

Finally a breakthrough?

The recent rise in the size of the estimates of aid effectiveness

By Hristos Doucouliagos¹ and Martin Paldam²

For Mak Arvin and Byron Lew eds., 2015. *Handbook on the Economics of Foreign Aid*. Edward Elgar. Should be out in December this year

Abstract:

The authors have previously surveyed the AEL, aid (empirical) effectiveness literature, using the technique of meta-analysis. We reached the result that the small positive effect of aid on growth found in the average study is mostly a publication selection bias. This present study concentrates on the new literature from 2007 onward which is divided into: Period (A) 2007-08: where the AEL showed aid ineffectiveness. Period (B) 2009-11: where the results are better. Three hypotheses may explain the upward kink in the results: (i) Aid effectiveness has increased. (ii) A breakthrough has occurred in the models and estimators. (iii) The improvement is an artifact. Unfortunately the data support the third hypothesis only.

Keywords: Aid effectiveness, meta-analysis

JEL codes: B41, F35

Acknowledgements: We are grateful to all of those who have discussed our meta-studies of aid effectiveness over the years. This is a literature that has come a long way. At present we thank the editor and the referee.

1. School of Accounting, Economics and Finance, Deakin University, 221 Burwood Highway, Burwood, Victoria 3125, Australia.

E-mail: douc@deakin.edu.au. URL: <http://www.deakin.edu.au/contact/staff-profile/?pid=435>

2. Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark.

E-mail: mpaldam@econ.au.dk. URL: <http://www.martin.paldam.dk/>.

1. Introduction: A disappointing literature with a recent hope

If development aid increases development, we say that aid is effective. It is an important and debated question if this is the case. The debate has led to the empirical *AEL*, Aid Effectiveness Literature, which by now consists of more than 200 papers containing more than two thousand estimates of aid effectiveness.³ The results vary greatly.

1.1 *The AEL, aid effectiveness literature*

The AEL operationalizes aid effectiveness to mean that the effect of aid on growth is robustly positive and economically significant in a statistical analysis of the data.⁴ At a first look aid does not work: The univariate correlation between aid and growth is zero or even marginally below zero (see section 3.2).

Most researchers share a prior: We want aid to work for moral/political reasons, so the AEL has refused to consider the zero-correlation result as the final word. Also, aid (net ODA) amounted to 125 billion US \$ in 2011. It is produced by a complex multinational sector of the economy referred to as the *aid industry* that sponsors a good deal of the AEL research. The industry has an interest in results showing that aid works. It is well known that sponsor interests create priors that bias research in many fields. Researcher priors and sponsor interests mean that there is a risk of bias in the AEL. This risk is not unique to the AEL, but it needs to be acknowledged nevertheless.

The economics profession has plenty of tools in the box allowing it to put structure on univariate relations, and this has generated the AEL. The papers have controlled the relation for: (i) the effect of about 60 other variables; (ii) simultaneity; (iii) outlying countries and exceptional years; (iv) interaction terms, etc.

On the face of it this effort has borne fruit as most AEL-models give positive estimates of aid effectiveness, see Figure 1 below. However, the models have proved highly unstable. Consequently, the hunt for the good model that is both robust and gives the “right result” of aid effectiveness has continued.

3. The papers coded goes from the start of this literature to the end of 2011, but as we cover working papers most of the papers found in journals in 2012-13 are included as well.

4. It is often claimed that this operationalization is too narrow. This will be discussed at the end of the paper. It should be mentioned that the AEL considers averages of the variables over 3-7 years to reduce activity effects, lagged reactions, and the difference between commitment and disbursement.

1.2 *The literature – a survey of our meta-studies*

The highly variable results and the risk of bias in the AEL make it difficult to survey the results. The authors recently completed a major project analyzing the AEL using meta-analysis that is developed precisely to extract the best average from such literatures.⁵

The technique of meta-analysis is old in medicine, where an experiment is an independent and expensive clinical trial. In economics (notably macro) data sets often overlap, so studies are dependent, and an experiment is a regression that costs next to nothing. Hence, many regressions are run before the best is *selected* for publication. Therefore, the scope for publication selection bias is larger in economics than in medicine, where such bias is often found. The tools of meta-analysis have been developed for economics in the last decade (Stanley and Doucouliagos 2012). The new tools have led to a wave of meta-studies.

The standard tool for examining the distribution of a set of estimates is the funnel plot (Stanley and Doucouliagos 2010). Theoretically funnels should be narrow (corresponding to the average t-ratio) and symmetric – empirical funnels are typically much wider and asymmetric than theory predicts. This points to *polishing*; *i.e.*, to the selection of results with high t-ratios and *censoring* at places predicted by the priors of researchers. This has led to the FAT, which tests for asymmetry and the PET meta-average that corrects the average for the asymmetry detected, see section 2.3 on the FAT-PET.

In the AEL both research priors and sponsor interests suggest a censoring bias at zero, so that the profession publishes too few negative results. Thus, it is clear what kind of asymmetry to expect. The distribution of the results often has an asymmetry precisely as expected; see Figure 2b and Table 2 below.

We commenced our quest using the search window from the start of the literature till 2005 (see D&P 2008), where 68 comparable studies were found.⁶ The simple mean of the estimates is fairly small but positive. When the PET meta-average is used to control for the asymmetry the coefficient becomes insignificant both economically and statistically. Consequently, we reached the sad result that the AEL has failed to show that aid works, on average.

Since 2005 the AEL has increased significantly, but D&P (2011) show that the new data did not change the ineffectiveness result. Furthermore, the many attempts in the new

5. The papers in our project are listed in section 2 of the references. They are referred to as D&P (year).

6. That study is supplemented with D&P (2010) that analyzes the main AEL-models using interacted terms. The *good policy model* (where aid and a good-policy index are interacted) of Burnside and Dollar (2000) and the *medicine model* (where aid is interacted with itself) from Hansen and Tarp (2000). Our study concludes that the two models have not survived replication outside the original data-set.

literature trying to sort out causality by the various methods proposed in the econometric literature have not changed the result, as further discussed in section 2.4.⁷

The last paper in our project is D&P (2013a) where we consider the literature trying to catch the reverse causality: The effect of growth on aid. Here the models are less numerous, but the result is much the same as the aid effectiveness result: A small positive but insignificant coefficient is found. This corroborates the causality results in D&P (2011).

We have had the good fortune that our study D&P (2008) has been critically replicated by Mekasha and Tarp (2013). They fully validated our basic analysis.⁸

These meta-studies have concluded that the AEL has consistently found an effect measured as the partial correlation of aid on growth of about 0.03, when it is corrected for publication selection bias. However, since 2008 the average results of the AEL have been better than in the previous decade.

1.3 The purpose of the paper: Is the upward kink in the AEL results from 2009 genuine?

The purpose of this paper is to analyze why an upward kink in the AEL results has occurred in recent years. Three hypotheses are considered: (i) it is a breakthrough in aid effectiveness, (ii) it is a breakthrough in the analytical methods of the AEL, and finally (iii) it is an increase in publication selection bias.

The analysis proceeds as follows. Section 2 looks at the path over time in the AEL and discusses the logic of our chronological categorization. This section also reports meta-analyses for the various periods. Then alternative evidence is considered: Section 3 looks at the path of income and aid, and section 4 looks at hypothesis (i) that aid policies have improved. Section 5 returns to the AEL and discusses hypothesis (ii) that a breakthrough has occurred in modeling and estimators, while section 6 discusses the increasing evidence on publication selection bias. Finally, section 7 pulls together the evidence from the meta-studies and from the other sources, showing that the results from the various sections tally rather well.

7. D&P (2011) found encouraging evidence that some *types* of aid may be effective in generating growth.

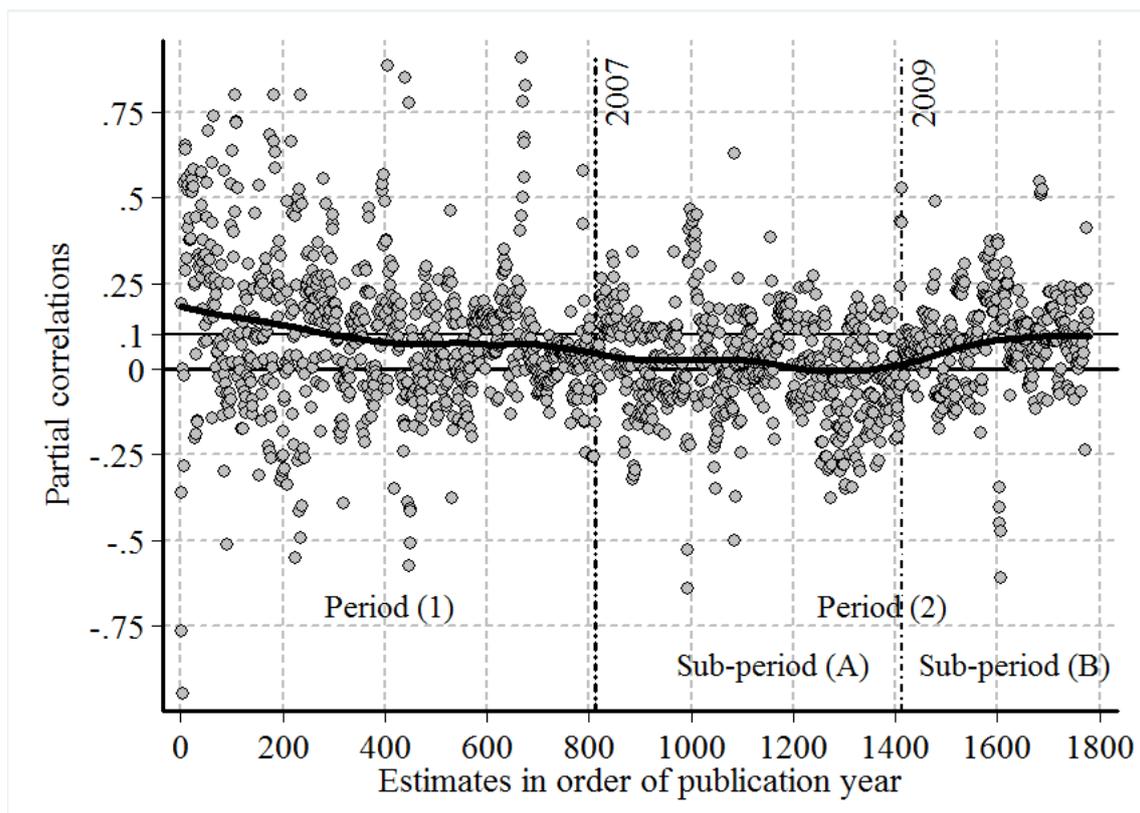
8. However, in addition Mekasha and Tarp (2013) challenged our multivariate MRA results (not discussed at present). Our response to their challenge can be found in D&P (2013b).

2. An analysis of the periods of the AEL

This section looks at the 1,777 estimates of the effect of aid on growth we have compiled from the AEL⁹. They are made comparable by a conversion of regression coefficients into partial correlations.

Section 2.1 shows the time path of the estimates and divides that path into different periods. Section 2.2 discusses the curious case of the falling trend from 1970 to 2008. Section 2.3 is a brief explanation of the meta-tools used, while section 2.4 considers some of the most common objections to meta-analysis. Section 2.5 looks at the funnel graphs showing the distributions of the results, while sections 2.6 and 2.7 report a basic meta-analysis. Section 2.8 tries to interpret the partial correlations in economic terms.

Figure 1. All estimates with kernel curve showing the movements of the average



Note: The kernel-curve is an Epanechnikov kernel with bandwidth 100. This is a moving average with a fixed bandwidth – smoothed by a polynomial showed by Epanechnikov to be optimal.

9. The data are drawn from 141 studies, from the start of the AEL in 1970 to the end of 2011, see D&P 2012, 2013b. They can be downloaded as a Stata file from <http://www.martin.paldam.dk/Meta-AEL.php>. These data have been rechecked 3 times so they are used in all calculations in the present paper as well; even if some of the latest working papers included have been published and new working papers have been published.

2.1 *The time path of the estimates in the AEL: Choosing the periods*

The path of the estimates is shown on Figure 1. The first thing the reader should note is the amazing variability of the estimated effect. Partial correlations are by definition restricted to the interval from -1 to $+1$, and almost the full interval is used.

The time axis is not equidistant as the step-length depends on the density of the estimates. The ordering of the papers is only precise for the year of publication, but in several cases working papers are included. This messes up the ordering a little as the lag between the working paper and the eventual publication is one to five years. Consequently a formal test for structural breaks is not presented. Instead the technique of kernel regression is used to identify the average time-path. The curve is always positive, though it comes down to touch zero in 2008. We interpret the curve as V-formed, with a well-defined kink at the end of 2008.¹⁰ Also, note that from 2000 the curve keeps below 0.1.

The period covered by D&P (2008) is taken as the old period, and two more years with broadly similar results are added, so that the old period (1) goes to the end of 2006. This paper concentrates on the new period (2) commencing in 2007. The new research has two parts with a kink in late 2008. This finding is the basis for sub-periods (A) and (B) used in sections 2.5 and 2.6. (A) 2007-08 are the dark years where the estimates are almost zero. Period (B) 2009-11 is after the kink. Here the results are better, suggesting aid effectiveness.

2.2 *The curious trend till late 2008*

The 38 years from 1970 to late 2008 before the kink have a falling path that is rather linear. It starts at almost 0.2 and eventually goes to 0. This corresponds to a fall of about $(0.2/38 \approx) 0.005$ per year. This is 2.5 % of the 0.2. If this is a fact, it implies that the aid industry unlearned by doing at a rate of 2.5 % per year.

This appears outright impossible. Studies from many fields show learning by doing. Learning might stagnate after some time, but a steadily negative learning by doing seems to be unique (see Thompson 2010). We conclude that the negative trend is no fact, but an artifact.

Aid-data started in the mid-1960s. Since then data has accumulated. With more data the variation in the estimates around the average falls. Imagine that most of the results selected for publication are chosen from the best estimates made. Then the fall in the reported aid effectiveness reflects the increased data sample.

10. The kernel-curve was estimated for bandwidth, $bw = 25, 50, 100, 150, 200, 250$ and 400 . For $bw < 100$ the curve is more volatile reflecting individual studies, and for $bw \geq 150$ the 'legs' of the V become more linear.

Thus, the curious trend points to publication selection bias. We know of no more likely explanation.¹¹

Therefore, the curious trend indicates that biases due to priors and industry interests should be a serious concern in the AEL. We suggest that the bias is due to a reluctance to publish negative values. This suggestion is confirmed by the formal test in section 2.6.

2.3 *The methods used: The funnel, the FAT-PET MRA*

MRA is **Meta Regression Analysis**, *i.e.*, a regression run on regression coefficients (or a transformation thereof). FAT is **Funnel Asymmetry Test**; PET is **Precision Estimate Test** – also known as the PET meta-average.

The coded estimates of aid effectiveness is denoted by b_i , where $i = 1, \dots, 1,777$. Each b_i comes with a standard error, s_i , allowing us to calculate the precision, $p_i = 1/s_i$, and the t-ratio, $t_i = b_i/s_i$. This is the data needed for the basic level one of the meta-study.¹²

The funnel is the (p_i, b_i) -scatter, where the b s are at the horizontal axis. It is a graphical display of the distribution of the b s. If the range on the p -axis is divided into segments, the observations in the lower segment, *i.e.*, at lower precisions, should scatter the most. However, the average b_i in each segment should be the same. This gives the characteristic funnel-shape observed on Figure 2.

Publication selection bias means that too few estimates with the wrong (negative) sign are published. Most such estimates occur at the bottom of the funnel. As p grows, fewer estimates have the wrong sign. Censoring is rarely complete, but it will cause funnel asymmetry¹³ so that the average at the lower p -segments is larger. This is precisely what the FAT is testing.

The FAT-PET is a smooth curve through the average at each p -level that converges to the true average (the PET) as precision rises:

$$(1) \quad b_i = \beta_M + \beta_F s_i + u_i, \quad \text{where } \beta_M \text{ is the PET and } \beta_F \text{ is the FAT. A division by } s_i \text{ yields}$$

11. The only alternative explanation we have met is that aid agencies start by picking the low hanging fruits. The project literature seems not to confirm this hypothesis. Also, the literature deals with a period of 60 years, where the projects that are ‘hanging low’ have changed rather much.

12. Level one of the meta-analysis is the four steps: literature collection; coding; study of funnel and calculation of the FAT-PET. It is well defined and robust. Level two studies the width of the funnel and asks questions to the literature. Depending upon the findings in level one, one may also use the augmented meta-average, but when a substantial censoring bias is found the average augmented meta-average maybe biased; see Paldam (2013b).

13. Note that in other fields the asymmetry is different. For example, if all regression coefficients were positive but priors suggest an elastic response, then the censoring will occur at say an elasticity of 1, with relatively few inelastic responses reported.

(1b) $t_i = \beta_M p_i + \beta_F + v_i$, that is used in the estimates.

The residuals are u_i and v_i ($\approx u_i/s_i$ ¹⁴). See Stanley and Doucouliagos (2012).

With asymmetry ($\beta_F \neq 0$) this MRA becomes a hyperbola when depicted on the funnel graph as s_i is exchanged with $1/p_i$ in equation (1). The hyperbola converges to β_M when precision rises. This is shown on the two funnels of Figure 2. The mathematical form is somewhat arbitrary, and studies have been made, giving the curve more or less curvature, notably by adding a squared term.¹⁵ As the FAT-PET MRA has to go through the average at each p -segment and converge to the top for large ps , the choice of a reasonable curvature is limited and the result is robust to such changes.

On Figure 2 the mean and the PET meta-averages are drawn as vertical lines, while the FAT-PET-curve drawn is the estimate from Table 2 that converges to the PET-line.

Both the model and the estimates are more similar within than between papers, so it is preferable to cluster the estimates by paper. This is done by coding a paper number, and the regressions are then run for cluster robust standard errors, with the paper number as the cluster. This increases the standard error of the estimates substantially as seen from Table 2.¹⁶

2.4 Common questions

Meta-analysis is fairly new in economics and the method has given rise to many questions and objections. Once the papers in a literature are coded it is often possible to operationalize these questions and provide answers for (at least) that literature.

One of the most commonly posed questions is how much the quality of the papers matter for the results reached. The standard way to analyze this question is to code the impact factor of the journal (or other outlet) in which the paper has appeared. If we assume that authors try to publish their papers as well as possible, and that journals try to choose the best papers, this is a measure of quality. We tried this variable in the first round of studies, and as usual in meta-analysis it changed results marginally only; see Table 6 in D&P (2008) for a comprehensive analysis.

Another point we have explored is the effect of estimators. Many economists consider the choice of the right estimator to be crucial; however it is rare that meta-studies show that a

14. The residuals v_i are likely to have less heteroskedasticity than the u_i residuals, which is why (1b) is the preferred formulation for estimation purposes.

15. We added a squared s -term to the FAT-PET MRA, as recommended in Stanley and Doucouliagos (2012 section 6.3), but the squared term did not produce statistically significant coefficients.

16. Given the large number of studies and estimates, it is unnecessary to use the wild bootstrap to correct standard errors.

substantial part of the (large) variation between the results is due to the choice of estimator. In the AEL we found no signs that the choice of estimator matters for the results; see section 2.6 in D&P (2008).

This is particularly important as regards the attempts made to correct the estimate for simultaneity, see section 3 of D&P (2011). If aid is termed h and growth is g then the AEL looks for an estimator of $\beta = \partial g / \partial h$ that corrects for the aid allocation effect $\alpha = \partial h / \partial g$ of growth on aid, as explained in Table 1. Authors, who want to find a positive value of β , tend to believe in case II, where the reverse causal effect $\alpha = \partial h / \partial g$ is negative, so that standard estimates of β are downward biased. This belief leads to the *claim* that the right (positive) result will appear if the right method is used to correct for simultaneity. The standard method is TSIV (two stage instrument variable estimation). So the claim is that if the right instruments could be found they would cause the estimate of a nice positive $b \approx \beta$.

Table 1. Simultaneity of aid effectiveness and aid allocation

AEL: the effect of aid on growth		Bias from aid allocation effect	
Effect	Estimated by	If $\alpha > 0$	If $\alpha < 0$
(2) $\beta = \partial g / \partial h$	$g = bh + \text{controls}, b \approx \beta$	Case I: $b > \beta$	Case II: $b < \beta$
AAL: the effect of growth on aid		Bias from aid effectiveness effect	
Effect	Estimated by	If $\beta > 0$	If $\beta < 0$
(3) $\alpha = \partial h / \partial g$	$h = ag + \text{controls}, b \approx \beta$	$a > \alpha$	$a < \alpha$

D&P (2011) looks at 219 attempts to control for simultaneity – to the extent that they move the estimate it is downwards. This tallies with the findings in D&P (2013a) that analyzes aid allocation as a function of aid. It finds that it is neither theoretically clear nor empirically confirmed that $\alpha < 0$. As of now, the best estimate of α is that it is small and positive. Thus, we are rather in case I.

2.5 The funnels for the two sub-periods: 2007-2008 and 2009-2011

Figures 2a and b show the funnel for the two sub-periods (A) and (B). The visual impressions in this section are formally tested in the following section.

It is clearly visible that the points are shifted to the right on Figure 2b compared to Figure 2a. This illustrates the difference between the two sub-periods. However, note that the top of the funnels differs much less than the base.

Figure 2a. The funnel plot of the 600 estimates from period (A) 2007-08

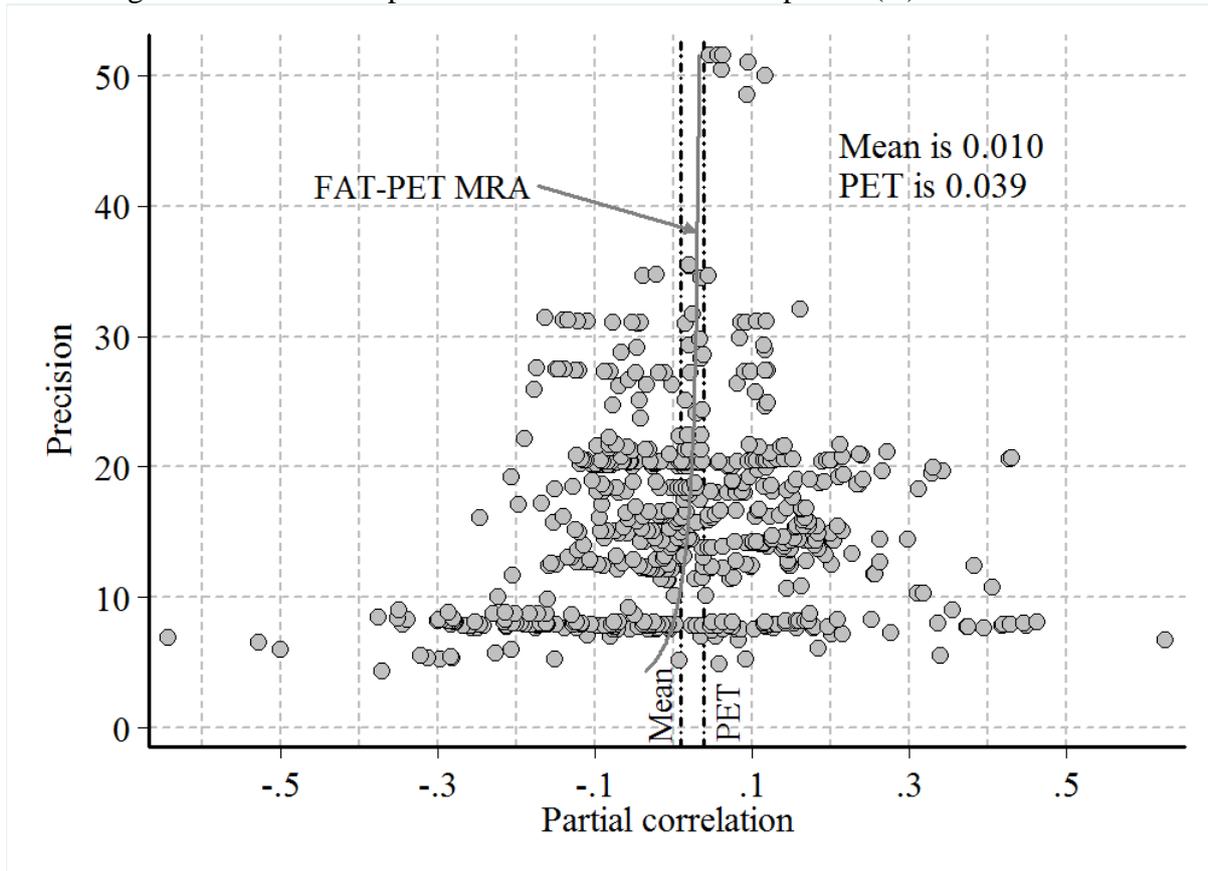
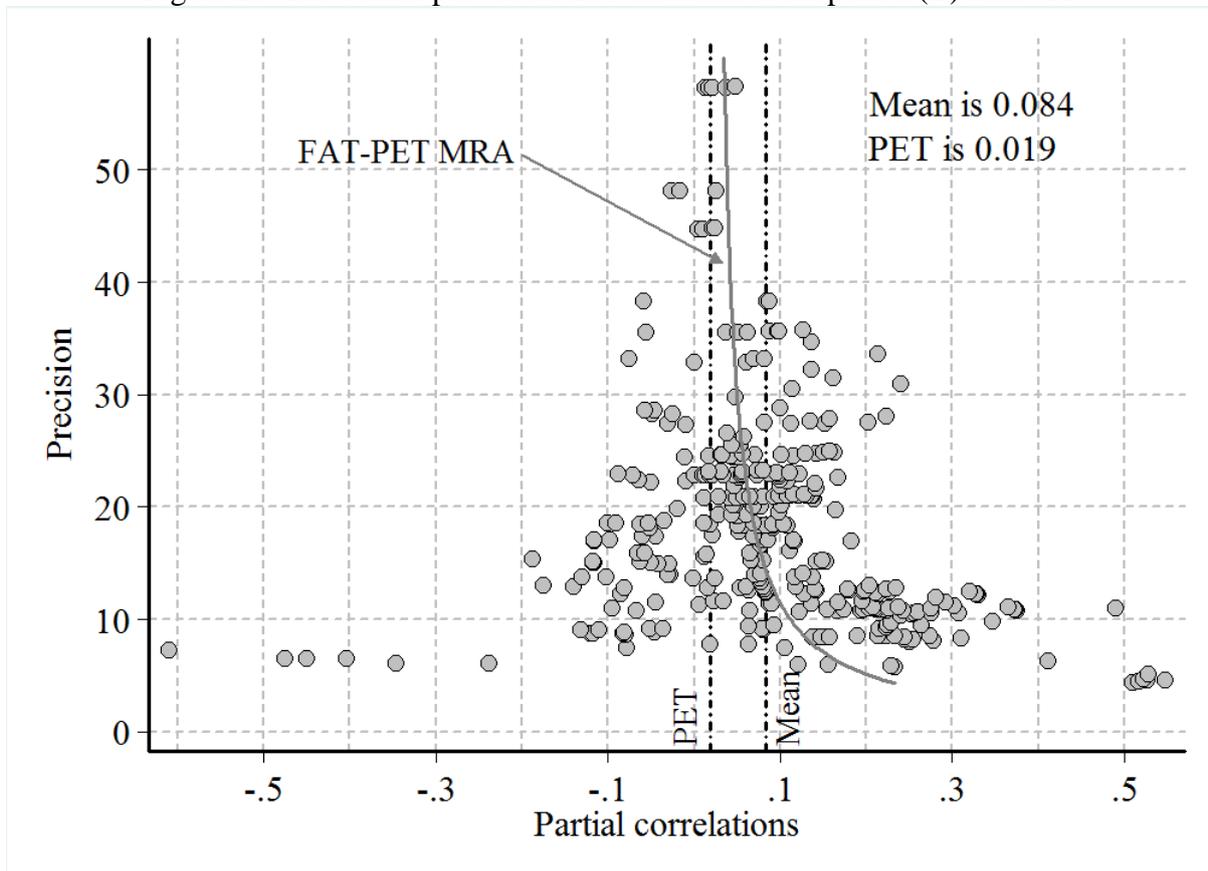


Figure 2b. The funnel plot of the 266 estimates from period (B) 2009-11



Both sub-figures 2 look alike at a first glance, but upon closer inspection Figure 2b is less symmetrical than Figure 2a. On Figure 2a the FAT-PET MRA is almost vertical, so the mean and the PET meta-average are close, but on Figure 2b a more substantial curvature occurs, as the mean is further from the top of the funnel in Figure 2b. Consequently, the PET meta-average is some distance away from the mean on Figure 2b. Even though the mean is quite different on Figures 2a and 2b, the PET is similar. This suggests that the PET produces results that are more robust than the mean.

2.6 The basic meta-analysis of the periods suggested by Figure 1

Table 2 has three panels from left to right and three sections from top to bottom.

The left hand panel gives the periods and the number of observations and papers.

The middle panel (shaded in gray): Column (5) reports the (arithmetic) mean followed by two t-ratios: Column (6) is (t1) the average t-ratio for the estimates, while column (7) is (t2) the t-ratio across the estimates.¹⁷ Column (8) reports the average precision, which is a function of the number of estimates.

Table 2. Analyzing all estimates and the various periods

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Observations				Arithmetic mean			FAT-PET MRA						
Period	N	Papers	Mean	(t1)	(t2)	<i>p</i>	PET	(t)	(trc)	FAT	(t)	(trc)	
Top section: All 1777 estimates													
All	1970-2011	1777	141	0.066	(0.9)	(14.6)	15.3	0.031	(5.8)	(2.1)	0.379	(4.0)	(1.4)
Middle section: All estimates divided into old and new													
(1)	1970-2006	811	84	0.100	(4.5)	(12.8)	13.5	0.035	(4.5)	(3.5)	0.589	(4.6)	(3.2)
(2)	2007-2011	866	56	0.038	(0.7)	(7.7)	16.8	0.036	(4.9)	(1.3)	0.070	(0.5)	(0.1)
Bottom section: The new estimates starting in 2007 divided in two sub-periods													
(A)	2007-2008	600	30	0.010	(0.3)	(1.6)	15.9	0.039	(3.6)	(1.0)	-0.316	(-1.6)	(-0.5)
(B)	2009-2011	266	26	0.084	(1.3)	(11.1)	18.2	0.019	(2.1)	(0.9)	0.931	(4.8)	(1.9)

Note: (t1) is the average t-ratio of the estimates, (t2) is the cross estimate t-ratio. In the meta-analysis (trc) is the robust, cluster corrected t-ratio, where the clusters are the papers. Estimates in bold are statistically significant ($t > 1.9$) by both t-tests. Estimates in both bold and italics are statistically significant by one of the t-tests.

17. Calculated from the mean and the standard deviation of the N estimates. As the standard deviation has to be divided by the square root of N , to yield the standard error, the t2-ratio increases rapidly with N .

The right hand panel gives a basic meta-analysis. The PET is the meta-average and the FAT is the funnel asymmetry test. Both of these estimates come with two t-ratios: (t) is the usual t ratio, and (trc) is the robust clustered t-ratio that is always lower. On average it is almost 3 times smaller.

The top section shows the results for all 1,777 estimates. Here the results are as in D&P (2011). For all observations the mean is 0.066, and the FAT is borderline significant. The PET meta-average is 0.031, so that the publication selection bias is half of the mean. The coefficient of 0.031 is statistically significant even if controlled for clustering. The result is the same as in our previous meta-studies, but with 1,777 observations even small effects become statistically significant. However, a partial correlation of 0.031 is of negligible practical importance: it is not economically significant.

The middle section considers the division in the old and the new period: *The old period (1)*: These are almost the same results as D&P (2008), but by adding the 16 papers from 2005-06 the PET becomes statistically significant, and so does the FAT. The mean estimate is 0.100, but 2/3 of that is a publication selection bias, so the PET meta-average is 0.035, which is about the same as for all observations. *The new period (2)*: Here the mean is 0.038, but the FAT is statistically insignificant, so the mean is almost the same as the PET.

The point to note is that the PET meta-average is practically the same for period (2) as for period (1). *We assess this to be an important finding.*

The bottom section divides the new estimates of period (2) in the two sub-periods (A) and (B) that are our main interests at present. They look different, but the FAT-PET MRA says the difference is in the publication selection bias, not in the genuine effect.

Sub-period (A) has a mean of 0.010 as expected from Figure 1. This is a small correlation indeed. Here the FAT is negative, but statistically insignificant. As the reader can see on Figure 2a the FAT-PET-curve is almost vertical, so the PET barely moves and fails to become statistically significant. However, the PET is actually close to the ‘usual’ value from the three previous analyses.

Sub-period (B) has a mean of 0.084 – this is 8.4 times more than in period (A) and almost as much as in period (1). So the upward kink seems significant. The FAT, however, becomes larger than in any of the previous analyses. So the PET falls and becomes statistically insignificant. Consequently, our meta-analysis shows that all of the improvement in aid-effectiveness after 2008 is actually due to a large increase in publication selection bias.

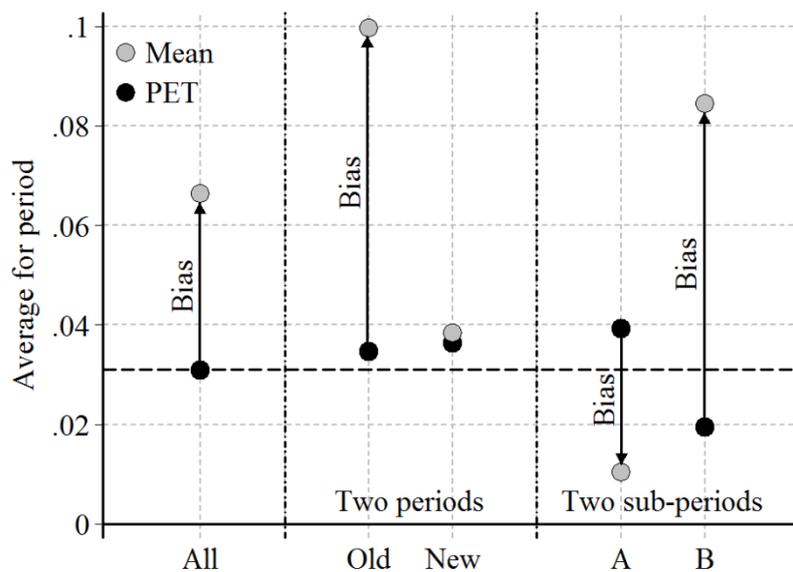
It should be noted that many of the results in periods (2) use long time series so that precision rises from period (1) to (A) to (B), as seen in column (8) of Table 2.

2.7 Comparing means and PETs for the five periods

For an easy overview of the results the reader should consult Figure 3. It shows the five estimates of the mean and the PET. The PET estimates for all periods are so similar that it cannot be rejected that they are the same. The best estimate is thus the one for *All*, indicated with the dashed horizontal line. Thus, we conclude that the true value of aid effectiveness is 0.03 ± 0.01 . This is the same result as in D&P (2008, 2011 and 2013b).

The mean moves much more than the PET. We interpret the difference as a publication selection bias. For *All* the bias is near to the PET in size, so that the mean is twice the PET, but the bias is quite volatile for the sub-periods. It is as large as it ever was in the latest sub-period.

Figure 3. Comparing the estimates of the mean and the PET from Table 2



Consequently, the meta-evidence supports hypothesis (iii) that the increase in aid effectiveness from period (A) to period (B) is a statistical artifact.

Section 2.2 claimed that the economic theory of learning suggested that effectiveness must increase over time. It also showed that the findings in AEL from 1970 to 2008 contradicted that claim. The PET does not show a significant increase, it rather suggests that the degree of aid effectiveness is constant over time.

A note of caution should be added. By now only a few studies have simulated publication selection bias to study the behavior of the mean and PET under controlled circumstances; see Stanley (2008), Paldam (2013b and 2013c), Stanley and Doucouliagos

(2014). It appears that the mean is always exaggerated in the direction of the prior. When theory predicts that a certain parameter is positive, the bias becomes positive too (see also Doucouliagos and Stanley, 2013). The PET is (much) closer to the true value than the mean, under a broad range of circumstances, but it is only a perfect estimate of the mean under rather special circumstances.

2.8 *How large is a partial correlation of 0.03?*

A partial correlation of 0.03 is tiny, even when so many studies have accumulated that the estimate becomes statistically significant. Under Cohen's (1988) widely used guidelines, a simple correlation that is less than 0.10 is deemed to be small. Doucouliagos (2011) constructs similar guidelines for partial correlations based on the observed distribution of over 22,000 partial correlations reported in empirical economics. This results in a threshold of 0.07 for all partial correlations and 0.10 for the empirical growth literature. Regardless of the benchmark, the aid-growth correlation is too low to be of economic significance.

The following sections consider alternative evidence, asking if there is evidence to expect a change in aid effectiveness from period (A) to period (B). Then section 7 concludes by considering all results.

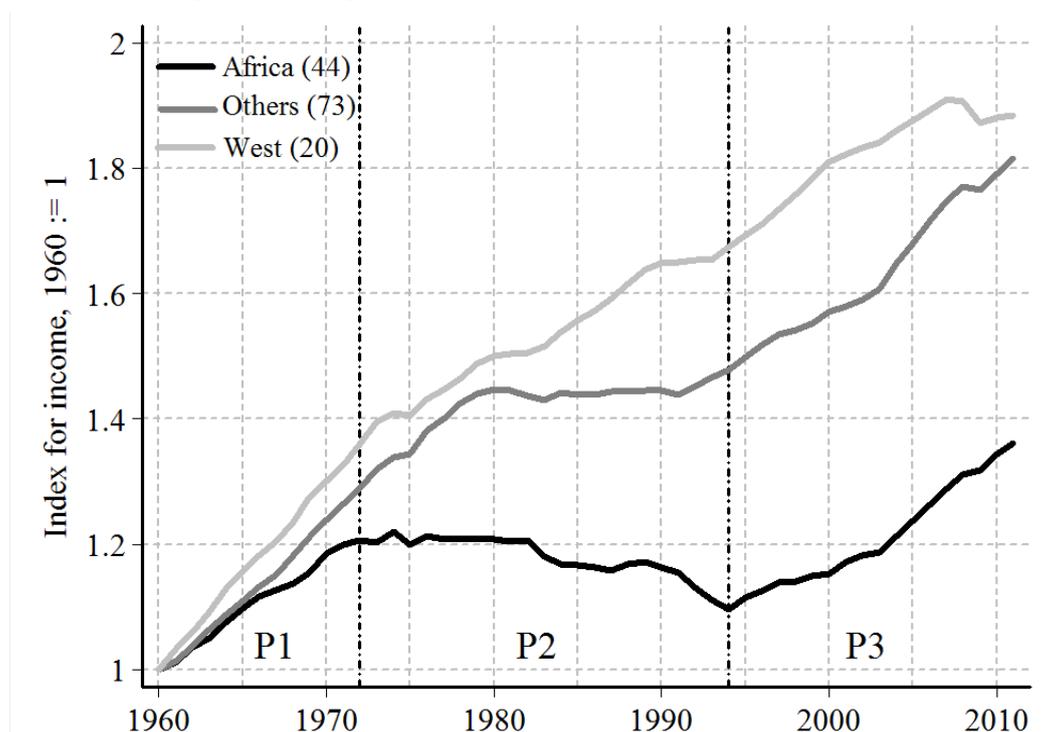
3. The trends in growth and aid

This section presents a few stylized facts. Section 3.1 compares economic development in the world and the path of development aid. Section 3.2 shows the zero-correlation result.

3.1 Economic development in the world and the path of development aid

The path of an index for the logarithm to real GDP per capita in three major country groups is shown on Figure 4. The index is set to 1 in year 1960 for all countries covered. The group ‘Others’ mainly covers middle income countries that receive little aid, but also some low income countries that have much the same path as Sub-Saharan Africa, which is the group with most of the largest aid recipients.

Figure 4. The path of income 1960-2011 for 137 countries

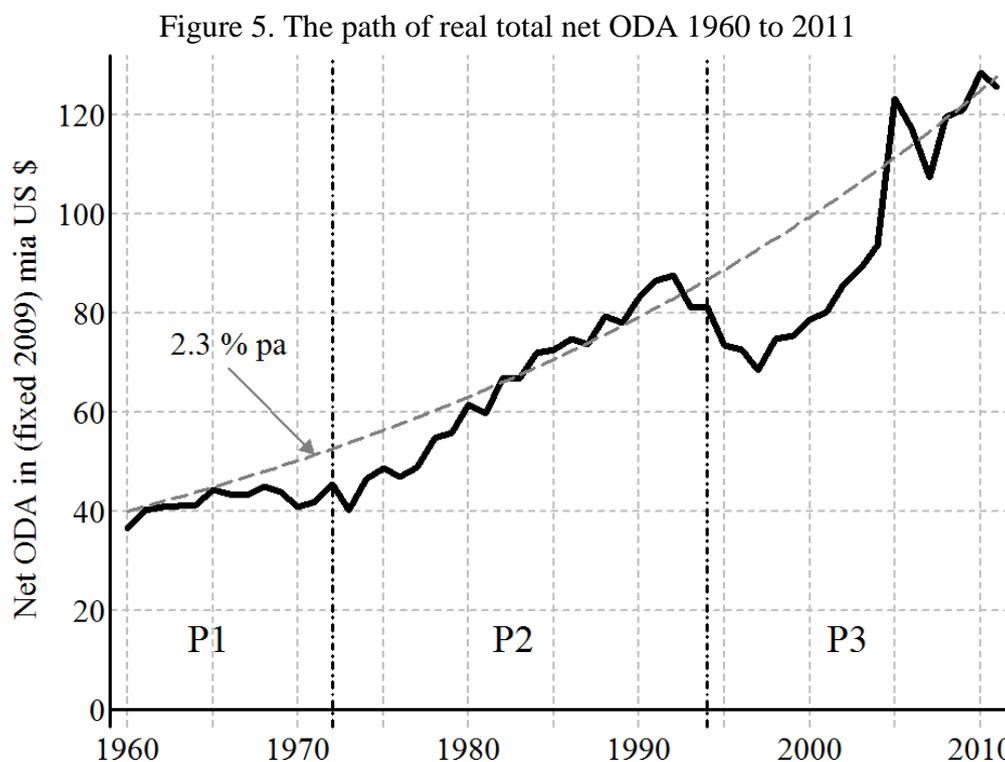


Note: Income is the logarithm to GDP per capita in fixed PPP prices. The periods P1, P2 and P3 apply to the poorest group. The data are for all 137 countries with full data in Maddison (2003), updated by PEN World Tables. Africa covers Sub-Saharan Africa only. Source: Paldam (2011).

The development path for Africa has a clear cyclical path with three periods: (P1) Reasonable growth before the oil crisis (or just before); (P2) Negative growth from 1972 to 1994; and (P3) reasonable growth after 1994. This cyclical path is confronted with a total of 13

explanations found in the literature in Paldam (2011). The analysis deals with Africa only, so it will not be repeated here.

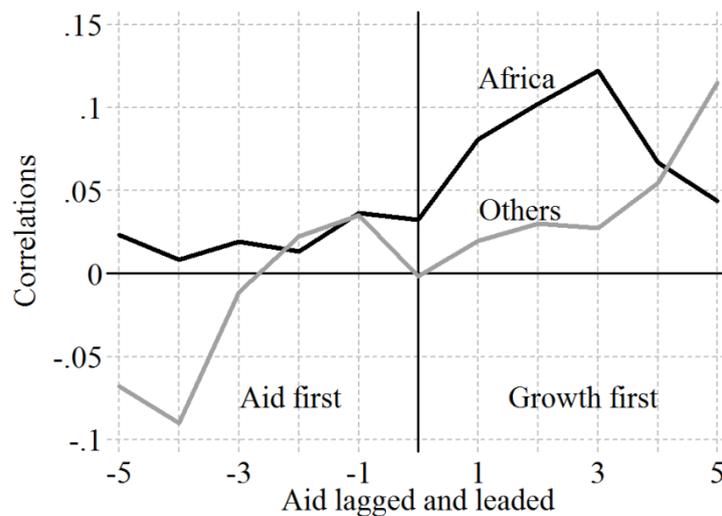
The said analysis of Africa does not consider if aid can explain the pattern, and this was no omission. The net ODA series from the OECD in fixed 2009 prices is shown on Figure 5. When the figure is compared with Figure 4 it is obvious that there is a problem: Aid was fairly stable in period (P1), in (P2) aid rose and growth turned down, and, finally, in (P3) aid fell and growth resumed. Especially for Africa this connection looks bad as shown by Easterly (2006, p 39-45) and Paldam (2007). Hence, it is difficult to argue that aid has been an important factor in the new growth in the poorest countries. And, of course, aid cannot explain the high growth of China and India as the two countries have received negligible aid.



Source: Data from the OECD downloaded from Center of Global Development. Most studies use 5-year averages, so the data provide 10 such periods. The gray dotted line represents constant growth. Data are in fixed 2009 billion US\$.

The gray dashed line of Figure 5 corresponds to a constant rate of increase in the quantity of aid of 2.3 % per year. This equals the population growth rate in Sub Saharan Africa, so it is not an impressive growth rate. The path of aid follows the constant increase path, except from 1992 to 2004 where aid was about 20 % below the trend.

Figure 6. Correlograms of the first differences to the data of Figures 4 and 5



Note: The data for the two country groups are the first difference to income, which is the real growth rate of the GDP per capita and the real growth rate of Net ODA. The correlations are smoothed by an MA(3) process to make them easier to read.

The aid flows include loans on concessional terms. These loans, and other loans, accumulated to an average debt of about 100 % of GDP in the mid-1990s, exactly when growth resumed. Almost 80 % of that debt has now been written off. However, the main debt reduction happened during a couple of years around 2005.

Figure 6 shows how the curves on Figures 4 and 5 are related. It is done by a correlogram using 5 lags to either side between the first differences to the series. There is no sign that aid can explain growth, while growth seems to contribute to explain aid. It should be noted that the correlation is positive, so that more growth attracts some extra aid.

3.2 The zero-correlation result

The World Data from the World Bank contains data for the aid-share defined as net ODA in % of GNI and the real growth rate (of GDP) per capita for the same countries. The data covers a total of 1059 5-year averages (with at least 4 reported values) as downloaded 26/1-2014.¹⁸ The data covers 153 countries that have received aid for at least one of the 10 periods. The average recipient country has received aid in 6.9 periods or about 35 years.

18. World Data change constantly even far back. Since we downloaded these data three years ago for D&P (2011) the correlations have got marginally smaller.

Figure 7a. The distribution of the 5-year averages for aid and growth, 1961-2000

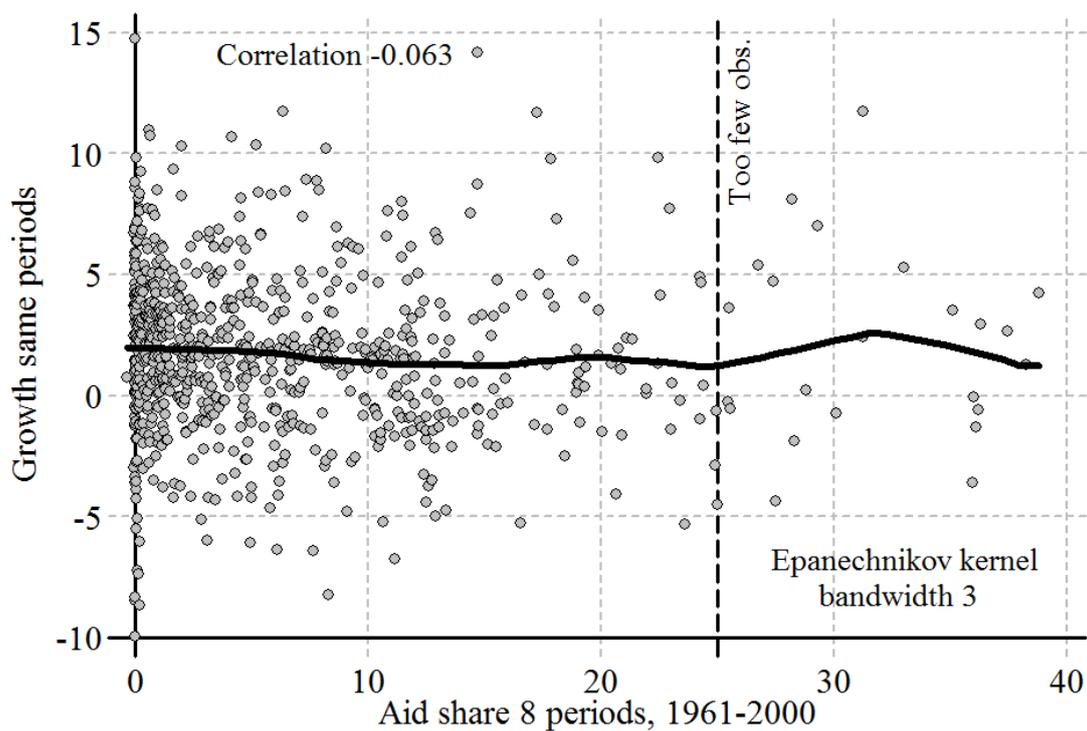
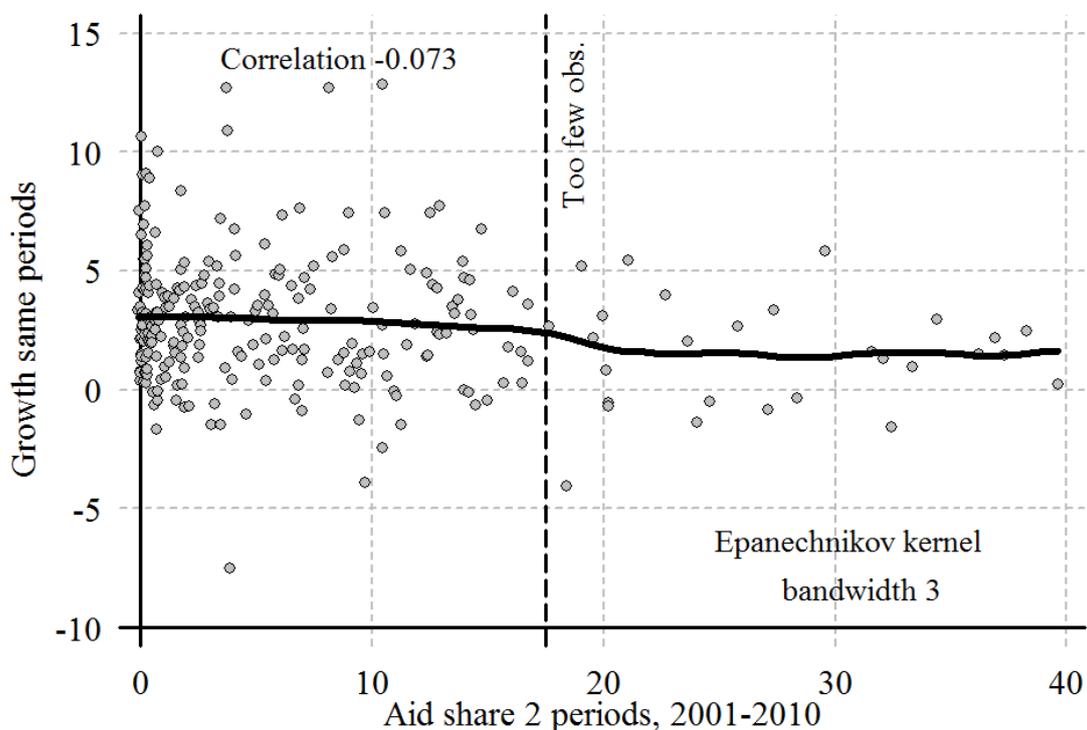


Figure 7b. The distribution of the 5-year averages for aid and growth, 2001-2010



Note: The dashed black lines indicate when the kernel is supported by so few observations that its form is unreliable. When the bandwidth is reduced the curve becomes wobbly at the high end. When the bandwidth is increased the curve becomes (even) more linear – always with a falling trend. This figure is an updated version of Figure 1 from D&P (2011). A few points outside the frame are deleted.

Table 3. Correlations between the aid share and the growth rate

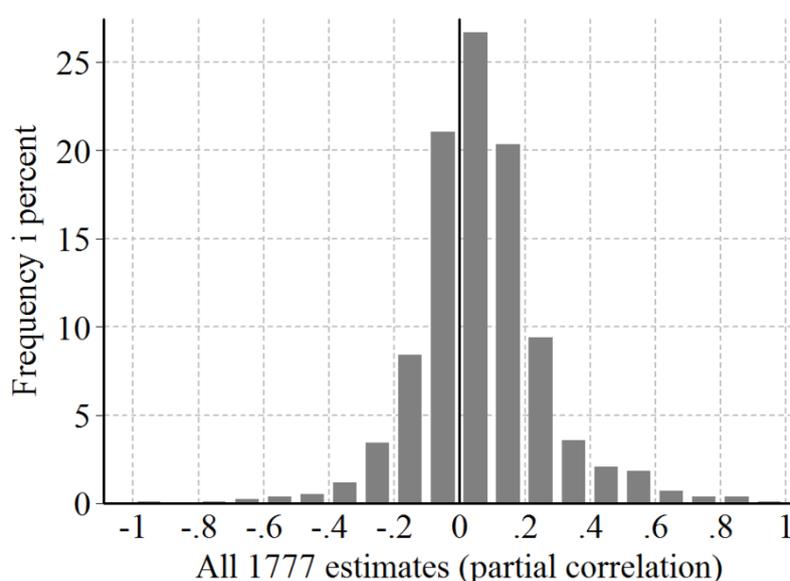
Years	N	Correlation	Years	N	Correlation
1961-65	51	-0.018	1991-95	133	-0.038
1966-70	76	-0.048	1996-00	134	0.036
1971-75	85	-0.008	2001-05	134	-0.152
1976-80	94	-0.132	2006-10	134	0.008
1981-85	105	0.090	Averages	1059	-0.038
1986-90	113	-0.123	Correlation for all merged		-0.057

Note: Years are averages over the five years.

These observations are depicted on Figure 7, where 7a shows all observations from 1961-2000 and 7b shows the observations from 2001-10. A few observations are outside the frames of the graphs. Aid effectiveness translates into a positive slope on the kernel-curve, but the slope is marginally negative. The form of the curve suggests if we should look for a linear or non-linear relation. Both lines deviate very little from a straight line.

The impression from Figure 7 is confirmed by Table 3. The correlations between aid and growth are marginally smaller than zero. For the period 2001-05 it is significantly negative, but if it is treated as 1 out of 10 episodes, it is not significant.

Figure 8. The distribution of the aid effectiveness estimates



Note: The frequency distribution is skewed to the right and normality is rejected. The Skewness/Kurtosis test for normality finds that skewness and kurtosis are consistent with normality with probabilities below 0.01 %, so the joint probability of normality is similarly rejected.

Figure 8 shows the frequency distribution of the 1777 estimates of aid effectiveness reached from these data. It is wonderful to see how wide a range of results our tools allow us to reach from the data just presented. We surely have powerful tools! But wonderful as it is, it also makes you wonder if we are not engaged in a game that is too easy to play. It seems to suggest that a well educated economist can take two uncorrelated variables, select a subsample and a model, and then reach any results desired and publish it in a standard journal!

4. The first hypothesis: The kink is true

The first hypothesis is that while aid did not work before, it now does. This will surely require a change in aid policy. Both national and multinational aid agencies often announce new policies, normally in the form of verbose strategy papers full of broad philosophical principles, many expressions of good will and carefully worded political compromises.

Critics claim that such changes of policies are too numerous and vague to have an effect out there in ‘the bush’ where they are implemented.¹⁹ Proponents claim that they gradually change attitudes and hence practice, but even under the best of circumstances it is a slow process. Most aid projects have an implementation span of 3-5 years or even longer as they are often continued as follow-up projects, especially in the poorest countries.

4.1 *Does the Paris Declaration herald new policies increasing aid effectiveness?*

In the last decade there has been a lot of discussion about aid effectiveness. We cannot review the strategy declarations of all aid agencies, but many aid agencies consider the Paris declaration on aid effectiveness in 2005 by the OECD to be important. Table 4 is OECD’s own summary.

Most donors will probably agree that in some form they did already followed most of these policies well before 2005, so that it was mostly a clarification and codification of policies already followed. There have, at most, been some minor adjustments within the eight year time span since then.

Table 4. The five points of the Paris declaration, summarized by the OECD

1 Ownership	<i>Developing countries set their own strategies for poverty reduction, improve their institutions and tackle corruption</i>
2 Alignment	<i>Donor countries align behind these objectives and use local systems</i>
3 Harmonization	<i>Donor countries coordinate, simplify procedures and share information to avoid duplication</i>
4 Results	<i>Developing countries and donors shift focus to development results and results get measured</i>
5 Mutual accountability	<i>Donors and partners are accountable for development results</i>

Source: <http://www.oecd.org/dac/effectiveness/parisdeclarationandaccraagendaforaction.htm>.

19. The reader should consult Easterly (2007, Chapter 5, notably pp 161-165) describing the production of declarations and reports by multilateral aid bureaucracies, also reporting the amazing number of pages. His description sounds as if it was written by Jonathan Swift, though Easterly claims that his inspiration came from Yosemite Sam.

4.2 *Back to the path on Figures 4, 5 and 7*

Figure 5 is provided with a thin dashed line representing a simple (linear) trend connecting the path in the aid quantity from 1970 to 1992 to the path from 2005 to 2011. It has a 2.3 % real annual rise in the volume of aid, corresponding to the population growth rate in the 50 poorest countries. There is a dip of about 20 % below that line in the 11 years from 1994 to 2004. The quantity of aid was 20 % below the trend during this period.

Aid effectiveness is a quality measure, so perhaps the fall in the quantity has led to an increase in the quality. It appears that donors often have problems using their budgets in poor countries, so that the marginal project may be relatively inefficient. However, most of this effect was already included in studies before 2009. The new studies from 2009 onwards add data from after 2005 when aid was back on the old trend.

Figure 4 shows that development turned up in the period from 1990-95. In the poorest countries this occurred in 1993-94. Figure 5 shows that this period was precisely the period where aid fell. So when the period from 1990-95 was entered in the data, it might have given a downward effect on the aid effectiveness estimated, but we have not been able to confirm that possibility in the data. Furthermore, Figure 7b and the correlations for the periods 2001-10 in Table 3 do not support the notion that aid has become more effective recently.

To sum up, we have found no reason to believe that the upward kink in reported aid effectiveness is a true kink.

5. The second hypothesis: The kink is a breakthrough in our models

The main reason for the avalanche of new papers is that the authors have looked for a breakthrough. The second hypothesis says that this has occurred at long last. The models were wrong in the past, resulting in biased (too small) estimates, but now better models have been found. According to this view, the new tools would have shown all along that aid works.

5.1 *The nature of technical progress*

Consider the AEL as a production process made by a group of about 140 researchers. They compete and collaborate in a complex way, producing the output taken to be the average line drawn as the kernel on Figure 1.

In general, technical progress in products produced by groups of researchers typically has a rather smooth path as it is due to many small improvements. For example, new cars keep driving longer per liter of gasoline. This is not due to one major breakthrough, but to numerous small improvements.

Even if a breakthrough occurs, it has to be implemented, and this too is often a gradual process. If there is a breakthrough, future models will change after some little time. This will appear as a soft kink of the type observed on Figure 1.

5.2 *Can a breakthrough be identified?*

A breakthrough is something that happens in one paper and then spreads, so to give the idea of a breakthrough a fair chance we compare all papers with $r > 0.05$ in period (A) and $r > 0.10$ in period (B).²⁰

To study if there is a breakthrough, it should be demonstrated that it actually exists in these models. The two parts of Table 5 confirm the analysis in Section 2, as the results in Table 5B are more numerous.

Authors typically announce the new features in their paper so that the referees and editor and the readers notice. If the paper is a breakthrough, other papers will refer to it. Hence, Table 5B lists the new feature claimed by the authors, or assessed by us, of each of the 14 possible breakthrough papers from period (B).

20. The precise reference is found on the URL: <http://www.martin.paldam.dk/Meta-AEL.php>. We have closed the data collection in the late 2011.

Table 5A. Key features of papers with the highest average from period A ($r > 0.05$)

Paper	Authors	r	Coverage	Estimators	Innovation
130	Elbadawi, Kaltani, Schmidt-Hebbel	0.37	1970-04	GMM	Post-conflict
96	Minou, Reddy	0.20	1990-00	GMM	Type of aid
94	Loxley, Sackey	0.17	1973-04	OLS	Africa
81	Feeny	0.16	1980-01	OLS, FE, RE	Melanesia
73	Bhandari, Pradhan, Dhakel, Upadyaya	0.08	1993-02	ARDL	Transition economies
85	Hudson, Mosley	0.08	1975-03	OLS/2SLS	Aid volatility

Note: r is the average partial correlation per paper. The selection of papers is described in text.

Table 5B. Key features of papers with the highest average from period B ($r > 0.10$)

Paper	Authors	r	Coverage	Estimators	Innovation
107	Abidemi, Abidemi, Olawale	0.53	1981-08	OLS	Nigeria
140	Sakyi	0.52	1984-07	ARDL	Ghana
138	Neanidis, Varvarigos	0.24	1973-07	OLS, GMM	Aid volatility & types of aid
114	Asteriou	0.21	1975-02	ARDL	South Asia
112	Arndt, Jones, Tarp	0.17	1970-00	IV-LIML	New instruments in IV-estimates
136	Muhammad, Quyyum	0.15	1960-08	ARDL	Pakistan
141	Selaya, Thiele	0.14	1962-01	GMM	
83	Fielding, Knowles	0.14	1970-93	OLS	per capita rather than % GDP
120	Bjerg, Bjørnskov, Holm	0.13	1960-00	GLS	Debt burden
125	Chervin, van Wijnbergen	0.12	1966-01	OLS/2SLS	Aid volatility
137	Ndambendia, Njoupouognigni	0.11	1980-07	ARDL	Africa
111	Annen, Kosempel	0.11	1970-04	OLS/GMM	Type of aid
117	Balimoune-Lutz	0.11	1970-97	2SLS	Africa. Considers social cohesion
115	Bah, Ward	0.10	1980-08	OLS/GMM	Small Island states

Note: see Table 5A.

We have put the papers in a strict time-order. It is not the order today, as papers go through a publication history of variable length. Between the first draft and the final publication is at least one year, but it may be as long as five years. Here the paper often changes. Papers are normally presented at conferences with other writers in the field, and they are subjected to refereeing, where referees mention citations the author has omitted, etc. Also, many stories are told about papers that did contain a new feature that was disliked by the old writers in the literature, who were the referees, and consequently it was delayed in publication. Thus, it is necessary to fix the order at a certain date.

Table 5 shows no noticeable breakthrough that can explain the kink. There is a tendency to use more focused samples. The use of more nuanced samples is consistent with the results of our meta-analysis: aid does not appear to generate growth for developing countries on average, but it may have beneficial effects for specific countries and time periods.

Also, there is growing use of ARDL (autoregressive distributed lag models). It is trivially true that most aid projects have activity effects. However, aid effectiveness is defined to mean that aid generates growth.²¹ Thus, activity effects should not be included. ADRL models are normally estimated on annual data that catches activity effects, so these models are likely to give an upward bias.

Neither a more selective sample nor ARDL models are a breakthrough in aid effectiveness modeling. Thus, we have found no sign that a new model has been found that heralds a breakthrough.²²

We caution to add that ingenuity and progress are continual processes. Perhaps the work of Clemens *et al.* (2012) or Brückner (2013) will prove to be a breakthrough when sufficient independent replication has been made, or perhaps some as yet unreported approach will provide the breakthrough. We cannot exclude this possibility. However, we can rule out that the upward link observed in Figure 1 is the result of an existing breakthrough.

21. While a railroad is being built it creates employment. However, the railroad may not generate growth once it is completed, if it does not come to transport a sufficient numbers of passengers and goods.

22 Some of the studies from period B also tend to use slightly more recent data.

6. The third hypothesis: The kink is an artifact

Section 2.6 showed that the standard test for publication selection bias suggested that bias has increased since 2008 – and this increase fully explained the kink in the results. The subject of publication selection bias is still often taken to be embarrassing, but it is important for our science (and all others) that it is recognized as a fact of life – this is increasingly done.²³

The introduction considered two priors: The political/moral wish for aid to work and the prior caused by sponsor interests. The way priors are created and change belongs in the field of psychology. Our tools are suited to tell a story about the external facts. The first prior comes from altruism and expressive behavior, and we have found no reason to suppose that it has changed. Sponsor interests are an external reason that could be analyzed if the data were available. It is a strong tradition in medical journals to demand that writers inform about sponsorships and possible conflicts of interest. This is also emerging in economics, but the data is in most cases not available.

The FAT-PET provides evidence of publication selection.²⁴ Doucouliagos and Stanley (2013) set guidelines for the magnitude of this bias. A FAT that is less than 1 is deemed to have “little to modest” selectivity.²⁵ Hence, the AEL does not suffer from serious selection bias. Nevertheless, as demonstrated in section 2.6 above, even this modest amount of bias is sufficient to half a small average partial correlation from 0.066 to 0.031.

The FAT-PET can be extended to investigate the selection process in more detail by estimating: $b_{ij} = \beta_M + \beta_F s_{ij} + \sum \alpha_j s_{ij} K_{ji} + u_i$, where K is a vector of data, estimator and specification choices. See D&P (2012) for an example of such a model that explains publication bias as a complex process involving modeling choices in the AEL.

23. Google Scholar yields about 3 million hits for ‘publication bias’. The word is often used in a loose way, but many are to empirical studies (mostly in medicine). Many of these show evidence of such bias. The term ‘sponsor interests’ gives 320’000 hits which again contain many studies that often find biases.

24. We have had direct experience with this process. An earlier version of D&P (2011) passed the first round of the review process in a top 20 economics journal but not the second round. A dissenting referee insisted that we change our conclusion of aid ineffectiveness. As scientists we want to be guided by data and evidence rather than the preferences of a referee. However, we do have the luxury of being tenured professors.

25. If the FAT coefficient is greater than 1 but less than 2, then selectivity can be considered to be “substantial” and it is “severe” if $FAT > 2$.

7. Conclusions

In previous meta-studies we have demonstrated that the AEL, Aid Effectiveness Literature, shows little effectiveness, but since 2009 the results have been better. The purpose of this paper was to see if these results are the breakthrough we have all hoped for. Unfortunately, it appears that the upswing is either fully due to an increase in the publication selection bias or it reflects the effects of random sampling.

Section 7.1 summarizes the quantitative evidence. Section 7.2 deals with an embarrassing issue of publications bias, and section 7.2 gives some interpretations of the result.

7.1 *Summing up*

The AEL is now 43 years old. Our study covers 141 papers with 1,777 estimates of the effect of aid on growth. When corrected for censoring the average is $+ 0.03 \pm 0.01$. This result has proved remarkably stable over time. Thanks to the large number of estimates the average is statistically significant, but it is economically negligible.

The benefit-cost ratio of aid projects measures the growth contribution of the projects. The cut-off value is typically 10 %. Half of all aid projects are normally assessed as successful, so half of all projects have a higher and the other half a lower ex post benefit-cost ratio. Thus, the average benefit-cost ratio of aid projects is around 10 % or perhaps a little lower. This means that 1 extra percentage point of aid (of GDP) should produce 0.1 to 0.07 percentage points of growth.²⁶ The result from the AEL (see section 2.8) is substantially smaller than that. This is the famous micro-macro problem of aid that has been known since Mosley (1986).

7.2 *Let us face it: Economists are human!*

Our results join others that show that publication selection bias is a serious issue in economics. For recent discussions and references see Ioannidis and Doucouliagos (2013) and Paldam (2013a).²⁷ Surprisingly many think that this is an embarrassing issue that should not

26. The average Sub-Saharan African country has an aid share of about 14 %. Thus aid should generate 1 to 1.5 % growth. The average per capita real growth rate in Africa since 1950 is 1.0 %, so aid should explain it all. As aid is quite volatile across countries, it should be easy to show aid effectiveness.

27. At present about 500 meta-studies have been done in economics. Most find funnel asymmetries which point to censoring; see Doucouliagos and Stanley (2012). The discussion is not confined to economics; see the discussion in medicine started by Ioannidis (2005). The Economist (2013) recently had a special issue on 'how science goes wrong' with a discussion of publication biases and the lack of replicability in many sciences. It

be mentioned in polite society, but the key purpose of research is truth finding – not politeness. If too much is swept under the carpet, so that people stumble over ridges, it is still more embarrassing. Also, the problem will *not* die from benign neglect – in the AEL it has *increased* since 2009. It is much better to recognize the problem and try to adjust for it, with the well tested tools of meta-regression analysis.

We believe that all economists know how the problem comes about: Think of the situation when you have got your data set ready in the computer. Now it is frightfully easy to run regressions, and you cannot help doing a lot. Then you have to choose which results to publish. In a field with strong emotional issues and lots of money at stake, it is hard to imagine that priors can be kept completely at bay, even when many researchers try.

The bias found in the AEL is moderate in size, but on average – over the 1,777 observations – it still accounts for half of the average found. Also, the bias is not constant, but when it is corrected for, the bias corrected average becomes almost constant. We also want to stress that our findings do not suggest that all or even the average study has engaged in selection bias. It only takes a small part of the evidence base to be afflicted by selection to have an effect. Indeed, the FAT suggests that the bias in the AEL is relatively modest. Nevertheless, it has a real impact. Moreover, even without any correction for selection bias, taking all reported estimates as the whole evidence base and at face value, the average effect of aid on growth is trivially small.

7.2 *Interpretation: How should the negligible effects of aid be understood?*

Four points will be made:

The first point is that the majority of the AEL looks for aid effectiveness on average: Has aid contributed to growth in the average country to which it has been applied? It is entirely possible that aid has had an important positive effect on growth in some countries and under certain conditions. Also, it is possible that some types of aid might have contributed even if aid in aggregate has not.

The second point is that the AEL examines only one goal of aid, not all announced ones. The AEL deals with the goal of development measured as economic growth. Aid agencies live in a politized environment in rich ‘post material’ countries, where it is not politically correct to say that the key aim of aid is to generate economic growth. As a result

stressed how serious it is that these issues are swept under the carpet. The reader may also look at Bohannon (2013) for other aspects of the problem.

more complex goals of aid are often declared.²⁸ Aid should reduce poverty, generate good governance and a better civic society, lead to equality between genders, reduce pollution and improve the climate, etc. These goals shift with political agendas and most are so lofty that it is difficult to assess if they are achieved. And on most of these issues only a few empirical papers exist.

Thirdly, one may ask why the effect is so negligible. One explanation may be precisely that with many and partly contradictory goals it is difficult to get very far. Another problem is fungibility: The marginal project caused by the aid is not necessarily the project financed by aid. Often recipient countries give nice projects – which would be done anyhow – to donors, to liberate funds to something else that may be less nice. Also, there are the problems of Dutch disease, executive capacity, etc. discussed in D&P (2009).

Fourthly, it should be stressed that meta-analysis covers the past, not the future. If the reader agrees that it should be possible to get more development for the money spent on development aid, it is obvious that reforms are necessary.

28. The theory of bureaucratic behavior predicts that when the activity of a bureau is unsuccessful at its declared goal then it prefers to change the goal to something less clear rather than change the activity, see Niskanen (1993).

References:

Section 1

- Bohannon, J., 2013. Who is afraid of the Peer Review? *Science* 342(#), 60-65
- Burnside, C., Dollar, D., 2000. Aid, policies and growth. *American Economic Review* 90(#), 847-68
- Clemens, M.A., Radelet, S., Bhavnani, R.R., Bazzi, S., 2012. Counting chickens when they hatch: Timing and the effects of aid on growth. *Economic Journal* 122(#), 590-617
- Doucouliaqos, H., 2013. How Large is Large? Preliminary and Relative Guidelines for Interpreting Partial Correlations in Economics. (Presented at the MAER-Net Colloquium in Greenwich 2013)
- Doucouliaqos, H., Stanley, T.D., 2013. Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity. *Journal of Economic Surveys* 27(#), 316-39
- Easterly, W., 2006. *The white man's burden*. Oxford UP, Oxford
- Economist, 2013. *How science goes wrong*. Issue of October 19th to 25th p 11 and p 21-24
- Hansen, H., Tarp, F., 2000. Aid effectiveness disputed. *Journal of International Development* 12(#), 375-398. Also pp 103-128 in Tarp, F., Hjertholm, P., eds. 2000. *Foreign aid and development. Lessons learnt and directions for the future*. Routledge Studies in Development Economics: London
- Ioannidis, J., Doucouliagos, C., 2013. What's to know about the credibility of empirical economics? *Journal of Economic Surveys* 27(#), 997-1004
- Ioannidis, J.P. A. 2005. Why Most Published Research Findings Are False. *PLOS Medicine*, 696-701.
- Klein, D.B., DiCola, T., 2004. Institutional Ties of Journal of Development Economics Authors and Editors. *Econ Journal Watch* 1(#), 319-30
- Maddison, A., 2003. *The world economy: Historical statistics*. OECD, Paris. Updated versions till February 2010 available from Maddison homepage: <http://www.ggdc.net/maddison/>
- Mekasha, T.J., Tarp, F., 2012. Aid and growth. What meta-analysis reveals. *Journal of Development Studies*, 49(#), 564-83
- Mosley, P., 1986. Aid effectiveness: The micro-macro paradox. *Institute of Development Studies Bulletin* 17(#), 214-25
- Niskanen, W.A. Jr., 1994. *Bureaucracy and Public Economics*. Edward Elgar, Aldershot (U.K.), 6 older essays
- Paldam, M., 2007. Mere hjælp til Afrika. Hvad vil man opnå? *Økonomi og Politik* 80(#), 2-20
- Paldam, M., 2011, 2015. The cycle of development in Africa. A story about the power of economic ideas. *Swiss Journal of Economics and Statistics* 147(#), 427-59. Updated in Christensen, B.J., Kowalczyk, C., eds., *Globalization: Strategies and Effects*. Springer, New York
- Paldam, M., 2013a. Regression Costs Fall, Mining Ratios Rise, Publication Bias Looms, and Techniques Get Fancier: Reflections on Some Trends in Empirical Macroeconomics. *Econ Journal Watch* 10(#), 136-56
- Paldam, M., 2013b. How do partly omitted control variables influence the averages used in meta-analysis in economics? Economics Working Papers, Aarhus University 2013-22
- Paldam, M., 2015. Simulating publication bias. Conference paper
- Penn World tables, URL: <http://www.rug.nl/research/ggdc/data/penn-world-table>

- Stanley, T.D. & Doucouliagos, H., 2014. Meta-Regression Approximations to Reduce Publication Selection Bias. *Research Synthesis Methods*. forthcoming.
- Stanley, T.D. and 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, Abingdon (UK)
- Stanley, T.D., Doucouliagos, H., Giles, M., Heckmeyer, J.H., Johnson, R.J., Laroche, P., Nelson, J.P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R.S., Rost, K., 2013. Reporting Guidelines for Meta-Regression Analysis in Economics. *Journal of Economic Surveys* 27(#), 390-4
- Thompson, P., 2010. Learning by Doing. Chapter 10 pp 429-76 in Hall, B.H., Rosenberg, N., eds., *Handbook of The Economics of Innovation*, Vol. 1. North-Holland, Amsterdam

Section 2. The papers in our project²⁹

- D&P, 2006. Aid effectiveness on accumulation. A meta study. *Kyklos* 59(#), 27-54
- D&P, 2008. Aid effectiveness on growth. A meta study. *European Journal of Political Economy* 24(#), 1-24
- D&P, 2009. The aid effectiveness literature. The sad results of 40 years of research. *Journal of Economic Surveys* 23(#), 433-61
- D&P, 2010. Conditional aid effectiveness. A meta study. *Journal of International Development* 22(#), 391-410
- D&P, 2011. The Ineffectiveness of Development Aid on Growth: An update. *European Journal of Political Economy* 27(#), 399–404
- D&P, 2012. The robust result in meta-analysis of aid effectiveness: A response to Mekasha and Tarp. Deakin University Economics Working Paper 2012/4.
- D&P, 2013a. Explaining development aid allocation by growth: A meta study. *Journal of Entrepreneurship & Public Policy* 2(#), 21-41
- D&P, 2013b. The robust result in meta-analysis of aid effectiveness: A response to Mekasha and Tarp. *Journal of Development Studies* 49(#), 584–87 (short version of D&P 2012)

29. D&P denote the authors. The papers and the latest revised data set are available from the URL: <http://www.martin.paldam.dk/Meta-AEL.php>.