Center for Global Development

Working Paper Number 171
May 2009

Blunt Instruments: On Establishing the Causes of Economic Growth

Samuel Bazzi and Michael Clemens

Abstract

Despite intense concern that many instrumental variables used in growth regressions may be weak, invalid, or both, top journals continue to publish studies of economic growth based on problematic instruments. Doing so risks pushing the entire literature closer to irrelevance. We illustrate hidden problems with identification in recent prominently published and widely cited growth studies using their original data. We urge researchers to take three steps to overcome the shortcomings: grounding research in somewhat more generalized theoretical models, deploying the latest methods to test sensitivity to violations of the exclusion restriction, and opening the "black box" of the Generalized Method of Moments (GMM) with supportive evidence of instrument strength.

The views expressed in this paper are those of the author and should not be attributed to the board of directors or funders of the Center for Global Development. This paper was made possible in part by support from the William and Flora Hewlett Foundation.

www.cgdev.org

Samuel Bazzi and Michael Clemens. 2009. "Blunt Instruments: On Establishing the Causes of Economic Growth."
CGD Working Paper 171. Washington, D.C.: Center for Global Development.
http://www.cgdev.org/content/publications/detail/1422132/

Blunt instruments: On establishing the causes of economic growth

Samuel Bazzi* Michael A. Clemens†‡

May 20, 2009

Abstract

Despite intense concern that many instrumental variables used in growth regressions may be weak, invalid, or both, top journals continue to publish studies of economic growth based on problematic instruments. Doing so risks pushing the entire literature closer to irrelevance. We illustrate hidden problems with identification in recent prominently published and widely cited growth studies using their original data. We urge researchers to take three steps to overcome the shortcomings: grounding research in somewhat more generalized theoretical models, deploying the latest methods to test sensitivity to violations of the exclusion restriction, and opening the "black box" of the Generalized Method of Moments (GMM) with supportive evidence of instrument strength.

*Email: sbazzi@ucsd.edu. Mailing address: Department of Economics; University of California, San Diego; 9500 Gilman Drive #0508; La Jolla, CA 92093-0508.

†Email: mclemens@cgdev.org. Mailing address: Center for Global Development; 1800 Massachusetts Ave. NW, 3rd floor; Washington, DC 20036.

‡JEL No. F35, C12, O4. We gratefully acknowledge support from the Hewlett Foundation. We thank Aart Kraay, Valerie Ramey, James Rauch, and David Roodman for helpful discussions. Any errors that remain are exclusively ours. Nothing herein necessarily represents the views of the Center for Global Development, its board, or its funders.

1 Introduction

One of the great challenges in economic research lies in identifying the causes of economic growth; one of the great challenges in applied economics lies in distinguishing correlation from causation. Together these mean that running regressions seeking the causes of growth is not for the meek. But the research questions involved hold such central importance that applied economists have built a large literature asserting that this or that factor causes growth.

Those growth regressions are greeted these days with a skepticism bordering on allergy. The wave of international growth empirics begun by Baumol (1986) and advanced by Barro (1991) stalled in the mid-1990s, as regression specifications proliferated and new results sometimes contradicted old. Lindauer and Pritchett (2002) wrote an unequivocal "obituary for growth regressions," and one of the literature's prominent contributors flatly concluded,

> "Using these regressions to decide how to foster growth is ... most likely a hopeless task. Simultaneity, multicollinearity, and limited degrees of freedom are important practical problems for anyone trying to draw inferences from international data. Policymakers who want to promote growth would not go far wrong ignoring most of the vast literature reporting growth regressions" (Mankiw 1995).

Since that nadir the literature has rallied somewhat. Researchers have become more assiduous in checking the robustness of results to the choice of regression specification (Sala-i-Martin 1997; Sala-i-Martin, Doppelhofer, and Miller 2004). They have explored concerns about parameter heterogeneity, measurement error, and influential observations (Temple 1999). They have expanded their samples as the succession of years and improvements in information technology have brought a flood of new data (Bosworth and Collins 2003).

Beyond this, researchers have taken greater care in testing whether the observed correlates of growth in fact *cause* growth. As some of their microeconomist colleagues have begun running randomized controlled trials around the world, growth empiricists have energetically sought natural macroeconomic quasi-experiments to identify the causal portion of the relationships they observe across countries. Architects of growth regressions published in top journals have used cross-country instrumental variables for governance quality,[1] trade,[2] and foreign aid,[3] among several other growth determinants. New develoments in econometrics have assisted this search for better identification, especially the advent of sophisticated dynamic panel Generalized Method of Moments (GMM) estimators (which entered the growth literature with Caselli, Esquivel, and Lefort 1996).

These efforts are a welcome alternative to nihilism. But in parallel with them, the economics literature in general has showed increasing concern with the strength and validity of instrumental variables in practice (surveyed by Murray 2006). Close investigations have suggested that many cross-country instruments may be weak, invalid, or both, in widely-cited studies on the growth effects of governance or trade (e.g. Rodríguez and Rodrik 2000; Brock and Durlauf 2001; Dollar and Kraay 2003; Glaeser, La Porta, López-de-Silanes, and Shleifer 2004; Albouy 2008; and Kraay 2008). Notwithstanding the popularity of instrumental variables in recent growth empirics, Durlauf, Johnson, and Temple (2005) conclude in their exhaustive review on the subject that "the belief that it is easy to identify valid instrumental variables in the growth context is deeply mistaken. We regard many applications of instrumental variable procedures in the empirical growth literature to be undermined by the failure to address properly the question of whether these instruments are valid".

Undeterred, researchers continue to publish instrument-based growth empirics in top journals. This suggests that either the latest work is using better instruments than before, or that earlier critiques of these methods have gone unheeded.

We argue that the field of growth empirics continues to pay inadequate attention to concerns about instrument validity and strength. This point is best made by example—not in order to single out particular papers, but to concretely characterize

[1] These include cross-country instrumental variables based on exogenous deaths of national leaders while in office (Jones and Olken 2005), colonial-era settler mortality (Acemoğlu, Johnson, and Robinson 2001), a Soviet-era survey of ethnolinguistic fractionalization (Mauro 1995), distance from the equator (Hall and Jones 1999), and Pacific-basin wind patterns (Feyrer and Sacerdote 2008).

[2] These include cross-country instruments based on geographic characteristics (Frankel and Romer 1999, Frankel and Rose 2002)

[3] These include cross-country instruments based on political ties, economic policies, and country size (Burnside and Dollar 2000). Boone (1996) also uses instruments based on political ties and country size in related work examining the impact of aid on investment growth.

a general phenomenon. We consider a range of regressions that recently passed the rigors of peer review and were accepted for publication in general-interest journals and top field journals. First, we discuss instruments around which have grown sizeable subliteratures, but which are shown to be invalid in each of their applications *by each of their other applications.* We examine one such instrument in detail. Second, we show that many important applications of dynamic panel GMM to growth empirics are founded on weak instruments, and are therefore incapable of testing the hypotheses they wish to explore.

We do not suggest a return to nihilism. Here we advocate three ways that growth empirics can surmount these difficulties: by basing instrumental variable regressions on theory sufficiently general to comprise other published results with the same instrument, by using the latest methods to probe sensitivity to violations of the exclusion restriction, and by opening the "black box" of GMM with complementary methods to assess instrument strength. We discuss each in detail below.

2 When strong instruments are invalid

To pass a rigorous peer review, each growth study employing an instrumental variable offers theoretical and empirical reasons to believe that the instrument is not substantially correlated with the regression's error term. It is well known that this is difficult to establish. There can be a multiplicity of theoretical arguments for and against any given exclusion restruction, the true error term is unobserved in all applied settings, and empirical tests of overidentifiying restrictions have notoriously low power—among other reasons. What is not as well known is that collectively the literature establishes the invalidity of some instruments that growth econometricians now use widely, calling into question broad classes of their findings.

Suppose that growth is determined by

$$g = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon, \tag{1}$$

where g is growth, the x_i are a set of k potentially endogenous determinants of growth, the β are parameters to be estimated and ε is an error term. Suppose we have an instrumental variable z such that $E[z\varepsilon] = 0$ but $E[zx_i] \neq 0\ \forall i$. We now try to estimate k separate regressions

$$g = \beta_0 + \beta_i x_i + \varepsilon_i,\ i = 1, ..., k \tag{2}$$

in each case instrumenting for x_i with z. But unless for every i it is the case that $\beta_j = 0\ \forall j \neq i$, we have $E[z\varepsilon_i] \neq 0\ \forall i$, and the instrument z is invalid in *every* regression (2). In other words, if existing research has shown that z is a strong instrument for a variable x_j not included in a regression of the form (2) and $\beta_j \neq 0$, then z need not be a valid instrument for x_i.

In this case, in the terminology advocated by Deaton (2009), z is "external" but it is not "exogenous". Any estimate $\widehat{\beta}_i$ will be biased to an unknown degree in an unknown direction, throwing into question the credibility of all results from the regressions (2). As Durlauf, Johnson, and Temple (2005: 639) point out, "Since growth theories are mutually compatible, the validity of an instrument requires a positive argument that it cannot be a direct growth determinant or correlated with an omitted growth determinant".

2.1 Original sins

A prominent example is the widespread use of "legal origins" in growth regressions, a practice that has become the subject of frequent grumbling at conference coffee breaks. A flotilla of recent cross-country growth regressions has employed an indicator of the origin of a country's legal system (British, French, Scandinavian, and so on) as an instrument in a variety of regression specifications—each one of which suggests that the instrument is invalid in all of the other specifications. These studies have passed the rigors of peer review at general-interest journals and top field journals.

Friedman, Johnson, Kaufmann, and Zoido-Lobatón (2000) use legal origin as an instrument for five separate measures of the "the quality of economic institutions" (corruption, tax rates, over-regulation, etc.) in regressions with the size of the unofficial economy as the dependent variable—which could directly affect growth. Djankov, La Porta, López-de-Silanes, and Shleifer (2003) use legal origin as an instrument for "the degree of formalism of the legal procedure", which they argue causes a decline in the quality of the legal system (its honesty, impartiality, ability to enforce contracts, and so on) that could be a major determinant of growth. Lundberg and Squire (2003) use legal origin as an instrument for inflation, the inequality of land ownership, and several other variables that they argue directly affect growth. If any two of these studies are correct, growth is determined by a form of equation (1) that renders instrumentation in the IV regressions (2) invalid.

It does not stop there. Alfaro, Chanda, Kalemli-Özcan, and Sayek (2004) use legal origin as an instrument for private sector credit, bank credit, and stock market capitalization, which they argue condition the effect of Foreign Direct Investment on growth. Levine, Loayza and Beck (2000) similarly use legal origin to instrument for three separate proxies for financial intermediation, all of which they argue cause economic growth. Glaeser, La Porta, López-de-Silanes, and Shleifer (2004) use legal origin as an instrument for "executive constraints" and average years of schooling in the population, with the level of income per capita as the dependent variable. Beck, Demirgüç-Kunt, and Levine (2005) use legal origin as an instrument for "the relative size of the small and medium enterprise sector", which could be associated with growth. There are other examples.

The findings of these studies suggest that 1) at least one of them specifies the second stage correctly, since they reject the null of zero coefficients in the second stage, but therefore 2) instrumentation can be valid in at most one of them, and at worst none.

2.2 Size matters—through various channels

We turn to another instrument in widespread use, and dwell on it at greater length because its problems are less broadly recognized. Several recent cross-country studies published in general-interest journals and top field journals rest their identification strategies on the correlation of populaton size with some endogenous variable. In each case, the authors give plausible reasons why population size is not only a strong instrument but uncorrelated with their regressions' error terms: the fact that growth regressions do not generally find population scale effects (Rose 2006; Easterly 2009).

When viewed collectively, however, these studies exhibit a problem that confutes their careful arguments in support of instrument validity: Given that none of these studies include the other studies' endogenous variables as regressors, if population size is a strong and valid instrument in even one of these studies, then it is invalid in *all* of the others. In other words, Deaton's (2009) conjecture that measures of country size can affect growth through multiple channels has empirical support.

This pattern emerges in several recent and prominently published regressions. Some investigators use population size (among other geographic characteristics) as an instrument for trade, as a determinant of the level of income per capita (Frankel and Romer 1999; Frankel and Rose 2002)[4] or its growth (Spolaore and Wacziarg 2005). Others regress growth not on the level of trade but on an indicator of the mix of goods exported, instrumented by population size (Hausmann, Hwang, and Rodrik 2007), without controlling for the level of trade. Still others use population size as an instrument to identify the effect of foreign aid on democracy (Djankov, Montalvo, and Reynal-Querol 2009), which many studies find to correlate with growth in some fashion.[5] Another approach uses country size—measured by area and level of GDP, but strongly correlated with population—to instrument for receipts of foreign direct investment (FDI) as a determinant of growth (Borensztein, de Gregorio, and Lee 1998).

The exclusion restriction necessary for population size to be a valid instrument for each of these endogenous variables is falsified by each of the other studies. Regardless of any theoretical and empirical case for instrument validity made by each paper in the group, population size can only be a strictly valid instrument in one of them at best, and none of them at worst. The degree to which each estimate is thereby biased could be small or large, but should not be ignored.

[4] The Frankel and Romer (1999) instrument actually contains information beyond country size. While controlling for log population and area in the structural equations of tables 2 and 3, Frankel and Romer demonstrate that their instrument remains strong with a F statistic over 10. However, upon more rigorous examination of the exclusion restrictions implicit in this instrument, Frankel and Rose (2002) conclude that among the six plausibly exogenous geographic determinants of trade flows used to construct their predicted trade instrument, log population is the only one that violates the implicit overidentifying restrictions used in constructing the instrument. See footnote 15 of Frankel and Rose (2002). This result supports our claims in this section about the non-excludability of size.

[5] For investigations of the effect of democracy on growth, see for example Barro (1996), Tavares and Wacziarg (2001), Giavazzi and Tabellini (2005), Rodrik and Wacziarg (2005), Persson and Tabellini (2006), Persson and Tabellini (2007), and Papaioannou and Siourounis (2008).

2.3 Strength in numbers, but not validity

The problem extends further than this, however, in a way that is not generally recognized. Many studies recur to multiple instruments, responding to criticism by pointing out that allegations of invalidity or weakness only apply to some of the instruments. It is common to gloss over the problem that the most valid instruments in the basket could be the weakest, and that the strongest could be the least valid.

Building on the above discussion of the population size instrument, it is possible for a study whose identification strategy appears to rest on an array of instruments to rely in fact entirely on population size. In work recently published in a general-interest journal, Rajan and Subramanian (2008) execute cross-section regressions of growth on foreign aid receipts, with aid instrumented by a complicated instrument constructed from aid-recipient population size, aid-donor population size, colonial relationships, and language traits.[6] Rajan and Subramanian assert, "Our instrument ... contains information that is not just based on recipient size" (footnote 16).[7]

[6] Rajan and Subramanian construct their instrument in a "zero-stage" specification by regressing bilateral aid flows as a fraction of recipient GDP on recipient and donor characteristics. They use the resulting coefficients to calculate predicted bilateral aid flows. They sum these predicted bilateral flows across donors to arrive at predicted total aid receipts for each recipient country as a fraction of recipient GDP. This predicted total, a constructed instrument for true aid receipts, becomes the excluded instrument in a series of two-stage least squares regressions of economic growth on aid receipts and a set of control variables. The instrument is: $a_{dr} \equiv \frac{A_{dr}}{Y_r} = \sum_{i=0}^{7} \beta_i I_{i,dr} + \sum_{i=0}^{5} \beta_{i+8} (\ln P_d - \ln P_r) I_{i,dr} + \upsilon_{dr}$, where A_{dr} is dollars of aid given by donor d to recipient r, Y_r is the GDP of r, β_0 through β_{13} are regression coefficients, P_d is donor-country population, and P_r is recipient-country population. The I's are a set of time-invariant country dummy variables describing the country dyad: a current or past colonial relationship (I_1); a current or past colonial relationship with the United Kingdom (I_2), France (I_3), Spain (I_4), or Portugal (I_5); common language (I_6); and a current colonial relationship (I_7). Finally, $I_{0,dr} = 1 \ \forall \ d, r$ and υ_{dr} is an error term. The estimated coefficient vector $\widehat{\beta}$ is then used to generate predicted bilateral flows $\overline{a}_{dr}$, which are summed across donors to create the constructed instrument $\overline{a}_r = \sum_d \overline{a}_{dr}$, which then instruments for aid receipts $a_r \equiv A_r/Y_r$ in the cross-section growth regression $g_r = \gamma_1 a_r + X_r' \Theta + u_r$, where g_r is real GDP per capita growth, X_r is a vector of country characteristics, γ_1 is a regression coeffcient, Θ is a vector of regression coefficients, and u_r is an error term.

[7] They justify this claim (in their table 5, panel C) by using one measure of country size (population) as an implicitly excluded instrument in the construction of $\overline{a}_r$ and, in a robustness check, showing that the constructed instrument $\overline{a}_r$ retains strength when a different measure of country size (land area) is included as an additional explicitly excluded instrument. But the only way to accurately assess whether or not $\overline{a}_r$ contains information beyond population size is to test whether or not it retains significance when population itself is included as a separate instrument as Frankel and Romer (1999) do (see footnote 4 above), and as we do below.

But the instrument contains, in fact, almost no information beyond the size of the recipient's population. In Rajan and Subramanian's data, for the period 1970-2000, the in-sample correlation of log population and the constructed instrument is -0.93. In the periods 1980-2000 and 1990-2000, this correlation is -0.95. In effect, Rajan and Subramanian are instrumenting for aid with population alone.

This problem deserves additional discussion, since it is common in applied work to rest identification on a group of instruments without making explicit which of them bears the burden of identification and therefore the key burden of validity. Frankel and Romer (1999) demonstrate that their geography-based instrument contains information beyond country size by including log population and log area as additional instruments in the second-stage.[8] Taking this minimalist approach, we explore in tables 1 and 2 the role of population as an instrument using the original data of Rajan and Subramanian.

Table 1 shows that all instrumentation power comes from the population instrument. Column 1 exactly reproduces a representative cross-section regression (Rajan and Subramanian Table 4, column 2). Instrumentation is very strong, as indicated by the Cragg and Donald (1993) F and Kleibergen and Paap (2006) rk statistics at the bottom of the table.[9] Column 2 includes log population in the second stage, and instrument strength collapses.[10] Column 3 discards Rajan and Subramanian's

[8] Although their approach did not go far enough to convince some skeptics (Rodríguez and Rodrik 2001, Kraay 2008), the validity of their instrument and robustness of their results have, in fact, withstood more demanding specifications and overidentifying restrictions (Noguer and Siscart 2005).

[9] Because the authors use heteroskedasticity-robust standard errors in all regressions, use of the Kleibergen-Paap statistics is more appropriate; it generalizes the Cragg-Donald statistic to the case of non-i.i.d. errors, allowing for heterskedasticity, autocorrelation and/or cluster robust statistics (Baum, Schaffer and Stillman 2007). In the special case of a single endogenous regressor considered here, though, the Cragg-Donald and Kleibergen-Paap F statistics are respectively simply the standard non-robust and robust first-stage F statistics. The results of section 3 below rely on versions of both of these statistics that generalize to the case of more than one endogenous regressor (see Stock and Yogo 2005 for an extended discussion).

[10] According to Stock and Yogo (2005), the Cragg-Donald F statistic must exceed 5 if we are to be confident at the 5% level that the bias to the coefficient estimate on the aid variable is less than 30% of the OLS bias. Critical values have not been tabulated for the Kleibergen-Paap F statistic since the specific thresholds depend on the type of violation of the i.i.d. assumption, which invariably differ widely across applications. Nevertheless, standard first-stage F statistics well below unity are unlikely to exceed even the most generously low thresholds for unbiasedness. Additionally, we can rely on the Kleibergen-Paap Lagrange-Multiplier test of *under*identification–a generalization of the Anderson canonical correlations test to non-i.i.d. errors for which one can utilize standard hypothesis testing procedures based on the $\chi^2_{z_1-x_1+1}$ distributed statistic where z_1 is the number of excluded

constructed instrument altogether and uses log population alone as an instrument for aid, giving results nearly identical to those in column 1. Table 2 shows only the first-stage F statistics from the other Rajan and Subramanian cross-section regressions: first in exact replication of their results, then with population deleted from the construction of their instrument, then with the instrument constructed based only on population and its interactions. In all cases, aid is weakly instrumented when information about population is absent from the constructed instrument, and strongly instrumented when only those variables containing information about population are present.

The Rajan and Subramanian cross-section method is indistinguishable from instrumenting exclusively with aid-recipient population. Their discussion of the validity of any other variable in the instrument matrix, then, is irrelevant. What matters is the validity of the instrument that strongly identifies causation. Since that is only country size, the Rajan and Subramanian analysis faces the same problem faced by the other papers resting on the population instrument: All of the aforementioned papers that use the population instrument invalidate its usage in these studies, since the regressions there do not control for the level of trade, the mix of goods exported, FDI, or democracy. And the Rajan and Subramanian exercise does not resolve important questions about the validity of the population instrument in all of the other papers that use it because those papers do not control for aid receipts in the second stage.

This problem extends beyond straightforward cross-section models to dynamic panel regressions as well. As an example, we consider the 10 year panel regressions in Hausmann, Hwang, and Rodrik (2007). The authors utilize two estimators: a pooled 2SLS estimator with log population and log area as instruments, and the Blundell and Bond (1998) dynamic panel GMM estimator with instrumental variables that include log population and log area as well as the standard set of lagged covariates employed in this popular estimation strategy (see section 3 for a detailed discussion of the estimator). Using the original data of Hausmann, Hwang and Rodrik, Table 3 demonstrates how assumptions about the excludability of country size can drive

instruments and x_1 is the number of endogenous regressors. Under the null hypothesis, the equation is underidentified, which offers, in general, a much lower hurdle than the weak identification tests of Stock and Yogo (2005). With a p-value of 0.77, this test in column 2 fails to reject the null.

identification of key parameters of interest even in a dynamic panel setting with numerous non-size-based instruments. Columns 1 and 2 replicate the results from Table 9, columns 6 and 8 respectively of Hausmann, Hwang and Rodrik (2007).[11] Column 3 removes log population and log area from the difference equation instrument matrix, which leaves instrument validity and inference largely unchanged. Column 4 removes log population and log area from the levels equation instrument matrix, while column 5 removes the size variables from the instrument matrices in both equations. Column 6 relaxes the assumption that country size is excludable and the results are similar to the preceding columns which both dropped the size instruments from the levels equation. Despite the wide array of plausibly valid internal instruments in lagged levels and differences, the identification of interest depends solely upon the excludability of country size from the structural equation in levels, which closely corresponds to the overidentified linear, pooled 2SLS specification in column 1. Simple comparisons of the Hansen overidentification test (J test p-value) across the six specifications in this table provide further evidence towards this conclusion.

That country size merely strongly identifies cross-sectional variation in trade composition frustrates interpretation of the causal, time-varying covariance among the endogenous variables of interest. The dynamic causal mechanism in this scenario is reduced to a pooled model as in Boone's (1996) early use of log population as an instrumental variable for foreign aid receipts in a 2SLS regression with investment growth as the outcome of interest.[12] Hausmann, Hwang and Rodrik (2007) point out this precise problem in their pooled 2SLS regression, but the paper does not indicate that the additional instruments employed in the GMM estimator might not solve the problem.[13] Despite the large number of non-size-based instruments in the Blundell and Bond (1998) estimator, the short 10 year panel in Hausmann, Hwang and Rodrik (2007), like their cross-sectional counterparts in Rajan and Subramanian (2008),

[11] Despite utilizing their exact *Stata* code and original dataset, the system GMM replication in column 2 differs slightly from the published results.

[12] Much like Rajan and Subramanian (2008), country size proved to be the sole source of identification in Boone's (1996) main 2SLS (and cross-section) results. Boone's other instruments, political proxies for relationships with major donor countries, are extremely weak predictors of aid flows. Results available upon request.

[13] "The variables used as instruments [log population and log area] fail the overidentification test in columns (2) and (6) [pooled 2SLS], most likely because they are persistent series akin to country fixed effects in a panel. Reassuringly, columns (4) and (8) show that the GMM setup where lagged levels and differences are used as instruments passes both the overidentification test and exhibits no second order correlation"(Hausmann, Hwang, and Rodrik 2007, footnote 9).

depends crucially on size-based instruments shown invalid in this setting by other studies.

We can go beyond the mere suspicion that residuals in some of these studies are correlated with the endogenous variables in the other studies. Table 4 shows this within the Rajan and Subramanian framework. Here we perform a series of OLS regressions, each with a candidate growth determinant on the left-hand side that has been omitted from the Rajan and Subramanian regressions. The right-hand side variables in each case are the second-stage regressors used by Rajan and Subramanian, plus log population. The table reports the point estimate and standard error for the coefficient on log population in each case. Log population has a statistically significant partial relationship with several variables that are plausible growth determinants. These include trade (Frankel and Romer 1999), foreign direct investment (Borensztein, de Gregorio, and Lee 1998), education expenditure (Bosworth and Collins 2003), inequality (Forbes 2000), government consumption (found to correlate with country size by Alesina and Wacziarg 1998, and acknowledged as a robust growth determinant by Doppelhoffer, Miller, and Sala-i-Martin 2004), alongside multiple others.[14]

3 When valid instruments are weak

So far we have discussed cases of strong instruments whose invalidity is difficult to detect. We flip now to cases of plausibly valid instruments whose weakness is difficult to detect.

The advent of GMM estimators in panel growth regressions has been something of a ray of hope for growth empiricists. These estimators take advantage of a much larger array of exclusion restrictions than does two-stage least squares: The dynamic panel estimator of Arellano and Bond (1991) instruments for current-period differences in endogenous variables with a matrix of multiply lagged levels of predetermined

[14] A further complication arises when one considers relaxing the assumption of linearity in the endogenous variables of interest. Suspend disbelief and assume that population size is a valid instrument for one and only one of the abovementioned studies, in Rajan and Subramanian (2008), for example. Since this instrument has been shown to provide the sole source of identification even in 2SLS regressions with additional instruments, it is implausible and arguably impossible that one could identify endogenous nonlinearities, via specifications including endogenous quadratic, cubic, or interaction terms. See Appendix B.

right-hand side variables, and the related system estimator of Blundell and Bond (1998) additionally instruments for current-period levels with once- or twice-lagged differences. These estimators offer a battery of plausibly valid instruments in most cross-country panels used in growth empirics.[15]

A crucial question goes unexplored in many applications of this new econometric technology: How much of the variance in the endogenous variables is explained by the instruments? A standard test for weak instruments in dynamic panel GMM regressions does not currently exist, so measuring instrument strength empirically is nontrivial.[16] Skeptical researchers have been primarily concerned with biases stemming from weak instruments in the Arellano and Bond "difference" estimator and violations of the exclusion restrictions in the Blundell and Bond "system" estimator.[17] What most have failed to address, however, is the potential for weak instruments in system GMM, which, although generally more robust to weak instruments than the difference estimator, can also suffer from serious weak instrument biases (Bun and Kiviet 2006, Bun and Windmeijer 2007, and Hayakawa 2007). In practice, many applications of these estimators, and particularly the system estimator, simply assume that they provide strong instrumentation.

Weak instruments could seriously affect inference in system GMM estimation of empirical models that are particularly common in the cross-country growth literature, namely short panels with a small number of cross-sectional units and large unobserved heterogeneity (Bun and Windmeijer 2007). The finite-sample biases of the system estimator are a function of a) the ratio of the number of instruments to the number

[15] Bond, Hoeffler, and Temple (2001) characterize the appropriateness of these exclusion restrictions in the particular context of estimating the canonical Solow growth model. Caselli, Esquivel and Lefort (1996) and Levine, Loayza and Beck (2000) were respectively the first to employ the Arellano and Bond (1991) and the Blundell and Bond (1998) estimators in the empirical growth literature.

[16] See Stock and Wright (2000) and Stock, Wright and Yogo (2002) on why the weak instrument diagnostics for linear IV regression do not carry over to the more general setting of GMM.

[17] Bobba and Coviello (2007), for example, demonstrate that the null result in Acemoğlu, Johnson, Robinson and Yared (2005) is reversed upon augmenting the weakly instrumented difference estimator with the levels equation in the system estimator. The implausibly high Hansen J test p-value (1.0) suggests that the system GMM estimates might be subject to invalid exclusion restrictions, the presence of which could be masked by the use of "too many instruments" in the full instrument matrix; see Roodman (2009b). Roodman also demonstrates the sensitivity of the results of Levine, Loayza and Beck (2000) to the exclusion of lagged differences in the levels equation of the system estimator run on smaller instrument matrices.

of cross-sectional units, b) number of moments exploited, c) the autoregressive coefficients on all lagged covariates in the "first-stage" regressions, and d) the ratio of the variance of the time-invariant country effects to the variance of the idiosyncratic shocks. Careful examination of the underlying data-generating process, then, might yield enough insight to characterize the magnitude of weak instruments bias in this increasingly commonly used and less-often criticized GMM estimator.

Below we test instrument strength in two influential sets of growth regressions based on system GMM and recently published in top field and general-interest journals. We follow a simple approach to assessing instrument strength in dynamic panel GMM regressions advanced heuristically in various settings by Blundell and Bond (2000), Roodman (2009a) and Dollar and Kraay (2003) and analytically by Bun and Windmeijer (2007) and Hayakawa (2007). We construct the Holtz-Eakin, Newey, and Rosen (1988) GMM instrument matrix used in the difference estimator and in the levels equation of the system estimator, and carry out the corresponding regressions using two-stage least squares. This allows a simple and transparent test of instrument strength in a closely related setting. If instrumentation of contemporaneous differences by once, twice or multiply lagged levels is weak, and instrumentation of contemporaneous levels by lagged differences is weak, this casts great doubt on the ability of GMM estimators to produce strong identification as used in these settings.

Extending this straightforward approach advanced by Bun and Windmeijer (2007) for the case of a single endogenous variable,[18] we examine particularly whether the additional moment conditions used in system GMM are actually strong enough to compensate for the well-established weak instruments problem in difference GMM estimation of growth models (Bond, Hoeffler and Temple 2001).

[18] We appeal to the results of Blundell, Bond, and Windmeijer (2000) and Stock and Yogo (2002) in justifying our extension of the Bun and Windmeijer analytics to the case of multiple endogenous regressors. In particular, we do not examine the quality of identification in the individual "first-stage" GMM regressions in isolation, but rather, we rely on the Cragg-Donald matrix version of the F statistic to test whether the instruments jointly explain enough variation in the multiple endogenous regressors to conduct meaningful hypothesis tests of causal effects.

3.1 Financial intermediation: Abundant instruments versus strong instruments

Table 5 employs this method to revisit the panel GMM results of Levine, Loayza, and Beck (2000) using the original data.[19] Column 1 reproduces a representative regression of growth on "liquid liabilities" (their Table 5, column 1). Column 2 gives the results of the closest reproduction of this regression we could achieve using the authors' dataset, and the results match relatively well.[20] Column 3 carries out the same regression using simple pooled OLS. In columns 4 and 5, we purge the country fixed effects from the regression by first-differencing (FD) and within-transformation (FE), the OLS analogues to the difference and levels equations. Theoretical evidence on dynamic panel bias (Nickell 1981, Hsiao 1986) suggests that pooled OLS and fixed effects OLS should produce respectively the upper and lower bounds for a consistent point estimate on the lagged dependent variable. Bond (2002) specifically shows why consistent system or difference GMM estimates should lie squarely within these theoretical bounds.

The fact that the point estimate on lagged GDP per capita in the replicated regression of column 2 lies slightly above the pooled OLS estimate in column 3 suggests that the system estimator may be producing upwardly biased results (though this is not the case in the published regressions). While weak instruments typically bias difference estimates downward, Bun and Windmeijer (2007) demonstrate how system estimates are generally biased upward with the biases increasing in the ratio of the variance of the time-invariant heterogeneity to the idiosyncratic shocks. In column 5, that ratio of variances is nearly 25, which is many orders of magnitude larger than the one-to-one ratio upon which Blundell and Bond (1998) predicated the system esti-

[19] This paper conducts similar regressions with three different endogenous measures of financial intermediation: "liquid liabilities" (currency plus demand and interest-bearing liabilities of banks and non-bank financial intermediaries) as a fraction of GDP; "commercial-central bank" (assets of deposit money banks divided by assets of deposit money banks plus central bank assets); and "private credit" (credit by deposit money banks and other financial institutions to the private sector as a fraction of GDP). Here we analyze the results for liquid liabilities alone, but we conducted the same analysis for all three variables with substantively identical conclusions.

[20] This replication is due to Roodman (2009b). It remains unclear why replications of the Levine, Loayza, and Beck (2000) regressions do not match more closely, especially considering that all replications employ the original data provided by the authors (see Appendix A). Some of the discrepancy could be due to differences in the estimator as deployed in the original *Gauss* program `DPD98` used by the authors relative to the `xtabond2` program for *Stata*, which we utilize here.

mator's consistency. Column 6 regresses differenced growth on differenced regressors, instrumented by lagged regressor levels to correspond to the difference estimator. Both the Cragg-Donald F and Kleibergen-Paap F statistics and the lower-hurdle Kleibergen-Paap LM test of underidentification (see footnote 10) show that instrumentation is very weak, far too weak for instrumentation to remove a substantial portion of OLS bias.

An additional problem lurks below the surface: The sample contains 77 countries, and 75 different instrumental variables are used in the system estimator.[21] The large number of instruments may actually result in a failure to expunge the endogenous components of the right-hand side variables, thereby biasing the coefficient estimates towards those from the OLS estimator (Roodman 2009b). In the limiting case, a 2SLS regression that had one instrument for each observation would show strong instrumentation but would produce coefficients exactly equal to those produced by OLS, and would not address endogeneity bias at all. Until recently the literature has offered little guidance on the appropriate number of instruments relative to the number of groups and time periods.[22]

Roodman (2009b) introduces a practical method for addressing this problem of "too many instruments" in dynamic panel GMM estimation. He suggests first restricting the number of lagged levels used in the instrument matrix for the difference equation,[23] but since Levine, Loayza and Beck (2000) restrict their original matrix to a single lag, we must try an alternative approach. By "collapsing" the instrument matrix, we can effectively combine the instruments into smaller sets while retaining the same information from the original 75 column instrument matrix.[24] Roodman suggests that a liberal rule of thumb for identifying potential cases of "too many instruments" is to become concerned when the number of instruments is close to the number of groups, as in the present case. Column 7 shows the results with the

[21] In both the levels and difference equations, 35 lagged regressors are used as instrumental variables, as well as the 5 period dummies included in the structural equation.

[22] See Han and Phillips (2006) for a thorough treatment of the asymptotic theory of many weak instruments in the general GMM setting.

[23] As conventionally applied in the empirical literature, the instrument matrix for the levels equation in the system estimator contains only one lagged difference for each endogenous variable in levels as additional moments would be redundant (Arellano and Bover 1995)

[24] The "collapsed" matrix contains one instrument for each lag of the instrumenting variable instead of one instrument for each period *and* lag of the instrumenting variables.

instrument matrix collapsed; instrumentation predictably becomes even weaker. Instrumentation this weak—no matter how valid—is incapable of testing hypotheses about coefficients in the main regression.

To test for weak instruments in the system estimator, we must also examine the levels equation independently of but in the same manner as the difference equation treated in preceding columns. Columns 8 and 9 conduct this parallel exercise for the levels equation. Since the difference equation is so weakly instrumented, the burden of strong identification in the system estimator relies on the levels equation moments. In column 8, the level of growth is regressed on the level of the regressors in a two-stage least squares framework, instrumented by the same lagged differences as in the levels equation of the system GMM estimator. Once again, instrumentation is far too weak to address any substantial portion of OLS bias, and when the instrument matrix is collapsed in column 9 the problem only worsens.

3.2 Weak aid or weak instruments?

Table 6 repeats this analysis for an entirely different set of regressions. It revisits the dynamic panel results of Rajan and Subramanian (2008) using the original data. Columns 1 and 2 exactly replicate their main Arellano-Bond (Table 9, column 1) and Blundell-Bond (Table 10, column 1) results.[25] Column 3 shows the simple pooled OLS result, which appears remarkably similar to the system estimate in the preceding column. Columns 4 and 5 purge country fixed effects from the regression in column 3 via first-differencing (FD) and within-transformation (FE), bringing the results close to those in the difference estimator in column 1. This per se is suggestive evidence

[25] Rajan and Subramanian include the second through seventh lags as instruments in both specifications. They note that they are employing up to eight lags, but given that their panel consists of eight periods and only the four five year periods since 1985 are actually used due to missing data on their institutional quality measure, their difference and system GMM specifications naturally do not include eighth lagged levels as instruments for any of the endogenous regressors. Also, although they claim to include an additional set of time-invariant, excluded instruments in their main difference-equation specifications (geography, ethnic fractionalization, Sub-Saharan Africa and East Africa) a *Stata* coding error results in their being dropped from the "first- stage" equations regressing differenced endogenous variables on lagged levels. To be consistent with their published results, we exclude these four time-invariant dummies from the Arellano-Bond regression in column 1 and the difference equation in the Blundell-Bond regression in column 2, as well as the 2SLS analogues in subsequent columns. Including them results in an immaterial improvement in the weak identification statistics (results available upon request).

that instrumentation in these panel regressions is too weak to improve on OLS. While the difference GMM point estimate on lagged GDP per capita lies below the fixed effects estimate, the system estimate in column 2 places the coefficient just within the range of plausibly consistent estimation. However, this suggestive evidence still neglects potential problems with weak instruments, which could manifest differently than traditional dynamic panel biases. Given that the ratio of the variance of the time-invariant individual effects to the variance of idiosyncratic errors is above 9 in column 5, it is unlikely that the heuristic bounding exercise of Bond, Hoeffler, and Temple (2001) can capture both the necessary and sufficient conditions for consistent estimation within the system GMM framework.

Thus, following the approach above, in column 6 we estimate the difference component of the system estimator in a 2SLS regression with exactly the same sequential moment conditions and exclusion restrictions. The Cragg-Donald F statistic falls well below unity, suggesting substantial finite sample biases, and although the Kleibergen-Paap F statistic appears high, the perceived strength turns out to be a statistical artifact of including up to seven lags. If on the other hand, one simply collapsed the 120 column instrument matrix with up to seven lags, the F statistics fall even lower in column 7.[26] The relatively higher Kleibergen-Paap F statistic of 3.8 suggests that the standard Cragg-Donald test statistic might be sensitive to heteroskedasticity. But since we still cannot reject the null of underidentification based on the Kleibergen-Paap LM test, identification is still too weak to conduct meaningful hypothesis tests based on the difference equation alone. Columns 8 and 9 repeat the same exercise for the levels equation component of the system estimator. Column 8 demonstrates weak instruments in the standard wide instrument matrix, and collapsing exacerbates the problem still further. These results suggest that the similarity between the biased OLS estimates in columns 3-5 and the dynamic panel GMM estimates in columns 1 and 2 is not a coincidence.

In cross-country dynamic panel growth regressions with multiple endogenous co-

[26] One could also solely or additionally reduce the size of the instrument matrix by simply restricting the number of lags as in Levine, Loayza and Beck (2000). Restricting the number of lags to one or two produces Kleibergen-Paap F statistics below unity, and we strongly fail to reject underidentification. These results are available upon request. Although Rajan and Subramanian note that restricting the lag depth in the GMM estimation of columns 1 and 2 does not affect their null results, these simple test statistics constructed from analogous 2SLS procedures point to more fundamental specification problems.

variates, it is unlikely that the weak instruments problem in the Arellano and Bond (1991) estimator can be solved by straightforward recourse to the richer system estimator of Blundell and Bond (1998). The latter estimator is an optimally weighted average of difference and levels equations with the weights on the levels equation moments increasing in the weakness of the difference equation instruments. With a weakly instrumented levels equation, the system estimates can exhibit biases of similar orders of magnitude to uncorrected OLS variants. The findings above are not peculiar to the specifications used in these two studies. Upon finding that the system GMM estimator results are either unchanged or differ (often in the right direction) from the difference estimator, researchers often rest their case without questioning the strength of either estimators' large instrument sets. While the estimation of growth models in a dynamic panel framework is intrinsically appealing, researchers applying these popular GMM estimators should be wary of interpreting not only significant but also null results. In the absence of more rigorous, albeit heuristic, tests such as those advocated above, the hasty application of popular statistical programs for system and difference GMM will not lead to improvements in our understanding of what *causes* growth.

4 Lessons

We demonstrate that invalid and weak instruments continue to be commonly used in the growth literature. This suggests that the warnings of Durlauf, Johnson, and Temple (2005) and others on this subject have gone unheard. Weak and/or invalid instruments do not assist researchers in conducting meaningful hypothesis tests about the causes of growth. Continued use of problematic instruments in the growth literature risks pushing all of its findings further towards irrelevance.

Many of the papers discussed here contain explicit policy implications based on their results, including Levine, Loayza, and Beck (2000) and Rajan and Subramanian (2008). Without strong and valid identification of causal effects, such exercises carry no policy implication. Nevertheless, these studies remain valuable contributions to the literature for other reasons—especially their innovations in method.

Economists will and should pursue pressing research questions on growth while methods remain imperfect. But we suggest a handful of guidelines for the next generation of growth empirics:

1. *Generalize the theoretical underpinnings of an instrument to account for other published results with the same instrument.* When an instrument has been used elsewhere in the literature, new users of that instrument bear the burden of showing that other important findings using that instrument do not invalidate its use in the new case. This can be done using a somewhat more generalized model that comprises causal pathways explored elsewhere with that instrument. Accounting for all plausible pathways through a "unified growth theory" is too high a standard, but accounting for the most prominent published pathways should be a minimum standard.

2. *Deploy the latest tools for probing validity.* Perfect instruments for growth determinants will remain elusive, but many underutilized tools exist to shine brighter light on the instruments we have. Imbens (2003) lays out a transparent method of assessing the sensitivity of a growth effect estimate to a given degree of correlation between instrument and error. Kraay (2008) and Conley, Hansen, and Rossi (2008) explore how to conduct second-stage inference accounting for prior uncertainty about the excludability of the instrument. Ashley (2009) shows how the discrepancy between OLS and IV estimates can be used to estimate the degree of bias under any given assumption about how badly the exclusion restriction holds.

3. *Open the black box of GMM.* It is no longer sufficient to assert that the mere use of system GMM adequately addresses the risk of weak instrumentation. As applied econometricians wait for an analog of the Stock and Yogo (2005) weak instrument diagnostics suitable for dynamic panel GMM estimation, its use must be complemented by supportive evidence that the instruments explain a sufficient degree of the variance of the endogenous regressors (and not simply because so many instruments are used). Papers exploring growth determinants should explore the strength of candidate instruments in analogous two-stage least squares regressions, should explore robustness to Roodman's (2009b) "collapsing" of the instrument matrix, and should explore methods robust to weak instruments.[27]

[27] See, among others, Stock and Wright (2000), Kleibergen (2005), Kleibergen (2007) and Kleiber-

References

Acemoğlu, Daron, Simon Johnson, and James A. Robinson (2001), "The Colonial Origins of Comparative: An Empirical Investigation", 91 (5): 1370-1400.

Acemoğlu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared (2005), "From education to democracy?" *American Economic Review Papers and Proceedings* 95: 44-49.

Albouy, David Y. (2008), "The colonial origins of comparative development: An investigation of the settler mortality data", NBER Working Paper 14130 (Cambridge, MA: National Bureau of Economic Research).

Alesina, Alberto and Romain Wacziarg (1998), "Openness, country size, and government", *Journal of Public Economics* 69 (1): 305-321.

Alfaro, Laura, Areendam Chanda, Şebnem Kalemli-Özcan, and Selin Sayek (2004), "FDI and economic growth: the role of local financial markets", *Journal of International Economics* 64 (1): 89-112.

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

Arellano, Manuel, and Stephen Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies* 58(2): 277-297.

Arellano, Manuel and Olympia Bover (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models", *Journal of Econometrics* 68 (1): 29-52.

Ashley, Richard (2009) "Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis," Journal of Applied Econometrics 24 (2): 325-337.

gen and Mavroides (2009) on weak-instrument robust inference in general GMM settings. While this literature does not directly address the specific dynamic panel estimators employed in the growth literature, their methods can be applied with some effort. For a discussion on the application of some of these approaches, see Baum, Schaffer and Stillman (2007).

Barro, Robert J. (1991), "Economic Growth in a Cross Section of Countries", *Quarterly Journal of Economics* 106 (2): 407-443.

Barro, Robert J. (1996), "Democracy and Growth", *Journal of Economic Growth* 1 (1): 1-27.

Baum, Christopher F., Mark E. Schaffer, and Stephen Stillman (2007), "Enhanced routines for instrumental variables/GMM estimation and testing", *Stata Journal* 7 (4): 465-506.

Baumol, William J. (1986), "Productivity Growth, Convergence, and Welfare: What the Long-run Data Show", *American Economic Review* 76 (5): 1072-1085.

Beck, Thorsten, Aslı Demirgüç-Kunt, and Ross Levine (2005), "SMEs, growth, and poverty: Cross-country evidence", *Journal of Economic Growth* 10 (3): 199-229.

Blundell, Richard and Stephen Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics* 87 (1): 115-143.

Blundell, Richard and Stephen Bond (2000), "GMM Estimation with Persistent Panel Data: An Application to Production Functions", *Econometric Reviews* 19 (3): 321-340.

Blundell, Richard W, Stephen R. Bond and Frank Windmeijer (2000), "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator"in B. Baltagi (ed.) *Advances in Econometrics, Volume 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, (JAI Elsevier Science).

Bobba, Matteo and Decio Coviello (2007), "Weak instruments and weak identification in estimating the effects of education on democracy", *Economics Letters* 96 (2): 301-306.

Bond, Stephen R. (2002), "Dynamic panel data models: a guide to micro data methods and practice", *Portuguese Economic Journal* 1 (2): 141-162.

Bond, Stephen R., Anke Hoeffler and Jonathan Temple (2001), "GMM Estimation of Empirical Growth Models", CEPR Discussion Paper No. 3048.

Boone, Peter (1996), "Politics and the effectiveness of foreign aid", *European Economic Review* 40 (2): 289-329.

Borensztein, Eduardo, Jose De Gregorio, and Jong-Wha Lee (1998), "How does foreign direct investment affect economic growth?" *Journal of International Economics* 45 (1): 115-135.

Bosworth, Barry P. and Susan M. Collins (2003), "The empirics of growth: An update", *Brookings Papers on Economic Activity* 2: 113-206.

Brock, William A. and Steven N. Durlauf (2001), "Growth empirics and reality", *World Bank Economic Review* 15 (2): 229-272.

Bun, Maurice J. G. and Frank Windmeijer (2007), "The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models", Discussion Paper No. 07/595 (Bristol, UK: University of Bristol Dept. of Economics).

Bun, Mauric and J.F. Kiviet (2006), "The Effects of Dynamic Feedbacks on LA and MM Estimator Accuracy in Panel Data Models", *Journal of Econometrics* 132 (2): 409-444.

Burnside, Craig, and David Dollar (2000), "Aid, Policies, and Growth", *American Economic Review* 90 (4): 847-868.

Caselli, Francesco, Gerardo Esquivel, and Fernando Lefort, (1996), "Reopening the Convergence Debate: A New Look at Cross Country Growth Empirics", *Journal of Economic Growth* 1 (3): 363-389.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2008), "Plausibly exogenous," Working Paper, Booth School of Business (Chicago: University of Chicago).

Cragg, John G. and Stephen G. Donald (1993), "Testing identifiability and specification in instrumental variable models", *Econometric Theory* 9(2): 222-240.

Dalgaard, Carl-Johan, Henrik Hansen, and Finn Tarp (2004), "On the empirics of foreign aid and growth", *Economic Journal* 114 (496): 191-216.

Deaton, Angus S. (2009), "Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development" Working Paper 14690 (Cambridge, MA: National Bureau of Economic Research).

Djankov, Simeon, Rafael la Porta, Florencio López-de-Silanes, and Andrei Shleifer (2003), "Courts", *Quarterly Journal of Economics* 118 (2): 453-517.

Djankov, Simeon, Jose G. Montalvo, and Marta Reynal-Querol (2008), "The curse of aid",*Journal of Economic Growth* 13 (3): 169-235.

Dollar, David and Aart Kraay (2003), "Institutions, trade, and growth", Journal of Monetary Economics 50 (1): 133-162.

Durlauf, Steven N., Paul A. Johnson, and Jonathan R. W. Temple (2005), "Growth Econometrics", in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1B (Amsterdam: Elsevier).

Easterly, William (2009), "Can the west save Africa?" *Journal of Economic Literature*, forthcoming.

Feyrer, James, and Bruce Sacerdote (forthcoming), "Colonialism and Modern Income", *Review of Economics and Statistics.*

Frankel , Jeffrey, and David Romer (1999), "Does Trade Cause Growth?"*American Economic Review* 89 (3): 379 Ð399.

Frankel, Jeffrey, and Andrew Rose, (2002) "An Estimate of the Effect of Common Currencies on Trade and Income", *Quarterly Journal of Economics* 117 (2): 437-466.

Friedman, Eric, Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobatón (2000), "Dodging the grabbing hand: the determinants of unofficial activity in 69 countries," *Journal of Public Economics* 76 (3): 459-493.

Forbes, Kristin (2000), "A reassessment of the relationship between inequality and growth", *American Economic Review* 90 (4): 869-887.

Giavazzi, Francesco and Guido Tabellini (2005), "Economic and Political Liberalizations", *Journal of Monetary Economics* 52 (2): 1297-1330.

Glaeser, Edward L., Rafael La Porta, Florencio López-de-Silanes, and Andrei Shleifer (2004), "Do Institutions Cause Growth?" *Journal of Economic Growth* 9 (3): 271-303.

Hall, Robert and Charles I. Jones (1999), "Why Do Some Countries Produce So Much More Output per Worker Than Others?"*Quarterly Journal of Economics* 114 (1): 83-116.

Han, Chirok and Peter C.B. Phillips (2006), "GMM with Many Moment Conditions", *Econometrica* 74 (1): 147-192.

Hansen, Henrik and Finn Tarp (2001), "Aid and growth regressions," *Journal of Development Economics* 64(2): 547-570.

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

Hayakawa, Kazuhiko (2007), "Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models", *Economics Letters* 95 (1): 32-38.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988), "Estimating vector autoregressions with panel data", *Econometrica* 56 (6): 1371-95.

Kleibergen, Frank (2005), "Testing Parameters in GMM Without Assuming That They Are Identified,"*Econometrica* 73 (4): 1103-1123.

Kleibergen, Frank (2007), "Generalizing Weak Instrument Robust IV Statistics towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics."*Journal of Econometrics* 139, 181-216.

Kleibergen, Frank and Sophocles Mavroeidis (2009), "Inference on Subsets of Parameters in GMM Without Assuming Identification "mimeo, Brown University.

Imbens, Guido W. (2003), "Sensitivity to exogeneity assumptions in program evaluation," American Economic Review 93 (2): 126-132.

Jones, Ben, and Benjamin Olken (2005), "Do Leaders Matter? National Leadership and Growth Since World War II", *Quarterly Journal of Economics* 120 (3): 835-864.

Kleibergen, Frank and Richard Paap (2006), "Generalized reduced rank tests using the singular value decomposition", *Journal of Econometrics* 133(1): 97-126.

Kraay, Aart (2008), "Instrumental Variables Regressions with Honestly Uncertain Exclusion Restrictions", *World Bank Policy Research Working Paper* No. 4632.

Levine, Ross, Norman Loayza, and Thorsten Beck (2000), "Financial intermediation and growth: Causality and causes", *Journal of Monetary Economics* 46 (1): 31-77.

Lindauer, David L. and Lant Pritchett (2002), "What's the big idea? The third generation of policies for economic growth", *Economía* 3 (1): 1-39.

Lundberg, Mattias and Lyn Squire (2003), "The simultaneous evolution of growth and inequality", *Economic Journal* 113 (487): 326-344.

Mankiw, N. Gregory (1995), "The growth of nations", *Brookings Papers on Economic Activity* 1995 (1): 275-310.

Mauro, Paulo (1995), "Corruption and growth", *Quarterly Journal of Economics* 110 (3): 681-712.

Murray, Michael P. (2006), "Avoiding invalid instruments and coping with weak instruments", *Journal of Economic Perspectives* 20 (4): 111-132.

Nickell, Stephen (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica* 49 (4): 1417-1426.

Noguer, Marta, and Marc Siscart (2005), "Trade raises income: a precise and more robust result", *Journal of International Economics* 65 (2): 447-460.

Papaioannou, Elias and Gregorios Siourounis (2008), "Democratisation and growth", *Economic Journal* 118 (532): 1520-1551.

Persson, Torsten and Guido Tabellini (2007), "The Growth Effect of Democracy: Is It Heterogenous and How Can It Be Estimated?"NBER Working Paper 13150 (Cambridge, MA: National Bureau of Economic Research).

Persson, Torsten and Guido Tabellini (2006), "Democracy and Economic Development: the Devil Is In the Details", *American Economic Review Papers and Proceedings* 96: 319-324.

Rajan, Raghuram and Arvind Subramanian (2008), "Aid and growth: What does the cross-country evidence really show?" *Review of Economics and Statistics* 90 (4): 643-665.

Rodríguez, Francisco and Dani Rodrik (2001), "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence"in Ben Bernanke and Kenneth S. Rogoff, eds., *Macroeconomics Annual 2000*, (Cambridge, MA: MIT Press for NBER).

Rodrik, Dani and Romain Wacziarg (2005), "Do Democratic Transitions Produce Bad Economics Outcomes?" *American Economic Review Papers and Proceedings* 95: 50-56.

Roodman, David (2009a), "How to Do xtabond2: An Introduction to Difference and System GMM in Stata," *Stata Journal* 9 (1): 86-136.

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

Rose, Andrew K. (2006), "Size really doesn't matter: In search of a national scale effect", *Journal of the Japanese and International Economies* 20 (4): 482-507.

Sala-i-Martin, Xavier (1997), "I just ran two million regressions", *American Economic Review* 87 (2): 178-183.

Sala-i-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller (2004), "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach," *American Economic Review* 94 (4): 813-835.

Spolaore, Enrico, and Romain Wacziarg (2005), "Borders and Growth", *Journal of Economic Growth* 10 (4): 331-386.

Stock, James H., and Jonathan H. Wright (2000), "GMM with Weak Identification", *Econometrica* 68 (5): 1055-1096.

Stock, James H. and Motohiro Yogo (2005), "Testing for weak instruments in linear IV regression", in J. H. Stock and D. W. K. Andrews (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg* (New York: Cambridge University Press).

Tavares, Jose and Romain Wacziarg (2001), "How Democracy Affects Growth", *European Economic Review* 45 (3): 1341-1375.

Temple, Jonathan (1999), "The new growth evidence", *Journal of Economic Literature* 37 (1): 112-156.

Table 1: Second-stage cross-section regressions, 1970-2000

Dependent Variable $N = 78$	Growth (1)	Growth (2)	Growth (3)
Aid/GDP	0.096	0.911	0.078
	(0.070)	(4.083)	(0.066)
Initial Log Population		1.604	
		(7.923)	
Initial GDP/capita	−1.409	1.061	−1.438
	(0.435)	(12.782)	(0.403)
Initial Policy (Sachs-Warner)	2.139	2.541	2.154
	(0.619)	(1.981)	(0.583)
Initial Life Expectancy	0.076	0.368	0.069
	(0.039)	(1.460)	(0.038)
Geography	0.606	0.601	0.581
	(0.259)	(0.714)	(0.249)
Institutional Quality	4.077	−0.041	4.071
	(2.328)	(22.475)	(2.295)
Initial Inflation	−0.005	−0.020	−0.005
	(0.005)	(0.069)	(0.004)
Initial M2/GDP	0.010	0.006	0.011
	(0.020)	(0.058)	(0.019)
Initial Budget Balance/GDP	0.016	0.117	0.012
	(0.036)	(0.488)	(0.035)
Revolutions	−1.406	−4.567	−1.395
	(0.656)	(16.113)	(0.625)
Ethnic Fractionalization	0.788	4.518	0.658
	(0.851)	(18.927)	(0.820)
Constant	5.505	−56.789	6.163
	(3.527)	(312.142)	(3.102)
Excluded Instrument	$\overline{a}_r$	$\overline{a}_r$	ln(pop)
Cragg-Donald F †	31.63	0.13	36.30
Kleibergen-Paap F †	36.13	0.07	32.14
Kleibergen-Paap LM test (p-val)	<0.01	0.77	<0.01

† In this special case of a single endogenous regressor, the Cragg-Donald and Kleibergen-Paap F statistics reduce respectively to the standard non-robust and robust first-stage F statistics. Heteroskedasticity-robust standard errors in parentheses. All specifications include dummies for sub-Saharan Africa and East Asia.

Table 2: Instrumentation strength in cross-section regressions

"Zero-Stage" Specification		Replication	Colonial variables only	Population variables only
Period		(1)	(2)	(3)
1970-2000	CD F stat	31.63	<0.01	35.90
(N=78)	KP F stat	36.13	<0.01	31.62
	KP LM test (p-val)	<0.01	0.98	<0.01
1980-2000	CD F stat	29.37	1.41	40.54
(N=75)	KP F stat	31.26	1.41	39.65
	KP LM test (p-val)	<0.01	0.28	<0.01
1990-2000	CD F stat	8.52	1.69	12.86
(N=70)	KP F stat	6.95	1.18	9.00
	KP LM test (p-val)	<0.01	0.29	<0.01

CD = Cragg-Donald. KP = Kleibergen-Paap. † In this special case of a single endogenous regressor, the Cragg-Donald and Kleibergen-Paap F statistics reduce respectively to the standard non-robust and robust first-stage F statistics.

Table 3: Partial regression coefficients when ln Population is regressed on other growth determinants and RS covariates, 1970-2000 cross-section.

Dependent Variable	ln Population regressor		
	Coefficient	*Std. err.*	*N*
Aid/GDP	−1.925	0.340	78
Trade/GDP	−13.680	2.497	77
FDI/GDP	−0.537	0.183	77
Education Expenditure/GDP	−0.423	0.179	75
Gini Coefficient	−2.452	0.991	62
Government Consumption/GDP	−1.399	0.352	78
Manufacturing Value Added/GDP	1.529	0.398	76
Military Personnel/Total Labor Force	−0.263	0.123	78
Private Capital Flows/GDP	−2.548	1.057	77
Public Debt Service/GNI	−0.396	0.229	73
Savings/GDP	3.245	1.502	78

Heteroskedasticity-robust standard errors in parentheses. The point estimates and standard errors on the additional right hand side covariates have been omitted, but are available upon request.

Table 4: The non-excludability of country size in Hausmann, Hwang and Rodrik (2007)

Dep. var.	Growth	Growth	Growth	Growth	Growth	Growth
Estimator	IV†	GMM†	GMM	GMM	GMM	GMM
Size IV?	Yes	Yes	Yes, lev eq.	Yes, diff eq.	No	Yes
Size excl.?	Yes	Yes	Yes	Yes	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)
log initial	−0.0384	−0.0133	−0.0146	0.0032	0.0110	0.0107
GDP/cap	(4.37)	(1.54)	(1.66)	(0.23)	(0.97)	(1.13)
log initial	0.919	0.0427	0.0466	0.0082	−0.0174	−0.0170
EXPY	(4.54)	(2.28)	(2.56)	(0.21)	(0.76)	(0.78)
log human	0.0045	0.0046	0.0035	0.0002	0.0068	0.0054
capital	(1.75)	(0.64)	(0.51)	(0.02)	(1.22)	(0.94)
log area						−0.0039
						(3.20)
log population						0.0074
						(3.25)
constant	−0.420	−0.222	−0.242	−0.0672	0.0691	0.0038
	(4.25)	(1.98)	(2.22)	(0.30)	(0.64)	(0.04)
N	299	299	299	299	299	299
Hansen J test (p-val)	0.0001	0.09	0.09	0.11	0.19	0.19
# Countries	79	79	79	79	79	79
# Instruments	8	18	18	18	16	18
CD F stat	17.47	—	—	—	—	—
KP F stat	15.20	—	—	—	—	—

CD = Cragg-Donald. KP = Kleibergen-Paap. GMM = *system* GMM estimator. † Original, published Table 9, Columns 6 and 8 (Hausmann, Hwang and Rodrik 2007). Heteroskedasticity-robust t-statistics (following the original paper) in parentheses.

Table 5a: Weak instruments in Levine, Loayza, and Beck (2000)

Dep. var.	Growth	Growth	Growth	Growth	Growth
Estimator	GMM[†]	GMM[‡]	OLS	OLS-FD	OLS-FE
Collapsed IV	No	No	No	No	No
	(1)	(2)	(3)	(4)	(5)
Liquid liabilities	2.952	3.030	1.687	0.876	0.738
	(0.001)	(0.002)	(0.000)	(0.221)	(0.356)
Log initial GDP/capita	−0.742	−0.263	−0.353	−13.235	−7.312
	(0.001)	(0.698)	(0.047)	(0.000)	(0.000)
Government size	−1.341	0.025	−1.110	−1.881	−1.986
	(0.001)	(0.988)	(0.033)	(0.125)	(0.150)
Openness to trade	0.325	−0.187	0.243	1.138	0.646
	(0.169)	(0.839)	(0.427)	(0.448)	(0.645)
Inflation	1.748	1.274	−0.009	−3.368	−3.502
	(0.001)	(0.390)	(0.992)	(0.001)	(0.000)
Avg yrs sec. school	0.780	−0.122	0.447	0.100	0.939
	(0.001)	(0.851)	(0.238)	(0.824)	(0.155)
Black market premium	−2.076	−2.313	−1.530	−0.542	−0.426
	(0.001)	(0.000)	(0.000)	(0.319)	(0.478)
Constant	0.060	−4.516	−1.202		56.028
	(0.954)	(0.409)	(0.470)		(0.000)
N	353	353	353	328	353
# Countries	—	77	77	77	77
# Instruments	—	75	—	—	—
IV: Lagged Levels	Yes	Yes	—	—	—
IV: Lagged Diffs	Yes	Yes	—	—	—
CD F stat	—	—	—	—	—
KP F stat	—	—	—	—	—
KP LM test (p-val)	—	—	—	—	—

CD = Cragg-Donald. KP = Kleibergen-Paap. GMM = *system* GMM estimator. We follow the original paper in reporting p-values in parentheses. [†] Original, published Table 5, Column 1 (Levine, Loayza, and Beck 2001) [‡] Replication.

Table 5b: Weak instruments in Levine, Loayza, and Beck (2000), continued

Dependent variable	Growth	Growth	Growth	Growth
	Difference Equation		Levels Equation	
Estimator	2SLS	2SLS	2SLS	2SLS
Collapsed IV Matrix	No	Yes	No	Yes
	(6)	(7)	(8)	(9)
Liquid liabilities	−1.208	−12.398	2.733	1.963
	(0.549)	(0.712)	(0.011)	(0.423)
Log initial GDP/capita	−12.243	−10.754	0.647	2.293
	(0.000)	(0.206)	(0.335)	(0.325)
Government size	0.052	8.622	−2.140	−8.899
	(0.980)	(0.481)	(0.126)	(0.204)
Openness to trade	2.817	0.511	−0.417	6.629
	(0.189)	(0.901)	(0.718)	(0.404)
Inflation	−1.353	−34.177	−1.031	2.822
	(0.669)	(0.513)	(0.581)	(0.487)
Avg yrs secondary school	0.959	−7.874	−1.047	−2.429
	(0.509)	(0.663)	(0.150)	(0.325)
Black market premium	−0.408	2.455	−1.926	−0.340
	(0.664)	(0.814)	(0.003)	(0.878)
Constant			−4.769	−24.238
			(0.389)	(0.269)
N	328	328	353	353
# Countries	77	77	77	77
# Instruments	40	12	40	12
IV: Lagged Levels	Yes	Yes	No	No
IV: Lagged Diffs	No	No	Yes	Yes
CD F stat	0.57	0.05	0.62	0.23
KP F stat	1.75	0.57	1.45	0.23
KP LM test (p-val)	0.22	0.57	0.81	0.21

We follow the original paper in reporting p-values in parentheses. CD = Cragg-Donald. KP = Kleibergen-Paap.

Table 6a: Weak instruments in Rajan and Subramanian (2008)

Dependent variable	Growth	Growth	Growth	Growth	Growth
Estimator	GMM-DIF†	GMM-SYS†	OLS	OLS-FD	OLS-FE
Collapsed IV Matrix	No	No	No	No	No
	(1)	(2)	(3)	(4)	(5)
Aid/GDP	−0.151	−0.054	−0.037	−0.236	−0.262
	(0.077)	(0.114)	(0.053)	(0.066)	(0.067)
Initial GDP/capita	−8.347	−2.456	−1.514	−13.245	−7.960
	(1.543)	(1.057)	(0.517)	(1.839)	(1.307)
Policy	−1.774	1.370	0.428	−1.131	−1.062
	(0.933)	(1.015)	(0.587)	(0.535)	(0.660)
Life Expectancy	−0.393	0.086	0.049	−0.136	−0.082
	(0.183)	(0.098)	(0.067)	(0.132)	(0.117)
Institutional . Quality	6.953	2.748	2.961	6.409	4.245
	(2.767)	(2.579)	(1.561)	(1.789)	(1.901)
Log Inflation	−1.985	−1.498	−1.854	−1.742	−1.916
	(0.671)	(0.663)	(0.489)	(0.409)	(0.440)
M2/GDP	−0.002	0.010	−0.004	−0.001	−0.007
	(0.032)	(0.021)	(0.015)	(0.039)	(0.032)
Budget Bal. /GDP	0.164	0.101	0.100	0.184	0.106
	(0.082)	(0.075)	(0.063)	(0.057)	(0.059)
Revolutions	−0.972	−0.073	−0.487	−1.104	−1.049
	(0.625)	(0.992)	(0.374)	(0.559)	(0.517)
Ethnic Frac.		0.129	0.173		
		(1.809)	(1.114)		
Geography		0.496	0.522		
		(0.353)	(0.276)		
N	167	239	239	167	239
# of countries	68	72	72	68	72
# of instruments	120	158	—	—	—
Lags used	2-7	2-7	—	—	—
IV: Lagged levels	Yes	Yes	No	No	No
IV: Lagged diffs	No	Yes	No	No	No
CD F stat	—	—	—	—	—
KP F stat	—	—	—	—	—
KP LM test (p-val)	—	—	—	—	—

CD = Cragg-Donald. KP = Kleibergen-Paap. † Exact replication of original published Table 9, Column 1 and Table 10, Column 1 respectively. Heteroskedasticity-robust standard errors in parentheses. Following Rajan and Subramanian, we include but suppress the point estimates on a constant and dummies for Sub-Saharan Africa and East Asia in columns 2, 3, 5, 8 and 9.

Table 6b: Weak instruments in Rajan and Subramanian (2008), continued

Dependent variable	Growth	Growth	Growth	Growth
	Difference Equation		Levels Equation	
Estimator	2SLS	2SLS	2SLS	2SLS
Collapsed IV Matrix	No	Yes	No	Yes
	(6)	(7)	(8)	(9)
Aid/GDP	−0.220	−0.355	0.116	0.470
	(0.071)	(0.123)	(0.089)	(0.854)
Initial GDP/capita	−11.060	−10.535	0.117	10.193
	(1.644)	(2.295)	(1.307)	(18.102)
Policy	−1.515	−1.869	−0.095	5.429
	(0.696)	(0.887)	(0.816)	(6.634)
Life Expectancy	−0.369	−0.232	0.014	−2.086
	(0.147)	(0.185)	(0.159)	(2.549)
Institutional Qual.	6.537	6.570	4.356	34.008
	(1.864)	(3.098)	(3.560)	(96.093)
Log Inflation	−1.921	−1.120	−1.927	−2.168
	(0.424)	(0.840)	(0.902)	(4.999)
M2/GDP	−0.002	−0.025	0.018	−0.015
	(0.039)	(0.052)	(0.028)	(0.109)
Budget Bal./GDP	0.246	0.354	0.014	0.422
	(0.066)	(0.110)	(0.156)	(0.948)
Revolutions	−1.072	−1.396	0.032	1.436
	(0.697)	(0.962)	(1.159)	(7.369)
Ethnic Frac.			0.570	−2.196
			(1.407)	(6.256)
Geography			0.333	−0.339
			(0.360)	(3.114)
N	167	167	239	239
# of countries	68	68	72	72
# of instruments	120	52	41	17
Lags used	2-7	2-7	2	2
IV: Lagged levels	Yes	Yes	No	No
IV: Lagged diffs	No	No	Yes	Yes
CD F stat	0.66	0.43	0.41	0.06
KP F stat	9.89	3.83	0.99	0.05
KP LM test (p-val)	0.52	0.23	0.82	0.47

CD = Cragg-Donald. KP = Kleibergen-Paap.

*** Not for publication ***

Appendix A: Sources of additional data

The original Rajan and Subramanian dataset was kindly provided by the authors. The analysis here required it to be supplemented with population data. The original dataset contained population ratios from zero-stage regressions but not separate figures for period-initial receiving country population. For the zero-stage regressions, the only database with sufficiently complete country coverage was the International Monetary Fund's online *International Financial Statistics* (accessed Sept. 9, 2007), which had populations of all aid recipient countries in the Rajan and Subramanian dataset, except for Bermuda, Kiribati, Turkmenistan, and Uzbekistan, which come from the World Bank's *World Development Indicators 2007.* In the main regressions, the extreme breadth of country coverage is not needed and we took population from the *Penn World Tables 6.1*, since real GDP/capita came from that source. The correlation between the two sources' population estimates is near unity.

All growth determinants used in Table 6 come from the World Bank's *World Development Indicators 2007* (Aid/GDP, Trade/GDP, FDI/GDP, Education Expenditure/GDP, Gini Coefficient, Government Consumption/GDP, Manufacturing Value Added/GDP, Military Personnel/Total Labor Force, Private Capital Flows/GDP, Public Debt Service/GNI, and Savings/GDP).

The Levine, Loayza and Beck (2000) dataset was obtained on 10 July 2008 from the World Bank website `http://go.worldbank.org/40TPPEYOC0`.

The Hausmann, Hwang and Rodrik (2007) dataset was kindly provided by the authors.

Appendix B: Weak identification of nonlinear effects

If there are diminishing returns to capital in an economy, the effect of aid on growth can be nonlinear and concave. Assuming a linear relationship can easily cloud such a relationship: the best linear fit to a concave parabola has slope zero. Beyond this clear theoretical reason to test for nonlinear effects, several important aid-growth regressions published in the past decade have tested for and found a nonlinear relationship (e.g. Hansen and Tarp 2001; Dalgaard, Hansen and Tarp 2004). In a small part of one table, Rajan and Subramanian attempt to test for a nonlinear relationship between aid and growth, but their identification strategy does not allow this. The instrumentation in these regressions is extremely weak. They do not report this.

The first row of Table B.1 shows the Cragg-Donald and Kleibergen-Paap statistics for three regressions in Rajan and Subramanian (2008) Table 4 (Panel A), where the aid effect is assumed linear. Instrumentation is strong. The next row shows the same statistics for three regressions in their Table 7 (Panel A), which include a squared aid regressor, and use $\overline{a}_r$ and its square as the only excluded instruments. The inclusion of the squared term causes instrumentation strength to collapse in the periods 1980-2000 and 1990-2000, which is not reported in RS. Strength is retained in the 1970-2000 period, but solely due to the presence of Guinea-Bissau in the sample for that period (Guinea-Bissau is omitted from the sample in RS's other two periods). Without Guinea-Bissau, in the third row, no useful degree of instrumentation strength is present regardless of periodization. All instrumentation in these nonlinear regressions, then, depends on a single country in a single period. The RS instrument does not allow a meaningful test of a nonlinear effect of aid on growth.[28]

There is no escape from this problem within the RS framework: The instruments independent of country size (I_1-I_7) do not explain aid variance, and the only strong instrument (population) is plausibly invalid. The only way to advance the literature is to find new instruments—better natural experiments to isolate the true effect of aid.

Table B1: Weak instruments in nonlinear specifications

Period	Test statistic	1970-2000	1980-2000	1990-2000
Linear specification	CD	31.62	29.37	8.52
(RS Table 4A)	KP	36.13	31.26	6.95
Quadratic specification	CD	13.70	0.01	0.14
(RS Table 7A)	KP	13.10	0.02	0.31
Quadratic specification	CD	0.41	0.01	0.14
without Guinea-Bissau *	KP	0.28	0.02	0.31

[28] One alternative procedure would be to carry out two separate zero-stage regressions, with regressands of linear aid and squared aid, to create two constructed instruments. This does not, however, improve instrumentation strength (results available on request).

CD = Cragg-Donald. KP = Kleibergen-Paap. * Guinea-Bissau is only included in the 1970-2000 regressions in RS.

Note that the coefficients on aid/GDP and aid/GDP squared were reported in Table 7 (Panel A) of Rajan and Subramanian (2008), but the first-stage F statistics were not reported.