A Comment on “Counting Chickens When They Hatch”

David Roodman[1]

July 14, 2013

Abstract

Clemens, Radelet, Bhavnani, and Bazzi (CRBB, 2012) argue that contradictions among the most-cited cross-country studies of the impact of foreign aid can be reconciled if certain methodological modifications are brought to all. The shared finding is then a positive correlation between receipts of foreign aid and subsequent growth in investment or GDP, which CRBB conclude is best interpreted as an impact of aid on growth. I reanalyze CRBB’s reanalyses and come to opposite conclusions. One CRBB modification, narrowing the aid variable to assistance most likely to affect growth in the short run, does not make much difference: “early-impact aid” does not have more early impact than “non-early impact aid.” Two other modifications—lagging the aid variables and differencing the data to remove fixed effects—are together done in a way that departs from standard practice and falls short of CRBB’s aim to temporally separate aid and growth observations. This leaves the aid variables vulnerable to contemporaneous endogeneity. Addressing this issue, as well as serial correlation, produces evidence of zero or negative Granger causation from aid/GDP to growth. And even if the CRBB regressions are taken at face value, few of them peg the impact of aid at a level statistically different from zero.

[1] I thank Michael Clemens and Samuel Bazzi for discussions and Aart Kraay, Mead Over, and Arvind Subramanian for comments. All views expressed are attributable only to the author.

Last year four men I have proudly called colleagues published a paper that unifies influential—and incompatible—studies of the impact of foreign aid (Clemens, Radelet, Bhavnani, and Bazzi, or CRBB, 2012). As CRBB document, the cross-country aid-growth literature has split over whether aid appears to generally "work" in the sense of expanding economic activity on average; or does not work; or works only under certain circumstances, such as good macroeconomic management. The controversy has attracted great interest among policy makers, and CRBB seem to offer resolution.

Anyone who reads CRBB should, I think, be impressed with its effort, discipline, scholarship, concision, and humility. CRBB build a powerful case that influential analyses reaching disparate conclusions will, after certain methodological modifications rooted in economic and econometric theory, converge to a similar result: more aid is followed by higher growth. "The most plausible explanation is that aid causes some degree of growth in recipient countries, though the magnitude of this relationship is modest, varies greatly across recipients, and diminishes at high levels of aid."

While I respect the care with which CRBB reach this view, I do not share it. Their delicate conclusion about the effect of aid is entirely plausible; so I differ on whether patterns in the CRBB data make that conclusion appear more certain than it otherwise would.

A core theme in CRBB is the need to factor into study design the likely *time profile* of aid's impacts. In replicating influential studies—Boone (1996), Burnside and Dollar (2000), and Rajan and Subramanian (2008)—CRBB modify the specifications in three ways that reflect their sensitivity to the time dimension. According to CRBB, these changes vivify the coefficient on aid and strengthen the interpretation of that coefficient as impact. I see problems in each step in their logic chain.

One step—narrowing the aid variables to forms of assistance, such as for roads, that can be expected to raise growth within a few years—actually does not make much difference. The "early-impact aid" that CRBB expect to cause more growth in the near term appears if anything to have *less* early impact than other aid. This cuts against CRBB's contention that some studies fail to detect the impact of aid on growth because they do not track the right subflow. And if CRBB's prior is correct that non–early impact aid does not quickly raise growth, the finding reported here that non-early aid is to a comparable degree soon followed by growth raises a question about whether the CRBB regressions are detecting causality from aid to growth, or some other causal relationship.

CRBB's second and third specification modifications are to "lag" aid in order to reduce its endogeneity and strengthen Granger-causal interpretation, and to difference the data to remove bias from omitted, persistent country-level variables (fixed effects). Both steps are prudent. But the implementation departs from standard practice in panel econometrics and by the same token falls short of CRBB's intent to temporally separate aid observations from growth observations. Loosely speaking, the change in growth from the late 1990s to the early 2000s is regressed on the change in early aid/GDP from the early 1990s to the late 1990s; so the timeframes of the two variables of interest overlap. Contemporaneous, endogenous causation cannot be ruled out after all. For instance, Indonesia's growth plunge in the late 1990s may have increased its aid/GDP in the period, thus the change in aid/GDP from the early 1990s to the late 1990s, even as the shock increased subsequent growth through a V-shaped recovery. Rising growth would then have followed rising aid/GDP, even without causality form aid to growth. In econometric language, differencing makes predetermined variables presumptively endogenous. Best practice is to replace or instrument such variables with still-deeper lags of themselves. Twice-lagging aid as a regressor—the solution most consistent with CRBB—points to an aid-growth association that is zero or mildly negative.

At any rate, even if the CRBB regressions are taken at face value, I find their results less compelling than CRBB do. The reasons appear to be two:

- For each of the three papers they replicate, CRBB run a variety of specifications that combat endogeneity more or less aggressively. CRBB look for consistent findings across specifications. In contrast, I focus on the most rigorous ones. In doing so, I argue, I impose CRBB's standards on CRBB's results.
- CRBB's preferred regressions model the impact of aid as parabolic. They synopsize the estimated coefficients on aid/GDP and $(aid/GDP)^2$ by reporting the aid level at which the aid-growth association peaks, not the magnitude of the association at that level. In simpler language laden with causal interpretation, they report the *aid level of maximum impact*, but not the *maximum impact*. The latter is much more relevant and much less often statistically significant.[2]

Followers of this literature know that Clemens, Radelet, Bhavnani, and Bazzi, like me, have taken a public stance on the ability of cross-country econometrics to reveal the impact of aid—my stance more pessimistic than theirs. Easterly, Levine, and Roodman (2004) and Roodman (2007) challenge much of the preceding literature as fragile. Roodman (2008) accepts that the results of some studies are more than statistical happenstance but raises specification issues that are argued to undercut interpretations of those results as causation from aid to growth. In the first version of CRBB, Clemens, Radelet, and Bhavnani (CRB, 2004) argue powerfully that early-impact aid boosts economic growth. The reader may presume that all participants are biased by entrenched positions.

Yet I daresay all have also learned from the debate. I hope the present analysis will continue that process and will help others more objective reach their own conclusions.

1. CRBB in thumbnail

Rather than building a fresh data set and specification, as in CRB, CRBB reanalyze the three most-cited studies in the aid-growth literature: Boone (1996), Burnside and Dollar (BD, 2000), and Rajan and Subramanian (RS, 2008). Boone and RS come from what CRBB term the "null strand" of the literature, which cannot detect any impact of aid on growth of investment/GNI (Boone) or GDP/capita (RS). BD leads the "conditional strand," finding that aid works in countries with "good" economic policies.[3] All data sets examined are panels; Boone's period length is ten years, BD's four, and RS's five.[4]

[2] For economy of language, I will sometimes use "impact" to describe and aid-growth association, without intending an interpretation of causality from aid to growth.
[3] This classification, while helpful, can blur the distinction between design and inference. RS, for example, run "conditional" regressions like BD's but reach "null" conclusions.
[4] RS also perform long-period cross-section regressions, but these receive less attention, probably because CRBB tend to favor panel regressions, which allow more careful handling of timing effects.

For each subject paper, CRBB begin with pooled panel Ordinary Least Squares (OLS) regressions using the subject's controls, sample, and data set—in the case of Boone, rebuilt with great effort. CRBB then expand the data sets with more-recent periods and stepwise introduce specification changes, most of which reflect a focus on the timing of the impacts of aid.

The CRBB unification project superficially resembles Roodman (2007), which copies specification choices such as period length and control set from one aid-growth paper to another. But the CRBB strategy is nearly orthogonal in that the main project is to introduce specification changes largely absent from the subject studies:

1. restricting the aid variable to those kinds of aid that can be most expected to affect growth within a few years (CRB 2004), called "early-impact aid";
2. lagging the aid variable by one period to reduce endogeneity;
3a. first-differencing the data to remove omitted-variable bias from time-invariant factors;
3b. since differencing renders the predetermined variable initial log GDP/capita presumptively endogenous, instrumenting this variable with its lag (Anderson and Hsiao 1982).

CRBB introduce one more change, which is important to their results but not bolted to the paper's conceptual frame:

4. regressing on $(\text{aid/GDP})^2$ along with aid/GDP in order to model nonlinear relationships.

As none of the regressions instruments for aid, lagging is the main strategy for reducing endogeneity. Thus the notion of causality pursued is akin to that of Granger (1969).

CRBB summarize their results as showing a positive correlation between past aid/GDP and current growth. After some robustness tests, they conclude that causation of growth by aid best explains the correlation. In sum, roughly speaking, if the three subject studies had matched CRB (2004)'s key specification choices, they would have matched CRB's results.

2. Early impact aid vs. non–early impact aid

As mentioned, an intelligent innovation in CRB and CRBB is the narrowing of the aid variable to assistance whose impact should be most detectable in a short-period panel, i.e., aid whose impacts on growth arrive quickly. Thus the title, "Counting Chickens When They Hatch." CRBB argue that when the aid variable is so narrowed, its significance grows. That is compelling circumstantial support for the notion that aid's impact is there to be found, if properly sought. But the prior that "early impact aid" is especially predictive of growth turns out not to hold.

CRBB operationalize early-impact aid by including budget support as well as investment in infrastructure, agriculture, and industry. They exclude technical cooperation, humanitarian assistance, and most aid for social sectors such as health and education, on the idea that these will only affect growth in the long run, if at all. In ambiguous cases, such as budget support, they err on the side of inclusion, since under their prior, mixing non-early aid into early aid should conservatively bias estimates of the near-term impact of early aid toward zero.

The CRBB datasets include several aid variables as fractions of GDP or GNI: receipts of early aid for each country and period; receipts of the conventional, broader aggregate, net Overseas Development Assistance (net ODA);

and repayments made on past ODA loans, which are what is netted out of net ODA. From these data, I compute:

Gross ODA = Net ODA + Repayments

Non–early impact aid = Gross ODA – Early-impact aid

I then substitute non-early aid for early aid in CRBB regressions.

Table 1 shows some replication regressions, using early aid, and an equal number with non-early aid. As in CRBB, the top part of the table regresses on aid in linear form while the bottom does so in quadratic. All regressions are run on CRBB's "extended samples," which add more years to the original studies' sampling frames. From left to right, the table has three panes, one for each of CRBB's subject studies. The first two columns of each pane copy specifications evidently preferred by CRBB, which are characterized by their expanded data sets and the early-impact aid variable. The first regressions in these pairs are OLS in levels, without country dummies. The second are in differences: for Boone they are still OLS while for BD and RS they are Anderson-Hsiao (1982), as explained in the next section.[5] The right halves of the three panes replace early aid with non-early aid.

To help interpret the quadratic-specification results, the table also shows an *F* test for the two aid/GDP coefficients; the aid/GDP level of maximum impact; the total impact at that level; and the marginal impact at the mean aid/GDP level. Mean aid/GDP is itself reported near the bottom.

In the linear specifications, non-early aid is slightly more apt than early aid to predict subsequent growth, with two coefficients significant instead of 0.1 instead of one. The contrast is greater in the quadratic specifications. The maximum impact and the marginal impact at the average aid level are statistically significant in four non-early aid regressions and one early aid regression. And in five of six cases, non-early aid has a higher point estimate for maximum impact.

Since early and non-early aid are substantially correlated, one might proxy for the other when only one is included a regression. So Table 2 regresses on both at once. Non-early aid wins the head-to-head contest by producing more significant coefficients.

It can be argued that the positive findings for non-early aid do not weaken CRBB's case that early aid also predicts growth; indeed, the additional results seem to reinforce the proposition that aid in general leads to growth. But the importance of the early aid distinction lies in its ability to discriminate between competing causal stories. For example, CRBB worry that the excluded variable population size could simultaneously affect growth and aid/GDP, leading to omitted-variable bias. A neutral prior would be that this causal story would affect all subcategories of aid equally. A finding that early and non-early aid/GDP have quite different relationships with growth would denigrate this prior in favor of the impact story. But that is not what we find.

[5] The Boone panel is too short for Anderson-Hsiao.

Table 1. Early vs. non–early impact aid

	Boone				Burnside & Dollar				Rajan & Subramanian			
	Early-impact aid/GDP		Non–early aid/GDP		Early-impact aid/GDP		Non–early aid/GDP		Early-impact aid/GDP		Non–early aid/GDP	
	OLS in levels	OLS in diff	OLS in levels	OLS in diff	OLS in levels	A-H	OLS in levels	A-H	OLS in levels	A-H	OLS in levels	A-H
Linear specification												
Aid/GDP, lagged	0.697***	0.016	0.382***	0.270	0.013	–0.195	0.139**	0.120	0.108	0.053	0.044	0.026
	(0.187)	(0.378)	(0.123)	(0.256)	(0.076)	(0.261)	(0.054)	(0.109)	(0.091)	(0.166)	(0.057)	(0.116)
Arellano-Bond *z*	3.40***		3.35***		1.89*	–0.94	1.88*	–1.26	2.16**	–0.18	2.20**	–0.19
Kleibergen-Paap *F*						21.62		27.22		5.12		5.91
Quadratic specification												
Aid/GDP, lagged	0.577	0.615	0.696*	1.118***	0.147	0.324	0.248**	0.310	0.380**	0.513	0.323***	0.082
	(0.537)	(0.516)	(0.352)	(0.406)	(0.167)	(0.316)	(0.113)	(0.295)	(0.169)	(0.384)	(0.114)	(0.246)
$(\text{Aid/GDP})^2$, lagged	0.559	–3.659**	–1.289	–3.422***	–1.152	–3.567*	–0.575	–0.686	–2.243	–4.239	–1.361***	–0.264
	(2.089)	(1.819)	(1.201)	(1.093)	(0.980)	(2.003)	(0.421)	(0.799)	(1.463)	(2.927)	(0.433)	(0.808)
F test of aid/GDP vars (*p*)	0.000***	0.059*	0.014**	0.006***	0.008***	0.102	0.013**	0.137	0.007***	0.477	0.000***	0.877
Maximum-impact aid level	–0.517	0.084*	0.270**	0.163***	0.064**	0.045	0.216***	0.226***	0.085***	0.060***	0.119***	0.155
	(–2.401)	(0.048)	(0.134)	(0.024)	(0.030)	(0.031)	(0.078)	(0.083)	(0.031)	(0.018)	(0.019)	(0.228)
Maximum impact	–0.149	0.026	0.094***	0.091**	0.005	0.007	0.027***	0.035	0.016**	0.015	0.019**	0.006
	(–0.829)	(0.035)	(0.035)	(0.041)	(0.007)	(0.012)	(0.010)	(0.030)	(0.008)	(0.014)	(0.009)	(0.023)
Marginal impact at average	0.609	0.363	0.591**	0.810**	0.103	0.179	0.212**	0.266	0.293**	0.344	0.239***	0.066
	(0.421)	(0.430)	(0.263)	(0.319)	(0.133)	(0.263)	(0.090)	(0.247)	(0.125)	(0.281)	(0.091)	(0.205)
Arellano-Bond *z*	3.51***		3.49***		1.99**	1.51	1.93*	1.50	2.28**	–2.15**	1.86*	–2.31**
Kleibergen-Paap *F*						23.81		29.44		4.94		6.47
Average aid/GDP	0.029	0.034	0.041	0.045	0.019	0.020	0.031	0.032	0.019	0.020	0.031	0.030
Observations	147	71	147	71	380	323	380	323	343	268	343	268

Results for controls, including log initial GDP/capita, not shown. First two Boone quadratic regressions match CRBB Table 5, columns 11 & 12. First two BD quadratic regressions match CRBB Table 7, columns 9 & 11. First two RS quadratic regressions match CRBB Table 9, columns 9 & 11. Anderson-Hsiao (A-H) regressions instrument lagged growth with twice-lagged growth. Maximum-impact aid level and maximum impact are the coordinates of the extrumemum of the parabola given by the coefficients on Aid/GDP and $(\text{Aid/GDP})^2$. The extremum of the parabola $y = ax^2 + bx$ is $(x, y) = (-b/2a, -b^2/4a)$. Maximum-impact aid level for first BD regression does not match CRBB because a bug in CRBB's code causes controls to be differenced. Arellano-Bond test is for first-order serial correlation in levels, for OLS, and second-order correlation in differences, for A-H. Kleibergen-Paap *F* is a heteroskedasticity-robust measure of instrument strength. Heteroskedasticity-robust standard errors in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

Table 2. Regressions on early aid/GDP and non-early aid/GDP at once

	Boone		Burnside & Dollar		Rajan & Subramanian	
	OLS in levels	OLS in diff	OLS in levels	A-H	OLS in levels	A-H
Linear specification						
Early aid/GDP, lagged	0.448*	−0.271	−0.137	−0.189	0.097	0.052
	(0.242)	(0.381)	(0.095)	(0.246)	(0.105)	(0.165)
Non-early aid/GDP, lagged	0.246*	0.336	0.175***	0.126	0.020	0.022
	(0.139)	(0.242)	(0.064)	(0.113)	(0.066)	(0.114)
Arellano-Bond z	3.27***		1.80*	−1.18	2.16**	−0.22
Kleibergen-Paap *F*				23.65		5.16
Quadratic specification						
Early aid/GDP, lagged	−0.280	−0.029	−0.113	0.244	0.181	0.579
	(0.572)	(0.626)	(0.228)	(0.357)	(0.211)	(0.473)
(Early aid/GDP)^2, lagged	4.803*	1.258	−0.269	−2.914	−0.664	−4.920
	(2.733)	(4.595)	(1.217)	(2.166)	(1.642)	(3.986)
Non-early aid/GDP, lagged	1.014**	1.214**	0.290**	0.225	0.272*	−0.042
	(0.443)	(0.602)	(0.145)	(0.327)	(0.145)	(0.267)
(Non-early aid/GDP)^2, lagged	−3.375**	−4.118	−0.596	−0.449	−1.237**	0.304
	(1.662)	(2.622)	(0.467)	(0.897)	(0.571)	(0.937)
Arellano-Bond z	3.38***		1.85*	1.59	1.88*	−2.33**
Kleibergen-Paap *F*				26.23		5.14
Observations	147	71	380	323	343	268

Results for controls, including log initial GDP/capita, not shown. Arellano-Bond test is for first-order serial correlation in levels, for OLS, and second-order correlation in differences, for A-H. Kleibergen-Paap *F* is a heteroskedasticity-robust measure of instrument strength. Heteroskedasticity-robust standard errors in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

3. Fixed effects, lags, and Anderson-Hsiao regression

The second and third steps CRBB take in light of timing concerns are to lag aid observations by one period and to difference the data to remove time-invariant variables. Both steps are standard. Lagging should reduce the endogeneity of aid. Differencing removes omitted country-level traits that simultaneously cause aid and growth. However, the implementation of these steps departs from standard practice in panel econometrics. By the same token, it falls short of CRBB's goal of temporally separating aid and growth. The upshot is a significant endogeneity concern that pertains to all the regressions in differences, including the Anderson-Hsiao ones that are the paper's most rigorous.

3.1. Econometrics

The various CRBB estimators (OLS in levels, OLS in differences, Anderson-Hsiao as they implement it) are all biased in distinctive ways, most of which CRBB note. Here I review why. I focus on the BD and RS regressions because they are simpler to discuss and are the basis of the most rigorous results.[6]

A typical CRBB regression equation can be cast as:

$$\Delta y_{it} = (\alpha - 1)y_{i,t-1} + \beta_1 d_{i,t-1} + \beta_2 d_{i,t-1}^2 + \boldsymbol{\gamma}'\mathbf{x}_{it} + \eta_t + \mu_i + \epsilon_{it} \quad (1)$$

where y_{it} is log GDP/capita of country i in period t, d is disbursements of aid/GDP, β_2 might be constrained to 0, $\mathbf{x}$ is a control set, η is fixed over space, μ is fixed over time, and ϵ is a mean-zero error that is assumed at a minimum to be independent across countries.[7] This is a dynamic panel data model. CRBB treat $y_{i,t-1}$ and $d_{i,t-1}$ as predetermined: potentially affected by growth innovations in periods $t-1$ and earlier, but not by innovations in period t or later. CRBB treat $\mathbf{x}$ as strictly exogenous, meaning $\mathrm{E}[\mathbf{x}_{is}\epsilon_{it}] = 0$ for all i, s, t.[8]

An immediate challenge to estimating this model is that the μ_i are "in" the regressor $y_{i,t-1}$ since equation (1) holds at $t-1$ too. Allowing the μ_i to remain in the error of a regression will almost certainly drive that error into correlation with $y_{i,t-1}$. This positively signed endogeneity will bias the estimate of α upward.[9] The knock-on effects for other parameters are unclear. In this model, the nonlinearity in d but not y creates some unusual possibilities. Perhaps the relationship between μ and y is nonlinear; the dependence of d on y almost certainly is since y is log GDP/capita and d is aid/GDP. The $d_{i,t-1}^2$ term could proxy for or pick up either of these linkages.

Perhaps more important, d may be endogenous to μ through other channels. CRBB highlight the fixed factor of population size, more precisely initial and average log population for the study period. Less-populous countries receive disproportionate aid (see BD). Meanwhile, population size probably affects growth through channels that bypass aid and other factors controlled for in CRBB's subject studies. CRBB cite Frankel and Romer (1999) on the amount of trade as a channel, Hausmann, Hwang, and Rodrik (2007) on the composition of exports, and Spolaore and Wacziarg (2005) on political integration with neighbors. The profusion of such stories linking fixed traits to growth again makes it unclear which way the endogeneity of $y_{i,t-1}$ to μ would bias β_1 and β_2.

The μ_i are econometric nettles. Consistency requires extracting them from the error. An intuitive strategy is to introduce country dummies. But in short panels, this causes the number of estimated parameters to grow as fast as the number of countries, which can also undermine consistency. Nickell (1981) shows that the effect is not innocuous in dynamic models.[10] Alternatively, using a random-effects estimator would delink the parameter count from the panel width—but at the cost of assuming that the μ_i are uncorrelated with aid/GDP and other regressors, that is, at the cost of assuming away the very concern that drives us.

[6] The Boone regressions put investment on the left but GDP/capita and its square on the right, which are complications from the point of view of dynamic panel econometrics. The Boone time dimension is too short to support the econometric fixes described below. And the errors contain serial correlation that I fail in the appendix to eradicate.

[7] Adding $y_{i,t-1}$ to both sides of (1) yields an AR(1) form:

$$y_{it} = \alpha y_{i,t-1} + \beta_1 d_{i,t-1} + \beta_2 d_{i,t-1}^2 + \boldsymbol{\gamma}'\mathbf{x}_{it} + \eta_t + \mu_i + \epsilon_{it}$$

[8] CRBB also control for reflows on aid loans and treat these as predetermined. I leave these variables out of the discussion for simplicity but in the empirics treat them in the same way as I treat d.

[9] This is more easily seen in the equation in note 7.

[10] CRBB err slightly in implying that dynamic panel bias arises in OLS in levels or OLS in differences. Rather, it occurs in the Least-Squares Dummy Variables estimator—OLS with fixed-effect dummies.

In their most rigorous regressions, CRBB take the common tack of differencing (1) to expunge the μ_i:[11]

$$\Delta\Delta y_{it} = (\alpha - 1)\Delta y_{i,t-1} + \beta_1 \Delta d_{i,t-1} + \beta_2 \Delta\left(d_{i,t-1}^2\right) + \boldsymbol{\gamma}'\Delta\mathbf{x}_{it} + \eta_t + \Delta\epsilon_{it} \qquad (2)$$

This step too comes at a cost, but one that is perhaps more manageable: it makes predetermined variables presumptively endogenous. For example, since in (1) Δy_{it} is endogenous to ϵ_{it}, $\Delta y_{i,t-1}$ is endogenous to $\epsilon_{i,t-1}$. The positive correlation between $\Delta y_{i,t-1}$ and $\epsilon_{i,t-1}$ can induce, with reference to (2), a negative correlation between the regressor $\Delta y_{i,t-1}$ and the error $\Delta\epsilon_{it} = \epsilon_{it} - \epsilon_{i,t-1}$.[12]

A similar argument says that differencing renders the other predetermined variables, $d_{i,t-1}$ and $d_{i,t-1}^2$, presumptively endogenous. Roodman (2008) confirms that growth negatively Granger-causes aid/GDP: when GDP/capita goes up, aid/GDP goes down. So we should expect the $d_{i,t-1}$ in $\Delta d_{i,t-1} = d_{it} - d_{i,t-1}$ to correlate negatively with the $\epsilon_{i,t-1}$ in $\Delta\epsilon_{it} = \epsilon_{it} - \epsilon_{i,t-1}$. That makes $\Delta d_{i,t-1}$ positively endogenous to $\Delta\epsilon_{it}$and should inflate the aid/GDP coefficient estimates when regressing in differences. Indeed, in 15 of 17 cases in CRBB's core tables (numbered 4–9), moving from OLS in levels to OLS in differences increases the impact coefficients. Also, like all endogeneity, this endogeneity should bias standard errors downward.

A standard response to the way differencing renders the predetermined endogenous is to instrument the newly suspect variables with their own lagged levels or differences. In the simplest implementation, this yields the Anderson-Hsiao (1982) levels and differences estimators, which are exactly identified instrumental variables estimators than can be implemented with Two-Stage Least Squares (2SLS). In the case at hand, we would instrument $\Delta y_{i,t-1}$, $\Delta d_{i,t-1}$, and $\Delta\left(d_{i,t-1}^2\right)$ with $y_{i,t-2}$, $d_{i,t-2}$, and $d_{i,t-2}^2$ or $\Delta y_{i,t-2}$, $\Delta d_{i,t-2}$, and $\Delta\left(d_{i,t-2}^2\right)$ (Bond, Hoeffler, and Temple 2001). CRBB choose differences, presumably because this leads to stronger instrumentation.

But CRBB instrument $\Delta y_{i,t-1}$ only, not $\Delta d_{i,t-1}$ and $\Delta\left(d_{i,t-1}^2\right)$. They thus go against the guidance of Bond (2002): "If x_{it} is assumed to be endogenous then it is treated symmetrically with the dependent variable y_{it}."

As a result, not even the most rigorous CRBB regressions cleanly implement a key piece of CRBB identification strategy, which is to temporally separate aid and growth observations in time. The uninstrumented variable of interest $\Delta d_{i,t-1}$ is computed from measurements taken in periods $t-1$ and $t-2$. The dependent variable Δy_{it} is taken from periods t and $t-1$, resulting in temporal overlap. So, for example, CRBB's treatment of $\Delta d_{i,t-1}$as exogenous in (2) requires us to assume that for an Asian nation in the late 1990s, a negative shock to GDP/capita did not systematically affect both the change in aid/GDP from the early 1990s to the late 1990s and the change in growth from the late 1990s to the early 2000s. Yet is seems likely that that the negative GDP shocks increased aid/GDP (reducing the denominator and perhaps increasing the numerator) and paved the way for V-shaped recoveries. Notice how the timeframes of the variables of interest overlap in this scenario. And notice how, as a result, the late-1990s negative growth shocks could create the appearance of elevated aid/GDP leading to ele-

[11] The η_t need not be differenced because this would not affect their span.
[12] Required here is the reasonable assumption that ϵ_{it} is $\mathrm{I}(0)$. As a counterexample, if growth is a random walk conditioning on controls then a correlation with $\epsilon_{i,t-1}$ would not imply a correlation with $\Delta\epsilon_{it}$.

vated growth.[13]

To recap, different approaches to estimating (1) and (2) lead to different types of bias and inconsistency. OLS in levels without country dummies can bias the aid impact estimates in an unclear direction. OLS in differences inflates the apparent impact of aid. Anderson-Hsiao can reverse the latter bias, but as CRBB do it, Anderson-Hsiao won't fully address the endogeneity biases in differences.

3.2. Revised empirics

The analytical frameworks of CRBB and standard panel econometrics point to two fixes for the endogeneity of $\Delta d_{i,t-1}$ and $\Delta\left(d_{i,t-1}^2\right)$: instrument them with deeper lags of themselves; or regress directly on those lags instead.

Table 3 attempts the first fix, with four alternate instrument sets: the twice-lagged change in aid/GDP, the twice-lagged level, the thrice-lagged level, and the twice- and thrice-lagged levels together.[14] These combinations exhaust all obvious possibilities when limiting the lag depth to three periods. In quadratic regressions, squared terms are included as instruments too. Unfortunately, the instruments prove uniformly weak. Only one regression produces a Kleibergen-Paap *F* statistic above 4, CRBB's rough threshold (column 12: Rajan and Subramanian, linear specification, early aid/GDP, instrumenting with twice- and thrice-lagged levels). And that overidentified regression also produces a marginal Hansen *J* test of instrument validity ($p = 0.12$). Still-deeper lags of $d_{i,t-1}$ and $d_{i,t-1}^2$ could be adduced as instruments, but they can be expected to be even weaker, and their inclusion in 2SLS would force further sample shrinkage.

Or we can regress directly on deeper lags. In so doing, we model growth as a function of twice- rather than once-lagged aid/GDP:

$$\Delta\Delta y_{it} = (\alpha - 1)\Delta y_{i,t-1} + \beta_1 \Delta d_{i,t-2} + \beta_2 \Delta\left(d_{i,t-2}^2\right) + \boldsymbol{\gamma}'\Delta\mathbf{x}_{it} + \eta_t + \Delta\epsilon_{it}$$

Precisely because $\Delta d_{i,t-2}$ and $\Delta\left(d_{i,t-2}^2\right)$ are extremely weak instruments for $\Delta d_{i,t-1}$ and $\Delta\left(d_{i,t-1}^2\right)$, dropping the latter from regressions is unlikely to much bias the coefficient estimates for the former.

Table 4 reports regressions on twice-lagged instead of once-lagged aid/GDP. As before, the regressions are run within CRBB's "extended samples," which add newer data to the original studies' time frames. But the need for twice-lagged aid shortens the available panel. Despite the smaller samples, and presumably reduced endogeneity bias, the estimates are more precise than corresponding ones in Table 1. And, surprisingly, the coefficients on $\Delta d_{i,t-2}$ are negative, sometimes with statistical significance.

To check the robustness of these novel results, I incorporate data available in the CRBB data sets for countries they exclude from their regressions. In addition, I revise the specifications in two other minor ways. In the RS data set, I correct an apparent error in the M2/GDP control that results in some values being off by a factor of 100. And I cluster all standard errors by country, as in CRB and RS, and essentially as in BD.[15] [16] In terms of sam-

[13] CRBB consider a scenario akin to this one and seem to rule it out as a source of their results (their Table 10). But their test does not address the regressions in differences.

[14] Throughout, repayments/GDP and its square are instrumented analogously.

[15] BD use Newey-West standard errors of undocumented lag length.

ple size, the widening of the panel more than offsets the shortening wrought by twice-lagging aid. The results for these "super-extended" samples (Table 5) tell the same story as before.

It is worth emphasizing that twice-lagging the aid/GDP variables is more than an econometric gambit: it is a change in structural model. The new model has the disadvantage of leaving out the very causal relationship of interest to CRBB—from aid/GDP to GDP/capita a few years later. But comments in that spirit can cut both ways. CRBB's preferred regressions leave out the zero-period (contemporaneous) impact and the two-period effect, and the three-period effect, etc. They also exclude impacts transmitted (linearly) through the controls. The new regressions study a different section of aid's impact, one that is no more or less canonical. The major difference is that this section can be studied more rigorously.

[16] For the BD regressions, the added countries are Barbados, Burkina Faso, Burundi, Chad, China, Republic of Congo, Cyprus, Iran, Israel, Jordan, Liberia, Lesotho, Malta, Mauritania, Mauritius, Nepal, Panama, Papua New Guinea, Rwanda, Singapore, South Africa, and Uganda. For the RS regressions, the added countries are Bangladesh, Barbados, Burundi, Chad, Guatemala, Guinea-Bissau, Honduras, Lesotho, Liberia, Mali, Malta, Mauritania, Mauritius, Nepal, Rwanda, Swaziland, Tanzania, and Zambia. In the latter case, some of the countries are in the original study's panel (RS, Appendix 2).

Table 3. Anderson-Hsiao regressions with various instrument sets, CRBB extended samples

	Burnside & Dollar								Rajan & Subramanian							
	Early-impact aid/GDP				Non–early aid/GDP				Early-impact aid/GDP				Non–early aid/GDP			
Instruments for aid (Δd_{t-1}):	Δd_{t-2}	d_{t-2}	d_{t-3}	d_{t-2}, d_{t-3}	Δd_{t-2}	d_{t-2}	d_{t-3}	d_{t-2}, d_{t-3}	Δd_{t-2}	d_{t-2}	d_{t-3}	d_{t-2}, d_{t-3}	Δd_{t-2}	d_{t-2}	d_{t-3}	d_{t-2}, d_{t-3}
Linear specification																
Aid/GDP, lagged	2.255	−0.012	−0.554	0.307	0.259	0.506	10.634	0.314	0.543	0.898	1.047	0.491	10.506	0.971	0.944	0.169
	(1.889)	(0.912)	(2.425)	(0.893)	(0.419)	(0.815)	(96.043)	(0.388)	(1.109)	(0.589)	(0.913)	(0.494)	(69.052)	(0.769)	(1.745)	(0.344)
Arellano-Bond z	−1.05	−0.86	−0.63	−0.84	−0.87	−1.13	−0.10	−0.95	−1.09	−1.13	−1.30	−1.15	0.15	−1.09	−0.63	−1.13
Kleibergen-Paap F	0.69	2.48	0.42	1.17	1.49	0.37	0.00	0.93	0.46	2.54	0.39	4.66	0.01	0.82	0.12	1.93
Hansen J (p value)				0.206				0.619				0.120				0.004***
Quadratic specification																
Aid/GDP, lagged	2.871	1.004	6.565	2.047**	3.590	1.099	−6.387	0.922	−2.134	1.720*	2.161	1.667**	8.099	1.149	4.307	−0.290
	(1.982)	(1.481)	(8.991)	(1.039)	(4.495)	(1.384)	(17.657)	(1.148)	(9.379)	(1.030)	(2.260)	(0.816)	(40.532)	(0.863)	(8.050)	(0.474)
(Aid/GDP)^2, lagged	−10.513	−6.713	−54.378	10.167**	−9.617	−3.243	10.134	−2.628	33.615	−7.425	−11.543	−11.608*	2.837	−1.694	−21.272	1.119
	(8.435)	(5.939)	(72.102)	(4.300)	(11.242)	(3.465)	(39.595)	(2.779)	(108.278)	(8.577)	(32.475)	(7.024)	(42.769)	(2.099)	(32.747)	(1.727)
F test of aid/GDP vars (p)	0.000***	0.752	0.928	0.080*	0.910	0.802	0.989	0.670	0.000***	0.254	0.081*	0.098*	0.999	0.649	0.781	0.308
Maximum-impact aid level	0.137***	0.075	0.060*	0.101***	0.187***	0.169***	0.315	0.175***	0.032	0.116	0.094	0.072***	−1.427	0.339	0.101**	0.130
	(0.031)	(0.059)	(0.035)	(0.021)	(0.029)	(0.053)	(0.439)	(0.053)	(0.043)	(0.085)	(0.170)	(0.016)	(−18.880)	(0.259)	(0.040)	(0.111)
Maximum impact	0.196	0.038	0.198	0.103	0.335	0.093	−1.006	0.081	−0.034	0.100	0.101	0.060**	−5.779	0.195	0.218	−0.019
	(0.126)	(0.082)	(0.323)	(0.067)	(0.452)	(0.139)	(−1.988)	(0.119)	(−0.192)	(0.068)	(0.096)	(0.028)	(−75.987)	(0.170)	(0.484)	(−0.040)
Marginal impact at mean	2.413	0.730	4.197	1.604*	2.960	0.891	−5.723	0.749	−0.733	1.425*	1.680*	1.183**	8.270	1.047	3.019	−0.223
	(1.630)	(1.268)	(6.278)	(0.872)	(3.763)	(1.168)	(−15.15)	(0.970)	(−5.016)	(0.769)	(1.009)	(0.552)	(41.917)	(0.764)	(6.090)	(−0.388)
Arellano-Bond z	−1.47	1.52	0.36	0.11	0.38	1.52	−0.43	0.65	−0.22	−2.92***	−0.18	−2.35**	0.01	−2.92***	0.46	−2.29**
Kleibergen-Paap F	0.79	1.95	0.14	2.33	0.19	0.85	0.06	0.39	0.04	2.12	0.02	3.24	0.01	0.86	0.21	2.03
Hansen J (p value)				0.194				0.495				0.274				0.050**
Mean aid/GDP	0.022	0.020	0.022	0.022	0.033	0.032	0.033	0.033	0.021	0.020	0.021	0.021	0.030	0.030	0.030	0.030
Observations	276	323	276	276	276	323	276	276	214	268	214	214	214	268	214	214

Results for controls, including log initial GDP/capita, not shown. Arellano-Bond test is for first-order serial correlation in differences. Kleibergen-Paap F is a heteroskedasticity-robust measure of instrument strength. Hansen J is a test for instrument invalidity. Heteroskedasticity-robust standard errors in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

Table 4. Anderson-Hsiao regressions with aid/GDP twice-lagged, CRBB extended samples

	Burnside & Dollar		Rajan & Subramanian	
	Early-impact aid/GDP	Non–early aid/GDP	Early-impact aid/GDP	Non–early aid/GDP
Linear specification				
Aid/GDP, twice-lagged	−0.302*	−0.072	−0.108	−0.330***
	(0.157)	(0.121)	(0.144)	(0.088)
Arellano-Bond z	−0.58	−0.72	−0.75	−0.69
Kleibergen-Paap *F*	42.96	44.49	22.36	22.30
Quadratic specification				
Aid/GDP, twice-lagged	−0.339	−0.404*	−0.483	−0.254
	(0.307)	(0.226)	(0.365)	(0.209)
$(\text{Aid/GDP})^2$, twice-lagged	0.291	1.314**	3.614	−0.309
	(1.753)	(0.603)	(3.504)	(0.830)
F test of aid/GDP vars (*p*)	0.143	0.692	0.409	0.001***
Maximum-impact aid level	0.582	0.154***	0.067**	−0.411
	(3.055)	(0.042)	(0.026)	(−1.417)
Maximum impact	−0.099	−0.031	−0.016	0.052
	(−0.446)	(−0.023)	(−0.012)	(0.221)
Marginal impact at mean	−0.327	−0.324*	−0.339	−0.272
	(−0.248)	(−0.195)	(−0.241)	(−0.167)
Arellano-Bond z	0.10	0.16	−1.75*	−1.86*
Kleibergen-Paap *F*	41.69	44.28	30.95	23.90
Mean aid/GDP	0.020	0.030	0.020	0.029
Observations	276	276	215	215

Results for controls, including log initial GDP/capita, not shown. Anderson-Hsiao (A-H) regressions instrument lagged growth with twice-lagged growth. Arellano-Bond test is for second-order correlation in differences. Kleibergen-Paap *F* is a heteroskedasticity-robust measure of instrument strength. Heteroskedasticity-robust standard errors in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

Table 5. Anderson-Hsiao regressions with aid/GDP twice-lagged, "super-extended" samples

	Burnside & Dollar		Rajan & Subramanian	
	Early-impact aid/GDP	Non–early aid/GDP	Early-impact aid/GDP	Non–early aid/GDP
Linear specification				
Aid/GDP, twice-lagged	–0.076	0.016	–0.099	–0.142
	(0.127)	(0.089)	(0.147)	(0.107)
Arellano-Bond z	–1.52	–1.56	–1.57	–1.46
Kleibergen-Paap *F*	12.37	11.13	11.90	12.31
Quadratic specification				
Aid/GDP, twice-lagged	–0.037	–0.333*	–0.444	–0.465*
	(0.246)	(0.187)	(0.358)	(0.270)
(Aid/GDP)^2, twice-lagged	–0.350	0.946**	3.295	1.017
	(1.373)	(0.380)	(2.995)	(0.783)
F test of aid/GDP vars (*p*)	0.588	0.734	0.687	0.397
Maximum-impact aid level	–0.053	0.176***	0.067***	0.229***
	(–0.546)	(0.033)	(0.025)	(0.063)
Maximum impact	0.001	–0.029	–0.015	–0.053**
	(0.017)	(–0.022)	(–0.013)	(–0.026)
Marginal impact at mean	–0.053	–0.263	–0.285	–0.384*
	(–0.193)	(–0.159)	(–0.234)	(–0.212)
Arellano-Bond z	0.98	1.61	–1.44	–1.08
Kleibergen-Paap *F*	12.11	10.49	19.15	13.96
Mean aid/GDP	0.023	0.037	0.024	0.040
Observations	381	381	269	269

Results for controls, including log initial GDP/capita, not shown. Anderson-Hsiao (A-H) regressions instrument lagged growth with twice-lagged growth. Arellano-Bond test is for second-order correlation in differences. Kleibergen-Paap *F* is a cluster-robust measure of instrument strength. Standard errors clustered by country in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

3.3. What to keep and what to dump

The econometric considerations in the previous subsection drive one to discard more and more information in an effort to banish endogeneity. Perhaps in dumping so much potentially tainted identifying information, we dump nearly all identifying information; perhaps we are rigorous to the point of rigor mortis.

There can be no certainty about the how best to make the trade-off between consistency and efficiency. Yet in the case at hand CRBB's standards argue against their own regressions and in favor of those in Table 4 and Table 5.

How so? Boone, BD, and RS all instrument aid. CRBB take a strong stand against the practice, at least as performed in those papers. CRBB point out that a single variable, the log of country population, explains most of the instrument strength in these studies.[17] As mentioned above, CRBB cite research suggesting that population affects growth not only through aid, but also through trade and regional political integration. This makes population a suspect instrument in regressions that do not control for such factors. "Relying on country population as an instrument throws into serious doubt the validity of the entire instrumentation strategy, and therefore of all regression results."

Population is essentially a fixed factor. Some 98% of its variation in CRBB samples is cross-country. This means that the component of variation in aid/GDP that is explained by population, whose exogeneity CRBB strongly question, is present in all CRBB regressions in levels. In general, the fear of fix factors is what drove the growth literature cross-sections to panels (Islam 1995, Temple 1999).

Meanwhile, the last subsection showed that CRBB's differencing leaves even lagged aid/GDP vulnerable to contemporaneous endogeneity, which goes against their concern with proper handling of timing effects.

Thus none of CRBB's regressions meet their econometric requirements: removing fixed effects and temporally separating aid and growth. The regressions in Table 4 and Table 5 handle both issues.

4. Maximum-impact aid level vs. maximum impact

Results from quadratic specifications are less intuitive than those from linear ones, since they allow marginal impact to vary with the aid level. To help interpret quadratic regressions, CRBB report the implied aid/GDP level at which the aid-growth association peaks. They do not report what would seem more relevant: the strength of the association at that point of maximum association—or, in coarser language, the impact at the point of maximum impact. The latter is much more relevant to understanding the impact of aid. And it turns out to be statistically significant much less often. CRBB's reporting of the turning-point aid level thus misleads about the certainty of the aid-growth association, however inadvertently.

The evidence for my claim has already been presented. In four of six cases in Table 1, the CRBB-reported aid level of maximum association, or "aid level of maximum impact," for early aid/GDP differs from zero at $p \leq 0.1$ (columns 1, 2, 5, 6, 9, 10). In contrast, the "maximum impact" is significant in just one, the RS regression in levels (column 9). (This is also the only early aid/GDP regression in which the marginal impact at the mean aid/GDP level is significant.)

As a result, if one looks squarely at CRBB's preferred regressions, one will not see much evidence of a positive aid-growth association. The more-rigorous Anderson Hsiao regressions (lower parts of columns 6 and 10, Table 1) put the maximum impact at 0.7% of GDP per year ($p = 0.53$) and 1.5% ($p = 0.27$) respectively.

5. Other specification issues

Interpretations of all the results presented so far can be strengthened and refined by addressing two more econometric issues. The first is serial correlation, which CRBB do not mention despite their central emphasis on

[17] In the case of RS, the reference is to their long-period cross-section regressions, not their short-period panel ones, which instrument with lags.

timing effects. Serial correlation is a first-order concern in panel econometrics and scrambles Granger inferences. The second issue is how best to study nonlinearity in the aid-growth relationship. Just as CRBB introduce specification improvements absent from their subject studies, I adduce further modifications to explore these issues. The appendix provides details, which this section summarizes.

To attempt to remove the serially correlated component of growth from the error I add to the regressor sets all quadratic combinations of controls, meaning $\mathrm{vec}(\mathbf{xx}')$, not just $\mathbf{x}$. This stratagem appears to succeed in the BD and RS regressions and leaves their results for aid intact.

And taking a cue from CRBB, I perform semiparametric analogs of regressions. These also tend to confirm the parametric results, and even rehabilitate the linear specifications in the top parts of earlier tables as meaningful guides to overall relationships.

The next two figures illustrate. They form a five-graph sequence that starts from a specification in Table 1 that supports CRBB (column 9: RS, OLS in levels, matching CRBB Table 8, regression 9). The sequence then introduces CRBB's specification changes, then mine. Figure 1 starts. It shows the smoothed fit for the association between GDP/capita growth and lagged early aid/GDP after removing the effects of controls. It also provides 95% confidence intervals for the nonparametric fit, and plots linear and quadratic fits. Over the bulk of the sample, to the left, the association is positive. And it is statistically significant, in that point estimates where early aid/GDP = 4–8% are above the confidence intervals where early aid/GDP = 0%. Meanwhile, the quadratic fit more closely matches the nonparametric one over the majority of the data.

The sequence continues in Figure 2. First the data are differenced as in CRBB, producing that clear positive slope that is predicted to contain upward bias (corresponding to CRBB Table 8, regression 10). Then lagged growth is instrumented (Table 1, column 10, and CRBB Table 8, regression 11). Then aid variables are twice-lagged (as in my Table 4, column 3). Then quadratic controls are added to reduce serial correlation (Table 7 in the appendix, column 3). The aid-growth association in this pane, the most rigorous in the sequence, is mildly negative over most of the sample.

Figure 1. Semiparametric regression of GDP/capita growth on early aid/GDP, in context of RS specification (CRBB Table 9, regression 9)

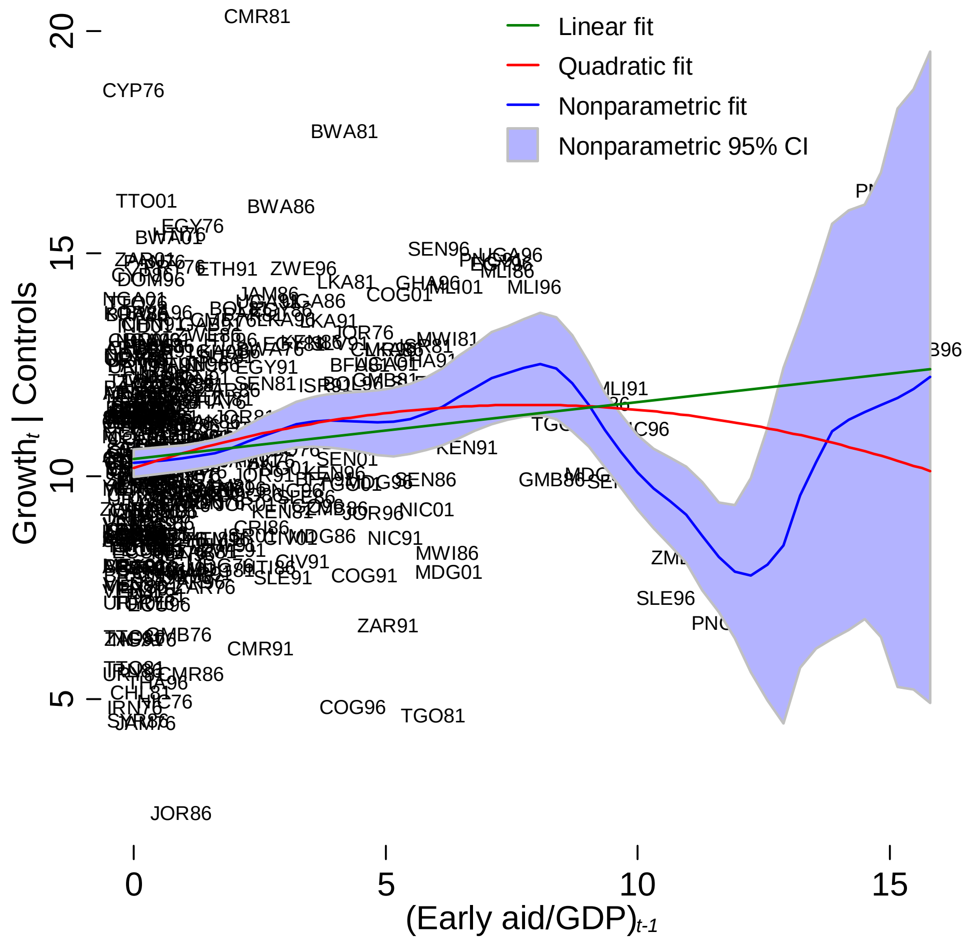
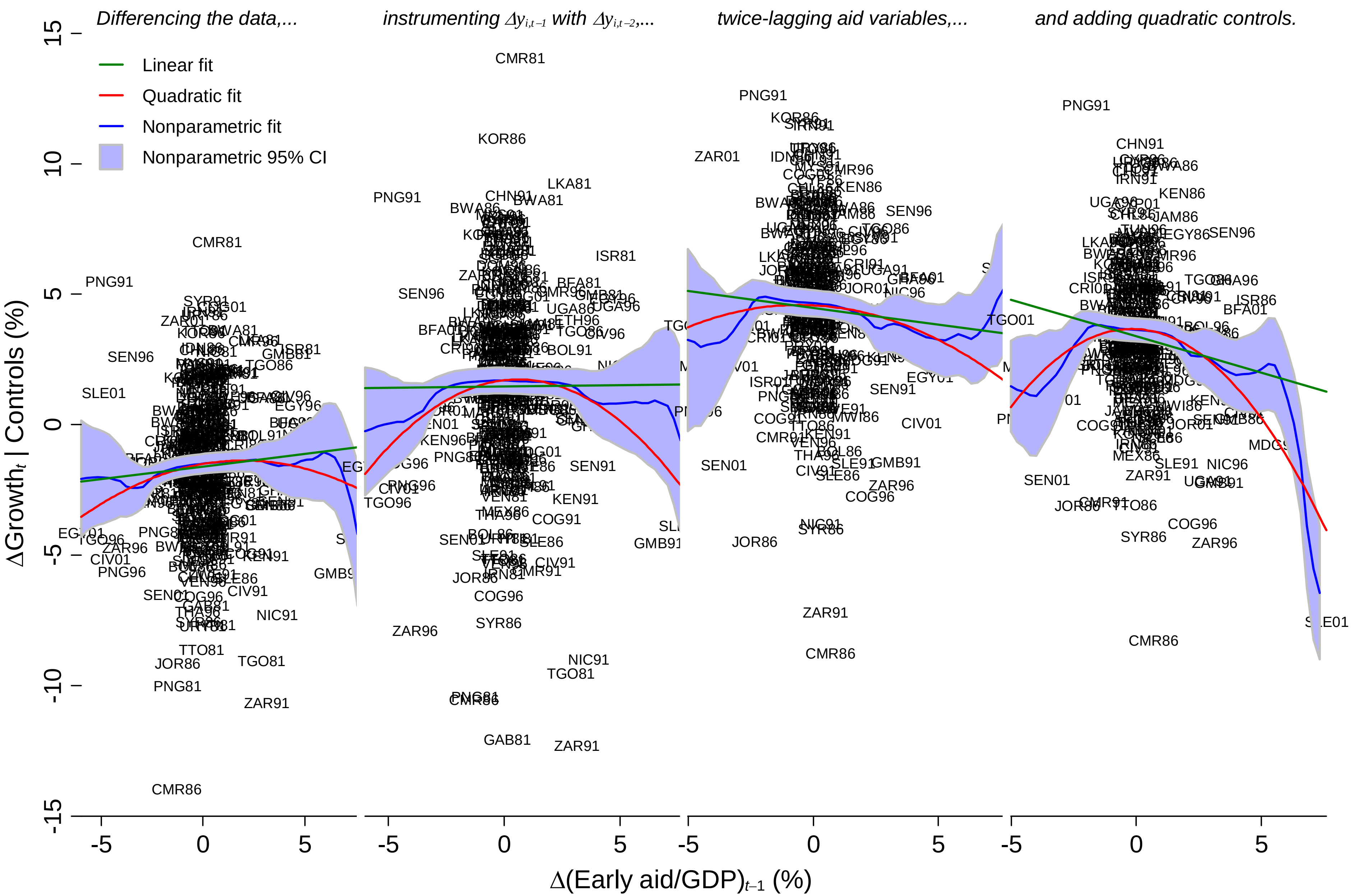


Figure 2. Semiparametric regression of GDP/capita growth on early aid/GDP, in context of RS specification (CRBB Table 9, regressions 10 –11)



6. Conclusion

CRBB is a must-read for anyone interested in the impact of foreign aid. It provides an excellent review of the literature, raises and addresses important econometric issues, introduces creative innovations, and makes the case for an important statistical regularity across data and control sets.

Despite the paper's intelligence and skill, its inference that aid probably causes growth does not appear well-founded to me. Of CRBB's three timing-motivated specification changes, the first, if anything, reduces the aid-growth association. The other changes, when done rigorously, point to an association that is zero or negative.

Arguably this critique is not an exercise in excessive perfectionism, but rather in holding CRBB regressions to CRBB standards. The regressions in levels harbor potential endogeneity from omitted time-invariant variables such as population size, which CRBB cite in setting aside other studies. The regressions in differences can harbor what CRBB seek to expunge: endogenous contemporaneous causation between growth and aid/GDP.

And even if the CRBB regressions are taken as valid, most of the relevant impact measures are not statistically significant: coefficients on early aid/GDP in the linear specifications, the maximum total impact, and the marginal impact at the mean are rarely statistically significant—except perhaps in the OLS-in-differences regressions which are known to be biased away from zero.

The strongest argument I see on behalf of the CRBB interpretation of the CRBB evidence is not very strong. It is that while fixed factors such as population are *present* in the early aid/GDP variable in the OLS-in-levels regressions, threatening endogeneity, they may not *dominate* in the way they would when instrumenting that variable with population, which is the practice CRBB actually criticize. Among the OLS-in-levels regressions, the BD and RS ones can be rid of serial correlation, and of these the RS one returns a significant impact if specified nonlinearly, as in Figure 1.

Though I focus on CRBB, I do not see it as particularly unreliable. BD's results are fragile (Easterly, Levine, and Roodman 2004). Bazzi and Clemens (2009) document significant problems in RS: the cross-section regressions depend on the dubious excludability of population size; the instruments in the panel GMM regressions are weak. Even leaving aside all the econometric issues discussed here and in CRBB, fundamental questions of identification remain in all observational cross-country studies. Variables such as GDP are poorly measured. Controls too may be endogenous. Errors may be correlated across countries. Overall we arguing men have been more effective on offense than defense

In the end, no definitive interpretation of the evidence is possible. The best we can do is transparently map assumptions to conclusions. If are prepared to assume that unobserved, persistent country traits do not substantially affect both growth and aid, then we might conclude with 95% confidence that the maximum impact of early-impact aid is between –0.9% and +1.9% of GDP (regression 5 of Table 1) or between 0.0% and +3.2% (regression 9). Similarly, if we believe that a shock to GDP in one period does not, on average, influence the change in aid/GDP coming into the period and the change in growth going out of it, we can embrace the results of CRBB's Anderson-Hsiao regressions, with their wider confidence intervals.

I am wary of the required assumptions, as I read CRBB to be. But once the assumptions are spelled out, the reader can judge for herself.

References

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

Angrist, Joshua D., and Krueger, Alan B. 1999. Empirical Strategies in Labor Economics. In O. C. Ahsenfelter and D. Card, eds., *Handbook of Labor Economics* 3. North Holland.

Bazzi, Samuel, and Michael A. Clemens. 2009. Blunt Instruments: A Cautionary Note on Establishing the Causes of Economic Growth. Working Paper 171. Center for Global Development.

Bond, Stephen. 2002. Dynamic Panel Data Models: A Guide to Microdata Methods and Practice. Centre for Micro Data Methods and Practice. London.

Bond, Stephen R., Anke Hoeffler, and Jonathan Temple. 2001. GMM Estimation of Empirical Growth Models. Discussion Paper 3048. Centre for Economic Policy Research. London.

Boone, Peter. 1996. Politics and the Effectiveness of Foreign Aid. *European Economic Review* 40 (2), 289–329.

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

Clemens, Michael A., Steven Radelet, and Rikhil R. Bhavnani. 2004. Counting Chickens When They Hatch: The Short-term Effect of Aid on Growth. Working Paper 44. Center for Global Development.

Clemens, Michael A., Steven Radelet, Rikhil R. Bhavnani, and Samuel Bazzi. 2012. Counting Chickens When They Hatch: Timing and the Effects of Aid on Growth. *Economic Journal* 122: 590–617.

Easterly, William, Ross Levine, and David Roodman. 2004. Aid, Policies, and Growth: Comment. *American Economic Review* 94(2): 774–80.

Frankel, Jeffrey A., and David Romer. 1999. Does trade cause growth? *American Economic Review*, 89 (3): 379–99.

Granger, Clive W. J. 1969. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* 37 (3): 424–38.

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

Hausmann, Ricardo, Jason Hwang, and Dani Rodrik. 2007. What you export matters. *Journal of Economic Growth* 12 (2): 1–25.

Imbens, Guido W., and Joshua D. Angrist. 1994. Identification of Local Average Treatment Effects. *Econometrica* 62(2): 467–75.

Islam, Nazrul. 1995. Growth Empirics: A Panel Data Approach. *Quarterly Journal of Economics* 110(4): 1127–70.

Kleibergen F, Paap R. 2006. Generalized Reduced Rank Tests using the Singular Value Decomposition. *Journal of Econometrics* 127(1): 97–126.

Rajan, Raghuram G., and Arvind Subramanian. 2008. Aid and Growth: What Does the Cross-Country Evidence Really Show? *The Review of Economics and Statistics* 90(4): 643–65.

Robinson, P.M. 1988. Root-*N*-Consistent Semiparametric Regression. *Econometrica* 56(4): 931–54.

Roodman, David. 2007. The Anarchy of Numbers: Aid, Development, and Cross-Country Empirics. *World Bank Economic Review* 21(2): 255–77.

Roodman, David. 2008. Through the Looking Glass, and What OLS Found There: On Growth, Foreign Aid, and Reverse Causality. Working Paper 137. Center for Global Development.

Temple, Jonathan. 1999. The New Growth Evidence. *Journal of Economic Literature* 37(1): 112 – 56.

Appendix. Improving the specifications

Here I discuss two specification issues raised in Roodman (2008), as they pertain to CRBB.

Serial correlation

Arellano and Bond (1991) is the most-cited paper in panel econometrics. Though viewed as the source of an estimator, its purpose is to develop and assess specification tests for estimators in the Anderson-Hsiao family. Of particular concern is serial correlation, which undermines consistency when instrumenting with lags. With reference to (2) in the main text, if there is first-order serial correlation, then the $y_{i,t-2}$ in the instrument $\Delta y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$ may be presumed correlated with $\epsilon_{i,t-2}$ and thereby with the $\epsilon_{i,t-1}$ in the error $\Delta\epsilon_{it}$. Arellano and Bond introduce a test for serial correlation in short panels with cross-sectionally independent errors. When a specification fails the test, researchers should consider instrumenting with deeper lags instead and then testing for longer-order serial correlation.

Lack of serial correlation is also essential to Granger's conception of causality, which CRBB take as their guide star. Granger's feasible test, for example, assumes white noise errors. Despite focusing on timing effects, CRBB do not mention the issue. Nor, in most cases, do they adjust their standard errors for it as in BD, RS, and CRB.

Though not discussed above, the tables in the main text include an Arellano-Bond *z* test for every regression for which it is feasible. The test is for first-order correlation in the levels regressions. It is for second-order in the Anderson-Hsiao regressions, since that is the order of correlation that would undermine the exogeneity of the instrument $\Delta y_{i,t-2}$ and, in Table 4 and Table 5, the regressors $\Delta d_{i,t-2}$ and $\Delta\left(d_{i,t-2}^2\right)$.

Serial correlation is prevalent in the CRBB regressions and in my extensions of them, the biggest exceptions being the BD regressions with twice-lagged aid (left half of Table 4). Fixed effects might explain the serial correlation in levels; but the phenomenon generally persists after differencing.

The issue may not compromise CRBB's results or mine. In the BD and RS specifications, regressing on all quadratic combinations of controls—$\mathrm{vec}(\mathbf{xx}')$ as well as $\mathbf{x}$—extracts much of the persistent component of growth from the error. Table 6 demonstrates: it is like Table 1 except in adding quadratic controls. As in equation (1), the controls $\mathbf{x}$ here exclude time dummies and GDP/capita. Nearly all the strong results in Table 1 remain stable or strengthen even as serial correlation is reduced.

The identification issues raised in the main text still pertain. So Table 7 analogously updates the more-credible Table 4. The texture of the results again remains similar. Note that the Kleibergen-Paap (2006) heteroskedasticity-robust *F*, a measure of instrument strength, improves substantially in the RS Anderson-Hsiao regressions when moving from once-lagged aid (columns 10 and 12 of Table 6) to twice-lagged (columns 3 and 4 of Table 7). Meanwhile, the serial correlation subsides. Overall, the most rigorous and relevant parametric impact estimates in this paper are those for BD in Table 4 (where serial correlation is not clearly detected in the first place) and for RS in Table 7 (where serial correlation is removed).[18]

[18] Parallel tables for the "super-extended" samples are in the spreadsheet on this paper's web page.

Table 6. Regressions including quadratic controls

	Boone				Burnside & Dollar				Rajan & Subramanian			
	Early-impact aid/GDP		Non–early aid/GDP		Early-impact aid/GDP		Non–early aid/GDP		Early-impact aid/GDP		Non–early aid/GDP	
	OLS in levels	OLS in diff	OLS in levels	OLS in diff	OLS in levels	A-H	OLS in levels	A-H	OLS in levels	A-H	OLS in levels	A-H
Linear specification												
Aid/GDP, lagged	0.687***	0.256	0.350***	0.591	–0.013	–0.267	0.138*	0.120	0.097	0.072	0.033	0.015
	(0.177)	(0.525)	(0.123)	(0.383)	(0.083)	(0.283)	(0.073)	(0.114)	(0.100)	(0.217)	(0.066)	(0.119)
Arellano-Bond *z*	2.74***		3.36***		0.63	–0.78	0.66	–1.14	–0.04	0.35	–0.10	0.37
Kleibergen-Paap *F*						19.32		23.97		1.16		1.31
Quadratic specification												
Aid/GDP, lagged	0.296	0.573	0.431	1.653***	0.222	0.228	0.298**	0.316	0.537***	0.503	0.331**	0.184
	(0.523)	(0.708)	(0.406)	(0.565)	(0.174)	(0.355)	(0.149)	(0.324)	(0.198)	(0.672)	(0.140)	(0.261)
$(\text{Aid/GDP})^2$, lagged	1.957	–3.212	–0.335	–4.351***	–1.973*	–3.376	–0.750	–0.725	–3.855**	–3.935	–1.342**	–0.647
	(2.153)	(2.238)	(1.319)	(1.502)	(1.104)	(2.268)	(0.533)	(0.914)	(1.518)	(4.923)	(0.544)	(0.784)
F test of aid/GDP vars (*p*)	0.000***	0.281	0.009***	0.026**	0.001***	0.128	0.041**	0.118	0.004***	0.251	0.002***	0.554
Maximum-impact aid level	–0.076	0.089	0.643	0.190***	0.056***	0.034	0.199***	0.218**	0.070***	0.064***	0.123***	0.142
	(–0.215)	(0.075)	(1.962)	(0.024)	(0.021)	(0.039)	(0.067)	(0.090)	(0.015)	(0.021)	(0.024)	(0.090)
Maximum impact	–0.011	0.026	0.139	0.157**	0.006	0.004	0.030**	0.034	0.019**	0.016	0.020*	0.013
	(–0.052)	(0.051)	(0.305)	(0.060)	(0.007)	(0.010)	(0.014)	(0.032)	(0.008)	(0.024)	(0.011)	(0.024)
Marginal impact at average	0.409	0.352	0.403	1.261***	0.147	0.091	0.252**	0.269	0.387**	0.347	0.249**	0.145
	(0.404)	(0.599)	(0.304)	(0.441)	(0.138)	(0.295)	(0.121)	(0.269)	(0.151)	(0.485)	(0.111)	(0.220)
Arellano-Bond *z*	2.68***		3.42***		0.69	1.27	0.62	1.34	0.05	–0.48	–0.31	–0.51
Kleibergen-Paap *F*						20.14		27.73		1.09		1.62
Average aid/GDP	0.029	0.034	0.041	0.045	0.019	0.020	0.031	0.032	0.019	0.020	0.031	0.030
Observations	147	71	147	71	380	323	380	323	343	268	343	268

Results for controls, including log initial GDP/capita, not shown. Anderson-Hsiao (A-H) regressions instrument lagged growth with twice-lagged growth. Maximum-impact aid level and maximum impact are the coordinates of the extrumemum of the parabola given by the coefficients on Aid/GDP and $(\text{Aid/GDP})^2$. Arellano-Bond test is for first-order serial correlation in levels, for OLS, and second-order correlation in differences, for A-H. Kleibergen-Paap *F* is a heteroskedasticity-robust measure of instrument strength. Standard errors clustered by country in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

Table 7. Anderson-Hsiao regressions with aid/GDP twice-lagged and quadratic controls, CRBB extended samples

	Burnside & Dollar		Rajan & Subramanian	
	Early-impact aid/GDP	Non–early aid/GDP	Early-impact aid/GDP	Non–early aid/GDP
Linear specification				
Aid/GDP, twice-lagged	−0.316**	−0.115	−0.167	−0.317***
	(0.153)	(0.139)	(0.154)	(0.094)
Arellano-Bond *z*	−0.28	−0.51	−0.20	−0.32
Kleibergen-Paap *F*	33.61	34.72	8.29	8.09
Quadratic specification				
Aid/GDP, twice-lagged	−0.389	−0.529**	−0.202	−0.301
	(0.337)	(0.266)	(0.373)	(0.227)
$(\text{Aid/GDP})^2$, twice-lagged	0.600	1.653**	0.246	−0.068
	(2.013)	(0.667)	(3.694)	(0.944)
F test of aid/GDP vars (*p*)	0.107	0.522	0.389	0.003***
Maximum-impact aid level	0.324	0.160***	0.411	−2.230
	(0.846)	(0.037)	(5.493)	(−32.699)
Maximum impact	−0.063	−0.042	−0.042	0.336
	(−0.124)	(−0.028)	(−0.489)	(5.162)
Marginal impact at average	−0.365	−0.428*	−0.193	−0.305*
	(−0.267)	(−0.230)	(−0.245)	(−0.178)
Arellano-Bond *z*	−0.32	−0.20	−0.73	−0.79
Kleibergen-Paap *F*	33.58	35.10	13.17	9.23
Average aid/GDP	0.020	0.030	0.020	0.029
Observations	276	276	215	215

Results for controls, including log initial GDP/capita, not shown. Anderson-Hsiao (A-H) regressions instrument lagged growth with twice-lagged growth. Arellano-Bond test is for second-order correlation in differences. Kleibergen-Paap *F* is a heteroskedasticity-robust measure of instrument strength. Standard errors clustered by country in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

Linear, quadratic, and semiparametric estimation

Roodman (2008) worries about the use of quadratic aid and interaction terms in CRB, Hansen and Tarp (2001) and other papers. To rebut this concern, CRBB move toward semiparametric analysis. It is a fruitful direction, which gives deeper insight into the results already presented in regression tables. On balance, semiparametric analysis backs up CRBB's assertion that their results are not a fragile artifact of an interaction between the quadratic specification and outliers. But it does show that neither linear nor quadratic modeling is invariably superior.

Perhaps the best way to put my concern about quadratic specifications is this: aid variables are always or nearly always positive. That tends to make them collinear with their squares. In regressions, identification of the coefficients on these linear and quadratic aid terms arises from the (small) unique components of variation in each (Angrist and Krueger 1999, pp. 1310–14). A regression will then identify the Local Average Treatment Effect (LATE; Imbens and Angrist 1994) of aid/GDP and $(\text{aid/GDP})^2$ for these small components of variation associated with extreme and perhaps unrepresentative observations. If the regression functional form is precisely correct—

if parameters are homogenous—then extrapolation of the LATE to the full sample will be consistent. But in the messy world of cross-country econometrics, the model is unlikely to be exactly right. The smaller the identifying variation, the greater the risk that the LATE will be misleading. In short, parabolic fits can be sensitive to outliers. This does not make quadratic fits "wrong," but it generates a caution about stability.

In response, CRBB make a scatter plot. Referring to equation (2), but suppressing Δ's for clarity, their procedure can be described as linearly partialling the differenced controls $\mathbf{x}$ out of y and d_{t-1}, then plotting the residuals $\mathrm{P}_{\mathbf{x}}^{\perp} y$ against the residuals $\mathrm{P}_{\mathbf{x}}^{\perp} d_{t-1}$. For the selected regression, CRBB's plot reassures with an apparently strong positive association not dominated by a few outliers.

But this analysis appears to contain three significant issues. First, semiparametric analysis requires more than a scatter plot. Patterns that appear obvious to the naked eye may not be statistically significant, or even present. This is why semiparametric analysis entails regression too, with the familiar tools of point estimates and confidence intervals.

Second, an apparent coding error causes CRBB to plot $\mathrm{P}_{\mathbf{x}}^{\parallel} y$ and $\mathrm{P}_{\mathbf{x}}^{\parallel} d_{t-1}$—the components of y and d parallel to rather than orthogonal to $\mathbf{x}$.[19]

Third, $\mathrm{P}_{\mathbf{x}}^{\perp} y$ and $\mathrm{P}_{\mathbf{x}}^{\perp} d_{t-1}$ are a nonstandard basis for nonparametric analysis. A more established approach for partially linear models such as CRBB's is Robinson's (1988). It works as follows. We wish to fit the equation:

$$y = \boldsymbol{\gamma}'\mathbf{x} + f(d) + \epsilon \tag{3}$$

where $\mathbf{x}$ and d are exogenous to the error ϵ and f is an unknown function. We take expectations conditional on d:

$$\mathrm{E}[y|d] = \boldsymbol{\gamma}'\mathrm{E}[\mathbf{x}|d] + f(d)$$

and subtract them from (3):

$$y - \mathrm{E}[y|d] = \boldsymbol{\gamma}'(\mathbf{x} - \mathrm{E}[\mathbf{x}|d]) + \epsilon \tag{4}$$

We estimate $\mathrm{E}[y|d]$ and $\mathrm{E}[\mathbf{x}|d]$ using nonparametric regression, such as a kernel-weighted moving average, producing $\mathrm{E}_N[y|d]$ and $\mathrm{E}_N[\mathbf{x}|d]$. This leads to a feasible estimator of $\boldsymbol{\gamma}$ in (4): we regress $y - \mathrm{E}_N[y|d]$ on $\mathbf{x} - \mathrm{E}_N[\mathbf{x}|d]$ using OLS, producing $\hat{\boldsymbol{\gamma}}$. Returning to (3), we estimate f with nonparametric regression of $y - \hat{\boldsymbol{\gamma}}'\mathbf{x}$ on d.

We can generalize the procedure to the case when $\mathbf{x}$ is endogenous and enough instruments $\mathbf{z}$ are available. Then, we estimate $\boldsymbol{\gamma}$ in (4) with a linear GMM estimator based on $\mathbf{z}$. The regressions of interest are exactly identified, so GMM estimators coincide and we implement with 2SLS.

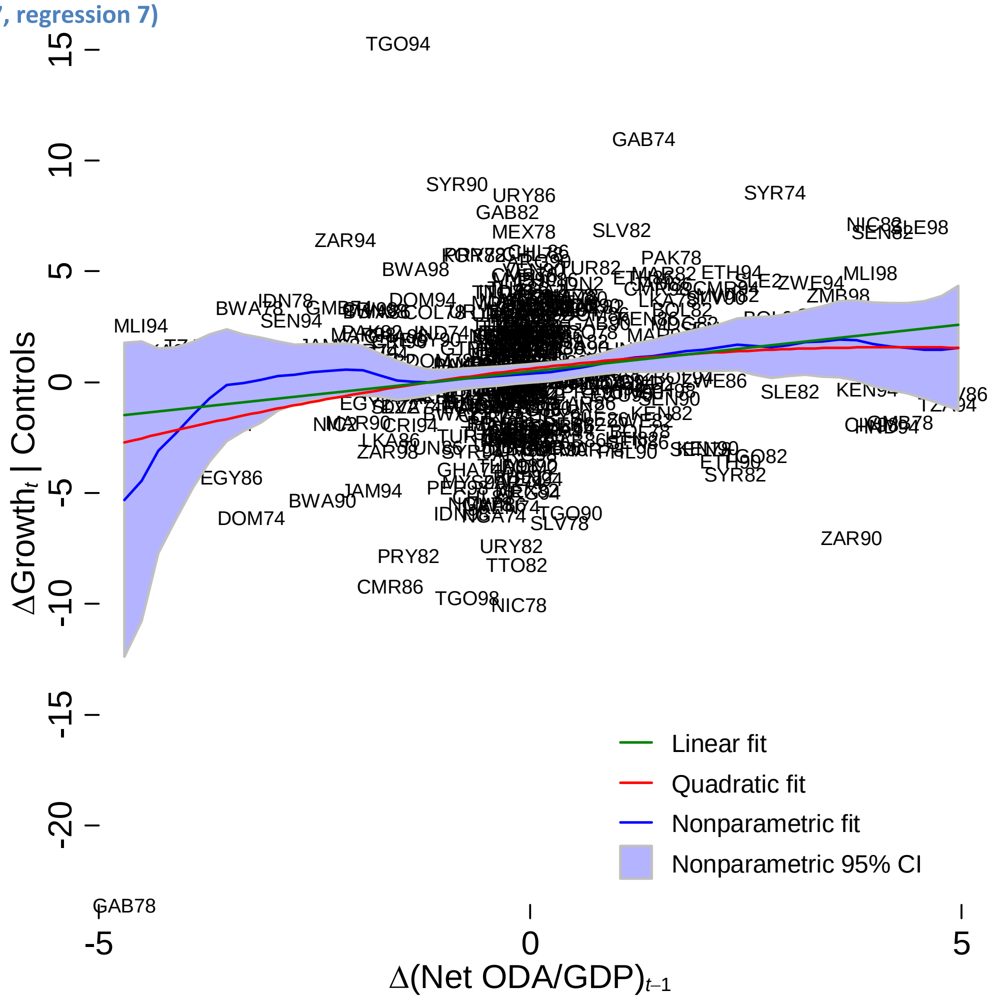
In contrast, CRBB move toward studying $\mathrm{E}[\mathrm{P}_{\mathbf{x}}^{\perp} y|\mathrm{P}_{\mathbf{x}}^{\perp} d]$. This is intuitive, but by the assumptions above, $\mathrm{E}[\mathrm{P}_{\mathbf{x}}^{\perp} y|\mathrm{P}_{\mathbf{x}}^{\perp} d] = \mathrm{E}[\mathrm{P}_{\mathbf{x}}^{\perp}(\boldsymbol{\gamma}'\mathbf{x} + f(d) + \epsilon)|\mathrm{P}_{\mathbf{x}}^{\perp} d] = \mathrm{E}[\mathrm{P}_{\mathbf{x}}^{\perp} f(d)|\mathrm{P}_{\mathbf{x}}^{\perp} d] + \mathrm{E}[\epsilon|\mathrm{P}_{\mathbf{x}}^{\perp} d] = \mathrm{E}[\mathrm{P}_{\mathbf{x}}^{\perp} f(d)|\mathrm{P}_{\mathbf{x}}^{\perp} d]$. This is not $\mathrm{E}[f(\mathrm{P}_{\mathbf{x}}^{\perp} d)|\mathrm{P}_{\mathbf{x}}^{\perp} d] = f$, the function of interest, if f is nonlinear.

[19] In lines 554 and 556 of the public main_regs_final.do Stata program, the "predict" commands lack a "residuals" option.

Figure 3 applies Robinson's procedure to the specification behind CRBB's scatter plot (BD, net ODA, OLS in differences). In the first stage of the procedure, $\mathrm{E}_N[y|d]$ and $\mathrm{E}_N[\mathbf{x}|d]$ are estimated as moving-average fits with Stata's "lpoly" command, taking defaults (Epanichnikov kernel, Stata's "rule of thumb" algorithm for setting bandwidth). The second-stage regression, shown, is done the same way. It is formatted like Figure 1, with 95% confidence intervals for the smoothed fit, and linear and quadratic fits as well.

The figure supports CRBB in suggesting a significant, positive aid-growth association in this specification over the bulk of the sample. Point estimates toward the right of the main data mass are above the confidence intervals toward the left, or nearly so, and vice versa. And the figure shows that this quadratic fit does not wildly chase outliers. That said, the overall relationship is well captured by the linear fit too.

Figure 3. Semiparametric regression of GDP/capita growth on Net ODA/GDP, in context of BD specification (CRBB Table 7, regression 7)



Similar graphs for the RS early aid specifications are in the main text. As noted, they line up well with the parametric results. In some cases, such as the second pane of Figure 2, the parabola appears to fit the data better than the line over the main data mass. In others, such as in the last pane, outliers appear to distort the quadratic fit. Overall, the semiparametric regressions usefully contextualize the linear and quadratic results.