November 28 - 2023

Three sets of evidence about aid effectiveness

The micro-macro paradox of aid revisited

Martin Paldam, Aarhus University[1]

Abstract

Aid effectiveness is the effect of aid on development, i.e., economic growth. The three sets of evidence are: (i) Micro project evaluations find a moderately high efficiency. (ii) Univariate macro estimates find the zero-correlation result. (iii) Multivariate macro estimates of the AEL, aid effectiveness literature, find highly variable results with a small meta-average. While a large literature covers (i) and (iii), (ii) is rarely mentioned. The micro-macro paradox is that micro results are much better than the macro results. A list of possible explanations that may reduce the paradox is provided, but the effect-size of most of these possibilities is hard to assess.

Keywords: Development aid effectiveness, project evaluations, economic growth

Jel.: F35, H43, O11, O12

[1] Department of Economics and Business, Fuglesangs Allé 4, 8210 Arhus V, Denmark.
E-mail mpaldam@econ.au.dk. Home page http://www.martin.paldam.dk.

The paper was discussed at the *Political Economy of Aid* conference in Dublin in October 2023. It has greatly benefitted from the discussion. It is also influenced by a long cooperation with Chris Doucouliagos.

1. Introduction

For almost 40 years – since Mosely (1986) – the literature on aid effectiveness has struggled with the micro-macro paradox. That is, many aid projects are successful, but it is hard to find a connection between the success of poor countries and the aid they receive. This paper shows that the paradox is as strong as ever.

A huge literature deals with the aims and effects of development aid.[2] The authors of this literature are researchers, civil servants, journalists, and politicians – often in mixtures. The size of the literature is easy to explain – poverty in poor countries is a huge problem. *Aid* is a tool against this problem, but it is unclear if it works. It is easy to point to countries that have absorbed much aid with no apparent effect. The lack of clarity calls for research.

Aid has an LDC recipient and an HIC donor, see Table 1. The donor may want to influence much in the recipient country, but the LDC treasures its independence. The two parts used to agree that the goal of development aid is development.

Development is a complex concept. It consists of confluent changes in most socio-economic variables. The best aggregate measure is *income*, y, which is the logarithm to *gdp* (real GDP/GNI per capita), and *growth*, g, of the real *gdp*. Aid is measured by the aid share, α = ODA/GNI, which is the OECD measure of Official Development Aid over Gross National income. These measures dominate the literature. They are also used at present. Thus, aid effectiveness is $\beta = \partial g/\partial \alpha$, where the variables may be lagged, as will be discussed.

1.1 The elusive goal of aid

This definition of aid effectiveness has been greatly discussed. One reason relates to Ronald Inglehart's distinction between pre-materialistic and post-materialist values, where surely the LIC recipients have the first while the HIC donors have the second set of values. For the two parts to agree on the goals requires compromises.

(i) The post-materialistic values give economic growth a low preference, and this is certainly reflected in the rhetoric of altruism in the high-income countries. A large literature has criticized GDP as a welfare measure, but the many attempts to make an alternative have resulted in measures that are strongly correlated with the GDP. Until now no operational alternatives have been found.

[2] Google Scholar gave almost five million hits to 'development aid' in November 2023. An early survey of the literature is Cassen (1985, 1994). The reference list to the second edition covers 24 pages. In the thirty years since then, the stream of literature has swelled. See also the reference lists in the 32 chapters of Arvin and Lew (2015).

(ii) An early compromise is to add the word *sustainable* to development. This is surely a nice addition, but what the addition means is a major discussion. Perhaps it means that the long run should receive a higher weight.

(iii) Another reason follows from the discussion below: Given that aid and growth have zero-correlation in a simple univariate relation you must find other goals if you want aid to go on. Consequently, aid has been provided with many other goals, of which most are quite lofty and difficult to use,[3] so the literature on the purpose of aid is large.

Aid is only about ½% of the GNI of the donor countries. The aid recipients are mainly LICs, where aid is substantial, such as 10% of GNI (see Figure 4), or about 30% of government budgets. Given these magnitudes, it should be possible to find a measurable macro variable that is improved by aid. However, while we wait for such measures, the literature has concentrated on economic growth, where much data exists.

1.2 The replication crisis in the social sciences. Meta-analysis as an answer?

All sciences know that results need repeated replication to be believable. Mueller-Langer *et al.* (2019) find that only 0.1% of economics papers are replicated. Thus, replication is rare, and when done it often gives embarrassing results. There is indeed a replication crisis in the social sciences.[4] The technique of meta-analysis partly replaces clean replications, and it has been used in the aid effectiveness literature as discussed in section 4.

Table 1. Terms and variables used. Note the definitions of *a* (*aid*) and *g* (*growth*)

LDC	Less Developed Country, LDCs consist of LICs and MICs
LIC	Low Income Country, most are in Sub-Saharan Africa
MIC	Middle Income Country
HIC	High Income Country, or an international organization financed by such countries
GDP/GNI	Gross Domestic Product, Gross National Income. For most countries: GDP ≈ GNI
gdp	GDP per capita in fixed PPP prices. Source: Maddison Project
$y = ln\ gdp$	*Income*, the natural logarithm to *gdp*
g	*Growth*, the growth rate for *gdp*, close to the first difference of income
ODA	Official Development Aid, concessional flow from HICs to LDCs. Source: OECD
α	*Aid,* ODA as a share of GNI. Source: World Development Indicators
N	Number of observations, notably (*aid*, *income*) pairs
$\beta = \partial g/\partial a$	Aid *effectiveness*, where *g* and α may be lagged or averaged over various periods
AEL	Aid Effectiveness Literature. The app 200 papers with estimates of β, section 4

The country classification is the one used by the World Bank. Variables are in italics. The (*aid*, *growth*) data discussed are all overlapping observations from the two sources.

[3] Sometimes it is proposed that aid aims at improving civil society. It is an unclear concept, but data exists for two proxies for civil society: The level of democracy and corruption. A handful of papers study the democracy-aid relation and the corruption-aid relation. These papers give very small effects.

[4] In November 2023, Google scholar has almost 600,000 hits to "replication crisis" of which 1/3 is to economics.

A key reason for the crisis is that the classical empirical method (theory/model/-regressions) is flexible and thus susceptible to the influence of priors and interests. The flexibility may even explain why this method is increasingly popular.[5]

The ideal purpose of aid generates strong ***priors*** for results showing that aid works. In addition, the annual aid flows may now have reached $200 billion. This generates strong ***interests*** in keeping the flow running, also for most researchers in development, who consult in the aid industry. Both the priors and interests are pro aid. Consequently, studies of aid effectiveness may exaggerate the results. The AEL, Aid Effectiveness Literature, discussed in section 4 does find a moderately small result in the average. Meta-analysis tries to detect and correct publication biases, and in the case at hand the analysis does find that half the mean result in the AEL is an exaggeration; see the five papers by Chris Doucouliagos and the author in the reference list.

The survey compares ***three methods*** to assess aid effectiveness. Section 2 looks at the results from ***micro*** project evaluations. Section 3 covers the univariate ***macro*** evidence, which is rarely discussed, so here some estimates are reported. Section 4 turns to the AEL of multivariate ***macro*** models estimating aid effectiveness. The AEL is large, but as mentioned it has been submitted to meta-studies giving robust results. While the results from each method are clear, the results are contradictory across methods, as claimed by the micro-macro paradox. Section 5 discusses how the contradictions may be resolved, and section 6 concludes.

2. Micro: Project evaluations: social rates of returns

Two studies are normally made of each aid project. It is approved based on a ***feasibility study*** of its expected social costs and benefits, where the benefits are calculated as the increase in the GDP it may generate. This also applies to the projects included in larger programs. When the project is handed over to the recipient, an ***evaluation study*** follows. It should reassess the costs and benefits. The micro evidence on aid effectiveness is a summary of project evaluations.

Results of different projects are made comparable by a calculation of the ***SRR***, social rate of return, which is the internal rate of the return of the social costs and benefits. It appears that many donors use an SRR = 10% as the decision criteria, so that only projects expected to

[5] Paldam (2021b) provides some statistics on the methods used in economics, 1997-2017, from a sample of 3,415 papers chosen to be (reasonably) representative. Papers using the classical empirical method increased from 20% in 1997 to 34% in 2017. Paldam (2018) demonstrates that if researchers behave as predicted by economic theory, it creates publication bias when many researchers have the same priors and interests. The typical bias is an exaggeration, as is commonly found by meta-studies, see Ioannidis et al. (2017) and Doucouliagos et al. (2018).

have a higher rate should be approved.

Project studies are made by consultants that should fulfill the ***terms of reference*** agreed upon by the donor and the recipient. Consultants want the terms to be clear and simple, and they want to please their customers to get more business. As many donors have problems finding enough projects, and recipients want activity, they may both be keen on giving 'easy' terms of reference to consultants. Also, there are many political pressures on the aid process both from donors and recipients.

2.1 Six complications making the evaluations differ from feasibility studies

Consequently, the evaluation should check if the expected rate of return has been realized. This is often complicated. Eight complications – C1 to C8 – will be mentioned.

(C1) ***Long effects***. Many projects are meant to have long consequences. A project increasing the quality of primary school teachers is unlikely to influence *growth* in the first two decades, but if it is successful, there will be a positive effect for a long time once the effect starts. Such effects are difficult to detect and can only be assessed after a long time. However, evaluations need to be made when the project is handed over to the recipient. Consequently, the evaluation can only check if the expected activity has been implemented. If a teacher training college has been built, it can be checked that it is up and running as planned.

(C2) ***Externalities***. Projects often have externalities that are difficult and expensive to trace. Externalities that are included are specified in the terms of references.[6] Sometimes non-included externalities appear during the project implementation.

(C3) ***Activity vs development effects***. During the implementation, most projects have activity effects, which should be sorted out from the development effects. Think of the proverbial project where a hole is dug in the ground the first week and filled up the next. This has an activity effect both weeks, but no development effect. The macro literature does not distinguish between activity and development effects, but it is done in project studies.

(C4) ***Low executive capacity***. Projects succeed more often in some countries than in others. The countries where projects often fail are typically relatively poor and stagnating, i.e., the countries needing aid the most. These countries have little executive capacity, so projects need extra expenditures for capacity building. This reduces the SRR of the projects.

(C5) ***Unforeseen consequences.*** During the implementation, things often change in

[6] It is considered important for many projects how they influence the income distribution. The direct distributional effects during the implementation may be estimated, but the full dynamic effects are difficult to catch.

ways that were not predicted. Long ago Hirschman (1967) studied a set of large World Bank projects. He concluded that both costs and benefits became larger and often different from the expected ones. Evaluations of such cases need much more than a check of the predictions in the feasibility study.

(C6) ***Endless projects***. Many development projects are continued as 'follow-up projects' and 'renovation projects,' making them difficult to evaluate.[7] Typically, the evaluation says that the project was 'almost' fine, but that it needs some follow-up.[8] The well-known problem of ownership transfer is involved here. If the reason the project failed is that the recipient country did not take ownership, it does not help if the donor resumes ownership.

Projects fail for three reasons. The first is that the project was poorly planned. I have seen a large dam that failed because the river carried enough water to require a small dam only. The second is poor implementation. I have seen schools built of cement that crumbled after a couple of years. The third is that the recipients have other priorities. I have seen a teacher training college where classrooms were used for corn storage. Such cases are genuine failures. In addition, projects may fail for reasons external to the project, i.e., due to natural or political disasters or policy changes.

(C7) ***Natural or political disasters***. Floods, earthquakes, wars, and riots destroy some projects. How the detrimental effects of such events should be included in project assessments is not obvious, but the likelihood of various disasters may be included in the evaluation. In addition, decisions on the location of a project may increase or decrease tribal tensions. Such decisions may affect the probability that the recipient takes ownership.

(C8) ***Policy changes***. A trickier situation is that projects that were justified under one regime of public regulations made no sense when the regulations were changed. There have been large movements in regulatory regimes in most LDCs, going toward more liberal ones.[9]

2.2 Typical results

With all complications, a large literature evaluates projects. The World Bank has an evaluation department (Independent Evaluation Group). It issues an annual report and a dozen special studies every year. Most development agencies have similar divisions.

[7] Paldam (1994) is a post-evaluation study of 37 Danish development projects in nine countries, five years after they were completed. About half of the projects were continued as follow-up projects.

[8] A common argument is that if the follow up-project is not accepted, the project costs are wasted. This is surely a nice way to say that the project failed.

[9] It is difficult for a consultant to hint at the possibility that a policy may change. To handle the political problem, some cost-benefit manuals recommend that projects are analyzed both in current prices and in world market prices (calculated at project location), which are close to general equilibrium prices.

It is not controversial that realized social rates of return of projects have a wide distribution, where about half passes the criterion. About one third passes handsomely, and about one third fails badly, while the last third is in the gray zone where they almost pass. Thus, the average project may have a social rate of return that is close to 10%, but probably not quite 10%. Below, an average (macro) number is needed, so it is assessed that the average social rate of return of development projects is SRR ≈ 8%.

Given that aid is only 10% of what happens in the recipient country, it is arguable that the non-aid activity is much more important. However, aid is meant to increase development, so it should not be compared with all of GDP, but only the part of GDP that is for investments including social investment. This part is much smaller – maybe 25% of GDP. Thus, aid may amount to almost 1/3 of the development efforts.

Some aid projects are not development projects. They may aim at supporting cultural exchange, friendship, or try to build civil society. For example, in connection with the Arab Spring 2010-12, Arab countries received such aid. Recently, a lot of aid has been used for keeping refugees in camps near their country of origin. In addition, there is emergency and food aid. Thus, the SRR for the full ODA flow is a bit lower than 8% − perhaps SRR ≈ 6%.

Consider a country that receives 10% in aid. With a social rate of return of 6%, the growth generated will be a product of the two, i.e., ½%. Given that the average LDC has about 2½% growth, the contribution of aid should be about 25% of the growth, i.e., it would have been 2% without aid, but now it is 2½%. A factor that should give an effect of one fifth of the growth observed should be easy to detect, especially as aid is volatile.

The micro-results are based on cost-benefit techniques that in principle analyze the costs and benefits from the start of the project to infinity and distinguish between activity effects and development effects. The next two sections deal with macro-results made by a comparison of the macro-data for *aid* and *growth*. They typically look at data for one to two five-year periods and do not distinguish between activity and development effects.

3. Macro 1: Univariate studies, *the zero-correlation result*

The relation between *aid* and *growth* is a ***univariate*** relation. When it is analyzed by simple univariate tools, it is difficult for priors and interests to play a role. Other variables may blur the picture, but when *N* is increased the standard error falls, and the robust results stand out.

3.1 Correlations with 11 leads/lags, for all data and for Sub-Saharan Africa only

The six correlograms in Figures 1a and b show the correlation of *aid* and *growth* with 11 lags. The Appendix demonstrates that the data for *aid* and *growth* have different distributions and some outliers that may dominate the calculations of the relationship. Thus, the calculations are repeated after truncation of outliers, as explained in the note to the figure.

Figure 1a. Correlogram for the annual data for *aid* and *growth* (defined in Table 1)

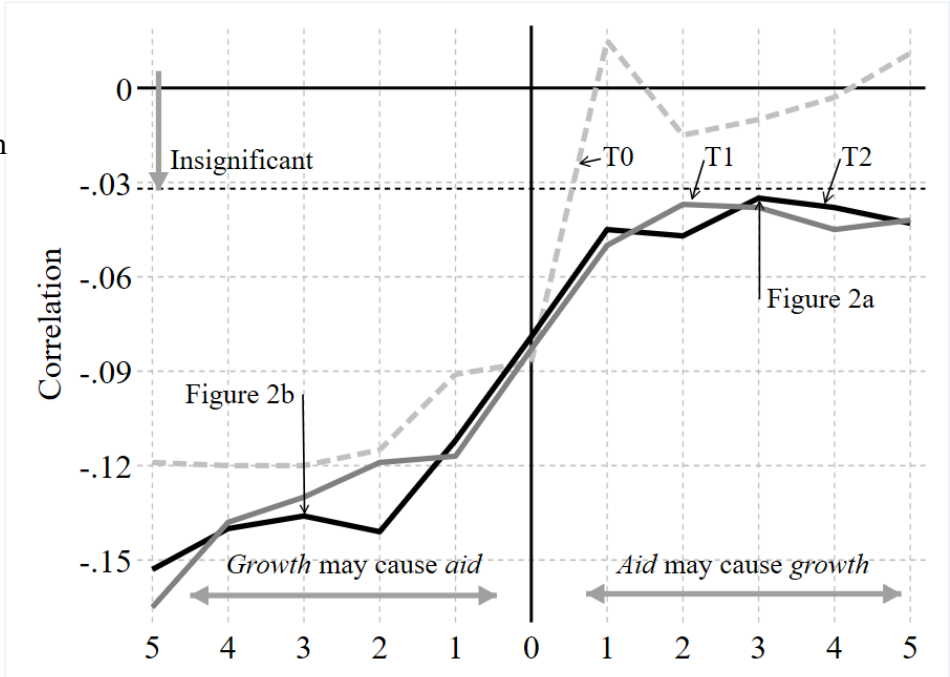


Figure 1a. All aid recipients. *N* is from 5,512 to 4,681

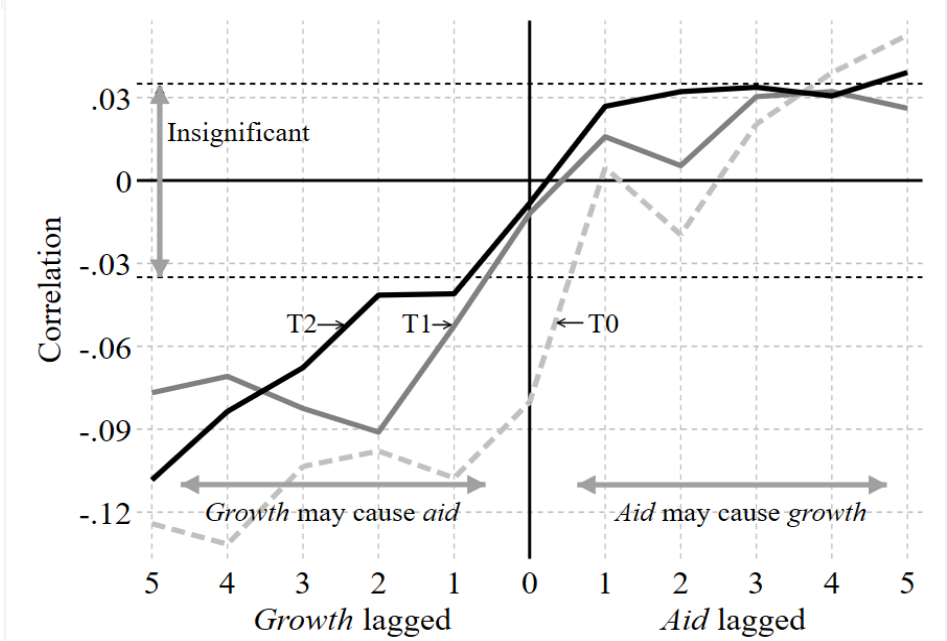


Figure 1b. Sub Saharan Africa. *N* is from 2,048 to 1,715

The high number for *N* is for the unlagged observation, before truncations, while the low number is for 5 lags after two truncations. The three curves are: T0: All data are used. T1: The data are truncated for aid larger than 50% or growth outside the interval +15%. This is about 5% of the data. T2: The data are further truncated for aid larger the 25%. This is a further 3% of the data.

All six curves in Figure 1 have roughly the same shape, so the 6 *x* 11 = 66 correlations presented on the figure show a robust picture, where only 14 – or 21% – are positive, while 52 are negative. The main difference between the two curve sets is that the curves on Figure 1b shift up by 0.04 correlation points. The lags indicate causality as noted. The vertical axis, for the unlagged relation, divides the graphs into two parts that suggest two causal regimes.

Left of the axis, the curves point to causality from *growth* to *aid*. All six times five average correlations are negative and significant. They suggest that low *growth* attracts *aid*. The level of the effect is −0.13 on Figure 1a, while it is −0.09 on Figure 1b. Both are significant.

Right of the axis, the curves point to causality from *aid* to *growth*. The level of the effect is −0.03 on Figure 1a, while it is +0.02 on Figure 1b. Here 16 of the correlations are negative, while 14 are positive, and most are insignificant. This is the ***zero-correlation result***. There is some indication of an activity effect for *aid* lagged one.

The fact that the left-hand and the right-hand side on Figure 1 is so different suggests that the aid to growth and the growth to aid relations can be estimated separately when lags are added. Thus, the AEL literature discussed below has no simultaneity issue.

3.2 Kernel-regressions for the best correlations on Figure 1

The two best points in Figure 1a are for three lags to either side for Cor3. The data for these points are further analyzed by a kernel regression $y = K(x, bw)$ that says if the variable x can explain y. A kernel regression may be understood as a smoothed moving average curve for a fixed bandwidth, bw. It is purely descriptive and assumes no economic theory and no functional form. The figures look much the same for the relations with 1, 2 and 4 lags.

Figure 2a for $g = K(\alpha_{-3}, 2.5)$ shows the zero-correlation result. It is possible to draw a horizontal line within the 95% confidence interval as shown. Thus, it is not rejected that the correlation is zero, though it was (just) rejected on Figure 1a. The AEL discussed in section 4 reports a model where aid squared was important for the result. Figure 2a shows why this variable has vanished.

Figure 2b is more interesting, as it shows a clear non-linear connection. In the range of growth from -5% to 8%, aid falls from 6% to 3.7%, i.e., by 2.3 percentage points. In this interval the slope is $\partial a/\partial g$ = -0.19. Outside the interval, the curve is flat. Thus, the best liner approximation for full range of growth from -15 to +15, the slope $\partial a/\partial g$ is smaller. Given that the density of growth rates is small at the linear ends of the scale, it is likely that the weighted mean of the slope is close to -0.13, which is the correlation on Figure 1a.

Figure 2. Kernel regressions for aid explaining growth and growth explaining aid

Figure 2a. Can *aid* explain *growth*? The point to the right on Figure 1a for three lags.

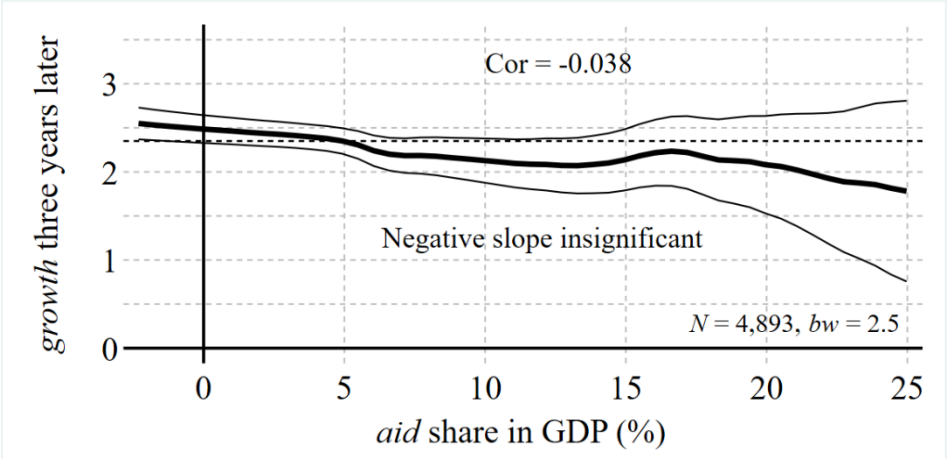
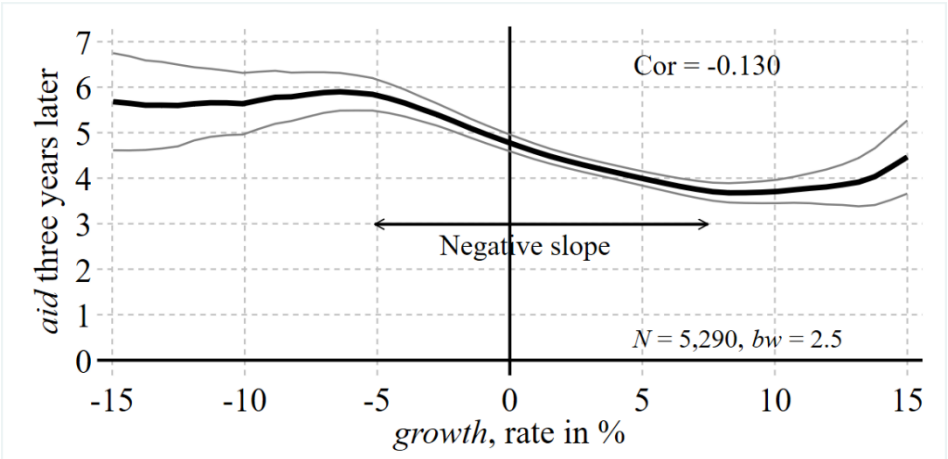


Figure 2b. Can *growth* explain *aid*? The point to the left on Figure 1a for three lags.



Both graphs are estimated with the stata-command *lpoly*, with the defaults and bw = 2.5. The scatter is suppressed, and the two thin gray lines are the 95% confidence intervals. The full set of all such kernels for all points of the T2 correlogram are provided in Paldam (2021c).

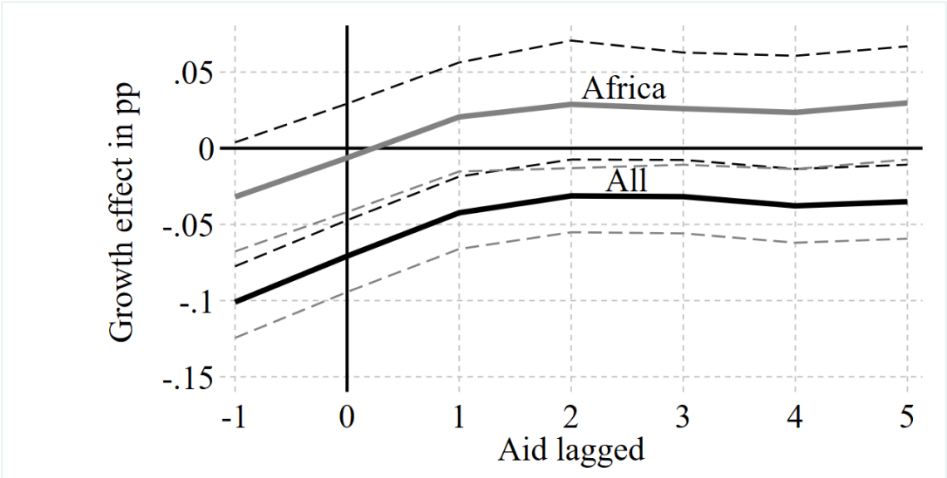
3.3 Regressions estimating the two T2-lines on right-hand part of Figure 1

Figure 3 is another way to assess aid effectiveness. Here the right-hand part of the two Cor-lines from Figure 1 is recalculated as the effect size from regression (1).

(1) $g = a + b\,\alpha + \epsilon,$ where ϵ is noise, and a and d are coefficients to be estimated

The regression curves are almost the same as the correlation curves. The upward shift in the curve when the estimates are done for the African countries compared with the curve for all aid recipients is 0.06 percentage points. The two sets of confidence intervals overlap, so it cannot be concluded that the curves are different. The right-hand side shows that the positive effect of *aid* on *growth* is insignificant. The result from equation (1) is consequently that *aid* has no significant effect on *growth*. That is bad according to the priors and interests of most researchers in the field. Hence, they are trying to do better.

Figure 3. Regression (1) $g = a + b\,\alpha_L + \epsilon$ for aid lagged explaining growth



Corresponds to the right-hand part of Figure 1. Calculated for the data used for T2. The effects are in pp, percentage points. The regression coefficients are almost the same as the correlations in this case. The estimates are surrounded by 95% confidence interval calculated as $b \pm 2se$.

4. Macro 2 multivariate studies: The family of AEL models

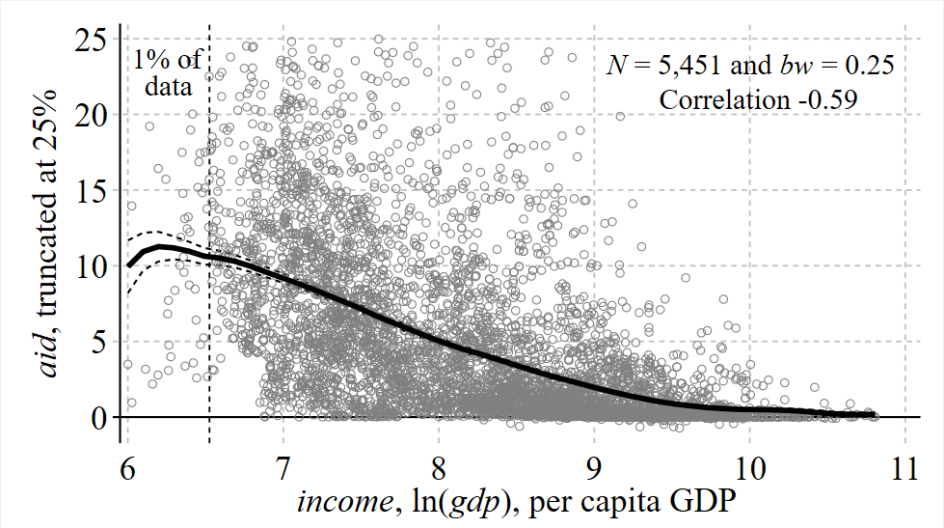
The multivariate analysis is known as the AEL, Aid Effectiveness Literature. It is a strangely isolated part of the cross-country growth regression field, which is known to be somewhat problematic. Wide cross-country datasets include countries at all income levels, and hereby they come to represent the Grand Transition from poor LICs to wealthy HICs. It is a highly confluent process that changes everything in society. Macro socio-economic variables have transitions, i.e., they change as a function of income from an LIC-level to a different HIC-level.

4.1 The transition in the aid share

Aid data have a typical transition shown on Figure 4, where the level of *aid* changes from about 11% in LICs to zero in HICs. Most transitions are similarly strong and fuzzy. The correlation of -0.59 only rises to -0.60 if the 1,500 observations for HICs with *income* from 10 to 12 and zero *aid* are added. These correlations are typical for socio-economic time series.[10]

[10] This section draws on the author's recent book Paldam (2021). It analyzes nine transitions in various fields. As much as possible they are analyzed in two dimensions: (a) wide cross-country samples, (b) long time series. The equivalence claim is that (a) and (b) show the same transition. It is confirmed in all cases examined. The agricultural transition looks amazingly like Figure 4.

Figure 4. The transition of *aid*, i.e., *aid* as a function of *income*



The data has $N = 5{,}646$, where 195 have *aid* larger than 25%. They are deleted as extreme. This has no visible effect on the transition curve estimated. The curve is surrounded by a 95% confidence interval. The intervals become wide for the last 1% data to the left, so they are consistent with a flat curve.

In the case of Figure 4, causality is clear. Countries do not get wealthy because aid falls, but when countries get wealthy, aid falls. Many other factors influence aid, to give the observed fuzziness. Other transitions may have some simultaneity, though most of the causality is normally from development to the other variables. [11]

Thus, most macro variables are confluent in cross-country datasets. The confluence is what is termed ***development***. It means that the variables all explain each other. Thus, when *growth* is explained by *aid* and some other variables, these variables are likely to be correlated with income. Hence, the explanatory variables have multicollinearity. Section 3 showed that growth is poorly explained by aid, so even moderate multicollinearity may have a relatively large effect on the small coefficient on aid. Consequently, if a researcher plays around with, e.g., 12 variables that are all correlated with development, the aid effectiveness coefficient will move around substantially, as indeed it does, see Figure 5. Thus, if you seek, you shall find.

4.2 The big family of model variants of the AEL

The aid effectiveness literature rarely refers to the general literature on cross-country growth regressions. It works to improve equation (1) from section 3.3 by an expansion to model (2) that adds a set of ad hoc control variables. Giving a family of model variants:

[11] Development is measured as the path of the GDP that is the aggregate of all variables. Thus, they all contribute somewhat to the GDP, but most give a small contribution, so the simultaneity is limited.

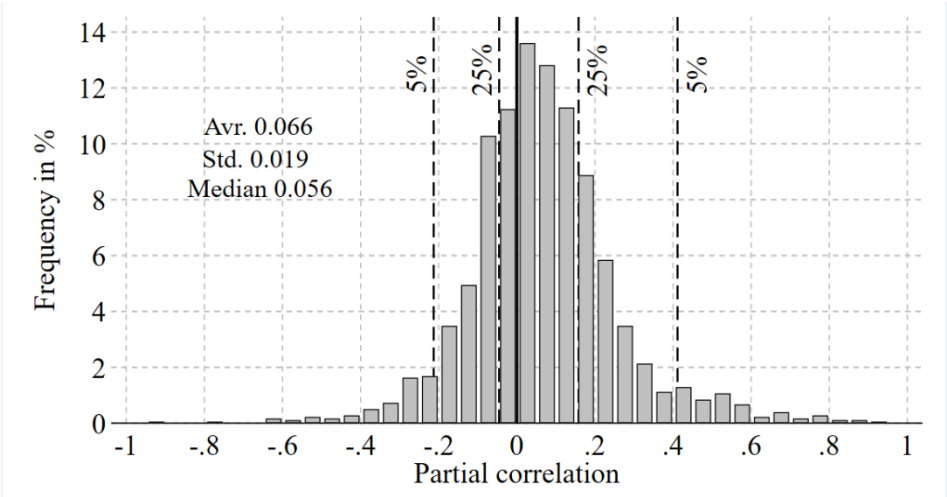
(1) $g = a + b\,\alpha + \varepsilon$, univariate relation, where $b \approx 0$

(2) $g = a + b\,\alpha + [c_1\,\gamma_1 + \ldots + c_n\,\gamma_n] + \varepsilon$, where [] holds control variables

Many datasets have been selected, the data have been averaged over 5 and 10 years, and they have been lagged. While OLS is the most common estimator, others have been used, notably estimators correcting for simultaneity. However, most variants are due to the controls that change from one paper to the next. They are typically ad hoc variables chosen after some experiments. Perhaps the average number of experiments per published result is $n = 25$?

About 5000 annual observations of the aid share are available, and about 2000 estimates of aid effectiveness have been published. If $n = 25$, it means that 50,000 regressions have been run on the 5,000 observations. That is a mining-ratio of 10 regressions per observation. As most relations use 5-year averages of the data, the mining-ratio is even 50 regressions per observation. Thus, we deal with a heavily mined field. Figure 5 shows the results. The reader should contemplate the distribution of the 48'000 unpublished results. I think that it is obvious that they have a lower average. Thus, the AEL is a ***swarm of partial replications*** of model variants, where only a few models are exactly the same.

Figure 5. The distribution of 1,779 estimates of aid effectiveness



Tests for skewness and kurtosis are highly significant, so normality is rejected.

Meta-analysis is a technique developed to sum up such literatures. The first 141 papers, reporting 1,779 estimates of aid effectiveness, have been subjected to such analysis. To make the results comparable, they have been converted to partial correlations, which correspond to the correlations in section 3. The frequency distribution of the published AEL results in Figure

5 shows the partial correlation. It has a theoretical range from -1 to +1, so the results are all over the scale. However, half of the estimates are between -0.05 and 0.18. The wideness of the distribution shows why it is dangerous to trust a single study. Thus, the figure illustrates the replication crisis in economics. [12]

The logic of Meta-analysis builds upon the idea that the estimated coefficient, *b*, and the standard error, *s*, should be independent, so the (b, 1/p) scatter (known as the funnel) should be symmetric. The standard test from Stanley (2008) detects an asymmetry and uses it to correct the mean giving an estimate of the meta-average; see Table 2.

Table 2. The mean and the basic FAT-PET for the 1,779 estimates

Mean	The FAT-PET MRA		
	PET meta-average	FAT asymmetry test	R^2
0.066 (14.6)	0.029 (3.5)	0.403 (5.2)	0.015

The parenthesis contains t-ratios. The calculations are made for the 1,779 estimates converted to partial correlations to be comparable, from the first 141 papers; see Doucouliagos and Paldam *op cit*. The Funnel Asymmetry Test (FAT) shows that the mean result is exaggerated, i.e., that the literature suffers from publication bias.

Table 3. The 22 controls used for the 1,779 estimates in order of their frequency

Control Number	Control variable	Included in *N*	%	Control Number	Control variable	Included in *N*	%
1	2 eq. growth savings	44	2	12	Capital controls	483	27
2	2 eq. growth aid	58	3	13	Policies	530	30
3	Aid *x* institutions	70	4	14	Ethno-linguistic index	605	34
4	Human capital	238	13	15	Inflation	644	36
5	FDI	224	13	16	Financial development	731	41
6	Size of government	250	14	17	Trade openness	740	42
7	Population size	292	16	18	Regional dummies	789	44
8	Aid squared	333	19	19	Aid instability	815	46
9	Fiscal stance	409	23	20	OLS	1,000	56
10	Aid *x* policy	411	23	21	Income	1,274	72
11	Aid lagged	463	26	22	Africa	1,535	86

The 22 controls are: Nine variables are self-explanatory. They are (4), (6), (7), (8), (11), (12), (14), (15) and (17). (1) and (2) are two-equation models, with either a growth and a saving equation, or a growth and an aid equation. (3) *Aid* interacted with an institutional variable. (5) Foreign direct investments. (10) *Aid* interacted with a measure for good policies. (13) A term for quality of policy. (16) Measure for financial deepening, such as bank balances over GDP. (20) Estimator, most non-OLS regressions are TSIV-regressions or GMM-regressions trying to account for simultaneity. (22) Only regional dummy is for Africa. Some estimates are from Africa only. Here Africa is coded blank.

[12] The meta-studies have been replicated by Mekasha and Tarp (2013, 2019), who did have some objections in the first paper, but in the second paper most objections have vanished, see Paldam (2023) for a study of the objections.

Fortunately, the meta-result has proven robust, see Doucouliagos and Paldam (2015). The meta-study identified the twenty-two control variables listed in Table 3. The average number included in the models is 6, and they can be selected in $\binom{22}{6} = 74{,}613$ ways, each giving a different estimate of β, so it is no wonder that the literature contains a plethora of model variants.[13] If each published estimate is selected from more than 25 experiments, it is possible that most of the 74,613 combinations have been tried.

The first meta-study of the AEL covered 68 papers, and here the PET was insignificant. When the number of studies grew to 141, the PET meta-average stayed rather constant. However, significance grew, and now the meta-average is significant.

4.3 The standard critique of meta-analysis

One of the key demands of a good meta-study is that it includes *all studies* with estimates of the effect analyzed, i.e., aid effectiveness β. The critique is that one should only look at 'good' studies, which may be studies in high quality journals, studies using frontline econometrics, or perhaps studies the reviewer likes, maybe because they confirm his priors.

It obviously goes against the whole purpose of meta-analysis to select a sample of studies. However, meta-studies allow the analyst to determine whether a certain sub-sample of 'good' studies gives different results. This is done by coding a variable for the 'good' studies and estimate the effect of that variable.

Many meta-studies have included the impact factor of the journal as a quality measure. This variable is rarely significant. This has sometimes been interpreted as saying that we only need to look at the results in top journals. However, top journals only publish a few studies in each field. Hence, the results are unlikely to be stable.

Also, a great many meta-studies analyze if a certain technique matters for the results. It might happen, but it is quite rare. The first paper introducing a methodological innovation often shows that it has a significant effect. But once it has been used in more papers, it often appears that the variable measuring the effect of the method fails to become significant.

In the AEL, many authors have claimed that it is important to control the results for simultaneity. For reasons shown in section 3, this may not be the case, and when a variable for studies adjusting for simultaneity is included that variable shows no effect, see Doucouliagos and Paldam (2011).

[13] When two estimates are compared, some may speak of two *models*, but if they only differ by some ad hoc control variables, it is better to speak of *model variants*. Reported robustness experiments are often done with model variants, where a handful of ad hoc controls are used.

4.4 A few additional results

The AEL has a substantial publication bias, as seen by the estimate of FAT in Table 2. The meta-average is a bit less than half of the mean of the results. The studies have also found that researchers working for the aid industry find larger results than other researchers do. This effect is not large, but it is hard to sort out priors and interests of researchers in this field, and most researchers seem to have both priors and some connection to aid flows. The bias is much as expected and not a rare finding in meta-studies; see the meta-meta studies Ioannidis et al. (2017) and Doucouliagos et al. (2018). It is a good rule-of-thumb to expect an exaggeration of 2 times in economics. It is no wonder that there is a replication crisis in economics.

It is also worth mentioning that two of the most celebrated controls are no. 3, aid times a good policy index, and no. 8, aid squared. The articles presenting the new control found fine results, so they were both warmly received by the aid industry,[14] and soon many papers used the variables as seen in Table 3. However, both failed at replications using other data than the original ones, and they disappeared from the literature. This shows why replication is crucial. As already noted by Aristoteles: "One swallow does not a summer make."

The results for the said two variables are typical. The first 10 of the 22 controls are used in less than 25% of the studies, and we have found no systematic pattern in the inclusion. Thus, the control set used has not stabilized, but is constantly changing. The publication bias suggests that they are used when they 'work' and not otherwise; where the term 'work' means that they help produce results desired by the researcher.

Consequently, the AEL can be summed up as: (i) aid has a small effect on development, and (ii) the estimation model contains a shifting handful of variables that are correlated with *aid* due to the general confluence of development. This is a fine case of multicollinearity, where the explanatory variables get unstable coefficients. This surely applies to the coefficient to *aid* that looks as shown in Figure 5. When combined with strong priors and interests, it gives a publication bias as found.

5. The inconsistency of the results: The micro-macro paradox

Thus, the literature on aid effectiveness suffers from a contradiction between the satisfactory micro-result and the weak macro result. After almost 40 years, the contradiction is as large as

[14] Each of the two aid agencies of the authors even financed a book where the article played a prominent role; see Burnside and Dollar (2000) and Hansen and Tarp (2000).

ever. Five possible explanations – (E1) to (E6) – may explain the paradox.

(E1) ***Long-run***. Macro-studies rarely extend above 5 years, while cost-benefit studies in principle have a time horizon to infinity, where the future costs and benefits are discounted to the present day. Many projects start with an investment that is all costs and brings benefits only sometime after the project is completed. If the analysis has a 5-year horizon only, it may catch only a (small) fraction of the benefits. Thus, aid effectiveness is underestimated.

(E2) ***Activity effects***. Macro-studies always catch the activity effects that are not the purpose of the project. Thus, aid effectiveness is overestimated.

(E3) ***National accounting***. The quality of national accounting is low in many LDCs. This will affect the macro-results downward, but not the micro-results.

The last three explanations argue that the micro-studies do not catch important negative externalities that are far from the project itself. They argue that the micro-results are too high.

(E4) ***Fungibility***. Aid is often used to finance good projects that would have been implemented anyhow. Herby funds are set free to make other projects, so the project financed by aid is different from the project caused by aid, which is likely to be less good. Hence, the effect of the aid is lower than calculated from the social cost benefit analysis, perhaps even considerably lower. In a corrupt environment,[15] the corrupt will know that aid projects are more carefully monitored than other projects, so they may take their cut elsewhere.

(E5) ***Dutch disease***. Aid primarily goes to pay for public expenditures. Some such expenditures are investments that may increase future growth, and this is accounted for in the cost-benefit study. But it does not account for the exchange rate effect. Aid is an inflow of foreign capital to pay for public expenditures. It will cause some revaluation, which harms the development of business in the recipient country, and thus future growth.

(E6) ***Low executive capacity***. Aid projects are known to make big demands on executive capacity, thus there is less available to other projects. Consequently, they may suffer.

The five reasons given suggest that explanations may be found closing the gap, or at least some of the gap.

[15] Corruption is a socio-economic variable that has a strong transition, see (Paldam, 2021). LICs are more corrupt than HICs.

6. Conclusions

The three surveys in sections 2 to 4 show that different approaches give different results as regards the effect of aid. The attempt in section 5 to list factors that may close the gap produced a list of six possibilities that may work, but neither of these possibilities are easy to quantify. As of now, it is only safe to say that little can be said with any precision about the effect of aid, except that it is not very high.

The AEL method of multivariate regression, which most economists seem to prefer, is shown to be particularly uncertain. The field of cross-country regression has increasingly been found to be too flexible. A main problem is that most socio-economic variables contain transitions, so that they change systematically from one level in poor countries to a different level in wealthy countries. They are strong both in long time series and in wide cross-country samples. Consequently, development is a strongly confluent process. Combined with the low fit in general, this makes coefficients unstable, and thus susceptible to the influence of priors and interests. Consequently, considerable weight should be placed upon the robust univariate results.

References:

Arvin, M., Lew, B., eds., 2015. *Handbook on the Economics of Foreign Aid.* Edward Elgar, Cheltenham

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

Cassen, R. and associates, 1985, 1994. *Does Aid Work?* Oxford UP, Oxford, UK

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

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

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

Doucouliagos, H., Paldam, M., 2011. The ineffectiveness of development aid on growth: An update. *European Journal of Political Economy* 27, 399-404

Doucouliagos, H., Paldam, M., 2015. Finally, a breakthrough? The recent rise in the size of the estimates of aid effectiveness Cpt. 20, pp 325-49 in Arvin and Lew (2015)

Doucouliagos, H., Paldam, M., Stanley, T.D., 2018. Skating on thin evidence: Implication for public policy. *European Journal of Political Economy* 54, 16-25

Hansen, H., Tarp, F., 2000. Aid effectiveness disputed. *Journal of International Development* 12, 375–398

Hirschman, A.O., 1967. *Development Projects Observed.* Brookings, Washington, DC

Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H(C), 2017. The Power of Bias in Economics Research. *Economic Journal* 127, 236-65

Maddison Project, URL: https://www.rug.nl/ggdc/historicaldevelopment/maddison/

Mekasha, T.J., Tarp, F., 2013. Aid and growth. What meta-analysis reveals. *Journal of Development Studies* 49, 464-83

Mekasha, T.J., Tarp, F., 2019. A meta-analysis of aid effectiveness. Revisiting the evidence. *Politics and Governance* 7(2), 5-28

Mosley, P., 1986. Aid effectiveness: the micro–macro paradox. *Institute of Development Studies Bulletin* 17, 214-25

Mueller-Langer, F., Fecher, B., Harhoff, D., Wagner, G.G., 2019. Replication studies in economics - How many and which papers are chosen for replication, and why? *Research Policy* 48(1), 62-83

Paldam, M., 1994. *Dansk u-landshjælp. Altruismens politiske økonomi*. Aarhus, Aarhus universitetsforlag

Paldam, M., 2018. A model of the representative economist, as researcher and policy advisor. *European Journal of Political Economy* 54, 1-15

Paldam, M., 2021a. *The grand pattern of development and the transition of institutions*. Cambridge UP, New York and Cambridge UK

Paldam, M., 2021b. Methods used in economic research. An empirical study of trends and levels. *Economics* 15, 28-42

Paldam, M., 2021c. The zero-correlation result. The univariate relation between aid and growth. Working paper available from http://martin.paldam.dk/Papers/Aid/Zero-correlation.pdf

Paldam, M., 2023. Meta-mining. The political economy of meta-analysis. *Kyklos* 76, 125-40

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

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

Stanley, T.D., Doucouliagos, H., 2014. Meta-regression approximation to reduce publication selection bias. *Research Synthesis Methods* 5 (1), 60–78.

World development indicators, URL: https://databank.worldbank.org/home.aspx

Appendix: The distribution of aid and growth

This appendix shows how the data looks for the aid recipients. The data are all overlapping data from the two sources listed in Table 1. Both distributions have long tails, with extreme observations. The vertical dashed lines show two truncations of outliers used in the calculations. Both distributions are skewed. Sub-Saharan Africa is the poorest group of countries, and thus the countries of the group are relatively large aid recipients as expected.

Figure A1. The distribution of the *aid* data

Figure A1a.
All data

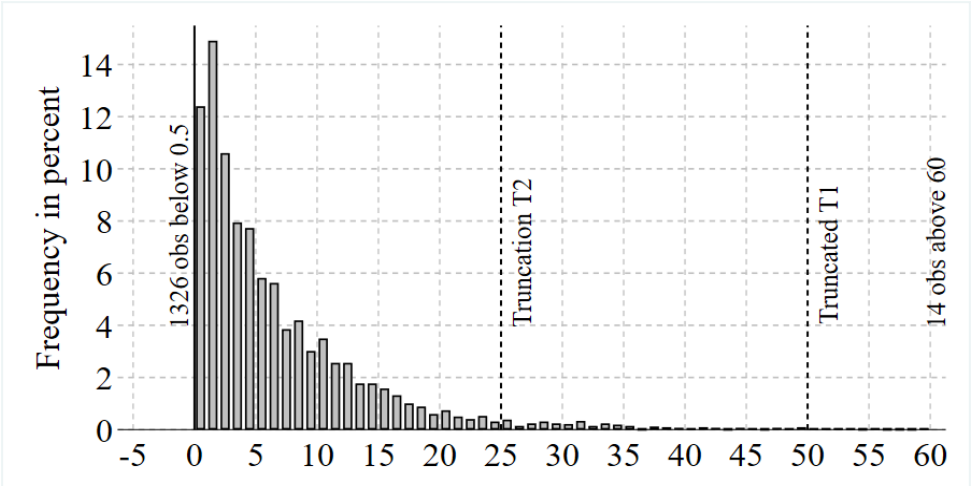
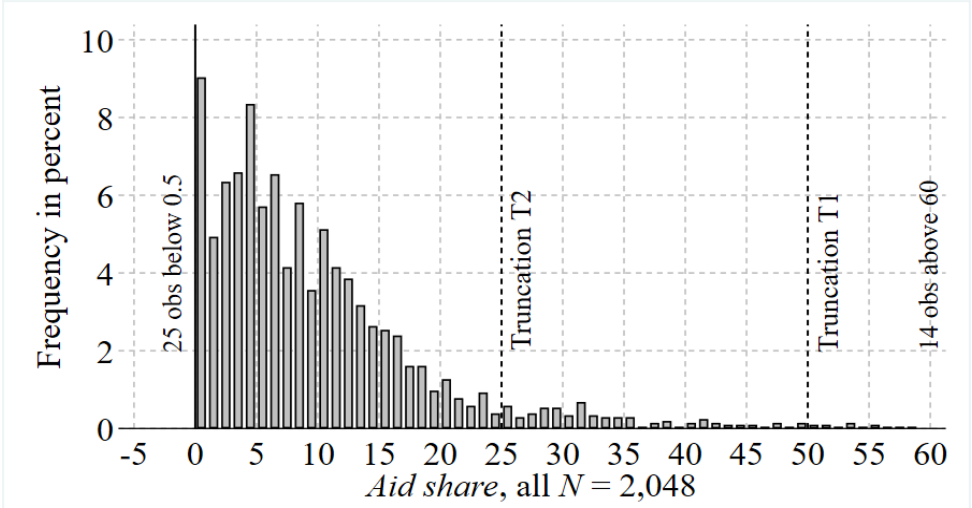


Figure A1b.
Sub-Saharan
Africa data



The N = 1,326 data with *aid* below 0.5 have 99 below zero, and 486 between 0 and 0.1. There are only 25 observations from Africa below 0.5%. The truncation definitions are from Figure 1.

The growth data are less skewed, especially when truncated for the extreme tails. For the African data, the growth tragedy 1972-94 (with negative growth in the average country)

shifts the distributions to the left compared with all LDCs.

Figure A2. The distribution of the *growth* data

Figure A2a.
All data

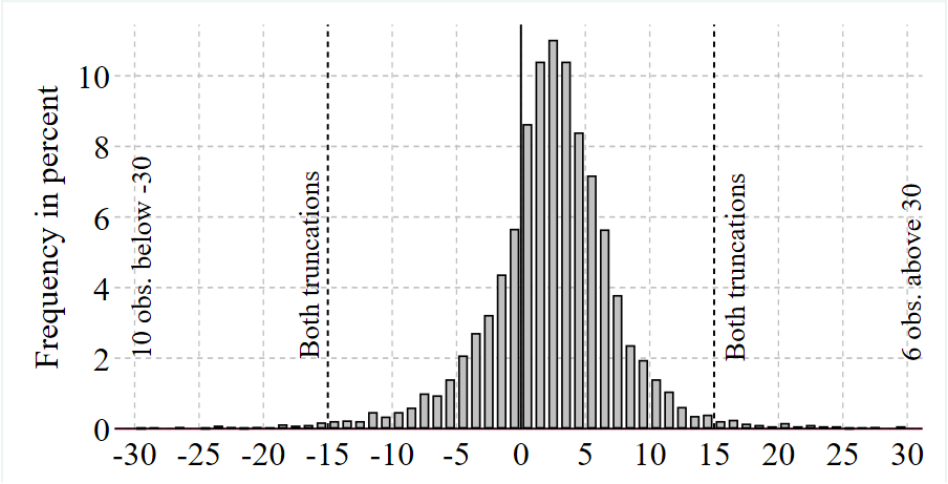


Figure A2b.
Sub-Sahara
Africa data