21/10-2004

Can the two new aid-growth models be replicated?

Peter Sandholt Jensen and Martin Paldam

School of Economics and Management, University of Aarhus, Denmark[1]

Abstract: The recent literature on the aid-growth relation discusses two competing models from the same family: The Good Policy Model, where the key feature is policy times aid, and the Medicine Model, where it is aid squared. Both models are reduced forms using roughly the same set of variables, and have been reached on a sample of about 1/3 of the available data. We first simplify the models so that they can be replicated on as much of the available data as possible. Within the sample the Good Policy Model proves fragile, while the Medicine Model is more robust. Neither model replicates outside the original data sample. Further, we apply a semi-parametric regression technique to test for an unknown functional form of the aid-growth relation. It rejects that aid is statistically significant. The evidence in favor of an aid-growth relationship is weak for the models of this family.

Keywords: Aid effectiveness, growth, semi-parametric panel regression

Jel: C14, C23, F35, O4

---

1. Mail: Building 322, DK-8000 Aarhus C, Denmark. Phone and e-mail for PSJ +45 8942 1615 and <psjensen@econ.au.dk> and for MP +45 8942 1607 and <mpaldam@econ.au.dk>. We are grateful to Howard White, and to Phillip Harms, Chris Doucouliagos and other discussants when the paper was presented at EPCS-04 in Berlin.

I. Introduction: Two models from the same family

For long the relation from development aid to economic growth was found to be dubious – much like the relation from resources to growth. However, during the last 5 years the discussion has been dominated by two new and optimistic models belonging to a certain *family* of models. Both have few *substantial variables* and reach a *key empirical finding*, which leads to a clear and optimistic *policy prescription*. If aid is *redirected,* it will do much (more) good:

The most influential is ***the Good Policy Model***.[2] Growth is explained by three substantial variables aid, a good policy index and their interaction. The key finding is a positive interaction term. It claims that LDCs choose economic policies independently of expected aid, and then aid gives the policies an extra push. Good economic policies become better and bad ones worse. Therefore, aid should be concentrated on countries with good policies.

The Medicine Model.[3] Growth is explained by two substantial variables: aid and aid squared. The key finding is a positive aid term and a negative aid squared term. It claims that aid helps all countries, but only up to a point, *Ω*. More aid is increasingly harmful. Consequently, aid should be distributed as evenly as possible and never exceed the optimal dose.

The main point in favor of the good policy model is that it tallies well to the intuition of practitioners, while the medicine model fits the data better. The empirical support for both models comes from a study of a data set CFS-56 (see table 2), which only covers about 30% of the existing observations for aid and growth. Both models are thus reached from the mining of a particular sub-sample of the data available. Consequently, it is ideal that the remaining 70% of the data are available to test if the models replicate. This is what we do at present.

A problem immediately arises. The reason why the models use so little of the available data is that the authors wanted to control their models for many potentially relevant effects. Few of the controls wanted are available for all countries and years desired. Thus we can only replicate the original models for slightly more than the CFS-56 data. However, we have developed two simplified versions that can be replicated on (almost) all of the data available.

The first simplified versions are reached by stripping the models down to the ***minimal versions*** needed for generating its key finding. The second simplified versions are the ***base***

2. It was proposed by C. Burnside and D. Dollar (see references).

3. First Hadjimichael et al. (1995) found that aid squared becomes significant in some aid-growth models. It was further developed in the papers by the group of F. Tarp, H. Hansen and C.J. Dalgaard (references), and in a somewhat different set up by R. Lensink and H. White. The model is sometimes termed the *Aid Laffer curve*, but our name has more precise connotations.

models reached by replacing the controls with fixed effects for countries, as shown in table 1.

We first make *within-sample* replications for the CFS-56 sample, where the two simplified versions of the Good Policy Model differ in a very revealing way, while The Medicine Model gives similar results in the two simplified versions. Secondly, we make *out-of-sample* replications of the results on (nearly) all the available aid-growth data using the simplified versions. Here the results are poor for both models. The Medicine Model claims that the relation is nonlinear in the aid variable, and we also use a new technique, where the base model contains a semi-parametric estimate of the aid-growth term. It allows us to test whether aid affects growth irrespective of the shape of the relation, and to see how the best aid-growth shape looks for models within this family.

The two models are closely related by being reduced form models of the causal relation from aid to growth, and they are only controlled for a certain set of additional variables. The family of models is defined in table 1 overleaf. We will briefly mention the effects of some additional controls, but only in the notes. At present we have decided to remain within the family, simply to keep the paper within bounds. This turns into a problem in section VI where we want to say something more general about the aid term.

This also means that we refrain from giving a survey of the whole literature. Broader surveys are, e.g., White (1992), Jepma (1995), Paldam (1997a; 109-129) and Hansen and Tarp (2000). They all show that the effect of aid varies widely from one study to the next.[4]

The choice of keeping to a reduced form relation means that the explicit channels from aid to growth are not modeled. This, e.g., applies to the link from aid to the share of the public sector and from that share to growth, or from aid to the savings rate and from that rate to growth. This is a debatable modeling strategy, but it is the one of the model family examined.

Section II surveys the new literature, our method and choice of models. Section III considers the data sets. Section IV gives the replications of the two models within the CFS-56 sample. Section V holds the out-of-sample replications. Section VI looks at the semi-parametric results for a general aid-term. Finally, section VII draws the conclusions and suggests some extensions. The countries included in the data sets are listed in the Appendix.

4. A recent survey by Hansen and Tarp (2000) referred to 72 estimates on the aid-growth relation. While 40 estimates found that aid increased growth, other 31 estimates found an insignificant result, and 1 found that aid harmed growth. Our research indicates that the last result is reported too rarely – several of the models presented in our paper give negative coefficients to aid if the country dummies are deleted.

II. The controversy and the two models

The variables and models discussed are listed in table 1 that defines the family of models included in the main text – the family is a subset of the larger group of Barro-type growth regressions (see Barro, 1997). Note that the time unit is 4 years. The *substantial models* are $g_{ti} = \mu_1 h_{it-j} + \gamma_0 \Gamma_{it} + \gamma_1 \Gamma_{it} h_{it}$ and $g_{ti} = \mu_1 h_{it-1} + \mu_2 h_{it-1}^2$, and the *substantial results* are thus the μ's and the γ's. When these models are estimated they are supplemented with a set of additional "control" variables.

Table 1. Variables and models discussed

i	country index	Y_{it}	GDP in PPP terms, start of unit
t	time index, one unit is 4 years	y_{it}	gdp, Y_{it} per capita, start of unit
g_{it}	real growth rate, average of 4 years	$\mathbf{D_i}$	fixed effects for time
h_{it}	aid in percent of GDP, same average	$\mathbf{D_t}$	fixed effects for counties
$\Phi(h_{it})$	generalized aid term	$\mathbf{x_{it}}$	"nuisance" controls
Γ_{it}	good policy index	$\mathbf{r_i}$	"necessary" controls, $\mathbf{r_i} \subset \mathbf{x_{it}}$

The **x** set contains 7 variables: (x_1) institutional quality index from Keefer and Knack (1995), (x_2) South of Sahara Africa dummy, (x_3) East Asia dummy, (x_4) political assassinations, (x_5) ethnical fractionalization, (x_6) the product of x_4 and x_5 and (x_7) financial depth M2/GDP.

(1)	$g_{ti} = \mu h_{it-j} + \boldsymbol{\alpha}'\mathbf{x_{it-j}} + \beta y_{it} + u_{it}$	Main idea
	Good Policy Model	Versions:
(2a)	$g_{ti} = \mu_1 h_{it} + \gamma_0 \Gamma_{it} + \gamma_1 \Gamma_{it} h_{it} + \boldsymbol{\alpha}'(\mathbf{x_{it}}, \mathbf{D_t}) + \beta y_{it} + u_{it}$	original
(2b)	$g_{ti} = \mu_1 h_{it} + \gamma_0 \Gamma_{it} + \gamma_1 \Gamma_{it} h_{it} + \boldsymbol{\alpha}'(\mathbf{r_i}, \mathbf{D_t}) + \beta y_{it} + u_{it}$	minimal version
(2c)	$g_{ti} = \mu_1 h_{it} + \gamma_0 \Gamma_{it} + \gamma_1 \Gamma_{it} h_{it} + \boldsymbol{\alpha}'(\mathbf{D_i}, \mathbf{D_t}) + \beta y_{it} + u_{it}$	base model
	Medicine Model	Versions:
(3a)	$g_{ti} = \mu_1 h_{it-1} + \mu_2 h_{it-1}^2 + \boldsymbol{\alpha}'\mathbf{x_{it-1}}, \mathbf{D_t}) + \beta y_{it} + u_{it}$	original
(3b)	$g_{ti} = \mu_1 h_{it-1} + \mu_2 h_{it-1}^2 + \boldsymbol{\alpha}'(\mathbf{r_i}, \mathbf{D_t}) + \beta y_{it} + u_{it}$	minimal version
(3c)	$g_{ti} = \mu_1 h_{it-1} + \mu_2 h_{it-1}^2 + \boldsymbol{\alpha}'(\mathbf{D_i}, \mathbf{D_t}) + \beta y_{it} + u_{it}$	base model
(4)	$g_{ti} = \Phi(h_{it-j}) + \boldsymbol{\alpha}'[\mathbf{D_i}, \mathbf{D_t}] + \beta y_{it} + u_{it}$	Generalized base model

Notes: Lowercase Greek letters are coefficients, bolded variables are vectors and u_{it} residuals. The convergence term, β, should be negative. The minimal version contains only the controls necessary to generate the substantive results (the μ's and γ's). Note that ***r*** contains only *x'ses*, which are constant over time.

We start by asking why it is so easy to get different results. Then the two new positive models are surveyed, and their policy implications are discussed.

II.1 The aid-growth relation: Why are results so different?

The question in the headline can be answered at two planes. The first is socio-political: Aid is a field where researchers have as well strong feelings as interests and are willing to go quite far torturing the data to make it confess, see further in section II.2. The second is that it is doable; aid effectiveness is a field where it is easy to vary the research in 3 dimensions:

(1) Aid data are of two types: The **ODA**-data (Official Development Aid) from the OECD, and two **EDA**-data sets (Effective Development Aid) made by adjusting each loan in the ODA-set with the gift element: The **CFS**-set from Chang, Fernandez-Arias and Serven (1998), and the **ELR**-data from Easterly, Levine and Roodman (2003). Section III discusses the three data sets. We use all three sets in the empirical sections.

(2) The two models between 3-4 substantial variables. However, many "nuisance" variables may in principle distort the relation so that they have to be controlled. This allows many thousand model variants, see section II.2. We have chosen to stick to the control set used by Burnside and Dollar (2001) and Dalgaard and Hansen (2001). We have also tried other controls, e.g., the ones of Lensink and White (2001), but to keep the paper tidy additional controls are mentioned in notes only.

(3) Both substantial models contain a second order term: The Good Policy Model uses aid times good policy, Πh, while the Medicine Model uses aid squared, h^2. By including non-linearities the number of model variants increases dramatically. We include a section analyzing the form of the aid-term using a semi-parametric technique, which finds the best continous form of the term, see section VI.

*II.2 The never ending story of the **x**-set, counter causality and moral hazard*

The theory of growth and the empirical literature on cross-country panel regression models are separated by a large gap, causing the problem of identifying exactly which variables belong in any equation (see Sala-i-Martin, 1997). Hundreds of variables that may or may not enter the **x**-set in relations of the type discussed have been proposed. We have decided to stick to the variables used in the published versions of the two models.

The two models have one advantage. When they are controlled for fixed effects for time, the only variables that matters for the substantive results are controls that are *constant for each country*. These variables control for country differences, precisely as can be done – in a more general way – by fixed effects for countries.

The choice between specific variables and fixed effects to control for country differences is difficult: (1) It is useful to know precisely which country differences that matter, but a large **x**-

set allows a search among millions of models, and thus makes moral hazard a key problem. Fixed effects controls for country differences in a general way, which is manipulation proof. (2) Few controls are available for all aid-data observations, so each control included reduces the data in the tests. This is why Burnside and Dollar used only half the CFS-data. The use of fixed effects for countries allows replications of the models on all the data where the substantial variables of the models are available.

The two models both aim at answering the following policy question: What happens to growth if aid to a country is increased? This is a time series question. The controls give the conditions that affect the answer – the ideal is that these conditions are as simple and well understood as possible. Fixed effects for countries claim that all country differences can be taken as one shift of the level. Hereby they convert the data to time series as much as possible.[5] We see no good reason why the two models should need to be controlled for a different set of variables. The model that needs the simplest controls is thus the superior one.

We want to analyze the causal relation from aid to growth, but it is possible that causality is from growth to aid. The many studies of the determinants of aid (see Paldam, 1997a; 120-122) do not suggest that the growth-to-aid relation is strong, but we cannot a priori reject reverse causality. Hence, it is important to control for counter-causality when the aid-growth relation is estimated. Two methods are available: (1) Aid is lagged by one time unit relative to the growth explained, or systematically by GMM-technique.[6] (2) The relation is estimated by a 2SLS-technique with suitable instruments in the first step. To find suitable instruments is not easy. Also they enter almost as the controls in the **x**-set and add to the moral hazard problem. Consequently, method (1) is our preferred method.

Historians routinely check all messages for excess concordance with the interest of the sender – a practice known as *source criticism*. This practice is less accepted in economics though the moral hazard problem is well known.[7] In the literature on macro aid effectiveness, the research is often financed from development aid budgets. The Good Policy model was made by World Bank researchers and produced results much in line with the Washington Consensus. The Medicine model has been advocated by a team of researchers financed by

---

5. Barro (1997; p 36-42) argues against the use of fixed-effect for two reasons: (1) By reducing panel-data to time series information is lost. But if this information is irrelevant, it should be discarded. (2) It increases the measurement error for the convergence term. It is not our subject at present. However, the conditional convergence term is negative (as it should) in all estimates given and significant in about 1/3 of the ones.

6. The original articles do not use the GMM estimator, and it proves to matter little. So we present the OLS estimates is the tables and report the GMM-results in the text and in notes.

7. See however the new literature on Meta-Analysis started by Card and Krueger (1995). A recent survey is Doucouliagos and Laroche (2004).

Danish Development Aid, and produced results much in accordance with the thinking of the funding agency.

II.3 Recent empirical evidence and the two theories: Good policy or medicine?

The micro evidence is that app. 50% of all development projects are successful and virtually none are harmful, see e.g. Cassen (1994) or Paldam (1997a). Thus the average project should give a positive contribution to growth. The contrast between the weak macro and positive micro evidence is known as the *micro-macro paradox* of aid (since Mosley, 1986).

The EDA-based research started with Burnside and Dollar (1997, 2000), presenting the Good Policy Model, which in addition to the new aid data introduced the good policy index discussed in section III.3 below. The result of the research was that aid only increased growth in countries following good policies (as defined). This has appealed to many development practitioners and has been widely reported (see World Bank, 1999).

Hansen and Tarp (2000) used ODA-data for the countries of the CFS-56 data set, and showed that an inclusion of aid squared made the interaction term insignificant. Dalgaard and Hansen (2001) showed the same using the CFS-56 data. Lensink and White (2001) reached similar conclusions, though they noted that the result is fragile to the countries included. They explained it in a model with endogenous growth. Aid finances government spending that is productive,[8] but has negative incentive effects elsewhere in the economy. At low levels of aid, the positive marginal effect of aid dominates, but with high aid the negative effect dominates.[9]

The parametric assumptions used to estimate the Medicine Model do not follow from the theory, and Hansen and Tarp (2000; p.118) note that we are in fact dealing with an unknown functional form, see section VI.

Even when we concentrate on the two models mentioned, it should be mentioned that a number of studies have appeared using other conditioning sets. Some examples are: Svensson (1999) finds that aid interacted with democracy significantly explains growth. Collier and Dehn (2001) finds that export shocks interacted with the change of aid explains growth. Chauvet and Guillaumont (2002) argued that that aid is more effective in politically stable economies; Chauvet and Guillaumont (2001) found that vulnerability to external shocks determines aid effectiveness rather than good policy. Collier and Hoeffler (2002) examine the role of conflict.

---

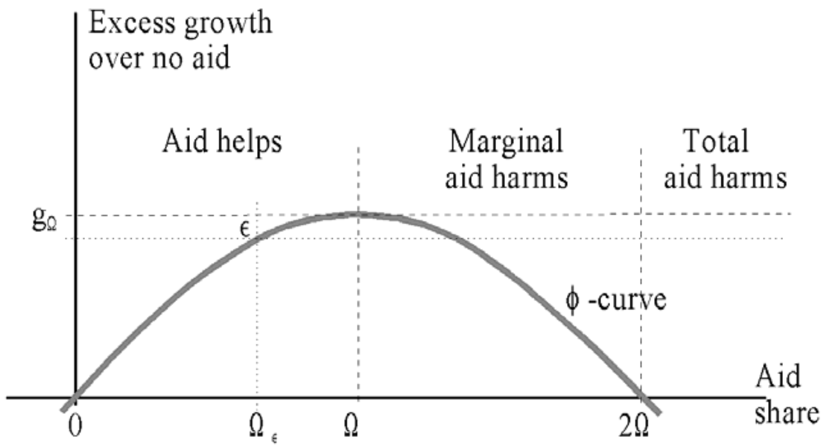
8. Gørgens, Paldam and Würtz (2003) find no signs that public regulations increase growth.
9. Paldam (1997b) is a study of Greenland that has received aid of about 50% of GDP since the early 1950's. It shows how far distortions and aid dependency can go in practice.

II.4 The reverse policy implications of the Good Policy Model and the Medicine Model

The *Good Policy Model* has the policy implication that aid should be concentrated on the countries following good policies. Burnside and Dollar even calculate the (large) gain for the world if aid is redirected accordingly.

The reverse policy conclusions emerge from the Medicine Model, where the Φ-function is quadratic as shown on Figure 1. Here the growth effect is independent of the policy of the recipient country: (a) If the aid share exceeds Ω the growth generated decreases. (b) The marginal growth contribution decreases from zero to Ω. Thus aid should be distributed to make the aid shares of the recipients as equal as possible. A lot hinges upon the position of the Ω-point.[10]

Figure 1. Optimizing the dose in the Medicine Model



In estimates on the CFS-56 data, where both aid and aid squared are significant, Ω is close to 5%, while figure 2a suggests that it may be lower. That corresponds to 12% in the ODA-data which is a little lower than what is found in the closest matching ODA-sample (ODA-55). The average aid share is below the Ω-point, but in the WDI (2003) data for 2001 no less than 24 countries did receive ODA-aid in excess of 12%, and 5 are even above 2Ω, where it would be better with no aid at all. Hence, a substantial growth gain would result from a redirection of aid, if the model is true, as is the case for the Good Policy Model.

10. The welfare argument is that transfers from DCs with low marginal with high marginal utilities give a world welfare gain. If we set the marginal loss in DC to ϵ (measured in "growth equivalents") then aid should not stop in Ω, but already in Ω_ϵ. The Φ-curve is flat around its maximum Ω, so even a small ϵ may be visible on the horizontal axis. When Ω is found to be between 5 and 6 we thus choose the lower value.

III. The data

First the three sets of aid-data are discussed, and then we turn to the good policy index. Note the two Appendix tables listing the countries included in the different samples. Table 2 surveys the various data-sets used in the regressions.

Table 2. Aid data samples

Name	Source	Variant	Period		n
ODA	Official, WDI (2003) used as	ODA-full	From 1966	All available in WDI (2003)	756
	source	ODA-55		Sample for the CFS-56 countries	472
CFS	Chang, Fernandez-Arias and	CFS-56	1970-1993[a)]	Used for both models	269
	Serven (1998). EDA-set	CFS-42		42 unused countries	216
		CFS-full		All 98 countries with updates	546
ELR	Easterly, Levine, Roodman	ELR-full	1970-1997	All data in sample	586
	(2001). Updated version of	ELR-m3		3 wild observations excluded	583
	CFS, extended by ODA-data	ELR-56		Sample for the CFS-56 countries	330

Note: The number of observations, each covering a unit of 4 years, is n. The countries of each sample are listed in the Appendix. Note that ODA-55 has one country less than CFS-56 and ELR-56 as Somalia was deleted from the Penn World Tables.[a)] On the home page of Chang, Fernandez-Arias the data start in 1975, but Burnside and Dollar give series starting in 1970.

III.1 The aid data: ODA and EDA

The ODA-data are the net disbursements to LDCs of (nonmilitary) grants and loans with a grant element above 25% by official agencies of the members of the Development Assistance Committee (DAC) and certain Arab countries. Data are from World Development indicators (WDI, 2003). No less than n = 756 observations are available using a 4-year time unit.

The EDA-data are produced by individual researchers from the ODA series by weighting each loan or grant by an estimated gift element. The CFS-98 data set by Chang, Fernandez-Arias and Server (1998) is the first such set. The published sample covers the period 1975-93 for 133 LDCs, but thanks to missing GDP data the "effective" sample is 98. The CFS-56 of Burnside and Dollar (1997, 2000) is an early version of that set.[11] It includes 56 countries

11. We have used the CFS-98 from Burnside and Dollar to get as close to the original models as possible.

only as discussed. Thus 98-56 = 42 countries were excluded. Furthermore, more growth rates are now available so one more time unit of the CFS-data can now be used for the estimates.

Easterly, Levine and Roodman (2003) updates the CFS-data-set, so that observations are available for more countries and the period 1970-1997.[12] Due to reclassification of data, some variables are no longer available for all countries. Therefore the data set only grows to n = 586 observations. Further, the ELR data set for the first time unit 1970-73 and for last time unit 1994-1997 have been extrapolated from the correlation between EDA and ODA. This generates *three wild observations*. The most extreme is the aid/GDP-ratio of –12.73% for the Seychelles, 1970-3, which in the CFS-data it is no less than +19%. Two other wild observations are Guinea Bissau with –5.71% and –4.59% for Gambia. As the Seychelles had low growth in the following period, this observation makes a difference**.**

The average real growth rate of GDP per capita is calculated over 4 periods using local currency as in the other two data-sets. Initial GDP per capita is real GDP per capita in 1996 prices from the latest version of the Penn World Tables. For our aid variable, we use nominal ODA relative to nominal GDP as our aid.

The Appendix lists the countries of the 3 samples. We have tried to determine if the EDA-sample is skew relative to the full ODA-set of countries, but found no major skewness.

III.2 Are EDA or ODA data better as the dependent variable in the models analyzed

The ODA-data measures the gross resource flow, while the EDA-data considers only the net flow. We use both definitions in the replications for two reasons: It is unclear which the better variable is, and they are highly correlated anyhow (see table 3).

A rational expectations view of the Barro-Ricardo type suggests that only net grants affect the behavior of agents.[13] Thus the EDA-data are the proper ones. However, a large body of evidence suggests that politics has a very short time horizon.[14] This argues that the short-run gross resource flow determines behavior, and hence that the ODA-data are better. The argument can be supported by the observation that the LDC government deciding to accept the aid surely does so in order to undertake activities.

12. We are grateful for information from D. Roodman. It appears that the ELR-team decided not to make ad hoc adjustments, but to use the data generated by the procedure followed even if that led to some “strange” observations in the data set.

13. Barro (1974) is the original proposition, while Ricciuti (2003) surveys the ensuing discussion and empirical studies. The proposition has not been totally rejected, but it appears not to hold to more than to 25-50%.

Table 3. Simple correlation coefficients between measures of aid

	CFS	ELR	ODA
CFS	1	0.847	0.826
ELR	-	1	0.792
ODA	-	-	1

Note: Data are for the period 70-93.

Table 3 shows the correlation of the 3 measures. The lowest of the three is 0.79 between the ODA and the ELR-data, but this is only due to the 3 "wild" observations. The high correlations suggest that models using the different measures should reach qualitatively similar results. The average ratio between the ODA-data and the CFS-data (the pure EDA-data) is app. 2.4. This suggests that the Ω-points reached by the ODA-variable should be 2.4 times higher than to the EDA-coefficients if the same relation is estimated on the two data sets.

III.3 The good policy index

The good policy index, Γ, is from Dollar & Burnside (2000) that claim it is exogenous:

(5) $\Gamma = 1.28 + 6.85$ Budget Surplus -1.40 inflation $+2.16$ Trade Openness

The weights have been calculated from a growth regression including the three variables in the index as well as a number of variables. The device of combining the policy variables into a good policy index is appealing from the point of view of exposition and policy, as it can be used to say that only countries following Bank/Fund macroeconomic advice can be helped. However, it is obviously a rather arbitrary construct, which has been analyzed and criticized in several studies.[15] We use the index in our replications. Unfortunately, the 3 variables are not available for all years and countries for which we have aid and growth data, but we can expand samples with about 60% in replications of the Good Policy Model, relative to the original sample.

The reader should note that the good policy index gives a highly significant coefficient of about 1 in all regressions, where it is included. Good policies increase real growth by 1 percentage point. The problem discussed is if it interacts with or is independent of aid.

14. The literature mainly deals with DC's, see e.g. Paldam (2003), but it is likely to generalize.
15. See Dalgaard and Hansen (2001), Lensink and White (2000) and Easterly, Levine and Roodman (2003).

IV. Within-sample replications of the two models

Both models were originally estimated on the aid CFS-56 data (see table 2). They are published with references to a homepage with the data used, and the estimates are easy to replicate (we do not present the recalculations).[16] After the replication we simplified the models in the two ways discussed: We stripped the models down to the *minimal version* and replaced all controls with the fixed effects of the *base model*. These results are presented in the tables of this section.

We estimated the minimal models using OLS and heteroscedasticity consistent errors as in the original papers. We further estimated the fixed effects model with the within groups estimator and difference and system GMM estimators. The GMM estimators are consistent for fixed T and N going to infinity, which is not the case for the within groups estimator with fixed effects. We usually report the results from within groups estimation indicating in the text what difference it makes to use a GMM estimator. It turns out that only in one case is there a serious difference (see below).[17]

IV.1 The Good Policy Model

The model is given in table 1, which also lists the original **x** set of 7 controls.[18] The substantial results – μ_1, γ_0 and γ_1 – are almost independent of the last four controls, but they fall if any of the three first controls – x_1 to x_3, which have no time dimension – is deleted.

The variables x_4, x_5 and x_6 are made to catch the effect of civil disturbances and war. Such events are likely to reduce both growth and aid giving a bias in the estimates of the effect of aid. However, even when they are significant they do not affect the substantial coefficients of the models. This is in accordance with the findings of Brunetti (1998).

16. The original data are used in the within-sample replications even when some observations (e.g. the GDP-data) have been marginally revised. Also, t-ratios in the replications are "only" adjusted for heteroscedasticity and not for clustering, to get as close as possible.

17. In the case of the good policy model, we treat policy and aid as exogenous as in the original paper by Burnside and Dollar(2001) when using the GMM. Instruments for initial GDP are the second lag of GDP and all further lags. In the models in which aid lagged is included, we use similar instruments. Tests of over-identifying restrictions in all cases, accept the null and tests of serial correlation indicate that the residuals have the desired properties in all cases. For the ODA sample, we restrict the number of instruments to avoid singularity of the covariance matrix for the moments.

18. For easy reference they are: (x_1) institutional quality index (x_2) Africa dummy, (x_3) East Asia dummy, (x_4) political assassinations, (x_5) ethnical fractionalization, (x_6) x_4 times x_5 and (x_7) financial depth.

Table 4. The Good Policy Model estimated on CFS-56 data

Model	(1)	(2)	(3)	(4)
Aid data	CFS-56	CFS-56	CFS-56	CFS-56
Period	70-93	74-93 Lag	70-93	74-93 Lag
h_{it}, aid share - L	–0.01 (0.04)	0.27 (1.27)	0.32 (1.32)	0.69 (1.68)
Γ_{it}, good policy	**0.68 (3.63)**	**0.68 (2.85)**	**1.04 (3.58)**	**1.10 (4.28)**
$\Gamma_{it}h_{it}$, interacted - L	**0.18 (2.53)**	–0.02 (0.18)	–0.13 (0.99)	**–0.20 (2.11)**
y_{it}, gdp (GDP-level)	–0.65 (1.15)	–0.42(0.63)	–2.07 (1.55)	–2.47 (1.61)
x_1, institutions	**0.73 (4.26)**	**0.76 (3.86)**	not in	not in
x_2, Africa	**–2.09 (2.70)**	**–2.61 (3.29)**	not in	not in
x_3, Orient	**1.38 (2.46)**	**1.67 (3.61)**	not in	not in
Time dummies	yes	yes	yes	yes
Country dummies	not in	not in	yes	yes
Number of obs	270	234	267	230
R^2	0.39	0.36	0.53	0.55

Note: Bold indicates significance at the 5% level. L indicates that aid is lagged one time unit, in columns (2) and (4). Our panel regressions need 2 observations for each country so 3-4 observations cannot be used. Brackets contain t-statistics.

The results are given in table 4. Column (1) gives virtually the same results as in the original article.[19] Column (3) shows what happens if the 3 specific controls for country differences are replaced with fixed effects. Here all substantial effects disappear and signs even change. Consequently, we know precisely what it is that drives the substantial results of the model. It is the country differences that are *not* controlled for by the institutional quality index, the Africa dummy and the East Asia dummy. We find this unconvincing.

The Good Policy Model is uncontrolled for reverse causality.[20] We argued above that the most tidy procedure is to lag aid as done in column (2) and (4) of the table. This turns the coefficients to the interaction term more negative and in column (4) it is even significantly negative. The reader may ask if (1) or (4) is the most reasonable model, and consequently if the “true” interaction term is +0.18 or –0.20. Re-estimating the fixed effects good policy model with the GMM difference estimator, we obtain similar results, with the difference that aid when being lagged now is significant in the model in column 4 of table 4. Thus the Good Policy Model is a fickle construct.

19. It also states that 5 observations were deleted for being too extreme. We have followed this procedure. The inclusion of these observations reduces the significance, but it does not change the results very much.
20. It was controlled for by 2SLS-estimation in the working paper, but the instruments were not convincing.

IV.2 The Medicine Model

The Medicine Model turns out to be easy to reproduce on the CFS-56 data. It is fairly robust to the controls, but it needs either a 2SLS-estimate or a lag. Table 5 shows results of OLS-estimates – for the model looking most like the ones of table 4, for easy comparability.

Table 5. The Medicine Model estimated on CFS-56 data

Model	(1)	(2)	(3)	(4)
Aid data	CFS-56	CFS-56	CFS-56	CFS-56
Period	70-93	74-93 L	70-93	74-93 L
h_{it}, aid share - L	0.28 (0.70)	**0.87 (2.34)**	0.50 (0.86)	**1.32 (2.32)**
h_{it}^2, aid squared - L	–0.02 (0.31)	**–0.065 (2.26)**	–0.04 (0.81)	**–0.12 (2.81)**
y_{it}, gdp	–0.59 (1.05)	–0.39 (0.59)	–2.03 (1.47)	–2.13 (1.48)
x_1, institutions	**0.89 (4.77)**	**0.98 (4.74)**	not in	not in
x_2, Africa	**–2.29 (3.01)**	**–2.91 (3.65)**	not in	not in
x_3, Orient	**2.54 (4.78)**	**2.99 (5.10)**	not in	not in
Time dummies	yes	yes	Yes	yes
Country dummies	not in	not in	Yes	yes
Number of obs	270	234	267	269
R^2	0.31	0.32	0.49	0.52

Note: See note to table 4.

The coefficients to the three controls are much the same as before, but now they can be replaced by the fixed effect without much change to the two substantive effects: μ_1 to h_{it-1} and μ_2 to h_{it-1}^2. The size of the two effects reported by Dalgaard and Hansen (2001) using 2SLS-estimation and a larger set of controls are 1.35 to aid and –0.13 to aid squared, so our base model replication (4) is very close. We consequently use that model for the out-of-sample replications, as it can be replicated on all available data. The key finding from table 5 is that both substantive coefficients μ_1 and μ_2 to aid and aid squared are fairly stable. Clearly, the Medicine Model is far superior to the Good Policy Model when it comes to robustness in the within-sample replications.[21]

21. Lensink and White (2001) use a different set of controls and use ODA data. The two main new controls are the debt share with a negative coefficient and enrolment in secondary school with a negative coefficient (!) as well. Estimating this extended model on the CSF-56 or CSF-98 data, we reach similar conclusions except that the human capital indicator turns out to be insignificant.

Finally, it is worth pointing out that when the calculated parables from the 4 estimates are drawn – as sketched on figure 1 – they all look similar with a Ω-point of between 5 and 7%. The one for the model in column (4) is included as the quadratic curve on figure 2a below. Re-estimating the equation in column (4) with the GMM-estimators, the aid terms are still significant with the right signs.

V. Out-of-sample replications of the two models

We now want to replicate the two models on the remaining 70% of the data. This is most difficult for the Good Policy Model. Here we base the replications on the models in columns (1) and (3) in table 4. For the Medicine Model we use column (4) in table 5 for the replications. It allows us to use all available aid data in the replications.

V.1 Replications on the full CFS-data set

The CFS-data contains 42 countries not included in the CFS-56 data, and more years have been added to the growth data, so we are able to replicate both models on more data.

Table 6. The Good Policy Model estimated on the CFS data

Model, equal to	(1) = (t4,1) [a)]	(2)	(3) = (t4,3) [a)]	(4)
Aid data	CFS-56	CFS-62	CFS-56	CFS-69
Period	70-93	70-93	70-93	70-93
h_{it}, aid share	–0.01 (0.04)	0.05 (0.46)	0.32 (1.32)	0.12 (0.66)
Γ_{it}, good policy	**0.68 (3.63)**	**0.84 (3.37)**	**1.04 (3.58)**	**1.12 (4.31)**
$\Gamma_{it}h_{it}$, interacted	**0.18 (2.53)**	0.06 (0.94)	–0.13 (0.99)	–0.07 (1.33)
y_{it}, gdp	–0.65 (1.15)	–0.08 (0.17)	–2.07 (1.55)	**-2.82(2.27)**
x_1, institutions	**0.73 (4.26)**	0.27 (1.78)	not in	not in
x_2, Africa	**–2.09 (2.70)**	–0.12 (1.73)	not in	not in
x_3, Orient	**1.38 (2.46)**	**1.84 (2.81)**	not in	not in
Time dummies	yes	yes	yes	yes
Country dummies	not in	not in	yes	yes
Number of obs	270	307	267	337
R^2	0.39	0.30	0.49	0.46

Note: See note to table 4. (2) and (4) are not cleaned for outliers. [a)] Column "(1) = (t4,1)" means equal to table 4 column (1) , and column "(3) = (t4,3)" means table 4, column (3).

Table 7. The Medicine Model estimated on CFS-data

Model, equal to	(1) = (t5,4)	(2)	(3)
Aid data	CFS-56	CFS-42	CFS-full
Period	74-93	74-97	74-97
h_{it-1}, aid share	**1.32 (2.32)**	0.26 (1.17)	**0.60 (2.95)**
h_{it-1}^2, do squared	**–0.12 (2.81)**	**–0.02 (2.53)**	**–0.035 (3.81)**
y_{it}, gdp	–2.13 (1.48)	**–0.78 (3.48)**	**–2.41 (2.40)**
Time dummies	yes	yes	**yes**
Country dummies	yes	yes	**yes**
Number of obs	269	216	546
R^2	0.52	0.38	0.43

Note: See note to table 4.

Table 6 shows the results for the Good Policy Model. Unfortunately, neither the good policy index nor the index for the quality of institutions is available for all the additional CFS observations, but the sample still expands with about 20%. Clearly, the model does not replicate.

The replication of the Medicine Model is presented in table 7. Column (2) shows what happens if the estimate is replicated on the "unmined" CFS-42 data. The quadratic term is still significant, but it is much smaller, and the coefficient to aid unsquared is now insignificant. If it is disregarded, aid is harmful at any level. If it is included the Ω-point is 6.5.

Column (3) presents the estimate for all 98 countries and all years now available. The result is precisely as expected from column (1) and (2), Both coefficients are significant due to the original 56, but only half as large as before, due to the added observations. Thus in this sample, we still get some evidence in favor of the Medicine Model, but the Ω-point moves to 8.5. Using GMM-estimators make little difference to the results.

V.2 Replications on the ELR and ODA-data sets

These data sets are larger than the CFS data set. This should allow us to reach higher levels of significance if either model replicates, but the results are much weaker for both models.

Table 8 holds the replications of the Good Policy Model. Due to lack of data for the good policy index and the institutional quality index we "only" manage to do our replications with about 400 observations, but the results all fail to support the model. The interacted term, $\Gamma_{it}h_{it}$, is insignificant throughout. We have also – unsuccessfully – tried to replicate the Good Policy Model on ELR-56 and ODA-55 data, which covers the 56 countries of the CFS-56 data set, but for more years. The results are parallel to what Easterly et al. (2003) found, and we have added the additional (negative) evidence of the ODA data set.

Table 8. The Good Policy Model estimated on ELR- and ODA-data

Model	(1)	(2)	(3)	(4)	(5)	(6)
Aid data	ELR-full	ELR-m3	ODA-full	ELR-full	ELR-m3	ODA-full
Period	70-97	70-97	66-97	70-97	70-97	66-97
h_{it}, aid share	0.02 (0.16)	0.012 (0.10)	0.01 (0.35)	0.18 (1.09)	0.18 (0.92)	0.0015 (0.03)
Γ_{it}, good policy	**0.77 (3.86)**	**0.78 (3.66)**	**0.89 (4.51)**	**0.88 (3.28)**	**0.88 (3.29)**	**1.06 (4.11)**
$\Gamma_{it}h_{it}$, interacted	0.07 (1.05)	0.07 (0.96)	–0.00 (0.27)	0.03 (0.25)	0.03 (0.25)	–0.02 (1.36)
y_{it}, gdp	–0.17 (0.41)	–0.18 (0.42)	–0.66 (1.62)	–1.08 (1.18)	–1.08 (1.13)	**–2.46 (2.43)**
x_1, institutions	0.21 (1.66)	0.21 (1.65)	**0.90 (4.51)**	not in	not in	not in
x_2, Africa	–1.19 (1.92)	–1.18 (1.89)	**–1.54 (2.64)**	not in	not in	not in
x_3, Orient	**2.19 (3.84)**	**2.18 (3.66)**	**1.77 (3.74)**	not in	not in	not in
Time dummies	yes	yes	yes	yes	yes	yes
Country dummies	not in	not in	not in	yes	yes	yes
Number of obs	380	379	397	413	412	427
R^2	0.30	0.30	0.33	0.41	0.42	0.50

Note: See note to table 4.

Table 9. The Medicine Model estimated on ELR- and ODA-data

	ELR data (EDA)		ODA-data
Model	(1)	(2)	(3)
Aid data	ELR-full	ELR-m3	ODA-full
Period	73-97	73-97	66-01
h_{it-1}, aid share	**0.21 (2.58)**	0.18 (0.62)	0.095 (1.62)
h_{it-1}^2, do squared	–0.003 (0.48)	0.001 (0.07)	–0.001 (1.26)
y_{it}, gdp	**–3.04 (3.48)**	**–3.13 (3.09)**	**–2.76 (3.51)**
Time dummies	yes	yes	Yes
Country dummies	yes	yes	Yes
Number of obs	586	583	755
R^2	0.43	0.43	0.47

Note: See note to table 4.

Table 9 shows the results for the Medicine Model. Our base model allows us to use all observations available. The quadratic term fails in all regressions, and aid un-squared fails in all but one regression. It is the full ELR-data set, but it is due to the 3 "wild" observations that probably should be deleted. When they are deleted the term fails. For these samples, it makes no difference to use GMM-estimators. We never get a positive and significant aid-policy term.

The ODA sample covers a longer period and includes 110 countries. Here the linear and the quadratic term are both insignificant when using OLS, though they have the same signs as in the CFS-56 data set. For the GMM difference estimator, the terms are not significant at the 5% level. With the system estimator, only the linear term is significant, whereas the squared term fails. Excluding the 55 original countries, this result no longer holds.

We have also replicated the results for the ELR-56 and the ODA-55 data set for the countries of the CFS-56 data, but for more years (regressions are not included). The results are once again insignificant, but the results for ODA-55 are close to the ones of Hansen and Tarp (2000) approaching significance at the 10% level for aid un-squared. However, the extra year added is enough to make significance fall below the 10% level.[22]

VI. The form and significance of the general aid term

We now replace the arbitrary parametric form for the aid-growth relation with, (4) $g_{ti} = \Phi(h_{it-j}) + \boldsymbol{\alpha}'[\mathbf{D_i}, \mathbf{D_t}] + \beta y_{it} + u_{it}$, where $\Phi(h_{it-j})$ can take any continuous form. First the method will be introduced, then the results are presented and finally a few concluding remarks are added.

VI.1 A semi-parametric term in a panel regression with fixed effects[23]

The technique approximates $\Phi(h)$ by a weighted sum of continuous functions, with estimated weights. The functions are cubic splines with four equidistanced knots in an interval on the real axis chosen such that all data can be included. The fixed effects for countries are treated as usual. We estimated the relationship with both OLS and GMM-estimators.[24]

Each regression produces a "normal" set of coefficients to the linear terms and a graph for the aid term. The graphs show the semi-parametric aid-growth relation and its point-wise 95% confidence bands, which are wider, where there are few observations. It also includes the

22. When we use the controls of Lensink and White (2001) we can only replicate the model for n = 601, for ODA-data and n =520 or N=518 when excluding wild observations for the ELR-data. In all three cases the aid squared term and aid unsquared fail.

23. The method is explained in Gørgens, Paldam and Würtz (2003), which also refers to the proofs. The ACH-test is from Aerts, Claskens and Hart (1999).

24. The GMM-estimates are very similar to the OLS-estimates, but much more imprecise. There are however certain problems with the estimation: 1) instruments need to be dropped 2) Because of singularities in certain matrices, the ACH-test cannot be computed 3) two-step estimates and system estimates are not available, because of similar problems as mentioned under 2).

fitted values from a linear regression and the relevant aid squared models referred to.

The Φ(h)-term is tested by two ACH specification tests: ACH test 1 compares the model estimated with a null of a model with no aid term. The critical values used are asymptotic values from Hart (1997). If we find evidence of a relationship, we go on to the second test: ACH test 2 tests the null of the linear model against a general nonlinear alternative.

As the output for each regression includes a bulky graph, we only present the results for four main cases: The original CFS-56 sample, the CFS-98 data set, the ELR and the ODA sample. In addition we add the regression on the more reasonable ELR-m3 data.

VI.2 Results: Main table and discussion of the results based on CFS-data

The 5 AHC (1) tests in table 10 tell a sad story of insignificance. The only marginally signify-cant result for the aid term is at the 10% level. As expected it is for the CFS-56 sample, when it is extended the test fails. However, both CFS-regressions reject the model with the linear term only against a general nonlinear alternative at the 5% level. Furthermore, we note that the t-tests in the quadratic model and the ACH-tests disagree as will be discussed in VI.4.

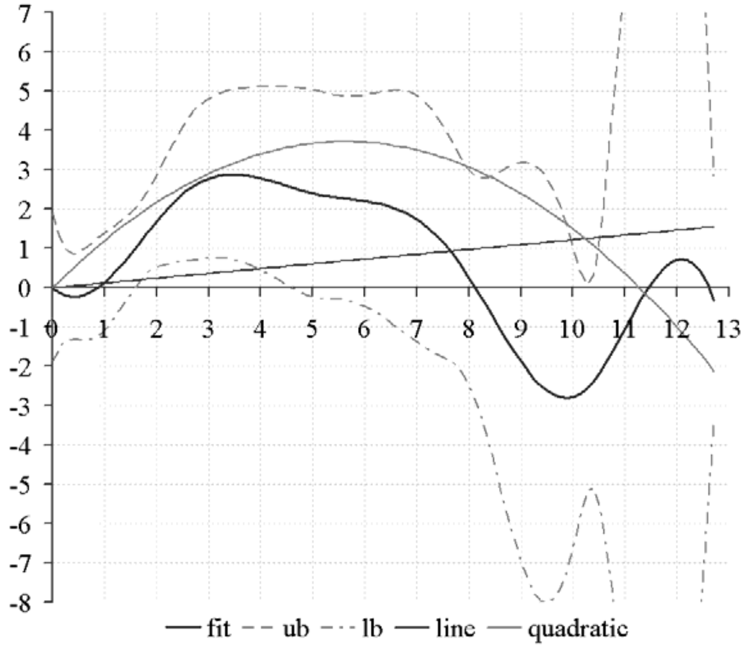
The Φ(h)-shapes on figures 2a and b both have a positive section for aid shares between 1% and 8%, but they do move very differently after 10%, though both eventually turn negative. The two significance bounds suggest that both curves have a positive peak between 3% and 5%, but this is a dubious conclusion given that the Φ(h)-shape as such is insignificant.

Table 10. The semi-parametric model estimated on 5 data-sets

	EDA-data				ODA-data
Model / Corresponds to	(1) / (t5,4)	(2) / (t7,3)	(3) / (t9,1)	(4) / (t9,2)	(5) / (t9,3)
Aid data	CFS-56	CFS-98	ELR-full	ELR-m3	ODA-full
Period	74-93	74-97	74-97	74-97	66-01
Φ-term to aid	Fig 2a	Fig 2b	Fig 3	Not given	Fig 4
y_{it}, gdp	–2.32 (1.66)	**–2.61 (2.48)**	**–3.31 (3.17)**	**–3.11 (2.97)**	**–2.59 (3.27)**
Time dummies	yes	yes	Yes	yes	yes
Country dummies	yes	yes	Yes	yes	yes
ACH-test 1 for aid term	3.27 [a)]	2.94 [b)]	2.75 [c)]	1.88	1.72
ACH-test 2 for not linear	**5.58**	**4.47**	n/a	n/a	n/a
Number of obs	269	546	586	583	756
R^2	0.53	0.43	0.44	0.44	0.47

Note: See note to table 4. The critical values for the ACH-test are 4.18 (5% level) 3.22 (10% level). In (a) and (b) the t-tests of both aid and aid squared are significant in the corresponding parametric regression. For (c) only the aid term is significant in the corresponding regression.

Figure 2a. Aid-term in the base model on the CFS-56 data, n = 269



Note: The size of the graph is marked by a box on the other graphs. The upper and lower 95% bounds of the fit are “ub” and “lb”.

Figure 2b. Aid-term in the base model on the CFS-98 data, n = 546

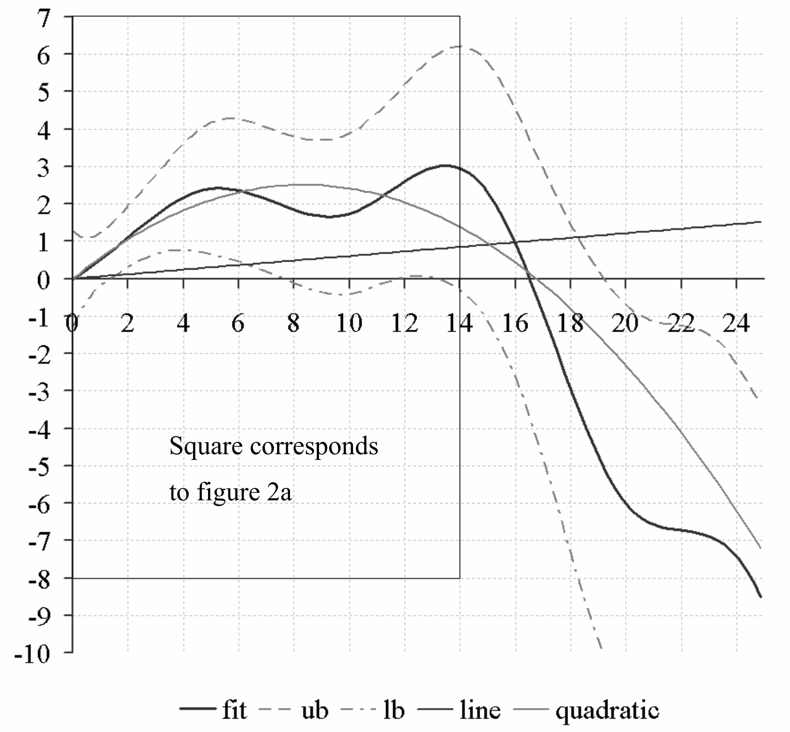
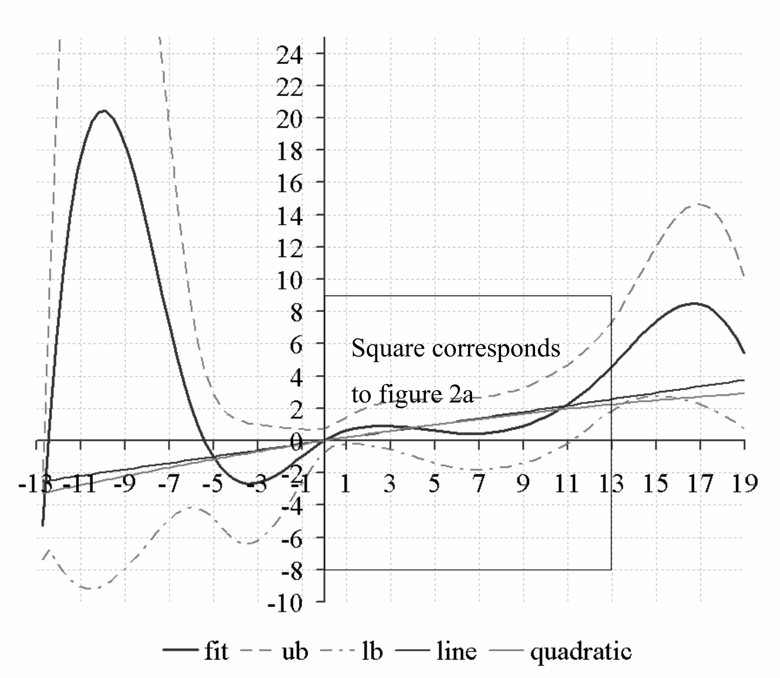
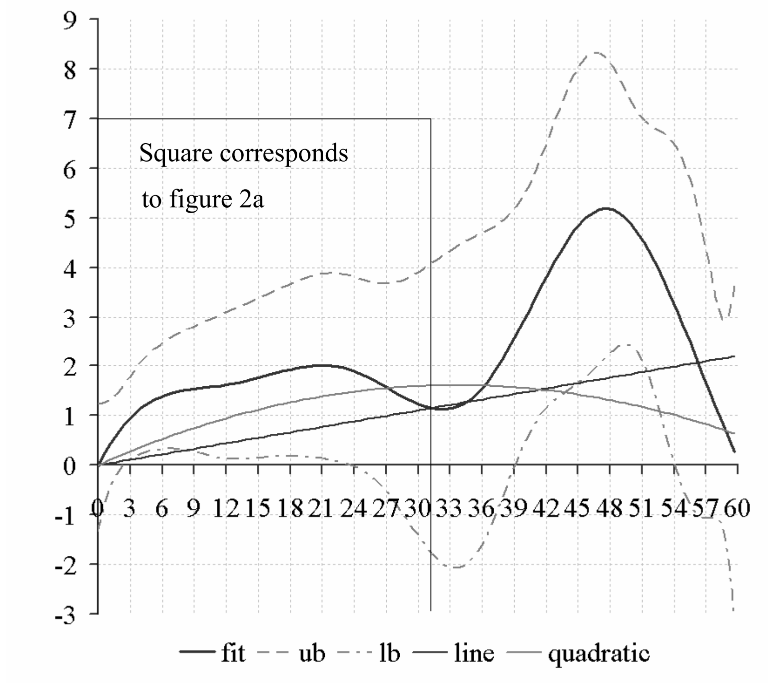


Figure 3. Aid-term in the base model on the ELR-full data, n = 586



Note: The "crazy" and insignificant peak at –10 is due to the 3 "wild" observations.

Figure 4. Aid-term in the base model on the ODA-full data, n = 756



Note: The aid-axis of the box showing the section corresponding to figure 2a is multiplied by 2.4.

VI.3 Results based on ELR and ODA-data

For the ELR data set, we get a strange shape (due to the 3 "wild" observations) suggesting that countries which are repaying debt rather than receiving aid get a lot of growth. However, the ACH-test rejects the relationship between aid and growth. The coefficient on the linear aid term is significant by the t-test, when all observations are included, but rejected when the 3 "wild" observations are removed from the data set. Thus it appears that the ACH-test is less sensitive to the wild observations than the t-statistics. Using the ELR-56 subset, we also find evidence of no relationship. This case does not include the wild observations.

Finally, for the full ODA sample we get a strange two-humped curve. However, the relationship is insignificant. This is also the case when we use only the 55 countries from the DB-56 set. From a visual inspection of the 4 figures and additional ones not presented.[25]

A common trait of the estimated relationships is that they all have a positive section at low levels of aid, and many but not all of the curves have a negative tail as in the CFS data. However, these results are rejected by the tests – mostly rather decisively.[26]

VI.4 A statistical comment: The disagreement of the tests

The ACH-tests in table 10 and the t-tests in the matching parametric regressions disagree in three out of 5 cases (see notes to table). This is puzzling, but it is possible as both are asymptotic tests.

Consider first columns (1) and (2). We here supplement the ACH-test 1 with the ACH-test 2, which has the linear model as the null. It rejects the linear model in columns (1) and (2) like the t-test. Thus it is possible to achieve significant results using t-statistics with coefficients that go both ways, while the ACH-test shows that the model as such is not improving. In column (3) the 3 "wild" observations give a significant coefficient with the t-test, but not with the ACH-test. Thus the ACH-test is less sensitive to outliers than the t-test.

We conclude that the ACH-test 1 on the generalized aid-term is the proper way to test if aid affects the growth rate.

---

25. We have run the semi-parametric regressions for all cases given in tables 5, 7 and 9. The results for cases not included are much as could be expected.

26. The models of table 10 have been used for several experiments. Firstly, we included the controls of Lensink and White (2001). They improved the fit of the aid term marginally: The null of no relationship is rejected at the 5% level using the ACH-tests in the ODA sample, neither is the linear model rejected. The coefficient to lagged aid is 0.061 and it is significant at the 5% level from the t-statistic. For the ELR-sample, the aid-term is still insignificant. If both aid and the debt-GDP ratio are lagged, all results are as in the table. If the debt-GDP ratio is endogenous to growth, the lagged value seems more appropriate. Secondly, we included the domestic savings ratio. It failed for all aid-data sets, and made aid insignificant in the regressions.

VII. Conclusions: Weak results and the "do no harm" criterion

After the gloomy results of the macro literature on aid effectiveness from its start in the 1950s till the mid 1990s two optimistic models appeared: The Good Policy Model where aid helps in countries with governments that pursue sound economic policies, and the Medicine Model where aid helps up to a point after which it turns harmful.

The papers presenting both theories are written after a thorough examination of a data set that covers only about 30% of available evidence. Our paper has studied the robustness of the models within the sample and if they replicate in the remaining 70% of the data. Even in the within-sample study the Good Policy Models prove fickle, while the Medicine Model is remarkably robust. However, in the out-of-sample replications both models fail. What is even worse is that a generalized aid-term proves insignificant in the large data sets available.

Our findings are thus consistent with the possibility that the recent discussion of aid effectiveness builds upon the mining of a fluke in a particular subset of the data.

Thus the results are rather negative to the family of models examined, and they revive the micro-macro paradox of aid. One may argue that growth is not the only goal of aid and maybe it can be demonstrated that some of the other goals are better reached. Also, we have not, of course, rejected all other models (see note 21).

However, we have found no evidence that moderate aid harms growth and the poverty of the poor countries is a terrible malady, so perhaps we should heed the advice Hippocrates gave to the medical profession 2500 years ago (in Epidemics, Bk. I, Sect. XI): "... *to help, or at least to do no harm*."

References:

Aerts, M., Claeskens, G., Hart, J.D., 1999. Testing the Fit of a parametric Function. *Journal of the American Statistical Association* 94 (447), 869-879

Barro, R.J., 1974. Are Government Bonds Net Wealth? *Journal of Political Economy* 82, 1095-1117

Barro, R.J., 1997. *Determinants of Economic Growth. A Cross-Country Empirical Study*. MIT Press: Cambridge, MA

Brunetti, A., 1998. Political Variables in Growth Regressions. Cpt 6, 117-135 in Broner, S., Paldam, M., eds., *The Political Dimension of Growth*. Macmillan for the IEA: Houndmills, UK

Burnside, C., Dollar, D., 1997, 2001. Aid, Policies and Growth. *American Economic Review* 90: 847-868. Earlier in Policy Research Working paper 1777, The World Bank, Development Research Group, Washington, DC

Card, D., Krueger A.B., 1995. Time-Series Minimum-Wage Studies: A Meta-Analysis. *American Economic Review,* 85, 238-43

Cassen, R., 1986, 1994. *Does Aid Work?* Clarendon: Oxford

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

Collier, P., Dehn, J., 2001. Aid, Shocks and growth, World Bank Policy Research Working paper 2688

Collier, P., Hoeffler, A., 2002. Aid, policy and growth in Post-conflict societies, World Bank Policy Research Working paper 2902

Dalgaard, C.J., Hansen, H., 2001. On Aid, Growth and Good Policies. *The Journal of Development Studies* 37, 17-41

Doucouliagos, C. and Laroche, P. (2003) Publication Bias in Industrial Relations Research, Proceedings of the 14th Association of the Industrial Relations Academics of Australia and New Zealand (AIRAANZ) Conference, Melbourne, 4-7 February

Easterly, W., Levine, R. Roodman, D., 2003. New data, New doubts. NBER Manuscript paper 9846

Guillaumont, P., Chauvet, L., 2001. Aid and Performance: A reassessment. *Journal of Development Studies* 37, 66-92

Guillaumont, P, Chauvet L, 2002. Aid and growth revisited: policy, economic vulnerability and political stability

Gørgens, T., Paldam, M., Würtz, A., 2003. How Does Public Regulation Affect Growth? Working paper, Aarhus University

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

Hansen, H., Tarp, F., 2000. Aid Effectiveness Disputed. In F. Tarp and P. Hjertholm, eds. *Foreign Aid and Development. Lessons Learnt and Directions for the Future*. Routledge Studies in Development Economics: London

Hansen, H., Tarp, T., 2001. Aid and growth regressions. *Journal of Development Economics* 64, 547-570

Hart, J.D., 1997. *Nonparametric Smoothing and Lack-of-Fit tests.* New York: Springer Verlag

Heston, A., Summers, R., Aten, B., 2002. Penn World Tables Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP)

Jepma, C.J., 1995. *On the Effectiveness of Development Aid*. Manuscript, Den Haag

Keefer, P., Knack, S., 1995. Institutions and Economic Performance: Cross-country Tests Using Alternative Institutional Measures. *Economics and Politics* 7, 207-227

Lensink, R., White, H., 2000. Assessing Aid: A Manifesto for Aid in the 21st Century? *Oxford Development Studies* 28, 5-17

Lensink, R., White, H., 2001. Are There Negative Returns to Aid? *Journal of Development Studies* 37, 42-65

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

Paldam, M., 1997a. *Dansk u-landshjælp. Altruismens politiske økonomi*. Aarhus UP: Aarhus, Denmark

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

Paldam, M., 2003, Are Vote and Popularity Functions Economically Correct? Essay 3, 49-59 in Rowley, C.K., Schneider, F., eds., *The Encyclopedia of Public Choice, Vol I.* Kluwer Academic Publishers, Dordrecht.

Ricciuti, R., 2003. Assessing Ricardian Equivalence. *Journal of Economic Surveys* 17, 55-78

Roodman, D., 2003. The Anarchy of Numbers: Aid, Development and Cross-country Empirics, Center for Global Development Working paper 32

Sala-i-Martin, X.,1997. I Just Ran Two Million Regressions, American Economic Review 87, 178-83

Svensson, J. ,1999. Aid, Growth and Democracy. *Economics and Politics* 11, 275-297

White, H., 1992. The macroeconomic impact of development aid: A critical survey. *Journal of Development Studies* 45, 155-163

WDI, 2003. *World Bank Indicators* 2003, CD-ROM

World Bank, 1999. *Assessing Aid: What Works, What Doesn't and Why*. Oxford UP: Oxford, UK

Appendix table 1 of 2: Countries included in samples

	CFS-56	CFS-full	ELR-full	ODA-full		CFS-56	CFS-full	ELR-full	ODA-full
Albania				I	Fiji		I	I	I
Algeria	I	I	I	I	Gabon	I	I	I	I
Angola		I	I	I	Gambia	I	I	I	I
Antigua & Barbuda				I	Ghana	I	I	I	I
Argentina	I	I	I	I	Grenada		I	I	I
Armenia				I	Guatemala	I	I	I	I
Bangladesh		I	I	I	Guinea		I	I	I
Barbados		I	I	I	Guinea Bissau		I	I	I
Belize		I	I	I	Guyana	I	I	I	I
Benin		I	I	I	Haiti	I	I	I	I
Bhutan			I		Honduras	I	I	I	I
Bolivia	I	I	I	I	Hong Kong			I	I
Botswana	I	I	I	I	Hungary			I	I
Brazil	I	I	I	I	India	I	I	I	I
Bulgaria			I	I	Indonesia	I	I	I	I
Burkina Faso		I	I	I	Iran		I	I	I
Burundi		I	I	I	Iraq			I	
Cambodia				I	Israel				I
Cameroon	I	I	I	I	Jamaica	I	I	I	I
Cape Verde		I	I	I	Jordan		I	I	I
Central African Rep.		I	I		Kenya	I	I	I	I
Chad		I	I	I	Korea	I	I	I	I
Chile	I	I	I	I	Lao PDR			I	
China		I	I	I	Lebanon				I
Colombia	I	I	I	I	Lesotho		I	I	I
Comoros		I	I	I	Liberia		I	I	
Congo, D.R. (Zaire)	I	I	I	I	Macao				I
Congo, Rep.		I	I	I	Madagascar	I	I	I	I
Costa Rica	I	I	I	I	Malawi	I	I	I	I
Cote d'Ivoire	I	I	I	I	Malaysia	I	I	I	I
Croatia				I	Mali	I	I	I	I
Cyprus				I	Malta		I	I	
Czech Rep.			I	I	Mauritania		I	I	I
Dominica				I	Mauritius		I	I	I
Dominican Rep.	I	I	I	I	Mexico	I	I	I	I
Ecuador	I	I	I	I	Mongolia			I	
Egypt	I	I	I	I	Morocco	I	I	I	I
El Salvador	I	I	I	I	Mozambique		I	I	I
Equatorial Guinea				I	Myanmar		I	I	
Ethiopia	I	I	I	I	Namibia				I

Appendix table 2 of 2: Countries included in samples

	CFS-56	CFS-full	ELR-full	ODA-full		CFS-56	CFS-full	ELR-full	ODA-full
Nepal		I	I	I	Somalia	I	I	I	
Nicaragua	I	I	I	I	Sri Lanka	I	I	I	I
Niger	I	I	I	I	St. Kitts & Nevits		I	I	I
Nigeria	I	I	I	I	St. Lucia		I	I	I
Oman		I	I		Sudan		I	I	
Pakistan	I	I	I	I	Suriname			I	
Panama		I	I	I	Swaziland		I	I	I
Papua New Guinea		I	I	I	Syria	I	I	I	I
Paraguay	I	I	I	I	Tanzania	I	I	I	I
Peru	I	I	I	I	Thailand	I	I	I	I
Philippines	I	I	I	I	Togo	I	I	I	I
Poland			I	I	Tonga		I	I	
Romania			I	I	Trindidad & Tobago	I	I	I	I
Russian Federation			I	I	Tunisia	I	I	I	I
Rwanda		I	I	I	Turkey	I	I	I	I
Samoa		I	I	I	Uganda		I	I	I
Saudi-Arabia			I		Ukraine				I
Sct. Vincent & Grenadines		I	I	I	Uruguay	I	I	I	I
Senegal	I	I	I	I	Vanuatu		I	I	I
Seychelles		I	I	I	Venezuela	I	I	I	I
Sierra Leone	I	I	I	I	Yemen				I
Singapore			I	I	Zambia	I	I	I	I
Solomon Islands		I	I		Zimbabwe	I	I	I	I

Note: The letter "**I**" indicates inclusion of a country in the sample. Two observations from Sao Tome and Principe have been excluded as they are so extreme in the ODA sample that they cause perfect colinearity when using the semi-parametric estimator with four equidistanced knots.