Stockholm Institute of Transition Economics

WORKING PAPER

April 2011

No. 11

Aid Effectiveness: New Instruments, New Results?

Emmanuel Frot and Maria Perrotta

STOCKHOLM INSTITUTE OF TRANSITION ECONOMICS

Stockholm Institute of Transition Economics (SITE) · Stockholm School of Economics · Box 6501 · SE-113 83 Stockholm · Sweden

Aid Effectiveness: New Instrument, New Results?

Emmanuel Frot* Maria Perrotta†

First version, May 2009. This version, April 2010

Abstract

Despite a voluminous literature on the topic, the question of whether aid leads to growth is still controversial. To observe the pure effect of aid, researchers used instruments that must be exogenous to growth and explain well aid flows. This paper argues that instruments used in the past do not satisfy these conditions. We propose a new instrument based on *predicted* aid quantity and argue that it is a significant improvement relative to past approaches. We find a significant and relatively big effect of aid: a one standard deviation increase in received aid is associated with a 1.6 percentage points higher growth rate.

PRELIMINARY, DO NOT CITE.

1 Introduction

Foreign assistance has been disbursed for decades and is today still seen as a major tool of development policy, and while all promises of increasing aid flows are likely not to be fulfilled, the trend is clearly towards an expansion. If there seems to be near unanimity among policy makers about the positive role of aid,[1] the academic

*SITE, SSE, P.O. Box 6501, SE-113 83 Stockholm, Sweden. Email: emmanuel.frot@hhs.se.

†SITE, SSE, P.O. Box 6501, SE-113 83 Stockholm, Sweden. Email: maria.perrotta@hhs.se. The authors wish to thank Ethan Kaplan, Jakob Svensson, Rajeev Dehejia, Philippe Aghion, Pamela Campa, Martin Berlin, all the participants at the IIES and Economics Department seminars at Stockholm University as well as three anonimous referees.

[1]For instance, the Monterrey Consensus, adopted by Heads of State and Government after the 2002 United Nations International Conference on Financing for Development, states that "Official Development Assistance (ODA) [...] is critical to the achievement of the development goals and targets of the Millennium Declaration", that "we recognize that a substantial increase in ODA and

community has not found any robust evidence that aid contributes to development.[2] The aid effectiveness literature is large and mostly inconclusive. The results vary widely in size and sign, and have often been proven not to be robust and often reversed by new estimations. The so-called third generation of aid and growth studies, which established some influential and widely cited results in the 90s, has recently been criticized, mainly on two points: the unsatisfactory instrumentation strategies and the "black box" way in which they use General Method of Moments (GMM) estimations.[3] These two points relate to the two fundamental issues the researcher must confront when designing an empirical strategy to deal with the question of aid effectiveness. The first is the identification of the causal effect of aid on growth, unconfounded by simultaneity and reverse causality. The second is the consistent estimation in a dynamic panel setting. We offer our main contributions on these two points.

We propose a new instrument and argue that it is a significant improvement relative to past approaches. It takes the "supply side" approach, that makes use of variables linked to the aid allocation process (mostly historical and political variables), one step further. Our identification strategy is similarly based on predicted aid flows; however, unlike existing studies, we exploit a source of variation that we argue not to be subject to the same criticisms, namely that of being directly correlated with the outcome. This source of variation is related to the temporal order in which donor-recipient partnerships are established: Frot (2009) shows that *when* a partnership is established and *how long* it lasts are of importance for aid quantities. In addition to being exogenous to growth, we show that our instrument is highly correlated with actual aid levels.

On the second point, we keep our estimation strategy as simple and transparent as possible. Given that standard panel estimators (fixed effect estimators) are biased in dynamic settings, we make use of the GMM estimators in order to account for individual level fixed effects. But we rely exclusively on our "external" instrument for the identification of the aid coefficient. In addition, we test the validity of the

other resources will be required if developing countries are to achieve the internationally agreed development goals and objectives, including those contained in the Millennium Declaration", and that "we urge developed countries that have not done so to make concrete efforts towards the target of 0.7 per cent of gross national product (GNP) as ODA to developing countries".

[2]Many authors argue that aid failed to achieve growth. Easterly (2006) gives a detailed presentation of the arguments. Easterly (2007) summarizes them in an article entitled "Was Development Assistance a Mistake?".

[3]Bazzi and Clemens (2009) make this point very effectively and provide many examples. In releasing his Stata package to perform GMM estimations, Roodman (2009a) warned about the risks of using it unwittingly.

instruments created by the GMM procedure and, as a consequence, we are able to comment on the validity of the GMM approach to estimate aid efficiency.

To give a preview of the results, we find a significant and moderate effect of aid on growth: in our sample, a 1 percent increase in received aid is associated with a 0.06 to 0.13 percent higher growth rate.

The paper is organized as follows: in the next section, we spell out what are the empirical challenges that the question of aid effectiveness presents, and highlight how the literature has dealt with them in some important contributions. In Section 3 we describe in detail how our instrument is built; we then briefly discuss our methodological choices in terms of estimators and present the results in Section 4. In Section 5 the robustness of the results is assessed. Section 6 concludes the paper. All variable definitions and data sources are to be found in a data appendix at the end of the paper.

2 Estimation pitfalls and previous literature

A vast literature has focused on the effect of aid on GDP growth, controlling for various variables. Some version of the following equation is implicitly or explicitly derived from a standard growth model a la Solow, and brought to the data:

$$\Delta y_{it} = \alpha_t + \beta y_{it-1} + \gamma a_{it-1} + \sum_k \delta_k x_{kit} + \mu_i + \xi_{it}. \tag{1}$$

In equation (1), i and t respectively index the countries and time periods (five-year intervals, usually), y is the (natural logarithm of) GDP, and Δ indicates its variation, an approximation for the growth rate, α is a constant that might change over time, γ is the coefficient of interest, the effect of aid a (also in logs), x_k are additional explanatory variables, and the error term consists of an unobserved country-specific effect μ_i and a random noise ξ_{it}.[4]

To estimate this equation, researchers have to deal with the familiar problem of reverse causality. Aid is to a larger extent allocated to low performing countries, such that low growth 'causes' high aid quantities. This simple observation makes the causal link from aid to GDP growth impossible to establish by looking at simple partial correlations between these two variables. To observe the causal effect of aid, researchers used instruments that must be exogenous to growth and explain aid flows

[4]Most papers in the literature estimate the effect of aid, expressed as a share of GDP, on growth. We prefer to use aid levels, for reasons exposed later, and for consistency with the literature also run our regressions using aid as a share of GDP in Section 5.

Table 1: Instruments in the aid effectiveness literature

Boone (1996)	Burnside and Dollar (2000)	Hansen and Tarp (2001)	Dalgaard et al. (2004)
Log population	Log of initial income	Egypt dummy	Aid (-1)
Friends of US	Policy index	Arms imports (-1)	Aid^2 (-1)
Friends of OPEC	Log population	Policy (-1)	Aid*inflation (-1)
Friends of France	Arms imports/Tot. imports, (-1)	Policy^2 (-1)	Aid*openness (-1)
Aid (-2)	Sub-Saharan African dummy	Policy*Log population	Aid*share of land in tropics (-1)
	Egypt dummy	Policy*Initial GDP per capita	M2/GDP (-1)
	Franc zone dummy	Policy*Initial GDP per capita^2	Budget surplus (-1)
	Central America dummy	Policy*aid (-1)	Inflation (-1)
	Policy*aid^2 (-1)	Openness (-1)	
	Aid(-1)		
	Aid^2 (-1)		

Note: Instrumental variables for aid used in four influential papers. -1 and -2 indicate lags.

well. Rajan and Subramanian (2008) and Bazzi and Clemens (2009), among others, review the past literature and question the validity of the instruments used in past studies. Table 1 lists the instruments used in four influential papers.[5]

These papers typically instrument aid with many variables without any clear identification strategy. Burnside and Dollar (2000) explain that theirs is based on the aid allocation literature, the so-called *supply-side* approach, but it is difficult to argue that any of their instruments satisfies the required exogeneity assumption. Deaton (2010) criticizes the whole literature by mentioning that neither the Egypt dummy nor population, though they are aid determinants, can plausibly be exogenous. These variables are *external* to growth but assuming that they do not have any influence on growth except through aid flows is not very plausible. Moreover, the Egypt dummy is problematic as the source of variation is unlikely to teach us anything about the effect of aid on growth in a general way. The variation between Egypt and non-Egypt countries, or for that matter between Franc-zone countries and non Franc-zone countries, is not very useful. Unfortunately, similar criticisms apply to all instruments listed in Table 1. None of them is exogenous to growth. Even the fraction of land in the tropics, used by Dalgaard et al. (2004), is correlated with institutions which, in turn, affect long-run development, as shown by Acemoglu et al. (2001). Lagged aid variables, either interacted with other exogenous regressors or not, also constitute a dubious choice if growth is serially correlated. Similarly, the assumption that a control such as *policy* has a contemporaneous effect on growth but none in the next period, except through aid, is hard to defend.

[5]The instruments used in Boone (1996) are found in Table 4 of his paper, those of Burnside and Dollar (2000) in Table 1 and those of Hansen and Tarp (2001) in Table 1. Dalgaard et al. (2004) reproduce previous specifications but their own set of instruments is found in Table 3 of their paper, Clemens et al. (2004) use the same set of instruments as Hansen and Tarp (2001).

Rajan and Subramanian (2008) recognize these issues and adopt a slightly different approach based on donor-recipient pair characteristics, instead of using recipients' characteristics. Donors choose aid allocation based on poverty considerations, but also because of history and influence. The authors here capture historical relationships through colonial links and commonality of language. Influence is proxied by the relative size of the donor and the recipient. The larger the donor, the larger its influence. Relative size is also interacted with historical variables as influence is likely to be further increased if historical links are strong. Aid quantities are estimated at the donor-recipient level and then summed across donors to find the recipient predicted aid quantity. Rajan and Subramanian (2008) then use this instrument to revisit most of the past evidence on aid effectiveness and find little robust evidence of any link between aid and growth.

This identification strategy improves upon past studies but is still not entirely convincing. Historical variables are unlikely to be exogenous to growth and are correlated with traditional growth determinants, as shown by Bertocchi and Canova (2002). Acemoglu et al. (2001) have also demonstrated how colonial origins are of importance for growth through institutional quality. A second concern with the instruments of Rajan and Subramanian (2008) is their limited variation since historical variables are simple dummies. In addition, they still include population in their set of instrumental variables, despite its drawbacks. In fact, Bazzi and Clemens (2009) show that their identification almost exclusively relies on population size because the other instruments are weak, to the point of being irrelevant. Therefore, Rajan and Subramanian (2008) face the same problem of invalid instruments as earlier papers.

A second challenge for the researcher is the fact that the process of economic growth calls for a dynamic model, in which current values depend on past realizations. This is why the lag value of income figures as a regressor in equation (1). One immediate problem in the estimation of such a model is that lagged values of the dependent variable (and potentially of the other regressors) are correlated with the fixed effect in the error term. This makes the OLS and 2SLS estimators inconsistent.[6] Sure enough, the fixed effect estimator is consistent; but with five-year intervals over forty years of data it is not possible to rely on asymptotic properties[7], although this point has often been overlooked. To deal with this issue, researchers have made use of a class of estimators built for the purpose, namely the GMM estimators. This procedure consists of first-differencing the data, as opposed to the fixed effect trans-

[6] Only the coefficient on income is plagued by this problem. On the other hand, the presence of one inconsistent coefficient also biases the other coefficients in the regression, moreover in a direction that is difficult to predict.

[7] Asymptotics require $t \to \infty$, while here we have $t = 8$ at most!

formation that demeans them (subtracts the sample mean). Endogenous variables are then instrumented using their own lagged values. The main advantages of these estimators are that they deal with individual level fixed effects without incurring the bias to which standard panel estimators (chiefly the fixed effect transformation) are subject in dynamic settings. Furthermore, they offer "internal" solutions for dealing with endogenous regressors. In particular, the Arellano and Bond (1991) original "difference" estimator instruments for current period differences in endogenous variables using their own multiple lagged levels. The more efficient Blundell and Bond (1998) "system" estimator, which exploits the moment conditions from a system of the differenced equation plus the original level equation, additionally instruments for current period levels using lagged differences. This wealth of plausibly valid instruments is never submitted to the standard weak-instrument diagnostics, so there is no guarantee for their relevance; and the problems for inference of using many weak instruments are very serious and very well known.[8] Moreover, the exclusion restrictions on which these methods rely are more demanding than what is often assumed (in particular for the "system" method; see Roodman (2009a) for a discussion of these issues).

Our approach is hence to exclusively rely on our external instrument for the identification of the aid coefficient. Endogenous variables for which we do not have an external instrument[9], mainly income, are instrumented using their lagged values, but we are very careful in keeping the number of instruments as low as possible by collapsing the instrument matrix, as recommended in Roodman (2009a). Moreover, in Section 5, we replicate the GMM instrumentation in a traditional IV setting, so that we can use the whole standard battery of tests for instrument strength. In the absence of a test for instrument strength in a GMM setting, this approach is used by Bazzi and Clemens (2009), following Blundell and Bond (2000), Bun and Windmeijer (2010) and Roodman (2009a). The "difference" estimator has often been criticized on the grounds that it is biased because of weak instrumentation. It was then recommended, as in Bond et al. (2001), to use the "system" estimator, which is considered to be more robust to weak estimation. However, recent research (see Bun and Windmeijer (2010) and Hayakawa (2007)) suggests that "system" GMM estimators may not fare any better and can be seriously biased. An additional contribution of the paper is therefore to assess the validity of the GMM approach in the aid effectiveness literature. In addition to not taking instrument strength for

[8] See Stock and Yogo (2005) and Staiger and Stock (1997). Stock and Wright (2000) and Bun and Windmeijer (2010) look at this issue in the context of GMM estimators.

[9] We use partnership characteristics to build instruments for aid but also for trade flows, see sections 3 and 5.1.

granted, we also statistically test the exclusion restrictions on which the "system" estimator relies. Papers on aid effectiveness typically eschewed these tests, whereas the restrictions are far from trivial.[10] Our results cast serious doubts on the ability of GMM estimators to identify the relevant effects, and suggest that the (consistent) two-stage least squares estimator, biased but free from weak instrumentation issues, should be considered first.

3 The instrument

This section focuses on describing in more detail our new instrument, which is the main contribution of this work.

3.1 Design

Total aid A_{it} to recipient i in year t can be decomposed as

$$A_{it} = \sum_j s_{ijt} D_{jt}, \tag{2}$$

where donors are indexed by j, D_{jt} is j's total aid budget in year t and s_{ijt} is the share of this budget allocated to recipient i. Each donor-recipient pair (i, j) in a given year t is characterized by two features: the date when the partnership was established, and how long this partnership existed. The latter is the difference between t and the entry date and is referred to as τ_{ijt}. We call κ_{ij} i's entry date position in an ordered sequence of all partnerships established by j. For instance, $\kappa_{ij} = 1$ for recipients that received aid from j in the first year j started to give aid, and so on.[11] More formally, define η_{ij} as the first year j gives aid to i and π_j as the first year the donor disburses aid to any country. The entry date order κ_{ij} is then defined as

$$\kappa_{ij} = \eta_{ij} - \pi_j + 1. \tag{3}$$

Donor portfolio expansion implies that aid shares are bound to fall on average. In order to make aid shares neutral with respect to portfolio size we define normalized

[10] On the other hand, Bond et al. (2001) argue that they must be satisfied when estimating a Solow growth model.

[11] To be precise, our data only starts in 1960, so the ordered sequence of recipients' cohorts is approximate. This is a data limitation which is akin to censoring, but on an independent variable; the econometric literature has surprisingly little to say about how to deal with this issue; see Manski and Tamer (2002) and Rigobon and Stoker (2009) for contributions on this issue.

aid shares σ_{ijt}

$$\sigma_{ijt} = s_{ijt} - \frac{1}{N_{jt}}, \tag{4}$$

where N_{jt} is the number of recipients that have received aid from donor j at least once before year t. Normalized shares are hence deviations from an equal sharing rule among all recipients.

Predicted aid shares are then the OLS fitted values of

$$\sigma_{ijt} = a + b\kappa_{ij} + c\tau_{ijt} + u_{ijt}. \tag{5}$$

Predicted aid shares for any observation (i.e. a given (i, j) pair in a given year) are fully defined by their entry date order and partnership length. In other words, to any partnership characterized by entry date of order κ and length τ we associate a predicted aid share $\hat{\sigma}_{\kappa\tau} \equiv \hat{a} + \hat{b}\kappa + \hat{c}\tau$. $\hat{\sigma}_{\kappa\tau}$ is not related to i, j, or t: it is the *typical* share (in fact, the average share) that any recipient gets from a donor if their partnership was established in the κ^{th} year of activity of this donor, τ years ago.

The instrument for aid is the predicted aid quantity

$$\hat{A}_{it} = \sum_j \hat{\sigma}_{\kappa_{ij}\tau_{ijt}} D_{jt}. \tag{6}$$

In words, we first estimate the predicted aid share each donor allocates to each recipient, based on the pair characteristics. We then multiply these predicted aid shares by the donors' aid budgets to obtain a predicted aid quantity for each recipient. The intuition is as follows. The instrument artificially recreates a situation where a country receives more aid in a given period, independently of the "fundamentals" of its economy, but rather for one or more of the following reasons: because it on average had an earlier order of entry with respect to other recipients in the donors' portfolio; because it was in the (average) partnership for a longer period of time; finally, because the (average) donor's budget for aid happened to be larger that year. Unlike actual aid A_{it}, $\hat{A}_{it}$ is not influenced by shocks to economic performance in the recipient country,[12] so it is not affected by reverse causality; moreover, we will argue in the following section that it is a strong instrument, relevant for predicting actual

[12] On the other hand, it is affected by shocks to the donor's economy, through the aid budget. For example, a boom year for one or more donor countries can lead to larger aid budgets and at the same time larger trade flows; if some of the recipients are also trade partners, which is often the case, we might erroneously attribute to aid the beneficial effects that come from other channels. However, we think that year effects do a good job of controlling for these instances. Moreover, in the robustness checks, we control for trade, which we consider to be the main potential alternative channel from having a partnership to growth.

aid flows, and that its only effect on growth occurs through the actual aid flows it proxies.

3.2 Properties

For $\hat{A}_{it}$ to be a good instrument, it must be the case that entry date order and length are strong determinants of aid shares. Frot (2009) shows that this is indeed the case, and we reproduce some of his results here. Using data on aid recipients, we group recipients into six cohorts based on entry dates: recipients with an entry date of one, then with entry dates between two and five, six and ten, eleven and fifteen, sixteen and twenty, and above twenty one. Figure 1 presents the average normalized share received by recipients in each cohort in each year.[13] In other words, Figure 1 shows how much recipients in each cohort get in deviation from equal sharing.

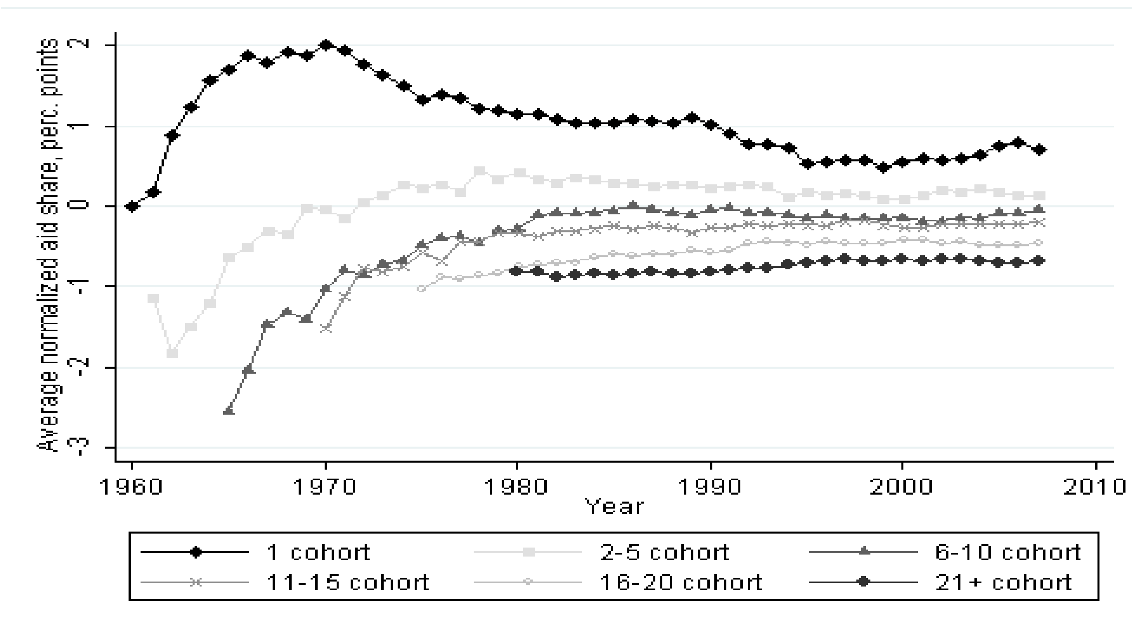

Figure 1: Average aid share in deviation from equal sharing, by recipient cohort

As shown by the figure, early entrants into donors' portfolios are on average receiving larger aid shares. There is some convergence across cohorts but even many

[13]Donors enter the market in different years, and sometimes exit the market. These changes make comparing the cohort averages difficult, so for Figure 1 we restrict the sample to donors that have been present from 1960 to 2007.

years after portfolios have been formed, it is still the case that entry dates and aid shares are correlated. Stratification by cohorts is visible in any year, and seems to have reached a certain persistence level.

Figure 1 does not alone offer enough evidence that entry dates play a decisive role in determining aid shares, neither does it exclude the case that other factors are behind the correlation between entry date order and aid receipts. It is likely that donors created partnerships that prioritized poor countries and heavily populated countries, and that such countries have received larger aid shares because of these characteristics, and not because of their entry dates. However, Frot (2009) also shows that the explanatory power of entry dates is robust to controlling for these characteristics. In order to disentangle these different possible effects, the normalized aid share of each recipient is regressed on a set of controls. The following equation is estimated:

$$\sigma_{ijt} = \alpha + \beta\tau_{ijt} + \gamma\tau_{ijt}^2 + \delta\kappa_{ij} + \boldsymbol{x_{ijt}\varphi} + \varepsilon_{ijt} \tag{7}$$

where κ_{ij} is entry date order, τ_{ijt} is the number of years the partnership has existed ($\tau_{ijt} = t - \eta_{ij} + 1$), $\boldsymbol{x_{ijt}}$ is a vector of controls including recipient GDP per capita, recipient population size, a dummy variable for whether donor and recipient shared a colonial relationship, and the distance between i and j, and ε_{ijt} is an error term uncorrelated with the independent variables. The variable τ_{ijt}^2 enters the equation to allow for convergence among countries with different entry dates. The exact functional form of the dependence of the normalized share σ_{ijt} on κ_{ij} is debatable. Equation (7) assumes that it is linear. Figure 1 suggests something more complex, with a falling effect of entry dates on aid shares (curves get closer when one moves downward vertically). To capture such non-linearities we also estimate equation (7) by adding κ_{ij}^2 as a regressor. Table 2 presents the results. Column (1) shows that entry dates are indeed affected by recipient and recipient-donor characteristics, as expected: donors did prioritize countries with a larger population, lower GDP per capita, geographically closer to them and countries with which a colonial relationship had been in place.

The remaining columns indicate that, as suggested by Figure 1, earlier entrants indeed receive larger aid quantities, even after controlling for such recipient and recipient-donor characteristics. Columns (4) and (5) acknowledge the censored nature of aid shares that are bound to lie between 0 an 1, and thus present censored regression estimates. The effects are sizable. Consider two hypothetical aid recipients A and B from the same portfolio. A and B's characteristics are identical, except that A's entry date is one and B's is ten (corresponding roughly to a one-standard deviation difference). The difference in A and B's aid shares in year 20 (20 years after they started receiving aid) is 0.99 percent using estimates from column (3), and 1.45

Table 2: Determinants of aid shares

	(1) Entry	(2) Aid share	(3) Aid share	(4) Aid share	(5) Aid share
GDP per capita	.00036***		-.00014***		-.00025***
	(.000065)		(.000018)		(.0000075)
Population, mil	-.012***		.0055***		.0059***
	(.0010)		(.0013)		(.00013)
Colony	-3.91***		2.69**		2.96***
	(1.04)		(1.03)		(.061)
Distance	.17**		-.057**		-.083***
	(.077)		(.022)		(.0040)
Entry		-.12***	-.11***	-.17***	-.15***
		(.016)	(.019)	(.0041)	(.0054)
Entry, squared		.0030***	.0032***	.0035***	.0036***
		(.00048)	(.00054)	(.00015)	(.00019)
Length		.062***	.091***	.061***	.11***
		(.0084)	(.011)	(.0033)	(.0043)
Length, squared		-.0011***	-.0018***	-.00091***	-.0019***
		(.00017)	(.00020)	(.000079)	(.000096)
Constant	6.98***	-.12	.019	-.44***	-.12**
	(1.03)	(.082)	(.26)	(.033)	(.055)
Observations	71620	132798	71620	132798	71620
Recipients	113	130	113	130	113
Donors	29	56	29	56	29
R^2	.057	.019	.098	.007	.028

Note: Robust standard errors clustered at the donor level in parentheses. Columns (4) and (5) estimate a censored-normal regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

percent from column (5). This is as large as between a quarter and 40 percent of the standard deviation of the shares distribution. To put this number into perspective, we compare it with the GDP differential that would result in such a difference. In other words, for B to have the same aid share as A, how much smaller should its per capita per capita GDP be? From the estimates of Table 2, B's income per capita would have to be USD 7071 to 5814 lower than that of A, using columns (3) and (5), respectively. The mean income per capita in the sample is USD 1712, with a standard deviation of USD 2043, so this difference is extremely large. This implies that entry dates have a large effect when compared to per capita GDPs. The small percentage difference is also significant in monetary terms, as it represents between USD 14 and 20 million (in 2006 USD). Entry dates, together with partnership length, are therefore good predictors of aid shares, on top of more traditional determinants of aid. In the next section, we will report more evidence of predicted aid indeed being a strong instrument for aid.

Returning to our question, we are ultimately interested in the effect of aid, as predicted by entry date order, on growth. Hence, we also need to ensure that there are no other confounding effects that go from entry dates to growth through other channels than aid, i.e. that exclusion restrictions are satisfied. For example, it might be the case that early entrants do not only receive more aid, but also larger trade flows, which in turn affect growth. In such a case, we would erroneously attribute to aid the better growth performance observed. A response to this concern is to control for those potential factors correlated with entry dates and affecting growth in the growth regression and show that aid has an independent effect on top of them. This is done for trade flows.[14] We also show that the direct correlation between entry order and growth, although present, is very weak, and there is no strong evidence against the claim that it might come entirely and only through aid.

[14]An issue with directly including trade flows in the estimation of equation (1) is that they, too, are affected by reverse feedback with the growth rate, the left-hand side variable. Our approach is to instrument them, too, in a similar way as we did for aid flows, using entry dates and partnership length. The predicted total trade flows are then included in the equation. Initially, we also took the same approach for inward foreign direct investment flows, but then abandoned this part of the analysis due to serious limitations in the bilateral FDI data.

4 Results

4.1 Preliminary stage

As mentioned above, our strategy consists of first estimating aid shares by regressing actual shares on entry dates and partnership length (and their squares).[15] We then compute predicted aid quantities $\hat{A}_{it}$ by summing up predicted aid shares multiplied by donors' aid budgets. The predicted aid quantity is then used as an external instrument in the "second stage" growth regression (i.e. in equation (1)).

4.2 Baseline results

Table 3 reports the OLS and IV estimation of equation (1) with and without country fixed effects.[16] The equation includes a number of controls which are frequently used in the literature: population size; a measure of schooling[17]; inflation as a measure of macroeconomic policies; liquid assets (M2/GDP), commonly used as a measure of financial depth; institutional quality, measured by the International Country Risk Guide (ICRGE) index; the Sachs et al. (1995) index of openness. We also include ethno-linguistic fractionalization and regional dummies for Sub-Saharan Africa and quickly growing East Asia, when possible. These controls are those most commonly used in the aid effectiveness literature, and allow us to draw comparisons with past studies.

The log-log specification adopted in equation (1) implies that the coefficient on aid is the elasticity of GDP with respect to aid. We start by not instrumenting the aid variable, and present naive estimates, with and without country fixed effects in Table 3. Column (1) confirms the traditional finding that, when not instrumented, aid has no effect on GDP growth. The inclusion of country fixed effects only reinforces this conclusion. However, as argued above, there is little to learn from regressions where aid is not instrumented. We move on to columns (3) and (4) where aid is instrumented using our instrument of predicted aid quantities. Because a major concern in the literature is the weakness of instrumentation for aid, we provide two statistics. The first is the p-value of the Angrist and Pischke (2009) test of excluded

[15]The specification we use to predict aid shares corresponds to Table 2 column (4). The correlation between predicted and actual shares is 46%, 48% between predicted and actual aid quantities.

[16]All regressions in the paper include year effects. Refer to the data appendix for all variable definitions and their sources.

[17]This is the Barro and Lee (2010) average years of primary schooling. Whether we use primary or secondary schooling does not make much difference.

Table 3: OLS and IV regressions

	(1) OLS	(2) FE OLS	(3) 2SLS	(4) FE 2SLS
Log GDP, lagged	-0.020	-0.23***	-0.030*	-0.23***
	(0.016)	(0.047)	(0.018)	(0.050)
Log aid, lagged	0.018	0.017	-0.017	0.10***
	(0.011)	(0.017)	(0.050)	(0.039)
Log population	0.023	-0.16	0.052	-0.23**
	(0.021)	(0.12)	(0.039)	(0.10)
Inflation	-0.10***	-0.11***	-0.11***	-0.10***
	(0.025)	(0.028)	(0.025)	(0.028)
Money, lagged	0.00069	0.0028**	0.0011	0.0029***
	(0.00060)	(0.0013)	(0.00086)	(0.0011)
Schooling	0.00029	-0.0013	-0.0028	0.033
	(0.012)	(0.036)	(0.012)	(0.037)
Institutional quality	0.013**	0.0077	0.014**	0.0042
	(0.0055)	(0.0073)	(0.0056)	(0.0070)
Openness	0.097***	0.075***	0.10***	0.078***
	(0.017)	(0.027)	(0.018)	(0.028)
Ethno. fractionalization	-0.083		-0.097	
	(0.057)		(0.063)	
East Asia	0.014		0.018	
	(0.026)		(0.028)	
Sub-Saharan Africa	-0.020		-0.0086	
	(0.038)		(0.047)	
Observations	347	347	347	344
Countries	61	61	61	58
AP test (p-val)			0.046	0.00088
KP F stat			4.16	12.3
R^2	0.32	0.38	0.29	0.28

Note: AP: Angrist-Pischke. KP: Kleibergen-Paap. DThe dependent variable is the growth rate. All regressions include year effects. Robust standard errors clustered at the recipient level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

instruments.[18] The second is the Kleibergen and Paap (2006) Wald statistic. Both are tests of instrument weakness.

Column (3) is the two-stage least square (2SLS) specification. It fails to find any significant effect of aid on GDP growth, but it is likely that omitted country fixed characteristics make the error term not orthogonal to the control variables, biasing the estimates. In addition, the Kleibergen and Paap Wald statistic is quite low.[19] In column (4), we include country fixed effects to avoid the bias due to their omission. The consequence for the aid coefficient is quite dramatic. It is much larger than in column (2) and comfortably passes the five percent significance threshold. The weak instruments statistics now confirm that our instrument is highly correlated with aid. The null hypothesis of the Angrist and Pischke test is strongly rejected, and the Kleibergen and Paap Wald statistic is much higher than in column (3). These results indicate that the inclusion of fixed effects is important for the validity of our approach.[20] The estimated effect implies an elasticity of GDP with respect to aid of 0.10. This elasticity is relatively moderate. Another way of interpreting the result is that a 1 percent change in aid increases GDP growth by approximately 0.10 percent. The first-stage regression for the fixed effect regression is shown in column (1) of Table 4. It confirms that predicted aid is a strong predictor of actual aid.

The results in Table 3 are problematic because of the correlation between lagged income and the error term, due to the strong persistence in income and the individual specific component in the error term. We can sign this bias for the lagged income coefficient, but not so easily for the other variables. It is nevertheless useful in order to evaluate the performance of the GMM estimator. In the OLS setting, the coefficient on lagged income is upward biased, whereas Nickell (1981) proved that the within group estimator is downward biased. We know that the true coefficient lies somewhere in this range, and this remark allows us to evaluate if the GMM estimator succeeds in removing the bias. In columns (1) and (2) of Table (5), we rely on the difference GMM estimator in order to remove the dynamic bias, in asymptotic terms. This method estimates the model in differences, to get rid of the fixed effects. The lags of endogenous regressors, which are exogenous to the first difference of the error term, are used to instrument for their first difference. In column (1), we instrument aid with its lags, as is usually done in the literature. In column (2), we use our

[18] With a single endogenous regressor, this statistic is simply the F-statistic of the first stage.

[19] Although critical values only exist for the Cragg-Donald Wald statistic, which is not robust to heteroskedasticity, the 25% maximal IV size value is 5.53, which suggests that the Kleibergen-Paap statistic is indeed low.

[20] This test is based on the F-statistic of the first stage, so the stark improvement is not surprising: the model including country fixed effects performs much better than that without.

Table 4: First stages

	(1) Aid	(2) Aid	(3) Trade
Log GDP, lagged	-0.088	-0.11	1.07***
	(0.27)	(0.27)	(0.13)
Log predicted aid, lagged	1.42***	1.69***	-0.025
	(0.41)	(0.40)	(0.13)
Log predicted trade, lagged		-1.12***	1.06***
		(0.35)	(0.21)
Log population	0.63	0.71	-0.31
	(0.70)	(0.69)	(0.25)
Inflation	-0.073	-0.078	0.0097
	(0.10)	(0.10)	(0.048)
Money, lagged	0.0016	0.00067	0.0068***
	(0.0052)	(0.0052)	(0.0025)
Schooling	-0.28*	-0.30*	0.047
	(0.16)	(0.16)	(0.10)
Institutional quality	0.056***	0.056***	-0.0099
	(0.021)	(0.019)	(0.015)
Openness	0.053	0.023	0.11
	(0.13)	(0.13)	(0.081)
Countries	58	58	58
R^2	0.35	0.37	0.78
Observations	344	343	343

Note: Column (1) is the first stage of the regression in Table 3 column (4). Columns (2) and (3) are the first stages of the regression in Table 7 column (4). The instruments for aid and trade are built from fitted values of the preliminary stage estimated at the bilateral level, and then aggregated at the country level. All regressions include country and year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

instrument for aid. Both coefficients are insignificant, but the estimates of the GDP coefficient cast serious doubts on the validity of the approach. The first-differenced GMM estimate is well below the within groups estimate of Table 3, which can already be expected to be strongly downward biased, given the small time dimension of the dataset. This signals that the GMM estimate is also biased, possibly because of weak instruments.[21] The first-differenced GMM estimator is therefore not very informative and for this reason, we use the system GMM estimator in the next two columns. This is a more efficient method, developed by Blundell and Bond (2000); it uses the moment conditions from the same difference equation *but also* from the original level equation at the same time. This method is valid under the assumption that the GMM instruments (i.e. the lagged differences) are exogenous to the error term in the level equation. This can be tested using a Hansen test, denoted in the regression tables as "level eq."; we report the test p-value. If this test fails, the validity of the system GMM approach is questionable, and the results should be interpreted with caution. Finally, the p-value of the Hansen J test of overidentification is reported; its null hypothesis must not be rejected for the GMM exclusion restrictions to be valid.

The GDP coefficient, both in columns (3) and (4), now lies in the expected range, which confirms that system GMM estimators are more appropriate. Estimates in column (3) do not make use of our aid instrument. The null hypothesis of the exogeneity of the GMM instruments for the levels equation is not rejected by the Hansen test. On the other hand, the number of instruments and countries is of the same order of magnitude, such that the p-value of the test is likely to be upward biased, as underlined by Roodman (2009b). Column (4) instruments aid with our instrument. The coefficient of aid is now significant, although only with a p-value of 6.9 percent. It is also smaller than in Table 3. The Hansen tests of overidentification restrictions and of exogeneity that the GMM instruments for the levels equation fail to reject their null hypotheses, suggesting that the assumptions required for the estimators to be valid are satisfied. Taken together, columns (4) of Tables 3 and 5 indicate an elasticity of GDP to aid between 0.057 and 0.10.

GMM estimators come with several caveats about their validity, however. The first concerns the risk of having too many instruments. Roodman (2009b) showed how Hansen tests tend to fail to reject the null hypothesis when the instrument count is large. A rule of thumb is that instruments should not exceed the number of countries, which is the case in our estimations. Relying on our external aid instrument in column (4) reduces the instrument count, but it still remains close to the number

[21] Here we mean the GMM instruments. We do not rely on them for the aid variable, and we know that our instrument for aid is actually strong.

Table 5: GMM regressions

	(1) Diff. GMM	(2) Diff. GMM	(3) Sys. GMM	(4) Sys. GMM
Log GDP, lagged	-0.38***	-0.54***	-0.016	-0.086*
	(0.12)	(0.14)	(0.027)	(0.050)
Log aid, lagged	-0.014	0.016	0.018	0.057*
	(0.030)	(0.040)	(0.017)	(0.031)
Log population	-0.073	-0.027	0.022	0.057
	(0.14)	(0.16)	(0.033)	(0.047)
Inflation	-0.11***	-0.076***	-0.086**	-0.080**
	(0.033)	(0.027)	(0.034)	(0.038)
Money, lagged	0.0019	0.0030**	0.0016**	0.0016*
	(0.0012)	(0.0012)	(0.00077)	(0.00090)
Schooling	-0.052	-0.047	-0.0072	0.058
	(0.050)	(0.063)	(0.027)	(0.051)
Institutional quality	0.0095	0.014**	0.013**	0.019**
	(0.0073)	(0.0072)	(0.0050)	(0.0085)
Openness	0.080***	0.088***	0.11***	0.11***
	(0.031)	(0.032)	(0.022)	(0.022)
Instruments	60	40	74	48
Countries	58	58	61	61
Hansen J test (p-val)	0.55	0.27	0.76	0.27
Hansen test (p-val), lev.			0.95	0.11
AR(1)	0.016	0.12	0.00081	0.0010
AR(2)	0.57	0.28	0.80	0.67
Observations	286	286	347	347

Note: Instruments for the differences equation are log GDP lagged twice in all specifications, and log aid lagged twice in columns (1) and (3). Instruments for the levels equation are log GDP lagged and differenced once in columns (3) and (4), and log aid lagged and differenced once in column (3). In columns (2) and (4), log predicted aid is used as an instrument. All regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of countries. Roodman (2009b) recommends that the instrument count is reduced as a robustness check. One way of doing this is to collapse the instrument matrix. Estimates with this collapsed matrix are reported in Column (1) of Table 6. The coefficient on aid is now smaller, and insignificant, but maybe more importantly, the Hansen tests now strongly reject the overidentification restrictions and the validity of the instruments for the level equation.[22] This suggests that the p-value of these tests in Table (5) were inflated by the number of instruments, and that system GMM estimators are based on questionable assumptions.

There is one more concern: unlike IV regressions in Table 3, no test of instrument strength is available in a GMM setting. Bazzi and Clemens (2009) argue that weak instruments are a major concern with these estimators, and suggest a replication of the GMM instrumentation in a traditional IV setting, where such tests exist.[23] We follow their advice and re-create the matrix of GMM instruments for the difference and system equations, reporting the estimations in columns (2) and (3). We can now report the Kleibergen-Paap statistics about instrument strength and the Kleibergen-Paap LM test of underidentification. Column (2) shows that the Wald statistic is very low for the difference equation, and that even the underidentification null hypothesis cannot be rejected. Since we know from Table 3 that predicted aid is not a weak instrument for aid, these signs of weak instrumentation must be due to the GMM instruments. This implies, among other things, that lagged GDP levels are very weak instruments for GDP differences. This is not very surprising, given that the difference GMM estimators performed poorly. Such weakness usually justifies the use of system GMM, as it is believed to be more robust. The fact that difference GMM performed so poorly implies that the identification relies heavily on the levels equation. Instrument strength in this equation is therefore crucial for the whole system GMM. But column (3) actually reveals that the instrumentation of this equation is even worse than for the difference equation.

All in all, from Table 6, we conclude that our system GMM estimates suffer from two severe drawbacks. First, exogeneity tests are rejected when the set of instruments is shrunk. Second, GMM instruments appear to be extremely weak. These points lead us to infer that the GMM approach may not improve the fixed effects specification. Instruments with this level of weakness imply that conclusions about estimated coefficients are fragile. Given these results, Table 3 column (4) remains our preferred specification.

[22]Collapsing the instrument matrix when using the GMM instruments for aid also leads to the rejection of these null hypotheses.

[23]Blundell and Bond (2000), Bun and Windmeijer (2010), Hayakawa (2007) and Roodman (2009a) make the same recommendation.

Table 6: Instrument collapsing and weak instruments

	(1) Collapse	(2) Difference	(3) System
Log GDP, lagged	-0.044	-1.00***	1.35
	(0.062)	(0.21)	(8.47)
Log aid, lagged	0.023	0.054	-0.85
	(0.043)	(0.049)	(6.08)
Log population	0.035	0.33	-0.71
	(0.063)	(0.26)	(4.23)
Inflation	-0.088**	-0.075***	-0.93
	(0.040)	(0.019)	(5.35)
Money, lagged	0.0018**	0.0043***	-0.000099
	(0.00090)	(0.0011)	(0.012)
Schooling	0.013	0.067	-1.38
	(0.059)	(0.085)	(8.99)
Institutional quality	0.012	0.018***	-0.12
	(0.010)	(0.0054)	(0.84)
Openness	0.10***	0.049*	0.13
	(0.024)	(0.027)	(0.35)
Instruments	22		
Countries	61	58	61
Hansen J test (p-val)	0.013		
Hansen test (p-val), level	0.0044		
AR(1)	0.0014		
AR(2)	0.79		
KP LM test (p-val)		0.24	0.88
KP F stat		2.05	0.011
Observations	347	286	340

Note: KP: Kleibergen-Paap. In column (1) system GMM is used, like in Table 5, column (4), but the matrix of GMM-type instruments is collapsed. Column (2) and (3) are 2SLS regressions where variables are instrumented using GMM-type instruments. In column (2) all variables are differenced once, and instrumented using log predicted aid and lagged log GDP levels. In column (3), instruments are log predicted aid and differenced log GDP. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Robustness

5.1 Instrument exogeneity

Exogeneity of our unique aid instrument cannot be statistically tested in the absence of other valid instruments. Exogeneity will be violated if predicted aid affects GDP growth through other channels than aid. This will happen if aid partnerships are correlated with other variables that influence GDP. Our concern is that our instrument captures a larger effect, from many causes, with aid only being one of its components. If we do not control for the other components, we will wrongly attribute the causal effect in its entirety to aid. For instance, countries engaged in a long-term aid partnership may also exchange valuable information about innovation or technological progress that have nothing to do with aid, but that reflect the specific nature of the relationship between the two countries. It is very difficult to directly control for these exchanges but other channels may capture these effects. An important variable very likely to be influenced by partnership characteristics is trade. We would expect that two countries engaged in a very strong aid partnership would also engage in other economic exchanges, and that trade would be a prominent one. If our instrument is just a correlate of trade, then it is likely that the effect we are measuring comes from trade, but not from aid.

To control for this possibility, we include trade, defined as the sum of exports toward and imports from donor countries, in the previous specifications. We construct a trade instrument using the same strategy as for the aid instrument. Using aid entry dates, we compute a predicted trade quantity for each bilateral trade partnership and obtain a predicted trade quantity by summing these up.[24]

Table 7 shows that controlling for trade only marginally changes the results. The effect of aid in the 2SLS fixed-effect regression has a similar size and is significant. Our trade instrument also appears to be strong, as is confirmed by the Angrist-Pischke *p*-value of the first-stage regression for trade, and the relatively high Kleibergen-Paap F statistic. The first two stages are presented in columns (2) and (3) of Table 4. We also included our trade variable in the GMM estimations. As in Table 6, these results cast serious doubts on the validity of the GMM approach in this setting, but we gain no new insight from this exercise.[25]

A stronger concern would be that some unobserved trait of the recipient country

[24]This is done because otherwise the simultaneity between trade and growth would once more bias the estimations. A reason for using the aid partnership entry dates to instrument trade flows is that we are especially interested in capturing the part of those flows that correlates with our aid instrument.

[25]Estimation tables are available in an appendix.

Table 7: OLS and IV regressions, with trade flows

	(1) OLS	(2) FE OLS	(3) 2SLS	(4) FE 2SLS
Log GDP, lagged	0.020	-0.20***	0.032	-0.33***
	(0.023)	(0.061)	(0.090)	(0.084)
Log aid, lagged	0.015	0.016	0.014	0.088***
	(0.011)	(0.017)	(0.021)	(0.033)
Log trade, lagged	-0.050**	-0.030	-0.065	0.092
	(0.024)	(0.040)	(0.10)	(0.059)
Log population	0.022	-0.17	0.022	-0.20*
	(0.022)	(0.12)	(0.023)	(0.10)
Inflation	-0.11***	-0.11***	-0.11***	-0.10***
	(0.023)	(0.028)	(0.025)	(0.028)
Money, lagged	0.0014**	0.0027**	0.0016	0.0021*
	(0.00069)	(0.0012)	(0.0016)	(0.0013)
Schooling	0.0030	-0.0047	0.0037	0.023
	(0.0100)	(0.036)	(0.012)	(0.037)
Institutional quality	0.013**	0.0077	0.013**	0.0065
	(0.0053)	(0.0077)	(0.0054)	(0.0065)
Openness	0.093***	0.075***	0.092***	0.069**
	(0.016)	(0.027)	(0.021)	(0.029)
Ethno. fractionalization	-0.046		-0.034	
	(0.049)		(0.099)	
East Asia	0.039		0.045	
	(0.031)		(0.055)	
Sub-Saharan Africa	-0.0090		-0.0055	
	(0.037)		(0.039)	
Observations	346	346	346	343
Countries	61	61	61	58
AP test (p-val), aid			0.000000061	0.00010
AP test (p-val), trade			0.030	0.0000036
KP F stat			2.49	12.3
R^2	0.33	0.39	0.33	0.27

Note: AP: Angrist-Pischke. KP: Kleibergen-Paap. All regressions include year effects. Robust standard errors clustered at the recipient level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

that promotes growth also has a direct effect on the starting date and/or the duration of the donor-recipient relationship, i.e. the building blocks of our instrument. For example, if donors are reluctant or unable to establish a partnership in places with despotic rulers or in places with persistent conflicts, this could at the same time delay the entry of those countries into donors' portfolios and limit growth. This would result in a negative correlation between entry date and growth, biasing upward the coefficient on aid in the main regression. In the first column of Table 8 we see that, indeed, countries with a later entry date did experience a lower growth rate. This simple correlation disappears, though, after controlling for the initial level of GDP and population size, arguably strong determinants of subsequent growth rates. In column (3) we also control for total aid received. The idea would be to check if, although the time of entry in a development cooperation partnership has an effect on growth, this effect goes through aid and only through aid. The direct inclusion of aid quantity in this regression is problematic, given the endogeneity of aid to growth, so we do not put too much weight on this last model.

In the regressions reported in columns (1) to (3), the observations are at the partnership level: this implies that each recipient country has many entry dates (one for each donor) and only one growth rate for each time period. In columns (4) to (6), we collapse the observations at the recipient country level, using the aid quantities as weight for donor countries: therefore, each recipient will only have one average entry date, which will be earlier if the most important donors in terms of aid given started their partnership with this country earlier, and vice versa. Even the simple correlation disappears in this setting. There results show that entrants with different entry dates do not on average differ from a GDP growth point of view and thus, is further suggestive evidence that our instrument is indeed exogenous.

5.2 Outliers

Easterly et al. (2004) showed how aid effectiveness results could be sensitive to the exclusion of a few outliers. We make use of the Hadi (1992) procedure to exclude outliers from the sample. Both with the within groups estimator and in the system GMM regressions, we find larger effects of aid than when all observations are used. The elasticity of GDP with respect to aid increases by 60 percent, with and without controlling for trade.

Figure 2 shows the two partial regression plots of GDP growth on instrumented aid, with and without outliers. We conclude that outliers tended to bias our estimates downward and hence, the GDP elasticity with respect to aid is possibly much larger.[26]

[26]Estimation tables are available in an appendix.

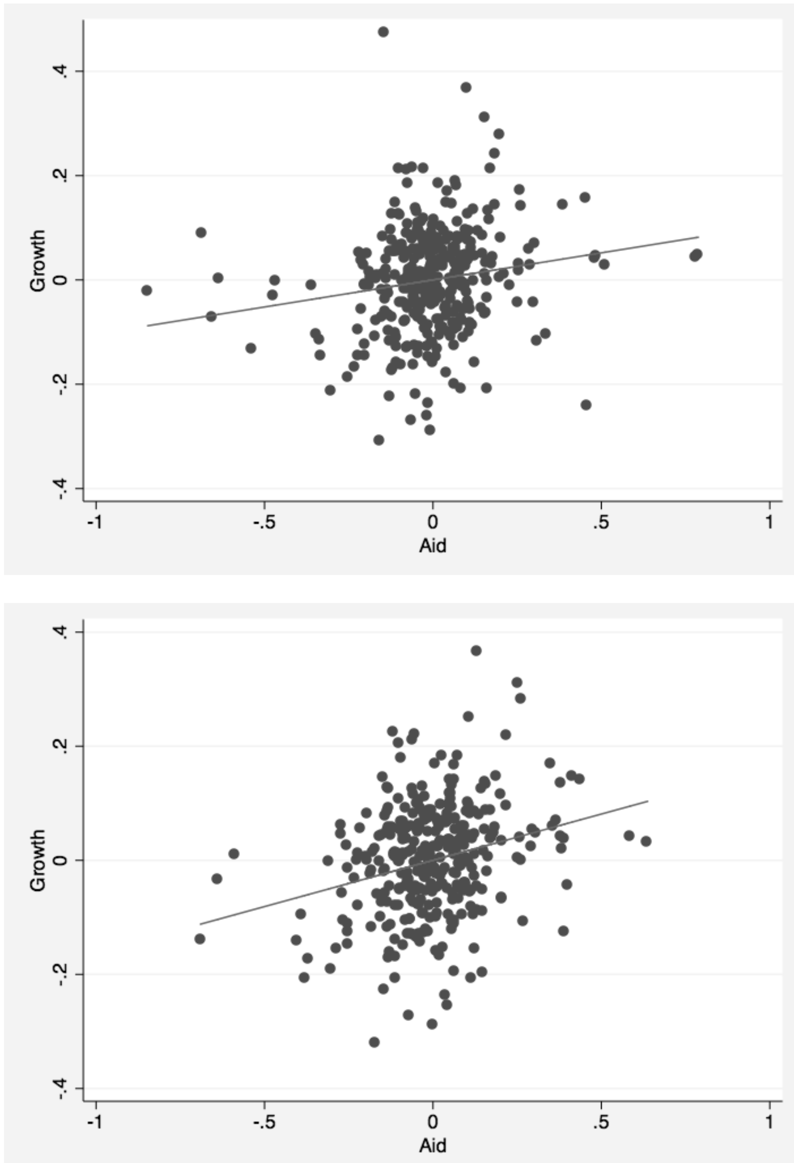

Figure 2: Partial regression plot of growth on aid, including and excluding outliers

Table 8: Correlation between entry date and growth

	Whole Sample			Collapsed		
	(1)	(2)	(3)	(4)	(5)	(6)
Entry	-0.00050**	-0.00025	-0.00018	-0.0024	-0.0030	-0.0028
	(0.00020)	(0.00022)	(0.00022)	(0.0023)	(0.0039)	(0.0040)
Log GDP, lagged		-0.0031**	0.0031*		0.00085	0.0034
		(0.0015)	(0.0018)		(0.0072)	(0.0085)
Log population		0.0054***	-0.011***		-0.0024	-0.0091
		(0.0013)	(0.0024)		(0.0086)	(0.012)
Log aid			0.021***			0.0092
			(0.0022)			(0.013)
Observations	20974	20974	20974	812	812	812
Countries	112	112	112	112	112	112
R^2	0.041	0.042	0.048	0.052	0.052	0.053

Note: The dependent variable is the growth rate over five years. All regressions include year effects. Columns (1)-(3) include one observation for each recipient-donor pair every five years; observations in columns (4)-(6) are the weighted average for each recipient and five-year period, where each donor is weighted with the total aid quantity donated to that specific recipient during the five-year period. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3 Sample size

Our previous specifications include control variables that are commonly found in the aid effectiveness literature. However, limited data availability sharply reduces the sample size. Our dataset contains 130 countries but the regressions rely on 61 countries at most. Larger sample size comes at the cost of omitting some growth determinants and hence, potentially biases the aid coefficient. On the other hand, the aid instrument, if truly exogenous, should remove the correlation between aid and the error term even in the presence of omitted variables. This provides an indirect test of instrument validity, in addition to extending the estimation to many more countries. The most parsimonious specification with only lagged GDP, aid, and population as controls, allows us to use data on 108 countries, a dramatic increase. Aid is not significant in any of the regressions.

Because we include as few controls as possible in these regressions, there may be strong outliers in these specifications. We put this result to the test of excluding outliers, once more following the Hadi procedure. Table 9 confirms that these data points strongly influence the results, despite representing a very small group of observations (the procedure excludes 7 observations). This is visually confirmed by the partial regression plots of growth on aid shown in the Appendix. First, aid becomes significant, even when it is not instrumented. This result is not robust to the inclusion of additional controls, as shown earlier, and thus has little meaning in itself. More interesting is the within-groups estimate when aid is instrumented, in column (4). The coefficient is significant, and its size almost the same as with the controls (see Table 7 column (4)). This is further encouraging evidence of our instrument being valid.[27]

Table 9: OLS and IV regressions excluding outliers, large sample

	(1) OLS	(2) FE	(3) 2SLS	(4) FE-2SLS
Log GDP, lagged	0.0046	-0.21***	-0.0057	-0.21***
	(0.0063)	(0.039)	(0.010)	(0.038)
Log aid, lagged	0.022***	0.023*	-0.022	0.085**
	(0.0073)	(0.013)	(0.033)	(0.037)
Log population	-0.011	0.022	0.021	-0.079
	(0.0097)	(0.089)	(0.026)	(0.11)
Countries	108	108	108	104
AP test (p-val)			0.00039	0.0000061
KP F stat			13.4	22.7
R^2	0.073	0.23	0.025	0.18
Observations	710	703	710	696

Note: AP: Angrist-Pischke. KP: Kleibergen-Paap. All regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, we use GMM estimators on the same large sample. The difference GMM estimator once more fails to produce correct estimates of the lagged GDP coefficient; while using system GMM, the aid coefficient is very close to the one in Table 5. This once more tends to confirm that our results are robust. As before, though, GMM estimates appear to be quite fragile. On the other hand, we find it encouraging that all our system GMM specifications find an elasticity close to 0.05.

These robustness tests confirm our earlier results that aid has a significant and positive impact on growth. The elasticity of GDP with respect to aid is found to lie between 0.05 and 0.16, depending on the estimators used and the exclusion of outliers from the regression sample.

[27]In fact, the coefficient on aid is virtually unchanged even without controlling for population and initial GDP. Tables can be obtained from the authors upon request.

5.4 Aid as a share of GDP

We depart from the aid effectiveness literature by measuring aid in constant dollars, while past research traditionally used aid as a share of GDP.[28] This departure was done to avoid introducing additional endogeneity in the aid variable. It is indeed peculiar to strive to remove reverse causality from GDP to aid by using instrumental variables and then re-introducing GDP as a denominator. We prefer to instead use aid quantities. This offers other advantages: first, the log-log specification directly estimates the elasticity of GDP with respect to aid; moreover, since lagged log GDP enters equation (1), the particular case with aid as a share of GDP can be seen as a special case of equation (1), albeit in its log form, while our aid-quantity specification would be the more general case.

Nevertheless, and despite the fact that instrumentation is likely to be more problematic, we feel that we cannot completely ignore the past convention and, in Table 10, we present results where the aid variable is expressed in GDP percentage points. The trade variable is also computed as a share of GDP, while other controls are the same as in previous tables.

Columns (1) and (2) are based on the whole sample, and columns (3) and (4) exclude outliers identified by the Hadi procedure. The first two columns show that aid has no effect on GDP, but the next two reveal that this is due to very few outliers (only 13 observations are excluded from column (2) to column (4)). As in Table 3, the aid coefficient differs significantly from zero only when instrumented. Removing outliers does not only increase the aid coefficient, it also improves the identification, as shown by the Angrist-Pischke test and the Kleibergen-Paap statistic.[29]

In Table 11, system GMM estimators are used to remove the bias induced by the dynamic nature of the specification.[30] Column (1) presents results based on the full sample, and column (2) excludes the outliers. Columns (3) and (4) control for trade. In both specifications, aid turns out to be significant once outliers are excluded, with p-values of 6.1 and 7.8 percent in columns (2) and (4), respectively. The size of the coefficient is smaller than with the within groups estimator. Although the Hansen

[28] Another departure is the definition of the growth variable that can be measured between the beginning and the end of the time period, or as an average of yearly growth rates. We return to this point in Appendix 7.2, as the results are not affected by this change of definition.

[29] We do not present results using OLS and 2SLS estimators, however aid coefficients are not significantly different from zero in any of them, with and without outliers.

[30] We focus on system GMM rather than difference GMM for the same reason as in Section 4. The lagged log GDP coefficients with difference GMM are well below their FE estimates, such that the difference GMM estimator must be severely biased and thus is not reliable. Tables are available from the authors upon request.

Table 10: OLS and IV regressions, aid as a share of GDP

	(1) FE	(2) FE 2SLS	(3) FE	(4) FE 2SLS
Log GDP, lagged	-0.23***	-0.28***	-0.23***	-0.15**
	(0.046)	(0.053)	(0.046)	(0.065)
Aid, share of GDP	0.15	-1.19	0.21	2.15**
	(0.25)	(0.80)	(0.31)	(1.02)
Log population	-0.15	-0.099	-0.16	-0.26***
	(0.12)	(0.15)	(0.12)	(0.081)
Inflation	-0.11***	-0.12***	-0.12***	-0.11***
	(0.027)	(0.024)	(0.028)	(0.030)
Money, lagged	0.0028**	0.0023	0.0024*	0.0026**
	(0.0013)	(0.0015)	(0.0012)	(0.0010)
Schooling	-0.0012	-0.061	0.0038	0.058
	(0.034)	(0.044)	(0.032)	(0.044)
Institutional quality	0.0085	0.0068	0.0066	0.0026
	(0.0072)	(0.0059)	(0.0061)	(0.0053)
Openness	0.075***	0.075***	0.064***	0.060**
	(0.027)	(0.026)	(0.023)	(0.024)
Countries	61	58	60	57
AP test (p-val)		0.028		0.0011
KP F stat		5.08		11.8
R^2	0.38	0.28	0.39	0.25
Observations	347	344	340	331

Note: AP: Angrist-Pischke. KP: Kleibergen-Paap. All regressions include year and country fixed effects. Outliers, identified through the Hadi procedure, are excluded from the sample in columns (3) and (4). Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

tests do not reject the GMM approach, we remain wary of these estimations where instrumentation is very weak.

Finally, in Table 12, we check that using aid as a share of GDP does not solve the issue previously encountered in Table 6. We replicate those specifications using the new aid variable. Column (1) runs the system GMM estimation collapsing the instrument matrix, and fails to reject the validity of the system GMM assumption. On the other hand, columns (2) and (3) show that the GMM instruments, both for the difference and system equations, are very weak.[31]

Our conclusions are therefore mostly robust to the change in aid measurement. When properly instrumented for, aid has a positive and significant effect on GDP. Our estimates with this new variable range from 1.18 to 2.15. These can be related to our former estimates. If γ_1 and γ_2 are the aid coefficients using log aid and aid as a share of GDP, then computing marginal effects, we should have $\gamma_1 = \frac{A_{t-1}}{Y_{t-1}}\gamma_2$. The mean of aid per GDP in the regression sample is 0.052, such that the corresponding γ_1 lies between 0.062 and 0.11. The actual estimates are between 0.057 and 0.16, so the two specifications lead to similar results.

6 Conclusions

In this article, we proposed a new instrument for identifying the causal effect of aid on growth. This instrument takes the supply side approach that relates to the aid allocation decision a step further, for the first time using a source of variation that is not just external but exogenous to growth. As far as possible, the instrument is shown to be valid and strong. We claim that this is an improvement from a stream of papers that relied on weak and non-exogenous instruments.

When it comes to the estimation strategy and the choice of estimator, we make simple and clear methodological choices, explain and motivate them step by step and probe their validity as best as we can. In particular, we do not take it for granted that GMM estimators provide strong instruments and thus solve any dynamic bias. On the contrary, we show that they should be used with much caution as the cure may be worse than the disease. Instrument weakness is so prominent that estimates are at best fragile, at worst misleading.

The effects uncovered by our identification strategy are statistically significant and robust to various specifications. They indicate an elasticity of GDP with respect to aid that lies around 0.10.

[31] This is also the case when outliers are excluded.

Table 11: GMM regressions, aid and trade as shares of GDP

	(1) Sys. GMM	(2) Sys. GMM	(3) Sys. GMM	(4) Sys. GMM
Log GDP, lagged	0.037	0.028	0.041	0.038
	(0.041)	(0.037)	(0.034)	(0.039)
Aid, share of GDP	0.94	1.18*	0.89	1.26*
	(0.61)	(0.63)	(0.56)	(0.71)
Trade, share of GDP			-0.045	-0.060
			(0.16)	(0.21)
Log population	-0.0082	0.0040	-0.017	-0.0056
	(0.034)	(0.026)	(0.029)	(0.034)
Inflation	-0.11***	-0.10***	-0.11***	-0.11***
	(0.022)	(0.021)	(0.022)	(0.021)
Money, lagged	0.0015**	0.0012**	0.0015**	0.0011
	(0.00061)	(0.00063)	(0.00076)	(0.00080)
Schooling	-0.013	0.00095	-0.018	0.0038
	(0.029)	(0.021)	(0.028)	(0.023)
Institutional quality	0.0082	0.0088	0.0076	0.0069
	(0.0055)	(0.0054)	(0.0050)	(0.0061)
Openness	0.095***	0.097***	0.091***	0.089***
	(0.022)	(0.024)	(0.024)	(0.028)
Instruments	48	48	49	49
Countries	61	61	61	60
Hansen J test (p-val)	0.62	0.83	0.68	0.83
Hansen test (p-val), level	0.57	0.61	0.65	0.51
AR(1)	0.00095	0.00015	0.00082	0.00012
AR(2)	0.93	0.81	0.94	0.76
Observations	347	335	347	334

Note: Instruments for the differences equation are log GDP lagged twice. Instruments for the levels equation are log GDP lagged and differenced once. Log predicted aid and trade are used as instruments in all regressions. Columns (2) and (4) exclude outliers. All regressions include year effects. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Instrument collapsing and weak instruments, aid as share of GDP

	(1) GMM Collapse	(2) 2SLS-Difference	(3) 2SLS-System
Log GDP, lagged	0.20	-1.19***	0.27
	(0.20)	(0.20)	(0.22)
Aid, share of GDP	2.68	-0.57	4.04
	(2.52)	(1.06)	(3.44)
Log population	-0.12	0.44	-0.15
	(0.14)	(0.30)	(0.15)
Inflation	-0.14***	-0.066***	-0.23*
	(0.051)	(0.020)	(0.13)
Money, lagged	-0.00016	0.0039**	-0.0016
	(0.0019)	(0.0015)	(0.0025)
Schooling	-0.023	0.066	0.0065
	(0.048)	(0.10)	(0.065)
Institutional quality	-0.0037	0.019***	-0.0066
	(0.012)	(0.0051)	(0.016)
Openness	0.074*	0.045	0.020
	(0.041)	(0.035)	(0.072)
Instruments	22		
Countries	61	58	61
Hansen J test (p-val)	0.65		
Hansen test (p-val), level	0.41		
AR(1)	0.0032		
AR(2)	0.86		
KP LM test (p-val)		0.54	0.15
KP F stat		1.99	1.20
Observations	347	286	340

Note: KP: Kleibergen-Paap. Column (1) presents GMM estimations, (2) and (3) are 2SLS regressions where variables are instrumented using GMM-type instruments and log predicted aid. Instruments for the differences equation are log GDP lagged twice. Instruments for the levels equation are log GDP lagged and differenced once. Log predicted aid is used as an instrument in all the regressions. All the regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *The American Economic Review 91*(5), 1369–1401.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Arellano, M. and S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies 58*(2), 277–297.

Barro, R. J. and J.-W. Lee (2010, April). A new data set of educational attainment in the world, 1950-2010. Working Paper 15902, NBER.

Bazzi, S. and M. Clemens (2009, May). Blunt instruments: On establishing the causes of economic growth. Working Paper 171, Center for Global Development.

Bertocchi, G. and F. Canova (2002, December). Did colonization matter for growth? an empirical exploration into the historical causes of africa's underdevelopment. *European Economic Review 46*(10), 1851–1871.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Blundell, R. and S. Bond (2000). GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews 19*(3), 321–340.

Bond, S., A. Hoeffler, and J. Temple (2001, September). GMM estimation of empirical growth models. Discussion Papers 01/525, Nuffield College.

Boone, P. (1996). Politics and the effectiveness of foreign aid. *European Economic Review 40*(2), 289 – 329.

Bun, M. and F. Windmeijer (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *The Econometrics Journal 13*(1), 95–126.

Burnside, C. and D. Dollar (2000). Aid, policies, and growth. *The American Economic Review 90*(4), 847–868.

Clemens, M., S. Radelet, and R. Bhavnani (2004, November). Counting chickens when they hatch: the short term effect of aid on growth. Working Paper 44, Center for Global Development.

Dalgaard, C.-J., H. Hansen, and F. Tarp (2004). On the empirics of foreign aid and growth. *Economic Journal 114*(496), p191 – 216.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature 48*, 424–455.

Doucouliagos, H. and M. Paldam (2009). The aid effectiveness literature: The sad results of 40 years of research. *Journal of Economic Surveys 23*(3), 433–461.

Easterly, W. (2006, March). *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*. New York: Penguin Press.

Easterly, W. (2007). Was development assistance a mistake? *American Economic Review 97*(2), 328 – 332.

Easterly, W., R. Levine, and D. Roodman (2004). Aid, policies, and growth: Comment. *The American Economic Review 94*(3), 774–780.

Frot, E. (2009). Early vs. late in aid partnerships and implications for tackling aid fragmentation. Working Paper No.1, SITE.

Hadi, A. S. (1992). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society 54*(2), 761 – 771.

Hansen, H. and F. Tarp (2001). Aid and growth regressions. *Journal of Development Economics 64*(2), 547 – 570.

Hayakawa, K. (2007). Small sample bias properties of the system GMM estimator in dynamic panel data models. *Economics Letters 95*(1), 32–38.

Kleibergen, F. and R. Paap (2006, May). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133*(1), 97–126.

Manski, C. F. and E. Tamer (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica 70*(2), 519–546.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417–1426.

Rajan, R. G. and A. Subramanian (2008, October). Aid and growth: What does the cross-country evidence really show? *Review of Economics and Statistics 90*(4), 643–665.

Rigobon, R. and T. Stoker (2009). Bias from censored regressors. *Journal of Business and Economic Statistics 27*(3), 340–353.

Roeder, P. G. (2001). Ethnolinguistic fractionalization (ELF) indices, 1961 and 1985. *Department of Political Science, University of California at San Diego*.

Roodman, D. (2007). The anarchy of numbers: aid, development, and cross-country empirics. *The World Bank Economic Review 21*(2), 255.

Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal 9*(1), 86–136.

Roodman, D. (2009b). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71*(1), 135–158.

Sachs, J. D., A. Warner, A. Åslund, and S. Fischer (1995). Economic Reform and the Process of Global Integration. *Brookings papers on economic activity 1*(1995), 1–95.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*(5), 331–386.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear iv regressions. In J. Stock and D. Andrews (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*. New York: Cambridge University Press.

Wacziarg, R. and K. H. Welch (2008). Trade Liberalization and Growth: New Evidence. *The World Bank Economic Review 22*(2), 187.

7 Appendix

7.1 Data appendix

Time periods. Observations for all variables except GDP, aid and trade are five-year arithmetic averages. Time period 1 represents years 1961-1965. The last period (period 10) is 2001-2005.

Log aid. Aid is Official Development Assistance (ODA) and comes from the Donor Assistance Committee (DAC) database of the OECD, Table 2a. Because predicted aid is built from predicted aid shares, net ODA, which is the usual aid variable in the aid effectiveness literature and which is potentially negative, cannot be used. Aid is defined as gross ODA, minus gross debt relief. The latter is excluded because it artificially inflates aid numbers in very recent years, where large debt cancellations were granted. Aid is not averaged, but summed up over the time period. It is expressed in millions of 2006 USD. Aid from all donors whose activity is reported by DAC and to all developing countries, according to DAC definition, is considered.

Log trade. Trade at the bilateral level is defined as the sum of imports and exports. At the recipient country level, it is summed across donor countries. Data in current USD millions from the International Trade dataset, version 2.01, of the Correlates of War Project. It is converted in 2006 USD by deflating it with the Consumer Price Index of the US Bureau of Labor Statistics.

Aid and trade as shares of GDP. Data in current USD is divided by GDP in current USD.

Log GDP. GDP in 2000 USD is from the World Development Indicators. GDP is not averaged, but measured every fifth year (1965, 1970, 1975, etc.). We use this instead of averaging to avoid introducing serial correlation.

GDP per capita. In 2000 USD. Source: World Development Indicators.

Growth. Growth is defined as the difference $ln(y_{\tilde{t}}) - ln(y_{\tilde{t}-5})$, where $y_{\tilde{t}}$ is GDP in year $\tilde{t}$. Note here that $\tilde{t}$ indexes year and not time periods.

Log population. Population is measured in millions. Source: World Development Indicators.

Inflation. Natural logarithm of 1+consumer price inflation rate. Source: World Development Indicators.

Money. Ratio of M2 to GDP. Source: World Development Indicators.

Schooling. Average years of primary schooling attained. Source: Barro and Lee (2010).

Institutional Quality. Variable between 0 and 16, defined as the sum of "Corruption", "Law and Order", and "Bureaucracy Quality", from the International Country Risk Guide (ICRG) of the PRS Group. Data is not available before 1984. For earlier years, data from the first available year is used. By doing so we follow the practice in the literature (see Roodman (2007)).

Openness. Index constructed by Sachs et al. (1995) and Wacziarg and Welch (2008).

Ethnic fractionalization. Ethnolinguistic Fractionalization index. Source: Roeder (2001).

Regional dummies. Dummies for East Asia and Pacific, and Sub-Saharan Africa. Region definitions are from the World Development Indicators.

Colony. Dummy variable equal to 1 if the pair has ever had a colonial link. Source: CEPII.

Distance. Distance in thousands of kilometers between the two main cities of the country. Source: CEPII.

Table A.1: Summary statistics and source of variables

Variable	Mean	s.d.	Unit
Growth	.18	.20	Percentage
Aid	1.86	3.28	Constant 2006 USD bn
Aid, GDP share	.087	.12	Percentage
GDP	27.1	76.4	Constant 2000 USD bn
GDP per capita	1711	2043	Constant 2000 USD
Population	20.5	74.9	Millions
Inflation	.15	.28	Annual change, perc. points
Openness	.22	.39	0-1 index
Money	32.5	27.5	M2 as perc. of GDP
Trade	41.9	113.8	Constant 2006 USD bn
Schooling	2.98	1.65	Year
Institutional quality	6.55	.2.61	1-16 continuous variable
East Asia	.12	.33	Identifier
Sub-Saharan Africa	.35	.48	Identifier
Ethno-linguistic frac.	.53	.27	Index (0 to 1)
Aid share	1.18	3.86	Percentage
Entry	8.84	9.07	year
Length	17.6	11.7	year
Colony	.038	.19	Index (0 to 1)
Distance	8.41	3.84	Thousands of km

7.2 Definition of growth

As indicated in Appendix 7.1, growth is defined over five-year periods. The aid effectiveness literature traditionally measures growth as the average yearly growth rate during the time period, i.e. as $\frac{1}{5}\sum_{i=0}^{4}\frac{y_{\tilde{t}+i+1}-y_{\tilde{t}+i}}{y_{\tilde{t}+i}}$. The two growth rates are highly correlated so we do not expect this change to affect the results.[32] On the other hand, we want to ensure that our results are not driven by this modification, and for greater comparability with the existing literature, we here replicate some of our results with growth defined as the five-year average of yearly rates.

Panel A uses aid volumes, panel B aid as a share of GDP. To compare results with the five-year growth rate and with the average yearly growth rate, one should, using a first order approximation, multiply these by five. Column (1) of Table A.2 is the within groups estimator with aid instrumented. The coefficient on aid is still significant, and its size multiplied by five is equivalent to the same coefficient in Table 3, column (4). Column (2) presents results with the system GMM estimator, and once more they correspond to what we found with the five-year growth rate. The next two columns exclude outliers. In A.3, which reports the same estimations but with aid as a share of GDP, the aid coefficient is significant only after outliers are excluded from the sample, similarly to the results in Section 5.4. Tables A.2 and A.3 confirm that our findings are in no way driven by our alternative definition of growth.

7.3 Additional robustness checks

7.3.1 Instrument exogeneity

In Table A.4, we include our trade variable in the GMM estimations. In column (1), the GMM difference estimator is used: like in Table 5, the coefficient on lagged GDP is too low for the estimator to be valid. In column (2) the system GMM estimator is used, and the coefficient on aid is smaller than with the within groups estimator, but still significant. The Hansen tests do not reject the required conditions. On the other hand, the relatively large number of instruments is likely to decrease the power of these tests. For this reason, we collapse the instrument matrix in column (3). The Hansen tests are still valid, but the aid coefficient is no longer significant.

Collapsing the instruments is useful for having more accurate Hansen tests, but since fewer moment conditions are used, the estimator becomes less efficient. Hence, we take the following approach: we keep in mind from the collapsing exercise that the

[32]The correlation is 0.99 in the data.

Table A.2: Growth as an average

	(1) FE-2SLS	(2) Sys. GMM	(3) FE-2SLS	(4) Sys. GMM
Log GDP, lagged	-0.047***	-0.013	-0.044***	-0.018*
	(0.010)	(0.011)	(0.013)	(0.0091)
Log aid, lagged	0.022***	0.011*	0.033***	0.011*
	(0.0078)	(0.0064)	(0.012)	(0.0060)
Log population	-0.054**	0.0077	-0.057**	0.014
	(0.021)	(0.0094)	(0.025)	(0.0098)
Inflation	-0.019***	-0.018*	-0.019***	-0.016**
	(0.0056)	(0.010)	(0.0064)	(0.0073)
Money, lagged	0.00058**	0.00030*	0.00049**	0.00032*
	(0.00023)	(0.00017)	(0.00025)	(0.00019)
Schooling	0.0058	0.0074	0.0092	0.012*
	(0.0075)	(0.013)	(0.0077)	(0.0069)
Institutional quality	0.00066	0.0035**	0.000031	0.0037**
	(0.0014)	(0.0016)	(0.0016)	(0.0017)
Openness	0.016***	0.023***	0.016**	0.024***
	(0.0058)	(0.0044)	(0.0068)	(0.0046)
Instruments		48		48
Countries	58	61	58	61
Hansen J test (p-val)		0.27		0.29
Hansen test (p-val), level		0.10		0.059
AR(1)		0.00086		0.000086
AR(2)		0.92		0.52
AP test (p-val)	0.00088		0.0017	
KP F stat	12.3		10.9	
R^2	0.27		0.081	
Observations	344	347	336	339

Note: KP: Kleibergen-Paap. AP: Angrist-Pischke. For the GMM estimations, instruments for the differences equation are log GDP lagged twice; instruments for the levels equation are log GDP lagged and differenced once. Log predicted aid is used as an instrument in all regressions. In columns (3) and (4), outliers are excluded using the Hadi procedure. All regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Growth as an average, aid as sh. of GDP

	(1) FE-2SLS	(2) Sys. GMM	(3) FE-2SLS	(4) Sys. GMM
Log GDP, lagged	-0.057***	0.0072	-0.030**	0.0052
	(0.011)	(0.0076)	(0.013)	(0.0068)
Aid, share of GDP	-0.23	0.17	0.46**	0.23**
	(0.16)	(0.11)	(0.21)	(0.11)
Log population	-0.027	-0.0020	-0.060***	0.0010
	(0.030)	(0.0068)	(0.016)	(0.0049)
Inflation	-0.023***	-0.022***	-0.022***	-0.020***
	(0.0048)	(0.0044)	(0.0059)	(0.0041)
Money, lagged	0.00048	0.00028**	0.00053**	0.00025*
	(0.00031)	(0.00012)	(0.00022)	(0.00013)
Schooling	-0.013	-0.0038	0.012	-0.00047
	(0.0091)	(0.0055)	(0.0093)	(0.0041)
Institutional quality	0.0012	0.0016	0.00041	0.0018
	(0.0012)	(0.0011)	(0.0011)	(0.0011)
Openness	0.015***	0.020***	0.012**	0.020***
	(0.0053)	(0.0044)	(0.0049)	(0.0049)
Instruments		48		48
Countries	58	61	57	61
Hansen J test (p-val)		0.62		0.82
Hansen test (p-val), level		0.40		0.48
AR(1)		0.00070		0.000086
AR(2)		0.84		0.58
AP test (p-val)	0.028		0.0011	
KP F stat	5.08		11.8	
R^2	0.29		0.24	
Observations	344	347	331	335

Note: KP: Kleibergen-Paap. AP: Angrist-Pischke. For the GMM estimations, instruments for the differences equation are log GDP lagged twice; instruments for the levels equation are log GDP lagged and differenced once. Log predicted aid is used as an instrument in all regressions. In columns (3) and (4), outliers are excluded using the Hadi procedure. All regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: GMM regressions, with trade flows

	(1) Diff. GMM	(2) Syst. GMM	(3) GMM Collapse	(4) 2SLS-Diff.	(5) 2SLS-Sys.
Log GDP, lagged	-0.50***	0.038	0.45	-1.18***	0.35
	(0.17)	(0.041)	(0.40)	(0.31)	(0.25)
Log aid, lagged	0.016	0.041*	0.019	0.041	0.026
	(0.040)	(0.024)	(0.10)	(0.048)	(0.052)
Log trade, lagged	-0.032	-0.096**	-0.41	0.11	-0.33
	(0.081)	(0.044)	(0.35)	(0.15)	(0.26)
Log population	-0.037	0.015	-0.12	0.44	-0.079
	(0.17)	(0.022)	(0.13)	(0.29)	(0.071)
Inflation	-0.078***	-0.095***	-0.13*	-0.079***	-0.21*
	(0.029)	(0.031)	(0.068)	(0.020)	(0.11)
Money, lagged	0.0029**	0.0025**	0.0052	0.0039***	0.0037
	(0.0012)	(0.0012)	(0.0047)	(0.0013)	(0.0035)
Schooling	-0.034	0.013	-0.11	0.077	-0.074
	(0.065)	(0.022)	(0.15)	(0.090)	(0.080)
Institutional quality	0.014*	0.014**	0.00094	0.017***	0.0030
	(0.0073)	(0.0060)	(0.017)	(0.0056)	(0.011)
Openness	0.078**	0.11***	0.095**	0.051*	0.084**
	(0.033)	(0.031)	(0.046)	(0.030)	(0.042)
Instruments	41	49	23		
Countries	58	61	61	58	61
Hansen J test (p-val)	0.25	0.63	0.44		
Hansen test (p-val), level		0.77	0.29		
AR(1)	0.11	0.00070	0.0031		
AR(2)	0.26	0.46	0.35		
KP LM test (p-val)				0.15	0.33
KP F stat				2.13	0.33
Observations	285	346	346	285	339

Note: KP: Kleibergen-Paap. Columns (1), (2), and (3) present GMM estimations, (4) and (5) are 2SLS regressions where variables are instrumented using GMM-type instruments, log predicted aid, and log predicted trade. Instruments for the differences equation are log GDP lagged twice. Instruments for the levels equation are log GDP lagged and differenced once. Log predicted aid and trade are used as instruments in all regressions. All regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Hansen tests do not reject the system GMM assumptions, but when it comes to point estimates, we consider those in column (2) as our preferred because they are more efficient. In columns (4) and (5), we test the strength of the GMM-type instruments, similarly to what is done in Table 6. We find that they are very weak, as shown by the extremely low Kleibergen and Paap F statistic. The null of underidentification cannot be rejected with a reasonable confidence level. As in Table 6, these results cast some serious doubt on the validity of the GMM approach in this setting.

7.3.2 Outliers

In Table A.5, we focus on our key regressions and run them without the Hadi-identified outliers in order to check their robustness. Columns (1) and (2) use the within groups estimator and find larger effects of aid than when all observations are used. The difference is sizable. The elasticity of GDP with respect to aid increases by 60 percent in both specifications, with and without controlling for trade. System GMM regressions in columns (4) and (5) yield similar results, but the aid coefficient becomes significant when controlling for trade. This is encouraging but we are reluctant to draw any firm conclusions from regressions based on very weak instrumentations. We take away from Table A.5 that outliers tended to bias our estimates downward, so that the GDP elasticity with respect to aid is possibly much larger.

7.3.3 Sample size

Table A.7 presents results from the most parsimonious specification with only lagged GDP, aid, and population as controls.

Column (1) of Table A.7 shows that the difference GMM estimator once more fails to produce correct estimates of the lagged GDP coefficient. Column (2) applies the system GMM estimator, and the aid coefficient is very close to in Table 5. This once more tends to confirm that our results are robust. Column (3) collapses the instrument matrix, and reveals that the system GMM assumptions are likely to be violated. As previously, the GMM estimates appear to be quite fragile. On the other hand, we find it encouraging that all our system GMM specifications find an elasticity close to 0.05.

Figure A.7 illustrates the change in the estimated aid coefficient of the within group estimator when outliers are excluded from the sample. Figure A.7 reveals that a few observations lie very far from the main group and so drive the result. When these are excluded, the coefficient becomes positive and significant, as shown in section 5.3 and Table 9.

Table A.5: Excluding outliers

	(1) FE-2SLS	(2) FE-2SLS	(3) System GMM	(4) System GMM
Log GDP, lagged	-0.21***	-0.33***	-0.099**	0.018
	(0.063)	(0.12)	(0.049)	(0.048)
Log aid, lagged	0.16***	0.14***	0.050	0.049**
	(0.058)	(0.045)	(0.031)	(0.020)
Log trade, lagged		0.10		-0.099*
		(0.082)		(0.052)
Log population	-0.25**	-0.21*	0.085	0.042
	(0.12)	(0.12)	(0.053)	(0.036)
Inflation	-0.096***	-0.10***	-0.079**	-0.094***
	(0.033)	(0.032)	(0.035)	(0.035)
Money, lagged	0.0025**	0.0018	0.0017	0.0025**
	(0.0012)	(0.0013)	(0.0010)	(0.0012)
Schooling	0.049	0.038	0.065*	0.035
	(0.038)	(0.038)	(0.033)	(0.028)
Institutional quality	0.00087	0.0030	0.020**	0.019**
	(0.0082)	(0.0073)	(0.0094)	(0.0076)
Openness	0.080**	0.069**	0.12***	0.12***
	(0.034)	(0.035)	(0.024)	(0.026)
Instruments			48	49
Countries	58	58	61	61
Hansen			0.29	0.37
Hansen level			0.029	0.19
AR(1)			0.000094	0.000090
AR(2)			0.72	0.97
KP F stat	10.9	11.2		
Observations	336	336	339	339

Note: KP: Kleibergen-Paap. Instruments for the differences equation are log GDP lagged twice. Instruments for the levels equation are log GDP lagged and differenced once. Log predicted aid and trade are used as instruments in all regressions. All regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: OLS and IV regressions, large sample

	(1) FE-2SLS	(2) FE-2SLS	(3) System GMM	(4) System GMM
Log GDP, lagged	-0.21***	-0.33***	-0.099**	0.018
	(0.063)	(0.12)	(0.049)	(0.048)
Log aid, lagged	0.16***	0.14***	0.050	0.049**
	(0.058)	(0.045)	(0.031)	(0.020)
Log trade, lagged		0.10		-0.099*
		(0.082)		(0.052)
Log population	-0.25**	-0.21*	0.085	0.042
	(0.12)	(0.12)	(0.053)	(0.036)
Inflation	-0.096***	-0.10***	-0.079**	-0.094***
	(0.033)	(0.032)	(0.035)	(0.035)
Money, lagged	0.0025**	0.0018	0.0017	0.0025**
	(0.0012)	(0.0013)	(0.0010)	(0.0012)
Schooling	0.049	0.038	0.065*	0.035
	(0.038)	(0.038)	(0.033)	(0.028)
Institutional quality	0.00087	0.0030	0.020**	0.019**
	(0.0082)	(0.0073)	(0.0094)	(0.0076)
Openness	0.080**	0.069**	0.12***	0.12***
	(0.034)	(0.035)	(0.024)	(0.026)
Instruments			48	49
Countries	58	58	61	61
Hansen			0.29	0.37
Hansen level			0.029	0.19
AR(1)			0.000094	0.000090
AR(2)			0.72	0.97
KP F stat	10.9	11.2		
Observations	336	336	339	339

Note: KP: Kleibergen-Paap. Instruments for the differences equation are log GDP lagged twice. Instruments for the levels equation are log GDP lagged and differenced once. Log predicted aid and trade are used as instruments in all regressions. All regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: GMM, large sample

	(1) Diff. GMM	(2) Sys. GMM	(3) GMM-Collapse
Log GDP, lagged	-0.40**	-0.0036	0.067
	(0.18)	(0.053)	(0.053)
Log aid, lagged	0.071*	0.042**	0.0058
	(0.040)	(0.021)	(0.035)
Log population	0.65*	-0.013	-0.055
	(0.37)	(0.051)	(0.060)
Instruments	44	53	19
Countries	106	108	108
Hansen J test (p-val)	0.50	0.24	0.066
Hansen test (p-val), level		0.034	0.0053
AR(1)	0.13	0.012	0.011
AR(2)	0.52	0.25	0.22
Observations	609	717	717

Note: The dependent variable is the growth rate. Instruments for the differences equation are GDP lagged twice in all the specifications. Instruments for the levels equation are GDP lagged and differenced once. Predicted aid is used as an instrument. The matrix of instruments is collapsed in column (3). All regressions include year effects. Robust standard errors clustered at the country level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

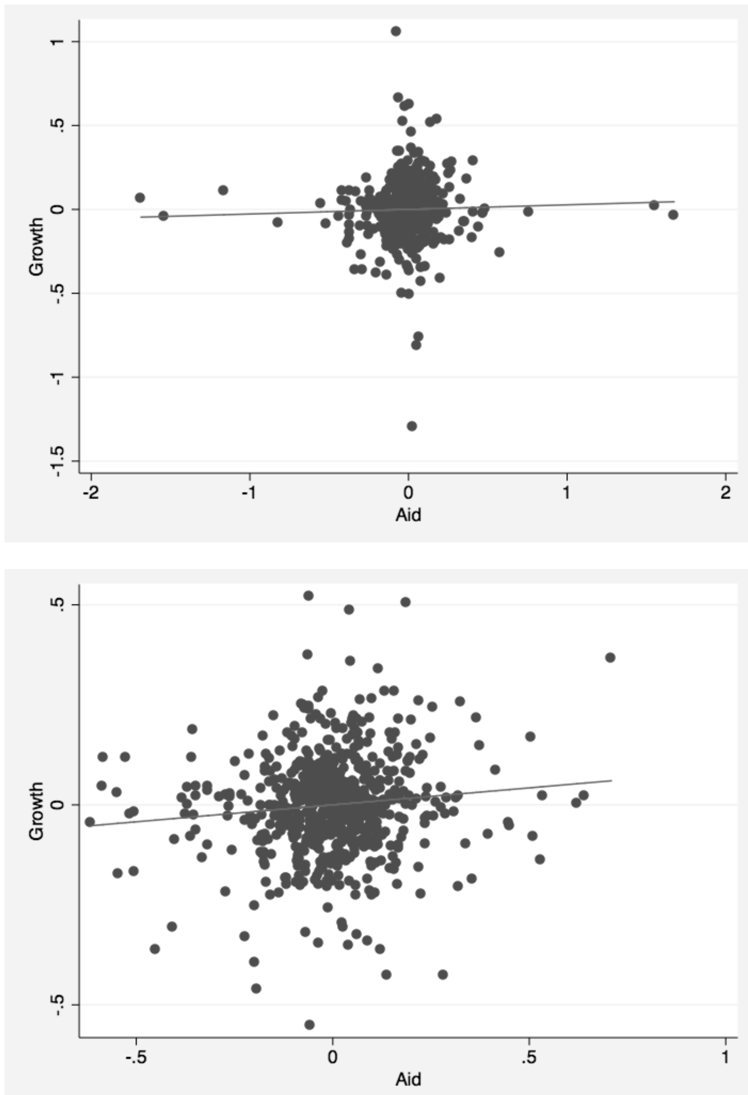

Figure A.7: Partial regression plot of growth on aid, including (top plot) and excluding (bottom plot) outliers, larger sample

Table A.8: List of countries

Sample with controls		**Large sample**	
Algeria	Malaysia	Angola	Palestinian Adm. Areas
Argentina	Mali	Barbados	Rwanda
Bangladesh	Mexico	Belize	Samoa
Bolivia	Morocco	Benin	Saudi Arabia
Botswana	Mozambique	Bhutan	Solomon Islands
Brazil	Niger	Burkina Faso	St. Lucia
Cameroon	Pakistan	Burundi	St.Vincent & Grenadines
Chile	Panama	Cambodia	Sudan
Colombia	Papua New Guinea	Cape Verde	Suriname
Congo, Dem. Rep.	Paraguay	Central African Rep.	Swaziland
Congo, Rep.	Peru	Chad	Timor-Leste
Costa Rica	Philippines	Comoros	Tonga
Cote d'Ivoire	Senegal	Djibouti	Vanuatu
Dominican Republic	Sierra Leone	Equatorial Guinea	Viet Nam
Ecuador	Sri Lanka	Ethiopia	
Egypt	Syria	Fiji	
El Salvador	Tanzania	Grenada	
Gabon	Thailand	Guinea	
Gambia	Togo	Guinea-Bissau	
Ghana	Trinidad and Tobago	Laos	
Guatemala	Tunisia	Lebanon	
Guyana	Turkey	Lesotho	
Haiti	Uganda	Libya	
Honduras	Uruguay	Madagascar	
India	Venezuela	Maldives	
Indonesia	Yemen	Mauritania	
Iran	Zambia	Mauritius	
Jamaica	Zimbabwe	Micronesia, Fed. States	
Jordan		Namibia	
Kazakhstan		Nepal	
Kenya		Nicaragua	
Liberia		Nigeria	
Malawi		Oman	

Note: The large sample corresponds to the regressions where the only controls are lagged log GDP, lagged log aid, and log population. Note that in addition to including more countries, the "large" sample also includes more observations for some countries than the sample with controls.