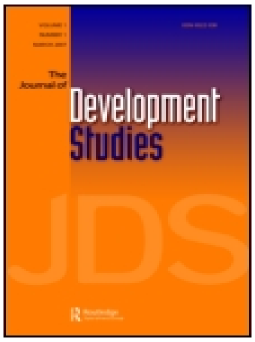

The Journal of Development Studies

ISSN: 0022-0388 (Print) 1743-9140 (Online) Journal homepage: https://www.tandfonline.com/loi/fjds20

Aid and Growth: What Meta-Analysis Reveals

Tseday Jemaneh Mekasha & Finn Tarp

To cite this article: Tseday Jemaneh Mekasha & Finn Tarp (2013) Aid and Growth: What Meta-Analysis Reveals, The Journal of Development Studies, 49:4, 564-583, DOI: 10.1080/00220388.2012.709621

To link to this article: https://doi.org/10.1080/00220388.2012.709621

Journal of Development Studies,
Vol. 49, No. 4, 564–583, April 2013

Routledge
Taylor & Francis Group

Aid and Growth: What Meta-Analysis Reveals

TSEDAY JEMANEH MEKASHA* & FINN TARP**
*UNU-WIDER and Aalto University, Helsinki, Finland, **UNU-WIDER, Helsinki, Finland and University of Copenhagen, Denmark

Final version received March 2012

ABSTRACT *Recent literature in the meta-analysis category where results from a range of studies are brought together throws doubt on the ability of foreign aid to foster economic growth and development. This article assesses what meta-analysis has to contribute to the literature on the effectiveness of foreign aid in terms of growth impact. We re-examine key hypotheses, and find that the effect of aid on growth is positive and statistically significant. This significant effect is genuine, and not an artefact of publication selection. We also show why our results differ from those published elsewhere.*

I. Introduction

The literature on the potential impact of aid on growth is large and multifaceted.[1] Hansen and Tarp (2000) identify three generations of literature, and more recently, a fourth generation has emerged (see Arndt et al., 2010). A distinctive aspect of this generation is the view that aid's aggregate impact on economic growth is non-existent. Doucouliagos and Paldam (2008) (henceforth DP08) reach a similar pessimistic conclusion in their various papers based on a meta-analytic approach and a database including 68 studies on the aid-growth link.

More specifically DP08 ask (i) whether the aid effectiveness literature has established that aid has an impact on economic growth and if so how large is the impact; and (ii) what explains the heterogeneity in reported aid-growth effects? DP08 apply different meta-analysis techniques,[2] and conclude that the aid effectiveness literature has failed to show that the effect of development aid on growth is positive and statistically significant. They also attribute the variation in the reported effect of aid on growth to different study characteristics (DP08: 13–18).

In relation to the aid-growth literature, DP08 is an example where studies which have emerged over a long time period and which rely on differing methodologies and data sets are analysed. The DP08 analysis has attracted attention in policy debates about aid so we decided to re-examine their core aid-growth analytical result.[3] This was motivated by three underlying concerns: (i) the need to specify and justify the underlying *econometric* model used; (ii) *statistical* choices related to measurement of the effect estimates and calculation of the weighted average (both in terms of methodology and choice of precision of coefficient estimates); and (iii) time consuming and tedious data entry and coding work that is not always straightforward to replicate for those interested in the results.

This study reports what we uncovered in the process, and expands the DP08 meta-analysis in various ways that better reflect the econometric, statistical and data challenges faced in this type of research. In doing so, we address two main research questions that are common to any standard

Correspondence Address: Finn Tarp, Department of Economics, Copenhagen University, Øster Farimagsgade 5, DK-1353 Copenhagen K, Denmark. Email: finn.tarp@econ.ku.dk

ISSN 0022-0388 Print/1743-9140 Online/13/040564-20
http://dx.doi.org/10.1080/00220388.2012.709621

meta-analysis: (i) whether the empirical effect (in our case the impact of aid on growth) is different from zero when one combines the existing empirical evidence; and (ii) if so, whether the effect is genuine or an artefact of so-called publication bias (also referred to as the 'file drawer' problem).

Meta-analysis, or regression of regression analysis, is normally used with the aim of synthesising the results from a group of studies while controlling for heterogeneity among studies. This methodology is usually applied in medical science research to assess the effectiveness of a well-defined healthcare intervention by combining data from primary studies that use randomised controlled trials. In recent years, however, meta-analysis has also been applied in economics and other fields of social science. One advantage of meta-analysis is that it can potentially address the subjectivity associated with traditional narrative literature surveys, and it may indeed provide a more systematic and objective (quantitative) assessment of an existing body of findings. Yet, the meta-methodology is by no means flawless (Stanley, 2001). Even if one accepts that meta-analysis is relatively more objective than narrative literature reviews, considerable room for subjectivity remains. For instance, in identifying the appropriate population of studies, authors often exercise personal judgment. Hence, bias from systematic selection of studies may follow. Moreover, decisions regarding data entry and coding, choice of a common metric for the effect size, statistical weighting of the effect estimates, model selection to conceptualise the meta-analysis and choice of moderator variables all involve different levels of personal judgement. Unless carefully handled, such judgement calls potentially lead to misleading conclusions and hence may jeopardise the relevance of meta-analysis as a quantitative tool for literature review (for example, Bullock and Svyantek, 1985; Wanous et al., 1989).

In this article, we bypass the general bias issue in literature search and rely on the exact same 68 studies as DP08. Though these studies are by no means an exhaustive list of papers in the aid-growth literature, we decided to stick to these 68 papers for the sake of comparison.[4] Besides, we use the same common metric to measure the effect size (that is, partial correlation) and the same set of moderator variables as DP08.[5] Despite these similarities we have relied on at least four different analytical choices.[6] First, we differ from DP08 in relation to model selection for the meta-analysis. In meta-analysis, one can rely on either a 'fixed effect' or a 'random effects' model depending on the assumption the meta-analyst makes regarding the nature of the true effect. DP08 argue that there is a single 'true' effect of aid on growth, which is common to all the 68 studies. This implies that they assume that random sampling error is the only factor behind the variation in reported effect estimates among studies. As a result of this assumption of 'effect homogeneity', DP08 mainly focus on a fixed effect meta-analysis. Our expectation is, in contrast, that the impact of aid on growth across the 68 studies is heterogeneous, and using both statistical tests and graphical tools we reject the effect homogeneity assumption.

One can also rule out the effect homogeneity assumption on theoretical grounds as the effect of aid on growth is a function of other factors. For instance, Burnside and Dollar (2000), Dalgaard et al. (2004), Hansen and Tarp (2001), and Chauvet and Guillaumont (2004), among many others, use interaction terms by which the partial effect of aid on growth is a function, not a constant. Moreover, the fact that the type of aid, the way it is delivered and the donor-recipient relationships differ across countries and have changed over time implies that the primary studies will target different population effect estimates. In sum, the effect homogeneity assumption of the fixed effect model cannot be expected to hold for the aid growth literature. Consequently, we conclude that random effects meta-analysis is more appropriate and shows that the underlying model choice does matter for the conclusions drawn.

Second, a major concern with the DP08 approach is the way the partial effect estimate is measured for papers that include non-linear terms like aid squared, aid-policy and aid-institutions. For papers that include one or more of these interaction terms the partial effect of aid will not be measured correctly if one ignores the coefficient of the non-linear term(s). To see this, consider the following growth regression:

$$G = \beta_o + \beta_1 * aid + \beta_2 * (aid * X) + \beta_3 * Z + \varepsilon \tag{1}$$

where X may be aid, policy or institutions and Z is a vector of other explanatory variables. In this case, the partial effect of aid on growth is given by $(\beta_1 + \beta_2 * X)$.[7] However, the data in DP08 relies on β_1 as the partial effect of aid. In the meta-analysis, this problem matters in particular for regressions that use the partial effect as a dependent variable. One case in point is in calculating the weighted average effect of aid on growth. We have recognised this issue by separately estimating the weighted average effect for papers that include one of the aforementioned interaction terms and for those that do not include any of these terms. As shown in Section III this choice matters for the results. DP08, on the other hand, ignored the issue.[8] Third, we differ from DP08 in the method used to calculate the weighted average effect of aid on growth and in our choice of the measure of statistical precision of coefficient estimates. In DP08 the weighted average aid-growth effect is calculated using sample size as weights under the assumption that studies with large sample size are more accurate. Accordingly, DP08 tune in on sample size as the preferred measure of statistical precision of parameter estimates. This choice is not, however, in line with established best-practice in standard fixed and random effects meta-analysis, which recommends that the inverse of the variance of estimates should be used as weights (that is, as the measure of statistical precision) when calculating the weighted average effect (pooled estimates) from an existing body of empirical literature. Sterne and Harbord, (2009) also note that the precision of an effect estimate cannot be fully captured by sample size. Other data characteristics are important in determining standard errors. Studies with very different sample sizes may have the same standard error and precision and vice versa. Consequently, in our estimations of the weighted average (combined) effect of aid on growth, we use the inverse of the variance of estimates as weights. As shown in Section III, the way the weighted average is estimated matters for the results. Moreover, in plotting the funnel plots used for visual inspection, we use the inverse of the standard error of the estimates as a measure of precision. DP08 use sample size; Sterne and Egger (2001) have demonstrated that this approach to measuring the precision in funnel plots is inappropriate.

Fourth, turning to data issues we began by re-entering all DP08 data and found reason for some recoding.[9] As a result, the number of observations used for the multivariate meta-regression-analysis (MRA) is increased from 471 to 519.[10] Nevertheless, we have followed DP08 throughout as closely as possible to make sure results are comparable. Thus, even if our revised data set does not exactly match that of DP08, the correlations between the two sets of data are high (Mekasha and Tarp, 2011: Table A9.1).

Before moving on to our analysis we highlight a general concern, which is a potential threat to the credibility of meta-analysis as a tool for quantitatively assessing an existing body of findings. This relates to differences in the quality of the primary studies. Meta-analysis combines results from different studies regardless of their quality, and this problem gets more pronounced in social science research where most studies are based on observational/non-experimental data. In contrast to controlled experiments, observational studies differ substantially in their model specification, econometric techniques, functional forms and research design leading to potential quality differences. Such differences are likely to lead to heterogeneity in effect estimates and unless properly captured, this heterogeneity may wrongly be interpreted as publication bias. It is therefore crucial to allow for quality differences in meta-analysis.

However, measuring (assessing) differences in qualities entails subjective judgment. It is nearly impossible to come up with a single yardstick against which quality of the primary studies is defined. Even if researchers agree on a single quality yardstick, how to take this into account in the meta-analysis is another challenge. One suggestion in this regard is to categorise studies as ‘good’ and ‘bad’ quality and do the meta-analysis either focusing only on the ‘good quality’ studies or undertake a separate meta-analysis for each category. But here subjectivity is an obvious issue. There is no way to objectively categorise studies as of ‘good’ versus ‘bad’ quality. Another suggestion is to use quality scoring as weights but this method faces strong criticisms on different grounds.[11] Yet another suggestion is to use ‘quality’ or ‘some components of quality’ as a moderator variable in the meta-regression analysis (MRA) and see whether there is a systematic difference in effect size between 'well-designed' and 'badly-designed' studies.[12] While defining 'quality' may still introduce subjectivity, controlling for 'some components of quality' can partly address this issue.[13] However, this can only be used in the case of multivariate MRA and in general leaves the problem unsolved in the calculation of the weighted average effect where moderators cannot be controlled for.

In sum, even if there is broad consensus regarding the importance of considering the quality of the primary studies, there is hardly agreement on how to measure quality and on the ways to incorporate it in the analysis. This makes the issue of quality differences across the primary studies a major caveat. In general, we need to bear the limitations in mind when trying to draw lessons from a meta-analysis and the more so for non-experimental research where the necessary tools to overcome the above challenges are very limited and are only starting to emerge. Moreover, unlike the case of random control trials where the treatment and its effect are well-defined, this is not always the case in observational macro-level studies like aid and growth. This accentuates the caution one needs to exercise when making inference from aid-growth meta-analysis.[14] So unless due care is taken, meta-analysis cannot per se guarantee an objective assessment of an existing body of findings. 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. Absence of evidence should only with great care be interpreted as evidence of absence (as noted by Temple, 2010).

This article is structured as follows. Section II deals with data and methodology, while detailed results are presented in Section III. Section IV concludes that meta-analysis, if applied meticulously, suggests a positive and statistically significant impact of aid on growth and importantly suggests there is no publication bias in the aid-growth literature. Various appendix tables in Mekasha and Tarp (2011) provide further background data and detail.

II. Data and Methodology

The data used here originate from 68 published and unpublished aid-growth studies identified by DP08 covering the period 1970–2004. Since each of the 68 studies reports one or more regressions, we have a total of 542 observations (regressions) to work with.[15]

The first step in any standard meta-analysis is to establish whether the size of the combined empirical effect in the literature under investigation is significantly different from zero or not. This is done by examining the pooled estimates (that is, the mean overall effect) of all the studies included. There are two approaches to calculating the pooled estimate, that is, the fixed effect model and the random effects model.[16]

In the fixed effect model it is assumed that all studies come from a population with a fixed average effect size, meaning that all studies are assumed to share a common true effect. Accordingly, the observed effect size[17] is assumed to vary from one study to another only because of random sampling error (within study variation). In contrast, in the random effects model, the assumption is that studies were drawn from populations that differ from each other in ways that could affect the treatment effect (Borenstein et al., 2007). In this case, the effect size will vary both due to sampling error (the fixed effect model) and due to true variation in effect size (between study variations).

Furthermore, in calculating the pooled estimate and hence the combined empirical effect, each effect size is weighted, the weight being the inverse of the variance from each study. In the case of the fixed effect model the weight is given by $1/v_{\mathrm{i}}$ where v_{i} is the within study variance. On the other hand, the weight in the random effects model is given by $1/(v_i + \tau^2)$ where v_{i} and τ^2 refer to the within and between study variances respectively. Having estimated the mean overall effect, the next step is to examine whether this observed effect is genuine or an artefact of publication bias (the so-called file drawer problem). The most commonly used tool to make a preliminary examination of the presence of publication bias is funnel plots, which are visual graphical images that illustrate the relationship between treatment effects estimated in individual studies (plotted on the horizontal axis) and a measure of study precision (shown on the vertical axis). The idea is that the precision (accuracy) in estimation of the underlying treatment effect (in our case the impact of aid on growth) increases as the study size grows. Consequently, small studies are expected to scatter widely at the bottom of the graph, while the spread is expected to narrow among larger studies at the top of the funnel. If there is no bias the plot will take the shape of an inverted funnel, and be symmetrical around the expected true effect. As indicated above, since sample size cannot fully capture the precision of reported effect size, our choice of measure of precision for the vertical axis in funnel plots follows Sterne and Egger (2001). They argue that standard errors (or their inverses) are the most appropriate measure of the precision of reported effect size.[18]

Even if funnel plots help in tracing publication bias or in general small study effects in the data, visual assessment of funnel plots is essentially subjective. Moreover, Sterne and Harbord (2009) note that funnel plot asymmetry does not necessarily arise from publication bias. Other potential reasons include, for instance, heterogeneity in underlying effects and/or low methodological quality of smaller studies. So, funnel plots should be seen as a generic means for investigating small study effects (if small studies show a larger treatment effect), not as a tool to diagnose a specific type of bias. It is therefore prudent to complement graphical observations from a funnel plot inspection with statistical tests for funnel plot asymmetry. Egger et al. (1997) provide the most commonly used test in the meta-literature; their test is regression-based to assess skewness in a funnel plot. This test starts by examining the relationship between study i's reported effect size ($Effect_i$) and its associated standard error (SE_i) as follows:

$$Effect_i = \alpha_0 + \alpha_1 SE_i + \varepsilon_i \tag{2}$$

According to Stanley (2005), one can divide this equation by SE_i to avoid potential problems of heteroscedasticity, rewriting Equation (2) as:

$$t_i = \frac{Effect_i}{SE_i} = \alpha_1 + \alpha_0 \frac{1}{SE_i} + \mu_i \text{ where } \mu_i \text{ is } \varepsilon_i / SE_i \tag{3}$$

The main idea behind this test is that, assuming a non-zero underlying effect and absence of publication bias, small studies will have a precision ($1/SE_i$) and a standardised effect ($Effect_i/SE_i$) close to zero. Large studies will have high precision and the standardised effects are expected to scatter around a regression line that passes approximately through the origin. The slope of this regression line estimates both the size and direction of the underlying effect. Failure of the regression line to pass through the origin implies publication bias. The size of the intercept gives a measure of asymmetry; the larger the deviation from zero the higher the asymmetry and hence bias in the effect size reported by the literature.

In sum, Equation (3) provides a basis for testing both funnel graph asymmetry and the presence of a genuine empirical effect beyond any publication bias. Stanley (2005) insists that the presence of an underlying genuine empirical effect, irrespective of publication bias, must be confirmed by another test. This is the so-called meta-significance test (MST), which verifies the authenticity of empirical effects by analysing the relationship between the natural logarithm of the absolute value of a study's standardised effect (t-statistics) and its degrees of freedom (df). The MST equation can be written as:

$$\ln(|t|) = \beta_0 + \beta_1 \ln(df) + \in_i \tag{4}$$

Equation (4) provides evidence of a genuine empirical effect if $H_0 : \beta_1 \leq 0$ is rejected. This test helps to identify a genuine empirical effect over and above any publication bias, and the line of thinking is clear in the following quote:

> Observing a positive association between df and the standardised test statistic throughout a given empirical literature is an additional means to confirm the authenticity of the effect in question. Without such a confirmation, seemingly positive findings reported in the literature may be the consequence of fortuitous misspecification or systematic publication biases. Without this or similar validation, a theoretical economic proposition should not be regarded as empirically corroborated or 'verified'. Seemingly strong empirical results across an entire literature might easily be the remnants of selected bias. (Stanley, 2005: 329)

III. Results and Discussion

In this section we first present the pooled estimate of the combined effect of aid on growth, and then turn to investigating whether the observed effect is genuine (authentic) or an artefact of publication bias.

The Weighted Average Effect of Aid on Growth

The first (and typically main) aim of any meta-analysis is to combine the available empirical evidence so as to establish whether the impact of an intervention is different from zero or not. Accordingly, in Table 1 we present the combined estimates of the impact of aid on growth (and the associated confidence intervals) from fixed and random effects meta-analysis. Both suggest a positive and significant effect of aid on growth (0.082 and 0.098 respectively) when the empirical evidence from the 68 studies is combined.

One difference with DP08 is the way we calculate the weighted average from the aid-growth literature as we follow standard practice and calculate the pooled estimate (in both the fixed and random effects models) using the inverse of the variance as weight. DP08 used the sample size as weight and found a weighted average of 0.08.[19] Such a simple weighted average calculation implicitly assumes away between study heterogeneity and is similar to a fixed effect. The results in Table 1 confirm this as the magnitude of the weighted average fixed effect estimate appears to be similar to DP08.

The fixed effect estimate is based on the assumption that there is a single true effect size (population treatment effect) inherent in all studies. This assumption is empirically testable and the fixed effect result can easily be challenged if there is heterogeneity of true effects across studies. Heterogeneity may not always be an issue, as in tightly controlled medical experiments (Schell and Rathe, 1992). As we rely on a wide-ranging set of 68 different studies with varying foci, quality of research design and analytical approach, heterogeneity is to be expected. This is indeed what the *Q*-test for heterogeneity reported in Table 1 suggests.[20] The presence of heterogeneity is also clearly confirmed graphically in Mekasha and Tarp (2011). The fixed effect model based on the homogeneity of effects assumption is clearly inappropriate in a meta-analysis

Table 1. Meta-analysis of the effect of aid on growth

Method	No. of regressions	Pooled estimate	95% CI lower	95% CI upper	P-value H_o: no effect
Overall fixed	537	0.082	0.076	0.089	0.000
Random	537	0.098	0.085	0.112	0.000

Notes: Test for heterogeneity: Q = 1791.745 on 536 degrees of freedom (p = 0.000) and the estimate of between studies variance=0.015. The number of regressions is 537 instead of 542 as four estimates do not have data on standard errors due to missing data, and we have also removed one regression from the study with ID38 as an outlier. We have also checked the sensitivity of the overall effect to the inclusion of the outlier and the results still hold. That is, 0.081 and 0.097 for the fixed and the random effects respectively.
Source: Authors' estimates.

of the aid and growth literature. Indeed, the effect homogeneity claim does not appear to be supported by the evidence inherent in the data.[21] It is therefore appropriate to focus on the random effects model.

The weighted average effect of aid on growth in Table 1 is positive and statistically significant with a magnitude of 0.098 in the random effects meta-analysis. As can also be seen from Table 1, the DP08 weighted average effect estimate does not fall in our 95 per cent confidence interval which indicates that we can reject their 0.08 estimate at the 5 per cent level of significance. As shown in Equation (1) the partial effect of aid on growth will not be measured correctly for papers that aim to capture non-linear effects of aid on growth. Table 2 shows how this matters for the result, including separately re-estimated weighted average effects by classifying the papers based on their treatment of non-linearity.

To illustrate, for papers that include the aid squared term overlooking $2\beta_2 * aid$ will overstate the weighted average effect of aid reported from these papers, because the expected sign of the coefficient of aid squared in (1) above is negative. This is consistent with the result reported in Table 2: the weighted average effect from papers that include the aid squared term is much higher than papers which do not include the aid squared term. In a similar fashion, for papers that include aid-policy and aid-institution interaction terms, the expected sign of the coefficient of the interaction term is positive. Hence, ignoring the $\beta_2 * X$ term in Equation (1) will understate the estimated weighted average effect of aid. Again, this is confirmed by the results in Table 2. Papers that include either aid-policy or aid-institution interaction terms appear to have a lower weighted average effect compared to papers that do not include these terms.

Table 2. Meta-analysis of the effect of aid on growth by classifying the studies based on the type of non-linear terms included in the papers

Type of non-linearity used in the papers:	No. of regressions	Combined effect estimate	95%CI lower	95% CI upper	P-value H_o: no effect
Studies with aid square					
Fixed	97	0.124	0.112	0.137	0.000
Random	97	0.131	0.110	0.153	0.000
Studies without aid square					
Fixed	441	0.064	0.056	0.072	0.000
Random	441	0.087	0.071	0.104	0.000
Studies with aid-policy					
Fixed	157	0.044	0.034	0.054	0.000
Random	157	0.044	0.027	0.060	0.000
Studies without aid-policy					
Fixed	381	0.113	0.104	0.122	0.000
Random	381	0.131	0.111	0.150	0.000
Studies with aid-institution					
Fixed	27	−0.112	−0.142	−0.081	0.000
Random	27	−0.112	−0.149	−0.075	0.000
Studies without aid-institution					
Fixed	511	0.091	0.084	0.098	0.000
Random	511	0.108	0.094	0.122	0.000
Studies with at least one of the three interaction terms					
Fixed	232	0.067	0.058	0.075	0.000
Random	232	0.066	0.051	0.082	0.000
Studies without the interaction terms					
Fixed	306	0.109	0.097	0.120	0.000
Random	306	0.138	0.113	0.162	0.000

Notes: The Q tests for heterogeneity for studies with and without conditionality are Q = 756.157 on 231 degrees of freedom (p-value = 0.00) and Q = 1106.690 on 305 degrees of freedom (p = 0.000) respectively.
Source: Authors' estimates.

The lower part of Table 2 reports the weighted average effect of aid separately for papers that include at least one of the above interaction terms and for those that do not include any of these interaction terms. The random effects estimate of the weighted average effect of aid for the latter group appears to be positive and statistically significant with a magnitude of 0.138. This magnitude is higher than the estimate found for papers that include at least one of the interaction terms. Moreover, this estimate is also higher than the one reported in Table 1 where issues with non-linearity are ignored.

Thus, overlooking the coefficients of aid squared, aid-policy and aid-institution interaction terms, in the calculation of the partial effect of aid on growth, leads to a biased weighted average effect from the aid-growth literature. While leaving out the coefficient of the aid squared term leads to an upward bias in the weighted average, the bias in the case of aid-policy and aid-institutions is downward. In light of this, the weighted average effect reported in DP08 is biased. To sum up, when one combines the existing empirical evidence from the 68 studies, the results suggest that the effect of aid on growth is about 0.14 and is statistically significantly different from zero.

Publication Bias versus Authentic Effect

Publication bias is typically said to exist when researchers, editors and reviewers tend to favour statistically significant findings causing studies that yield relatively small and/or insignificant results to remain unpublished (that is, remain 'in the file drawer'; see Stanley, 2005).[22] Whether this is indeed a problem in the aid-growth literature is not easy to say. In this literature, small and insignificant results have on several occasions drawn considerable academic and policy attention after which they have been shown not to be robust to even minor changes in data and methodology. Prominent examples include the 'micro-macro' paradox by Mosley (1986); the 'aid only works with good policy' hypothesis by Burnside and Dollar (2000); and the Rajan and Subramanian (2008) 'aid is insignificant' finding.[23] In any case, if a publication/small study bias exists it would tend to bias empirical effects, and as such must be carefully investigated with a view to disentangling any genuine empirical impact from publication effects. In line with established practice in the meta-literature we first use funnel plots to examine visually if the aid-growth literature seems to suffer from such bias.

Figure 1 presents a funnel plot using standard error as the measure of precision.[24] The vertical line at the centre of the funnel plot shows a summary estimate of the effect size from the 68 aid-growth studies. When there is no bias, estimates are expected to vary randomly and evenly around this estimate. The diagonal lines in the figure represent the 95 per cent confidence limits around the summary treatment effect for each standard error on the vertical axis.[25] These lines show the expected distribution space of studies in the *absence of heterogeneity*. That is, assuming

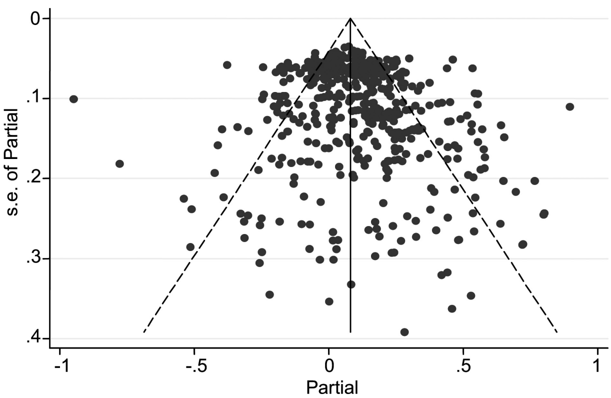

Figure 1. Funnel plot with pseudo 95 per cent confidence limits.

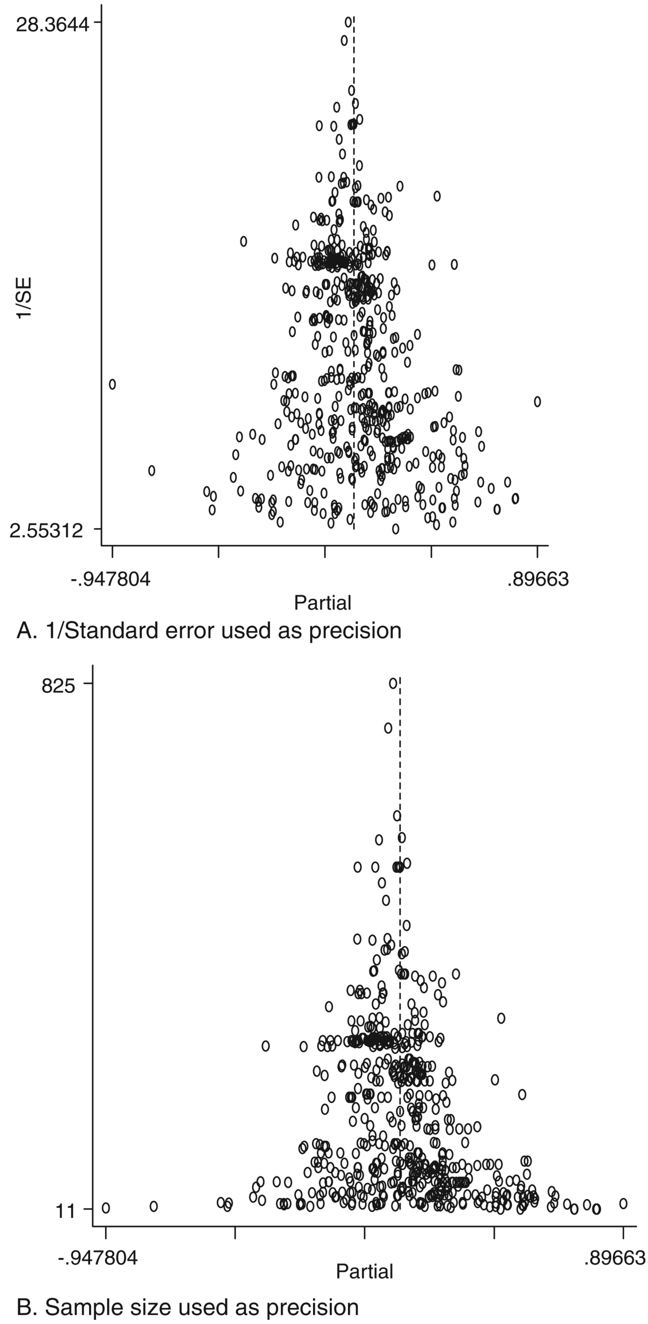

Figure 1.1. Funnel plots of the aid-growth literature.

there is no heterogeneity in the reported effect sizes among studies, 95 per cent of the studies should lie within the funnel defined by the diagonal lines.

As can be seen from the funnel plot in Figure 1, the estimates from the aid-growth literature are fairly randomly distributed around the fixed effect estimate. Although the distribution of the studies to the right of the funnel seems relatively more concentrated, there is no clear asymmetry in the funnel graph. This lack of asymmetry becomes clearly visible in Figure 1.1. Figure 1.1.A relies on the inverse of standard error as the measure of precision and is thus our preferred funnel plot. This figure depicts the clearly symmetrical distribution of the effect of aid on growth as estimated from the 68 studies. A similar impression is also observable in Figure 1.1.B with sample size as the measure of statistical precision (for comparison with DP08). In general, these funnel plots provide no basis to argue for a directional bias once one places the reference line at the correctly estimated overall empirical effect (see also Figure A3 in Mekasha and Tarp, 2011).

While the above funnel plot analysis provides no grounds to claim that a publication bias is present, it is premature to draw any firm conclusion about potential publication bias from this evidence. Even though funnel plots may be revealing, their interpretation is subjective and potentially ambiguous so statistical testing is required. The most commonly used statistical test of publication bias is the Egger et al. (1997) test, also known as the funnel asymmetry test (FAT) (Stanley, 2005). FAT basically estimates Equation (3), which is then expanded in a next stage to control for more explanatory variables.

The main variables of interest are the constant term and the coefficient of 'precision'. While the coefficient of 'precision' shows the magnitude and direction of any genuine underlying effect over and above any possible bias, the constant term depicts the existence and degree of the bias in the literature surveyed. The results of bivariate and multivariate meta-regression analysis are presented in Tables 3 and 4 respectively. In the bivariate FAT meta-regression-analysis (FAT-MRA) the dependent variable is the standardised effect of aid (t-statistics) regressed on the inverse of the standard error (that is, precision). Since more than one regression is taken from most of the studies, observations within a study are unlikely to be independent. To address this, standard errors are clustered on publications in all regressions.[26] For the sake of comparison, we also report heteroskedasticity consistent and heteroskedasticity and autocorrelation consistent (HAC) standard errors.

Despite the reasonable symmetry in the funnel plot discussed above, the result from the bivariate regression depicted in Table 3 seems to suggest presence of a positive and statistically significant publication bias. The positive sign of the bias suggests that small studies with high standard error tend to report a high partial effect of aid on growth, and hence a statistically significant effect. The FAT-MRA can also be relied on to identify genuine empirical effects of aid on growth regardless of publication bias (Stanley, 2008). In Table 3 this genuine empirical effect is captured by the coefficient of 'precision' and the FAT-MRA shows a positive and significant effect in column 1, but this does not appear to be the case when we apply HAC and clustered standard errors. The result reported in column 1 of Table 3 – concerning both publication bias and genuine empirical effect – is fundamentally the same as what DP08 find (they do not use HAC or clustered standard errors). This result is the main basis for their claim that aid is ineffective and publication bias is a problem in the aid-growth literature. Unlike DP08, however,

Table 3. Bivariate FAT meta regression analysis dependent variable=standardised effect (t-stat)

Variables	(1) Robust	(2) HAC	(3) Clustered
Bias coefficient			
Constant	0.794***	0.794***	0.794***
	(0.164)	(0.223)	(0.297)
Genuine effect of aid			
Precision	0.0245*	0.0245	0.0245
	(0.0142)	(0.01998)	(0.0260)
Observations	537	537	537
R-squared	0.005	0.005	0.005

Notes: Robust, heteroskedastcity and autocorrelation consistent and clustered standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.
Source: Authors' estimates.

Table 4. Multivariate FAT meta-regression analysis: reduced model dependent variable = standardised effect (t-stat)

Variables	(1) Robust	(2) HAC	(3) Clustered
Bias coefficient			
Constant	−0.232	−0.232	−0.232
	(0.321)	(0.308)	(0.350)
Genuine effect of aid			
Precision	0.166**	0.166**	0.166*
	(0.0733)	(0.0843)	(0.0924)
Publication outlet			
Working paper	−0.0697***	−0.0697***	−0.0697***
	(0.0167)	(0.0193)	(0.0184)
Cato	−0.202***	−0.202***	−0.202***
	(0.0324)	(0.0295)	(0.0282)
JDS	−0.0833***	−0.0833***	−0.0833***
	(0.0280)	(0.0271)	(0.0272)
JID	−0.0587***	−0.0587**	−0.0587*
	(0.0196)	(0.0239)	(0.0304)
EDCC	−0.146***	−0.146***	−0.146***
	(0.0389)	(0.0434)	(0.0501)
Applied economics	−0.116**	−0.116**	−0.116**
	(0.0545)	(0.0574)	(0.0519)
Author detail			
World Bank	−0.0853***	−0.0853***	−0.0853***
	(0.0204)	(0.0198)	(0.0178)
Gender	−0.0737***	−0.0737***	−0.0737**
	(0.0202)	(0.0258)	(0.0293)
Influence	0.0668***	0.0668***	0.0668***
	(0.0164)	(0.0167)	(0.0162)
Data			
Panel	0.105***	0.105***	0.105**
	(0.0379)	(0.0404)	(0.0426)
No. of years	−0.0106***	−0.0106***	−0.0106***
	(0.00162)	(0.00159)	(0.00152)
Asia	0.0303	0.0303	0.0303
	(0.0222)	(0.0222)	(0.0239)
Single country	0.491***	0.491***	0.491**
	(0.160)	(0.170)	(0.191)
y1960s	0.0547**	0.0547**	0.0547
	(0.0270)	(0.0289)	(0.0368)
y1990s	0.103***	0.103***	0.103***
	(0.0318)	(0.0329)	(0.0328)
Sub sample	0.0446**	0.0446***	0.0446**
	(0.0212)	(0.0169)	(0.0187)
Low income	−0.0879***	−0.0879***	−0.0879***
	(0.0284)	(0.0254)	(0.0328)
EDA	−0.0376**	−0.0376**	−0.0376**
	(0.0164)	(0.0176)	(0.0181)
Conditionality			
Aid square	0.0716***	0.0716***	0.0716***
	(0.0125)	(0.01015)	(0.0108)
Interaction institutions	−0.100***	−0.100***	−0.100**
	(0.0248)	(0.0291)	(0.0380)
Specification and control			
FDI	0.0909***	0.0909***	0.0909**
	(0.0258)	(0.0343)	(0.0417)

(*continued*)

Table 4. (*Continued*)

Variables	(1) Robust	(2) HAC	(3) Clustered
Theory	0.0415***	0.0415**	0.0415**
	(0.0155)	(0.0165)	(0.0191)
Average	0.0115***	0.0115***	0.0115***
	(0.00232)	(0.00211)	(0.00206)
Inflation	−0.0510**	−0.0510**	−0.0510***
	(0.0204)	(0.0198)	(0.0173)
Size of government	0.101***	0.101***	0.101***
	(0.0150)	(0.0142)	(0.0151)
Financial development	0.0345***	0.0345**	0.0345**
	(0.0129)	(0.0139)	(0.0142)
Region dummy	−0.0313**	−0.0313**	−0.0313**
	(0.0123)	(0.0130)	(0.0127)
Openness	−0.0706***	−0.0706***	−0.0706**
	(0.0185)	(0.0226)	(0.0274)
Per capita income	−0.0709**	−0.0709**	−0.0709*
	(0.0283)	(0.0318)	(0.0383)
Observations	518	518	518
Adj. R-squared	0.425	0.425	0.425

Notes: Q-test for heterogeneity: (chi2(518) = 1000; p > chi2 = 0.000). See Higgins and Thompson (2002) for the test of heterogeneity. Robust, Heteroskedastcity and Autocorrelation Consistent and clustered standard errors in parentheses. ***p < 0.01, **p < 0.05,*p < 0.1.
Source: Authors' estimates.

we do not believe that we should conclude and stop the analysis here. Digging deeper is revealing.

The above bivariate Egger et al. (1997) test is commonly criticised for leading to an inflated false-positive rate (high type I error), and such false positive results become a major issue especially when there is between study heterogeneity (see Ioannidis and Trikalinos, 2007). In a similar manner, Stanley (2005) argues that heterogeneity of effects may induce asymmetry into the funnel plots even in the absence of publication bias. This implies that failure to account for factors that can explain heterogeneity in research findings will potentially exaggerate the bias. As heterogeneity is evident in the aid-growth literature one should refrain from making inferences about publication bias based on the bivariate regression (Harbord et al., 2009). Instead, one needs to turn to a multivariate analysis.

In the aid-growth literature, drawing conclusions based on bivariate regression will obviously lead to misleading inferences for various reasons. The fact that some studies aim to estimate the direct impact of aid on growth while others focus on identifying the transmission channels (such as investment, health, education) makes effect estimates heterogeneous. That is, compared to the former, the direct effect of aid on growth is likely to be smaller in the latter case where the channels are already controlled for. Due to this, the reported effect estimates from the different regression models will obviously vary as a function of the controls included in the regressions, but this is not because of publication bias. Disregarding this fact in the FAT-MRA makes it look as if there is publication bias. Thus, one needs to incorporate information about the controls included in the underlying regressions of the primary studies in the biavariate FAT-MRA regressions. As Stanley (2005) also indicates, if such important information is not controlled for, the FAT-MRA will like any other econometric analysis suffer from omitted variable bias.

Accordingly, we expand the bivariate FAT-MRA model reported above into a more general FAT-MRA by including important explanatory variables that can potentially affect the reported variation (heterogeneity) in research findings.[27] We do not pretend to have insight on this point that goes beyond that of DP08. Accordingly, we first expand the FAT-MRA model by including all the 50 moderator variables they identified.The result from this regression can be found in

Mekasha and Tarp (2011), which shows that the magnitude of the precision coefficient improves and becomes significant in two of the three cases. Moreover and importantly, after controlling for factors that can potentially explain heterogeneity in reported effects, the bias coefficient (that is, the constant term) becomes insignificant in all cases. This suggests that once the moderator variables (study characteristics) are controlled for then there is no publication bias.

We also note that most of the variables included in the multivariate regression are also statistically insignificant. There is, in other words, a trade-off here between including all the 50 moderator variables in order to control for/explain heterogeneity versus potential multicollinearity and loss of degrees of freedom. Moreover, all controls are not equally important in contributing to the omitted variable bias and/or explaining heterogeneity. We therefore follow the General-to-Specific (GETS) modelling procedure by Krolzig and Hendry (2001) to systematically reduce the insignificant variables from the multivariate model. By doing so we eliminate 21 of the 50 moderator variables that appear to be non-important; the adjusted R^2 increase from 41 to 43 per cent, supporting the removal of the 21 moderators. The results from the reduced multivariate model are reported in Table 4.

As can be seen from the multivariate FAT-MRA results in Table 4, the genuine impact of aid on growth, as reflected in the coefficient of 'precision', is found to be positive and statistically significant in all three cases with a magnitude of 0.17. To put our results in perspective, we did a back-of-the-envelope calculation based on the estimates in Arndt et al. (2010).[28] This exercise shows that our finding is quite close to their estimate. Compared to the bivariate model, controlling for other variables which can potentially affect the reported variation of the effect of aid on growth greatly improves the magnitude of the genuine effect of aid. Moreover, in all the regressions the constant term, that is, the parameter used to test for existence of publication bias, becomes statistically insignificant. This is consistent with the result from the funnel plot and indicates lack of evidence to suggest presence of publication bias in the aid-growth literature.

On this basis we suggest that it is highly likely that the DP08 results suffer from omitted variable bias, noting that their conclusions are exclusively dependent on a bivariate FAT-MRA analysis. Mekasha and Tarp (2011: Table A3) also included a variety of robustness checks for the FAT-MRA results presented in Table 4. They include considering studies after the 1990s only; excluding studies that did not include African countries; and finally considering published studies only. In all cases the key finding presented in Table 4 holds.

The above evidence should, as Stanley (2005) puts it, be confirmed by a meta-significance test (MST) for authentic effect before firm conclusions are drawn. The MST test uses the relationship between the logarithms of a study's absolute value of t-statistics and the degrees of freedom to examine a genuine empirical effect. A genuine empirical effect is reflected in a positive and statistically significant coefficient of the log of degrees of freedom in Equation (4). The bivariate and multivariate results of our MST regressions are reported in Table 5 and Table 6 respectively.

Table 5. Bivariate MST meta regression analysis dependent variable=ln (t-stat)

Variables	(1) Robust	(2) HAC	(3) Clustered
ln(df)	0.00338	0.00338	0.00338
	(0.0474)	(0.0568)	(0.0635)
Constant	0.0637	0.0637	0.0637
	(0.219)	(0.258)	(0.277)
Observations	538	538	538
R-squared	0.000	0.000	0.000

Notes: Robust, heteroskedastcity and autocorrelation consistent and clustered standard errors in parentheses ***p < 0.01, **p < 0.05, *p < 0.1.
Source: Authors' estimates.

Table 6. Multivariate MST meta regression analysis: reduced model dependent variable = ln (t-stat)

Variables	(1) Robust	(2) HAC	(3) Clustered	(4) Clustered
Genuine empirical effect				
ln(df)	0.328***	0.328***	0.328***	
	(0.0847)	(0.0964)	(0.0820)	
ln(n)				0.365***
				(0.0942)
Publication outlet				
Working paper	−0.626***	−0.626***	−0.626***	−0.639***
	(0.145)	(0.138)	(0.140)	(0.139)
CATO	−1.390***	−1.390***	−1.390***	−1.402***
	(0.285)	(0.258)	(0.220)	(0.218)
JDS	−0.606**	−0.606**	−0.606**	−0.611**
	(0.235)	(0.254)	(0.265)	(0.263)
EDCC	−0.877	−0.877**	−0.877***	−0.867***
	(0.541)	(0.354)	(0.316)	(0.316)
AER	−1.029***	−1.029***	−1.029***	−1.035***
	(0.320)	(0.265)	(0.272)	(0.270)
Author details				
World Bank	−0.496**	−0.496**	−0.496**	−0.504**
	(0.203)	(0.194)	(0.213)	(0.212)
Gender	−0.400**	−0.400**	−0.400**	−0.402**
	(0.178)	(0.159)	(0.155)	(0.155)
Influence	0.334**	0.334**	0.334**	0.330**
	(0.135)	(0.129)	(0.129)	(0.130)
Data				
No. of years	−0.0357**	−0.0357**	−0.0357**	−0.0356**
	(0.0149)	(0.0168)	(0.0169)	(0.0169)
Africa	−0.286*	−0.286*	−0.286*	−0.297*
	(0.164)	(0.166)	(0.147)	(0.149)
Single country	1.426***	1.426***	1.426***	1.389***
	(0.300)	(0.298)	(0.252)	(0.249)
y1960s	0.399**	0.399*	0.399*	0.388*
	(0.201)	(0.217)	(0.233)	(0.231)
y1990s	1.016***	1.016***	1.016***	1.004***
	(0.203)	(0.211)	(0.209)	(0.207)
Conditionality				
Aid square	0.574***	0.574***	0.574***	0.573***
	(0.141)	(0.146)	(0.124)	(0.127)
Interaction institutions	0.822***	0.822***	0.822***	0.814***
	(0.216)	(0.217)	(0.217)	(0.216)
Specification and control				
FDI	0.576***	0.576***	0.576***	0.549***
	(0.173)	(0.145)	(0.137)	(0.137)
Gap model	0.294	0.294	0.294*	0.316**
	(0.211)	(0.185)	(0.149)	(0.151)
Theory	0.612***	0.612***	0.612***	0.618***
	(0.141)	(0.141)	(0.156)	(0.158)
Average	0.0530***	0.0530***	0.0530***	0.0540***
	(0.0128)	(0.0143)	(0.0119)	(0.0124)
Lag used	0.259	0.259	0.259	0.259
	(0.184)	(0.161)	(0.186)	(0.185)
Size of government	0.601***	0.601***	0.601***	0.596***
	(0.137)	(0.137)	(0.135)	(0.134)
Region dummy	−0.329**	−0.329**	−0.329***	−0.332***
	(0.148)	(0.124)	(0.0952)	(0.0952)

(*continued*)

Table 6. (*Continued*)

Variables	(1) Robust	(2) HAC	(3) Clustered	(4) Clustered
Openness	−0.275**	−0.275**	−0.275**	−0.276**
	(0.124)	(0.123)	(0.120)	(0.122)
Constant	−1.681***	−1.681***	−1.681***	−1.873***
	(0.434)	(0.470)	(0.354)	(0.403)
Observations	519	519	519	519
Adj. R-squared	0.203	0.203	0.203	0.202

Notes: Test for heterogeneity: (chi2(518)=550.16; P>chi2=0.317) Robust, Heteroskedastcity and Autocorrelation Consistent and Clustered standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.
Source: Authors' estimates.

As can be seen from the bivariate regression reported in Table 5, the coefficient of log of degrees of freedom (ln (df)) exhibits a positive sign, but is insignificant in all cases. This should come as no surprise. The results reported in Table 5 are from a bivariate regression, and it is likely that this bivariate MST-MRA suffers from omitted variable bias for reasons similar to those discussed above.

We therefore turn again to the DP08 explanatory variables used for the FAT-MRA in Table 4 and run a multivariate MST-MRA. The results from the full model are presented in Mekasha and Tarp (2011); the first three columns of Table 6 report the reduced form MST-MRA model after systematically removing insignificant variables using the GETS modelling procedure. Column 4 of Table 6 checks if the result remains the same when one uses the log of the number of observations (ln(n)) instead of the log of degrees of freedom (ln(df)) as a measure of estimation accuracy.

As Table 6 demonstrates, in all the multivariate MST-MRA regressions, the coefficient of estimation accuracy is positive and significant. This underpins the authenticity of the positive and significant effect of aid on growth observed in the FAT-MRA regressions. Moreover, similar to the FAT robustness checks, the key finding holds under robustness checks for the MST-MRA results (Mekasha and Tarp, 2011).

Our results from MST-MRA also stand in contrast to the conclusions of DP08. They found a negative and insignificant coefficient on ln(df) and suggested that there is a lack of evidence to support the idea that development aid has an effect on economic growth. Once again, this is based on a simple bivariate MST, which fails to take into account other explanatory variables. This negative conclusion on aid effectiveness does not survive when the bivariate model is expanded to the multivariate context.

IV. Conclusions

Our main aim was to contribute to the aid-growth literature and associated policy debates using meta-analysis of the 68 studies employed by DP08. We also use the same measure of effect size (partial correlation) and the same moderator variables for the multivariate analysis. There are four major differences: (i) model selection, that is, the choice between the fixed and random effects model; (ii) the way the effect size is treated for papers that include non-linear terms; (iii) choice of statistical weighting for the effect estimates both in calculating the weighted average effect and in the funnel plots; (iv) in the process of data entry and coding. Having fully replicated the analysis by DP08 and identified our differences, we expand their meta-analysis in various ways that better reflect the econometric, statistical and data challenges faced in this type of research. What did we find?

On the data issues, some recoding and filling in missing values resulted in an increase in the number of observations for our meta-analysis.

In relation to model selection, our results show that the fixed effect model assumption of a single true effect common to all studies is unrealistic in the aid-growth literature. Specifically, both graphical inspection and statistical testing reveal that there is heterogeneity in the estimate of the true effect of aid on growth across the 68 studies. Furthermore, the effect homogeneity assumption can be rejected, from the outset, on theoretical grounds. We thus emphasise that the random effects model is to be preferred as it allows for between study heterogeneity.

Accordingly, we calculated the weighted average effect of aid on growth (using the inverse of the variance as weight) relying on the random effects model. Our results show that the weighted average effect of aid on growth from the 68 studies is positive and statistically significant with a magnitude of 0.098. This finding stands in contrast to DP08, and based on the random effects results their estimate of 0.08 is rejected at the 5 per cent level of significance.

The partial effect of aid on growth for regressions that include interaction terms is not measured correctly as the coefficient of the interaction term(s) is not taken into account. This can potentially bias the weighted average effect of aid. By calculating the weighted average effect of aid separately for regressions with and without non-linear terms, the weighted average effect estimate of aid in the random effects model emerges as positive and significant with a magnitude of 0.14. This shows how disregarding the coefficient of the interaction terms in the calculation of the partial effect matters for the results. We thus suggest that future meta-analysis of aid and growth needs to find a way to properly incorporate the partial effect of aid from studies that include a non-linear term.

Having calculated the weighted average effect of aid on growth from the 68 studies, we moved on to check whether the observed effect is genuine or an artefact of publication bias using FAT-MRA and the General-to-Specific (GETS) modelling approach in choosing the important study characteristics (moderator variables) that help to explain the heterogeneity in research design across studies. The multivariate FAT-MRA results clearly suggest that publication bias is not a problem in the aid-growth literature once the heterogeneity is controlled for. The measure of publication bias obtained from the multivariate FAT-MRA model appears to be statistically indistinguishable from zero, which is in line with the reasonably symmetrical funnel plot depicted in Figure 1.1.A. In the same vein, the FAT-MRA results reported in Table 4 also confirm the positive and significant effect of aid on growth as depicted by the positive and statistically significant coefficient of precision.

The genuineness of the observed effect and hence the absence of publication bias in the aid-growth literature is further underpinned by the results of our MST-MRA regressions. As shown in Table 6, there is evidence of a clear empirical effect that goes beyond publication bias. Though the coefficient that verifies the authenticity of the impact of aid on growth is not significant in the bivariate MST, the authenticity of the observed positive and significant aid-growth impact becomes evident once we move to a multivariate setting. As shown in Mekasha and Tarp (2011), these findings are robust in different samples.

We also highlight the importance of heterogeneity in the true effect of aid on growth across the studies under review. As is evident from the Q-test for heterogeneity reported under Table 4, there still exists excess (unexplained) variation despite the inclusion of the relevant moderator variables. This confirms the presence of real heterogeneity in the true effect of aid on growth that goes beyond heterogeneity in research design. This is again consistent with the assumptions inherent in the random effects model and shows that the effect homogeneity assumption of the fixed effect model is not tenable.

To sum up, we have shown that the conclusions in DP08 do not hold when one applies meta-analysis rigorously to the aid-growth literature. We found a positive and significant effect of aid on growth and importantly found no evidence to suggest presence of publication bias. That said, and as pointed out from the outset, even if meta-analysis can potentially address the subjectivity associated with narrative literature reviews, it is far from flawless. For instance, subjectivity remains a threat in meta-analysis unless researchers carefully handle the judgement calls they encounter in various stages of the meta-research process. Moreover, differences in the quality of the primary studies and the lack of a reasonably objective tool to measure quality appear to be a major caveat, and especially for observational data-based studies. Differences in quality can lead to heterogeneity in effect estimates and unless properly captured, the heterogeneity can wrongly be perceived as publication bias. On top of this, and as our results show, there is real heterogeneity in the true effect of aid on growth that goes beyond methodological heterogeneity; that is, heterogeneity persists even after controlling for study characteristics. Given also the fact that meta-analysis is a method, which is more appropriate for data generated through Random Control Trials, the application to the aggregate aid-growth literature should only be undertaken with great caution.

One should not overstate the implications of the results from macro-level aid-growth meta-analysis. Nonetheless, such analysis, if applied rigorously according to best practice, can help in giving useful insight into the qualitative aspects of the research process; for example, to identify the presence or absence of publication bias in the literature under consideration. Besides, identification of the most relevant study characteristics that explain heterogeneity in the effect estimate can be relied on to improve research design of future primary studies. Regarding the quantitative results, although we do find evidence that is in line with Arndt et al. (2010), we remain vigilant in drawing strong implications. This is not only because of the limitations of applying meta-analysis to the macro-level aid-growth literature but also due to the fact that the estimate here is obtained by combining inherently heterogonous effect estimates. Moreover, the conclusions that emerge from the present review are obviously not the whole story about aid effectiveness. Economic growth, though important, is only one of the multifaceted development objectives of foreign aid. It should be noted that poverty reduction is now the main aim and target of most foreign aid programmes.[29] Finally, we fully agree with calls to improve the design and implementation of aid to the benefit of the poorest people in the poorest countries. Aid processes are complex and few would (and certainly not the present authors) dispute that they can be improved.

Acknowledgements

We thank Channing Arndt, Tony Addison, Christian Friis Bach, Andrew Berg, Carl-Johan Dalgaard, Henrik Hansen, Imed Drine, Sam Jones, Katarina Juselius, Carol Newman, John Rand, and Danielle Resnick for valuable comments and advice on the issues addressed in this article. The same goes for observations by participants in a number of UNU-WIDER organised conferences, including events with the African Economic Research Consortium (AERC) in Nairobi, the United Nations University Office in New York (UNU-ONY), DFID and Danida as well as Sida in London, Copenhagen and Stockholm respectively. Conferences organised by the Helsinki Centre of Economic Research (HECER) in Helsinki and the Nordic Conference in Development Economics (NCDE) in Copenhagen are also acknowledged. Finally, most insightful comments and critique from an anonymous referee have helped us sharpen our analysis and exposition, and the same goes for excellent editorial advice. Hristos (Chris) Doucouliagos and Martin Paldam kindly provided their data and also comments on an earlier version of this article. The usual caveats apply.

Notes

1. See, for example, Mosley (1986), White (1992), Tsikata (1998), Burnside and Dollar (2000), Morrissey (2001), Dalgaard et al. (2004), Tarp (2006), McGillivray et al. (2006), Rajan and Subramanian (2008) and Arndt et al. (2010), among many others.
2. These include Funnel Asymmetry Test (FAT), Meta Significance Test (MST), and a meta-regression analysis (MRA). As regards the MRA both fixed and random model effect results are reported by DP08, who opt for relying on the fixed effect approach (see DP08: 13).
3. Doucouliagos and Paldam (2011) expand the dataset and provide a brief update of DP08 but their focus as well as basic methodological choices and conclusions are the same.

4. This should not be taken as an approval of the list of papers identified by DP08. The literature is large and complex.
5. Even if we consider the entire moderator variable set used in DP08 to begin with, we eventually focus on the relevant ones using a General-to-Specific modelling approach to reduce the set.
6. In addition to showing how these differences matter for the results and conclusions, we began by fully replicating the results of DP08.
7. Note that in the case of the aid squared term, the partial effect is $\beta_1 + 2\beta_2 * aid$.
8. We are aware of the conditional aid effectiveness meta-analysis in Doucouliagos and Paldam (2010) (DP10), applying a similar meta-analysis as in DP08. In the 2010 paper, the authors conclude that the aid effectiveness literature has failed to establish non-linear terms like aid squared. This conclusion is, however, only as valid as the meta-methodology employed in the paper. Besides, from a total of 147 regressions that include an aid squared term, 100 show a negative and significant coefficient for the aid-squared term (See DP10: 400, Table 2). In view of this, even if one accepts the conclusion in DP10, it only implies that the coefficients of the interaction terms, on average, 'should be zero' rather than indicating that the coefficient of the interaction term from each paper 'is actually zero'. Therefore, the findings of DP10 do not address/justify the concern that we have pointed out regarding the treatment of interaction terms in DP08.
9. See Table A9.1 in Mekasha and Tarp (2011). Note also that in our data we do not include the variable 'Danida affiliation'. None of the three authors classified by DP08 as Danida affiliated (studies 12, 13, 33, 34 and 40) fell into this category when the studies were examined.
10. Note that we were able to increase the number to 519 by re-coding the values of the moderator variables which, for some studies, were wrongly coded as missing in DP08 (Table A8 in Mekasha and Tarp, 2011).
11. See Greenland (1994), Higgins and Green (2011) and Jüni et al. (1999).
12. Card (2011) suggests that if differences are found, conclusions must be restricted to those studies that the researcher thinks produce most valid results. But here one can argue that this suggestion cannot be appealing if the sample size for the meta-analysis is small.
13. Controlling for 'some aspects of quality' enables the researcher to tell which aspects of quality affect the reported effect size and hence can guide future design of primary studies, Card (2011).
14. Here it should be noted that meta-analysis of micro-level observational studies can be more informative as they have a well-defined treatment and better comparability compared to macro-level primary studies. For an example of micro-level meta-analysis see Havranek and Irsova (2011, 2012).
15. We removed one regression from the study (ID 30) as this regression is already included (coded) in study ID 29. In study ID 30, the author used the regression from study ID 29 purely for comparative purposes. Thus, correcting for this double coding leads to 542 observations rather than 543.
16. The terms fixed and random effects used in meta-analysis are quite different from the ones applied in standard panel data models in econometrics. In meta-analysis the difference between fixed and random effects models originate from the underlying assumption as regards the nature of the 'true' effects.
17. The term effect size refers to the magnitude of the effect observed in each study. In the meta-literature there are different metrics to measure this; the partial correlation coefficient being the most commonly used. As in DP08 we calculate the partial correlation coefficients of each study using $\sqrt{\frac{t^2}{t^2+df}}$ where t and df refer to t-statistics and degrees of freedom respectively.
18. We also present the funnel plots with sample size for comparison with DP08, but our preferred measure of precision follows Sterne and Egger (2001).
19. DP08 (pp. 8-10) indicate (but do not report) that the weighted average is statistically insignificant.Applying the standard fixed and random effects model on the *original DP08 data* shows that the aid-growth weighted average effect is positive and statistically significant both in the fixed and random effects model with a magnitude of 0.078 and 0.093 respectively.
20. The test involves $Q = \sum_{i=1}^{k} w_i (T_i - T)^2$ where T_i is the estimate of the effect magnitude, T is the weighted average and w_i is the weight (the inverse of the variance of T_i). Under the null hypothesis of homogeneity, Q is distributed as chi-square with degrees of freedom equal to the number of studies minus one.
21. Even when one applies the heterogeneity tests on the original DP08 data, there is no ground to accept the effect homogeneity assumption of the fixed effect model.
22. Also, small studies tend to have large standard errors leading to insignificant results. If this leads authors to strive to come up with large-sized effects in order to compensate for the high standard errors such a bias should be detected.
23. See Hansen and Tarp (2000), Hansen and Tarp (2001), Dalgaard et al. (2004), and Arndt et al. (2010).
24. When standard errors are along the vertical axis, the vertical axis is reversed (zero at the top), so as to put large studies at the top of the graph reflecting that larger studies have smaller standard errors.
25. The summary estimate of the effect size in Figure 1 is obtained from the fixed effect model (under the effect homogeneity assumption). This presents one limitation in funnel plot analysis. Vevea and Hedges (1995) explain why one should not necessarily associate asymmetry in the funnel plot with publication bias. Presence of heterogeneity can also potentially lead to such an asymmetry in the funnel plot.
26. In DP08 the results appear to be very sensitive to clustering, see Table A5 and discussion in Mekasha and Tarp (2011).
27. See for example Rose and Stanley (2005), Abreu et al. (2005) and Stanley (2005, 2008).

28. Note that these estimates are not directly comparable as the estimate in the present article is given as a partial correlation while the one in Arndt et al. (2010) takes an elasticity interpretation. We make the comparison by first changing the coefficient estimates from Arndt et al. (2010) to a partial correlation using the same formula used to convert the coefficient estimates of the primary studies included in this meta-analysis. Accordingly, we convert a total of 10 regressions from Arndt et al. (2010) to partial correlation and we get a weighted average of 0.173 which is the same as what we get in this meta-analysis. Note, however, that if we focus only on the most preferred regressions from Table 4 in Arndt et al. (2010), this weighted average effect will be 0.26.
29. See Feeny and Ouattara (2009), Feeny (2003) and Gomanee et al. (2005).

References

Abreu, M., de Groot, H.L. and Florax, R.J. (2005) A meta-analysis of beta-convergence: the legendary 2 per cent. *Journal of Economic Surveys,* 19(3), pp. 389–420.

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

Borenstein, M., Hedges, L. and Rothstein, H. (2007) Meta analysis: fixed effect vs. random effects. Accessed at http://www.sciencedownload.net/demodownload/Fixed%20effect%20vs.%20random%20effects.pdf.

Bullock, R.J. and Svyantek, D.J. (1985) Analyzing meta-analysis: potential problems, an unsuccessful replication, and evaluation criteria. *Journal of Applied Psychology,* 70(1), pp. 108–115.

Burnside, C. and Dollar, D. (2000) Aid, policies, and growth. *American Economic Review,* 90(4), pp. 847–868.

Card, N.A. (2011) *Applied Meta-Analysis for Social Science Research: Methodology in the Social Sciences* (New York: Guilford Publications).

Chauvet, L. and Guillaumont, P. (2004) Aid and growth revisited: policy, economic vulnerability and political instability, in: B. Tungodden, N. Stern and I. Kolstad (eds) *Toward Pro-Poor Policies – Aid, Institutions and Globalization* (Washington, DC: World Bank and Oxford University Press), pp. 95–109.

Dalgaard, C.-J., Hansen, H. and Tarp, F. (2004) On the empirics of foreign aid and growth. *Economic Journal,* 114(496), pp. F191–F216.

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

Doucouliagos, H. and Paldam, M. (2010) Conditional aid effectiveness: a meta-study. *Journal of International Development,* 22(4), pp. 391–410.

Doucouliagos, H. and Paldam, M. (2011) The ineffectiveness of development aid on growth: an update. *European Journal of Political Economy,* 27(2), pp. 399–404.

Egger, M., Smith, G.D., Schneider, M. and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal,* 315(7109), pp. 629–634.

Feeny, S. (2003) The impact of foreign aid on poverty and human wellbeing in Papua New Guinea. *Asia Pacific Development Journal,* 10(2), pp. 73–93.

Feeny, S. and Ouattara, B. (2009) What type of economic growth does foreign aid support? *Applied Economics Letters,* 16(7), pp. 727–730.

Gomanee, K., Girma, S. and Morrissey, O. (2005) Aid, public spending and human welfare: evidence from quantile regressions. *Journal of International Development,* 17(3), pp. 299–309.

Greenland, S. (1994) Invited commentary: a critical look at some popular meta-analytic methods. *American Journal of Epidemiology,* 140(3), pp. 290–296.

Hansen, H. and Tarp, F. (2000) Aid effectiveness disputed. *Journal of International Development,* 12(3), pp. 375–398.

Hansen, H. and Tarp, F. (2001) Aid and growth regressions. *Journal of Development Economics,* 64(2), pp. 547–570.

Harbord, R.M., Harris, R.J. and Sterne, J.A. (2009) An updated test for small-study effects in meta-analyses, in: J.A. Sterne (ed.) *Meta-Analysis in Stata: An Updated Collection from the Stata Journal* (College Station, TX: Stata Press), pp. 138–148.

Havranek, T. and Irsova, Z. (2011) Estimating vertical spillovers from FDI: why results vary and what the true effect is. *Journal of International Economics,* 85(2), pp. 234–244.

Havranek, T. and Irsova, Z. (2012) Publication bias in the literature on Foreign Direct Investment spillovers. *Journal of Development Studies*, 48(10), pp. 1375–1396.

Higgins, J.P. and Green, S. (2011) *Cochrane Handbook for Systematic Reviews of Interventions. Wiley Cochrane Series* (Chichester: John Wiley).

Higgins, J.P. and Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine,* 21(11), pp. 1539–1558.

Ioannidis, J.P. and Trikalinos, T.A. (2007) The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal,* 176(8), pp. 1091–1096.

Jüni, P., Witschi, A., Bloch, R. and Egger, M. (1999) The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association,* 282(11), pp. 1054–1060.

Krolzig, H.-M. and Hendry, D.F. (2001) Computer automation of general to specific model selection procedures. *Journal of Economic Dynamics and Control,* 25(6–7), pp. 831–866.

McGillivray, M., Feeny, S., Hermes, N. and Lensink, R. (2006) Controversies over the impact of development aid: it works; it doesn't; it can, but that depends. *Journal of International Development,* 18(7), pp. 1031–1050.

Mekasha, T.J. and Tarp, F. (2011) Aid and growth: what meta-analysis reveals. *WIDER Working Paper WP/22*, World Institute for Development Economics Research (UNU-WIDER), Helsinki.

Morrissey, O. (2001) Does aid increase growth? *Progress in Development Studies,* 1(1), pp. 37–50.

Mosley, P. (1986) Aid-effectiveness: the micro-macro paradox. *IDS Bulletin,* 17(2), pp. 22–27.

Rajan, R.G. and Subramanian, A. (2008) Aid and growth: what does the cross-country evidence really show? *The Review of Economics and Statistics,* 90(4), pp. 643–665.

Rose, A.K. and Stanley, T.D. (2005) A meta-analysis of the effect of common currencies on international trade. *Journal of Economic Surveys,* 19(3), pp. 347–365.

Schell, C.L. and Rathe, R.J. (1992) Meta-analysis: a tool for medical and scientific discoveries. *Bulletin of the Medical Library Association,* 80(3), pp. 219–222.

Stanley, T.D. (2001) Wheat from chaff: meta-analysis as quantitative literature review. *Journal of Economic Perspectives,* 15(3), pp. 131–150.

Stanley, T.D. (2005) Beyond publication bias. *Journal of Economic Surveys,* 19(3), pp. 309–345.

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

Sterne, J.A. and Egger, M. (2001) Funnel plots for detecting bias in meta analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology,* 54(10), pp. 1046–1055.

Sterne, J.A. and Harbord, R.M. (2009) Funnel plots in meta-analysis, in: J.A. Sterne (ed.) *Meta-Analysis in Stata: An Updated Collection from the Stata Journal* (College Station, TX: Stata Press), pp. 109–122.

Tarp, F. (2006) Aid and development. *Swedish Economic Policy Review,* 13(3), pp. 9–61.

Temple, J. (2010) Aid and conditionality, in: D. Rodrik and M.Rosenzweig (eds) *Handbook of Development Economics*, Volume 5 (Amsterdam: Elsevier), pp. 4416–4522.

Tsikata, T. (1998) Aid effectiveness – a survey of the recent empirical literature. *IMF Papers on Policy Analysis and Assessments 98/1*. Washington, DC: International Monetary Fund.

Vevea, J. and Hedges, L. (1995) A general linear model for estimating effect size in the presence of publication bias. *Psychometrika,* 60(3), pp. 419–435.

Wanous, J.P., Sullivan, S.E. and Malinak, J. (1989) The role of judgment calls in meta-analysis. *Journal of Applied Psychology,* 74(2), pp. 259–264.

White, H. (1992) What do we know about aid's macroeconomic impact? An overview of the aid effectiveness debate. *Journal of International Development,* 4(2), pp. 121–137.