Aarhus, 1/5-08

Development aid

The embarrassing gap between wishes and results

Martin Paldam, School of Economics and Management,
University of Aarhus, Denmark.

The following deals with the question: Does development aid generate development? In the language of economics it is: Has aid to the poor countries generated economic growth? This question is analyzed in a large literature surveyed in Doucouliagos & Paldam (2008b).

Technically this is an easy question to analyze. The statistical techniques for growth regressions are well known, and the data are ideal: Since aid programs started 45 years, no less than 6000 pairs of annual observations for aid and growth have been generated. The average aid share is about 7½%, and while some countries get plenty, others receive little.

From similar data in other fields, 10 other factors have been shown to have a robust impact on growth. When comparable, most of these factors are of a smaller magnitude than 7½ % of GDP. If aid gave development, it should be very easy to show, but it is not. This is well known in development research. If the world's leading experts in growth and development were asked, they would surely agree that research has been unable to show that aid is a significant factor of development.[1]

Table 1. The correlation of aid and growth 1227 periods of 5 years

Period	N	Cor	Period	N	Cor
1960-65	92	-0.12	1985-90	143	-0.12
1965-70	103	-0.00	1990-95	169	-0.00
1970-75	111	-0.01	1995-00	178	0.09
1975-80	122	0.06	2000-05	175	-0.02
1980-85	134	0.09	Average	1227	0.00

Note: It does not matter for the first two digits of the average if it is weighted with the number of observations or not. For all countries in the WDI (references), except OECD countries.

The central fact of the discussion is presented in Table 1: Aid and growth are uncorrelated.[2] When the standard growth models are applied to the data, the result is still very close to zero.[3] If the discussion of aid effectiveness was a normal scientific one, it would be boring as it would deal with the second decimal only: Can we find a tiny effect, or is there none at all?

1. Two strong priors

However, the discussion of aid effectiveness is far from boring. Hot scientific disagreement is often brought about by something else that gives it zest. In the case at hand, the controversy is due to two strong priors:

(i) The ***angel prior:*** Aid aims at doing good, and we all want to be on the side of the angels.

(ii) The ***gravy prior*:** Aid has now reached the $100 billion mark. About 10% is fees to consultants, including most development economists. We do not want to be kicked off the gravy train.

The two priors point in the *same* direction – the angel prior gives the researcher a moral justification for gravy seeking. Hence, research in aid effectiveness is a prior-ridden field, where the tension between the priors and the basic fact has generated a large and embarrassing literature. From time to time, a researcher succeeds in explaining away the fact, and a great joy erupts among the interested, but then new data are published, and the happy result collapses. See the story about the life cycle of good models below.

I term the aid effectiveness literature *embarrassing*. It ought to be the job of research to explain the facts – not to explain away the facts. When the latter endeavor is so transparently connected to wishful thinking and economic interest, it is truly toe curling.

The purpose of this paper is to look at the findings in the large literature, which covers app 250 studies. The following is a short summary of the results anyone, who tries to look beyond the veil of priors, ought to agree on.

2. From micro to macro: The paradox

On the micro level, it is uncontroversial (since Cassen, 1986 and 1994) that about 50% of all aid projects succeed. However, development is a macro level phenomenon. The macro aspect is built into aid projects (sector programs) through social cost-benefit analyses. For many years, the typical criterion for determining if an aid project was acceptable was that its contribution to growth had been at least 10%.

Consequently, half of the projects contribute more than 10%, and the failed half contribute less than 10%, so we expect that 10% holds in average. If it does, projects amounting to $7^1/_2$% of GDP should generate $^3/_4$ percentage points of growth per year. This is half the growth of the average less developed country.

Some aid is given for reasons unconnected to economic development, e.g. as emergency aid, aid to democracy and culture, etc. Even if we say that only $^2/_3$ of aid is aimed at development, it should still contribute about $^1/_3$ of the growth in the average poor country. This should be very easy to show. Table 1 is the simplest way to look at the macro data. It is clear that aid and growth are totally unrelated.

The micro findings as regards the individual projects and the macro result for the whole society are thus separated by an amazingly large gap. It is known as the *micro-macro paradox* of development aid (since Mosley 1987).

The tension between the zero-correlation result and the two priors has generated the large literature in the field. It has the motto: If you torture the data long enough, it will confess what we wish!

3. The large scale data mining and the reluctancy bias

The torture of the data is done by incessant and repeated mining. Scores of a thousand models have been applied to the data by the full battery of statistical techniques.

Growth regression is a field where it is notoriously easy to make model variants. Every researcher has a large amount of calculation power on his desk, and fine statistical packages are easily available. With these powerful tools, he can easily estimate e.g. 1'000 model variants. Each estimate provides one value of aid effectiveness. From Table 1, we know that approximately 950 of these estimates are insignificant. However, the remaining 50 are significant, and of these about 25 are positive. What happens if we choose the one that is the most positive? It will make everybody happy, but maybe we have all been conned?

A careful data search is a fine way to *propose* new regularities. However, new models discovered by data mining must be *independently replicated* (that is, by new authors on new data) to be trustworthy. This is well known by researchers in medicine, physics etc, but in the social sciences it is only accepted in principle – not in practice.

While we are searching the 1'000 model variants, the priors enter. Researchers must have publications to prosper. The 950 insignificant results are difficult to publish, and 25 results have the "wrong sign" – they will be rejected by all insiders. So it is tempting to steer the search process away from them too.

Hence, the attention of the researcher will somehow concentrate on the 25 good models and on finding reasons why one of the nice positive models is better than all others. The very word *research* refers to the process whereby you search and search again. A research result happens when a researcher stops searching. But when does he stop searching? Surely he stops when he is *satisfied* with the result. It is hard to believe that the two priors do not influence when satisfaction sets in. Hereby I do not suggest that researchers are dishonest, only that they are human.

The two priors thus cause the search process to be skew, and the problem in the field of aid effectiveness is that both priors are in the same direction. This has caused Hristos Doucouliagos[4] and me to suggest that this literature may suffer from a ***reluctancy bias***. The profession is reluctant to publish negative and insignificant results.

A set of statistical tests have been developed (see section 5) to check if a literature has a bias like that. When these tests are applied on the aid effectiveness literature, all the red warning lights come on! Our suggestion is strongly confirmed.

4. The causal direction: Referring to Figure 1

The gut reaction of any economist, looking at Table 1, is that it does not sort out causality. We want to study E: aid ➔ growth. But what we observe may in principle be caused by the reverse causality R: growth ➔ aid.

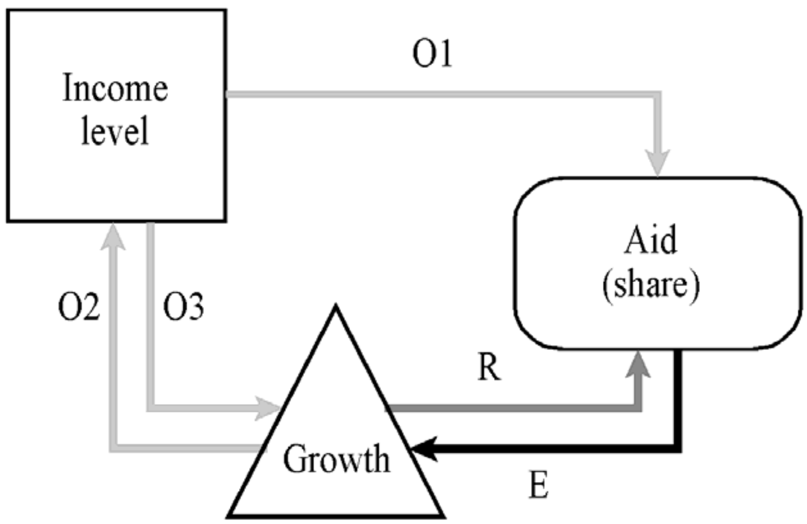
The literature has often discussed how the two causal relations can be sorted out: Maybe the world is so mischievous that the two causal links - E and R - produce coefficients that have the reverse sign and are precisely of the same numerical size. It would surely be a strange coincidence, but the two priors have ensured that it has been thoroughly researched.

Figure 1 shows the causal links that may potentially be involved. E is the aid effectiveness link. We want to study E, and hence to control for reverse causality R. In addition, some other causal links (O1 – O3) may come in and influence E and R. As we want to find something that makes E more positive, we hope that R is substantial and negative; but as we shall see, it is neither substantial nor negative. In the literature, no less than 4 possible R-mechanisms have been proposed, so that R = R1 + R2 + R3 + R4:

(R1) The poorer the country is, the more aid should it receive (O1), so if poor countries grew less than other countries (O3), this would generate reverse causality. Both (O1) and (O3) are small, so their product is very small. R1 is thus negative but tiny.

(R2) Poor countries get emergency aid, when they have a crisis. R2 is thus negative, but emergency aid is normally only a small fraction of the aid. In addition, it often comes with a delay. The calendar divides time into years, and this may cause R2 to turn sign; thus in average it is probably negative but small.

Figure 1. Five possible causal links in the aid-growth complex



(R3) The large development banks – the World Bank and its regional sisters – give their aid as cheap loans to good projects. Countries generate more such projects the faster they grow. R3 is thus positive.

(R4) It is often alleged that commercial interests influence the aid flows so that donors favor future markets. Rapidly growing countries are more promising in this respect. R4 is thus positive too, but probably quite small.

When all 4 effects R1 to R4 are added, it is likely that the effects cancel out and the net effect is really small. This is precisely what the aid allocation literature finds. The R-effect is estimated in 30 papers that contain 211 estimates (see Doucouliagos & Paldam 2007). The average result is positive, but very small.[5]

In spite of a considerable research effort analyzing the causality from growth to aid, nothing suggests that the zero correlation result hides two reverse effects with the reverse sign.[6] The reason for the zero correlation result is that *both* links are basically zero.

5. Meta studies of the literature

The aid effectiveness literature (AEL) consists of 110 studies, and the aid allocation literature (AAL) that explains the pattern in the aid flows has reached 170 studies. Some of these studies overlap, but there are still about 250 in total.[7] At least 200 man years of research have been put into this production. This is a large effort and it is worth taking seriously.

The meta analysis is a technique that studies the distribution of the published results in a literature about the same effect, such as the effectiveness of aid. It studies if the results converge to the same results, if there are breakthroughs in the research, and in addition it has developed tests to detect publication biases.

We have found 543 estimates of the effect of aid on growth (see Doucouliagos & Paldam 2008a) which are so well documented that they can be converted to the same scale. When their distribution is analyzed, we find that it has two properties that are unreasonable if we assume that researchers are robots seeking only truth.

(U1) Results ought to become clearer (more stable and significant), the larger the data set on which they are estimated. However, we find that the results get less and less significant, the larger the data set. This is unreasonable. (U2) The aid industry should learn from experience, as everybody else, but the estimated effect of aid falls over time. This is unreasonable as well.[8] The two unreasonable results turn reasonable if we take the reluctancy bias into consideration. That is, if we note that the published results discriminate against bad results.

A simple example where we can calculate everything[9] will show what happens. Table 2 considers a study of a data set of N flips with a coin. The good result is heads (you win). Imagine that you consider all outcomes of 4 flips with the coin and reject the 95% worst results! If N is 4, all the remaining good results are heads; the flipping effectiveness is thus 100%. If N is 10, the flipping effectiveness falls to 82.3%. It looks as if effectiveness goes down as N goes up, but what really happens is that the effect of the bias caused by reluctance to consider bad results goes down. However, the bias is still substantial even when N is 100.

This means that if you accept that researchers are human, everything falls together. (U1) shows the fall of the bias, the more data that are used. (U2) This makes it look as if aid effectiveness falls over time. When the distribution of the published results is adjusted for the reluctancy bias with the methods developed for this purpose,[10] the average result turns insignificant. The literature has not found a robust method to reject the zero correlation result.

Another way to understand the implication of the reluctancy hypothesis is to consider the characteristic fate of the most aid-positive models.

Table 2. The effect of the reluctancy bias, in the case of the head-tail game

Number of flips N	4	10	20	40	100	∞
Good 5% results:	100%	82.3%	72.8%	66.1%	60.3%	50%
Neutral choice: All results	50%	50%	50%	50%	50%	50%
Selectivity bias of good results	100%	64.7%	45.6%	32.3%	20.5%	0%

Note: The two top rows of percentages are the probability for head. The bias is in % of the neutral choice.

The life cycle of good models: When one such model is found, it is easy to publish, and it is heavily cited and popularized. It is received with great joy by the aid industry: Finally it has been showed that aid works! The diligent researcher, who found the model, is promoted, etc. But then after 3 years, new data appears, and the model collapses, and it is quietly swept under the carpet to all the others.[11]

6. The aid industry and the lack of research ethics

It is a fascinating experience to discuss development aid, for here one inevitably encounters a lobby with many discussants. I have discussed with dozens of those in the (Danish) media. Nearly all of these discussants live fully or partly on development aid. One of the most common arguments is that I should go easy with these facts as they may harm aid. This point of view is also common from people in research positions. These researchers thus reveal that they feel obliged to twist their own research so as to be as positive to aid as possible. This is precisely why the aid effectiveness literature has the reluctancy bias.

It is precisely because of the interests and the resulting biases that exist in research that *ethical standards* for research exist. They demand that researchers with economic interests in a field state this clearly when they publish in that particular field. I have studied 100 articles on aid effectiveness and tried to check who the researchers are. 75% present only a university affiliation. Fortunately many have home pages and can be “googled”. A surprising number of those I found turned out to be frequent travelers on the gravy train. I have even found cases of authors who had special research positions financed by aid money, but did not mention this curious coincidence in their papers on aid effectiveness!

7. Conclusion

Thus we have seen that the reason why the zero correlation result occurs is simple: It is because the two causal links between the variables development aid and development are both very close to zero. In spite of very good data and a large research effort, it has not been proven that development aid generates development.

It does not mean that there is no effect of aid at all. There may be good social consequences, though it appears that it has not been demonstrated. Further, we know that aid can alleviate emergency situations, food aid can feed the hungry, military aid can help the good guys defeat the bad ones, etc. But development is something else. It is not easy to generate. It is notoriously difficult for foreigners to give a country a treatment so that it starts to grow.

References:[12]

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

Barro, R.S., Sala-i-Martin, X., 1995, 2004. *Economic growth*. MIT Press, Cambridge, MA

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

Cassen, R., 1986, 1994. *Does Aid Work?* Oxford UP, Oxford UK

Chenery, H.B., Srinivasan, T.N., Behrman, J., 1988, 1989, 1995. *Handbook of Development Economics.* North-Holland, Amsterdam

Christensen, P.W., Doucouliagos, H., Paldam, M. 2007a. Master list of the AEL: The Aid Effectiveness Literature. Update 1/7-2007. Documentation working paper.

Christensen, P.W., Doucouliagos, H., Paldam, M. 2007b. Master list of the AAL: The Aid Allocation Literature. Update 24/1-2007. Documentation working paper.

Doucouliagos, H., Paldam, M., 2007. Explaining development aid allocation by growth: A meta study. Economics Working Paper 2007-13, Aarhus University

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

Doucouliagos, H., Paldam, M., 2008b. The aid effectiveness literature. The sad results of 40 years of research. *Journal of Economic Surveys* forthcoming

Easterly, W., Levine, R., Roodman, D., 2004. Aid, policies, and growth: Comment. *American Economic Review* 94, 774-780 (Comment to Burnside & Dollar, 2000) Earlier: New data, New doubts. NBER WP 9846

Hansen, H., Tarp, F., 2000. Aid effectiveness disputed. *Journal of International Development* 12, 375-98.

Herbertsson, T.T., Paldam, M., 2007. Does development aid help poor countries catch up? An analysis using the basic relations. *Nationaløkonomisk Tidsskrift* 145, 188-214

Jensen, P.S., Paldam, M., 2006. Can the two new aid-growth models be replicated? *Public Choice* 127, 147–75

Mosley, P., 1987. *Overseas Aid: Its Defence and Reform.* John Spiers, Brighton, Sussex

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 WP No 05/126) Accepted by *Review of Economics and Statistics*

Stanley, T. D., 2008. Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection. *Oxford Bulletin of Economics and Statistics* 70, 103-27

WDI, World Development Indicators. Data bank of the World Bank: http://devdata.worldbank.org/dataonline/

Notes:

---

1. Barro & Sala-i-Martin (2004) is the leading textbook on growth and development. It contains about 100 pages on the factors causing growth. It does not even mention aid. The same applies to the two large volumes of the Handbook of Growth (Aghion & Durlauf, 2005). Even in the 4 heavy volumes of the Handbook of Development (Chenery, Srinivasan & Behrman 1988-95) aid plays a small role.
2. Growth is the real GDP per capita. Aid is ODA in % of GDP. These data are averaged over 5 years to reduce the effect of natural calamities such as draught and the activity effects of aid.
3. See Rajan & Subramanian (2005) and Herbertsson & Paldam (2007).
4. Chris Doucouliagos is professor at Deakin University, Melbourne, Australia. He is one of the leading experts in meta-analysis. See section 5.
5. To increase E, as wished, R should be negative, but it actually is positive.
6. Econometric techniques have been developed to take into consideration that causality may go both ways. They have often been applied in this field. They do not in general give different results from standard models disregarding the reverse causality.
7. The most comprehensive bibliographies of the AEL and AAL are Christensen et al. (2007a) and (2007b) respectively. They are the data of Doucouliagos & Paldam (2008a and b) and a handful of other meta studies.
8. One reason why the available data grow is time. However, (U1) and (U2) can be sufficiently discriminated to show that both hold true independently of each other.
9. The values are calculated from the binominal distribution with the probability of ½ for head.
10. Several such methods have been presented. Normally they give very much the same results. See Stanley (2008) for a recent comparison of these methods.
11. The two most positive models from the last decade are Burnside & Dollar (2000) and Hansen & Tarp (2000). Both have had a life cycle as described. They failed the decisive replication test (see Easterly, Levine & Roodman 2004 and Jensen & Paldam 2005) and Doucouliagos & Paldam (2008b).
12. Papers written by the author can be downloaded till about 2 years after they are published from the *working papers* section at the URL: http://www.martin.paldam.dk.