

**Through the Looking Glass, and What OLS Found There:
On Growth, Foreign Aid, and Reverse Causality**
By David Roodman

Abstract

The cross-country literature on foreign aid effectiveness has relied on the use of instruments to distinguish causality from mere correlation. This paper uses simple non-instrumental techniques in the spirit of Granger to demonstrate that the main aid-growth connection is a negative causal relationship from growth to aid—aid, that is, as a fraction of recipient GDP. Coarsely, when GDP goes up, aid/GDP goes down. The endogeneity of aid, long-suspected, is real. Less understood is that adding certain common controls to regressions puts this relationship through the looking-glass, flipping both its sign and apparent direction: aid seems to cause growth. Ideally, instrumentation expunges the endogeneity shown here. In practice, estimates of aid's impact have run into problems. Autocorrelation in the errors is widespread, and can render endogenous lagged variables used as regressors or instruments. The pitfalls of “difference” and “system” include invalidity and proliferation of instruments. Multicollinearity in term pairs of interest, such as aid and aid² or “project” and “program” aid, can amplify endogeneity bias. The combination of specification problems and widespread fragility (shown in earlier work) leads to pessimism about the ability of cross-country econometrics to demonstrate aid effectiveness. This does not rule an average positive effect, nor does it contradict the fact that aid has saved millions of lives, but it does suggest that the average effect on economic growth is too small to be detected statistically.

The Center for Global Development is an independent think tank that works to reduce global poverty and inequality through rigorous research and active engagement with the policy community. Use and dissemination of this working paper is encouraged, however reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License. The views expressed in this paper are those of the author and should not be attributed to the directors or funders of the Center for Global Development.

Through the Looking-Glass, and What OLS Found There: On Growth, Foreign Aid, and Reverse Causality¹

David Roodman²

Center for Global Development

Abstract: The cross-country literature on foreign aid effectiveness has relied on the use of instruments to distinguish causality from mere correlation. This paper uses simple non-instrumental techniques in the spirit of Granger to demonstrate that the main aid-growth connection is a negative causal relationship from growth to aid—aid, that is, as a fraction of recipient GDP. Coarsely, when GDP goes up, aid/GDP goes down. The endogeneity of aid, long-suspected, is real. Less understood is that adding certain common controls to regressions puts this relationship through the looking-glass, flipping both its sign and apparent direction: aid seems to cause growth. Ideally, instrumentation expunges the endogeneity shown here. In practice, estimates of aid’s impact have run into problems. Autocorrelation in the errors is widespread, and can render endogenous lagged variables used as regressors or instruments. The pitfalls of “difference” and “system” include invalidity and proliferation of instruments. Multicollinearity in term pairs of interest, such as aid and aid² or “project” and “program” aid, can amplify endogeneity bias. The combination of specification problems and widespread fragility (shown in earlier work) leads to pessimism about the ability of cross-country econometrics to demonstrate aid effectiveness. This does not rule an average positive effect, nor does it contradict the fact that aid has saved millions of lives, but it does suggest that the average effect on economic growth is too small to be detected statistically.

¹ With apologies to Lewis Carroll.

² Research Fellow, Center for Global Development. I thank Carl-Johann Dalgaard, Henrik Hansen, and Finn Tarp for sharing data and programs; Selvin Akkus for enthusiastic research assistance; and Michael Clemens, Takashi Kurosaki, Andrew Powell, Steven Radelet, Yasuyuki Todo, David Wheeler, and participants in the July 2007 workshop at Research Institute of Economy, Trade, and Industry in Tokyo for helpful comments. Address for correspondence: droodman@cgdev.org.

It has been just over ten years since World Bank researchers David Dollar and Craig Burnside disseminated in a working paper the electrifying finding that foreign aid raises economic growth in countries with good economic policies. Since then, the aid community has become somewhat enamored of cross-country aid-growth regressions, then somewhat less so. But to this day, studies of the type remain influential. Practitioners in the \$80 billion-a-year aid enterprise care about their work and hanker for objective and persuasive evidence that they are helping.³ There are political pressures for the same. Cross-country regressions have promised to meet the need in grand fashion by looking at aid as a whole across the universe of poorer countries.

On a pessimistic note, Easterly, Levine, and Roodman (2004) and Roodman (2007a) show that fragility is widespread in the contemporary aid-growth econometric literature—the first finds it in Burnside and Dollar (BD, 2000) and the second in six more studies. However, the present paper makes the case that it would be wrong to dismiss the entire literature as fragile. Some studies have found robust, positive associations between aid and growth—at least certain kinds of aid, or in certain kinds of countries.

Here, we argue that reported associations between aid and growth are indeed more than noise filtered by publication bias. There *is* a clear aid-growth relationship in the four- or five-year timeframes normally studied in aid-growth panel regressions. But instead of being positive and running causally from aid to growth, it is negative and runs from growth to aid—aid, that is, as it is usually measured, as a fraction of recipient gross domestic product (GDP). Coarsely, when GDP goes up, aid/GDP goes down. More precisely, the short-term elasticity of a recipient's total aid with respect to its GDP is less than 1, so that the short-term elasticity of aid/GDP with respect to GDP (and GDP/capita) is negative.

³ Figure is for Net Aid Transfers from traditional donors in 2005 (Roodman 2006a).

The possibility of reverse causation from growth to aid is not news, though clear demonstrations are rare. What seems little appreciated is that adding certain common controls to regressions of growth on aid can put this negative link from growth to aid/GDP through the looking glass, making it look like a positive link flowing causally from aid to growth. This paper argues that that is happening, and explains how.

Having demonstrated that the dominant statistical relationship between aid and growth is endogenous (to growth), the paper then discusses a number of common specification problems that have undermined attempts to eliminate such endogeneity through instrumentation. These include autocorrelation, instrument proliferation, and multicollinearity. The paper concludes by positing that some of these specification difficulties are arising inevitably as researchers strive against limits to what econometrics can glean from the data. This does not rule an average positive effect of aid, nor does it contradict the fact that aid has saved millions of lives, but it does suggest that the average effect on economic growth is too small to be detected statistically.

1. The sign flip

1.1. *A small mystery*

Especially since the mid-1990s, fear of endogeneity has shaped the aid-growth literature, leading to routine use of instrumental variables estimators. However, the nature of that endogeneity—its sign, strength, and causality—has been little explored. In this section, we perform some simple regressions in the spirit of Granger causality testing in order to investigate the endogeneity. In the process, we unearth an odd pattern.

For the sake of an example, we adapt the preferred 2SLS regression in Hansen and Tarp (HT, 2001), the leading published study in the “unconditional” aid-growth literature, which tends

to argue that aid works on average, albeit with diminishing returns.⁴ In this specification, HT regress per-capita economic growth on recipient's aid/GDP and $(\text{aid/GDP})^2$ in a panel that covers 1974–93 with four-year periods. They copy Burnside and Dollar's controls but introduce different instruments, which are the one-period lags of aid/GDP, its square, the BD economic policy index, and quadratic and cubic terms involving these.⁵ To fit this regression into the simple OLS approach taken here, we discard the instrumented variables, aid/GDP and $(\text{aid/GDP})^2$, and bring some or all of the instruments into the regressor list in their place. OLS regressions on these variables are valid under essentially the same assumptions as the original.

Table 1 exhibits the pattern of interest. The first panel shows the results of regressing growth solely on aid/GDP and $(\text{aid/GDP})^2$, with no other controls, using the original HT data set. Column 1 in the panel corresponds most closely to the original HT regression in that it regresses on *lagged* aid/GDP and $(\text{aid/GDP})^2$, which are instruments in the HT regression. The second and third columns instead regress on current or future observations of the two aid variables. This change looks strange within HT's analytical framework—HT make no claim that current or future aid is exogenous to growth—but is akin to a Sims (1972) causality test for causality from growth to aid. (A proper Sims test is below.) The coefficient on the linear aid variable is negative throughout, with the largest coefficient and smallest standard error in the forward-observations regression.

For the second panel of Table 1, we include the full set of HT exogenous variables—regressors and excluded instruments—as controls. In the first column, the coefficient on aid/GDP

⁴ Regression 1.2 in HT (2001). HT's overall preferred specification is rather different, using difference GMM.

⁵ The controls are: log initial GDP/capita; assassinations/capita, ethnic fractionalization in 1960, and their product; dummies for Sub-Saharan Africa and fast-growing East Asia; money supply/GDP, lagged; the ICRG governance variable; and period dummies. Instead of entering Burnside and Dollar's economic policy index as a control, they enter its underlying variables: inflation, budget surplus/GDP, and Sachs-Warner openness. The instruments are the one-period lags of aid/GDP, $(\text{aid/GDP})^2$, policy, policy^2 , log population \times policy, log initial GDP/capita \times policy, (log initial GDP/capita) $^2\times$ policy, aid/GDP \times policy, and $(\text{aid/GDP})^2\times$ policy.

is now significant, positive, and close to the 0.241 in HT's 2SLS original. This contrasts with the negative coefficient in the regression above it in the table. But as one scans the second panel from left to right, the coefficients on both aid variables change monotonically, and flip sign.⁶ The clustered p value for the negative coefficient on forward aid/GDP is 0.12; as we will see, regressing on more distant forward lags only increases the significance.

In sum, without controls, the aid-growth relationship appears consistently negative, if more strongly for forward aid. Adding controls flips the sign on lagged aid.

This pattern has a precedent. In a similar analysis, Blomström, Lipsey, and Zejan (BLZ, 1996) use OLS to investigate causal relationships between growth and investment/GDP in a cross-country panel with 5-year periods. They question the conclusion of De Long and Summers (1991) about the importance of (equipment) investment for growth. Regressions like those above but substituting investment/GDP for aid/GDP produce almost exactly the same pattern—except with opposite signs throughout. Regressions on lagged, current, and forward investment/GDP all return a positive sign. Adding controls, country dummies in this case again produces a sign flip, putting a negative sign on lagged investment/GDP. (See Table 2.)

In their *Quarterly Journal of Economics* article, BLZ do not explain the sign flip. What could? In the working paper version of their study, BLZ (1993) conjecture that it reflects heavy investment in the run-up to the 1982 global debt crisis, after which growth plunged. As for the aid-growth sign flip, one partial explanation is that HT are right about the meaning of the positive coefficients on lagged aid with controls: aid works, with diminishing returns. Some other story must then explain the negative coefficients on the right side of Table 1. Perhaps donors

⁶ In the regression for the second panel, the same shift—from lagged to current to forward—is applied to all the variables whose lags are instruments in the original HT regression.

view countries that have recently achieved strong economic growth as less needy, and so give them less aid.

But these theories are rather complicated and ad hoc. And they do not explain the striking anti-symmetry between the aid and investment results, nor the consistently negative coefficients on aid when there are no controls. The overall pattern is elegant. It transcends the choice of regressor. Something mathematical appears to be at work.

1.2. More evidence

Before offering a unified theory, we report results from a fuller set of tests of the causal relationships between aid and growth, copying and augmenting the BLZ framework.

First, we expand Table 1. For statistical power, this time we work on a much larger data set than in the HT study, covering 1962–2001, again with four-year periods. But for simpler and less arbitrary tests, we hew closely to the BLZ model, dispensing with the quadratic aid term and HT’s complex control set. The first row in Table 3 regresses growth on the *change* in aid/GDP, without controls, in order to investigate temporal relationships between shocks to aid/GDP and shocks to GDP/capita. It shows a generally negative relationship, strongest for current and forward changes in aid/GDP. This suggests that growth negatively Granger-causes aid, with an effect that persists for at least one four-year period. The second row, like the first of Table 1, regresses growth on aid/GDP in levels rather than differences. Again, the coefficients start at essentially zero on the left and slide negative toward the right. But here, since aid/GDP in levels is an accumulation of past innovations, growth shocks in the current period leave traces in aid/GDP far into the future: the coefficient declines monotonically from left to right and gains significance. The third row of Table 3 copies BLZ in introducing country dummies into the regression on aid/GDP in levels. The coefficients on the once- and twice-lags of aid/GDP now become

markedly positive, at 1.1–1.7 standard errors above zero: there is a sign flip. If we instead control for period-initial GDP/capita (bottom row), there is some evidence of a sign flip, though this time the coefficients are only half a standard error north of 0.

In all the regressions on aid/GDP in levels, the strongest coefficients continue to appear on the most forward observations of the aid variable. As BLZ argue for the case of investment, this pattern suggests that the best model of the aid-growth relationship puts aid after growth. The positive coefficients on lagged aid/GDP in the presence of the controls are not as strong as with Table 1's quadratic aid term and full control set, though reasonably strong with country dummies. The weakness of the positive coefficients on lagged aid, next to the continuing strength of those on forward aid, suggest that whatever is at work behind the sign flip is second-order and indirect, perhaps an artifact of the first-order causality from growth to aid.

These tests show that the dominant aid-growth link runs temporally from growth to aid. But they do not rule out causation the other way. For a more formal examination of causality in both directions, we perform Granger (1969) and Sims (1972) tests (see Table 4), which are rare in the aid-growth literature.⁷ They paint a clear picture of Granger causality from growth to aid but not vice versa. Again sticking to the BLZ specifications, we set the lag limits involved to two periods. In the Sims tests, we include lags of the dependent variable to reduce serial correlation in the errors.⁸ *F* tests of the joint significance of the terms of interest are at the bottom of the table. In the Granger tests, lagged observations of aid/GDP add no explanatory power to a regression of growth on its own lags. But lagged growth does help predict aid/GDP even controlling

⁷ One example is Arvin and Barillas (2002), who find no overall relationship in either direction between aid/GNP and GNP/capita but do find positive causality from aid/GNP to GNP/capita and negative causality from GNP/capita to aid/GNP in Latin America and the Caribbean, and positive causality from GNP/capita to aid/GNP in Sub-Saharan Africa.

⁸ Simulation results in Geweke, Meese, and Dent (1983) favor causality tests with the two structures used here and in BLZ.

for its own lags. Similarly, in the Sims tests, future aid/GDP enters significantly when added to a regression of growth on past values of both variables, but the reverse is not true. Here too, the aid-growth results match BLZ's investment-growth results.

1.3. An explanation

There is little doubt that growth negatively Granger-causes aid/GDP (and positively Granger-causes investment/GDP). This is unremarkable, if seldom remarked. Crudely, it makes sense that when GDP goes up, aid/GDP goes down. More precisely, the short-term elasticity of aid with respect to GDP is less than unity, so that that of aid/GDP with respect to GDP (or GDP/capita) is negative. Whether out of inertia or policy, donors do not increase aid as fast as recipients' economies grow, or reduce it as fast as they shrink.⁹ In this sense, aid is counter-cyclical. Meanwhile, investment is pro-cyclical, collapsing during economic crises even more than GDP and rebounding quickly during recoveries. This difference between aid and investment is akin to the fact that Tables 1 and 2 are alike except with opposite signs throughout.

So the outstanding question, the small mystery, is this: what explains the sign flip on lagged aid/GDP and investment/GDP when controls are added? Does it require major additional assumptions about the real world? Or is there some parsimonious premise about the data generating process that suffices?

There is. Consider a growth regression on aid/GDP and country dummies. For intuition, imagine we observe Country A and Country B at period t . We translate the conditioning on country dummies into an assumption about this scenario by assuming that the two countries grow the same in total over the study period. Now suppose that in period $t - 1$, Country A receives more aid/GDP than Country B. All else equal, since aid/GDP contains information about a coun-

⁹ Mosley, Hudson, and Horrell (1987, Table 4) report point estimates for the elasticity of aid/GNP to GNP of -0.59 and -0.93 .

try's growth history, encoded with opposite sign, higher aid/GDP for A indicates a *worse* growth record through $t - 1$. Implicit here is an assumption about the initial conditions: that the difference between A's and B's GDP/capita at the start of the study period is zero—or, more generally, expected to be zero because the initial levels are not correlated with subsequent innovations in aid and growth. If A grows less than B through $t - 1$, yet grows the same over the study period, then A must grow faster after $t - 1$, leading us to expect higher growth at t in particular. Thus can higher aid/GDP in the past precede and appear to cause higher growth now. That is the sign flip. Much the same story can be told for investment/GDP, except that higher investment/GDP at $t - 1$ indicates a *stronger* growth record up to then and predicts lower growth thereafter.

The mental exercise for a regression that controls for initial GDP/capita instead of country dummies is more complicated, in an interesting way. The change in controls leads us to assume that A and B, instead of growing the same amount over the study period, have the same GDP/capita at the beginning of period t . Again, A has higher aid/GDP during period $t - 1$, thus lower growth previous to then. If A, by assumption, catches up to B in GDP/capita by the end of $t - 1$ despite a generally weaker growth history, then it must experience a growth spurt during $t - 1$. If we then adduce a new assumption, of positive serial correlation in growth innovations, that raises the expectation of growth in period t . Because this statistical argument involves a step not needed in the fixed effects scenario, from growth in $t - 1$ to growth in t , the correlation between lagged aid/GDP and growth should be weaker in this scenario, as we find it to be.

The general principle is this: If y Granger-causes x , simple and partial correlations of y with past x can have opposite signs if the partial correlation is conditioned on controls that contain information about y -history subsequent to the observation of the past x . Whether this phe-

nomenon occurs asymptotically depends on the nature of any Granger causality from x to y , the correlation structure of the innovations in x and y , the initial conditions, and the exact controls. The Appendix formally demonstrates the effect for the two scenarios just described. Although our concrete examples of x are “something over GDP,” i.e., computed as a ratio whose denominator is tightly linked to y , this is not necessary for the looking-glass effect.¹⁰ All that is needed is a clearly signed marginal elasticity between x and y . It could just as well happen with inflation, say, or credit to private entities as a fraction of total credit.

In a sense, this carefully motivated and developed argument simply illustrates the general principle that conditional and unconditional covariances mean different things.

Some blend of the two scenarios out offers a plausible model for many aid-growth regressions. The HT regression we started with, for instance, includes dummies for Sub-Saharan Africa and fast-growing East Asia, initial GDP/capita, inflation, and budget surplus/GDP, all of which contain information about growth history. Meanwhile, it appears that most such regressions are serially correlated. Easterly *et al.* (1993) and Pritchett (2000) emphasize the volatility of growth in developing countries over the long term. But there is a persistent short-term component too—think of China, which has grown year after year, defying the expectations of many, or African countries that have not grown, year after year, defying the hopes of many. This persistence, more precisely our inability to explain all of it with conditioning variables, creates a serially correlated error component in many 4- and 5-year panel growth regressions. Below, I find a coefficient of correlation in the errors of 0.1–0.2 in aid-growth regressions.

In sum, worries about the endogeneity of aid to growth, dating back at least to Mosley (1980), are well-founded. Underappreciated though is the precise nature of this endogeneity and

¹⁰ Here, we elide the distinction between GDP and GDP/capita, which is what y really is. But shocks to the two are highly correlated.

how common controls can give it the appearance of a positive aid-to-growth link. An early example may be Mosley (1980), which finds that in the poorest developing countries, contemporary aid/GDP is insignificant in a growth regression—but is positive and significant once lagged.¹¹ Such findings can be robust since they arise from real statistical relationships.

2. Specification issues in the aid-growth literature

The simple tests reported here imply that the dominant aid-growth relationship is not a positive one running from aid/GDP to growth but a negative one going the other way. They do not completely rule out aid affecting growth but do show that any such effect is statistically weak. That places a heavy burden on instrumentation strategies. Invalid instruments will pick up the causation from growth to aid. Weak instruments will pick up about as much of that as they will the reverse. As we have seen, where instrumentation falls short, reverse causation can not only obscure positive results of interest to researchers but generate them.

This section reviews some common specification issues in the aid-growth literature that hamper researchers' ability to remove the reverse causation. Two, autocorrelation and instrument proliferation, go directly to the question of instrument validity in some studies. The third is multicollinearity, which under certain plausible assumptions increases the sensitivity of regressions to endogeneity.

2.1. Autocorrelation

Serial correlation in the errors is prevalent in the pooled OLS and 2SLS regressions in this literature. One reason it is rarely reported may be that extreme growth volatility in oil-rich countries such as Iran or conflict-ridden ones such as Sudan obscures the general pattern, foiling standard autocorrelation tests. Table 5 shows that one can greatly increase the power of an autocorrelation

¹¹ Likewise, the OLS regressions on lagged aid in the Clemens, Radelet, and Bhavnani (2004) working paper, like the HT regressions in text, exhibit a sign flip if forward aid is used instead.

test by applying the Hadi (1992) outlier-identification procedure to a scatter of estimation residuals against their own lags, then excluding them from the test sample. For instance, the Arellano-Bond (1991) test for first-order serial correlation returns a p value of 0.54 for Burnside and Dollar's (2000) preferred specification. But excluding nine outliers drops it to 0.15. (Researchers commonly dismiss such p values as above conventional thresholds, but conservatism argues for high thresholds on such tests, not low ones.)

Autocorrelation can cause endogeneity bias when, in attempts to eliminate endogeneity, variables of interest are lagged in OLS regressions (Hadjimichael et al. 1995) or instrumented with their own lags (HT and Clemens, Radelet, and Bhavnani (2004)). If lagged aid is endogenous to lagged growth and the lagged growth innovation is correlated with the contemporary one, then lagged aid can be correlated with it too.¹² To indicate the degree of autocorrelation, the table also reports serial correlation coefficients for the restricted samples, which should be taken as suggestive to the extent that the regressions on which they are based are biased by the autocorrelation itself.

2.2. Instrument proliferation

The “difference” and “system” GMM estimators (Arellano and Bond 1991; Arellano and Bover 1995; Blundell and Bond 1998) dominate the most recent aid-growth literature (Dalgaard, Hansen, and Tarp 2004; Bobba and Powell 2007; Dreher, Nunnenkamp, and Thiele 2006; Mishra and Newhouse 2007; Rajan and Subramanian forthcoming), in part because of the availability of the “xtabond2” program for Stata (Roodman 2006b). The instrumentation strategies used in many of these studies raise several concerns.

¹² Dropping the lags of aid and aid² from the instrument sets in the 2SLS regressions in HT and Clemens, Radelet, and Bhavnani (2004) eliminates the significance of the coefficients on the aid terms. So identification appears to depend on these suspect instruments.

As typically used, the estimators generate instruments whose number is quadratic in the time dimension of the study panel. Instrument proliferation creates the risk of overfitting of endogenous variables. It also vitiates the Hansen J test for joint instrument validity (Bowsher 2002; Roodman 2007b). Thus if instrument proliferation is not stopped, the estimators can be *expected* to generate results that are simultaneously invalid and sport the appearance of validity. In addition, the assumptions necessary for the validity of the individual instruments for the levels equation in system GMM, spelled out in Blundell and Bond (1998), are non-trivial. They can be checked with difference-in-Hansen tests, but rarely are, at least in specifications in which the instrument collection has been curtailed to protect the power of the tests.

For example, in Michaelowa and Weber (2006) the significance of the aid term appears to go hand-in-hand with the instrument count. Mishra and Newhouse (2007)'s system GMM regressions estimate the coefficient on the lagged dependent variable to be almost exactly 1.0, a value at which the system GMM instruments are invalid (Blundell and Bond 1998). Dalgaard, Hansen, and Tarp (2004) and Rajan and Subramanian (forthcoming) rely on GMM regressions in which instruments rival observations in number. Both report that shrinking the instrument collection does not affect their results. Still, they illustrate that problematic practices have been the norm.

2.3. Multicollinearity

Textbooks typically warn that multicollinearity inflates the variance of OLS (e.g., Greene 2000, p. 256). This is easy to understand: nearly identical variables are hard to distinguish. But regressions with collinearity can in fact return statistically strong results (Spanos and McGuirk 2002). The intuition comes from basic econometrics. By the Frisch-Waugh Theorem, only the unique variation in a regressor helps identify its OLS coefficient. If the regressor is nearly multicollinear

with other variables, that unique variation is small; if that unique variation is nevertheless strongly correlated with the dependent variable, OLS will report a significant t statistic for the coefficient on the variable. If one regresses daily revenue of a shoe store on left shoe and right shoe sales, and an extra right shoe is recorded as being sold on a day of strong revenue, OLS will put a significant positive coefficient on right-shoe sales and nearly the same coefficient and t statistic on left-shoe sales—but with opposite sign. In contrast, regressing on either variable alone will return little significance. Thus near-collinearity can enhance the power of regressions to detect relationships involving small subcomponents of variables, a sort of “corona effect,” analogous to the way the moon’s transit across the nearly collinear sun during an eclipse reveals the sun’s corona.

The principle of course carries over to more sophisticated estimators in the GMM class. So, at the least, multicollinearity in variables of interest means that identification of their coefficients is driven by small components of variation within them. These components may be associated with specific observations—outliers—or they may be diffuse. If diffuse, the possibility remains, as with outliers, that the variation spotlighted represents a structural break in the data-generating process, and is therefore unrepresentative of the full sample.

Near-multicollinearity appears common in the aid literature starting in the late 1990s. In the 2SLS HT regression referred to in section 1, I find that the variance inflation factor (VIF) for aid/GDP, after projecting the endogenous variables onto instruments, is 8.6: instrumented aid/GDP is correlated $\sqrt{1 - 1/8.6} = 0.94$ with its projection onto the other regressors.¹³ The main source of collinearity is $(\text{aid/GDP})^2$. This is because aid/GDP has largely positive support and is right-skewed, and because projecting the two aid variables onto a common set of instruments

¹³ The VIF is $1/(1 - R^2)$, where the R^2 is for a regression of the variable on all the other right-hand-side variables. In a regression with two variables on the right, it is the same for both.

reduces unique variation in them. Given the probable invalidity of the instruments in HT, based on lags of the instrumented variables, this near-multicollinearity may be helping to detect an endogenous relationship. Multicollinearity is most likely present in all aid-growth studies that include aid/GDP and $(\text{aid}/\text{GDP})^2$ (Hadjimichael *et al.* 1995; Durbarry, Gemmell, and Greenaway 1998; Lensink and White 2001; Hudson and Mosley 2001; Clemens, Radelet, and Bhavnani 2004; Collier and Dollar 2004; Ali and Isse 2005).

And the potential for multicollinearity is not confined to these studies of diminishing returns to aid. Since Burnside and Dollar (2000), it has become common to interact aid with other terms, split aid into subcomponents, or introduce new terms that are inherently correlated with aid/GDP. All of these practices can generate near-multicollinear pairs. Burnside and Dollar interact aid and aid^2 with policy. Dalgaard, Hansen, and Tarp (2004) interact aid with the fraction of a country's area that is in the tropics. Other studies split aid into bilateral flows (from national governments) and multilateral flows (from the World Bank and other international agencies) (Ram 2004); technical assistance and non-technical assistance (Chauvet and Collier 2006); project aid and program aid (Cordella and Dell'Araccia 2003; Ouattara and Stroble 2004); "productive" and "unproductive" aid (Neanidis and Varvarigos 2007); and aid that appears geopolitically motivated and aid that does not (Reddy and Minoui 2006; Bobba and Powell 2007). Still others have regressed simultaneously on aid/GDP and a measure of aid instability or unpredictability that tends to scale with aid/GDP (Lensink and Morrissey 2001; Neanidis and Varvarigos 2007). Frequently both variables in the constructed pairs are highly significant with opposite sign, just as in our Table 1. Some of the coefficient magnitudes stretch credulity. Few of the studies report testing the variables of interest individually; Hudson and Mosley (2001) and Bobba and Powell (2007) are exceptions. Figure 1 shows the recurring pattern. Scanning the figure, it is hard to es-

cape the suspicion that regressions on two collinear aid terms have an inherent propensity to generate seemingly strong results, which is amplified by systemic tendencies (however unconscious) toward data mining and publication bias.

The effects of multicollinearity are complex and context-dependent. This section offers two illustrations and a bit of theory about how it can interact with endogeneity. First, we turn to the Easterly, Levine, and Roodman (ELR, 2004) testing of Burnside and Dollar (2000). Table 6 is an expanded version of one in ELR, showing sample-expansion tests of the BD OLS regressions that include both $\text{aid/GDP} \times \text{policy}$ and $(\text{aid/GDP})^2 \times \text{policy}$. In the original BD regressions (first panel of Table 6) the two variables enter with opposite signs and similar t statistics. And they are nearly multicollinear, as indicated by the VIFs. But the coefficients jump around as the sample changes in subsequent panels. ELR survey these results and declare the BD results fragile. However, the numbers are not pure noise. The two variables *always* enter with opposite sign and similar t statistics. And the t statistics are generally too strong to be caused by chance alone. But the sign on $\text{aid/GDP} \times \text{policy}$ is not stable. If one eliminates the collinearity by dropping $(\text{aid/GDP})^2 \times \text{policy}$, the large t statistics disappear throughout (results not shown). Thus the collinearity seems to make the estimator's true variance large, as the textbooks predicts, but making the reported variance small, as Spanos and McGuirk (2002) show is possible.

The second example is a regression in Dalgaard, Hansen, and Tarp (DHT, 2004). To the ELR dataset, DHT adduce a variable that is the fraction of a country's area that is in the tropics. For most countries, it is 0 or 1. Tropical location surfaces as a growth determinant in Gallup and Sachs (1999) and Sachs (2001, 2003). The causal channels may involve institutions and economic policies (Acemoglu, Johnson, and Robinson 2001; Easterly and Levine 2003). In DHT's growth regressions, aid and $\text{aid} \times \text{tropical area fraction}$ are quite significant, with nearly equal and

opposite signs, so that the derivative of growth with respect to aid *in* the tropics is not statistically different from 0. On average, aid seems to work outside the tropics but not in them.¹⁴

To give a sense of the data behind this result, Figure 2 shows the scatter plot of aid/GDP and aid/GDP×tropical area fraction in the DHT sample. For countries fully within the tropics, the two variables coincide. And most countries outside the tropics have received relatively little aid for their economic size so aid/GDP and aid/GDP×tropical area fraction are both small. As a consequence, the identification of the aid-tropics interaction derives mainly from the handful of non-tropical, heavily aided countries. Jordan stands out on the graph, but Egypt and Syria, and, to a lesser extent, Botswana, help make up the corona too. (Thirty percent of Botswana lies outside the tropics.) Table 7 shows that dropping either aid variable from a representative DHT regression greatly weakens the coefficient on the other. Leaving both in but excluding the four main corona countries from the sample eliminates the result.¹⁵

Aid may well have raised growth in the main corona countries. But other stories are plausible too. In the oil boom years of the 1970s and 1980s, Jordan received 5–15% of GDP in aid, mostly from other Middle Eastern nations, which could have provided short-term stimulus by bringing excess production factors into play. But regional economic linkages probably also figured in the temporary growth spurt. Jordan’s Port of Aqaba is an important transshipment point for oil, for example. At any rate, such cases deserve close examination. But the purpose of

¹⁴ Roodman (2007a) tests the DHT system GMM regression for fragility. Unlike the other six test subjects, the DHT regression comes through almost all tests with flying colors. A deficiency in the working paper version of Roodman (2007a), however, is that it does not subject GMM regressions (those of DHT and Hansen and Tarp 2001) to removal of outliers. The hindrance was that the partial scatter, the basis for the outlier search in that paper, is not well defined for GMM. The published version overcomes this hindrance by running a parallel 2SLS regression, identifying outliers from it in the usual way, then re-running the GMM regression excluding these outliers.

¹⁵ The original is regression 5 of their Table 3, using system GMM. The reproduction here in column 1 of Table 7 differs from the original in sample size. The original was executed with the DPD for Ox package (Doornik, Arellano, and Bond 2002). It turns out that an assumption built into this software—incomplete observations that create gaps in the time series must always be included in the data file rather than deleted—led the program to break time series for countries with gaps into two or more separate series. The xtabond2 module for Stata (Roodman 2006b), used here, does not behave this way.

econometric analysis is to abstract from idiosyncrasies, searching for patterns in large numbers of observations. Multicollinearity is a sign that the opposite may be happening, that strong idiosyncrasies are being mistaken for general patterns.

Beyond this general concern, under plausible circumstances multicollinearity can be shown to magnify endogeneity bias—in theory and in simulations. Separate research is developing this analysis more fully. Suppose aid/GDP has a small endogenous component. In a regression on aid/GDP and typical controls, endogeneity could be technically present but a minor concern. But adding a second, nearly collinear, term, such as $(\text{aid/GDP})^2$, greatly shrinks the unique variation in aid/GDP, and can do so such that the endogenous component becomes large within it. In particular, if aid/GDP consists of a large lognormal component unrelated to growth and small (endogenous) one *linearly* related to it, adding $(\text{aid/GDP})^2$ to a simple regression of growth on aid/GDP allows a much better fit, in which the endogenous component moves to the fore within the corona. Adding $(\text{aid/GDP})^2$ works out to multiply the reported coefficient and variance on aid/GDP by approximately the VIF. This inflates the t statistic on aid/GDP by about $\sqrt{\text{VIF}}$. Thus, adding the quadratic term increases the power to detect a real relationship, albeit linear and endogenous. In this way, multicollinearity can interact with the demonstrated endogeneity in the aid-growth relationship to produce more dramatic, yet invalid, results.

3. Conclusion

There appear to be almost no findings in the contemporary literature that a) find a significant effect of aid on growth, b) are robust, and c) are free of the methodological problems described here. The technical problems pointed out here may admit technical solutions, such as replacing aid/GDP with aid/capita. But given the prevalence of specification problems and fragility in this literature, some may be surface manifestations of deeper, less mutable problems. Prevalent auto-

correlation may be a sign of our continuing inability to fully explain cross-country differences in growth—why China keeps growing fast and Niger keeps not growing fast. That in turn hampers our ability to exploit within-country variation by, roughly speaking, treating successive periods in a country as independent experiments. Multicollinearity arises out of attempts to address statistically second-order but practical questions of which kinds of aid raise growth when and where. But if aid is a minor ingredient in the complex processes of economic growth, if it is often ineffective, and if the available datasets are small and marred by measurement error, even answering the first-order question of whether it works on average is challenging.

In reviewing the 50-year history of this literature, McGillivray *et al.* (2005) identify a structural break in the late 1990s, with Burnside and Dollar (2000). Before then, there is no consensus about the first-order effects of aid. After, “practically all studies...conclude that aid is associated with higher growth, one way or another.” What happened? The biggest change was the increase in econometric sophistication, moving from cross-sections to panels; from OLS to 2SLS and GMM; and from a single, linear aid regressor to interaction terms and aid subcomponents. The present analysis suggests that this sophistication has backfired, clouding rather than sharpening our vision. The implications for the practice of econometrics generally are worth pondering.

This is almost certainly not the end of the econometric quest for evidence on aid effectiveness. The future of the research appears to lie in asking smaller questions, and perhaps in performing experiments. Chen, Mu, and Ravallion (2006), for example, study how much the placement of World Bank–financed rural development projects in China can explain sub-national variation in household income ten years later. Whether the new cycle of investigation will bear more fruit than the last remains to be seen. But clear, reassuring evidence of the overall effect of aid on growth does not appear forthcoming.

Table 1. Varying lag distance and direction in OLS regressions on Hansen and Tarp (2001) exogenous variables

Variable	Observations for aid/GDP		
	Lagged	Current	Forward
No controls			
Net ODA/GDP	-0.066 (0.067)	-0.109 (0.067)	-0.178 (0.059)***
(Net ODA/GDP) ²	-0.014 (0.006)**	-0.003 (0.002)*	0.003 (0.002)*
Observations	278	332	276
Full controls			
Net ODA/GDP	0.234 (0.107)**	0.086 (0.070)	-0.102 (0.065)
(Net ODA/GDP) ²	-0.014 (0.006)**	-0.003 (0.002)*	0.003 (0.002)*
Observations	230	285	227

Standard errors clustered by country in parenthesis.

*significant at 10%; **significant at 5%.

Table 2. Regressions of GDP/capita growth on investment/GDP

	Observations for investment/GDP		
	Lagged	Current	Forward
No controls			
Investment/GDP	0.30 (0.09)***	0.60 (0.11)***	0.80 (0.09)***
Observations	404	404	404
Controlling for fixed effects			
Investment/GDP	-1.00 (0.25)***	-0.01 (0.25)	1.65 (0.24)***
Observations	404	404	404

Classical standard errors in parenthesis. Source: Blomström, Lipsey, and Zejan (1996). ***significant at 1%.

Table 3. Coefficient on Net ODA/GDP when varying lag distance and controls in OLS growth regressions

Aid variable	Twice-lagged	Lagged	Current	Forward	Twice-forward
No controls					
Δ Net ODA/GDP	0.001 (0.027)	-0.029 (0.025)	-0.065 (0.027)**	-0.083 (0.033)**	-0.024 (0.026)
Observations	780	902	1015	973	917
Net ODA/GDP	0.003 (0.022)	-0.006 (0.014)	-0.007 (0.016)	-0.030 (0.012)**	-0.051 (0.013)***
Observations	907	1020	1130	1075	934
Controlling for country fixed effects					
Net ODA/GDP	0.031 (0.027)	0.032 (0.019)	0.007 (0.033)	-0.058 (0.029)**	-0.067 (0.033)**
Observations	907	1020	1130	1075	934
Controlling for initial GDP/capita					
Net ODA/GDP	0.011 (0.019)	0.008 (0.019)	-0.006 (0.018)	-0.042 (0.019)**	-0.072 (0.018)***
Observations	825	914	997	928	817

Standard errors clustered by country in parenthesis. **significant at 5%; ***significant at 1%.

Table 4. Granger and Sims causality tests, GDP/capita growth, Net ODA

Dependent variable:	Granger tests		Sims tests	
	Net ODA/GDP		Net ODA/GDP	
	Growth	Aid	Growth	Aid
GDP/capita growth, forward 2				0.053 (0.056)
GDP/capita growth, forward 1				0.026 (0.047)
GDP/capita growth, current				-0.079 (0.061)
GDP/capita growth, lag 1	0.281 (0.064)***	-0.152 (0.036)***	0.192 (0.051)***	-0.131 (0.048)***
GDP/capita growth, lag 2	0.019 (0.034)	-0.042 (0.040)	0.085 (0.042)**	-0.090 (0.065)
Aid, forward 2				-0.056 (0.036)
Aid, forward 1				-0.048 (0.065)
Aid, current				0.038 (0.047)
Aid, lag 1	-0.015 (0.022)	0.907 (0.065)***	0.011 (0.052)	1.001 (0.181)***
Aid, lag 2	0.017 (0.031)	-0.128 (0.098)	0.011 (0.041)	-0.061 (0.174)
Observations	897	893	623	628
R^2	0.08	0.74	0.10	0.71
F test for causal effect on dependent variable (p value)	0.784	0.000	0.020	0.485

Standard errors clustered by country in parenthesis. Constant term not reported.

Coefficients of particular interest in bold. **significant at 5%; *** significant at 1%.

Table 5. Examples of serial correlation in the errors of aid-growth regressions

Study	Representative regression	Description	Arellano-Bond test		Serial correlation in errors, excluding residual-lagged residual outliers	
			Full sample (p value)	Residual-lagged residual outliers		Excluding outliers (p value)
Burnside & Dollar (2000)	5/OLS	All low- and middle-income countries; excluding 5 outliers; no aid ² xpolicy	0.54	Cameroon 1978–81, 1982–85, 1990–93; Ethiopia 1986–89; Gabon 1974–77, 1978–81, 1982–85; Nicaragua 1978–81, 1982–85	0.15	0.11
Hansen & Tarp (2001)	1.2	2SLS; includes aid ² but not aidxpolicy	0.74	Cameroon 1982–85; Ethiopia 1986–89; Gabon 1978–81, 1982–85; Nicaragua 1978–81, 1982–85	0.02	0.20
Collier & Dollar (2002)	Table 1, col. 2	OLS; includes aid ² and aidxpolicy but not aid; as revised in Collier & Hoeffler (2004)	0.62	Gabon 1978–81; Togo 1994–97; Sudan 1994–97	0.09	0.10
Clemens, Radelet, & Bhavnani (2004)	Table 4, col. 5	2SLS; excluding 4 outliers; full control set	0.03	Iran 1982–85; Togo 1994–97	0.002	0.19

All tests based on reproduction regressions on original data sets. Reproduction is exact for Burnside & Dollar and close for the others. The Arellano-Bond (1991) test for autocorrelation is run on the full sample for each regressions, and after excluding residuals in a scatter plot of the residuals against their own lags, identified using the Hadi (1992) procedure and a threshold p value of 0.1. Reported serial correlations are biased if original regressions are invalidated by that correlation.

Figure 1. Collage of aid-growth results consistent with multicollinearity

Aid/GDP*Policy Index	-	0.16 **	
	-	(0.06)	
(Aid/GDP) ² *Policy Index	-	-0.01 **	
	-	(0.00)	
FAIDOECD	0.17604	0.08021	2.195**
FAIDOECDSQ	-0.001963	0.00095	-2.064**
AIDGNP	0.0758 (3.22)	0.0775 (3.45)	0.108 (2.16)
			0.1466 (2.98)
AIDGNP2		-0.0005 (-0.79)	-0.0013 (-1.82)
Aid/GNP	0.541 (2.18)	1.155 (5.99)	
(Aid/GNP) ²	-0.0386 (2.83)	-0.0679 (5.69)	
(ODA/GDP) ²	-0.02 (1.60)	-0.04 (3.07)	
CPIA×ODA/GDP	0.31 (2.94)	0.18 (3.06)	
Aid	0.60** (4.70)	0.34** (3.27)	
Aid x fract. of land in tropics	-0.77** (3.10)	-0.48** (2.40)	
Aid/GDP	-0.661 (-1.321)	0.1561 (1.6743)	
Aid ²		-0.540 (-2.914)	
PG_AID		-0.554*** (0.200)	-1.198* (0.686)
PJ_AID		0.180** (0.080)	0.193 (0.207)
Lagged MA/GDP	49.0541*** [17.2589]	58.9424** [22.2058]	57.2925*** [17.4080]
			47.1262*** [13.3617]
RS predicted GA/GDP	-5.8826 [4.1722]	-4.8461 [4.2949]	-7.3931** [3.6465]
			-8.6034** [3.6930]
(ODA-TA)/GDP, years 1 to 4		-0.425 (2.88)***	-0.594 (2.88)***
TA/GDP, years 1 to 4		1.902 (2.44)**	2.427 (3.30)***
Aid * Productive expenditures	0.015 (0.000)	0.022 (0.000)	
Aid * Unproductive expenditures	-0.018 (0.051)	-0.036 (0.003)	

Sources, in order: Durbarry, Gemmill, and Greenaway 1998; Lensink and White 2001; Hudson and Mosley 2001; Collier and Dollar 2004; Dalggaard, Hansen, and Tarp 2004; Ali and Isse 2005; Ouattara and Strobl 2004; Reddy and Minoui 2006; Chauvet and Collier 2006; Neanidis and Varvarigos 2007.

Table 6. Easterly, Levine, and Roodman testing of Burnside & Dollar OLS regressions that include $\text{aid}^2 \times \text{policy}$

Sample		Low- and middle-income countries	Low-income countries only
Burnside and Dollar original	Aid/GDP \times policy	0.20 (2.07)**	0.27 (2.03)**
	(Aid/GDP) ² \times policy	-0.02 (2.22)**	-0.02 (2.45)**
	Observations	275	189
	VIF of aid/GDP \times policy	13.1	20.5
	VIF of (Aid/GDP) ² \times policy	9.7	12.8
ELR data, BD countries, 1970–93	Aid/GDP \times policy	0.31 (2.30)**	0.28 (1.81)*
	(Aid/GDP) ² \times policy	-0.05 (2.35)**	-0.05 (2.41)**
	Observations	274	183
	VIF of aid/GDP \times policy	10.2	15.1
	VIF of (Aid/GDP) ² \times policy	6.9	8.3
ELR data, full sample, 1970–93	Aid/GDP \times policy	-0.11 (1.10)	-0.27 (1.94)*
	(Aid/GDP) ² \times policy	0.02 (1.92)*	0.03 (2.34)**
	Observations	300	205
	VIF of aid/GDP \times policy	9.0	13.8
	VIF of (Aid/GDP) ² \times policy	6.0	7.4
ELR data, BD countries, 1970–97	Aid/GDP \times policy	0.19 (1.64)	0.15 (1.11)
	(Aid/GDP) ² \times policy	-0.03 (1.58)	-0.03 (1.56)
	Observations	322	216
	VIF of aid/GDP \times policy	10.0	14.3
	VIF of (Aid/GDP) ² \times policy	6.8	7.8
ELR data, full sample, 1970–97	Aid/GDP \times policy	-0.14 (1.31)	-0.27 (1.89)*
	(Aid/GDP) ² \times policy	0.03 (2.25)**	0.03 (2.35)**
	Observations	356	244
	VIF of aid/GDP \times policy	7.9	11.7
	VIF of (Aid/GDP) ² \times policy	5.2	6.1

Other regressors not reported. Heteroskedasticity-robust t statistics in parenthesis. *significant at 10%; **significant at 5%.

Figure 2. Aid/GDP vs. Aid/GDP×tropical area fraction, Dalgaard, Hansen, and Tarp (2004) dataset

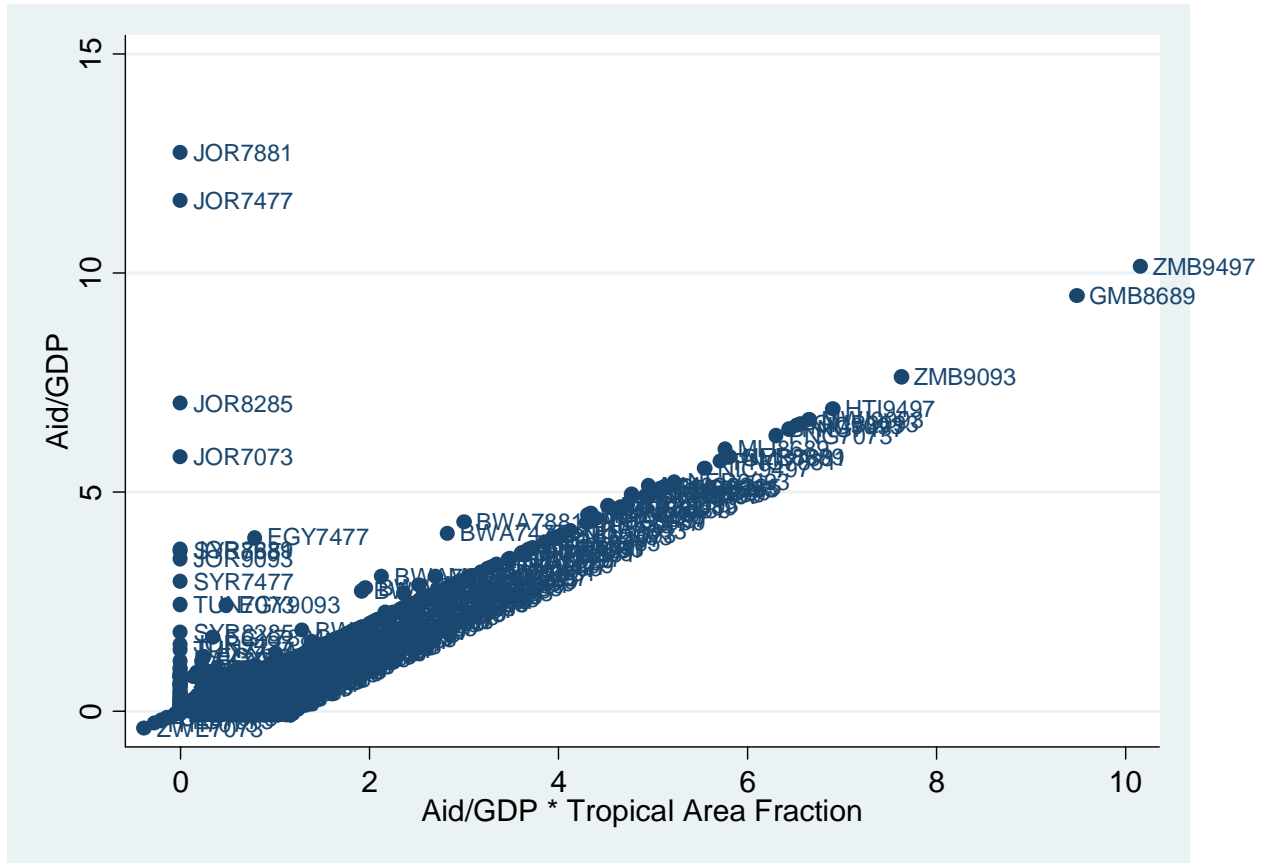


Table 7. Tests of regression 5, Table 3, of Dalgaard, Hansen, and Tarp (2004)

Dependent variable: GDP/capita growth

Aid/GDP	0.70 (5.33)***	0.25 (1.17)	0.40 (0.18)	
Aid/GDP × tropical area fraction	-0.97 (3.12)***		-0.23 (1.10)	-0.78 (0.34)
Log initial GDP/capita (PPP)	0.39 (0.59)	1.18 (2.91)***	0.45 (0.84)	0.22 (0.42)
Including Botswana, Egypt, Jordan, Syria?	Yes	Yes	Yes	No
Observations	332	332	332	309

The first column is close but not exact reproduction of the original. All regressions are "system GMM." For clarity, only regressors of interest are reported. Robust *t* or statistics clustered by country in parenthesis.

***significant at 1%.

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.
- Ali, A.M., and H.S. Isse. 2005. "An Empirical Analysis of the Effect of Aid on Growth." *International Advances in Economic Research* 11: 1–11.
- Arellano, M., and S. Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *Review of Economic Studies* 58: 277–97.
- Arellano, M., and O. Bover. 1995. "Another look at the instrumental variables estimation of error components models." *Journal of Econometrics* 68: 29–51.
- Arvin, B.M., and F. Barillas. 2002. "Foreign Aid, Poverty Reduction, and Democracy." *Applied Economics* 34: 2151–56.
- Blomström, M., R.E. Lipsey, and M. Zejan. 1993. "Is Fixed Investment the Key to Economic Growth?" Working Paper 4436. National Bureau of Economic Research. Cambridge, MA. August.
- Blomström, M., R.E. Lipsey, and M. Zejan. 1996. "Is Fixed Investment the Key to Economic Growth?" *Quarterly Journal of Economics* 111(1): 269–76.
- Blundell, R., and S. Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87: 11–143.
- Bobba, M., and A. Powell. 2007. "Aid Effectiveness: Politics Matters." Working Paper 601. Inter-American Development Bank. Washington, DC. January.
- Bowsher, C.G. 2002. "On testing overidentifying restrictions in dynamic panel data models." *Economics Letters* 77: 211–20.
- Burnside, C. and D. Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90(4): 847–68.
- Chauvet, L. and P. Collier, 2006. "Helping Hand? Aid to Failing States," Working Paper DT/2006/14, Développement, Institutions & Analyses de Long terme, Paris.
- Chen, S., R. Mu, and M. Ravallion. 2006. "Are there Lasting Impacts of Aid to Poor Areas? Evidence from Rural China." Policy Research Working Paper 4084. World Bank. Washington, DC.
- Clemens, M., S. Radelet and R. Bhavnani. 2004. "Counting Chickens When They Hatch: The Short-Term Effect of Aid on Growth." Working Paper 44. Center for Global Development. Washington, DC.

- Collier, P., and D. Dollar. 2002. "Aid Allocation and Poverty Reduction." *European Economic Review* 45(1): 1–26.
- Collier, P., and A. Hoeffler. 2004. "Aid, Policy and Growth in Post-Conflict Societies." *European Economic Review* 48 (5): 1125-45.
- Cordella, T., and G. Dell’Ariccia. 2003. "Budget Support versus Project Aid." Working Paper 03/88. International Monetary Fund. Washington, DC. April.
- Dalgaard, C., Hansen, H. and Tarp, F. 2004. 'On the Empirics of Foreign Aid and Growth', *The Economic Journal*, vol. 114(496) (June), pp. F191–F216.
- De Long, J.B., and L.H. Summers. 1991. "Equipment Investment and Economic Growth." *Quarterly Journal of Economics* 106(2): 445–502.
- Dreher, A., P. Nunnenkamp, and R. Thiele. 2006. "Does Aid for Education Educate Children? Evidence from Panel Data." Working Paper. Kiel Institute for the World Economy. Kiel, Germany. August.
- 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, School of Economics, Centre for Research in Economic Development and International Trade, University Park, Nottingham, United Kingdom.
- Easterly, W., M. Kremer, L. Pritchett, and L.H. Summers. 1993. "Good Policy or Good Luck?: Country Growth Performance and Temporary Shocks." *Journal of Monetary Economics* 32: 459–83.
- 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).
- 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.
- Geweke, J., R. Meese, and W. Dent. 1983. "Comparing alternative tests of causality in temporal systems: analytic results and experimental evidence." *Journal of Econometrics* 21: 161–94.
- Granger, C.W.J. 1969. "Investigating causal relation by econometric and cross-sectional methods." *Econometrica* 37: 424–438.
- Greene, W.H. 2000. *Econometric Analysis*. Prentice Hall: Upper Saddle River, NJ.

- 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, D.C.
- 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.
- Hudson, J., and P. Mosley. 2001. "Aid Policies and Growth: In Search of the Holy Grail." *Journal of International Development* 13(7): 1023-38.
- Lensink, R., and O. Morrissey. 2000. "Aid Instability as a Measure of Uncertainty and the Positive Impact of Aid on Growth," *Journal of Development Studies* 36 (3): 31-49.
- Lensink, R., and H. White. 2001. "Are There Negative Returns to Aid?" *Journal of Development Studies* 37 (6): 42-65.
- Ouattara, B., and E. Strobl. 2004. "Disaggregating the Aid and Growth Relationship." Economics Discussion Paper. University of Manchester.
- McGillivray, M., S. Feeney, N. Hermes, and R. Lensink. 2005. "It Works; It Doesn't; It Can, But That Depends...: 50 Years of Controversy over the Macroeconomic Impact of Development Aid," Working Paper 2005/24, World Institute for Development Economics Research, Helsinki.
- Michaelowa, K., and A. Weber. 2006. "Aid Effectiveness Reconsidered: Panel Data Evidence for the Education Sector." Discussion Paper 264. Hamburg Institute of International Economics. Hamburg.
- Mishra, P., and D. Newhouse. 2007. "Health Aid and Infant Mortality," Working Paper 07/100, International Monetary Fund, Washington, DC. April.
- Mosley, P. 1980. "Aid, Savings, and Growth Revisited," *Oxford Bulletin of Economics and Statistics* 42(2): 79-95.
- 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.
- Neanidis, K.C., and D. Varvarigos. 2007. "The Allocation of Volatile Aid and Economic Growth: Evidence and a Suggestive Theory," Discussion Paper WP 2007-07, Department of Economics, Loughborough University, Loughborough, UK.

Roodman, Through the Looking-Glass, and What OLS Found There

- Pritchett, L. 2000. "Understanding Patterns of Economic Growth: Searching for Hills among Plateaus, Mountains, and Plains." *World Bank Economic Review* 14(2): 221–50.
- Rajan, R., and A. Subramanian. 2005. "Aid and Growth: What Does the Cross-Country Evidence Really Show?" *Review of Economics and Statistics*, forthcoming.
- 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.
- Reddy, S.G., and C. Minoiu. 2006. "Development Aid and Economic Growth: A Positive Long-Run Relation." Department of Economic and Social Affairs, United Nations, New York, NY. September.
- Rodrik, D. 1999. *The New Global Economy and Developing Countries: Making Openness Work*. Policy Essay 24, Overseas Development Council, Washington, DC.
- Roodman, D. 2006a. "An Index of Donor Performance." Working Paper 67, revised. Center for Global Development, Washington, DC.
- Roodman, D. 2006b. "How to Do xtabond2: An Introduction to 'Difference' and 'System' GMM in Stata." Working Paper 103. Center for Global Development, Washington, DC.
- Roodman, D. 2007a. "The Anarchy of Numbers: Aid, Development, and Cross-country Empirics." *World Bank Economic Review* (forthcoming).
- Roodman, D. 2007b. "A Short Note on the Theme of Too Many Instruments." Working Paper 125. 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.
- Sims, C.A. 1972. "Money, income and causality." *American Economic Review* 62: 540–552.
- Spanos, A., and A. McGuirk. 2002. "The Problem of Near-multicollinearity Revisited: Erratic vs Systemic Volatility," *Journal of Econometrics* 108: 365–93.

Appendix. Formal demonstration of sign flip for two cases

This appendix defines two models in which unconditional and conditional covariances between two variables, one Granger-causing the other, can have opposite signs. The two correspond to the regressions in section 1.2 on aid/GDP or investment/GDP that control for country dummies or initial GDP/capita. In the first model, the controls consist of fixed effect dummies and all innovations in the variables are i.i.d. In the second, the control set is the analog of initial GDP/capita, and growth innovations are autocorrelated in a simple way.

Controlling for fixed-effect dummies, i.i.d. innovations

The two variables of interest are x and y , both $I(1)$ for each individual in a panel. y Granger-causes x , and with unambiguous sign, while x does not cause y . Yet Δy is regressed on x . The model is:

$$\begin{aligned} \Delta y_{it} &= u_{it} \\ \Delta x_{it} &= \phi(L)\Delta y_{it} + v_{it} \\ u_{it}, v_{it} &\sim i.i.d. N(0,1) \\ E[u_{it}x_{i0}] &= 0 \end{aligned} \tag{1}$$

where i and t index observations in the panel $N \times T$ panel, L is the lag operator, and $\phi(L) = \phi_0 + \phi_1 L + \phi_2 L^2 \dots$ is a polynomial whose non-zero coefficients all have the same sign and are absolutely summable. The variables are taken to be well-defined for all integers t but only observed for $t \in [1, T]$. Since each Δx_{it} is an unambiguously signed linear combination of past Δy , so is the “long difference” of each x_{it} with respect to x_{i0} for $t > 0$. So for each $t > 0$, we can write

$$x_{it} = x_{i0} + \psi^t(L)\Delta y_{it} + w_{it} \tag{2}$$

where $\psi^t(L)$ is a polynomial that varies with t whose non-zero coefficients are all of the same sign, and where $E[u_{it}w_{js}] = 0$ for $i \neq j$ or $t \neq s$. In particular,

$$\psi_j^t = \sum_{k=0}^{\min(j,t-1)} \phi_{j-k}. \quad (3)$$

Given a lag distance l , we are interested in the behavior of the OLS estimator in the large- N limit when regressing Δy_{it} on $x_{i,t-l}$, controlling for fixed effect dummies. In particular we are interested in the sign of the estimated coefficient on $x_{i,t-l}$. l can be positive, zero, or negative.

The sign of the coefficient of interest is the same as that of the covariance between Δy_{it} and $x_{i,t-l}$ after partialling out the dummies from both the variables or, equivalently, one of the them. Partialling them out demeans the variables, where the means are individual-specific. So let $\overline{\Delta y_i}$ denote the average of Δy_{it} over $[1, T]$ for any given i . We study the probability limit of the empirical covariance of $x_{i,t-l}$ and $\Delta y_{it} - \overline{\Delta y_i}$:

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \text{C}\hat{\text{ov}}[x_{i,t-l}, \Delta y_{it} - \overline{\Delta y_i}] &= \text{plim}_{N \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N x_{i,t-l} (\Delta y_{it} - \overline{\Delta y_i}) \\ &= \frac{1}{T} \sum_{t=1}^T \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left(x_{i,t-l} \Delta y_{it} - x_{i,t-l} \frac{1}{T} \sum_{s=1}^T \Delta y_{is} \right). \end{aligned}$$

Substituting with (1) and (2) makes this

$$\frac{1}{T} \sum_{t=1}^T \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left((x_{i0} + \psi^{t-l}(L)u_{i,t-l} + w_{i,t-l})u_{it} - (x_{i0} + \psi^{t-l}(L)u_{i,t-l} + w_{i,t-l}) \frac{1}{T} \sum_{s=1}^T u_{is} \right).$$

By assumption (and a Law of Large Numbers) the x_{i0} and the $w_{i,t-l}$ fall out in the limit since they are orthogonal to the u_{it} , giving

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^T \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left(\psi^{t-l}(L)u_{i,t-l}u_{it} - \psi^{t-l}(L)u_{i,t-l} \frac{1}{T} \sum_{s=1}^T u_{is} \right) \\ &= \frac{1}{T} \sum_{t=1}^T \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left(\sum_{r=0}^{\infty} \psi_r^{t-l} u_{i,t-l-r} u_{it} - \sum_{r=0}^{\infty} \psi_r^{t-l} u_{i,t-l-r} \frac{1}{T} \sum_{s=1}^T u_{is} \right) \end{aligned}$$

And since the u_{it} are spherical with unit variance, this simplifies in the limit to

$$\frac{1}{T} \sum_{t=1}^T \left(\psi_{-l}^{t-l} - \frac{1}{T} \sum_{r=1}^{\min(T, t-l)} \psi_{t-l-r}^{t-l} \right) \quad (4)$$

Here, ψ_{-l}^{t-l} is understood to equal 0 if $-l < 0$, i.e., $l > 0$. Indeed, in this case—if we are regressing Δy on lagged observations of x —the covariance is just

$$-\frac{1}{T^2} \sum_{t=1}^T \sum_{r=1}^{\min(T, t-l)} \psi_{t-l-r}^{t-l},$$

which is non-zero and has the opposite sign from the coefficients of $\psi(L)$. This is the sign flip.

Because our practical interest is in regressions where the variable of interest is lagged, the case of $l \leq 0$ —regression on forward observations of x —is less important. Nevertheless, it is worth investigating briefly. Substituting (3) into (4) leads after some algebra to

$$\text{plim}_{N \rightarrow \infty} \text{C}\hat{\text{ov}}[x_{i, t+|l|}, \Delta y_{it} - \overline{\Delta y}_i] = \sum_{k=0}^{|l|} \phi_k - \frac{1}{T^2} \sum_{t=1}^T \sum_{r=1}^{\min(T, t+|l|)} \sum_{k=0}^{t+|l|-r} \phi_k \quad (5)$$

Since the second term contains $\phi_{|l|+1}, \phi_{|l|+2}, \dots$ while the first one does not, if these coefficients are large enough, the whole expression can have sign opposite that of the ϕ_i , which is not the result we are after. But it does not seem very realistic for the coefficients on the deep lags to dominate. A realistic example in which this does not happen comes if we suppose that $k \geq |l| \Rightarrow \phi_k = 0$. That is, in the context of a regression of Δy_{it} on $x_{i, t+|l|}$, none of the innovations in y before t affect $x_{i, t+|l|}$. Then (5) works out to

$$\begin{aligned}
 \text{plim}_{N \rightarrow \infty} \text{C}\hat{\text{ov}}[x_{i,t+|l|}, \Delta y_{it} - \overline{\Delta y_i}] &= \sum_{k=0}^{|l|} \phi_k - \frac{1}{T^2} \left((1+|l|) + (1+|l|+1) + \dots + T + \underbrace{T + \dots + T}_{|l| \text{ times}} \right) \phi_0 \\
 &\quad - \frac{1}{T^2} \left((1+|l|-1) + (1+|l|) + \dots + T + \underbrace{T + \dots + T}_{|l|-1 \text{ times}} \right) \phi_1 \\
 &\quad - \dots - \frac{1}{T^2} (1+2+\dots+T) \phi_{|l|} \\
 &= \frac{1}{T^2} \sum_{k=0}^{|l|} \frac{(T+l+k)(T+l+k-1)}{2} \phi_k
 \end{aligned}$$

The coefficients on the ϕ_k are all positive, so we conclude that in this case a fixed-effects regression of Δy_{it} on $x_{i,t+|l|}$ will report the correct sign on the relationship between y and x .

Controlling for previous y , MA(1) innovations

In this model, we control for the “initial level” of y instead of fixed effect dummies. This time, for reasons motivated in section 1.3, we assume the innovations in y are positively serially correlated and in the simplest possible way, with an MA(1) structure. We also simplify $\phi(L)$ from before to a single constant (0^{th} -order polynomial) to keep the math manageable. And we require that $x_{i0} = \phi y_{i0} + w_i$ where w_i is a random term independent of everything else in the model. So the model can be written:

$$\begin{aligned}
 \Delta y_{it} &= \alpha_0 u_{it} + \alpha_1 u_{i,t-1} \\
 x_{it} &= \phi y_{it} + v_{it} \\
 u_{it}, v_{it} &\sim \text{i.i.d. } N(0,1) \\
 \alpha_0, \alpha_1 &> 0 \\
 \text{E}[y_{i0}] &= 0 \\
 \text{E}[u_{it} y_{i0}] &= 0
 \end{aligned} \tag{6}$$

The assumption that $\text{E}[y_{i0}] = 0$ (where the expectation is not conditional on i) merely allows us to exclude the constant term from the OLS regression we study.

Since this model does not distinguish between period-average and period-initial values, the analog to “period-initial” y is $y_{i,t-1}$. So that is what we control for. We look for a sign flip when regressing on x observed before $y_{i,t-1}$, i.e., on $x_{i,t-l}$ where $l \geq 2$. The covariance of interest is between $x_{i,t-l}$ and Δy_{it} after partialling out $y_{i,t-1}$:

$$\begin{aligned} & \text{plim}_{N \rightarrow \infty} \text{C}\hat{\text{ov}} \left[x_{i,t-l}, \Delta y_{it} - \frac{\text{C}\hat{\text{ov}}[\Delta y_{it}, y_{i,t-1}]}{\text{V}\hat{\text{ar}}[y_{i,t-1}]} y_{i,t-1} \right] \\ &= \text{plim}_{N \rightarrow \infty} \text{C}\hat{\text{ov}}[x_{i,t-l}, \Delta y_{it}] - \text{plim}_{N \rightarrow \infty} \text{C}\hat{\text{ov}}[x_{i,t-l}, y_{i,t-1}] \frac{\text{C}\hat{\text{ov}}[\Delta y_{it}, y_{i,t-1}]}{\text{V}\hat{\text{ar}}[y_{i,t-1}]} \end{aligned} \quad (7)$$

Consider first the case of $l = 2$. The first term above is zero because Δy_{it} , as an MA(1) series, depends only on shocks in periods t and $t - 1$, while $x_{i,t-2}$ depends only on earlier shocks. In the second term, $\frac{\text{C}\hat{\text{ov}}[\Delta y_{it}, y_{i,t-1}]}{\text{V}\hat{\text{ar}}[y_{i,t-1}]}$ can be shown, by substituting with definitions in (6), to converge to a strictly positive value, while $\text{C}\hat{\text{ov}}[x_{i,t-2}, y_{i,t-1}]$ converges to a non-zero value with the same sign as ϕ . Overall then, as before, the sign of the conditional covariance between Δy and lagged x is opposite that of the unconditional relationship.

For completeness, now consider the case of $l = -1$ —regressing on forward observations of x . The first term of (7) works out to

$$\begin{aligned} & \text{plim}_{N \rightarrow \infty} \text{C}\hat{\text{ov}}[\phi y_{i,t+1}, \Delta y_{it}] \\ &= \phi \text{plim}_{N \rightarrow \infty} \text{C}\hat{\text{ov}}[\alpha_0 u_{i,t+1} + (\alpha_0 + \alpha_1)(u_{i,t} + u_{i,t-1} + \dots + u_{i1}) + \alpha_1 u_{i0} + y_{i0}, \alpha_0 u_{it} + \alpha_1 u_{i,t-1}] \\ &= \phi(\alpha_0 + \alpha_1)^2. \end{aligned}$$

Similarly, the second term is $\text{plim}_{N \rightarrow \infty} \widehat{\text{Var}}[\phi y_{i,t-1}] \frac{\alpha_0 \alpha_1}{\widehat{\text{Var}}[y_{i,t-1}]} = \phi \alpha_0 \alpha_1$. Substituting into (7), the condi-

tional covariance is $\phi(\alpha_0 + \alpha_1)^2 - \phi \alpha_0 \alpha_1 = \phi(\alpha_0^2 + \alpha_0 \alpha_1 + \alpha_1^2)$, which has the same sign as ϕ , as

desired. The conditional and unconditional covariances of Δy_{it} and $x_{i,t+1}$ have the same sign.