0.09 and -0.02, for columns 5 and 6, respectively. If cross-sectional data are used (*Panel* = 0), aid effectiveness receives even less support.

6.   Conclusions: The sad findings are robust

In D&P08 we set out to determine what conclusions could be drawn from the vast literature on aid effectiveness. The literature reports an amazing range of results. Nevertheless, the meta-analysis methodology adopted allowed us to reach two conclusions:

(i) The basic level of analysis reached a very sad result: The literature shows aid ineffective-ness. To date, no stable model showing aid effectiveness has been found. The small positive effect found in the average study is mostly due to publication bias. This result is *fully* confirmed by M&T – their FAT-PET results are virtually identical to ours. We have made a third recoding and the prior inferences still stand unchanged. We consequently know that most of the positive effects found in the AEL are fickle and not genuine. This result severely limits the scope for further analysis, but it still makes sense to study the large variation in reported estimates at the second level.

(ii) We actually managed to explain a good deal of the variation. These results are less robust than the basic result. But they still tell an interesting story.

In spite of the results at the basic level, M&T treat the MRA variables as if they were all capturing genuine empirical effects. In their study M&T offer no justification for this dramatic contradiction. Even then the results M&T reach still show only a small effect. We show in this paper that modest changes to the modeling of the MRA produce results that are in accordance with the findings at the basic level. However, regardless of the model, the central and uncomfortable conclusion from D&P08 remains: The large global research effort has not managed to establish that aid is effective in generating growth.

Few could deny that the eradication of poverty is an urgent and pressing global issue. The key issue is not whether the developed world should assist; for surely it must. Rather, the central issue is the form that this assistance should take.

References:

Arndt, C., Jones, S., Tarp, F., 2010. Aid, growth and development: Have we come full circle? *Journal of Globalization and Development* 1 (2), Article 5 pp 1-27

Burnside, C., Dollar, D., 2000. Aid, policies and growth. *American Economic Review* 90, 847–68

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

Caudill, S.B., Holcombe, R.G. 1999. Specification search and levels of significance in econometric models. *Eastern Economic Journal* 25, 289-300

Christensen P.W., Doucouliagos, H., Paldam, M., 2nd edition, 2010. *Master list of the AEL: The Aid Effectiveness Literature.* Posted on http://www.martin.paldam.dk. It replaced 1st edition from 2007

Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. Erlbaum, Hillsdale, NJ.

Dalgaard, C.-J., Hansen, H., Tarp, F., 2004. On the empirics of foreign aid and growth. *Economic Journal* 114, 191-216

Doucouliagos, H., Paldam, M., 2006 (**D&P06**). Aid effectiveness on accumulation. A meta study. *Kyklos* 59, 227-54

Doucouliagos, H., Paldam, M., 2008 (**D&P08**). Aid effectiveness on growth. A meta study. *European Journal of Political Economy* 24, 1-24

Doucouliagos, H., Paldam, M., 2009 (**D&P09**). The aid effectiveness literature. The sad results of 40 years of research. *Journal of Economic Surveys* 23, 433-61

Doucouliagos, H., Paldam, M., 2010 (**D&P10**). Conditional aid effectiveness. A meta study. *Journal of International Development* 22, 391-410

Doucouliagos, H., Paldam, M., 2011 (**D&P11**). The ineffectiveness of development aid on growth: An update. *European Journal of Political Economy* 27, 399–404

Durlauf, S.N., Johnson, P.A., Temple, J.R.W., 2005. Growth econometrics. Chapter 8, 555-677 in Aghion, P., Durlauf, S., eds., *Handbook of Economic Growth*. North-Holland, Amsterdam

Hansen, H., Tarp, F., 2000. Aid effectiveness disputed. *Journal of International Development* 12, 375-98

Hunter, J.E., Schmidt, F.L., 2004. *Methods of meta-analysis: Correcting error and bias in research synthesis* (2nd ed.). Sage, Newbury Park, CA

Mekasha, T.J., Tarp, F., 2011 (**M&T**). Aid and growth. What meta-analysis reveals. UNU-Wider. Working Paper No. 2011/22. Helsinki

Mueller, D.C, 2003. *Public Choice III.* Cambridge University Press, Cambridge, U.K.

Schulze, R., 2004. *Meta-analysis: A comparison of approaches.* Hogrefe & Huber Pub., Göttingen

Stanley, T.D., 2005. Beyond publication bias. *Journal of Economic Surveys*, 19: 309-45

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, 103-27

Stanley, T.D., Doucouliagos, H., 2010. Picture this: A simple graph that reveals much ado about research. *Journal of Economic Surveys* 24, 170–91

Stanley, T.D., Doucouliagos, H., 2012. *Meta-Regression Analysis in Economics and Business*. Routledge, forthcoming

Stanley, T.D., Jarrell, S.B., 1989. Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys* 3, 54-67

Young, N.S., Ioannidis, J.P.A., Al-Ubaydli, O., 2008. Why current publication practices may distort science. *PLoS Med,* 5: doi:10.1371/journal.pmed.0050201.

World Development Indicators, URL: http://databank.worldbank.org/ddp/home.do?Step=1&id=4

World of Learning, Book and URL: http://www.worldoflearning.com

Ziliak, S.T., McCloskey, D.N., 2008. *The Cult of Statistical Significance: How the standard error costs us jobs, justice, and lives.* The University of Michigan Press, Ann Arbor