The Aid Effectiveness Literature:

The Sad Results of 40 Years of Research

Hristos Doucouliagos and Martin Paldam

December 2007

CWPE 0773

The aid effectiveness literature:
The sad results of 40 years of research

Hristos Doucouliagos and Martin Paldam

HD: School of Accounting, Economics and Finance, Deakin University, Melbourne Campus, 221 Burwood Highway, Burwood, Victoria 3125, Australia. Email: douc@deakin.edu.au

MP: School of Economics and Management, University of Aarhus, Building 1322, 8000 Aarhus C, Denmark. Email: mpaldam@econ.au.dk

Abstract: *The aid effectiveness literature (AEL) consists of empirical macro studies of the effects of development aid. By the end of 2004, it had reached 97 econometric studies of three families, which have been analyzed in one study for each family using meta-analysis. The AEL is an ideal subject for meta-analysis as it uses only a few formally similar models which try to catch precisely the same effects. Also, it is an area with strong beliefs – often generated by altruism – and interests. In this survey of the AEL, we shows that when the whole of the literature is examined, a clear pattern emerges in the results: after 40 years of development aid, the evidence indicates that aid has not been effective. We show that the distribution of results is significantly asymmetrical in a way that reflects the reluctance of the research community to publish negative results. The Dutch Disease effect of aid has been ignored but is a plausible explanation for aid ineffectiveness.*

JEL: B2, E21, E22, F35, O35

Keywords: Aid effectiveness, meta study, accumulation, growth

The aid effectiveness literature:
The sad results of 40 years of research

1.   Introduction: Analyzing the process of research in aid effectiveness

The welfare loss due to mass poverty in the LDC world is one of the largest problems facing mankind. Many in the rich countries want to do something to reduce the problem, and all DCs give development aid. In addition, many in the West want aid to be increased. This raises hope of a better world, but the issue has always been surrounded by doubts.

The fast growth of China and India causes mass poverty to fall more than ever before, but the two giants hardly receive aid. Poverty is falling much slower – if at all – in the main aid recipient countries.[1] Also, it is well-known that the correlation between aid and growth is essentially zero. The mixed casual evidence and the zero correlation result have caused many to doubt the effectiveness of aid.

The mixture of hope and doubts is the main reason for the large empirical ***AEL***, aid effectiveness literature, based on macro data. The AEL considers *the effect of aid on growth, savings and investment*. To study the effect on growth, a model, *g = g(h)*, of growth *g* as a function of *h*, the aid share, has to be estimated. Effectiveness means that $\mu = \partial g / \partial h$ is significantly positive. The AEL analyzes *well defined effects*.

Three recent meta-analyses of this literature (Doucouliagos and Paldam 2006, 2007 and 2008)[2] have concluded that aid has been ineffective. The aims of this paper are to survey the AEL and to step back from the details of the meta-analyses and to draw several inferences from the key results. In particular, we consider several important aspects not explored fully in the extant meta-analysis of the AEL. We are especially interested in the pattern of findings over time, the impact of priors and biases in the AEL and the apparent lack of learning by doing by the aid industry. Further, we offer the Dutch Disease effect as an explanation for aid ineffectiveness.

1.1   Meta-analysis as a quantitative study of a research process
The *Journal of Economic Surveys* published the first study on meta-regression analysis in economics (Stanley and Jarrell 1989). Since then, applications have risen at an exponential rate.[3] The purpose of a meta study is to perform a quantitative "forensic" analysis of the case

of *one literature*, as the AEL, which estimates the very *same effect*, such as μ. Research is a process that works by a mixture of: (a) *innovation*, theory, models and estimators, where new results are found, and (b) *independent replication* by other authors and data.[4] Through this process, new results are produced, and they gain or lose credibility. We all hope that the research process converges to truth. Applied to the AEL, meta-analysis tries to answer three questions about the research process:

Q1: Do the estimates in the AEL *converge* to something we can term truth?
Q2: Can we identify the main *innovations* causing/preventing the convergence?
Q3: Are there *biases* in the way the process of truth finding works in the AEL?

The data of a meta study is a coding of all estimates of one effect, such as μ, in the literature, where each one is provided with a (standardized) vector of information characterizing the methods by which the estimate is reached.

Obviously, Q3 provides information about the priors and the incentives of the research process in the case, i.e., the AEL. Some of these findings are embarrassing, or even painful. However, they correspond to the suspicions researchers routinely vent at the lunch table of *all* research institutes we know. All we do is to provide formal tests of fairly common suspicions, of which some are confirmed, and others are rejected.

1.2   Three perennial problems

A key problem in macroeconomics is that it has rather limited data, which are constantly mined by many researchers. In many fields, one may view the research community as a collective constantly fishing in the "common pool" of available degrees of freedom. Below, we shall demonstrate the amount of data mining done in the AEL, but we want to stress from the start that it is a general problem.

We all know from introspection that when we study an empirical question, we analyze the data till we are *satisfied* with the result. Results are thus generated by a *stopping rule* for a *search*. We all want to believe that we stop when we have reached *the truth*. However, what we believe is the truth is influenced by priors, and there are also incentives in research. It is not a trivial matter if the incentives are truth-finding-consistent.

Thus, *data mining*, *priors* and *incentives* create inevitable problems – and it is an important empirical question how much they matter. These problems ought to cause all of us to treat new results generated by an *innovation* with some skepticism until they have

successfully passed *independent replication*. Sciences such as physics, chemistry and medicine always demand independent replication. However, this is even more important in economics due to the lack of data and the very limited possibilities for making controlled experiments.

We have found that the literature on aid effectiveness provides an ideal case for meta-analysis. The effects analyzed are well-defined; the models are so simple that their differences can be quantified; aid is an emotional issue where priors are strong; and the multi-billion "aid industry" is heavily involved in the research.

In the AEL, priors and interests work in the same direction. We formulate this as the following *reluctance hypothesis*: Aid effectiveness is a field where researchers (and maybe also some journals) are *reluctant to publish negative results*. This is a hypothesis that may work both internally in the individual researcher through the stopping rule and externally through the whole process of publication. Reluctance creates a truncated distribution of results. This tilts the process of convergence, so it is a problematic phenomenon.

Table 1. Main conclusions from our three meta studies

Type	Causal link	Conditional on	Conclusion	Significance	Section in this paper
Family A	Aid ➔ investment		App. 0.25	Dubious, from 0	5
	Aid ➔ savings		App. -0.65	Dubious, from -1	
Family B	Aid ➔ growth		Positive, small	No	6
Family C	Aid|condition ➔ growth	Good policy	Rejected	No	7
		Aid itself (aid squared)	Positives small	Dubious	
		Eight more models	Not replicated		

Note: Aid either leads to increased investments or is crowded out through a fall in savings. Hence, the investment effect minus the savings effect should add to 1 when the BOP effect is added. Significance means that the investment effect is larger than 0, and that the savings effect is larger than -1.

1.3   A preview of the results

A thorough search showed that the AEL as of 1/1-2005 consists of 97 studies.[5] Doucouliagos and Paldam (2006) divide these studies into three main families of models. Each family has been subjected to a meta study. Consequently, we now know precisely what the AEL says. The results vary remarkably, but the aggregate results are sad as summarized in Table 1. The average effect of aid is positive, but it is small and of dubious significance statistically. With

the accumulation of more data, the results have gradually grown worse. The latest disappointment is the collapse of the once promising Good Policy Model when it was submitted to independent replication, as discussed in Section 6.

Section 2 of the paper looks at the data for aid and growth and shows that they have no correlation. It is argued that this contrasts with standard economic theory. Section 3 divides the AEL into three families of models, and gives the dimensions of the data mining that has taken place. Section 4 discusses biases and priors. Sections 5 to 7 present the results of the meta-analysis of the three families of models. Section 8 discusses an overlooked parallel to a literature that may explain the findings of the AEL. Section 9 concludes the paper.

2.   Absolute aid effectiveness and the aid paradox

Development aid programs started during the 1960s and have generated a large literature covering every aspect of aid one can think of. Two points should be evident from the start:

E1:   Aid agencies aim at social rates of return of approximately 10% in feasibility studies of their projects. If this is realized, an aid share of 1% (of GDP) gives a growth rate of 0.1%.

For many reasons, 10% is likely to be optimistic, but project evaluations typically find that half of all projects succeed, so we expect an effect between 0.05 and 0.1. Consequently, we term this interval the *range of reason*. The average aid share is about 7.5% so it follows that aid should contribute between 0.4 and 0.8 percentage point to the growth rate. As the average LDC growth rate is about 1.6%, aid should explain between 25% and 50% of the growth in the average country. This is substantial and should be very easy to show.

E2:   Both aid agencies and recipients have now had about 40 years of experience, where the process of learning by doing should have improved aid effectiveness.

Studies of the size of learning by doing typically find orders of magnitude of 1-2% per year (see Barro and Sala-i-Martin, 2004, 212-220). Over 40 years, this should cause an increase in effectiveness between 50% and 100%. This should be clearly visible as a *positive trend* in empirical estimates of μ e.g. the $\mu = \mu(t)$, and $\mu = \mu(N)$ curves, where t is time and $N = N(t)$ gives the number of observations. The fact that we find significantly negative trends in both $\mu = \mu(t)$, and $\mu = \mu(N)$ below is taken as evidence for the reluctance hypothesis.

Figure 1a. A scatter plot of the data for aid and growth
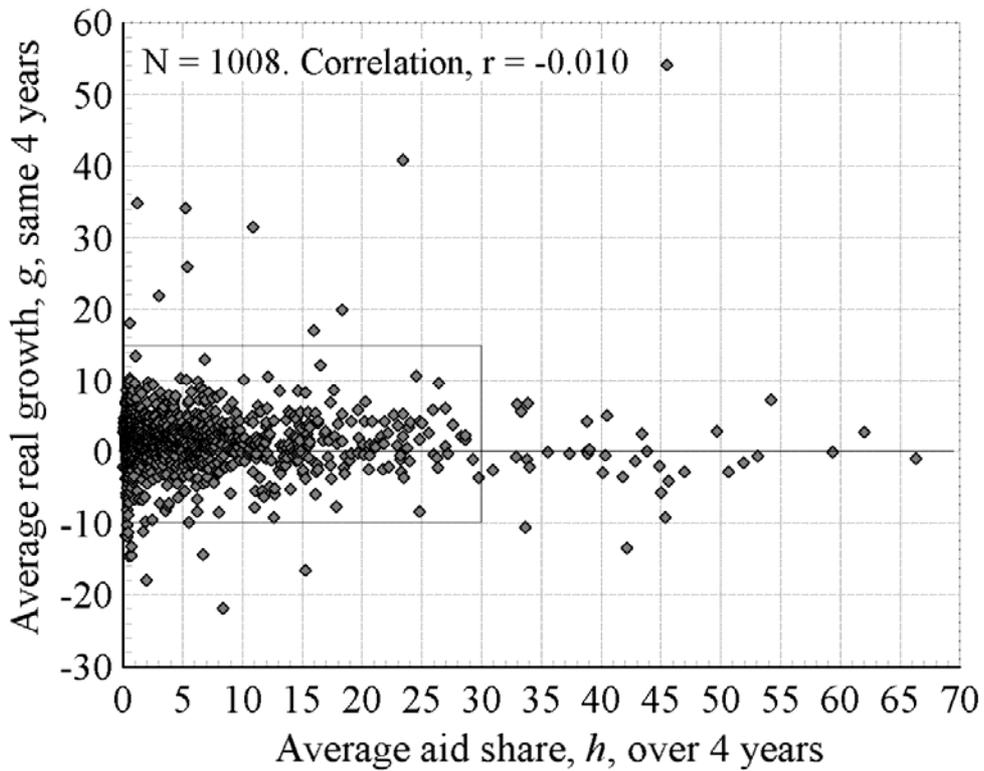
Note: The densely packed observations in the "box" are enlarged on Figure 1b.

Figure 1b. The box on Figure 1a enlarged
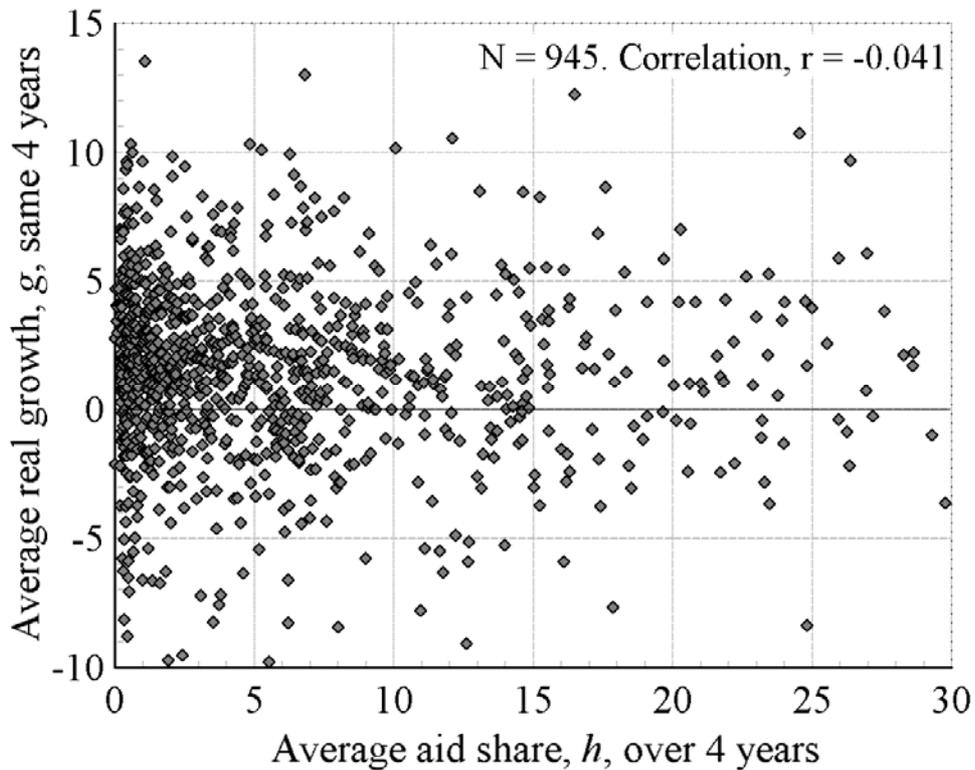
Note: An *Appendix* with similar graphs lagged to both sides is available, see Paldam (2005).

2.1   The data, definitions and the zero correlation result

Aid statistics started gradually during the 1960s, and since 1970, data have rapidly accumulated and now grow by about 140 observations per year. The available data are included in the World Development Indicators. WDI covers 156 LDCs, starting from 1960, so in 2000, we should have 7,200 observations or 1,560 observations when a 4-year average is calculated. About 35% are missing, but this still leaves the 1,008 4-year averaged observations shown in Figure 1. The figure shows two series averaged to the 10 four-year periods: 1961 to 64, 1965 to 68 … 1997 to 2000. The series are:

Growth, *g*, the real rate of growth of *gdp*, that is, GDP per capita.
Aid, *h*, as the share of development aid (ODA), *H*, of *GDP/GNI*, so that *h = H/GDP*.

The model, *g = g(h)*, may be uncontrolled or controlled for country heterogeneity, which we term *absolute* and *controlled aid effectiveness* respectively.[6]

Figure 1 shows the raw data, and Table 2 brings the simple regressions for absolute aid effectiveness. Aid effectiveness is rejected by the data. This has been known since Griffin (1970). It has been replicated regularly since then. It is, for example, documented in great detail by Mosley (1987), and recently in Rajan and Subramanian (2005) and Herbertsson and Paldam (2007).[7]

The zero correlation result has caused the literature on growth and convergence to ignore development aid. For example, neither the standard textbook by Barro and Sala-i-Martin (1998, 2004) nor the 1800 pages of the Handbook of Growth (Aghion and Durlauf, 2005) mention aid.

Table 2. Absolute aid ineffectiveness: Simple regressions between aid and growth

Same data as Figure 1		(1) Lag +1 growth before aid		(2) Lag 0 aid and growth		(3) Lag −1 aid before growth	
		Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
All data	Constant	**1.816**	**0.000**	**1.579**	**0.000**	**1.504**	**0.000**
	Effect/slope	**-0.039**	**0.023**	-0.010	0.935	0.003	0.364
	N	895		1,008		876	
	R^2	0.006		0.000		0.000	
In box	Constant	**1.843**	**0.000**	**1.676**	**0.000**	**1.578**	**0.000**
	Effect/slope	**-0.052**	**0.007**	-0.022	0.207	-0.010	0.559
	N	841		945		839	
	R^2	0.009		0.002		0.000	

Note: Bolded estimates are significant at the 5% level.

6

2.2   Four reasons why the absolute aid ineffectiveness result is puzzling

R1: *The micro evidence*. All aid programs have an evaluation process, and many studies have summarized the findings. Cassen (1986, 1994) is the classic survey, and the results are uncontroversial. About 50% of all development projects work, and very few of the remaining projects harm even if they fail. Simple aggregation thus leads to average aid effectiveness as per (E1) above. The contrast between macro level ineffectiveness and micro level effectiveness is known as the micro-macro paradox since Mosley (1986).[8] **Counter-R1:** Aid is fungible, and donors have a good chance of selecting non-marginal projects, so the marginal effect of aid is somewhat less effective than the projects financed by aid.

R2: *Standard macro theory*. Aid leads to a balance-of-payments improvement and to public spending. Public spending has an activity effect, and that effect can be permitted to run in the economy due to the balance-of-payments improvement. **Counter-R2:** A whole set of arguments exists why some of the activity effect may not have its full size, but is *crowded out*. The fullest crowding out is the *Ricardian Equivalence* by which loans – that have to be paid back – will have no effect, but lead to increased savings. However, development aid contains a gift, which does not have to be repaid, and thus should have an effect.

R3: *Standard growth theory*. Both the theory of growth and growth empirics show that increased accumulation causes growth. We know that aid finances development projects which are – in principle and often in practice – investments. **Counter-R3:** Accumulation is only one factor generating growth, and the marginal activity caused by aid is somewhat different from the activity financed. The link from aid to growth does not necessarily proceed via the accumulation effect generated.

R4: *The "why would they" argument*. Given standard rationality assumptions, an activity such as aid that has run for 40 years must do at least some of what it should. Why else would it continue? **Counter-R4:** The average aid share of the donors is actually quite small (about 0.3% of donor GDP) and has even decreased a little in the last decade due to *aid fatigue* caused by precisely the dissatisfaction with the small effects of aid. Also, it is obvious that while aid optimism was high during the first 1-2 decades, a business was created that has the usual stakeholder interests in the continuation of the activity. Moreover, the aid allocations literature suggests that at least some aid is given for non-humanitarian reasons – including commercial and security interests and human rights concerns.

Seen together, R1 to R4 suggest that aid should help – maybe not very much – but at least more than shown in Figure 1 and Table 2, which suggest a total lack of effect. This is the

aid paradox that has driven research in the field. One can see the research in aid effectiveness as an attempt to *overcome* the paradox.

2.3   Some additional observations

A look at Figure 1 allows us to draw two more conclusions: First, the data are ideal for an analysis of aid effectiveness. They are plentiful and have great variance. Aid shares have an average of 7½% of GDP. This is substantial relative to other quantities that are known to affect growth as per (E1) above. Secondly, the fact that the raw data show nothing means that any significantly positive (or negative) effect found must be due to the imposition of structure on the data. That is, results are due to the "frills" of the analysis, as is further discussed below.

The analysis of aid effectiveness takes each country to provide equally good information for the analysis. The data points from India and Mauritius in Figure 1 are thus of the same size. We also noted that the average aid share is 7½% in the average country. However, the aid received by the average citizen in the LDC world is much smaller for two reasons: (1) A well-known fact about the cross-country distribution of aid is that its share falls with the size of the population. Most countries are small, but the bulk of the population lives in large countries. The giants, India and China, have aid shares well below $^{1}/_{4}$% and almost 40% of the LDC population. (2) About a third of the information for the calculation of the true average is missing. It is likely that the missing values are below average. If aid shares are weighted with population sizes of the countries, the average share falls to about one third, i.e. to 2½% of GDP.[9] Thus, the cumulated aid – over the 40 years – corresponds to one annual income (GDP per capita) of the average citizen of the LDC world.

Finally, it should be mentioned that the standard ODA measure of aid is defined as unilateral transfers with a gift element above a threshold of 25%. However, Chang, Fernandez-Arias and Serven (1998) introduced the EDA measure of aid, where each grant is weighted by its gift element. EDA data are available for fewer countries and years than the ODA data, but the two data sets have a correlation of 0.83 when overlapping. Since the EDA data became available, some of the research has used EDA and some ODA data. Doucouliagos and Paldam (2008) have dealt with this complication by a special EDA dummy and by converting coefficients to partial correlation which are invariant to the pure shift in scale. The EDA dummy becomes negative. We interpret this as evidence that policy makers are myopic. They consider the size of the ODA, and largely disregard repayments, which are likely to be the problem of later governments anyhow.

The conclusion of this section is that the absolute aid ineffectiveness result stands. Ineffectiveness still stands if the aid effectiveness relation is extended with the standard control sets from the convergence literature.

The aid effectiveness discussion concerns controlled effectiveness. Here, the problem is if a *basic set of controls* has been found which turns the ineffectiveness result into effectiveness, in a way which is both robust and justified by economic theory.

3.   Structure and statistics of the AEL

This section first classifies the models of the AEL and shows that it falls in three groups with perfect formal homogeneity. The development in the production of papers is shown. And then we give statistics showing the ratio of estimates relative to the number of models estimated.

3.1   The structure of the AEL: Three families of models
The AEL has explored many models, but from a formal point of view it can be divided into three main families by their causal structure as shown in Figure 2.
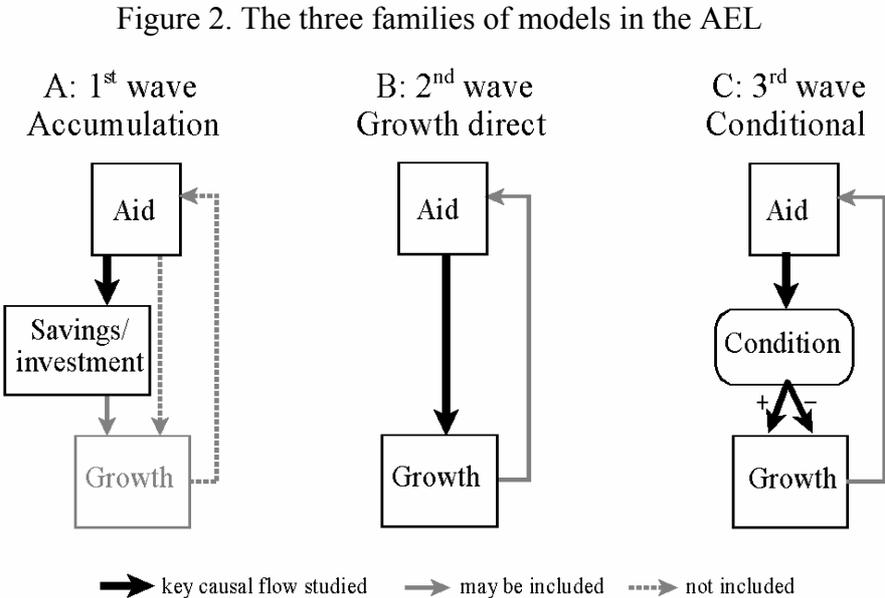
Figure 2. The three families of models in the AEL

Table 3 gives the formal structure of the models. Within each family the models have perfect *formal homogeneity*, as the models only differ in three ways: (i) the data sample on which they are estimated, (ii) the control set, which can be seen as a choice from a "master set" of 60

variables, and (iii) the estimator used. These differences can be coded, and this makes it ideal to submit each family to a meta study.

Table 3. The models and variables of the AEL

Family of models	Model – all models of each family has the format given:
A: Accumulation	$s_{it} = \boldsymbol{\alpha} + \mu\,h_{it} + \boldsymbol{\gamma_j}\,\boldsymbol{x'_{jit}} + u_{it}$ and $i_{it} = \boldsymbol{\alpha} + \mu\,h_{it} + \boldsymbol{\gamma_j}\,\boldsymbol{x'_{jit}} + u_{it}$
B: Growth	$g_{it} = \boldsymbol{\alpha} + \mu\,h_{it} + \boldsymbol{\gamma_j}\,\boldsymbol{x'_{jit}} + u_{it}$
C: Conditional growth	$g_{it} = \boldsymbol{\alpha} + \mu\,h_{it} + \delta\,z_{it} + \omega\,h_{it}\,z_{it} + \boldsymbol{\gamma_j}\,\boldsymbol{x'_{jit}} + u_{it}$

Variable	Definition	Variable	Definition
i	index for countries	$s_{it},\ i_{it}$	rate of savings/investments (of GNP/GNI)
t	index for time period (of 3-10 years)	g_{it}	real growth rate
j	index for control variables	h_{it}	aid share (of GNP/GNI)
a	constant, may be divided into	z_{it}	conditional variable
$\boldsymbol{\alpha} = (\alpha_i,\ \alpha_t)$	fixed effects for countries and years	$\boldsymbol{x_{jit}}$	vector of j control variables
$\mu,\ \delta,\ \omega,\ \gamma,$	coefficients to be estimated	u_{it}	residuals

Note: Most of the early models were pure cross-country models, i.e., had no time index, and some models are estimated for one country only, so they have no country index.

Figure 3. Production over time of papers in the AEL
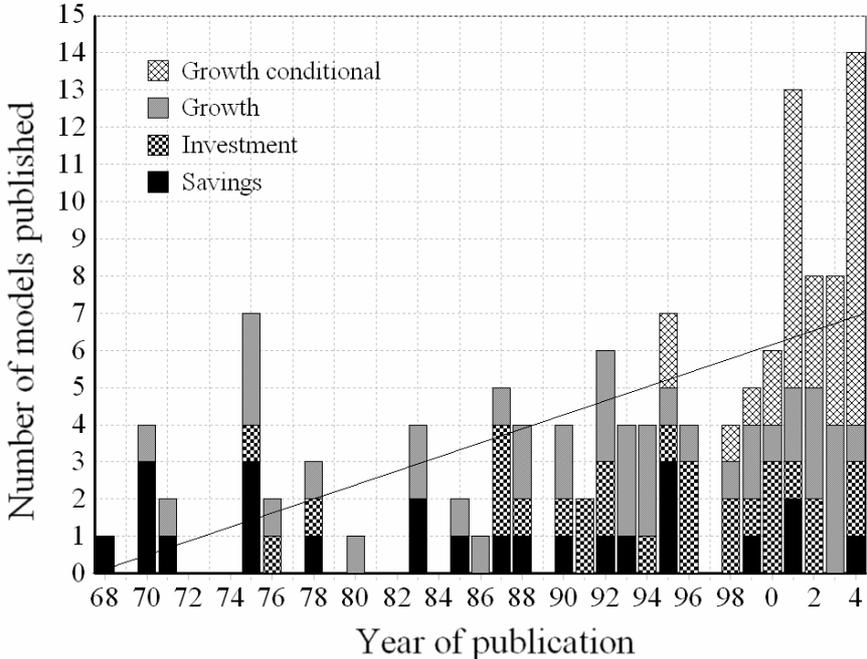
Note: The line included is a linear trend-line through the number of models published. It has a significant slope, but it exaggerates the slope, as the last 5 years include some working papers which may or may not be published later, while no working papers are included in the first 30 years.

3.1 The process of publication over time

Figure 3 shows the development over time in the production of the models of the AEL. It has a significantly rising trend. It started with a wave of type A models – first savings models and then gradually investment models. Then came the larger wave of type B models, and finally, since 1995, C type models have emerged. Papers in the later waves often contain estimates of models of one or even two of the previous families. Type C papers nearly always give results of type B as well. The present wave of papers is still on the upswing, so we are likely to see many more papers in the field.

Table 4. Statistics of the AEL

Regressions	A: Accumulation		B: Growth	C: Conditional			Proxy	Sum
	Savings	Investments		Good Policy	Medicine	Others		
Best-set	21	37	68	23	15	10	8	182
All-set	61	122	543	232	123	23	29	1,113
Sample size	1,890	3,872	11,312	5,834	4,681	663	2,264	30,516

Note: *Proxy* studies are done using data – such as capital inflows – instead of aid, but nevertheless they draw conclusions regarding aid. This was often done in the early papers where few aid data existed. *Best-set* is the regression preferred by the author of the paper; *All-set* is all regressions reported.

3.2 Statistics of papers and estimates: Dimensions of the mining done:

Table 4 gives the dimensions of the data set contained in the 97 papers. They hold 182 models of the 3 families. Thus, the average paper contains models from 1.88 families. The papers publish 1,113 regressions – that is 11 per paper. In the meta-analyses, two data sets are used:

The best-set: Each model provides one data point, the empirical result preferred by the author.
The all-set: Each empirical result – that is, each regression – is taken as a data point.

For the AEL, the best-set gives 182 data points, while the all-set provides 1,113 data points; but they have to be divided into groups as they examine the effect of aid on savings, on investment and on growth as well as on growth conditional. Half of the estimates (543) are aid-growth effects, μ, reached from estimates of the equation $g_{it} = \alpha + \mu h_{it} + \gamma_j x'_{jit} + u_{it}$. As this is the largest group of estimates of one effect, we shall use it as the key example below.

The numbers of estimates made relative to the numbers of data-points provide a measure of the data mining done. At present, almost 5,000 annual observations of, h, the share of aid in GDP, have been published, and for nearly all of these the corresponding, g, real growth rates per capita is also available. However, these data are normally averaged over 4-5

years, so we shall take it that the number of observations available is $N \approx 1000$, as shown on Figure 1. But, in fact the highest N reported in any of the 1,113 estimates is 825.

On subsets of these 1,000 observations, a total of 1,113 regressions are published. If we assume that at least 25 estimates have been made for each published regression, the AEL is based on at least 25,000 regressions on 1,000 observations. Another way to get the same result is to note that data of the sum of the samples analyzed is 30,516. That is almost 30 times the 1,000 data available. So we are dealing with a *mining ratio* of 25-30 in the AEL.

It would be a very interesting question to examine other well-defined literatures for their mining ratios: If we take the annual rates of inflation, p, and unemployment, u, in the OECD countries (app $N \approx 50 \ x \ 24 = 1200$) and ask how many Phillips curves, $\pi = \partial p / \partial u$, have been estimated on these data, we will surely get a mining ratio which is much higher.

3.3   *What does a mining ratio of 25-30 mean?*

Data mining is a *common pool problem*, where the pool is the degrees of freedom available. However, it is an unusual common pool in the sense that nothing seems to happen when the pool is exhausted, mining just goes on, and no red flags pop up. The individual researcher does not make a problematic dent in the degrees of freedom available by running, e.g., 250 regressions on 1,000 observations. However, the 104 authors of the AEL constitute a *mining collective*, who has mined the data thoroughly, and a dense net of cross-citations does exist in the field. Everybody has read some of the literature, and has thus joined the mining collective.

Data mining is a process that eats degrees of freedom, but it is forbiddingly difficult to calculate the precise amount of mining done and the resulting loss of degrees of freedom. The first data published are mined by most of the 97 papers, while more recent data are mined by the most recent papers only.

When a coefficient is presented with a t-ratio of e.g. 2.7, it is significant at the 1% level if $N > 50$, if we are considering one regression run on virgin data. Then it is provided with the *** notation. The reader may contemplate what a mining ratio of 25-30 does to such a nice t-test. Are we all involved in a con-game as argued in the classic Leamer (1983)? From the arguments of Leamer is a clear line to the concept of robustness testing, which will be discussed below.

Another way of approaching the problem is to point out that data mining reduces the likelihood of Type I errors (rejection of true model), but it increases the likelihood of Type II errors (acceptance of false model). Combined with the incentive to polish findings, this means that it is likely that some models are reached by polishing a random quirk in a certain data set.

We conclude: The development aid data have been so thoroughly mined that it is *highly likely that some Type II errors have occurred in the papers published*. This is why independent replication of models is essential for their credibility. As more data accumulate, the literature should reveal whether models survive. Also, the techniques of meta studies contain tests developed to detect some aspects of mining, though it is difficult to design fully mining-proof test.

4.   Meta analysis: Priors, interests and biases

The introduction listed 3 questions that meta studies are made to analyze in the case of a literature. One of these questions is: (Q3) Is the process of truth finding *biased*?

At the operational level, biases are generated by misspecification, by the choice of untypical data or the wrong estimator, etc. Such errors are often due to genuine ignorance about the data generating process, and as knowledge is gained they are reduced. However, this is likely to be a slow and stepwise procedure as the signs of coefficients often tends to be "established", so it can become difficult to publish results with a wrong signs.

However, as the acceptance of results by each researcher is the product of a stopping rule for a data search, it may be influenced by priors. If many researchers have the same priors, the whole process of research may be biased.

4.1   Priors commonly detected by meta studies

Table 5 lists the five most common priors. We believe that all researchers know these priors, both from observation and from introspection. However, it is difficult to assess their importance. Meta-analysis is a technique providing quantitative estimates of their importance. Not for the individual study, but for a set of studies.

The first is the prior of researchers and journals for clear results, which may lead to *polishing*: It is easier to publish clear results than wishy-washy ones, and researchers have strong incentives to publish.[10] Also, it is unsatisfactory to finish a project with no conclusion. Meta studies routinely detect polishing and so do we (e.g. Card and Krueger 1995 and Roberts and Stanley 2005). If polishing is *neutral*, it makes some results too large and others too small. Thus, it creates variation, but it may not influence the value to which the research process converges.

Table 5. Five priors analyzed in meta studies – they may also apply to journals

Prior	Prior generator	Bias found in the AEL
	Internal in researchers and journals: Reduced by academic competition	
Polishing	Researchers have to publish to flourish, and journals want clear results	Polishing causes results to be too good
Ideology	Authors may have an ideology that predicts an outcome	Some authors express political-ideological views, and find results in accordance
Goodness	We all want to be seen as good, and the activity analyzed may have a good purpose	To find a negative effect of aid is a critique of goodness. Causes reluctance
Author history	Previous writings of the author and his/her associates causes path dependence	50% of AEL authors participate in more than one paper. Several groups fight for *their* model
	External pressures and interests: Reduced by competing institutions.	
Institutional interests	Authors often work for an institution with an interests in the results	Much of the AEL is financed by the aid industry. Causes reluctance

The clearest *ideological/political factor* is that a handful of the early AEL writers state that they have a Marxist/left-wing ideology, or a Libertarian one. These authors have a prior for finding that aid harms the recipient country, though for different reasons. [11] Left-wingers explain the poverty of LDCs by exploitation by DCs, and see aid as a factor in that process. Libertarians note that aid is given to public sectors in the LDCs, and see it as an inducement for government growth, planning and ultimately for socialism. Authors of the above two persuasions normally report negative aid effectiveness.

Table 6. Some characteristics of the AEL authors

	Participation in			Origin of author	Nr
Papers	Number	Probability of more[a]		DC (OECD country)	73
1	75	No more	50.0 %	Mixed [b]	27
2	17	1 more	22.7 %	LDC	4
3	8	2 more	16.0 %	Financing of research	
4	3	3 more	8.0 %	University	72
5	1	4 more	3.3 %	International organization	17
6+	0	5+ more	0 %	Other aid	12
All	104			Other	3

Note: a. Probability that author appears no more in the AEL, in 1 paper more, etc.
b. Author with non-DC origin now working in DC (mainly the USA).
Another point to note is that only 9 of the 104 authors are female.

4.2   The goodness/interest tangle and the reluctance hypothesis

Most churches advocate aid to the poor of the world. So do idealists such as Bono (Paul Hewson), Jeffrey Sachs, etc. It is something genuinely *good*, in a hard and unequal world. So it is a tragedy if it does not work.[12] Thus, if a search process leads to a handful of significant coefficients to the aid-term, it may be morally better to report the positive ones.

The main institutional interests in the AEL are that those working for/financed by the aid industry may have a prior for showing that aid works. The interests may cause a bias for three reasons: (1) Loyalty within organizations. (2) Career pressures and other pressures of organizations on employees. (3) Selection/self-selection of organizations and employees.

Consequently, goodness and interests give priors that – in the AEL case – both predict the same outcome: Researchers will be reluctant to publish negative results. In this case it is difficult to untangle the nature of the bias found.[13]

However, this gives another moral problem: Imagine that the true effect is about zero, and there is a way to use the aid money better. If, under those assumptions, too many positive results are reported, the urge to search for the better way to eradicate poverty is reduced.

As suggested by table 5 it is not a problem that researchers have biases and interests if there are enough competing biases and interests. This is not the case in the AEL, not only are the internal biases of researchers in one direction but the institutional interests are overwhelmingly in the same direction.

Many donors reserve a small fraction of the aid budgets (of $106 billion in 2005) for development research. Even ½% of $100 billion is still $500 million. The social sciences receive perhaps 1/3, which is large relative to the finances available for research in development. In addition, perhaps as much as $10-15 billion is used for consultants, which include many economists. Consequently, a dense net of links exists from the aid industry to development research and development researchers.[14] In most areas of development research, this is fine, but it is likely to be a problem when researchers, who work for the aid industry, write on aid effectiveness. Here, the reader should be informed about the interests of the researcher.[15]

In the author statistics of Table 6, we are not able to fully identify the institutional interests of all 104 authors, though for the last decade many researchers have created home pages with biographic information, and most papers have a note of acknowledgement. 78 of the researches only give a university affiliation. However, many university researchers do get outside funds, and even if a certain paper did not use such funds the author may have other grants, so it is likely that many of the 78 belong in the next two rows of the table. At least 35% of the AEL researches work for the aid industry, and it is likely to be much more.

Thus, to the extent that funding affects the priors in the AEL research, it will do so in an *asymmetrical* way. The effect of the asymmetry of priors is an empirical question. We are glad to report that although it works in the direction predicted, and it often becomes significant, it is *not* very strong.

As seen from Table 6, half of the 104 authors of the AEL appear in 2-5 papers. Also, many are members of groups. Both individual authors and groups typically keep demonstrating that their model is right, and that the other group's model is wrong. It poses a problem for tests of significance in meta studies that group behavior reduces the degrees of freedom in the tests. This is a problem for the analysis covered in Section 7, but it is less of a problem in Sections 5 and 6 where the number of studies is high, and the results are clear anyhow.

Authors often suspect that priors apply to journals.[16] A journal may have an ideology or a history specializing in a certain type of studies, pushing a particular point of view. Further, some journals receive grants, and may not like to bite the hand that feeds them. We certainly know of authors who have been forced to shorten and sharpen papers by referees and editors. Hence, meta studies normally include both variables for author characteristics and for publication outlet. Normally, some of both types become significant.

4.3 A note on the techniques of meta-analysis

Most techniques in meta-analysis like averages and confidence intervals are fairly standard, but some are unusual: The **MST**, Meta Significance Test tests for the existence of a genuine empirical effect, and for excess significance in small samples. The **FAT**, Funnel Asymmetry Test for the existence of publication bias in the sense of reluctance.[17] They both analyze funnel plots as the one on Figure 4, giving the distribution of estimates of the same coefficient $\mu = \mu(N)$, and its *t*-tests as a function of sample size, *N*. They have been developed to detect 3 properties of the research process in the case of data dependencies:

Convergence: The result should converge to something that differs from zero. With dependent data that has to be tested as a relation between *t*-tests and *N*. This is the logic of the MST, which tests if the literature finds a genuine empirical effect.

Polishing: That is results are made *too* significant. The significance of the results should rise proportionally to ln(*N*), as *N* rises. Polishing is easier the smaller the sample, so polishing is detected if the significance of the published results fall slower than by ln(*N*), when *N* rises. It is a fact of life that people polish their goods to make them as shiny as possible to attract customers. This is tested by the MST.

Figure 4. Funnel plot of the 543 estimates of the aid-growth effect

The trend is the regression:
$$\mu = 0.310 - 0.043 \ln(N)$$
$$(7.1) \quad (-4.7)$$
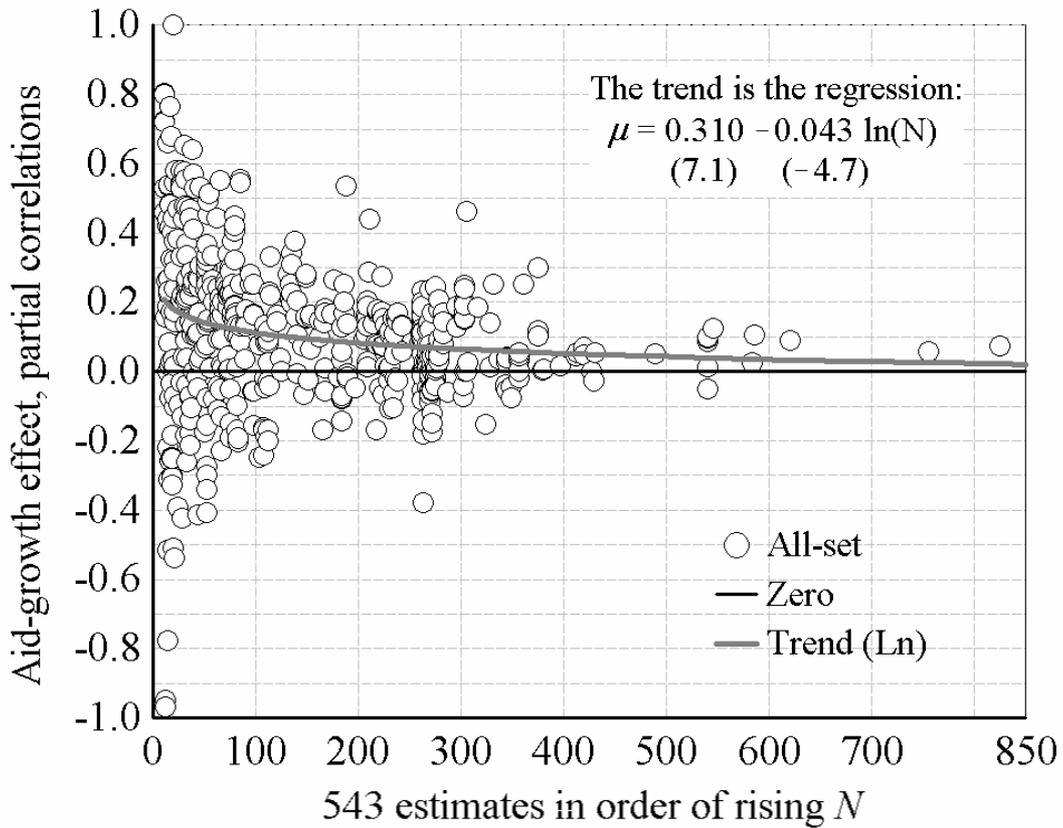
Asymmetries: The distribution of the estimated coefficients of the same effect should be symmetric around a line giving the true value. In our case with learning by doing as per item E2, from Section 2, the symmetry should be around a line with a *positive* slope. An asymmetry in the funnel means that the research process is systematically biased against certain results – e.g., as predicted by the reluctance hypothesis. [18]

The FAT is the formal test for asymmetry. In the AEL the asymmetry is visible to the naked eye as seen from Figure 4: The best log-linear line through the points has a highly significant negative slope. The same result appears if the horizontal axis is *t*. We take this as confirmation of the reluctancy hypothesis.

Finally, we should mention that we studied whether the use of more advanced econometrics had any effect on results by including binary dummies for techniques when trying to explain the pattern in the results (see below). These dummies had no effect. It might be due to the low level of significance throughout, but also we found only one instance where an author noted that results changed when a more advanced method was used. It was a case where the introduction of an econometric refinement changed a coefficient from (just) insignificant to

(just) significant. The advances in econometric techniques are not what changed the results, rather, the increase of data available did. This story appears to generalize.

The next three sections summarize our studies, following the classification in Figure 2.

5. Results from Family A: Does aid cause increasing accumulation?

Family A of the AEL started around 1970, when development economists used Harrod-Domar models. They saw accumulation as the crucial factor in growth. The savings rate and subsequently the balance of payments are thus the key constraints for growth (see Chenery and Strout, 1966). Aid was meant to finance accumulation. Hence, it should move both gaps outward, and thus contribute greatly to growth. This family of studies was the subject of the Doucouliagos and Paldam (2006) meta-analysis.

Table 7. Two related challenges to aid

	Marginal activity caused by aid	Origin of challenge
C1	Aid reduces domestic savings by the same amount	Griffin and Enos (1970) and Weisskopf (1972b)
C2	Aid increases public consumption by the same amount	Boone (1996)

Note: While (C2) ➔ (C1) the reverse causality does not hold.

5.1   The first challenge to aid: The savings effect is crowded out

Several papers just before 1970 started a discussion of savings functions in poor countries and the effect of external inflows.[19] As a part of the new savings literature, Griffin and Enos (1970) and Weisskopf (1972b)[20] demonstrated (on the scanty data then available) that aid flows decreased savings in the recipient countries by the same amount. The fungibility of aid permitted the marginal activity generated by aid *not* to be increased accumulation. If the key constraint for growth was accumulation, this was a major challenge to the justification of aid. This challenge corresponds to the one identified by Boone (1996).

The savings challenge led to a wave of studies – see Figure 3 – and this tradition has continued till this day, with many studies of the AEL having a section with regressions analyzing the effect of aid on the rate of savings or investments. As listed in Table 4, a total of 29 studies bring 90 savings regressions, and 37 studies bring 122 investment regressions. The standard savings-investment bookkeeping identity for an open economy looks as follows:

$$(1) \quad I - S = (I_P - S_P) + (I_G - S_G) = -XMB,$$

where *I* is investments, *S* is savings and *XMB* is the surplus on the goods and service balance, and the subscripts $_P$ and $_G$ indicate the private and the (general) government sector.

In this framework aid, *H*, is a device that allows *XMB* to turn negative by *H*. It is, of course, given for that purpose. This will allow investment to rise by the amount of *H*, provided that *S* does not fall. If *S* falls by *H,* the rise in *I* is crowded out. With the normalized variables $(s, i, h) = (S/Y, I/Y, H/Y)$ this gives the possible effects listed in Table 8.

The challenge of Boone (1996) is that aid leads to an increase in public consumption only. As follows from equation (1), this causes the government savings rate to fall correspondingly, and it is thus one mechanism that explains the challenge of Griffin and Enos.

Table 8. Interpreting possible effects of the aid on savings and the investment

Effectiveness	Super	Full	Some	No	Harmful
Crowding out	Less than none	None	Some	Full	More than full
Savings effect	effect > 0	0	0 < effect < −1	−1	effect < −1
Investment effect	effect > 1	1	1 < effect < 0	0	effect < 0

Note. The effects are expressed in percentage points of shares of the GDP.

5.2   The results: A large but probably not a full crowding out

Figure 5 shows the results. Both pictures show an amazing range of results, and thus give a rather unclear picture. The savings graph shows that there is considerable crowding out. However, there is probably not a full crowding out

Figure 5. The estimated effect of aid on either savings or investments
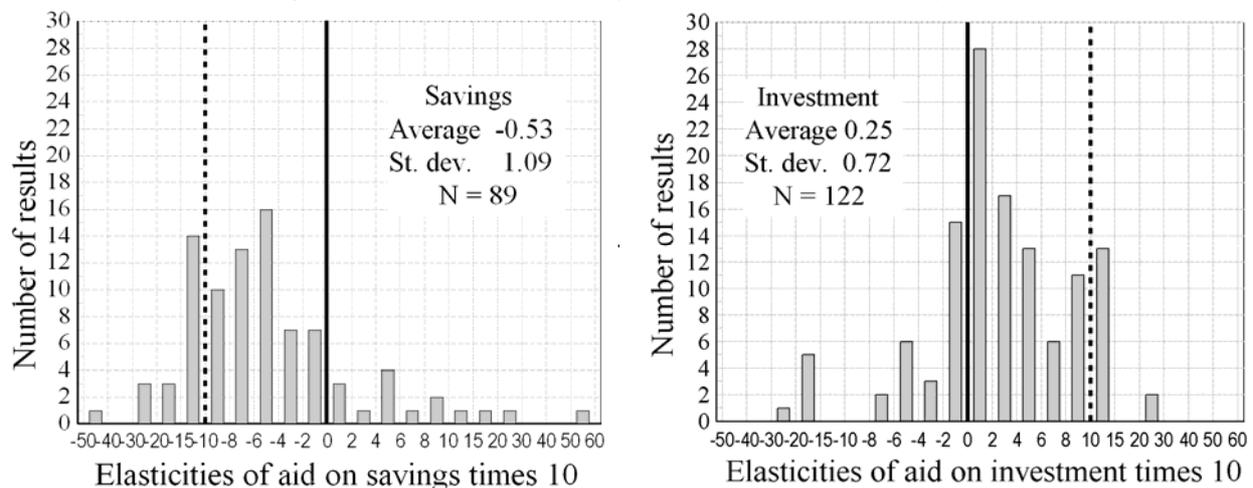
The investment graph is similar but shifted up by 1 as it should. It has its highest peak just above zero, but then there is a secondary peak around 1, so on average there is probably a positive effect, but it is clearly well below 1.

After further analysis, we conclude that aid increases accumulation by about 25% of the aid, and that most of the remaining 75% lead to an increase in public consumption, and hence to a fall in public savings. Both effects are of dubious significance relative to -1 and 0 respectively. However, accumulation is only one explanation for growth, so the total effect on growth depends on what the remaining 75% of the aid do to the economy. As it basically leads to public consumption, it is likely to be a problem because we know that public consumption has a negative effect on growth (see Barro and Sala-i-Martin 2004; pp 525-26). Also, papers in the AEL that include public consumption get a negative coefficient to that control.[21] Hence, from the analysis so far, it is unclear if aid leads to development.

If we consider growth the key goal of aid, then it is better to study the effect directly. This is done in the other families of the AEL, most directly in family B.

6. Results from Family B: Does aid cause increasing growth?

A bit more than half of the estimated aid effects are for the family of reduced form growth models. Here we have no less than 543 published estimates of aid effectiveness, all generated by models that fit the format of line B in Table 3 and repeated as formula (3) below. This family of studies was the subject of the Doucouliagos and Paldam (2008) meta-analysis.

6.1 Growth models: The relation between aid-growth and convergence models

This B-family of models is a subfamily of the large literature on cross-country growth models. That is, the literature summarized in Barro and Sala-i-Martin (2004; chaps. 10-12). The literature started as a study of convergence, using the well researched Barro equation:

(2) $\quad g_{it} = \alpha + \beta \log y_{it} + \gamma_{jit} x'_{jit} + u_{it}$, which holds a convergence term and a control set.

Without the control set no absolute convergence occurs, but with a suitable control set conditional convergence occurs. This equation was then amended by replacing (or supplementing) the convergence term, $\beta \ln y_{it}$, with the aid effectiveness term, μh_{it}

(3) $\quad g_{it} = \alpha + \mu h_{it} + \gamma_{jit} x'_{jit} + u_{it}$, which holds an aid effectiveness term and a control set.[22]

Obviously, very much can be said about (2) and (3), but at present we shall take the basic set-up as given, and just concentrate on the results, as summarized by the meta-analyses.

The family of Barro-type growth regressions is characterized by having a huge choice set for the x's. Theory gives only vague guidance in the choice. For each choice of x-set an estimate of either β or μ is generated, by equation (2) and (3) respectively.

The voluminous literature on Barro growth empirics has now tried about 400 controls, and of these about 60 has been tried in the AEL. The typical number of controls is 5. This gives $\binom{60}{5} \approx 5.5 \cdot 10^6$ or possible models to experiment with.[23] Even if the true value of μ is zero, 5% of these estimates will be significant at the 5% level, and of these, half will be positive. So it is crucial to test results for robustness, and in the end we are back to the demand for independent replication. The AEL contains some discussions of robustness in the papers from 1995 onward, but in our assessment it is still not a very prominent subject.

Figures 4 and 6 show the remarkably broad range of the results. We can read these figures as a monument to the ingenuity of our profession. From a set of raw data showing nothing (see Figures 1a and b), it has proved possible to generate the distribution of results shown. The 543 observations have a positive average, but it is small relative to the variation, so it is not surprising that it proves insignificant in the MST tests that adjust for dependencies between the results.

6.3 The development over time, reluctance, and summary

Figure 4 above and the corresponding graph with *t* at the horizontal axis (see Section 4.4 and Doucouliagos and Paldam 2008) show 4 points:

(C1) The variation is falling over time, as it should when the sample size grows.

(C2) The best-set is typically chosen among the more extreme points. So a relatively larger fraction of the best-set estimates is outside the "range of reason" than of the all-set. The FAT shows that there are too many significant results in small samples.

(C3) The average result is steadily decreasing. It is now +0.02 on Figure 4 and it does converge to zero using the appropriate MST test.

(C4) The funnel (on Figures 4) is not symmetrical around a horizontal or a rising average line. The asymmetry is confirmed by the FAT.

Items (C3) and (C4) show that the AEL is asymmetrical in a way that confirms the reluctancy hypothesis from Section 4.2. The majority of the authors of the AEL are (understandably) reluctant to publish negative results.

It is worth noting that if the same graph had been presented for the first 250 estimates only, we would have reported a trend line $\mu(t) = 0.174$ (4.8) $- 0.00008$ (0.3)t. Thus, from that figure, we would have had to conclude that aid is amazingly effective, and has no signs of asymmetry. Now, in the full perspective of all studies we have to draw a different conclusion. Clearly, small samples give results that may very well be misleading.

This gives three facts to consider when to assess the size of the true value of μ:

(F1) The trends of the results on the figures have reached values of 0.02 to 0.04, which is of small economic significance and is not statistically significant different from zero.

(F2) They have not converged to something "final", as the average aid effect continues to fall as newer evidence is accumulated.

(F3) There is a certain amount of reluctance to publish negative results, so the process of research is upward biased.

When these 3 facts are put together, it is clear that the AEL has not managed to overcome the zero correlation result. That is, it has not managed to come up with a robust and theoretically well founded control set that turns zero absolute effectiveness into a positive controlled effectiveness.

This brings us to the newest strand in the literature. Several new surveys, notably McGillivray *et al.* (2006), claim that this strand has finally resolved the aid effectiveness controversy to show that aid works. We show that this conclusion is incorrect.

7.   Results from Family C: Is the effect of aid on growth conditional?

The conditionality models accept the zero correlation result, but reject that it means that aid and growth are independent. Conditionality models interpret the result to say that aid helps in some cases, but not in others.[24] This is in line with the impressions of old hands in the aid business: Sometimes aid works and other times it fails. Old hands often claim that they can predict the outcome. This family of studies was the subject of the Doucouliagos and Paldam (2007) meta-analysis.

So perhaps one can find a criterion, *z*, which can be transformed to be symmetrical around zero, so that if *z* > 0 aid works, and if *z* < 0 aid harms. That is, the interacted variable, $h_{it}z_{it}$, gets a significant coefficient, ω, when model (4) is estimated:

(4) $g_{it} = \alpha + \mu h_{it} + \delta z_{it} + \omega z_{it} h_{it} + \gamma_{jit} x_{jit} + u_{it}$

The AEL has pointed to 10 candidates for the role of z during the last decade. Eight are examined in 1 or 2 studies, so they allow no meta study till now, but the other two models are the *Good Policy Model* covered by 22 studies (and 232 regressions), and the *Medicine Model* covered by 15 studies (and 123 regressions).

The models are propagated by two groups, with *institutional homes within the aid industry*. These groups fight for a model that is *liked* by their institution to the extent that it has supported its propagation. The most prolific of the groups is the *World Bank Group* around David Dollar and Paul Collier, which produced seven papers on the Good Policy Model.[25] The second is the *Danida Group* around Finn Tarp and Henrik Hansen,[26] which produced four papers on the Medicine Model.

7.1   Good policy as the criterion for a division of the sample

The Good Policy Model by Burnside and Dollar (1996; 2000) uses a special Good Policy Index as *z*. The index is a weighted sum of the budget surplus, the inflation rate and the trade openness, scaled to be symmetric around a zero mean for the sample of countries and years analyzed. The Good Policy index is outcome-related so it is almost a tautology that the coefficient δ to z_{it} becomes positive and significant when model (3) is estimated. However, it was non-trivial when Burnside and Dollar reported that ω to $h_{it}z_{it}$ became significant and positive. The implication is that aid to countries with good policies helps the country, and help to countries with bad policies harms the country.

How much this message has actually affected World Bank lending since 1995 is not known, but it has probably had an effect, especially since it was propagated in World Bank (1998), and it has certainly received a lot of attention in discussions since then. The model has been vigorously defended by researchers from the *World Bank Group* in no less than 7 papers, but it has been demonstrated in the ensuing literature that it is a fragile model, and when the standard tools of meta-analysis are applied to the 23 papers and 232 regressions, it appears that the key coefficient of the model (that is ω to $h_{it}z_{it}$) is insignificant. In fact the model is

unusually fragile to changes in sample, control variables,[27] etc. When subjected to independent replication it has always failed.

7.2 The Medicine Model: The effect of aid squared

The Medicine Model uses aid itself as the condition, so model (4) reduces to:

$$(5) \quad g_{it} = \alpha + \mu h_{it} + \omega h_{it}^2 + \gamma_{jit} x_{jit} + u_{it}$$

It has two interesting coefficients: μ and ω. The result of the proponents of the model is that $\mu > 0$ and $\omega < 0$. This produces a reverse parable for excess growth with a positive section between $h = 0$ and $h = 2h^*$, and with a maximum for $h = h^*$. The marginal contribution of aid to growth is -2ω.

The aid squared term is propagated by *the Danida group*[28] in four papers, and about 25% of the regressions in its support are actually found in papers of the group. This model is quite robust to reasonable changes in control variables as long as the original (g, h)-data sample is used. However, it fares less well when reestimated on other data sets. The 16 papers and 123 regressions analyzing this model have failed to prove decisively that the two coefficients are statistically different from zero, though the results are just around the level of significance. If the key criterion of independent replication is used, it fails.

In short, the two leading new conditional models may not be anything but the mining of an arbitrary quirk in the data. However, eight more models have been proposed and supported by some empirics. Time will tell if they hold up.

8.   A parallel literature: Resource rents and Dutch Disease

Imagine the following thought experiment: An economist, who knows everything but the AEL, is asked: What piece of economic theory would be your first choice for analyzing the macroeconomic effect of aid? We think this economist would point to the theory on the effects of a transfer of an external rent:

8.1 A theory with three names: The transfer problem, Dutch disease, the resource curse

Development aid is an *external rent* that enters into the domestic economy. Hence, we will have to use the theory that before 1950 was discussed as the *transfer problem*. Since then, it has mainly been discussed in connection with resource rents received from exported resour-

ces. Here, the theory is known as the *Dutch Disease,* or more ominously as the *Resource Curse.*[29] The key result is that while a transfer certainly does increase the income *level* of the recipient, it is "paid for" by a decrease in the *growth* rate, making it less of an advantage in the longer run than it appears at first.[30]

The resource rent received by the LDC world is a couple of times larger than the aid received, and it is even more unequally distributed. The typical natural resource deposit has a long exhaustion time, but resource prices fluctuate to give considerable variation over time; whether aid or resource rents fluctuate more is unknown. Both resource rent and development aid are received primarily by LDC governments, and they are used to finance public spending in much the same way. To the extent that development aid is fungible, it makes virtually no difference if the rent received comes as development aid or as a resource rent. Hence, the models used in the analysis should be similar, but we have found very little exploration of the links between the AEL and the Dutch Disease literature.[31]

8.2 The key role of the real exchange rate: A variable missing in the AEL

The Dutch Disease literature gives the exchange rate a main role by demonstrating that a rent transfer inevitably leads to a real revaluation of the currency of the recipient country. Hereby its international competitiveness is reduced. This causes losses to the economy outside the "booming" aid sector (that is the public sector)[32]. The macro effects of aid are thus less favorable than predicted by the micro-macro result (P2 of the introduction). The Dutch Disease literature predicts a *micro-macro paradox*. The interesting question is consequently not the sign of the Dutch Disease effect of aid, but only the *size* of the effect.

The reader is referred to the case of the unsatisfactory development of the ex-DDR after it became a heavily subsidized part of Germany after the reunification (see Sinn, 2004), and to the case of Greenland (see Paldam, 1997a). In the latter case, the Dutch Disease effect of a 50% aid share for half a century seems to have had a 50% effect on the real exchange rate. Thus, with a 7% aid share in the average LDC, it is possible that the real exchange rate has revalued by a similar amount. It will surely differ a lot from one country to the next, but it is – *a priori* – likely to be a substantial effect.

We note that the LDCs have had rather more inflation than the DCs, and more flexible exchange rates as well. So there is a story waiting to be told about the – small or big – role development aid has played in this development. It is actually documented in a few cases that aid has played an important role in the dynamics of prices and exchange rates, and hereby for the real economy. One well-documented case is that Tanzania was able to keep an unrealisti-

cally low exchange rate due to aid during the first half of the 1980s, with the predictable bad effects on the growth rate, till aid was temporarily stopped (see e.g. Paldam, 1997b).

This story is ruled out in the AEL by the way the models are set up. First it used the real 2-gap models (see Section 3), where the exchange rate is fixed and inflation has no effect by definition. Then it continued in the Barro tradition, which was developed from growth theory, which is real, though inflation is often used as an exogenous control variable.

9.   Conclusion

Development assistance began in earnest in the 1960s. It was met shortly after with academic and policy interest and the evaluation of aid as a vehicle for poverty reduction continues today. The AEL is huge and important. Recently, three meta-analysis have reviewed the accumulated findings from this literature – Doucouliagos and Paldam (2006, 2007 and (2008). They reached the sad result that aid has failed in its primary task. Our aim in this paper was to survey the AEL and patterns in this literature and review the findings from the three meta-analyses. Morover, we explored some aspects of the process through which research is conducted in the AEL.

The discussion above has operated on two levels, and so will the conclusion. We first look at the results as a case study of a research process, and then at the results found as regards aid effectiveness.

A careful analysis of the AEL reveals a highly significant *reluctancy bias*. If the typical AEL researcher finds a handful of significant results he/she will normally choose to present one of the most positive ones as the key result of the study. Considering the issues, this is not surprising, but it is a problem for truth finding/revelation.

When this result is combined with the standard finding that results are polished, so that the published results are a bit too good, we have a real possibility that results for a long time may fail to converge to truth. It is remarkable that we had to conclude that the AEL had not proved that aid is effective, even when 74% of the published aid-growth effects are positive.

The reader may ask how common asymmetries are in published results in empirical macro theory. Obviously the AEL deals with a highly emotional subject and is special in having a large asymmetry in the pressure of interests on the research. It is also unusual that the emotions and interests go hand in hand. However, it is not unusual that meta studies detect asymmetries on the funnel plots of all estimates of a certain result. It should be noted that

while the asymmetries are clear in this literature, they are actually lower than what has been found in some other literatures, especially water price elasticity (Stanly 2005).

Meta studies are a very useful tool providing a perspective on a literature estimating a certain effect and to allowing us to assess if it is converging.

We have demonstrated that the AEL has not managed to show that there is a significantly positive effect of aid. Consequently, if there is an effect, it must be small.

Development aid is consequently an activity that has proved difficult to do right. When something is difficult, it is of paramount importance that it is transparent, i.e., that it is done by simple, clear and easily controllable rules.

However, aid is surrounded by complex politics. In order to attract popular support in donor countries, it caters to all kinds of lofty and continuously shifting goals mixed up with stakeholder and strategic interests. In the aid discourse, the air is often stale and muggy from all the big, sweet and vague words that steadily shift. Even if it would cost some support in the short run, *it would prevent aid fatigue in the longer run if aid became more effective, and it could be found to work.*

References:

Aghion, P., Durlauf, S., eds., 2005. *Handbook of Economic Growth.* North-Holland, Amsterdam.

Ahmed, N., 1971. A note on the Haavelmo hypothesis. *Review of Economics and Statistics* 53, 413-14.

Anderson, S., Boettke, P., 2004. Intellectual tyranny of the status quo. The development set: The character of the Journal of Development Economics 2002. *Econ Journal Watch* 1, 306-18.

Barro, R.J., Sala-i-Martin, X., 1998, 2004. *Economic growth.* MIT Press, Boston.

Bauer, P.T., 1971. *Dissent on development: Studies and debates in development economics.* Weidenfeld and Nicolson, London. Sums up several older papers.

Boone, P., 1996. Politics and the effectiveness of foreign aid. *European Economic Review* 40, 289-329.

Burnside, C., Dollar, D., 2000. Aid, policies and growth. *American Economic Review* 90, 847-68. Working paper available since 1996.

Cassen, R., and associates. 1986, 1994. *Does aid work?* Oxford UP, Oxford, UK.

Chang, C. C., Fernandez-Arias, E., Serven, L., 1998. *Measuring aid flows: A new approach.* World Bank. URL: <http://www.worldbank.org/research/growth/ddaid.htm>

Chenery, H.B., Strout, A.M., 1966. Foreign assistance and economic development. *American Economic Review* 56, 679-733.

Corden, W.M., 1984. Booming sector and Dutch disease economics: review and consolidation. *Oxford Economic Papers* 35, 359-80.

Dewald, W.G., Thursby, J.G., Anderson, R.G. 1986. Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. *American Economic Review* 76, 587-603.

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

Doucouliagos, H., Paldam, M., 2007. Conditional aid effectiveness. A meta study.[32]

Doucouliagos, H., Paldam, M., 2008. Aid effectiveness on growth. A meta study. *European Journal of Political Economy, forthcoming.*[33]

Egger, M., Smith, G.D., Scheider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 316, 629-34.

Elbadawi, I.A., 1999. External aid: Help or hindrance to export orientation in Africa. *Journal of African Economics* 8, 578-616.

Friedman, M., 1958. Foreign economic aid: Means and objectives. *Yale Review* 47, 500-16.

Griffin, K., 1970. Foreign capital, domestic savings and economic development. *Bulletin of the Oxford University Institute of Economics and Statistics* 32, 99-112.

Griffin, K., Enos, J., 1970. Foreign assistance: Objectives and consequences. *Economic Development and Cultural Change* 18, 313-27.

Gylfason, T., Herbertson, T.T., Zoega, G., 1999. A mixed blessing: Natural resources. *Macroeconomic Dynamics* 3, 204-25.

Haavelmo, T., 1965. Comment to W. Leontieff. The rates of long-run economic growth and capital transfers from developed to underdeveloped areas. Section 13, 1039-1068 in *The Econonometric Approach to Development Planning.* Pontificia Academiae Scintiarum Scripta Aaria 28. North-Holland Publishing, Amsterdam.

Hadjimichael, M.T., Ghura, D., Mühleisen, M., Nord, R., Ucer, E.M., 1995. Sub-Saharan Africa: Growth, savings, and investment, 1986-93. IMF Occasional Paper, No. 118.

Herbertsson, T.T., Paldam, M., 2007. Does development aid help poor countries catch up? An analysis of the basic relations[19] *Accepted by the Nationaløkonomisk Tidsskrift/Danish Economic Journal.*

Hughes, H., 2003. Trade or aid? Which benefits developing countries more? *Economic Papers, 22,* 1-19.

Klein, D.B., DiCola, T., 2004. Investigating the apparatus. Institutional ties of the Journal of Development Economics authors and editors. *Econ Journal Watch* 1, 319-30.

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

Lensink, R., White, H., 2001. Are there negative returns to aid? *Journal of Development Studies* 37, 42-65.

McCullough, B.D., McGeary, K.A., Harrison, T.D. 2006. Do Economics Journal Archives Promote Replicable Research? Drexel University - Department of Economics & International Business Working Paper.

McGillivray, M., Feeny, S., Hermes, N., Lensink, R., 2006. Controversies over the impact of development aid: It works; it doesn't; it can, but that depends … *Journal of International Development* 18, 1031-50.

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

Mosley, P., 1987. *Foreign aid: Its defence and reform.* University of Kentucky Press, Lexington.

Mosley, P., 1987. *Overseas Aid. Its Defence and Reform.* Wheatsheaf Books, Brighton UK.

Paldam, M., 1997a. Dutch disease and rent seeking: The Greenland model. *European Journal of Political Economy* 13, 591-614.

Paldam, M., 1997b. *Dansk u-landshjælp. Altruismens politiske økonomi.* Aarhus UP, Aarhus.

Paldam, M., 2005. Appendix. Available from http://www.martin.paldam.dk.

Rahman, A., 1968. Foreign capital and domestic savings: A test of Haavelmo's hypothesis with cross-country data. *Review of Economics and Statistics* 50, 137-38.

Rajan, R.G., Subramanian, A., 2005. Aid and Growth: What Does the Cross-Country Evidence Really Show? IMF WP No. 05/127. June (see also IMF WP No 05/126).

Sachs, J., Warner, A.M., 1995. *Natural resource abundance and economic growth.* NBER Working Paper 5398. Published in several versions.

Sinn, H.W., 2004. *Ist Deutschland noch zu retten?* Econ-Verlag, Berlin.

Stanley, T.D., 2001. Wheat From Chaff: meta-analysis as quantitative literature review. *The Journal of Economic Perspectives* 15, 131-50.

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

Stanley, T.D., Jarrell, S.B., 1989. Meta-Regression Analysis: A Quantitative Method of Literature Surveys. *Journal of Economic Surveys 3,* 161-70.

Stanley, T.D., 2008 Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistic*s, forthcoming. Also, School Working Paper, Economics Series 2006/21, Deakin University.

Tarp, F., Hjertholm, P., eds. 2000. *Foreign aid and development. Lessons learnt and directions for the future.* Routledge Studies in Development Economics: London.

WDI, annual. *World development indicators.* CD from the World Bank, Washington DC.

Weisskopf, T.E., 1972a. An econometric test of alternative constraints on the growth of underdeveloped

countries. *Review of Economics and Statistics* 54, 67-78.

Weisskopf, T.E., 1972b. The impact of foreign capital inflow on domestic savings in underdeveloped countries. *Journal of International Economics* 2, 25-38.

World Bank, 1998. *Assessing aid: What works, what doesn't, and why.* A Policy Research Report. Oxford UP, New York.

Younger, S.D., 1992. Aid and the Dutch Disease: Macroeconomic management when everybody loves you. *World Development* 20, 1587-97.

Endnotes

1. The average African country had falling GDP per capita from 1980 to 1995, and received 16% of GDP in aid.

2. Readers should refer to the original meta studies for details.

3. The July 2005 issue of the *Journal of Economic Surveys* was devoted to the use of meta-regression analysis as a tool for detecting and correcting publication bias in empirical economics (see Stanley 2005).

4. Cases have been reported where even *dependent* replication – that is, replication by researcher B of the results of researcher A on the *same* data – has failed (see Dewald, Thursby and Anderson, 1986, and McCullough, McGeary and Harrison, 2006). Fortunately, this is not a problem we have encountered in the AEL.

5. Since then, at least 10 studies have appeared. They have not changed the result as far as we can tell. The master list of the AEL is posted at http://www.martin.paldam.dk.

6. The terminology *absolute and controlled effectiveness* is parallel to the terminology in the growth literature, which speaks about *absolute and conditional convergence*. In the AEL, the term *convergence* is used to designate models that contain a second order aid term.

7. In the convergence literature (mentioned in note 8), absolute convergence is rejected, but several control sets exist that are sure to turn the rejection into acceptance of conditional convergence. One set is fixed effects for countries, and another set is the Barro-set. Neither of these sets has a similar effect in aid effectiveness relations.

8. There are indeed several such micro-macro paradoxes. For example, there is the exchange rate disconnect puzzle where high volatility in real exchange rates is not related to fluctuations at the macroeconomic level, even though the business sector claims that there are real effects at the microeconomic level.

9. The missing observations for 30% of the countries cause some uncertainty, so the percentage is 2½ ± ½. For ease of calculations we have set it at 2½%.

10. A paper is harder to sell if the key coefficient α (t) = 0.25 (1.69), is than if it is 0.45 (2.57). Maybe, from looking at the residuals, we can discover why economic theory suggests that a couple of observations are omitted; or a variable has to be squared. Econometrics has many tools that can persuade data to confess.

11. The Marxist/left are a heterogeneous group with few economists among the *gurus*. However, the main *gurus* of the Libertarian school are economists. Some of those – notably Friedman (1958) and Bauer (1971) – have written very clearly about development aid.

12. It makes people happy to believe that they are working for mankind at the same time as they serve their own interests. In this sense, it can be argued that the finding of positive aid effectiveness provides happiness to the personnel of the aid industry.

13. Reluctancy pertains to two old considerations: One deals with giving *the benefits of doubt* to the good result that aid works. The second deals with the distinction between *telling the truth and telling the whole truth*. If research is *pure* truth finding both considerations are irrelevant excuses. However, researchers are human.

14. A recent study of one volume of the *Journal of Development Economics* shows that virtually all authors are associated with development agencies (see Klein and DiCola, 2004). In the same issue, Anderson and Boettke (2004) present some reflections about the effects of this dominance.

15. At this point we should declare that one of the authors of this paper has worked as a consultant to the World Bank.

16. For example, the hypothetical *Journal of Women's Liberation Studies* may accept a paper showing that women are *not* discriminated in a certain field, but the paper must then be of a very high quality, while it accepts papers of a more dubious quality if they show that women are harmed by discrimination.

17. The tests are developed in Card and Kreuger (1995), Egger *et al.* (1997) and Stanley (2001, 2005, 2008).

18. The test results are documented in some detail in Doucouliagos and Paldam (2008).

19. The discussion was started by a remark in Haavelmo (1965) about the savings function in LDCs, which led to a discussion, see Rahman (1968); Ahmed (1971); Griffin (1970); Weiskopf (1972a).

20. The challenge was supported by arguments building on the "imperialism" school of thought, and notably Weisskopf was a supporter of the (then) New Left, and saw his results as a confirmation of his views.

21. If public consumption is included among the controls, then – to calculate the effect of aid – one has to correct the aid effectiveness for the effect of aid on public consumption times the effect of public consumption on growth. Without that correction the estimate will exaggerate the effect of aid, as indeed we find that it does.

22. The same model – formally speaking – can equally well be justified from the Harrod-Domar framework. Often, it is just presented as the obvious tool for analyzing aid effectiveness, or the paper simply takes off from other papers in the field and argues that the new model presented is an improvement of the models cited.

23. Barro and Sala-i-Martin (2004; 543-59) describe an experiment where a sample of 85 million model variants were estimated to study the robustness of the coefficients.

24. Let us term growth above average *head* and growth below average *tail*. Imagine that the data for aid and growth are fully independent as suggested by the no frills regressions of Table 1, then in about 50% of the flips, the coin lands on head and in the remaining cases on tail. If you examine 100 conditions that changed during the

day you made a series of flips, you will surely find some connections, but they will not survive an independent replication.

25. This group always makes it clear that they are employees of an aid agency, and they have published Word Bank (1998) advocating their model. Three more papers have been produced by "renegade" members of the group who have left the World Bank and now refute the group's model.

26. Danida is the Danish Aid Agency. The group is the core of DERG (the Development Economics Research Group) at Copenhagen University which is/was financed by the Danida Research Fund. The coordinator of DERG is F. Tarp, who holds a special chair in Development financed by Danida, which also uses most of the members as consultants. The model was propagated by a Danida grant to Tarp and Hjertholm (2000), etc.

27. The control set of the model is: (i) Initial gdp, (ii) Institutional quality, (iii) Sub-Saharan Africa, (iv) East Asia. (v) Ethnic fractionalization, (vi) Assassinations, (vii) Ethnic fractionalization x Assassinations and (viii) M2/GDP lagged. The first 4 of these controls are crucial for the result.

28. The model was discovered by Hadjimichael, Ghura, Mühleisen, Nord and Ucer (1995) in a paper on Africa. It has also been advocated by Lensink and White (2001).

29. Some references are Corden (1984), Sachs and Warner (1995) and Gylfason *et al.* (1999).

30. This is a frequent theme in the aid versus trade literature, where the trade generates dynamism and efficiency in the economy, while aid is a rent with negative effects, see Hughes (2003).

31. Two exceptions are Younger (1992) and Eldabawi (1999) – two papers that have remained largely uncited in the AEL. See also Rajan and Subramanian (2005).

32. As mentioned in Section 5, about 75% of the marginal effect of aid is the expansion of public consumption; most of the rest is public investment.

33. Working paper from Department of Economics, Aarhus University, or from http://www.martin.paldam.dk.