PUCH 3237-S

6th June, 2005

Can the two new aid-growth models be replicated?

Peter Sandholt Jensen and Martin Paldam

Department of Economics, University of Aarhus (Bldng 322), DK-8000 Aarhus C, Denmark.

Phone for PSJ +45 8942 1560 and MP +45 8942 1607.

E-mail <psjensen@econ.au.dk> and <mpaldam@econ.au.dk>.

Published in *Public Choice* 127, 147–175 (2006)

Abstract: Recent aid effectiveness literature centers on two competing models from the family of conditional models: The Good Policy Model, where the key feature is policy times aid, and the Medicine Model, where it is aid squared. Both models were reached on a sample of 1/3 of the available data. The models are simplified to be replicatable on more of the data. Within-sample the Good Policy Model proves fragile, while the Medicine Model is more robust. Both models fail in out-of-sample replications. A semi-parametric technique is used to test for an unknown functional form of the aid-growth term. It rejects that aid is statistically significant.

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

Jel: C14, C23, F35, O4

Acknowledgement: We are grateful to P. Harms, H. Doucouliagos and other discussants when the paper was presented at EPCS-04 in Berlin, and to D. Roodman for information on the ELR-data. We also want to thank the careful referee.

I. Introduction: Two models from the "interaction" family

During the last 5 years the aid effectiveness discussion has been dominated by a new *family* of models, where development aid works on a certain condition, so that an aid interaction term is crucial. We analyze the two most important models from this family. Both have few *substantive 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 main point in favor of the first model is that it tallies well to the intuition of practitioners, while the second model fits the data better:

The Good Policy Model (see section II.1) is the most influential. It claims that aid gives policies an extra push. Good economic policies become better and bad ones worse. Therefore, aid should be concentrated on countries with good policies. Several donors (notably the World Bank) have modified their aid policies somewhat in accordance with this advice.

The Medicine Model (of section II.2) includes both aid with a positive sign and aid squared with a negative sign. Aid helps all countries, but only up to a point, h^*. More aid is increasingly harmful. Consequently, aid should be distributed proportionally to GDP and never exceed the optimal dose (h^*). This model has been used as a general defense for aid.

Both models have recently been pointed to as the most influential aid effectiveness models (see Hudson, 2004), and they have been disseminated by a development agency, see World Bank (1998) and Tarp and Hjertholm (2000). It is thus important to examine their empirical support; in particular as there is a large problem: 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 data for aid and growth. Consequently, it is important to study if the models replicate on the remaining 70% of the data. This is what we do at present.

One reason why the models use so little of the data is that the authors control the models for many *potentially* relevant effects – whether or not these variables are *actually* relevant for the key results. Few of the controls are available for all countries and years available. Thus we can only replicate the original models for slightly more than the CFS-56 data. However, we have developed two simplified versions of both models that can be replicated on (much) more of the data: One simplification is reached by stripping the models down to the ***minimal versions*** needed for generating its key finding. Another simplification is the ***base versions*** reached by replacing the controls with fixed effects for countries.

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 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 semi-parametric technique, which 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 newest and most comprehensive survey of the aid effectiveness literature is a set of meta-studies (Doucouliagos and Paldam, 2005). They cover 97 papers, by 104 authors, studying 143 models belonging to 3 families: (A) 45 aid to accumulation models, (B) 68 aid to growth models, and (C) 30 aid to growth conditional on a third variable – to which the two models examined belong. The results support all possible results.[1] However, when all studies are summed up by the tools of meta-analysis the result is that even when the effect of aid on growth is positive, it is small and statistically insignificant.

In the language of growth theory we can thus say that absolute aid effectiveness is rejected by the literature. The two models analyzed claim that there is conditional aid effectiveness. This is the important claim we examine.

Section II surveys the two models, our method and the logic of our two simplified model versions. 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, while 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.

II. The two models and the two simplified versions of each

The variables and models discussed are listed in table 1. Subsections II.1 and II.2 briefly present the two models, then II.3 looks at the great variability of aid effectiveness results and the moral hazard problem of growth regressions. Finally II.4 discusses the choices of controls in the two simplified versions used in the robustness tests.

1. The largest family (B) of studies reaches the following count: (i) 46% of the papers conclude that aid increases growth, (ii) 45% find no effect on growth, while (iii) 9% conclude that aid is harmful to growth.

Table 1. Variables and models discussed

i	country index	z_{it}	Good Policy Index
t	time index, one unit is 4 years	Y_{it}	GDP in PPP terms, start of unit
g_{it}	real growth rate, average of 4 years	y_{it}	initial log(Y_{it} per capita) for each period
g'_{it}	excess growth due to aid	$\boldsymbol{d_t}$	fixed effects for time, in all models
h_{it}	aid in percent of GDP, same average	$\boldsymbol{D_i}$	fixed effects for counties
*	maximum (g'*, h*) in Medicine Model	$\boldsymbol{x_{it}}$	specific controls for irrelevant variation
$\Phi(h_{it})$	generalized aid term	$\boldsymbol{r_i}$	necessary controls, $\boldsymbol{r_i} \subset \boldsymbol{x_{it}}$

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)	$g_{ti} = \mu h_{it-j} + \beta y_{it} + \boldsymbol{\alpha' x_{it-j}} + u_{it}$	Main idea
	Good Policy Model (Lag L = 0, 1)	Versions:
(2a)	$g_{ti} = \mu h_{it-L} + \delta z_{it} + \omega z_{it} h_{i-Lt} + \beta y_{it} + \boldsymbol{\alpha'(x_{it}, d_t)} + u_{it}$	Original
(2b)	$g_{ti} = \mu h_{it-L} + \delta z_{it} + \omega z_{it} h_{it-L} + \beta y_{it} + \boldsymbol{\alpha'(r_i, d_t)} + u_{it}$	Minimal version
(2c)	$g_{ti} = \mu h_{it-L} + \delta z_{it} + \omega z_{it} h_{it-L} + \beta y_{it} + \boldsymbol{\alpha'(D_i, d_t)} + u_{it}$	Base model
	Medicine Model (Lag L = 0,1), with maximum (g'*, h*)	Versions:
(3a)	$g_{ti} = \mu h_{it-L} + \omega h_{it-L}^2 + \beta y_{it} + \boldsymbol{\alpha'(x_{it}, d_t)} + u_{it}$	Original
(3b)	$g_{ti} = \mu h_{it-L} + \omega h_{it-L}^2 + \beta y_{it} + \boldsymbol{\alpha'(r_i, d_t)} + u_{it}$	Minimal version
(3c)	$g_{ti} = \mu h_{it-L} + \omega h_{it-L}^2 + \beta y_{it} + \boldsymbol{\alpha'(D_i, d_t)} + u_{it}$	Base model
(4)	$g_{ti} = \Phi(h_{it-1}) + \beta y_{it} + \boldsymbol{\alpha'(D_i, d_t)} + u_{it}$	Generalized base model

Notes: Greek letters are coefficients, vectors are bolded, and u_{it} are residuals. The minimal version contains only controls necessary to generate the substantive results. $\boldsymbol{r}$ contains only x's, which are constant over time. Some variables in both models are estimated with no lags and one lag. All estimated models contain fixed effects for time, but only base models have fixed effects for countries.

II.1 The Good Policy Model: Redirect aid to countries with good policies

The Good Policy Model from Dollar and Burnside (1996, 2000) has two relations, where the second (2d – not in table 1) defines the Good Policy Index, z_{it}. It is scaled to have an average for all countries of about 0, so that $z < 0$ in countries with bad policies and $z > 0$ in countries with good policies. It is important that z is found to be independent of aid:

(2a) $g_{it} = \mu h_{it-L} + \delta z_{it} + \omega z_{it} h_{it-L} + \beta y_{it} + \boldsymbol{\alpha'(x_{it}, d_t)} + u_{it}$

(2d) $z_{it} = 1.28 + 6.85$ Budget Surplus -1.40 inflation $+2.16$ Trade Openness

The substantive part of the model is $g'_{it} = \mu h_{it-L} + \delta z_{it} + \omega z_{it} h_{it-L}$, where g'_{it} is excess growth due to aid. The findings in the original model was that $\mu \approx 0$, $\delta \approx 1$ and $\omega \approx 0.2$. That is, the effect of aid works *exclusively* through the interaction term $z_{it}h_{it-L}$.

Two coefficients are trivial: The Good Policy Index, z, gives a significant coefficient of $\delta \approx 1$ in all regressions, but this is by construction. (1d) is found by a growth regression, and z is almost an outcome variable.[2] Aid has no effect, $\mu \approx 0$. This is well in line with the literature.

The important – non-trivial – finding is that aid interacts with z, i.e. that the interaction coefficient ω is significant and positive. Thus aid has a positive effect if and only if z is positive (policies are good), while aid harms if z is negative (policies are bad). The model thus has the policy implication that aid should be concentrated on the countries following good policies. Burnside and Dollar calculate the gain for the world if aid is redirected accordingly.

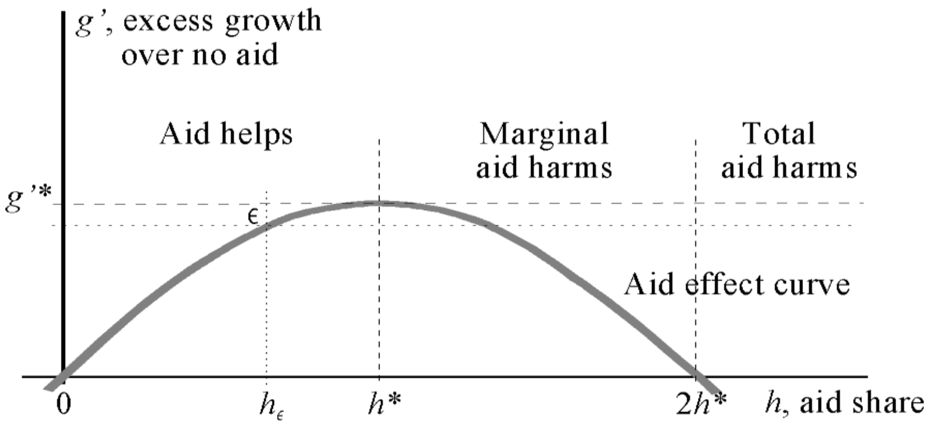
The model has been criticized and defended in no less than 22 papers, covered by Doucouliagos and Paldam (2005c). It has been widely disseminated by the World Bank (1998), the *Economist*, etc., and recently restated in Collier and Dollar (2004) and Burnside and Dollar (2004). Our tests use more of the available data than the previous studies.

II.2 The Medicine Model: Distribute aid proportionally to GDP and never give too much

The model is mainly cited in the version of Hansen and Tarp (2000) for the ODA data and Dalgaard and Hansen (2001) for the EDA data (see section III on the two data sets):

(3a) $g_{ti} = \mu h_{it-L} + \omega h_{it-L}^2 + \beta y_{it} + \boldsymbol{\alpha}'(\mathbf{x}_{it}, \mathbf{d}_t) + u_{it}$

Figure 1. Optimizing the dose of aid in the Medicine Model



2. The combination of policy variables to a good policy index is appealing from the point of view of exposition; but it is an arbitrary construct, which is criticized in several studies, see Doucouliagos and Paldam (2005c).

Here the substantive term is $g'_{ti} = \mu h_{it\text{-}L} + \omega h_{it\text{-}L}^2$. The key result is that $\mu > 0$ and $\omega < 0$. This gives an aid effectiveness curve as drawn on figure 1. The growth effect of aid is independent of the policy of the recipient country, and aid effect curve has a maximum, (g'^*, h^*). If $h > h^*$ the growth generated decreases. The marginal growth contribution of aid is $2\omega < 0$. Thus aid should be distributed to make aid shares of all recipients as equal as possible. Consequently a lot hinges upon the position of the h^*-point.[3]

The results preferred by Dalgaard and Hansen (2001) are $\mu \approx 1.35$ and ≈ -0.13, which gives (g'^*, h^*) = (3.5%, 5.1%). This is substantive excess growth, and as the estimates use EDA-data that (as will discussed) are smaller than the usual ODA-data, it corresponds to $h^* =$ 12% for the ordinary ODA-data. About 24 countries receive aid above this h^*, and 5 countries are even above $2h^*$, where it would be better with no aid at all. However, most replications find smaller values of h^*, and often neither μ nor ω are significant.

The Medicine Model has been analyzed in 15 papers,[4] which mainly work with the same data as Burnside and Dollar, and the tests are mostly done by nesting the two models and showing that the Medicine Model has the best fit. See Docouliagos and Paldam (2005c) for a review. Our study is the first to replicate the model on all available data.

II.3 Variability and moral hazard of the aid effectiveness relation

The aid effectiveness literature has – as mentioned – produced a wide range of results, including the ones of the two models discussed. One reason for the range is socio-political: Aid is a field where many researchers have strong feelings and interests. Therefore, they are willing to go quite far in torturing the data to make it confess. Another reason is that it is doable; it is easy to vary this 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 (2004). Section III discusses the three data sets. We use all three sets in the empirical sections.

---

3. The welfare argument for aid is that it is transfers from DCs with low marginal utilities to LDCs with high marginal utilities give a world welfare gain. If we set the marginal loss in DC to ϵ (appropriately measured) then aid should not stop in h^*, but already in h_ϵ. The g'-curve is flat around its maximum g'^* so even a small ϵ may be visible on the horizontal axis. When h^* is found to be between 5 and 6 we thus choose the lower value.

4. The model was originally discovered by Hadjimichael *et al.* (1995), see also Lensink and White (2001), who term it the Aid Laffer Curve. We prefer the name *Medicine Model*, due to its more precise connotations.

(2) Both substantive models contain a first order aid term, μh, and a second order interaction term: It is aid times good policy, ωzh, in the Good Policy Model, while it is aid squared, ωh^2, in the Medicine Model. By including non-linearities, the number of model variants further increases. Section VI analyzes the form of the aid-term using a semi-parametric technique, which finds the best continuous form of the term.

(3) The control set, $\boldsymbol{x}_{it}$, should in principle contain any variable that has an influence on growth, which is independent of aid. Thus a wide range of $\boldsymbol{x}_{it}$-sets are possible: The theory of growth and the empirical literature on cross-country panel regression models are separated by a gap. It is so wide that several hundred variables that may or may not enter the $\boldsymbol{x}$-set in relations of the type discussed have been proposed. Consequently millions of control-sets are possible, and they typically give a large range of results also for the substantive model, as pointed out by Levine and Renelt (1992) and Sala-i-Martin (1997). It has resulted in proposals for several coping strategies. Advantages and drawbacks of the proposed strategies are discussed in Jensen and Würtz (2005). In addition to statistical strategies there is also the old adage of applying judgment to the control-set chosen: Are these controls reasonable?

The meta-studies referred to show that the 104 authors of the aid effectiveness literature have used about 60 controls. The 5 authors of the two models use about 5 of these – plus an additional handful for robustness experiments. Why the said controls are chosen is barely discussed. The authors concentrate on the substantive parts.

With an unusually wide range of possible choices the *moral hazard problem* of statistical inference becomes unusually large. It is likely that the choices made are influenced by the results when researchers have *priors*.[5] Consequently the likelihood of making Type II errors (acceptance of false models) becomes large. It follows that all models in the field must be independently replicated on new data to be believable. This is what we do at present.

To ensure replicability we are forced to stick to the variables used in the published versions of the two models, and to accept the framework used. That is, we use the Barro-type cross-country regressions and aggregate variables for aid as well.[6]

5. The meta-studies referred to show that the results in this literature suffers from the usual priors: (i) authors have path dependencies, (ii) authors polish results to make them "better" and thus easier to sell to journals, and (iii) authors have interests. In particular, we find that about 35% of the authors work in/for the aid industry.

6. Several writers such as Mavrotas (2002) and Clemens, Radelet and Bhavnani (2004) argue that aid flows should be disaggregated, and each component given a different explanation, with a different time horizon. This is a promising new development, but it will not be pursued at present.

II.4 Two simplified versions of each model: The minimal and the base version

In order to replicate the models on as much of the non-mined data as possible they have to be simplified. This is easier to do with the Medicine Model than with the Good Policy Model, as it is limited by the availability of the 3 variables entering into the Good Policy Index. The simplification is done in two ways: The minimal version is made by stripping the models of unnecessary controls, and the base versions replace all controls with more general ones.

Both models are controlled, by $\boldsymbol{x_{it}}$, for some *potentially* relevant effects, which are not *actually* relevant for the key results. The most obvious way to simplify the models is thus to delete the unnecessary controls and keep only those *necessary*, the $\boldsymbol{r_i}$-set, for reaching the typical results for the substantive models. Note that the $\boldsymbol{r_i}$-set has no time dimension – the time dimension is fully handled by the fixed effect for time, $\boldsymbol{d_t}$ (and y_{it} see below). This simplification is easy to do, and it allows a considerable extension of the data available for replication. This is the *minimal version* of the models.

The argument for the base versions of the models starts from the idea that it would be still better if the specific controls of the $\boldsymbol{x_{it}}$-set, could be replaced by general ones. As the $\boldsymbol{r_i}$-set has no time the $\boldsymbol{D_i}$-set of fixed effects for countries are precisely that.

The choice between specific variables and general fixed effects to control for country differences is worth a few words: For specific variables speak that it is interesting to know precisely which country differences are crucial.[7] This allows the reader to assess if these controls are reasonable. However, specific variables have three costs of which we have already discussed moral hazard and data reduction. The third is that each specific control has en endogeneity problem.[8] Fixed effects have no moral hazard problem, such dummies always available and they are truly exogenous. Furthermore fixed effects convert the data – as much as possible – to one time series.[9] This is an important advantage as the two models both aim at answering the policy question: What happens to growth if aid to a country is increased? This is a time series question. This completes our argument for the *base version* of the models.

---

7. Barro (1997; 36-42) has another argument against fixed effects: It increases the measurement error for the convergence term, y_{it}. It is not our subject at present, and the term is usually negative in the estimates, anyhow.

8. An example of an obvious mistake is that several models adds the share of the public sector to the control set, and report a significant negative coefficient to that control. As most aid goes to the public sector this gives a substantial upward bias in the coefficient to aid.

9. 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 used the assumption that all country differences can be taken out as one shift of the level.

When we estimate *both* the minimal and the base version of the two models we may learn what drives the results. If the minimal version gives results, which disappears in the base version, as in IV.1, we know that the results are due to specific controls that are uncontrolled for in the minimal version. If the minimal version gives smaller coefficients than the base model, as in IV.2, we know that the specific controls are inadequate. Before we turn to the data two minor items should be mentioned.

One specific control is special. It is initial income, y_{it}, which cannot be replaced with fixed effects, and has thus been kept in all regressions, as is fixed effects for time, $\boldsymbol{d_t}$. Little happens to the substantive results if y_{it} is deleted; but we know from the literature on Barro-type regressions how y_{it} should behave, and it thus gives a small check on the estimates.

We analyze the causal relation from aid to growth, but it is possible that causality is from growth to aid. Studies of the determinants of aid (as Alesina and Dollar, 2000) do not suggest that the growth-aid relation is strong, but we cannot a priori reject reverse causality. Hence, we need to control for counter-causality in aid-growth regressions. Three methods are available: (1) Aid is lagged by one time unit relative to the growth explained. (2) The relation is estimated by a 2SLS-technique, or (3) by GMM-technique for dynamic panels. Finding suitable instruments is not easy, and 2SLS-estimation cannot be combined with fixed effect for countries. Also the instruments enter almost as the controls in the $\boldsymbol{x}$-set and add to the moral hazard problem. The original articles do not use (system and difference) GMM-panel estimators, but they are easy to apply, and we have re-estimated everything using GMM. It proves to matter little. So we present the OLS estimates in the tables and report the GMM-results in the text and in notes. Consequently, method (1) is our preferred method.

III. The data

First the three sets of aid data are defined, and then we discuss which one to prefer. The two Appendix tables list the countries included in the different samples. Table 2 surveys the various data sets used in the regressions.

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.

Table 2. Aid data samples

Name	Source	Variant	Period		n
ODA	Official, WDI (2003) used as source	ODA-full	From 1966	All available in WDI (2003)	756
		ODA-55		Sample for the CFS-56 countries	472
CFS	Chang, Fernandez-Arias and Serven (1998). EDA-set	CFS-56	1970-1993[a)]	Used for both models	269
		CFS-42		42 unused countries	216
		CFS-full		All 98 countries with updates	546
ELR	Easterly, Levine, Roodman (2004). Updated version of CFS, extended by ODA-data	ELR-full	1970-1997	All data in sample	586
		ELR-m3		3 wild observations excluded	583
		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 and Serven the data start in 1975, but Burnside and Dollar give series starting in 1970.

The EDA-data are produced 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 Serven (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 (2000) is an early version of that set.[10] It includes 56 countries 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 be used for the estimates.

Easterly, Levine and Roodman (2004) updates the CFS-data set, so that observations are available for more countries and the period 1970-97.[11] 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 the last time unit 1994-97 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, 1970-73, which in the CFS-data it is no less than +19%. Two other wild observations are Guinea Bissau with –5.71% and Gambia with –4.59%. 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

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

11. 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.

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 measure the gross resource flow, while the EDA-data consider only the net flow. We use both definitions in the replications for theoretical as well as empirical reasons:

Theoretically, it is unclear which variable to prefer. A rational expectations view of the Barro-Ricardo type suggests that only net grants affect the behavior of agents.[12] Thus the EDA-data are the proper ones. However, a large body of evidence suggests that politics has a very short time horizon (see e.g. Paldam, 2004). 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 some activities.

Empirically table 3 shows that they are highly correlated. 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. Doucouliagos and Paldam (2005) investigate the effect of the two definitions for the elasticity of aid on growth and find that the ODA-data give slightly better results.

The average ratio between the ODA-data and the CFS-data (the pure EDA-data) is app. 2.4. This suggests that the $h*$-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.

Table 3. Correlations 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 1970-93.

12. 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%.

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 on these data. After the replication (not presented) the models are simplified to the *minimal* and the *base versions*. All models presented in tables 4 to 9 are estimated by OLS and heteroscedasticity-consistent errors as in the original papers. The fixed effects model uses the within-groups estimator.

We further estimated (difference and system) GMM estimators to check all results. 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. The check is thus necessary, though it proves that inconsistency is of no consequence except in one case reported in the text.[13]

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 (L=0)	74-93 (L=1)	70-93 (L=0)	74-93 (L=1)
Aid effect, μ	–0.01 (0.04)	0.27 (1.27)	0.32 (1.32)	0.69 (1.68)
Good policy, δ	**0.68 (3.63)**	**0.68 (2.85)**	**1.04 (3.58)**	**1.10 (4.28)**
Interacted (L), ω	**0.18 (2.53)**	–0.02 (0.18)	–0.13 (0.99)	**–0.20 (2.11)**
GDP-level, β	–0.65 (1.15)	–0.42(0.63)	–2.07 (1.55)	–2.47 (1.61)
Institutions (x_1)	**0.73 (4.26)**	**0.76 (3.86)**	No	No
Africa (x_2)	**–2.09 (2.70)**	**–2.61 (3.29)**	No	No
Orient (x_3)	**1.38 (2.46)**	**1.67 (3.61)**	No	No
Time dummies, $\boldsymbol{d}_t$	Yes	Yes	Yes	Yes
Country dummies, $\boldsymbol{D}_i$	No	No	Yes	Yes
N, number of obs	270	234	267	230
R^2	0.39	0.36	0.53	0.55

Note: OLS regressions, brackets holds t-statistics, and bold show significance at the 5% level. L is the lag (if any) to *h* and *hz*. Panel regressions need 2 observations for each country, so 3-4 observations cannot be used.

13. On GMM: We treat policy as exogenous in the Good Policy Model as in the original paper. Instruments for initial GDP are the second lag of GDP and all further lags. Similar instruments are used in the Medicine Model. Tests of over-identifying restrictions always accept the null and tests of serial correlation are satisfactory. For the ODA sample, we restrict the number of instruments to avoid singularity of the covariance matrix of moments. It matters little if aid is unlagged and treated as endogenous or lagged and treated as predetermined.

IV.1 The Good Policy Model

A good replication should have the following two key features: The coefficient to the aid term $\mu \approx 0$, and the coefficient to the interacted term $\omega \approx 0.2$. In addition the effect of good policy $\delta \approx 1$ and the convergence term $\beta < 0$.

The model is given in table 1, which also lists the original $\boldsymbol{x}$ set of 7 controls.[14] The substantive results – μ and ω – 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 results are given in table 4. Column (1) gives virtually the same results as in the original article.[15] Column (3) shows what happens if the 3 specific controls for country differences are replaced with fixed effects. Here the substantive effects, μ and ω, disappear, and signs even change. Consequently, we know precisely what drives the substantive 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.[16] We argued above that the tidiest 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 by GMM, we obtain similar results, with the difference that the aid-policy interaction still has a negative coefficient, but it is only significant when the first step system-GMM estimator is used. Thus the Good Policy Model is a fickle construct.

IV.2 The Medicine Model

A good replication of the Medicine Model should have the following key features: The coefficient to the aid term $\mu > 0$, and the coefficient to the squared aid term $\omega < 0$. The size of the two effects reported by Dalgaard and Hansen (2001) using 2SLS-estimation and a large set of controls are 1.35 to aid and –0.13 to aid squared.

---

14. For easy reference they are: (x_1) institutional quality, (x_2) Africa, (x_3) East Asia, (x_4) political assassinations, (x_5) ethnical fractionalization, (x_6) x_4 times x_5 and (x_7) financial depth. 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. Though this might bias the estimates of the effect of aid, it does not happen, as expected from Brunetti (1998).

15. 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.

16. It was controlled for by 2SLS-estimation in the working paper, but the instruments were not convincing.

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 (L=0)	74-93 (L=1)	70-93 (L=0)	74-93 (L=1)
Aid share (L), μ	0.28 (0.70)	**0.87 (2.34)**	0.50 (0.86)	**1.32 (2.32)**
Aid squared (L), ω	–0.02 (0.31)	**–0.065 (2.26)**	–0.04 (0.81)	**–0.12 (2.81)**
GDP-level, β	–0.59 (1.05)	–0.39 (0.59)	–2.03 (1.47)	–2.13 (1.48)
Institutions (x_1)	**0.89 (4.77)**	**0.98 (4.74)**	No	No
Africa (x_2)	**–2.29 (3.01)**	**–2.91 (3.65)**	No	No
Orient (x_3)	**2.54 (4.78)**	**2.99 (5.10)**	No	No
Time dummies, $\boldsymbol{d}_t$	Yes	Yes	Yes	Yes
Country dummies, $\boldsymbol{D}_i$	No	No	Yes	Yes
N, number of obs	270	234	267	269
R^2	0.31	0.32	0.49	0.52

Note: See note to table 4. L is the lag (if any) to h and z^2. All regressions are OLS.

The model is easy to reproduce on the CFS-56 data; 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. The coefficients to the three controls are much the same as before, but now they can be replaced by the fixed effect. Regression (4) is a perfect replication of the substantive results of Dalgaard and Hansen and it can be replicated on all available data. The key finding from table 5 is that both substantive coefficients μ and ω are fairly stable. Clearly, the Medicine Model is superior to the Good Policy Model when it comes to robustness in the within-sample replications.[17]

When the parables from the 4 estimates are drawn – as sketched on figure 1 – they all look similar with the h^*-point between 5% and 7%. The one for the model in column (4) is included as the quadratic curve on figure 2a below.[18]

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

18. The aid terms are still significant with the right signs when re-estimated with GMM.

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.

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

Model	(1) = (t4,1) [a)]	(2)	(3) = (t4,3) [a)]	(4)
Aid data	CFS-56	CFS-62	CFS-56	CFS-69
Period	70-93 (L=0)	70-93 (L=0)	70-93 (L=0)	70-93 (L=0)
Aid effect, μ	–0.01 (0.04)	0.05 (0.46)	0.32 (1.32)	0.12 (0.66)
Good policy, δ	**0.68 (3.63)**	**0.84 (3.37)**	**1.04 (3.58)**	**1.12 (4.31)**
Interacted, ω	**0.18 (2.53)**	0.06 (0.94)	–0.13 (0.99)	–0.07 (1.33)
GDP-level, β	–0.65 (1.15)	–0.08 (0.17)	–2.07 (1.55)	–**2.82 (2.27**)
Institutions (x_1)	**0.73 (4.26)**	0.27 (1.78)	No	No
Africa (x_2)	**–2.09 (2.70)**	–0.12 (1.73)	No	No
Orient (x_3)	**1.38 (2.46)**	**1.84 (2.81)**	No	No
Time dummies, $\boldsymbol{d}_t$	Yes	Yes	Yes	Yes
Country dummies, $\boldsymbol{D}_i$	No	No	Yes	Yes
N, number of obs	270	307	267	337
R^2	0.39	0.30	0.49	0.46

Note: See note to table 4. No variable is lagged. All regressions are OLS. (2) and (4) contains outliers.

a. Column "(1) = (t4,1)" is table 4 column (1) , and column "(3) = (t4,3)" is table 4, column (3).

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 shows the results for the Good Policy Model. 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 by 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, μ, to aid is now insignificant. If it is disregarded, aid is harmful at any level. If it is included, the h*-point is 6.5.

Table 7. The Medicine Model estimated on CFS-data

Model	(1) = (t5,4)	(2)	(3)	(4)
Aid data	CFS-56	CFS-42	CFS-full	CFS-full
Period	74-93 (L=1)	74-97 (L=1)	74-97 (L=1)	70-93 (L=0)
Aid share (L), μ	**1.32 (2.32)**	0.26 (1.17)	**0.60 (2.95)**	0.23 (0.72)
Aid squared (L), ω	**–0.12 (2.81)**	**–0.02 (2.53)**	**–0.035 (3.81)**	–0.04 (0.13)
GDP-level, β	–2.13 (1.48)	**–0.78 (3.48)**	**–2.41 (2.40)**	–0.37 (0.72)
Institutions (x_1)	No	No	No	**0.77 (4.19)**
Africa (x_2)	No	No	No	**–2.47 (3.60)**
Orient (x_3)	No	No	No	**2.51 (4.74)**
Time dummies, $\boldsymbol{d}_t$	Yes	Yes	Yes	Yes
Country dummies, $\boldsymbol{D}_i$	Yes	Yes	Yes	No
N, number of obs	269	216	546	346
R^2	0.52	0.38	0.43	0.28

Note: See note to table 4. L is the lag (if any) to h and z^2. All regressions are OLS.

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 h^*-point moves to 8.5. Using GMM-estimators has little effects on 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 key coefficient, ω, to the interacted term, $z_{it}h_{it}$, 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 those of Easterly *et al.* (2004), and we have added the additional evidence of the ODA-data set.

Table 9 shows the results for the Medicine Model. The base model uses all observations available. Aid squared fails in all regressions, and aid fails in all but one regression. It is the full ELR-data set, but it is due to the 3 "wild" observations. When they are deleted, the term fails. In the minimal model with aid unlagged all aid terms are insignificant. For these samples, it makes no difference to use GMM-estimators.

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

Models	(1)	(2)	(3)	(4)	(5)	(6)
Aid data	ELR-full	ELR-m3	ODA-full	ELR-full	ELR-m3	ODA-full
Period	70-97 (L=0)	70-97 (L=0)	66-97 (L=0)	70-97 (L=0)	70-97 (L=0)	66-97 (L=0)
Aid effect, μ	0.02 (0.16)	0.012 (0.10)	0.01 (0.35)	0.18 (1.09)	0.18 (0.92)	0.0015 (0.03)
Good policy, δ	**0.77 (3.86)**	**0.78 (3.66)**	**0.89 (4.51)**	**0.88 (3.28)**	**0.88 (3.29)**	**1.06 (4.11)**
Interacted, ω	0.07 (1.05)	0.07 (0.96)	–0.00 (0.27)	0.03 (0.25)	0.03 (0.25)	–0.02 (1.36)
GDP-level, β	–0.17 (0.41)	–0.18 (0.42)	–0.66 (1.62)	–1.08 (1.18)	–1.08 (1.13)	**–2.46 (2.43)**
Institutions (x_1)	0.21 (1.66)	0.21 (1.65)	**0.90 (4.51)**	No	No	No
Africa (x_2)	–1.19 (1.92)	–1.18 (1.89)	**–1.54 (2.64)**	No	No	No
Orient (x_3)	**2.19 (3.84)**	**2.18 (3.66)**	**1.77 (3.74)**	No	No	No
Time dummies, $\boldsymbol{d}_t$	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies, $\boldsymbol{D}_i$	No	No	No	Yes	Yes	Yes
N, 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. No variable is lagged. All regressions are OLS.

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

	ELR-data (EDA)		ODA-data	ELR-data (EDA)		ODA-data
Model	(1)	(2)	(3)	(4)	(5)	(4)
Aid data	ELR-full	ELR-m3	ODA-full	ELR-full	ELR-m3	ODA-full
Period	73-97 (L=1)	73-97 (L=1)	66-01 (L=1)	73-97 (L=0)	73-97 (L=0)	66-01 (L=0)
Aid share, μ	**0.21 (2.58)**	0.18 (0.62)	0.095 (1.62)	0.13 (0.88)	0.07 (0.37)	0.07 (1.58)
Aid squared, ω	–0.003 (0.48)	0.001 (0.07)	–0.001 (1.26)	0.013 (0.89)	0.019 (0.47)	–0.008 (0.79)
GDP-level, β	**–3.04 (3.48)**	**–3.13 (3.09)**	**–2.76 (3.51)**	0.009 (0.59)	(0.16 (0.47)	**0.73 (2.21)**
Institutions (x_1)	No	No	No	**0.29 (2.63)**	**0.29 (2.64)**	**0.26 (2.56)**
Africa (x_2)	No	No	No	**–1.42 (2.84)**	**–1.39 (2.71)**	–0.82 (1.77)
Orient (x_3)	No	No	No	**3.19 (2.63)**	**3.16 (6.51)**	**3.41 (8.00)**
Time dummies, $\boldsymbol{d}_t$	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies, $\boldsymbol{D}_i$	Yes	Yes	Yes	No	No	No
N, number of obs	586	583	755	541	539	526
R^2	0.43	0.43	0.47	0.22	0.22	0.29

Note: See note to table 4. L is the lag (if any) to h and z^2. All regressions are OLS.

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 insignificant. 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 those 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.[19]

VI. The form and significance of the general aid term

We now replace the arbitrary parametric form of the aid-growth relation with: (4) $g_{ti} = \Phi(h_{it-1}) + \boldsymbol{\alpha}'(\mathbf{D_i}, \mathbf{d_t}) + \beta y_{it} + u_{it}$, where $\Phi(h_{it-1})$ 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

The Medicine Model discussed above makes a strong assumption on the functional form of $\Phi(h)$. This can be relaxed by using a semi-parametric method that relies on the Weierstrass approximation theorem (see e.g., Apostol, 1972; 322).[20] The theorem states that every continuous function on a compact interval can be uniformly approximated by a polynomial. Thus, one could in principle use a high order polynomial (referred to as basis functions in the literature) to approximate the unknown $M(h)$. We chose to use cubic splines with four equidistanced knots on the h-axis. Intuitively, this is mathematically similar to using a polynomial, but has been shown to have better finite sample properties. The fixed effects for countries are treated as usual.

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

19. 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 both aid and aid squared fail.

95% confidence bands, which are wider where there are few observations. It also includes the fitted values from a linear regression and the relevant aid squared models referred to.

The $\Phi(h)$-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. We estimated the relationship with both OLS and GMM-estimators.[21]

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

The 5 ACH (1) tests in table 10 tell a sad story of insignificance. The only marginally significant result for the aid term is at the 10% level. It is, as expected, for the CFS-56 sample.

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
Aid term, $\Phi(h)$	Fig 2a	Fig 2b	Fig 3	Not given	Fig 4
GDP-level, β	–2.32 (1.66)	**–2.61 (2.48)**	**–3.31 (3.17)**	**–3.11 (2.97)**	**–2.59 (3.27)**
Time dummies, $\boldsymbol{d}_t$	Yes	Yes	Yes	Yes	Yes
Country dummies, $\boldsymbol{D}_i$	Yes	Yes	Yes	Yes	Yes
ACH-1 for aid term	3.27 [a)]	2.94 [b)]	2.75 [c)]	1.88	1.72
ACH-2 for not linear	**5.58**	**4.47**	n/a	n/a	n/a
N, 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.

20. 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).

21. The GMM-estimates are very similar to the OLS-estimates, but less precise. There are however certain problems with the estimation: 1) Instruments need to be dropped. 2) Because of singularities in certain matrices, ACH-tests cannot be computed, and two-step estimates and system estimates are not available.

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.

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 that 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.[22]

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.[23]

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

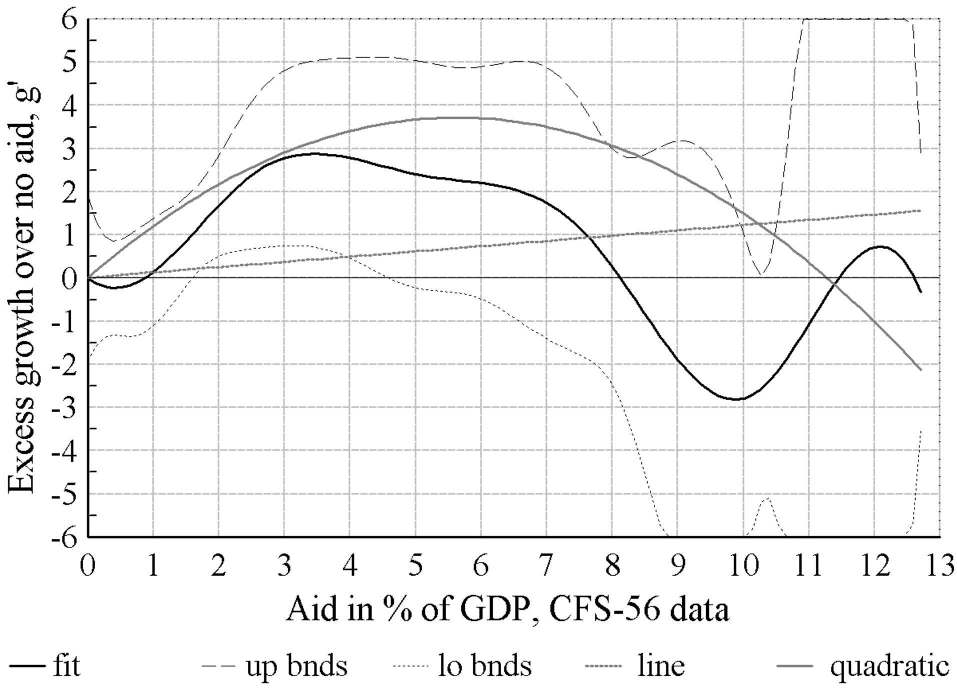
The ACH-tests in table 10 and the t-tests in the matching parametric regressions disagree in three out of 5 cases (see notes to table). This is puzzling, but it is possible as both are asymptotic tests.

---

22. Semi-parametric regressions were made for all cases of tables 5, 7 and 9, with results as the ones reported.

23. 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: In the ODA sample the null of no relationship is rejected at the 5% level using the ACH-tests, 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.

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 "up bnds" and "lo bdns". The parable is calculated from regression (4) in table 5.

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

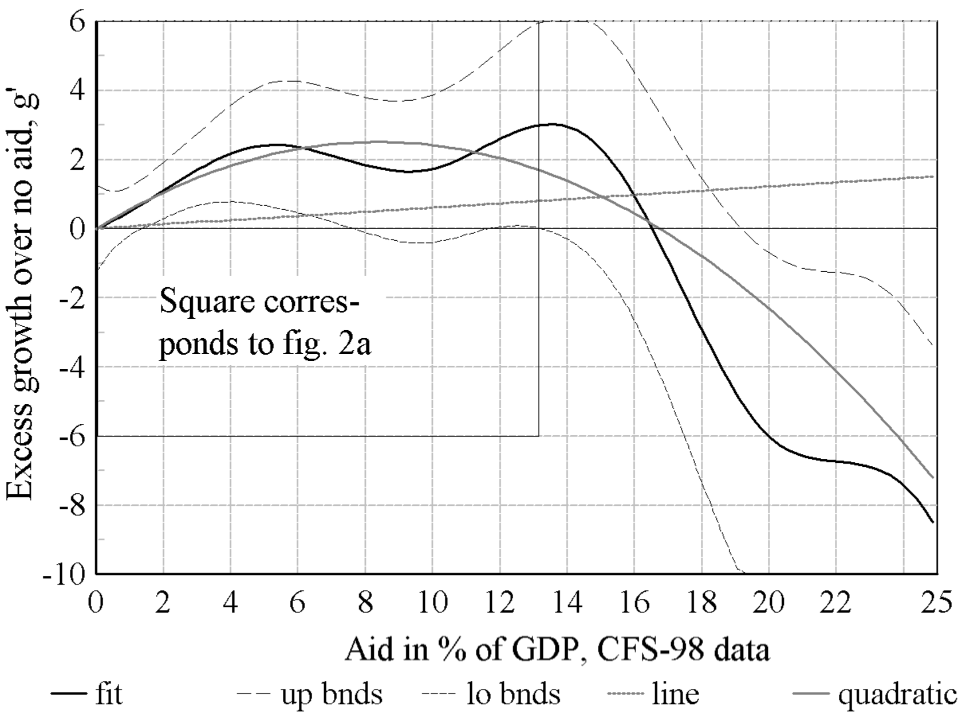
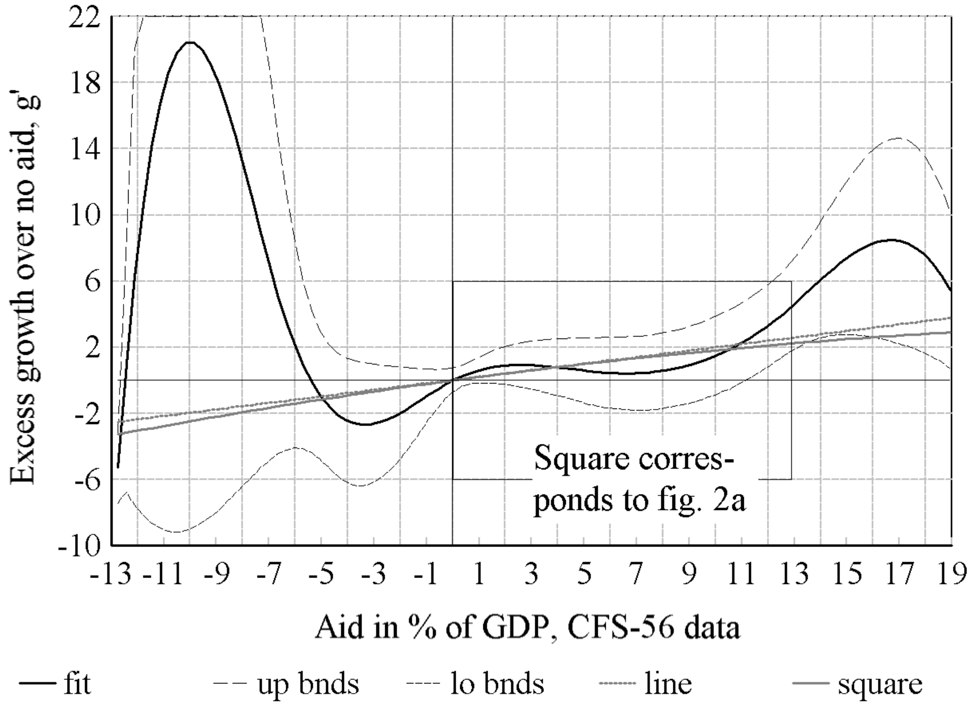
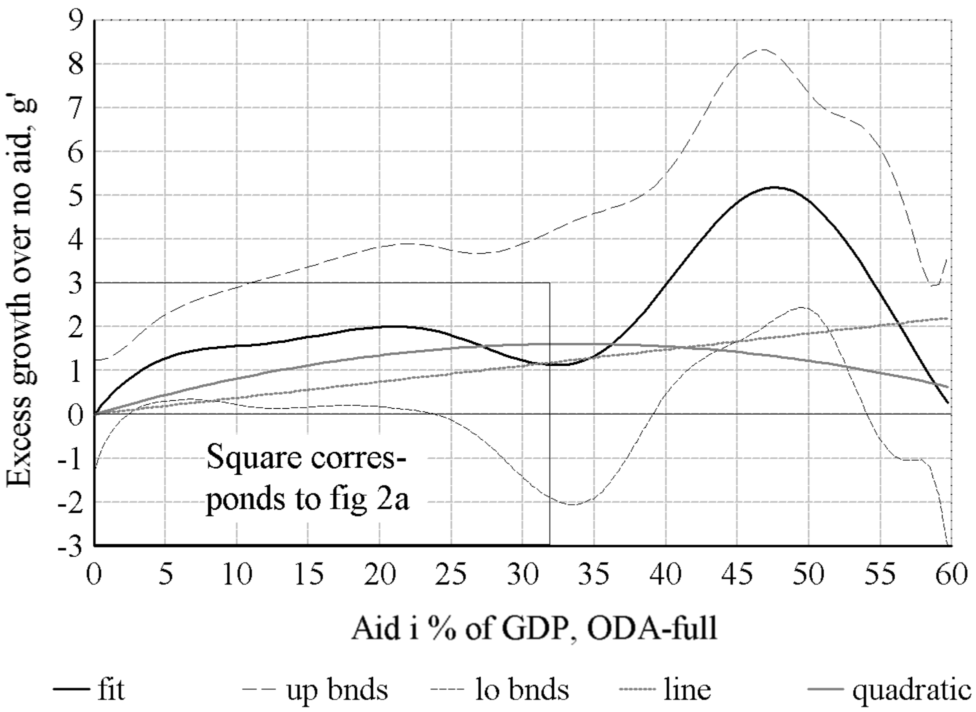


Figure 3. Aid-term in the base model on the ELR-full data, N = 586



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.

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.

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 whether 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 flukes in a particular subset of the data.

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 it is, as mentioned, arguable the aid should disaggregated, parts having different effects. 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), 869-79

Alesina, A., Dollar, D., 2000. Who gives foreign aid to whom and why? *Journal of Economic Growth* 5, 33-63

Apostol, T. M., 1972. *Mathematical Analysis. 2nd ed.* Addison Wesley, Reading, MA

Barro, R.J., 1974. Are Government Bonds Net Wealth? *Journal of Political Economy* 82, 1095-117

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., 2000. Aid, Policies and Growth. *American Economic Review* 90: 847-68. World Bank working paper in several versions from 1996

Burnside, C., Dollar, D., 2004. Aid, policies and growth: Reply. *American Economic Review* 94, 781

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>

Clemens, M., Radelet, S., Bhavnani, R., 2004. Counting chickens when they hatch: The short term effect of aid on growth. WP Nr 44. Center for Global Development

Collier, P., Dollar, D., 2004. Development effectiveness: What have we learnt? *Economic Journal* 114, 244-71

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

Doucouliagos, C., Paldam, M., 2005a. Aid effectiveness on accumulation. A Meta Study. Mimeo Economics Department, University of Aarhus, Denmark [24]

Doucouliagos, C., Paldam, M., 2005b. Aid effectiveness on growth. A Meta Study. Mimeo Economics Department, University of Aarhus, Denmark

Doucouliagos, C., Paldam, M., 2005c. Conditional aid effectiveness. A Meta Study. Mimeo Economics Department, University of Aarhus, Denmark

Easterly, W., Levine, R., Roodman, D., 2004. Aid, policies, and growth: Comment. *American Economic Review* 94, 774-780 (Comment to Burnside and Dollar, 2001) Earlier: New data, New doubts. NBER WP 9846

Gørgens, T., Paldam, M., Würtz, A., 2003. How Does Public Regulation Affect Growth? Mimeo Economics Department, University of Aarhus, Denmark

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. *Journal of International Development* 12, 375-39. Also pp 103-128 in Tarp and Hjertholm (2000)

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

Hudson, J., 2004. Introduction: Aid and Development. *Economic Journal* 114, 185-90, with contributions by Collier and Dollar, Dalgaard, Hansen and Tarp, and Mosley, Hudson and Verschoor

---

24. The three studies are referred to as one. They are available from http://www.martin.paldam.dk, from June – August 2005.

Jensen, P.S., Würtz, A., 2005. The ill-posed problem in growth empirics. Mimeo Economics Department, University of Aarhus, Denmark

Keefer, P., Knack, S., 1995. Institutions and Economic Performance: Cross-country Tests Using Alternative Institutional Measures. *Economics and Politics* 7, 207-27

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

Levine, R., Renelt, D., 1992. A Sensitivity Analysis of Cross-Country Growth Regressions. *American Economic Review* 82, 942-63

Mavrotas, G., 2002. Foreign aid and fiscal response: Does aid disaggregation matter? *Weltwirtschaftliches Archiv* 138, 534-59

Paldam, M., 2004. Are vote and popularity functions economically correct? Pp 49-59 in Rowley, C.K., and 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

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

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

WDI, 2003. *World Bank Indicators* 2003, CD-ROM, also on the internet

World Bank, 1998. *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.