doi: 10.1111/j.1467-6419.2008.00568.x

THE AID EFFECTIVENESS LITERATURE: THE SAD RESULTS OF 40 YEARS OF RESEARCH

Hristos Doucouliagos

Deakin University

Martin Paldam

University of Aarhus

Abstract. The aid effectiveness literature (AEL) consists of empirical macroeconomic estimates of the effects of development aid. By the end of 2004, it comprised 97 econometric studies of three families of related effects. Each family has been analyzed in a separate meta-analysis. The AEL is an ideal subject for meta-analysis as it uses only a few formally similar models to estimate the same underlying effects. It is also an area with strong beliefs, often generated by altruism. When this whole literature is examined, a clear pattern emerges. After 40 years of development aid, the preponderance of the evidence indicates that aid has not been effective. We show that the distribution of results is significantly asymmetric reflecting the reluctance of the research community to publish negative results. The Dutch disease effect on exchange rates provides a plausible explanation for the observed aid ineffectiveness.

Keywords. Accumulation; Aid effectiveness; Growth; Meta-study

1. Introduction: Analyzing the Process of Research in Aid Effectiveness

Poverty among the lesser developed countries (LDCs) is one of the largest problems facing mankind. Many in developed countries (DCs) want to do something to reduce the problem and wish to increase development aid. All DCs give some development aid. This raises the hope of a better world. Unfortunately, the economics of development aid has always been surrounded by doubts.

The fast growth of China and India causes mass poverty to fall more than ever before, but these two giants hardly receive aid. In contrast, poverty is falling much more slowly, if at all, in the main aid recipient countries.[1] Also, it is widely known in the field that the correlation between aid and growth is essentially zero – see Figures 1(a) and 1(b). This mix of casual evidence and the absence of a correlation between aid and growth have caused many to doubt the effectiveness of aid; see, for example, the recent books by Calderisi (2006) and Easterly (2006).

A mixture of hope and doubt is the main reason for the large empirical aid effectiveness literature (AEL). The AEL investigates the effect of aid on growth, savings and investment. To study the effect on growth, a model of growth, g, as a function of h, the aid share, has to be estimated – see equation (3) later. Aid effectiveness means that $\mu = \partial g / \partial h$ is significantly positive.

Three recent meta-analyses of this literature (Doucouliagos and Paldam, 2006, 2007a, 2008) have concluded that aid has been ineffective. In this paper, we wish to step back from the details of these meta-analyses, draw several overarching inferences and offer a more conventional survey of this important area in economics research. In particular, we consider several important dimensions of aid not yet fully explored. 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 suggest that exchange rate movements (i.e. the 'Dutch disease') might explain observed 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. The coming of age of meta-analysis was marked by a Special Issue of the *Journal of Economic Surveys*, which focused on the use of meta-regression analysis as a tool for detecting and correcting publication selection bias in empirical economics (Roberts and Stanley, 2005; Stanley, 2005). The purpose of a meta-study is to perform a quantitative 'forensic' analysis of a research literature, such as the AEL, which estimates the same effect, such as μ. Research is a mixture of (1) *innovation* of theory, models and estimators producing new results, and (2) the *independent replication* by other authors and data seeking to confirm earlier innovations.[2] Through this process, new results are produced, and they gain or lose credibility. All involved hope that the research process converges to some notion of '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 might term 'truth'?
Q2: Can we identify the main innovations, which cause or prevent convergence?
Q3: Are there biases along the way in uncovering 'truth' about aid effectiveness?

Obviously, Q3 requires information about researchers' priors and the incentives of the research process. In the case of the AEL, some of our findings may be embarrassing, or even painful. However, they only confirm the suspicions that researchers routinely vent at the lunch table of *all* research institutes and conferences. We provide formal tests of these common suspicions, some of which are confirmed while others are rejected.

1.2 *Three Perennial Problems*

A key problem in macroeconomics is that it has rather limited data, which are frequently used by many different researchers. In many fields, one may view the research community as a collective constantly fishing in the 'common pool' of available degrees of freedom.

We all know from introspection that when we study an empirical question, we analyze the data till we are satisfied with the result. Reported results are thus the product of a stopping rule. We all want to believe that we stop when we have reached some approximation to the truth. However, what we believe to be the approximate truth is influenced by our priors. Also, interests and institutions influence the stopping rule. It is not a trivial matter whether researchers' incentives and priors are consistent with objective scientific inquiry.

Thus, data mining, priors and incentives create perennial 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 *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 crucial in economics due to data limitations and the difficulties in conducting controlled experiments.

We have found that the literature on aid effectiveness provides an ideal case for meta-analysis. The effects analyzed are well defined and the models are so simple that their differences can be quantified. However, development aid is an emotional issue where priors are strong, and the multi-billion 'aid industry' has it own research agenda.

In the AEL, priors and interests work largely in the same direction. Aid effectiveness is a field where many researchers (and perhaps some journals) are reluctant to publish negative results. This *reluctance hypothesis* may work both at the micro-level through individual researcher's stopping rules and at the macro-level through the publication process. Reluctance creates a truncated distribution of empirical results. This distorts the process of convergence towards 'truth', perhaps impeding it entirely.

1.3 *A Preview of the Results*

A thorough search showed that the AEL as of 1 January 2005 consists of 97 studies.[3] 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 comprehensively what the AEL says. The results vary remarkably, but the aggregate results are disappointing as summarized in Table 1.

The two effects mentioned under family A indicate that aid does not lead to increased investment if it is crowded out by a fall in savings. Hence, the investment effect minus the savings effect should add to 1 when the balance-of-payments effect is added. Aid effectiveness here means that the investment effect is larger than 0, and that the savings effect is larger than -1.

The average effect of aid on investment and growth 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

Table 1. Main Conclusions from our Three Meta-Studies.

Type	Causal link	Conditional on	Conclusion	Significance	Section in this paper
Family A	Aid → investment	n.a.	Small positive	Dubious (from 0)	5
	Aid → savings	n.a.	≈ -0.65	Dubious (from -1)	
Family B	Aid → growth	n.a.	Small positive	No	6
Family C	Conditional	Good policy	Rejected	No	7
	aid → growth	Aid itself (aid squared)	Small positive	Dubious	

Note: The effects mentioned are coefficients to the aid share.

once promising good policy model (discussed in Section 7) when it was submitted to independent replication.

The next section, 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 5–7 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 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) should result in 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 a growth effect between 0.05 and 0.1. The average aid share is about 7.5%, so it follows that aid should contribute somewhere between 0.4 and 0.8 percentage points 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 developing country. This is a substantial share and should be easy to identify.

E2: Both aid agencies and recipients have now had about 40 years of experience, where the process of learning by doing or by trial and error 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, pp. 212–220). Over 40 years, this should cause an increase in aid effectiveness of roughly 50% or more. Such an improved effectiveness 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 as will be explained.

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). WDI covers 156 LDCs, starting from 1960, so in 2000 we should have 6200 observations or 1550 observations when a 4-year average is calculated. About 35% are missing, but this still leaves the 1008 4-year averaged observations shown in Figure 1(a). The figure shows two series averaged to the 10 4-year periods: 1961–1964, 1965–1968, ..., 1997–2000. The series are the real rate of growth, g, measured by GDP per capita and as a percentage; and aid, h, as the percent share of development aid, H, of GDP, $h = H/\text{GDP}$. The model, $g = g(h)$, may or may not be controlled for country heterogeneity, which we term *controlled aid effectiveness* or *absolute aid effectiveness*, respectively.[4]

Figures 1(a) and 1(b) show the raw data, and Table 2 reports 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. For example, it is documented in great detail by Mosley (1987), and recently in Rajan and Subramanian (2008) and Herbertsson and Paldam (2007).[5]

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 1860 pages of the *Handbook of Economic Growth* (Aghion and Durlauf, 2005) mention aid.

2.2 *Four Reasons Why the Absolute Aid Ineffectiveness Result is Puzzling*

R1: The 'why would they' argument. Given standard rationality assumptions, an activity such as aid that has continued for at least 40 years must do at least some of what it seeks. Why else would it continue? Against this 'if it is it must be rational' argument is the fact that the average aid share of the donors is actually quite small (only about 0.3% of donor GDP) and has even decreased a little in the last decade. This decline is often attributed to *aid fatigue*, which is caused by dissatisfaction with this small or nonexistent aid effect. See, for example, the analysis of the discussion of selected donor countries of their aid programs in Hook (1996). Also, it is obvious that while aid optimism was high during the first one to two decades, an industry was created that has the usual stakeholder interests in the continuation

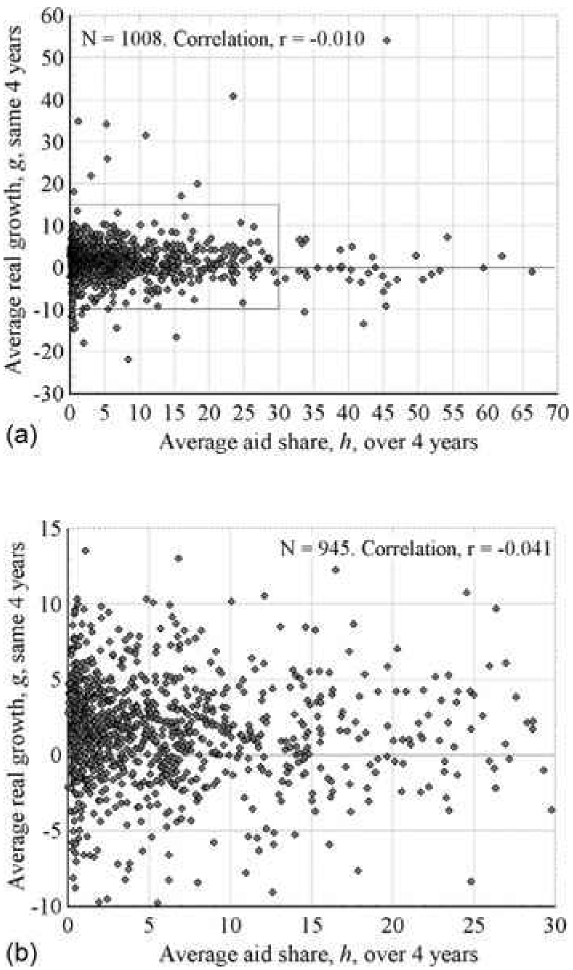


Figure 1. (a) Scatter Plot of Growth and Aid. (b) The Enlarged Box from Figure 1(a).

Notes: (a) The densely packed observations in the 'box' are enlarged in Figure 1(b). (b) An Appendix with similar graphs lagged to both sides is available; see Paldam (2005).

of their activity. Moreover, the aid allocations literature suggests that at least some aid is given for non-humanitarian reasons – including commercial, security and human rights interests.

R2: 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 his findings are uncontroversial. About 50% of all development projects work, and very few of the remaining projects cause harm, even if they fail. Simple

Table 2. Absolute Aid Ineffectiveness: Simple Regressions Between Aid and Growth.

		(1) Lag +1 Growth before aid		(2) Lag 0 Aid and growth		(3) Lag −1 Aid before growth	
Same data as Figure 1		Coeff.	*p*-value	Coeff.	*p*-value	Coeff.	*p*-value
All data	Constant	**1.816**	**0.000**	**1.579**	**0.000**	**1.504**	**0.000**
Figure 1(a)	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**
Figure 1(b)	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: Estimates in bold are significant at the 5% level.

aggregation thus leads to modest average aid effectiveness as discussed in E1 above. The contrast between macro level ineffectiveness and micro level effectiveness is widely known as the micro–macro paradox (Mosley, 1986). Against this view is that fact that aid is – at least partly – fungible: donors have a good chance of selecting projects which would have been implemented by the recipient anyhow. Consequently, the marginal project *caused* by the aid may differ from the project financed by the aid. The marginal project is likely to be less effective than the aid project.

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 that are both in principle and often in practice investments. On the other hand, accumulation is only one factor generating growth, and the marginal activity caused by aid is somewhat different from those activities privately financed. The link from aid to growth does not necessarily proceed via the accumulation effect generated, but may also work through channels such as improved human capital and health.

R4: 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 to this conventional view is a whole set of arguments about why some of the activity effect may not have its full size, but is *crowded out*.

The most extreme form of crowding out is 'Ricardian equivalence', where everything that has to be repaid is crowded out by increased savings. However, development aid contains a gift component, which does not have to be repaid; thus, the positive effect should be proportional to the gift component of the aid (Effective

Development Aid, EDA) and not to total aid (Official Development Aid, ODA), as further discussed below.

Seen together, R1–R4 suggest that aid should help, at least somewhat. But simple statistical evidence (recall Figures 1(a) and 1(b) and Table 2) suggests a total absence of any effect. This, in a nutshell, is the 'aid paradox', and it has driven much research in the AEL. One may interpret much of the research in aid effectiveness as an attempt to overcome this paradox.

2.3 *Some Additional Observations*

First, we note that the available data are ideal for an analysis of aid effectiveness. They are plentiful and have great variation. Aid shares have an average $7^1/_2\%$ of the recipient's GDP. This is substantial relative to other factors that are known to affect growth. Second, the fact that the raw data show no correlation means that any significantly positive (or negative) effect found must be due to the imposition of some more complex structure on the data. That is, results are due to the 'frills' of the analysis, as is further discussed below.

Second, 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 *et al.* (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 adding an EDA dummy to their meta-regression analysis and by converting coefficients to partial correlations that are invariant to the pure shift in scale.

The coefficient on the EDA dummy is *negative* (see Doucouliagos and Paldam, 2008), thus rejecting Ricardian equivalence as defined above. We interpret the finding as further evidence on the typical result in modern political economy that policy decision making is myopic.[6] Decision makers tend to have a short time horizon: they consider the size of the ODA, and largely disregard repayments, which are likely to be the problem of later governments.

Finally, the analysis of aid effectiveness assumes that each country provides equally important information from the point of view of research. The data points from India and Mauritius in Figure 1(a) are thus of the same size. Although the average aid share is $7^1/_2\%$, the aid received by the average citizen in an LDC is much smaller. It is a well-known fact that per capita aid falls with the size of the population (Doucouliagos and Paldam, 2007b). 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 yet nearly 40% of the LDC population. Also, about a third of the aid share data are missing. It is likely that the missing values are below average. If aid shares are weighted by population and the missing observations are assessed,[7] the average share falls by nearly two-thirds, i.e. to $2^1/_2\%$ of GDP.[8] Thus, the cumulated aid over the past 40 years corresponds to one year's worth of income (or GDP per capita) for the average citizen in an LDC.

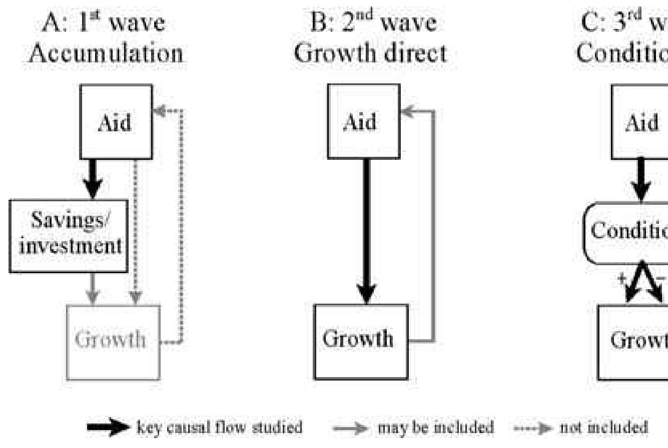


Figure 2. The Three Families of Models in the AEL.

Even after these considerations, absolute aid remains ineffective. This ineffectiveness will still hold even if the aid effectiveness relation is extended to include the standard set of control variables from the growth convergence literature. Typically, the AEL concerns controlled effectiveness. Thus, we turn to the question of whether the problem is, whether a *basic set of control* variables has been found which transfer this overall ineffectiveness result into controlled aid effectiveness in a way that is both robust and justified by economic theory.

3. Structure and Statistics of the AEL

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.

Table 3 outlines the formal structure of the models. Within each family the models have great similarity, as they only differ in three ways: (1) the data sample on which they are estimated, (2) the control set, which can be seen as a choice from a 'master set' of 60 variables, and (3) the exact econometric model employed. These differences are easily coded; thus, this area of research is ideal for meta-analysis.

3.2 *The Process of Publication Over Time*

Figure 3 shows the development over time in the production of the empirical estimates of the AEL. There is a significantly rising trend. It started with a wave of family A models – first savings models and then gradually investment models. Then came the larger wave of family B estimates, and lastly, since 1995, family C research has emerged. Papers in the later waves often contain estimates of the

Table 3. The Models and Variables of the AEL.

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

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)
α	constant, may be divided into	z_{it}	conditional variable
$\boldsymbol{\alpha} = (\alpha_i, \alpha_t)$	fixed effects for countries and years	$\boldsymbol{x_{jit}}$	control variables
		u_{it}	residuals
μ, δ, ω, γ	coefficients to be estimated		

Note: Many of the early models had no time index. Some models have no country index.

previous families. Family C papers nearly always give results of family B as well. The present wave of aid research is still on the rise, so we are likely to see many more papers in the AEL before the wave turns down.

3.3 *Data Mining and Its Discontents*

Table 4 reports some statistics of the 97 AEL papers. They hold 182 models of the three families. Thus, the average paper contains models from 1.88 families. The 97 papers of the AEL contain 1113 regression estimates, or 11 per paper. In our three meta-analyses, two data sets have been used:

the best-set, where each model provides the single empirical result preferred by the author
the all-set, where each reported regression estimate is taken as a data point

For the AEL, the best-set gives 182 data points, while the all-set provides 1113 data points. However, these estimates are further subdivided into groups that measure the effect of aid on savings, on investment, on growth and on growth conditional. Half of the data (543) are aid-growth estimates, μ, from the equation $g_{it} = \boldsymbol{\alpha} + \mu h_{it} + \boldsymbol{\Sigma\gamma_j x_{jit}} + u_{it}$. Because this is the largest group of estimates, it is our main focus below.

The number of estimates reported relative to the number of available aid-growth data points provides a measure of the magnitude of data mining. At present, there

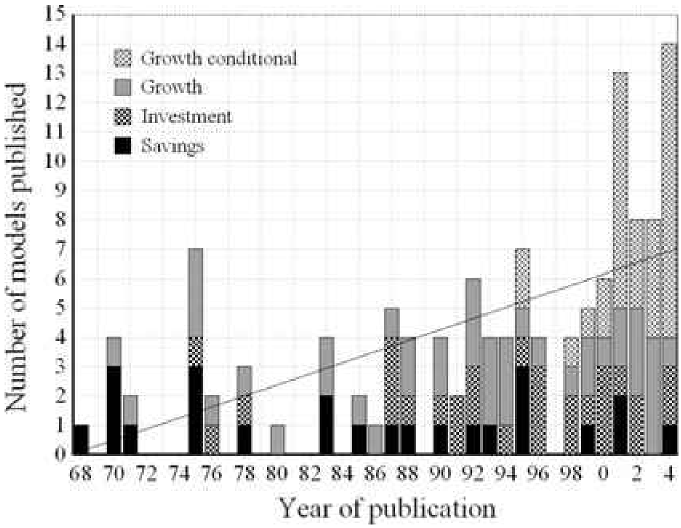


Figure 3. Production Over Time of Papers in the AEL.

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.

exist about 5000 annual observations of h, the share of aid in GDP. For nearly all of these values of h the corresponding real growth rate g per capita is also available. However, these data are usually averaged over 4- to 5-year periods; thus, the number of observations available is $N \approx 1000$, as shown on Figure 1(a). On subsets of these 1000 observations, a total of 1113 regression estimates are published. The highest N reported in any of the 1113 estimates is 825. The sum of

Table 4. Statistics of Reported Estimates in the AEL.

	A: Accumulation			C: Conditional				
Regressions	Savings	Investments	B: Growth	Good policy	Medicine	Others	Proxy	Sum
Best-set	21	37	68	23	16	10	8	182
All-set	61	122	543	232	123	23	29	1113
Sample size	1890	3872	11,312	5834	4681	663	2264	30,516

Note: *Proxy* studies use data, such as capital inflows, but nevertheless draw inferences regarding aid. *Best-set* is the regression estimate preferred by the author of the paper. *All-set* includes all of the reported regression estimates.

the sample sizes used to produce these 1113 regression estimates is 30,516, which is almost 30 times the 1000 data available. Thus, we are dealing with a *mining ratio* of 30, or perhaps much larger, in the AEL.[9]

Data mining is a *common pool problem*, where the pool is the degrees of freedom available. However, unlike the 'tragedy of the commons', overgrazing entails no obvious explicit costs. Even after the common resource, degrees of freedom, is exhausted, mining just goes on unabated. An individual researcher may, by herself, more than exhaust the available degrees of freedom: 'I just ran two million regressions' (Sala-i-Martin, 1997). However, this problem is only compounded when one realizes that all 104 authors of the AEL may be doing the same thing. These 104 authors constitute a *mining collective*, who have collectively mined the data thoroughly, as evidenced by the dense net of cross-citations in the AEL. Everybody has read some of the literature (i.e. they are knowledgeable about the latest mining techniques and where some of the remaining ore may be located) and has opted to join the mining collective.

Data mining is a process that eats degrees of freedom. However, it is impossible to know the precise scale of the mining operation and thereby the exact loss of degrees of freedom. The first data published have been mined by most of the 104 researchers, while more recent data are mined by the most recent papers only.

When a coefficient is presented with a t-ratio of say 2.7, it is significant at the 1% level if we are considering one regression run on virgin data. Such results are decorated with the *** epaulet. Needless to say, a mining ratio of 30 or more entirely invalidates the conventional interpretation of the t-test. Are we all involved in a con-game as suggested by Leamer (1983)? From Leamer's classic paper, there is a clear line to our current obsession for robustness testing.

Another way of approaching the problem is to point out that data mining increases the likelihood of type I errors, in the statistical sense. These type I errors correspond to the acceptance of a false aid-effect model. When combined with the natural tendency to polish one's findings, it is very likely that some published models are merely a random quirk in a certain data subset. The development aid data have been so thoroughly mined that it is highly likely that some type I errors have been published and that many 'significant' aid effects are not in fact statistically significant nor practically relevant. This is why independent replication of regression results is essential for their credibility. As more data accumulate, the literature should reveal whether models actually reflect the underlying phenomena. Can meta-analysis detect data mining and reveal the nuggets that may be buried among the tailing?

4. Meta-Analysis: Priors, Interests and Biases

Is the process of truth finding biased (Q3)? At the operational level, biases are generated by misspecification, by the choice of data, the wrong estimator or the wrong econometric model. Such errors may be due to genuine ignorance about the data-generating process. As experience is gained, some of these errors will probably be reduced. However, the learning process is likely to be slow and discontinuous. Along the way, the signs of coefficients tend to be 'established'; once established, publishing estimates with the 'wrong' signs becomes difficult. Because acceptance of results by each researcher is the product of a stopping rule, it is influenced by priors. If many researchers have similar priors, the entire research enterprise 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 direct observation and from introspection. However, it is difficult to assess their relative importance. Fortunately, meta-analysis provides quantitative estimates of these biases and provides statistical tests of their presence in an area of research.

First is the seeming harmless prior that researchers and journals desire clear results. However, this sensible desire may lead to polishing. It is a fact of life that people polish their goods to make them as shiny as possible to attract customers. It is easier to publish clear, or 'statistically significant', results than ambiguous

Table 5. Priors.

Prior	Source of prior	AEL realization
	Internal motivation potentially 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 hold an ideology that is consistent with a given outcome	Some authors express political–ideological views, and find results in accordance with these views
Goodness	Researchers want to be seen as 'good', and their activity to have a 'good' purpose	To find a negative effect of aid is to question this 'do-good' enterprise; hence the 'reluctance'
Author history	Previous writings of the author and her associates causes path dependence	50% of AEL authors participate in more than one paper. Several groups compete for the preeminence of their model
	External pressures and interests potentially reduced by competing institutions	
Institutional interests	Authors often work for an institution with an interest in the results	Much of the AEL is financed by the aid industry; hence generating 'reluctance'

Note: 'Reluctance' means that the author/journal is reluctant to accept negative results.

ones, and researchers have strong incentives to publish.[10] Also, it is unsatisfactory to finish a research project with no definitive conclusion. Meta-studies routinely detect selection for significance (i.e. 'polishing') and so have we (e.g. Card and Krueger, 1995; Roberts and Stanley, 2005; Stanley, 2005; Doucouliagos and Paldam, 2006, 2007a, 2008). If polishing were neutral, it would increase the number of significantly positive and significantly negative results reported. Thus, it increases variation, but not necessarily biasing the overall magnitude of the aid effect.

In the AEL, some authors have explicitly expressed a Marxist/left-wing ideology, while others have adopted a libertarian one.[11] Therefore, these authors have priors for finding that aid harms the recipient country, though for different reasons. Those on the left explain poverty in LDCs as exploitation by DCs and see aid as a contributing factor in this process. Libertarians note that aid is often given to public sectors in the LDC, and see it as an inducement for government growth, planning and ultimately for socialism. Authors with either of these ideologies usually report negative aid effectiveness.

4.2 *The Goodness/Interest Tangle and the Reluctance Hypothesis*

Most religions advocate giving aid to the poor, as do idealists such as Bono (Paul Hewson) and Jeffrey Sachs, for example. Helping the poor is generally considered to be *good* in a hard and often unequal world. So it is a tragedy if it does not work.[12] Thus, if data mining leads to all types of aid coefficients, it may be seen as morally superior to report the significantly positive ones.

Institutional interests can cause those working for or financed by the aid industry to have priors that aid works. These interests may cause bias for three reasons: (1) loyalty within organizations; (2) career pressures on employees; (3) selection/ self-selection of organizations and employees. In the AEL, 'goodness' and institutional interest might both cause researchers to be reluctant to publish negative results. In this case, it is difficult to untangle the nature of the biases found.

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

As suggested by Table 5, the effects of researchers' biases and interests can be minimized if there are enough competing interests and biases. However, such competing interests are not balanced in the AEL. For example, the studies by Marxists and libertarians may be outnumbered by institutional interests and by the desire to do 'good'.

Many donors reserve a small fraction of the aid budgets (of \$106 billion in 2005) for research. Even if only $^1/_2$% of \$100 billion is set aside it is still \$500 million. In addition, perhaps as much as \$10–15 billion goes to consultants, including economists.[13] Consequently, a dense net of links exists from the aid industry to development researchers.[14] When researchers write on aid effectiveness, the reader should be informed about the interests of the researcher.[15]

Table 6. Some Characteristics of the AEL Authors.

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

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

We are not able to fully identify the institutional interests of all 104 AEL authors (see Table 6). Seventy-two researchers only give a university affiliation. However, many university researchers also receive outside funding. Even if a given research project did not receive outside funding, these researchers may have other grants. In any case, at least 35% of the AEL researchers work for the aid industry, and the true proportion with ties to the aid industry is likely to be much higher.

Thus, to the extent that funding affects the priors of the AEL researchers, it will do so in an asymmetric way. The effect of the asymmetry of priors is an empirical question. We are glad to report that, although it works in the direction of over-reporting positive aid effects, this bias is not very strong in unconditional estimates of aid on growth (see Doucouliagos and Paldam, 2008, for details).

Authors often suspect that priors apply to journals. A journal may have a dominant ideology or a history of specializing in specific areas and thereby of promoting a particular point of view. Further, some journals receive grants, and may not like to embarrass their benefactors. Meta-analysis can include moderator variables for both author characteristics and for the publication outlet. Often, both types of moderator variables are found to be significant.

4.3 *Meta-Analytic Methods*

Generally, meta-analytic methods are conventional statistical methods. However, some are unique to meta-analysis. The meta-significance test (MST) tests for the existence of a genuine empirical effect, as does precision-effect testing (PET) (Stanley, 2005, 2008). The funnel asymmetry test (FAT) identifies the existence of publication selection or ‘reluctance’.[16] These tests examine funnel plots (see Figure 5(a) later) and the distribution of estimates of some coefficient, $\mu = \mu(N)$, and its t-values as a function of sample size, N, or precision, 1/SE. They may be used to detect three properties of the research process.

- *Convergence*: Results should converge to something that differs from zero, if there is a genuine empirical effect.
- *Polishing*: This means that reported results are too significant. 'Polishing' is easier the smaller the sample and the precision. Polishing may be detected when the reported effect increases with its standard error (Stanley, 2005, 2008). This is tested by FAT. Doucouliagos and Paldam (2008) find that reported aid effects do increase with their standard errors ($t = 5.03$; $p < .001$), which indicates polishing.
- *Asymmetries*: The distribution of the estimated coefficients of the same effect should be symmetric around the true value. An asymmetry in the funnel plot means that the research process is systematically biased. In the AEL the research process should favor positive aid effects to be consistent with our 'reluctance' hypothesis (Stanley, 2005, 2008). Again, this is confirmed by FAT (Doucouliagos and Paldam, 2008).

Finally, we should mention that we studied whether the use of more advanced econometrics had any effect on the econometric results, by testing whether the estimation technique explained between study heterogeneity, after controlling for other study characteristics, such as data and specification differences. We find that advances in econometric techniques are not the cause of variation in the results (see Doucouliagos and Paldam, 2008).

5. Does Aid Cause Increasing Accumulation? Results from Family A

Aid effectiveness research of family A 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 (Chenery and Strout, 1966). Aid was meant to finance accumulation. This family of aid research was the subject of the Doucouliagos and Paldam (2006) meta-analysis.

5.1 *The Savings Effect is Crowded Out*

As a part of the new savings literature, Griffin and Enos (1970) and Weisskopf (1972b) demonstrated (on the scanty data then available) that aid flows decreased savings in the recipient countries by the same amount. The fungibility of aid meant that the marginal activity generated by aid did *not* lead to 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 (Figure 3) and this tradition has continued to this day. Many studies still contain a section analyzing the effect of

Table 7. Interpreting Possible Effects of the Aid on Savings and Investment.

Effectiveness	Super	Full		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 GDP.

aid on the rate of savings or investments. As listed in Table 4, a total of 29 studies report 90 aid-savings effects, and 37 studies report 122 aid-investment effects. The standard savings–investment identity for an open economy is

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

where I is investment, S is savings and XMB is the surplus of goods and services, and the subscripts $_P$ and $_G$ indicate the private and the government sector, respectively.

In this framework aid, H, is a device that allows XMB to turn negative by H. This is, of course, why aid is given. Aid allows investment to rise by the amount of H, provided that S does not fall. If S falls by H, the potential rise in I is fully crowded out. With the normalized variables (s, i, h) = (S/Y, I/Y, H/Y), we summarize the possible effects in Table 7.

Boone's (1996) finding is that aid leads to an increase in public consumption only. It follows from equation (1) that this, in turn, causes the government savings rate to fall correspondingly and is one explanation of the puzzle identified by Griffin and Enos (1970).

5.2 *The Results: A Large but Probably not a Full Crowding Out*

Figure 4 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, but probably less than 100%.

The investment graph tells a similar story. It has its highest peak just above zero, but then there is a secondary peak around 1. On average there is 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. Most of the remaining 75% causes an increase in public consumption, and hence a fall in public savings. Both effects are of dubious significance. However, accumulation is only one explanation for growth. The total effect on growth depends on what the remaining 75% of the aid does to the economy. To the extent that it leads to public consumption, it is likely to be unproductive because public consumption is known to have a negative effect on growth (see Barro and Sala-i-Martin, 2004, pp. 525–526). Also, AEL papers that include public

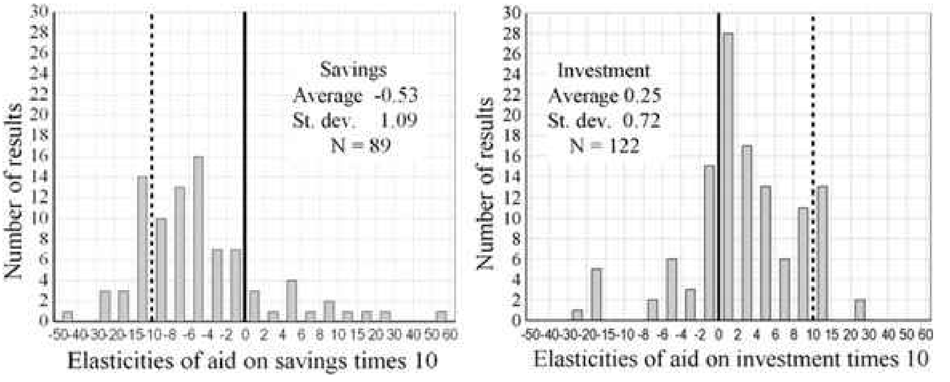


Figure 4. The Estimated Effect of Aid on Either Savings or Investments.

Note: The figures are reproduced from Doucouliagos and Paldam (2006).

consumption in a growth regression tend to find a negative coefficient.[17] From the analysis so far, it is unclear whether aid leads to development.

If we consider growth the key goal of aid, then it is better to study this effect directly. Thus we turn to family B AEL studies.

6. Does Aid Cause Increasing Growth? Results from Family B

As stated before, nearly half of the estimated aid effects (543) are from the family of reduced form growth models. This family of studies was the subject of Doucouliagos and Paldam's (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 (Barro and Sala-i-Martin, 2004, chs. 10–12). This literature started as a study of convergence, using the well-researched Barro equation:

$$g_{it} = \alpha + \beta \ln y_{it} + \sum\nolimits_j \gamma_{jit} x_{jit} + u_{it} \tag{2}$$

with a convergence term and a control set.

Without control variables, no absolute convergence occurs. With a suitable set of controls, 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}

$$g_{it} = \alpha + \mu h_{it} + \sum\nolimits_j \gamma_{jit} x_{jit} + u_{it} \tag{3}$$

and a control set.[18]

The family of Barro-type growth regressions is characterized by having a huge choice set for $\boldsymbol{x}_j$. Theory gives only vague guidance in the choice. For each selected $\boldsymbol{x}_j$ an estimate of either β or μ is generated. The voluminous literature on Barro growth empirics has now tried about 400 control variables, and of these 400 approximately 60 have been tried in the AEL. In contrast the average number of control variables used is approximately 5. This gives $\binom{60}{5} \approx 5.5 \times 10^6$ possible models with which to experiment.[19] Even if the true value of μ is zero, 5% (or more) of these estimates will be significant at the 5% level. Of these, half should be positive. So it is crucial to test results for robustness, which leads us back to the need for independent replication. The AEL contains some discussions of robustness, but it is still a neglected subject.

Figures 4 and 5 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 with zero correlation (see Figures 1(a) and 1(b)), it has proved possible to generate a rich distribution of all types of results. The 543 aid on growth estimates have a small positive average, so it is not surprising that it proves insignificant in explicit meta-regression (Doucouliagos and Paldam, 2008).

6.2 *The Development Over Time, Reluctance and Summary*

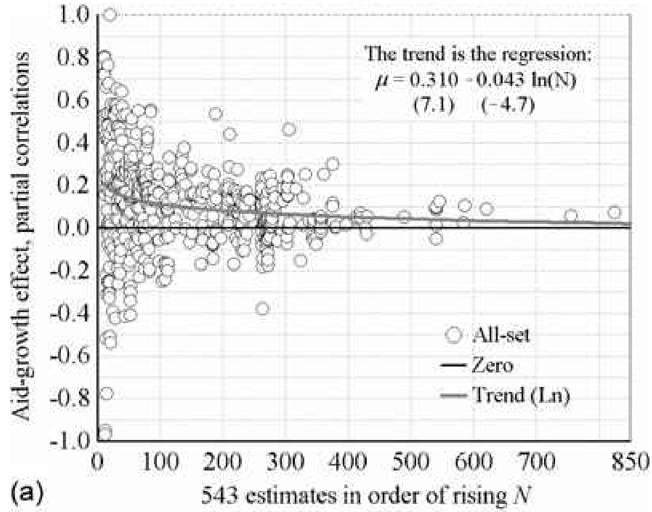
Figures 5(a) and 5(b) show how the 543 comparable estimates of aid effectiveness on growth look when depicted as a function of *N*, the number of observations, and of *t*, time. On each graph a simple regression is included. The two regressions have obvious multicollinearity: a main reason why *N* rises is because as time passes more observations are published.

Table 8 shows that to the extent that we can distinguish between the effects of *t* and *N,* both are significant, but the main effect is from *N.* Thus we conclude:

(C1) The variation is falling over time and with the sample size.
(C2) The best-set is typically chosen among the more extreme points.
(C3) The average result is steadily decreasing (Figure 5(b)). It is now +0.02 on the figures and it seems to converge to zero. PET confirms this small and insignificant aid effect (Doucouliagos and Paldam, 2008, p.11).
(C4) The funnel is not symmetrical around a horizontal or a rising average line (Figure 5(a)). The asymmetry is confirmed by the FAT.

The AEL is asymmetric in a way that confirms the reluctance hypothesis. The majority of the authors of the AEL are (understandably) reluctant to publish negative aid 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 $\mu(t) = 0.174(4.8) - 0.00008$ $(0.3)t$, and 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.



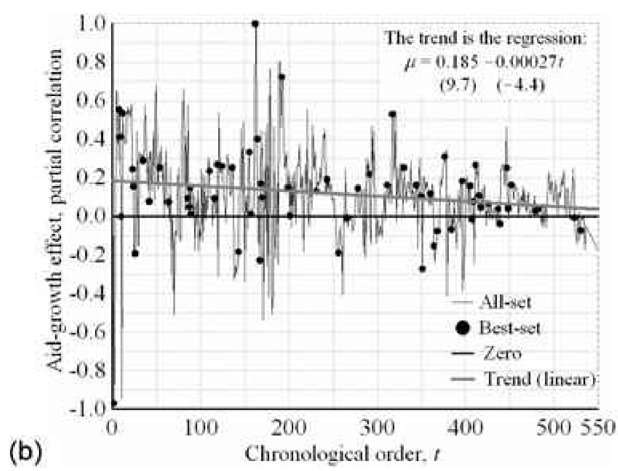


Figure 5. (a) Funnel Plot of the 543 Estimates of the Aid-Growth Effect. (b) Time Series Graph of Aid-Growth Effects: Looking for $\mu(t) \rightarrow \hat{\mu}$.

Note: Figures 5(a) and 5(b) are reproduced from Doucouliagos and Paldam (2008).

This gives three facts to consider when assessing the size of the true value of μ.

(F1) The trends of the results from Figures 5(a) and 5(b) have reached aid-growth effects of 0.02–0.04. These values are of no economic significance and are not significantly different from zero.

(F2) Estimates of the growth-aid effect have yet to converge to anything 'final', and the average aid effect continues to fall as more evidence is accumulated.

Table 8. The Relative Power of N and t in Explaining the Trends of Figures 5(a) and 5(b)

	Constant		Coefficient to ln N		Coefficient to t		Obs.	AR^2
(1)	0.31	(7.1)	−0.043	(−4.7)			543	0.035
(2)	0.19	(9.7)			−0.00027	(−4.4)	543	0.033
(3)	0.28	(5.9)	−0.26	(−2.1)	−0.00015	(−1.9)	543	0.039

Note: (1) is the line shown on Figure 5(a) and (2) is the line on Figure 5(b). The parentheses give t-ratios.

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

The AEL has yet to overcome the zero correlation result. That is, there is no robust and theoretically well-founded set of control variables that turns zero absolute effectiveness into a positive aid-growth effect.

Several new surveys, notably McGillivray *et al.* (2006), claim that family C research has finally resolved the aid effectiveness controversy and show that aid works. Unfortunately, this hopeful conclusion is incorrect.

7. Is the Effect of Aid on Growth Conditional? Results from Family C

One way to interpret the zero correlation result is that it shows that aid helps in some cases, and harms in others. Then it follows that we should look for the condition that results in one outcome or the other. Studies of aid-conditionality, family C, are the subject of a separate meta-analysis (Doucouliagos and Paldam, 2007a).

Aid-conditionality research searches for some criterion, z, scaled to have an average of about zero, where $z > 0$ causes aid to work and $z < 0$ causes aid to harm the economy. That is, the interacted variable, $h_{it}z_{it}$, has a significant positive coefficient, ω, when model (4) is estimated:

$$g_{it} = \alpha + \mu h_{it} + \delta z_{it} + \omega\, z_{it} h_{it} + \sum\nolimits_j \gamma_{jit} x_{jit} + u_{it} \tag{4}$$

The AEL has identified 10 candidates for the role of z over the last decade. Eight candidate mediator variables, z, are examined in only one or two studies; thus, meta-analysis is not possible. But the other two candidates are associated with models that are widely studied in the AEL. These are the *good policy model* covering 23 studies (and 232 estimates) and the *medicine model* covering 16 studies (and 123 estimates).

These two models are promoted by separate groups with institutional homes within the aid industry. The most prolific is the World Bank group, centered around David Dollar and Paul Collier, which has produced seven papers on the good policy model.[20] The second is the Danida group, centered around Finn Tarp and Henrik Hansen,[21] which has published four papers on the medicine model. We have subjected each of these to meta-analysis.

7.1 *Good Policy as the Criterion for a Division of the Sample*

The good policy model by Burnside and Dollar (2000) uses a weighted sum of the budget surplus, the inflation rate and 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 certain that the coefficient δ on z_{it} becomes positive and significant when model (4) is estimated. However, Burnside and Dollar's finding that ω on $h_{it}z_{it}$ is significant and positive is an important discovery. The implication is that aid to countries with these 'good' policies helps the country, and aid to countries with 'bad' policies harms the country's economy.

How much this message has actually affected World Bank lending over the last decade is not known, but it has probably had some effect. The model has been vigorously defended by researchers from the World Bank group in no less than seven papers, but it has also been demonstrated in the ensuing literature that Burnside and Dollar's (1996) positive finding is fragile. When the standard tools of meta-analysis are applied to the 23 papers and 232 conditional-aid effects, it appears that the key coefficient of the model (i.e. ω to $h_{it}z_{it}$) is insignificant. In fact, the estimated coefficient is unusually fragile to changes in sample, control variables, methods and models.[22] Burnside and Dollar's (1996) positive finding does not withstand independent replication.

7.2 *The Medicine Model: The Effect of Aid Squared*

The model was discovered by Hadjimichael *et al.* (1995) in a paper on Africa, but has been most eloquently defended by the Danida group. It has also been advocated by Lensink and White (2001). The medicine model uses aid itself as the condition, so model (4) reduces to

$$g_{it} = \alpha + \mu h_{it} + \omega h_{it}^2 + \sum\nolimits_j \gamma_{jit} x_{jit} + u_{it} \tag{5}$$

It has two important coefficients: μ and ω. Proponents of the model find that $\mu > 0$ and $\omega < 0$. This produces an inverted parabola for excess growth, which has a maximum at $h = h^*$ and a positive section between $h = 0$ and $h = 2h^*$. The *marginal* contribution of aid to growth is $2\omega h < 0$.

The aid-squared term is promoted by the Danida group in four papers. About 25% of its support is found in papers of this group. Equation (5) is robust to reasonable changes in control variables as long as the conventional *g*-*h* data are used. However, the medicine model fares less well when estimated on other data sets. When taken together, the 15 papers and 123 estimates of the medicine model have failed to prove decisively that the two key coefficients are statistically different from zero. Nonetheless, most of the reported coefficient estimates have *p*-values near 0.05. If we demand independent replication, this model also fails to garner support.

In short, the two leading conditional models may prove to be nothing more than the mining of a fortuitous quirk in the data. However, eight other conditional-aid models have been proposed and supported by some empirical research. Only time will tell if they hold up.

8. A Parallel Literature: Resource Rents and Dutch Disease

8.1 *The Dutch Disease, the Resource Curse and the Transfer Problem*

Development aid may be viewed as an external rent that enters into the domestic economy. Before 1950, such issues were studied under the name 'the transfer problem' and once referred to the question of how best to have Germany fulfill World War One reparation payments. Since then, it has mainly been discussed in connection with resource rents received from exporting resources. The associated theory is also known as the 'Dutch disease' or more often as the 'resource curse' (Corden, 1984; Sachs and Warner, 1995; Gylfason *et al.*, 1999). The key result is that while a transfer certainly does increase the *income level* of the recipient, it is 'paid for' by a compensating decrease in the *growth rate*. Thus, aid would be less of an advantage in the longer run than it might first appear.[23]

The total resource rent amount received by LDCs is double that of the development aid received and it is even more unequally distributed. The typical natural resource deposit has a long life, but resource prices fluctuate greatly. Both resource rents 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. Nonetheless, we have found very little exploration of the links between the AEL and the Dutch disease literature.[24]

8.2 *The Key Role of the Real Exchange Rate*

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 (i.e. the public sector).[25] The macro effects of aid are thus less favorable than conventionally predicted. The interesting question is consequently not the sign of the Dutch disease effect of aid, but only the *size* of the effect.

It is difficult to assess the size of the effect as it often comes with a lag. After German unification in 1990, the much poorer East Germany became a heavily subsidized part of Germany. This caused a rapid increase in the relative standard of living, but then after 7–8 years the relative growth of East Germany ceased, and for the last decade the gap has grown slightly (see Sinn, 2004). A similar case is

Greenland, which for half a century has received a Danish grant of 50% of GDP. This has caused a real revaluation of about 50% (see Paldam, 1997a).

Thus, with a 7% aid share in the average LDC, it is possible that the real exchange rate in these countries has revalued by a similar amount. Exchange rate movements will surely differ a lot from one country to the next, but they have the potential to have a substantial effect on aid effectiveness. Unfortunately, the use of exchange rates is ruled out in the AEL by the way the aid-growth relation is modeled.

We note that the LDCs have had more inflation than the DCs and more exchange rate variability as well. So there is a story here waiting to be told about the role development aid has played in exchange rate movements. In a few cases, we know that aid has played an important role in the dynamics of prices and exchange rates, and hereby for the real economy. For example, Tanzania was able to keep an unrealistically low exchange rate due to aid during the first half of the 1980s. However, this had predictably negative effects on the growth rate, until aid was temporarily stopped (Paldam, 1997b).

9. Conclusion

Development assistance began in earnest in the 1960s. Soon, thereafter, it became a widely researched topic, and the evaluation of aid as a vehicle for poverty reduction continues today. The AEL is huge and important. Recently, three meta-analyses have reviewed the accumulated research results from this literature – Doucouliagos and Paldam (2006, 2007a, 2008). Together, these meta-analyses reach the sad result that aid has failed in its primary mission. The purposes of this paper are to survey the AEL, to identify and discuss patterns in this literature and to review the findings from the three meta-analyses. Moreover, we explore the process through which research is conducted in the AEL.

A careful analysis of the AEL reveals a highly significant *reluctance bias*. If an AEL researcher finds several results he/she will be reluctant to report those that suggest that aid causes harm. Rather, the most significantly positive result is likely to be selected as the key finding for a given aid effectiveness study. Considering the issues, this is not surprising, but it is an impediment to uncovering the true effects of aid.

When this tendency is combined with the widespread practice of polishing one's findings so that the published results are clear and statistically significant, research can fail to converge for decades. It is remarkable that our meta-analyses could find no evidence of aid effectiveness, even though 74% of the published aid-growth effects are positive.

The reader may be wondering how common these publication selection biases are in empirical economics in general. Obviously, the AEL deals with a highly emotional subject, where special interests align with everyone's desire to do 'good'. Nonetheless, it is also common for meta-analyses to detect asymmetries in funnel plots, reflecting a similar selection of economic results. Dozens of areas of empirical economics have, by now, been investigated for signs of publication selection, and the majority of these contain evidence of substantial or severe biases (Doucouliagos and Stanley, 2007). Meta-analysis is a very useful tool and provides a unique perspective on empirical literatures. It allows us to assess if the results of a literature are converging, whether results suffer from publication selection, and whether there exists an underlying genuine empirical effect.

Our three meta-analyses of the AEL have failed to find evidence of a significantly positive effect of aid. Consequently, if there is an effect, it must be small. Development aid is an activity that has proved difficult to do right. Even though such a negative conclusion may cost some support for aid in the short run, it may ultimately prevent aid fatigue in the longer run. Our findings underscore the need to find more effective ways to employ development aid.

Acknowledgements

Our cooperation was supported by the Aarhus University Research Foundation. Pia Wichmann Christensen has been a very competent research assistant. This paper benefited from ideas and comments from Peter Sandholt Jensen, Tryggvi Thor Herbertsson, Niels Haldrup and a very careful referee. Useful comments were received from seminar participants at the University of Aarhus, Deakin University, University of Queensland, the Australian National University, Hendrix College and the Kiel Institute of World Economics, as well as from the Public Choice Society 2006 Meetings in New Orleans and Turku, and the Aarhus Colloquium of Meta-Analysis in Economics, Sønderborg, 2007.

Notes

1. The average African country had falling GDP per capita from 1980 to 1995, and received 16% of GDP in aid.
2. Cases have been reported where even *dependent* replication – i.e. replication by researcher B of the results of researcher A on the *same* data – has failed (see Dewald *et al.*, 1986; McCullough *et al.*, 2006). Fortunately, this is *not* a problem we have encountered in the AEL.
3. Since then, at least 10 new studies have appeared. They have not changed the result as far as we can tell. Christensen *et al.* (2007a) provide a comprehensive bibliography of the AEL. A bibliography of the even larger AAL, aid allocation literature, is in Christensen *et al.* (2007b).
4. 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 *conditional* is used to designate models that contain a second-order aid term.
5. In the convergence literature, 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 on aid effectiveness relations.
6. See the survey of the empirical results on mass political economy in Paldam (2004).
7. In a number of cases we know that data are missing because the country received no or very little aid due to a crisis (such as in Zimbabwe lately) making it difficult to get the aid to the recipients.

8. The missing observations for 30% of the countries cause some uncertainty, so the percentage is $2^1/_2\% \pm {}^1/_2\%$. To ease calculations we have set it at $2^1/_2\%$.
9. Another way to think about the mining ratio is to ask: what is the average number of regressions made for each one published? Imagine it is q. Then $M = 1113q$ measures the amount of data mining and $M/1000$ is the mining ratio. We believe that this approach will give a similar mining ratio to the one calculated in the main text.
10. A paper is harder to sell if the key coefficient has a nominal p-value of 0.10 rather than 0.01. Maybe, by 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 main supporter of the 'imperialism' school of thought in this literature is Weisskopf (see his 1972a, b). But also Griffin (1970) and Griffin and Enos (1970) have statements supporting the (then) New Left, and saw their results as a confirmation of his views. It is more common to find statements in support of Friedman (1958) and the various papers reprinted in Bauer (1971).
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, finding evidence of positive aid effectiveness engenders 'happiness' to the personnel of the aid industry.
13. The World Bank has an internal rule of thumb that 10% may be/should be used for consultancy fees. It is likely that it is a fairly general rule.
14. A recent study of one volume of the *Journal of Development Economics* shows that virtually all authors were associated with development agencies (see Klein and DiCola, 2004). In the same issue, Anderson and Boettke (2004) comment on the effects of this dominance.
15. At this point we should declare that one of the authors of this paper, Paldam, has worked as a consultant to the World Bank.
16. These tests are developed in Card and Kreuger (1995), Egger *et al.* (1997) and Stanley (2001, 2005, 2008).
17. 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.
18. The same model, formally speaking, can be justified from the Harrod–Domar framework.
19. Barro and Sala-i-Martin (2004, pp. 543–559) describe an experiment where a sample of 85 million model variants was estimated to study the robustness of the coefficients.
20. This group (which is now widely scattered) always made it clear that they were employees of an aid agency, and they have published World 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; see Easterly (2003), Easterly *et al.* (2004) and Svensson (1999).
21. Danida is the Danish Aid Agency. The group is the core of DERG (the Development Economics Research Group) at Copenhagen University, which is 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).

22. The control set of the model is (1) initial GDP, (2) institutional quality, (3) Sub-Saharan Africa, (4) East Asia, (5) ethnic fractionalization, (6) assassinations, (7) ethnic fractionalization × assassinations and (8) M2/GDP lagged. The first four of these controls are crucial for the result.
23. 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 consequences (Huges, 2003).
24. Two exceptions are Younger (1992) and Eldabawi (1999), but they remain largely ignored by the AEL. See also Rajan and Subramanian (2008).
25. 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.

References

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

Ahmed, N. (1971) A note on the Haavelmo hypothesis. *Review of Economics and Statistics* 53: 413–414.

Anderson, S. and 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: 306–318.

Barro, R.J. and Sala-i-Martin, X. (1998, 2004) *Economic Growth*. Boston, MA: MIT Press.

Bauer, P.T. (1971) *Dissent on Development: Studies and Debates in Development Economics*. London: Weidenfeld and Nicolson.

Boone, P. (1996) Politics and the effectiveness of foreign aid. *European Economic Review* 40: 289–329.

Burnside, C. and Dollar, D. (2000) Aid, policies and growth. *American Economic Review* 90: 847–868. Working paper available since 1996.

Calderisi, R. (2006) *The Trouble with Africa: Why Foreign Aid Isn't Working*. Basingstoke: Palgrave Macmillan.

Card, D. and Krueger, A.B. (1995) Time-series minimum-wage studies: a meta-analysis. *American Economic Review* 85: 238–243.

Cassen, R. and associates. (1986, 1994) *Does Aid Work?* Oxford: Oxford University Press.

Chang, C.C., Fernandez-Arias, E. and Serven, L. (1998) *Measuring Aid Flows: A New Approach*. World Bank. Available at http://www.worldbank.org/research/growth/ddaid.htm; last accessed July 2008.

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

Christensen, P.W., Doucouliagos, H. and Paldam, M. (2007a) Master list of the AEL: the aid effectiveness literature. Available at http://www.martin.paldam.dk; last accessed July 2008.

Christensen, P.W., Doucouliagos, H. and Paldam, M. (2007b) Master list of the AAL: the aid allocation literature. Available at http://www.martin.paldam.dk; last accessed July 2008.

Corden, W.M. (1984) Booming sector and Dutch disease economics: review and consolidation. *Oxford Economic Papers* 35: 359–380.

Dewald, W.G., Thursby, J.G. and Anderson, R.G. (1986) Replication in empirical economics: the *Journal of Money, Credit and Banking* project. *American Economic Review* 76: 587–603.

Doucouliagos, H. and Paldam, M. (2006) Aid effectiveness on accumulation. A meta-study. *Kyklos* 59: 227–254.

Doucouliagos, H. and Paldam, M. (2007a) Conditional aid effectiveness. A meta-study. Working Paper 2005-14, School of Economics and Management, Aarhus University.
Doucouliagos, H. and Paldam, M. (2007b) A meta-analysis of development aid allocation: the effects of income level and population size. Working Paper 2007-15, School of Economics and Management, Aarhus University.
Doucouliagos, H. and Paldam, M. (2008) Aid effectiveness on growth. A meta-study. *European Journal of Political Economy* 24: 1–24.
Doucouliagos, H. and Stanley, T.D. (2007) Theory competition and selectivity: a meta-meta-analysis. Presented at the Aarhus Colloquium of Meta-Analysis in Economics, Sønderborg, Denmark, September.
Easterly, W. (2003) Can foreign aid buy growth? *Journal of Economic Perspectives* 17: 23–48.
Easterly, W. (2006) *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. London and New York: Penguin.
Easterly, W., Levine, R. and Roodman, D. (2004) Aid, policies, and growth: comment. *American Economic Review* 94: 774–780 (Comment to Burnside and Dollar, 2000).
Egger, M., Smith, G.D., Scheider, M. and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 316: 629–634.
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: 500–516.
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. and Enos, J. (1970) Foreign assistance: objectives and consequences. *Economic Development and Cultural Change* 18: 313–327.
Gylfason, T., Herbertson, T.T. and Zoega, G. (1999) A mixed blessing: natural resources. *Macroeconomic Dynamics* 3: 204–225.
Haavelmo, T. (1965) Comment to W. Leontieff. The rates of long-run economic growth and capital transfers from developed to underdeveloped areas. In *The Econometric Approach to Development Planning* (Section 13, pp. 1039–1068). Pontificia Academiae Scintiarum Scripta Aaria 28. Amsterdam: North-Holland.
Hadjimichael, M.T., Ghura, D., Mühleisen, M., Nord, R. and Ucer, E.M. (1995) Sub-Saharan Africa: growth, savings, and investment, 1986–93. IMF Occasional Paper No. 118.
Herbertsson, T.T. and Paldam, M. (2007) Does development aid help poor countries catch up? An analysis of the basic relations. *Nationaløkonomisk Tidsskrift/Danish Economic Journal* 145: 188–214.
Hook, S.E. (ed.) (1996) *Foreign Aid Toward the Millennium*. Boulder, CO, and London: Lynne Rienner Publishers.
Huges, H. (2003) Trade or aid? Which benefits developing countries more? *Economic Papers* 22: 1–19.
Klein, D.B. and DiCola, T. (2004) Investigating the apparatus. Institutional ties of the *Journal of Development Economics* authors and editors. *Econ Journal Watch* 1: 319–330.
Leamer, E.E. (1983) Let's take the con out of econometrics. *American Economic Review* 73: 31–43.
Lensink, R. and White, H. (2001) Are there negative returns to aid? *Journal of Development Studies* 37: 42–65.
McCullough, B.D., McGeary, K.A. and Harrison, T.D. (2006) Do *Economics Journal* archives promote replicable research? Working Paper, Department of Economics and International Business, Drexel University, PA.

McGillivray, M., Feeny, S., Hermes, N. and Lensink, R. (2006) Controversies over the impact of development aid: it works; it doesn't; it can, but that depends. *Journal of International Development* 18: 1031–1050.
Mosley, P. (1986) Aid effectiveness: the micro–macro paradox. *Institute of Development Studies Bulletin* 17: 214–225.
Mosley, P. (1987) *Foreign Aid: Its Defence and Reform*. Lexington, KY: University of Kentucky Press.
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: Aarhus UP.
Paldam, M. (2004) Are vote and popularity functions economically correct? In C.K. Rawley and F. Schneider (eds), *Encyclopedia of Public Choice* (pp. 49–59). Dordrecht: Kluwer Academic.
Paldam, M. (2005) Appendix. Available at http://www.martin.paldam.dk; last accessed July 2008.
Rajan, R.G. and Subramanian, A. (2008) Aid and growth: what does the cross-country evidence really show? *Review of Economics and Statistics* 90: 643–665.
Roberts, C.J. and Stanley, T.D. (eds) (2005) *Issues in Meta-Regression Analysis and Publication Bias in Economics*. Oxford: Wiley-Blackwell.
Sachs, J. and Warner, A.M. (1995) Natural resource abundance and economic growth. NBER Working Paper 5398.
Sala-i-Martin, X. (1997) I just ran two million regressions. *American Economic Review* 87: 178–183.
Sinn, H.-W. (2004) *Ist Deutschland noch zu retten?* Berlin: Econ-Verlag.
Stanley, T.D. (2001) Wheat from chaff: meta-analysis as quantitative literature review. *Journal of Economic Perspectives* 15: 131–150.
Stanley, T.D. (2005) Beyond publication bias. *Journal of Economic Surveys* 19: 309–345.
Stanley, T.D. (2008) Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistics* 70: 103–127
Stanley, T.D. and Jarrell, S.B. (1989) Meta-regression analysis: a quantitative method of literature surveys. *Journal of Economic Surveys* 3: 161–170.
Svensson, J. (1999) Aid, growth and democracy. *Economics and Politics* 11: 275–297.
Tarp, F. and Hjertholm, P. (eds) (2000) *Foreign Aid and Development. Lessons Learnt and Directions for the Future*. Studies in Development Economics. London: Routledge.
WDI (2007) *World Development Indicators*. Washington, DC: CD from the World Bank.
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. New York: Oxford University Press.
Younger, S.D. (1992) Aid and the Dutch disease: macroeconomic management when everybody loves you. *World Development* 20: 1587–1597.