WORKING PAPER NUMBER 32
MAY 2007

Center for Global Development

The Anarchy of Numbers: Aid, Development, and Cross-country Empirics

By David Roodman

Abstract

Recent literature contains many stories of how foreign aid affects economic growth: aid raises growth in countries with good policies, or in countries with difficult economic environments, or mainly outside the tropics, or on average with diminishing returns. The diversity of these results suggests that many are fragile. I test 7 important aid-growth papers for robustness. The 14 tests are minimally arbitrary, deriving mainly from differences among the studies themselves. This approach investigates the importance of potentially arbitrary specification choices while minimizing arbitrariness in testing choices. All of the results appear fragile, especially to sample expansion.

The Anarchy of Numbers: Aid, Development, and Cross-country Empirics

David Roodman[1]
Research Associate, Center for Global Development

Center for Global Development
Working Paper #32

Final version: May 2007

Abstract: Recent literature contains many stories of how foreign aid affects economic growth: aid raises growth in countries with good policies, or in countries with difficult economic environments, or mainly outside the tropics, or on average with diminishing returns. The diversity of these results suggests that many are fragile. I test 7 important aid-growth papers for robustness. The 14 tests are minimally arbitrary, deriving mainly from differences among the studies themselves. This approach investigates the importance of potentially arbitrary specification choices while minimizing arbitrariness in testing choices. All of the results appear fragile, especially to sample expansion.

Keywords: Foreign assistance; economic growth; economic development; robustness testing

JEL codes: F350, O230, O400

[1] The author thanks William Easterly and Michael Clemens for advice, Craig Burnside, Lisa Chauvet, Jan Dehn, and Henrik Hansen for data and assistance, and Stephen Knack, Patrick Guillaumont, anonymous reviewers, and participants in an August 2003 seminar at the Center for Global Development for valuable comments. Responsibility for the final product rests with the author. Address for correspondence: David Roodman, Center for Global Development, e-mail: droodman@cgdev.org. This paper has appeared in the *World Bank Economic Review* 21(2), pp. 255–77, May 2007. The final version differs from the July 2004 one in two major respects. It no longer attempts to correct specification problems such as autocorrelation, focusing purely on the issue of fragility. And it now tests GMM regressions for robustness to outlier removal, by picking outliers from parallel 2SLS regressions. These changes in method explain the changes in results since the previous version.

In early 1981, economist Edward Leamer gave a speech at the University of Toronto in which he bemoaned the state of econometrics. Econometrics sought the status of a science, with regressions its analogue for the reproducible experiments of chemistry or physics. Yet an essential part of econometric "experimentation" was too often arbitrary, opaque, and unrepeatable. "The econometric art as it is practiced at the computer terminal involves fitting many, perhaps thousands, of statistical models….This search for a model is often well intentioned, but there can be no doubt that such a specification search invalidates the traditional theories of inference" (Leamer 1983). The way out of the quagmire, he argued, was for econometricians to explore larger regions of "specification space," systematically analyzing the relationship between assumptions and conclusions.

One econometric debate with hallmarks of the syndrome Leamer describes is that on the effectiveness of foreign aid in developing countries. Since Griffin and Enos (1970), econometricians have parried over the question of how aid affects economic growth in receiving countries. Prominent in the contemporary work, Burnside and Dollar (2000) conclude, "aid has a positive effect on growth in a good policy environment." Their evidence: the statistical significance in cross-country panel growth regressions of an interaction term of total aid received and an indicator of the quality of recipient economic policies (aid×policy). But Burnside and Dollar is just one voice among many. Collier and Dehn (2001), Collier and Dollar (2002, 2004), and Collier and Hoeffler (2004), corroborated their finding, while others challenged it. From the ongoing debate emerge several stories of the relationship between aid and growth, each of which turns on a particular quadratic or interaction term involving aid. The stories are not incompatible, but most papers support only one. Hansen and Tarp (2001) find that entering the square of aid drives out aid×policy, and makes the simple aid term significant too: aid works on average, but with dimin-

ishing returns. Guillaumont and Chauvet (2001) also fail to find significance for aid×policy, and instead offer evidence that aid works best in countries with difficult economic environments, characterized by volatile and declining terms of trade, low population, and natural disasters. In the same vein, Collier and Dehn (2001) find that *increasing* aid cushions countries against negative export price shocks. Collier and Hoeffler (2004) offer a triple-interaction term: aid works particularly well in countries that are recovering from civil war and that have good policies. Last, Hansen and Tarp, with Dalgaard, say that aid raises growth outside the tropics but not in them (Dalgaard, Hansen, and Tarp 2004).

These papers differ not only in their conclusions but their specifications too. Within the group, there are two different choices of period length in the panel data sets, three definitions of "policy," three of aid, and four choices of control variable set. Though probably none of the choices is made on a whim, these differences appear to be examples of what Leamer called "whimsy." From Leamer's point of view, the studies taken together represent a small sampling of specification space. And few include much robustness testing. Without further analysis, it is hard to know whether the results reveal solid underlying regularities in the data or are fragile artifacts of particular specification choices.

This paper examines the possibility of fragility systematically. Since by the laws of chance any regression can be broken with enough experimentation, it is essential for credibility that the testing suite itself be minimally arbitrary. The tests derive from two sources: the various choices already present in the original specifications; and the passage of time, which allows expansion of data sets (as in Easterly, Levine, and Roodman 2004). In all, regressions from 7 of the most prominent studies are subjected to this systematic test suite.

Section 1 reviews the approaches and conclusions of the studies that are tested for robustness. Section 2 describes the tests. Section 3 reports the results. Section 4 concludes.

I. History

The hope has often arisen that a turn to the numbers would shed light on the questions of whether and when foreign aid works. In the view of Hansen and Tarp (2000), this literature has gone through three generations. The first generation essentially spans 1970–72, and mainly investigates the aid-*savings* link. Influenced by the Harrod-Domar model, in which savings is the binding constraint on growth, aid-induced saving is assumed to lead directly to investment, thence to growth via a fixed incremental capital-output ratio. The second generation runs from the early 1970s to the early 1990s and directly investigates whether aid affects investment and growth.

Hansen and Tarp argue that the preponderance of the evidence from these first two generations shows that 1) aid increases total savings, but less than one-to-one; and 2) that aid increases investment and growth. They suggest that studies with more pessimistic results, such as Mosley, Hudson, and Horrell (1987), have gained disproportionate attention precisely because they are contrarian.

The third generation commences with Boone (1994) and continues to this day; it is the focus of this paper. The current generation has brought several innovations. The data sets cover more countries and years. Reflecting the influence of the new growth theory, regressors are typically included to represent the economic and institutional environment (sometimes together called the "policy environment"). The potential endogeneity of aid is addressed through instrumenting. And the marginal aid-growth slope is allowed to vary, through incorporation of such regressors as aid^2 and aid×policy. The data sets are almost always panels.

Burnside and Dollar (2000) test whether an interaction term of aid and an index of recipient economic policies is significantly associated with growth. Their panel is drawn from devel-

oping countries outside the former Eastern bloc, covering the six four-year periods in 1970–93. They incorporate some controls found significant in the general growth literature, namely: initial income (log real GDP/capita) to capture convergence; ethno-linguistic fractionalization (Easterly and Levine 1997), assassinations/capita, and the product thereof; the Knack-Keefer (1995) institutional quality variable, "ICRGE"; M2/GDP, to indicate financial depth, lagged one period to avoid endogeneity (King and Levine 1993); and dummies for sub-Saharan Africa and fast-growing East Asia.

Burnside and Dollar use a measure of aid called Effective Development Assistance (EDA, Chang, Fernandez-Arias, and Serven 1998). EDA differs in two major respects from the usual net Overseas Development Assistance measure (net ODA) tabulated by the Development Assistance Committee (DAC). First, EDA excludes technical assistance, on the grounds that it funds not so much recipient governments as consultants. Second, it differs in its treatment of loans. Net ODA counts disbursements of concessional (low-interest) loans only, but at full face value.[1] As a capital flow concept, it nets out principal but not interest payments on old loans. In contrast, EDA includes development loans, regardless how concessional (for example, near-commercial loans by the World Bank to middle-income countries such as Brazil), but counts only their grant element—that is, their net present value.

Concerned about limited statistical power, Burnside and Dollar combine their economic policy indicators into a single variable. They first run a growth regression without aid terms, but with all controls and three indicators of economic policy—log (1+inflation), budget balance/GDP, and the Sachs-Warner (1995) openness variable. All three policy variables differ

[1] DAC considers a loan concessional if it has a grant element of at least 25% of the loan value, using a 10% discount rate.

from 0 at the 0.05 level, so Burnside and Dollar form a linear combination of the three using their coefficients as weights.[2]

When Burnside and Dollar run their base specification, including aid and aid×policy the term of central interest, aid×policy, does *not* in fact enter significantly. However, they find that it becomes significant after either of two possible changes. Five outlier observations can be excluded (giving Burnside and Dollar's preferred specification). Or a quadratic interaction term can be added—aid^2×policy, in which case both aid×policy and aid^2×policy appear significantly different from 0, the first with positive sign, the second negative. Burnside and Dollar famously conclude that aid raises growth in a good policy environment, but with diminishing returns.

Burnside and Dollar's work has triggered responses, some critical, some supportive. Hansen and Tarp (2001) make one prominent attack. They modify the Burnside and Dollar two-stage least-squares (2SLS) regressions in several ways, most importantly by adding aid^2. Aid×policy is not significant in their results, but aid and aid^2 are, the first positive and the second negative. The implication is that aid is effective on average, but with diminishing returns—regardless of recipients' policies as far as the evidence goes. Hansen and Tarp then criticize both the Burnside and Dollar regressions and their own for failing to handle several standard concerns. There may be country-level fixed effects that correlate with both policies and growth. Failing to purge or control for all such effects could give spurious explanatory power to policies and aid×policy. Also, variables other than aid and its interaction terms, such as fiscal balance, could be endogenous and need instrumenting too. They deploy the Arellano-Bond (1991) Generalized Method of Moments (GMM) estimator, which is designed to handle these problems in short panels. Hansen

[2] They also add a constant term to the index, but this has no effect on the regression results of interest here.

and Tarp also add Δaid and Δ(aid^2) as regressors.[3] Their results on aid and aid^2 hold. And Δaid and Δ(aid^2) are significant too, again, the first with positive sign and the second negative.

Guillaumont and Chauvet (2001) tell a third story. They hypothesize that the economic vulnerability of a country influences aid effectiveness. They call economic vulnerability the "environment," not to be confused with Burnside and Dollar's "policy environment." In this story, aid flows stabilize countries that are particularly buffeted by terms of trade difficulties, other sorts of external shocks, or natural disasters. Guillaumont and Chauvet build an environment index out of four variables: volatility of agricultural value added (to proxy for natural disasters), volatility of export earnings, long-term terms of trade trend, and log of population (small countries being more vulnerable to external forces). Their specification is distinctive in using 12-year periods, and in its controls, which include population growth, mean years of secondary school education among adults, the Barro-Lee (2000) measure of political instability based on assassinations and revolutions, ethno-linguistic fractionalization, and lagged M2/GDP. In their OLS and 2SLS regressions, aid×environment appears with the predicted negative sign, indicating that aid works better in countries with worse environments. The term also drives out aid×policy.

Collier and Dollar (2002) corroborate Burnside and Dollar with a quite different data set and specification. Unlike Burnside and Dollar, they perform OLS only. They include former Eastern bloc countries, the Bahamas, and Singapore. They use net ODA rather than EDA. They study 1974–97 instead of 1970–93. They drop all Burnside and Dollar controls except log initial GDP/capita, ICRGE, and period dummies. But they add region dummies.[4] And they define policy as the overall score from the World Bank's Country Policy and Institutional Assessment

[3] This is equivalent to adding lagged aid and lagged aid^2 since the regressions also control for aid and aid^2.
[4] The regions are Europe and Central Asia, Middle East and North Africa, Southern Asia, East Asia and Pacific, Sub-Saharan Africa, and Latin America and the Caribbean, as defined by the World Bank.

(CPIA), which is a composite rating of countries on 20 aspects of policies and institutions.[5] Thye add aid^2 but then drop the linear aid term from their preferred specification as insignificant. After all the changes, aid×policy is again significant, as is aid^2, with a negative sign.

Stargin from the Collier and Dollar core regression, Collier and Hoeffler (2004) analyze how recent emergence from civil war influences aid effectiveness. Sticking to the four-year panel, they create three dummies to indicate how recently civil war ended. "Peace onset" is 1 in the period when a country goes from civil war to peace. "Post-conflict 1" is 1 the following period, and "post-conflict 2" the period after that—assuming civil war does not recur. Aid×policy×post-conflict 1 is significant in Collier and Hoeffler's preferred (OLS) specification: aid works particularly well in a good policy environment a few years after civil conflict.

Also corroborating Burnside and Dollar, Collier and Dehn (2001) hew closely to the Burnside and Dollar specification and data set, and tell a story that incorporates elements from Guillaumont and Chauvet. They find that adding variables incorporating information on export shocks renders Burnside and Dollar's preferred specification—the one with aid×policy but not aid^2×policy—more robust to the inclusion of Burnside and Dollar's five outliers. First, they add two variables indicating the magnitude of any positive or negative commodity export price shocks. They report that aid×policy is then significant at 0.01 for a regression on the full data set. The negative-shock variable is significant too, with the expected minus sign.[6] Then Collier and Dehn add four aid-shock interaction terms: lagged aid×positive shock, lagged aid×negative shock, Δaid×positive shock, and Δaid×negative shock. The first and last prove positive and sig-

[5] Collier and Hoeffler (2002) make a small correction to the Collier and Dollar data set, excluding 5 observations where a missing value had been treated as 0. The Collier and Hoeffler version of the Collier and Dollar regression is tested here.

[6] However, the reproduction using their data gives a *t* statistic of only 0.42 to aid×policy despite having the same R^2 and sample size, so their result may be an error. But the same, negative sign does appear on the negative-shock variable.

nificant in OLS, and the last, Δaid×negative shock proves particularly robust in their testing. The study buttresses Burnside and Dollar while suggesting that well-timed aid increases ameliorate negative export shocks. This matches the Guillaumont and Chauvet result in spirit. But where Guillaumont and Chauvet interact the *amount of aid* with the *stand*ard *deviation* of an index of export *volume* and other variables, Collier and Dehn's significant term involves the *change* in aid and the *change* in export *prices*.

Dalgaard, Hansen, and Tarp (2004) tell a novel aid-growth story. They focus on the share of a country's area that is in the tropics, as a determinant of both growth and the influence of aid on growth. This variable surfaces as a growth determinant in Bloom and Sachs (1998), Gallup and Sachs (1999), and Sachs (2001, 2003). The causal links may include institutions and economic policies (Acemoglu, Johnson, and Robinson 2001; Easterly and Levine 2003). Dalgaard, Hansen, and Tarp thus see tropical area as an exogenous "deep determinant" of growth. In the regressions, aid and aid×tropical area fraction are quite significant, the first with positive sign, the second with negative sign and similar magnitude. For countries situated completely in the tropics, the derivative of growth with respect to aid (the sum of the two coefficients) is indistinguishable from 0. Thus, on average, aid seems to work outside the tropics but not in them. The authors report that their new interaction term drives out both aid×policy and aid^2.

There are other third-generation studies (Hadjimichael *et al.* 1995; Durbarry, Gemmell, and Greenaway 1998; Lensink and White 2001; Svensson 1999; Chauvet and Guillaumont 2002; Burnside and Dollar 2004). This paper focuses on those already highlighted as being among the most influential and, with one exception, having been published. The exception is Collier and Dehn (2001) which is a pillar of the published Collier and Dollar (2004).

The testing here applies to what appear to be authors' preferred regressions. (See Table 1.) Country by country, the tested regressions generate a diversity of conclusions about the slope of growth with respect to aid at the margin. (See Table 2, illustrating for the 20 largest aid recipients in 1998.) As an example of the calculations here, the Burnside and Dollar structural equation is

$$\Delta Y = \alpha A + \beta A \times P + \gamma P + \mathbf{x}\delta + \varepsilon,$$

where Y is GDP/capita, A is aid, P is policy, $\mathbf{x}$ is a vector of controls, including initial GDP/capita, and ε is the error term. So the implied slope of growth respect to aid is

$$d(\Delta Y)/dA = \alpha + \beta P,$$

which depends on the recipient's policy level. Applying such formulas to 1998 data, the Burnside and Dollar regression generally predicts benefits from increasing aid while, at the other extreme, the Dalgaard, Hansen, and Tarp and Guillaumont and Chauvet regressions express pessimism. The question is what to make of such conclusions.

II. The Test Suite

There is some robustness testing in the recent literature on aid-growth connections, albeit focusing on Burnside and Dollar. Lu and Ram (2001) introduce fixed effects into the Burnside and Dollar regressions. Ram (2004) splits the aid variable into the components coming from bilateral and multilateral donors, and also tests alternative definitions of policy. Dalgaard and Hansen (2001) modify the choice of excluded outliers. Easterly, Levine, and Roodman (2004) extend the Burnside and Dollar data set to additional countries and an additional period, 1994–97. All these tests eliminate the key Burnside and Dollar result. The present study expands Easterly, Levine, and Roodman along two dimensions. It applies more tests. And it tests more studies.

In addition to fragility, the other bugaboo of econometrics is misspecification. Important questions can be raised about the validity of the regressions tested here. Some exhibit serial correlation in the errors.[7] The excludability and relevance of instruments are a legitimate concern. Regressors treated as exogenous may not be. And term pairs such as aid and aid^2 may be multicollinear. But for the sake of concision, this paper focuses on the problem of fragility.

The Tests

The tests applied to these third-generation aid-growth regressions constitute a more systematic sampling of "specification space" than has hitherto been made. To limit complexity and minimize arbitrariness, each test ideally involves changing just one aspect of the estimations at a time. The tests are summarized in Table 3. The first four groups of tests, relating to the controls, the definition of aid and policy, and period length, transfer one specification's choices to the others.[8] Last are tests that modify the sample by dropping outliers and/or expanding to new countries and periods.

The tests are:

1. *Changing the control set.* In his worries over whimsy, the specification choice that conerns Leamer is that of regressors. The studies examined use four different control sets, which give rise to four robustness tests, detailed in Table 3. Each substitutes an alternative control set for the original one and examines the effect on the significance of key terms.

   Like the authors of the original regressions, and in the spirit of avoiding arbitrariness, the robustness tests here use all complete observations available for developing countries (in-

[7] An earlier version of this paper attempted to address autocorrelation by further modifying the tested specifications, at the expense of complexity in presentation, and, arguably, "whimsy." In particular, most of the test regressions included the log of population as a control since Sargan-type tests suggested it was an improperly excluded instrument. This explains the difference in results between this and the earlier version.

[8] Those papers that instrument the variables of interest also differ in their choice of instruments. But since different variables (aid×policy in one regression, say, versus aid^2 in another), ought to be instrumented differently, the various instrument sets are less interchangeable and of less use for the approach in this paper.

cluding the countries of Eastern Europe). Because different variables are available for different subsets of countries, changing the regressor set changes the regression sample. One could perform variants of the tests that are restricted to the intersections of the old and new samples in an attempt to distinguish the effects of changing sample and changing variables. This course is not taken here because it would add to the complexity, would still cause sample changes, and would not answer the hypothetical, "What would the results have been if the original authors had used alternative controls?" The authors almost certainly would have used all available observations.

2. *Redefining aid.* All the studies take total aid received as a share of recipient GDP. But there are differences in defining both the numerator and denominator of the ratio. Burnside and Dollar, Collier and Dehn, and Dalgaard, Hansen, and Tarp use Effective Development Assistance in the numerator while the rest use net ODA. On the choice of denominator, there is also a split. Hansen and Tarp and Guillaumont and Chauvet use GDP converted to dollars using market exchange rates. The others use real GDP from the Penn World Tables. A country's relative price level strongly correlates with income per head, with the poorest countries having price levels 20–25% of that of the United States. Thus using purchasing power parities instead of exchange rates will cause the GDPs of the poorest countries to be measured as relatively larger and aid to them as relatively smaller as a share of GDP. This might have a significant effect on coefficient estimates for aid and its interactions

   With two options each for measuring aid and GDP, there are four possible combinations for aid/GDP. The literature includes all but EDA/exchange rate GDP, and these are the bases for three tests.[9] In fact, EDA/real GDP and ODA/real GDP are highly correlated (Dalgaard

[9] The published EDA data (Chang, Fernandez-Arias, and Serven 1998) cover only 1975–95. EDA as used here is extrapolated to the rest of 1970–2001 via a regression of EDA on net ODA.

and Hansen 2001), so switching from one to the other may not stress results much. (See Table 4.)

3. *Redefining good policy.* Three sets of "good policy" variables appear among the tested regressions: 1) Burnside and Dollar's combination of budget balance, inflation, and Sachs-Warner openness; 2) inflation and Sachs-Warner only (Hansen and Tarp GMM); and 3) CPIA alone (Collier and Dollar, Collier and Hoeffler). These generate three robustness tests. Using Burnside and Dollar's coefficients to form policy indexes (6.85 for budget balance, –1.40 for inflation, and 2.16 for Sachs-Warner), the first two policy definitions are correlated 0.98, but the third varies more distinctly. (See Table 4.) But in actually application of the tests, the Burnside-Dollar–style index-forming regression is rerun each time; it includes all regressors except aid and its interaction terms, and the coefficients on the policy variables are used to make the index, regardless of statistical significance.[10]
4. *Changing periodization.* All but Guillaumont and Chauvet use four-year periods. The lack of higher-frequency observations of the Guillaumont and Chauvet environment variable prevents adapting their 12-year regressions to a 4-year-period panel. But the other regressions can be tested on 12-year panels. Notably, key cross-section studies in the growth literature use periods of 10–25 years despite the small samples that result (Barro 1991; Mankiw, Romer, and Weil 1992; Sachs and Warner 1995).
5. *Removing outliers.* The tested Burnside and Dollar specification excludes five observations that are a) outliers in aid×policy and b) highly influential on the coefficient on that term. This raises a general question about the importance of outliers. To investigate, one robustness test reruns the reproductions of the original regressions after excluding outliers. Another does the

[10] The constant term in the policy index is computed in the same manner as in Burnside and Dollar. It is the predicted growth rate in the model when the policy variables and the period dummies are zero, and all other variables

same for the expanded-sample versions. (See below.) Following Easterly, Levine, and Roodman (2004), outliers are chosen by applying the Hadi (1992) procedure for identifying multiple outliers to the partial scatter of growth and a regressor of interest, using 0.05 as the cut-off significance level.[11] In 2SLS estimatations, regressors are first projected onto instruments.[12]

Outliers are not synonymous with influential observations. But even outliers that do not greatly influence coefficients of interest can substantially affect reported standard errors. In addition, outliers are the observations most likely to signal measurement problems or structural breaks beyond which the core model does not hold—both of which seem better reasons for exclusion than high influence. That said, outliers do not necessarily signal measurement problems or structural breaks. This is especially possible when the variable of interest is highly non-normal, such as the Collier and Dehn export price shock variable. In such cases, outliers may contain valuable information about the development process under rare circumstances. Since the two-dimensional partial scatter plot is not well-defined for GMM regressions, in those cases, analogous 2SLS regressions are run for the purposes of identifying outliers.

6. *Expanding the sample*. Easterly, Levine, and Roodman (2004) develop a dataset that extends that of Burnside and Dollar from 1970–93 to 1970–97 and adds six countries. For the present study, that data set has been extended to 2001 and improved in other respects. (See Appendix 1.) This allows a net expansion in both years and countries for all but Guillaumont and

---

take their sample-average values.

[11] Applying the Hadi procedure directly to a full, many-dimensioned data set typically identified 20% or more of observations as outliers.

[12] This test is even run on the Burnside and Dollar 5/OLS regressions, from which one set of outliers is already excluded. Regardless of the genesis of these regressions' results, it is interesting whether they are driven by a few observations in the remaining sample.

Chauvet regression, whose 12-year periods and unusual environment variable hinder expansion.

Issues in Interpreting Results

If Leamer's (1983) extreme bounds analysis is applied to the results of this testing, then a coefficient will be deemed robustly different from 0 only if it is significantly different from 0 in *every* test. However, as Sala-I-Martin (1997) argues, this definition of robustness seems extreme. For example, one could test robustness by averaging together all observations for each global region, generating samples of some 6 observations. Almost no regression would pass this test. One could argue that this test would be "unfair," i.e., too weak to generate meaningful results. But there is no sharp division between fair and unfair tests. Indeed, in this test suite, the 12-year-period test destroys every regression it can be applied to. It is not obvious whether the test is too strong or the regressions too weak. Thus robustness should be a continuous rather than dichotomous concept.

Sala-I-Martin offers his own procedure for assessing robustness. In essence, he estimates the cumulative distribution function for a coefficient of interest by running a large number of variants of the regression it comes from. The *robustness* of a coefficient is then the fraction of the density that is on one or the other side of zero. The validity of this concept is based on the assumption, however informal, that the set of regressions actually run is a *representative of all possible variants of the original regression*. For the collection of tests assembled here, that assumption is not valid. For example, one important subset of tests, those expanding the sample, cannot be applied to the Guillaumont and Chauvet regression. It does not seem plausible that the test results are representative both with and without this important subset of tests.

The sampling of specification space that is made here is *minimally arbitrary*, but cannot be assumed to be *representative* of all possible tests. Thus while Leamer's definition of robustness may be too harsh for this context, Sala-I-Martin's has its own limitations. This will be true even if one performs every possible combination of tests in the suite rather than just one at a time. In the end, it seems that human judgment applied to the full set of results must substitute for mechanical definitions of robustness. This in turn means there is some value in keeping the tests few enough for the human mind to embrace.

III. Results

The first step in the testing is to use the authors' data sets to reproduce their original results (see columns 1 of Table 5 and Table 6). All the reproductions exhibit the same pattern of results as the originals and all but one have the same sample size.[13] The Burnside and Dollar, Collier and Dehn, and Hansen and Tarp reproductions are perfect, and the rest are close. Since the purpose of the paper is to test robustness, the inexact matches are not a concern. If the results from the tested regressions are robust, they should withstand whatever minor changes in data or specification cause the discrepancies in the reproductions.

Table 5 and Table 6 report results on key terms in all tested regressions.[14] Blank cells indicate inapplicable tests. The test involving the definition of aid as EDA/real GDP, for example, is not applicable to regressions that originally use it. Using 12-year periods does not work for the Collier and Hoeffler regression, because the definition of their post-conflict 1 variable assumes

[13] The Dalgaard, Hansen, and Tarp regression was executed with the DPD for Ox package (Doornik, Arellano, and Bond 2002). It turns out that an undocumented limitation in this software—incomplete observations that create gaps in the time series must always be included in the data file rather than deleted—led to a slight mishandling of the data. The xtabond2 module for Stata (Roodman 2006), used here, does not have this limitation. This explains the difference in samples.

[14] Full results are available upon request.

4-year periods. Lack of higher-frequency data for Guillaumont and Chauvet's environment variable prevents short-period tests. A total of 77 robustness checks are run.[15]

Results for tests inspired by differences among the original regressions are in Table 5. The Collier and Hoeffler result on post-conflict 1×aid×policy (or the collinear post-conflict 1×aid) and the Dalgaard, Hansen, and Tarp results for aid and aid×tropical area fraction do best. Interestingly, all of these center on sharply bimodal variables: The Collier and Hoeffler post-conflict 1 dummy is 1 for only 13 of the 344 observations in their original sample, and there are negative shocks in 38 of the 234 Collier and Dehn observations. In the Dalgaard, Hansen, and Tarp sample, 233 of the 371 observations are 100% tropical and 68 are 0%, leaving 70 in between. Evidently regularities involving such variables are more resilient to specification changes.

Results from sample-modifying tests appear in Table 6. The first two result columns are based on regressions on the original authors' datasets—first for their full sample, second for the sample excluding outliers. The next pair of columns is analogous, for the expanded data set. The figures in Appendix 2 illustrate the sample-modifying results, and are reminders of the importance of checking for outliers. Except for Guillaumont and Chauvet, all the original OLS and 2SLS results depend on outliers for some or all of their significance. The dependence is particularly heavy for the regressions involving aid×policy. On the other hand, the *lack* of significance of most of the coefficients under the sample-expansion test is not driven by outliers. It is worth noting that the Collier and Dehn result on Δaid×negative shock, another interaction term involving a variable with a highly non-normal distribution, is arguably stronger than it looks. The coef-

[15] Initial testing revealed multicollinearity in the Collier and Hoeffler regression. In their preliminary regression 3.1 (not regression 3.4, which is tested here), they include the variables post-conflict 1, post-conflict 1×policy, and post-conflict 1×aid^2, along with the favored post-conflict 1×aid ×policy. In the reproduction of 3.1, post-conflict 1×aid×policy and post-conflict 1×aid have a partial correlation of 0.985, making the two statistically indistinguishable. Thus the Collier and Hoeffler results ought to be interpreted as pertaining to either post-conflict 1×aid ×policy or post-conflict 1×aid. Occam's razor argues for the latter.

ficient is reversed by the exclusion of outliers from the original sample. But it is arguably fallacious to draw conclusions about the role of shocks having excluded many of the most dramatic examples.

The overall pattern is clear-cut. The 12-year test is the toughest, probably because of the small samples, failing all regressions. The new-data test is not far behind, an important point given that the surgery it involves—a moderate sample expansion—is much less radical. Reading the tables by rows (test subjects) instead of columns (tests), we see that the Dalgaard, Hansen, and Tarp result on the aid-tropics link is the only one to come through the specificiation-modifying tests strongly. But it too falls down on the sample-modifying tests after outliers are removed. Four of the nine outliers are for the Jordan, covering 1974–89, a period in which that non-tropical country experienced high growth and received much aid from its neighbors. This confirms the conclusion of Rajan and Subramanian (2005) that the aid×tropics result is fragile too.

IV. Conclusion

Each of the papers examined here embodies a set of choices about model specification and data. Aid is measured a certain way. A certain epoch is studied. Periods have a certain length. And so on. Some of these choices imply assumptions about the world, such as, say, that aid is exogenous to growth. All limit the scope of a strict interpretation of the results. A question of great importance for the literature is, how many of such implied assumptions can be dropped without harming the conclusions?

The results reported here suggest that the fragility found in Easterly, Levine, and Roodman (2004) for Burnside and Dollar is the norm in the cross-country aid effectiveness literature. Indeed, in a counterpoint to the focus of Leamer (1983), Levine and Renelt (1992), and Sala-I-

Martin (1997) on the choice of controls as a source of fragility, it turns out that modifying the sample generally affects results most. For example, in the Collier and Dollar regression, half of the specification-modifying tests leave the *t* statistic at 1.49 or higher and two more lower it to near 1.00. (See Table 5.) But adding more years sends it to –0.19—and, after dropping outliers, to –0.81. (See Table 6.)

Does this mean that the various stories of aid effectiveness should be summarily dismissed? Are recipient policies, exogenous economic factors, and post-conflict status irrelevant to aid effectiveness? Are there no diminishing returns to aid? Is helping the neediest countries a hopeless task? No. There can be no doubt that some aid finances investment, and that domestic policies, governance, external conditions, and other factors these authors study influence the productivity of investment. Why then do such stories of aid effectiveness not shine through more clearly? The reasons are several. Aid is probably not a fundamentally decisive factor for development, not as important as, say, domestic savings, inequality, or governance. Moreover, foreign assistance is not homogeneous. It consists of everything from food aid for famine-struck countries to technical advice on building judiciaries to loans for paving roads. And much aid is poorly used—or, like venture capital, is good bets gone bad. Thus the statistical noise tends to drown out the signal.

Perhaps researchers will yet unearth more robust answers to the fundamental questions of aid policy. Or perhaps they have hit the limits of cross-country empirics. Either way, robust, valid generalizations have not and will not come easily. Despite decades of trying, cross-country growth empirics have yet to teach us much about whether and when aid works.

Table 1. Regressions Tested

Regression	Estimator	Former East bloc?	Controls	Study period	Years/ period	Definition of Aid	Definition of Policy	Outliers out?	Key significant term(s)
Burnside & Dollar 5/OLS	OLS	No	LGDP, ETHNF, ASSAS, ETHNF×ASSAS, ICRGE, M2, SSA, EASIA, period dummies	1970–93	4	EDA/real GDP	BB, INFL, SACW	Yes	aid×policy
Collier & Dehn 3.4	" "	" "	" "	1974–93	" "	" "	" "	No	aid×policy, Δaid×negative shock
Collier & Dollar 1.2[1]	OLS	Yes	LGDP, ICRGE, policy, period and region dummies	1974–97	" "	ODA/real GDP	CPIA	" "	aid×policy, aid^2
Collier & Hoeffler 3.4	OLS	" "	" "	" "	" "	" "	" "	" "	aid×policy× post-conflict 1
Hansen & Tarp 3.2	Difference GMM	No	LGDP, ASSAS, ETHNF×ASSAS, ICRGE, M2, period dummies	1978–93	" "	ODA/ exchange rate GDP	INFL, SACW	" "	aid, aid^2, Δaid, $Δaid^2$
Dalgaard, Hansen, and Tarp 3.5	System GMM	Yes	LGDP, policy, period dummies	1970–97	" "	EDA/real GDP[2]	BB, INFL, SACW	" "	aid, aid×tropical area fraction
Guillaumont & Chauvet 5.2	2SLS	No	LGDP, ENV, SYR, POPG, M2, PINSTAB, ETHNF, period dummy	1970–93	12	ODA/ exchange rate GDP	" "	" "	aid, aid×environment

[1]As revised in Collier & Hoeffler 1.1. [2]As extrapolated to 1970–74 and 1996–97 in Easterly, Levine, and Roodman (2004).

Abbreviations: LGDP=log initial real GDP/capita; ETHNF=ethno-linguistic fractionalization, 1960; ASSAS=assassinations/capita; ICRGE=composite of International Country Risk Guide governance indicators; M2=M2/GDP, lagged; SSA=Sub-Saharan Africa dummy; EASIA=fast-growing East Asia dummy; ENV=Guillaumont & Chauvet "environment" variable; SYR=mean years of secondary schooling among adults; PINSTAB=average of ASSAS and revolutions/year; BB=budget balance/GDP; INFL=log(1+inflation); SACW=Sachs-Warner openness; EDA=Effective Development Assistance; ODA=Net Overseas Development Assistance.

Source: Author's analysis based on sources described in the text.

Table 2. Marginal Impact of Aid According to Preferred Regression, Various Studies, for 20 Largest Aid Recipients of 1998

Country	B&D 5-OLS	Collier & Dehn	Collier & Dollar	Collier & Hoeffler	Hansen & Tarp GMM	Dalgaard *et al.*	Guillaumont & Chauvet
Bangladesh	0.20	–0.04	0.43	0.37	0.14	0.29	–0.72
	(0.15)	(0.15)	(0.20)	(0.20)	(0.13)	(0.18)	(0.37)
Bolivia	0.56	0.15	0.29	0.24	–0.02	–0.29	–0.33
	(0.22)	(0.22)	(0.16)	(0.15)	(0.08)	(0.33)	(0.17)
China	0.16	–0.06	0.53	0.46	0.20	0.66	
	(0.15)	(0.14)	(0.25)	(0.24)	(0.15)	(0.14)	
Cote d'Ivoire	0.56	0.15	0.11	0.07	–0.05	–0.29	–0.23
	(0.22)	(0.22)	(0.10)	(0.09)	(0.08)	(0.33)	(0.13)
Egypt	0.59	0.16	0.30	0.25	0.13	0.53	–0.55
	(0.23)	(0.22)	(0.15)	(0.14)	(0.13)	(0.14)	(0.28)
Ethiopia	0.30	0.01	0.20	0.71	–0.08	–0.29	0.04
	(0.16)	(0.16)	(0.12)	(0.15)	(0.08)	(0.33)	(0.12)
Haiti	0.15	–0.07	–0.23	–0.25	–0.28	–0.29	
	(0.15)	(0.14)	(0.13)	(0.12)	(0.10)	(0.33)	
India	0.13	–0.08	0.48	0.41	0.20	0.19	–0.97
	(0.15)	(0.14)	(0.22)	(0.22)	(0.15)	(0.20)	(0.51)
Indonesia	0.61	0.17	0.51	0.44	0.19	–0.29	–0.68
	(0.23)	(0.23)	(0.24)	(0.24)	(0.15)	(0.33)	(0.35)
Kenya	0.46	0.09	0.33	0.28	0.06	–0.29	–0.48
	(0.19)	(0.19)	(0.16)	(0.16)	(0.10)	(0.33)	(0.25)
Mozambique			0.06	0.53	–1.04	–0.20	–0.26
			(0.08)	(0.12)	(0.39)	(0.30)	(0.14)
Nicaragua	0.57	0.64	–0.09	0.51	–0.65	–0.29	–0.18
	(0.22)	(0.45)	(0.12)	(0.18)	(0.24)	(0.33)	(0.11)
Philippines	0.61	0.17	0.43	0.37	0.18	–0.29	–0.86
	(0.23)	(0.23)	(0.20)	(0.20)	(0.14)	(0.33)	(0.45)
Poland	0.54	0.13	0.42	0.36	0.17	0.69	
	(0.21)	(0.21)	(0.20)	(0.20)	(0.14)	(0.14)	
Russia	–0.03	–0.17	0.36	0.32	0.20	0.69	
	(0.17)	(0.13)	(0.17)	(0.17)	(0.15)	(0.14)	
Tanzania	0.44	0.99	0.09	0.05	–0.21	–0.29	
	(0.19)	(0.78)	(0.08)	(0.08)	(0.09)	(0.33)	
Thailand	0.62	0.18	0.57	0.49	0.20	–0.29	–0.80
	(0.23)	(0.23)	(0.27)	(0.26)	(0.15)	(0.33)	(0.42)
Uganda	0.50	1.56	0.16	0.11	–0.09	–0.29	–0.13
	(0.20)	(1.24)	(0.11)	(0.11)	(0.08)	(0.33)	(0.10)
Vietnam	0.19	–0.05	0.45	0.39	0.12	–0.29	0.00
	(0.15)	(0.14)	(0.21)	(0.21)	(0.12)	(0.33)	(0.00)
Zambia	0.08	–0.11	–0.38	–0.40	–0.43	–0.29	–0.32
	(0.15)	(0.13)	(0.21)	(0.19)	(0.15)	(0.33)	(0.17)

Robust standard errors in parentheses. All figures are based on reproductions of the original regressions. All pertain to 1994-97, except those the the Guillaument & Chauvet regression, which pertain to 1982-93. Aid is taken as a share of exchange rate GDP in the Hansen & Tarp and Guilaumont & Chauvet regressionis and of PPP GDP in the rest. Blank cells are causing by missing observations of underlying indicators.
Source: Authors' analysis based on sources described in the text.

Table 3. Robustness Tests

Test	Description
Changing controls	
BD controls	Control for LGDP, ETHNF, ASSAS, ETHNF×ASSAS, ICRGE, M2, SSA, EASIA, period effects, as in Burnside & Dollar, Collier & Dehn, Hansen & Tarp
CD controls	Control for LGDP, ICRGE, period and region effects, as in Collier & Dollar, Collier & Hoeffler
GC controls	Control for LGDP, ENV, SYR, POPG, M2, PINSTAB, ETHNF, period effects, as in Guillaumont & Chauvet
DHT controls	Control for LGDP, ICRGE, SSA, EASIA, period effects, as in Dalgaard, Hansen, and Tarp
Changing aid definition	
EDA/real GDP	Effective Development Assistance/real GDP, as in Burnside & Dollar, Collier & Dehn, Dalgaard, Hansen, & Tarp
ODA/real GDP	Net Overseas Development Assistance/real GDP, as in Collier & Dollar, Collier & Hoeffler
ODA/exchange rate GDP	Net Overseas Development Assistance/exchange rate GDP, as in Hansen & Tarp, Guillaumont & Chauvet
Changing policy definition	
INFL, BB, SACW	Inflation, budget balance, and Sachs-Warner openness, as in Burnside & Dollar, Collier & Dehn
INFL, SACW	Inflation and Sachs-Warner, as in Hansen & Tarp
CPIA	Country Policy and Institutional Assessment, as in Collier & Dollar, Collier & Hoeffler
Changing period length	
12-year	Aggregate over 12-year periods, as in Guillaumont & Chauvet
Changing sample and data set	
No outliers	Remove Hadi outliers in the partial scatter of the dependent variable and the independent variable of greatest interest
Expanded sample	New data set. Carried to 2001, except shocks data end in 1997 and Guillaumont & Chauvet environment variable not updated
Expanded sample, no outliers	Combine above two changes

Abbreviations as in Table 2.

Table 4. Simple Correlations of Aid and Good Policy Measures, Respectively, Four-Year Periods, on Available Observations

	EDA/real GDP	ODA/real GDP	ODA/exchange rate GDP
EDA/real GDP	1.00		
ODA/real GDP	0.97	1.00	
ODA/exchange rate GDP	0.78	0.82	1.00

	Inflation, budget balance, Sachs-Warner	Inflation, Sachs-Warner	CPIA
Inflation, budget balance, Sachs-Warner	1.00		
Inflation, Sachs-Warner	0.98	1.00	
CPIA	0.53	0.52	1.00

Source: Author's analysis based on sources described in the text.

Table 5. Coefficients on key terms under specification-modifying tests (original data set)

			Changing controls				Changing aid			Changing policy		
Specification	key term	Original	BD	CD	GC	DHT	EDA/ real GDP	ODA/ real GDP	ODA/ XR GDP	INFL, BB, SAC	INFL, SAC W	CPIA
Burnside & Dollar	Aid × policy	**0.19**		**0.15**	0.06	**0.20**		**0.16**	0.05		**0.29**	–0.03
		(2.61)		**(2.09)**	(1.07)	**(2.31)**		**(2.07)**	(1.96)		**(3.12)**	(–0.25)
		270		279	263	276		275	275		296	264
Collier & Dehn	Aid × policy	0.10		0.03	0.02	0.05		0.07	0.02		**0.12**	0.11
		(1.70)		(0.47)	(0.43)	(0.84)		(1.19)	(1.39)		**(2.02)**	(0.83)
	Δaid × negative shock	**0.04**		0.02	0.03	0.02		**0.04**	**0.01**		**0.03**	0.02
		(3.17)		(1.12)	(1.53)	(1.06)		**(2.67)**	**(4.27)**		**(2.04)**	(0.82)
		234		242	227	242		234	234		256	268
Collier & Dollar	Aid × policy	**0.14**	0.12		–0.07	0.12	0.17		0.03	0.05	0.06	
		(2.15)	(1.84)		(–1.91)	(1.87)	(1.70)		(1.50)	(1.00)	(1.10)	
		344	337		374	349	349		347	365	388	
Collier & Hoeffler	Post-conflict 1 × aid × policy	**0.18**	**0.18**		**0.13**	**0.18**	**0.29**		**0.04**	**0.18**	**0.18**	
		(3.92)	**(3.89)**		**(2.82)**	**(3.81)**	**(3.75)**		**(4.07)**	**(3.97)**	**(4.23)**	
		344	337		374	349	349		347	365	388	
Hansen & Tarp	Aid	**0.90**		1.10	**1.00**	1.10	–0.36	1.85		**0.94**		0.69
		(4.22)		(1.56)	**(2.57)**	(1.56)	(–0.25)	(1.30)		**(2.02)**		(1.53)
	Aid^2	**–0.02**		–0.02	**–0.03**	–0.02	–0.04	**–0.17**		–0.02		–0.02
		(–3.83)		(–1.19)	**(–2.81)**	(–1.19)	(–0.18)	**(–2.48)**		(–1.81)		(–1.76)
	ΔAid	**–0.70**		–0.47	–0.58	–0.47	–0.47	–1.46		**–0.71**		**–0.69**
		(–4.91)		(–1.42)	(–1.56)	(–1.42)	(–0.70)	(–1.74)		**(–2.63)**		**(–2.48)**
	$\Delta(Aid^2)$	**0.01**		0.01	0.02	0.01	0.07	0.12		0.01		**0.01**
		(3.64)		(1.10)	(1.66)	(1.10)	(0.53)	(2.24)		(1.86)		**(2.11)**
		213		213	181	213	214	214		213		215
Dalgaard, Hansen, & Tarp	Aid	**0.69**	**1.17**	**1.34**	**1.33**			**1.10**	**0.46**			
		(5.09)	**(6.44)**	**(6.62)**	**(5.72)**			**(5.23)**	**(4.04)**			
	Aid × tropical area %	**–0.98**	**–1.49**	**–1.79**	**–1.66**			**–1.17**	**–0.45**			
		(–3.16)	**(–8.58)**	**(–8.26)**	**(–6.72)**			**(–5.09)**	**(–3.79)**			
		371	354	371	315			371	365			
Guillaumont & Chauvet	Aid × environ-ment	–0.15	–0.13	–0.07		–0.09	–0.45	–0.31			**–0.15**	**–0.13**
		(–1.79)	(–1.73)	(–1.40)		(–1.54)	(–1.96)	(–1.77)			**(–2.13)**	**(–2.01)**
		68	71	73		73	66	66			68	69

All *t* statistics (in parentheses) are heteroskedasticity-robust; those for GMM regressions also autocorrelatior
Except for original Hansen and Tarp regression, all GMM standard errors incorporate the Windmeijer (2005) sample correction. Entries significant at 0.05 in bold.
Source: Authors' analysis based on sources described in the text.

Table 6. Coefficients on Key Terms under Data Set–Modifying Tests

Specification	key term	Original: Full sample	Original: No outliers	New data: Full sample	New data: No outliers
Burnside & Dollar	Aid × policy	**0.19**	–0.05	–0.03	–0.25
		(2.61)	(–0.45)	(–0.17)	(–1.27)
	Observations	270	263	446	436
Collier & Dehn	Aid × policy	0.10	0.11	0.06	0.10
		(1.70)	(1.11)	(1.07)	(0.73)
	ΔAid × negative shock	**0.04**	–0.06	**0.03**	–0.17
		(3.17)	(–1.33)	**(2.93)**	(–1.92)
	Observations	234	224	402	379
Collier & Dollar	Aid × policy	**0.14**	0.07	–0.01	–0.04
		(2.15)	(1.06)	(–0.19)	(–0.81)
	Observations	344	341	520	508
Collier & Hoeffler	Post-conflict × aid × policy	**0.18**	**1.18**	0.08	–0.06
		(3.92)	**(2.12)**	(1.91)	(–0.19)
	Observations	344	333	520	494
Hansen & Tarp	Aid	**0.90**	**0.96**	0.08	0.03
		(4.22)	**(2.19)**	(0.41)	(0.16)
	Aid^2	**–0.02**	–0.02	–0.001	–0.001
		(–3.83)	(–1.95)	(–0.57)	(–0.42)
	Aid, lagged	**–0.70**	**–0.70**	–0.13	–0.07
		(–4.91)	**(–2.73)**	(–0.92)	(–0.53)
	Aid^2, lagged	**0.01**	0.01	0.002	0.001
		(3.64)	(1.86)	(1.03)	(0.71)
	Observations	213	212	517	514
Dalgaard, Hansen, & Tarp	Aid	**0.69**	0.34	**0.89**	–0.69
		(5.09)	(0.20)	**(5.23)**	(–0.79)
	Aid × tropical area %	**–0.98**	–0.41	**–0.96**	0.63
		(–3.16)	(–0.25)	**(–3.65)**	(0.73)
	Observations	371	362	474	463
Guillaumont & Chauvet	Aid × environment	–0.15	–0.11		
		(–1.79)	(–1.96)		
	Observations	68	67		

All *t* statistics (in parentheses) are heteroskedasticity-robust; those for GMM regressions also autocorrelation-robust. Except for original Hansen and Tarp regression, all GMM standard errors incorporate the Windmeijer (2005) finite-sample correction. Entries significant at 0.05 in bold.

References

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

Anderson, T.W. and C. Hsiao. 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics* 18(1): 47–82.

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): 277–97.

Banks, A. 2002. "Cross-National Time-Series Data Archive." Databanks International, Bronx, NY.

Barro, R. J. 1991. "Economic Growth in a Cross Section of Countries." *American Economic Review* 106(2): 407–43.

Barro, R. J. and J. Lee. 2000. "International Data on Educational Attainment: Updates and Implications." Working Paper 7911. National Bureau of Economic Research, Cambridge, MA.

Bloom, D. and J. D. Sachs. 1998. "Geography, Demography and Economic Growth in Africa." *Brookings Papers on Economic Activity* 2: 207–73.

Boone, P. 1994. "Aid and Growth." Mimeo, London School of Economics, London.

Burnside, C. and D. Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90(4): 847–68.

Burnside, C. and D. Dollar. 2004. "Aid, Policies, and Growth: Revisiting the Evidence." Policy Research Working Paper O-2834. World Bank, Washington, DC.

Chang, C. C., E. Fernandez-Arias, and L. Serven. 1998. "Measuring Aid Flows: A New Approach." Working Paper 387. Inter-American Development Bank, Washington, DC.

Chauvet, L. and P. Guillaumont. 2002. "Aid and Growth Revisited: Policy, Economic Vulnerability and Political Instability." Paper presented at the Annual Bank Conference on Development Economics: Towards Pro-poor Policies, Oslo.

Collier, P. and J. Dehn. 2001. "Aid, Shocks, and Growth." Policy Research Working Paper 2688. World Bank, Washington, DC.

Collier, P. and D. Dollar. 2002. "Aid Allocation and Poverty Reduction." *European Economic Review* 45(1): 1–26.

Collier, P. and D. Dollar. 2004. "Development Effectiveness: What Have We Learnt?" The *Economic Journal* 114(496): F244–71.

Collier, P. and A. Hoeffler. 2004. "Aid, Policy and Growth in Post-Conflict Societies." *Euro-*

pean Economic Review 48(5): 1125–45.

Dalgaard, C. and H., Hansen. 2001. "On Aid, Growth and Good Policies." *Journal of Development Studies* 37(6): 17–41.

Dalgaard, C., H. Hansen, and F. Tarp. 2004. "On the Empirics of Foreign Aid and Growth." *The Economic Journal* 114(496): F191–F216.

Dehn, J. 2000. "Commodity Price Uncertainty in Developing Countries." Working Paper 12. Centre for the Study of African Economies, Oxford.

Development Assistance Committee (DAC). 2002. *Development Assistance Committee Online*. Paris.

Doornik, J.A., M. Arellano, and S. Bond. 2002. "Panel Data Estimation Using DPD for Ox." Available online at http://www.doornik.com/download/dpd.pdf.

Durbarry, R., N. Gemmell, and D. Greenaway. 1998. "New Evidence on the Impact of Foreign Aid on Economic Growth." CREDIT Research Paper 98r8. University of Nottingham, Nottingham.

Easterly, W. and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112(4): 1203–50.

Easterly, W. and R. Levine. 2003. "Tropics, Germs, and Crops: How Endowments Influence Economic Development." *Journal of Monetary Economics* 50(1): 3–39.

Easterly, W., R. Levine, and D. Roodman. 2004. "New Data, New Doubts: A Comment on Burnside and Dollar's 'Aid, Policies, and Growth (2000)'." *American Economic Review* 94(2).

Easterly, W. and S. Rebelo. 1993. "Fiscal Policy and Economic Growth: An Empirical Investigation." *Journal of Monetary Economics* 32(3): 417–58.

Gallup, J. L., and J. D. Sachs. 1999. "Geography and Economic Development." In Pleskovic, B. and Stiglitz, J. E. eds., Annual World Bank Conference on Development Economics, 1998 Proceedings. World Bank, Washington, DC, pp. 127–78.

Griffin, K. B. and J. L. Enos. 1970. "Foreign Assistance: Objectives and Consequences." *Economic Development and Cultural Change* 18(3): 313–27.

Guillaumont, P. and L. Chauvet. 2001. "Aid and Performance: A Reassessment." *Journal of Development Studies* 37(6): 66–92.

Hadi, A. S. 1992. "Identifying Multiple Outliers in Multivariate Data." *Journal of the Royal Statistical Society*, Series B 54: 761-777.

Hadjimichael, M. T., D. Ghura, M., Muhleisen, R. Nord, and E. M. Ucer. 1995. "Sub-Saharan

Africa: Growth, Savings, and Investment, 1986–93." Occasional Paper 118. International Monetary Fund, Washington, DC.

Hansen, H. and F. Tarp. 2000. "Aid Effectiveness Disputed." *Journal of International Development* 12(3): 375–98.

Hansen, H. and F. Tarp. 2001. "Aid and Growth Regressions." *Journal of Development Economics* 64(2): 547–70.

Holtz-Eakin, D., W. Newey, and H. S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56(6): 1371–95.

International Monetary Fund (IMF). 2003. *International Financial Statistics Database*. Washington, DC.

Jepma, C. J. 1991. "The Tying of Aid." OECD Development Centre, Paris.

Kaufmann, D., A. Kraay, and M. Mastruzzi. 2003. "Governance Matters III: Governance Indicators for 1996–2002." Policy Research Working Paper 3106. World Bank, Washington, DC.

Killick, T. 1998. "Aid and the Political Economy of Policy Change." Routledge, London.

King, R. G. and R. Levine. 1993. "Finance and Growth: Schumpeter Might be Right." *American Economic Review* 108(3): 717–37.

Knack, S. and P. Keefer. 1995. "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures." *Economics and Politics* 7(3): 207–27.

Leamer, E. E. 1983. "Let's Take the Con out of Econometrics." *American Economic Review* 73(1): 31–43.

Levine, R. and D. Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review* 82(4): 942–63.

Lensink, R. and H. White. 2001. "Are There Negative Returns to Aid?" *Journal of Development Studies* 37(6): 42–65.

Lu, S. and R. Ram. 2001. "Foreign Aid, Government Policies, and Economic Growth: Further Evidence from Cross-country Panel Data for 1970–93." *Economia Internazionale* 54(1): 15–29.

Mankiw, N. G., D. Romer, and D. N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *American Economic Review* 107(2): 407–37.

Mosley, P., J. Hudson, and S. Horrell. 1987. "Aid, the Public Sector and the Market in Less Developed Countries." *Economic Journal* 97(387): 616–41.

Rajan, R. and A. Subramanian. 2005. "Aid and Growth: What Does the Cross-Country Evidence Really Show?" Working Paper 05/127. International Monetary Fund, Washington, DC.

Ram, R. 2004. “Recipient Country’s ‘Policies’ and the Effect of Foreign Aid on Economic Growth in Developing Countries: Additional Evidence.” *Journal of International Development* 16(2): 201–11.

Rodríguez, F. and D. Rodrik. 2001. “Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence.” In Bernanke, B. and Rogoff, K. S., eds., National Bureau of Economic Research Macroeconomics Annual. MIT Press, Cambridge, MA.

Roeder, P.G. 2001. “Ethnolinguistic Fractionalization (ELF) Indices, 1961 and 1985.” Avaliable on http//:weber.ucsd.edu/~proeder/elf.htm, accessed May 2004.

Roodman, D. 2006. “How to Do xtabond2: An Introduction to “Difference” and “System” GMM in Stata.” Working Paper 103. Center for Global Development, Washington, DC.

Sachs, J. D. 2001. “Tropical Underdevelopment.” Working Paper W8119. National Bureau of Economic Research, Cambridge, MA.

Sachs, J. D. 2003. “Institutions Don’t Rule: Direct Effects of Geography on per Capita Income.” Working Paper W9490. National Bureau of Economic Research, Cambridge, MA.

Sachs, J. D. and A. Warner. 1995. “Economic Reform and the Process of Global Integration.” *Brookings Papers on Economic Activity 1995(1)*: 1–118.

Sala-I-Martin, X. X. 1997. “I Just Ran Two Million Regressions.” *American Economic Review* 87(2): 178–83.

Summers, R. and A. Heston. 1991. “The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–88.” *Quarterly Journal of Economics* 106(2): 327–68.

Svensson, J. 1999. “Aid, Growth and Democracy.” *Economics and Politics* 11(3): 275–97.

Van de Walle, N. and T. A. Johnston. 1996. “Improving Aid to Africa.” Policy Essay 11. Overseas Development Council, Washington, DC.

United Kingdom Department for International Development (DFID). 2000. “Eliminating World Poverty: Making Globalisation Work for the Poor.” White Paper on International Development Presented to Parliament by the Secretary of State for International Development by Command of Her Majesty, London.

United States Department of State (various years). *World Military Expenditures and Arms Transfers*. Washington, DC.

Wacziarg, R. and K. H. Welch. 2002. “Trade Liberalization and Growth: New Evidence.” Mimeo, Stanford University, Stanford, CA.

Windmeijer, F. 2005. “A Finite Sample Correction for the Variance of Linear Two-step GMM Estimators.” *Journal of Econometrics* 126: 25–51.

World Bank. 2003. World Development Indicators 2003 Database. Washington, DC.

Appendix 1. Data set construction

The data set used in this study is based on that of Easterly, Levine, and Roodman (2004). Some variables in that set have been slightly revised. Others have been added to match the data sets of the tested regressions. The period of coverage has been pushed back to 1958 and forward to 2001 where possible. All data were collected from standard sources, expect for countries' export price indexes, which were provided by Jan Dehn (see Dehn 2000). (See Table A1.)

Following are notes on the data set construction:

Revisions since Easterly, Levine, and Roodman (2004)

- Some observations for inflation are completed by using wholesale inflation where consumer price inflation was unavailable.
- The update of the Sachs-Warner variable is slightly revised under the influence of the independent update by Wacziarg and Welch (2002).
- Some missing values for Effective Development Assistance during 1975–95, the period of the EDA data set, are filled in in the same manner as missing values outside this period already were, via a regression of EDA on Net ODA.
- ICRGE now varies over time, rather than taking 1982 values throughout. Observations before 1982 are assigned 1982 values. In addition, the variable is revised in order to extend it after 1997. In 1998, the PRS Group stopped reporting two of ICRGE's original components, Expropriation Risk and Repudiation of Government Contracts. So these were dropped entirely from the variable, leaving Corruption, Bureaucratic Quality, and Rule of Law. On annual data, the revised ICRGE has a 0.97 correlation with the original.
- Some missing values for ethno-linguistic fractionalization are filled in from Roeder

(2001).

Expansion of period

- Data was collected where available for 1958–2001.
- The Collier and Dehn shocks variables are only updated to 1997 because the underlying data on export prices in Dehn (2000) cease in 1997.
- The Guillaumont and Chauvet environment variable is not updated, for lack of underlying data on its four components.
- The 1998–2001 values for the updated Sachs-Warner variable are based on 1998 data only. Currency Data International, the long-time source of black market premium data, which is one component of Sachs-Warner, shut down in 1999.

Table A2 documents the changes in sample that the new data set brings to the tested regressions.

Table A1. Construction of Data Set

Variable	Code	Data source	Notes
Per-capita GDP growth	GDPG	World Bank 2003	
Initial GDP per capita	LGDP	Summers and Heston 1991, updated using GDPG	Natural logarithm of GDP/capita for first year of period; constant 1985 dollars
Ethno-linguistic fractionalization, 1960	ETHNF	Roeder 2001	Probability that two randomly chosen individuals differ ethnically
Assassinations/ capita	ASSAS	Banks 2002	Assassinations/capita
Political instability, lagged	PINSTAB	Banks 2002	Simple average of ASSAS and revolutions/year
Institutional quality	ICRGE	PRS Group's IRIS III data set (see Knack and Keefer 1995)	Revised. Computed as the average of the three components still reported after 1997, dropping two.
M2/GDP, lagged one period	M2–1	World Bank 2003	
Sub-Saharan Africa	SSA	World Bank 2003	Codes nations in the southern Sahara as sub-Saharan
East Asia	EASIA		Dummy for China, Indone-

			sia, South Korea, Malaysia, Philippines, and Thailand
Central America	CENTAM	World Bank 2003	
Franc zone	FRZ	Burnside and Dollar 2000	Codes African nations in the CFA franc zone
Egypt	EGYPT		
Budget surplus	BB	World Bank 2003; IMF 2003	World Bank primary source. Additional values extrapolated from IMF, using series 80 and 99b (local-currency budget surplus and GDP)
Inflation	INFL	World Bank 2003; IMF 2003	log (1 + inflation). World Bank primary source. Wholesale price inflation from IMF used to fill gaps
Sachs-Warner, updated	SACW	Sachs and Warner 1995; Easterly, Levine, and Roodman 2004; Wacziarg and Welch 2002	Extended to 1998. Slightly revised pre-1993.
Positive and negative shock	POSSHOCK NEGSHOCK	Dehn 2000	Shocks are % price index changes. "Shock" threshold country-specific. Reconstructed based on underlying index data for 1957–97
Effective Development Assistance/real GDP	AID	Chang, Fernandez-Arias, and Serven 1998; OECD-DAC 2002; IMF 2003; World Bank 2003; Summer and Heston 1991	Available values for 1975–95 from Chang, Fernandez-Arias, and Serven. Missing values extrapolated based on regression of EDA on Net ODA. Converted to 1985 dollars with World Import Unit Value index from IMF, series 75. GDP computed like LGDP above
Net Overseas Development Assistance/real GDP	ODAPPPGDP	OECD-DAC 2002; IMF 2003; World Bank 2003; Summer and Heston, 1991	Like AID exception using net ODA from OECD-DAC
Net Overseas Development Assistance/nominal GDP	ODAXRGDP	OECD-DAC 2002; World Bank 2003	
Dummy for end of civil conflict in previous period	POSTCONFLICT1	Collier and Hoeffler 2004	
Tropical area fraction	TROPICAR	Gallup and Sachs 1999	
Population	LPOP	World Bank 2003	log (population)
Population growth	POPG	World Bank 2003	
Mean years of secondary schooling among	SYR	Barro and Lee 2000	

those over 25		
Arms imports/total imports, lagged	ARMS-1	U.S. Department of State, various years

All variables aggregated over time using arithmetic averages.

Table A2. Overview of Differences in Regression Samples, Original and Expanded Data Sets

Regression	Lost in expanded data set	Gained in expanded data set	Study period	
			Original	Expanded
Burnside & Dollar	Somalia, Tanzania	Burkina Faso, Bulgaria, China, Cyprus, Hungary, Iran, Jordan, Myanmar, Papua New Guinea, Poland, Republic of Congo, Romania, Singapore, South Africa, Uganda	1970–93	1970–2001
Collier & Dehn	" "	" "	1974–93	1974–97
Collier & Dollar, Collier & Hoeffler	None	Angola, Burkina Faso, Bulgaria, China, Cyprus, Czech Republic, Guinea, Guinea-Bissau, Iran, Jordan, Liberia, Mongolia, Mozambique, Myanmar, Namibia, Oman, Papua New Guinea, Poland, Republic of Congo, Romania, Somalia, Suriname, Uganda	1974–97	1974–2001
Hansen & Tarp	Somalia	Angola, Burundi, Benin, Burkina Faso, Barbados, Bulgaria, Central African Republic, Chad, China, Cyprus, Hungary, Iran, Jordan, Mauritania, Mauritius, Mozambique, Myanmar, Nepal, Papua New Guinea, Poland, Republic of Congo, Romania, Rwanda, Singapore, South Africa, Uganda	1978–93	" "
Dalgaard, Hansen, and Tarp	None	Bangladesh, Czech Republic, Guinea-Bissau, Singapore, South Africa, Tanzania	1974–97	" "
Guillaumont & Chauvet	None	None	1970–93	1970–93

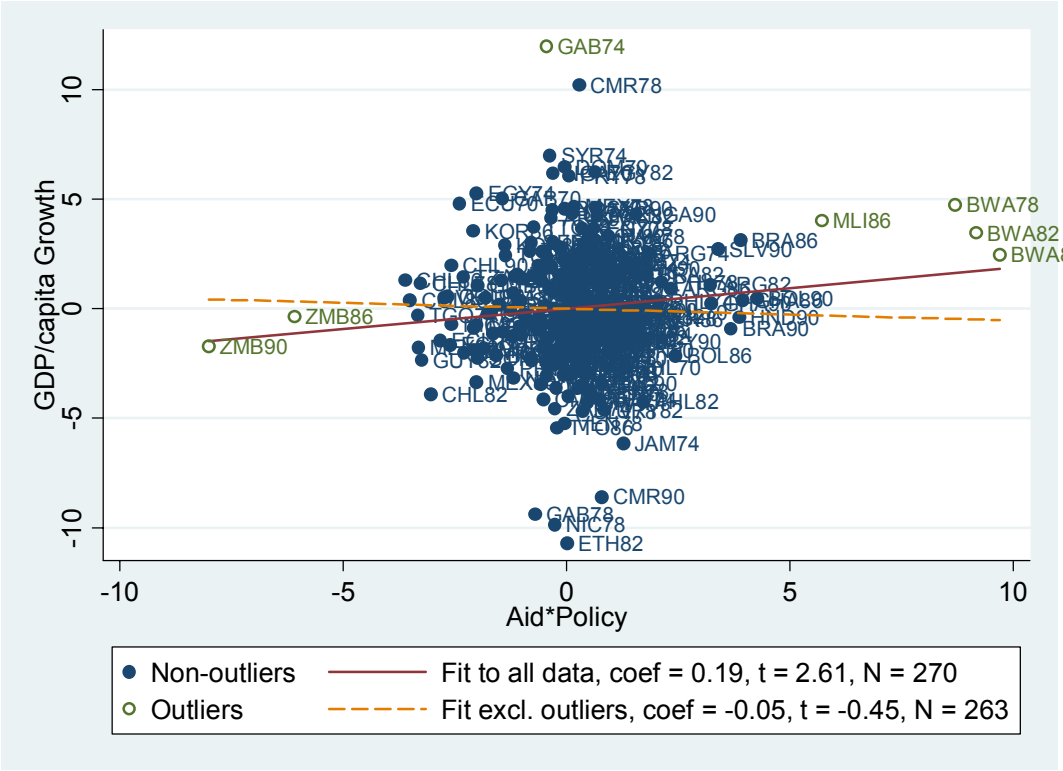
Source: Author's analysis based on sources described in the text.

Appendix 2. Partial scatters of growth against variables of interest

Each figure below is a partial scatter of GDP/capita growth versus a variable of interest in the context of a tested regression. The figures correspond to OLS and 2SLS results of table 6 in the main text. Data points are labeled by ISO three-letter codes and the last two digits of the starting year of the period. Outliers are marked separately, and two partial regression lines are shown, one for the full sample, one for the sample excluding outliers. Note that the second line in each graph is not the best fit to the non-outlier data points as plotted. Deleting observations causes the estimated coefficients to shift and all remaining data points in the partial scatter to move. The second line is the best fit to the data points in their post-exclusion positions, which are not shown.

Figure A1. B&D regression 5/OLS: Partial scatter of GDP/capita growth versus aid×policy

Original data



Expanded sample

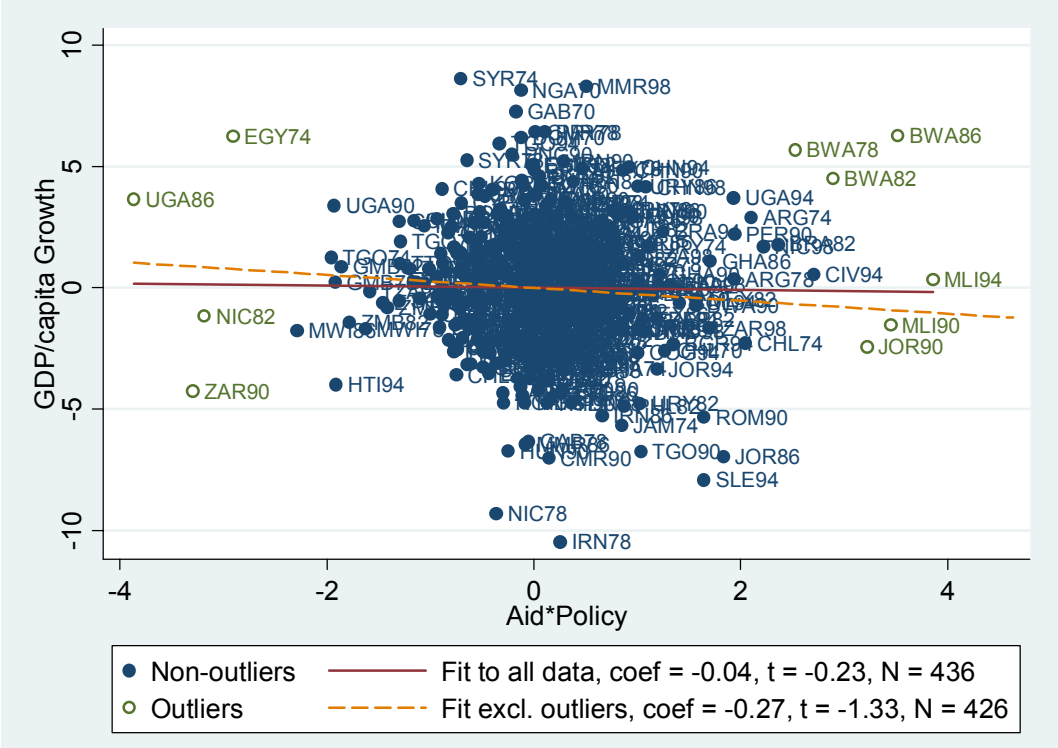
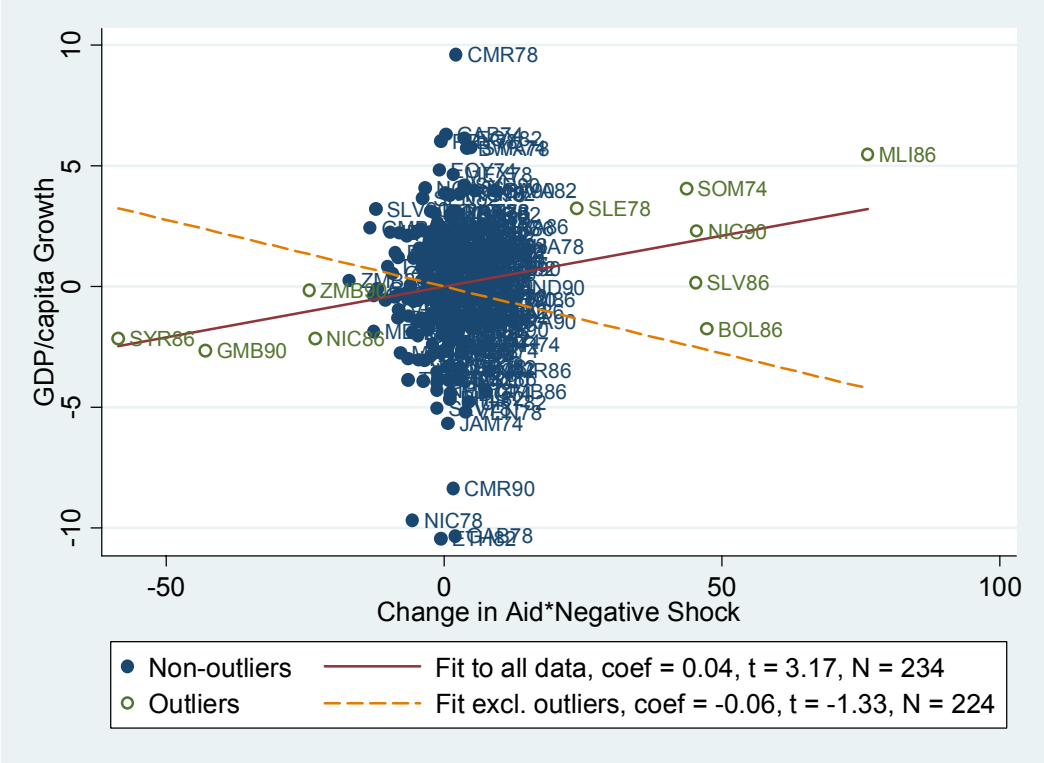


Figure A2. Collier & Dehn regression: Partial scatter of GDP/capita growth vs. Δaid×neg. shock

Original data



Expanded sample

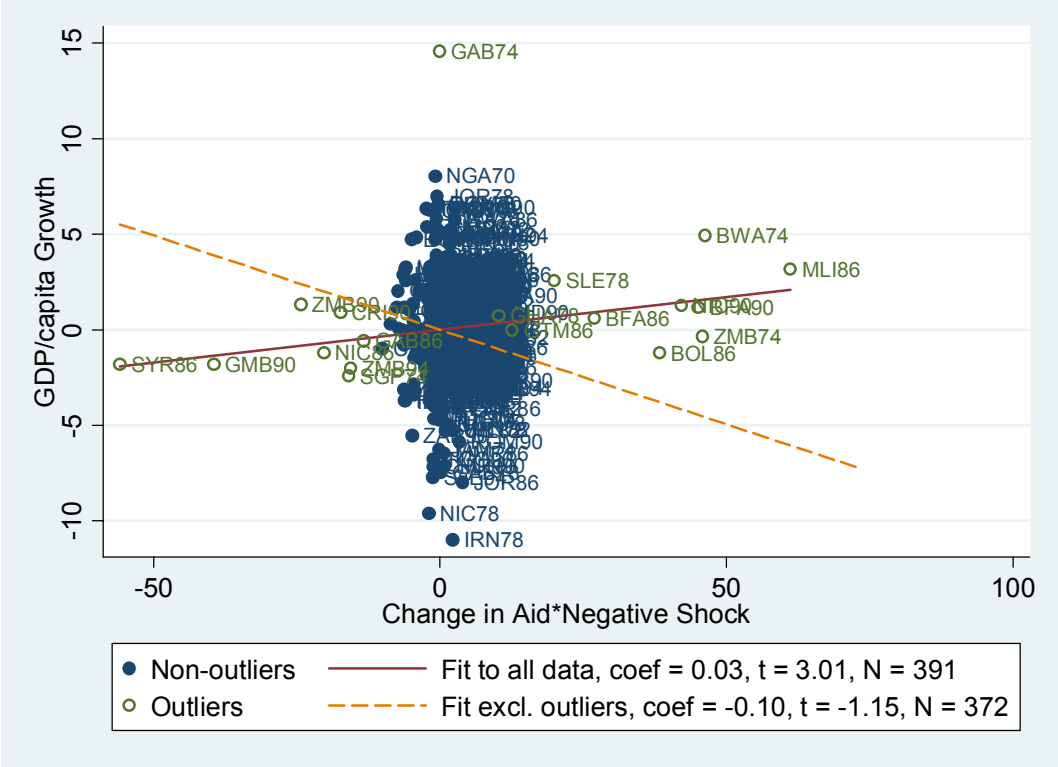
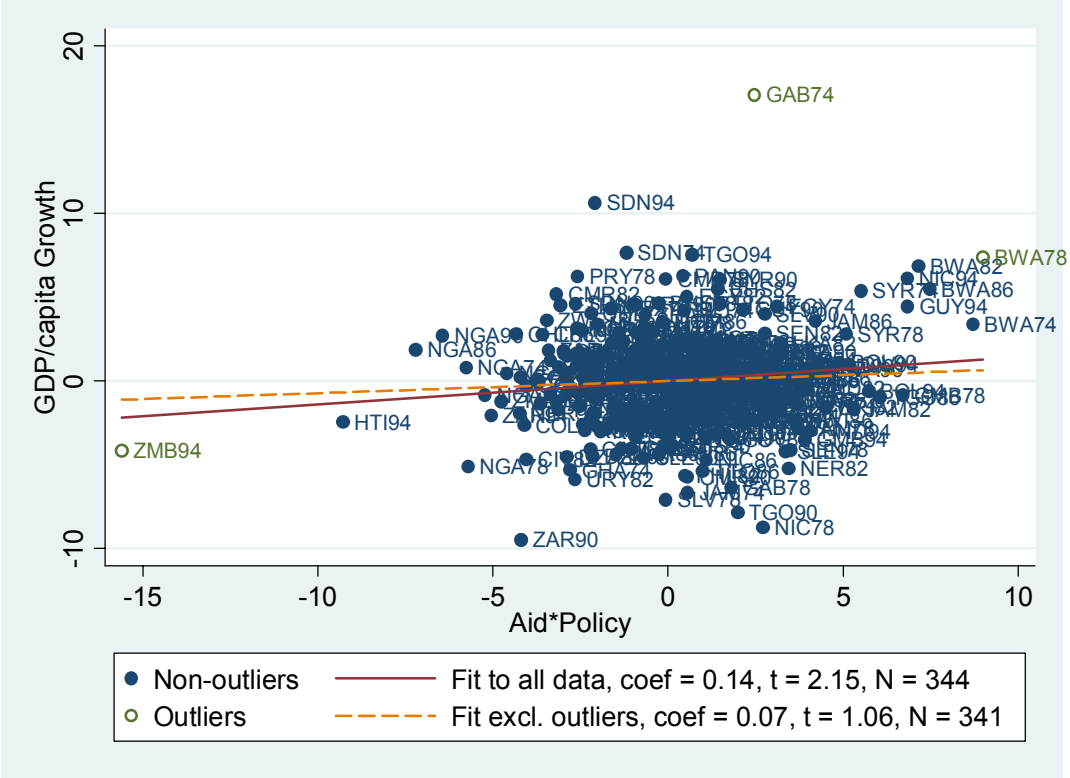


Figure A3. Collier & Dollar regression: Partial scatter of GDP/capita growth vs. aid×policy

Original data



Expanded sample

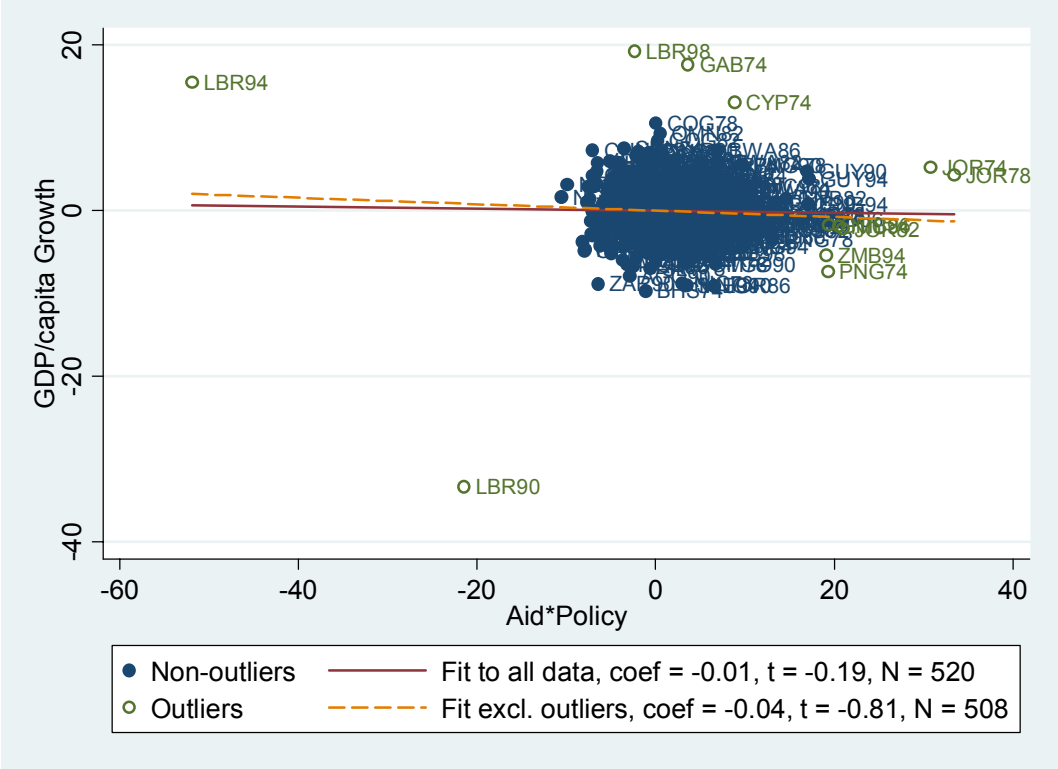
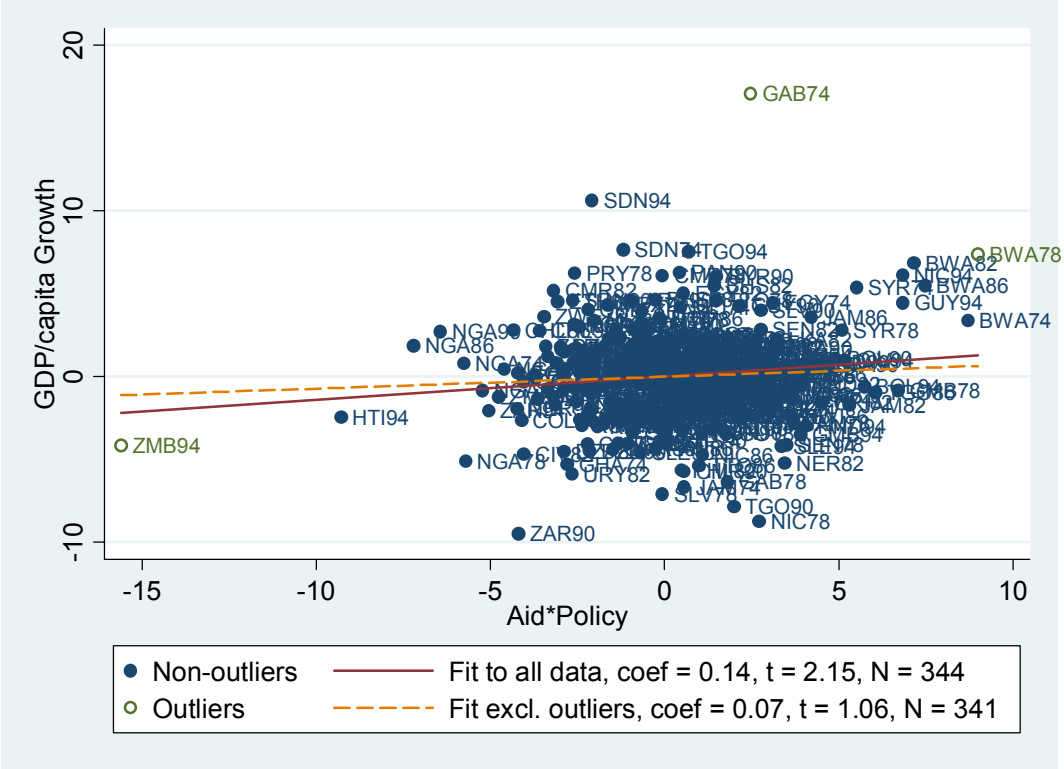


Figure A4. Collier & Hoeffler regression: Partial scatter of GDP/capita growth vs. Post-conflict 1×aid ×policy

Original data



Expanded sample

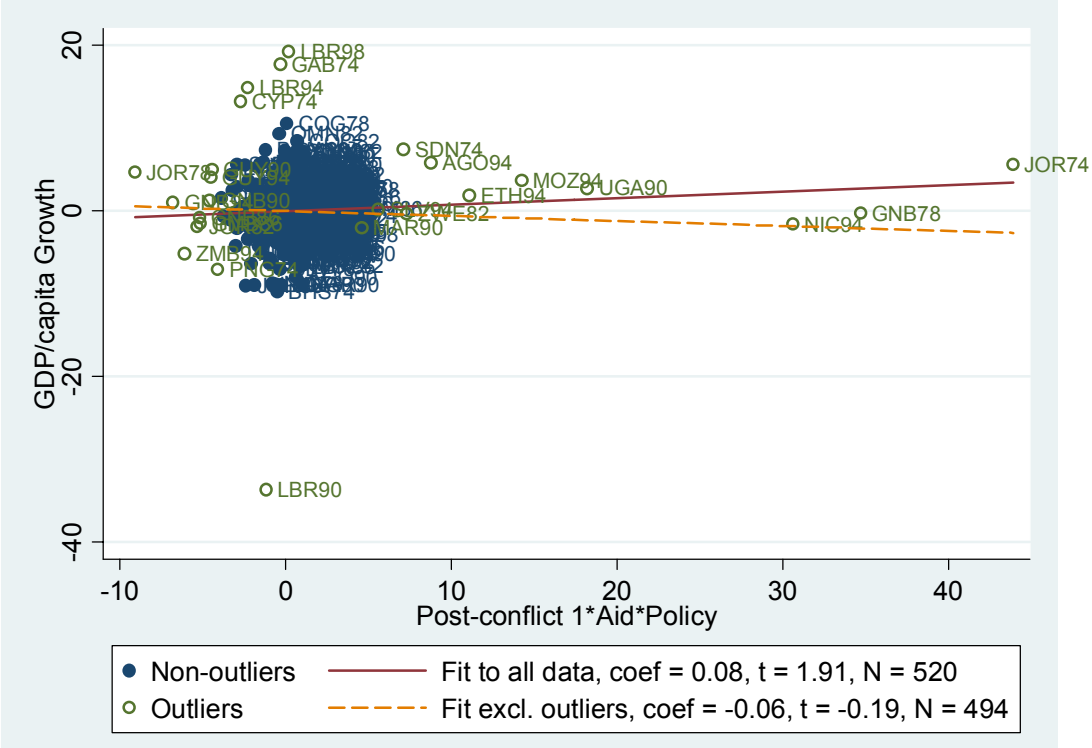


Figure A5. Guillaumont and Chauvet Regression 1.2: Partial Scatter of GDP/Capita Growth versus Aid×Environment

Original data