Not for Publication

Online-Only Appendix to "Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth"

A Testing for underidentification and weak instruments

We provide here additional details on the test statistics and inference procedures used in the paper to assess the strength of identification in regressions based on instrumental variables procedures. These weak instruments test statistics are often reported in empirical applications. However, the inferential implications, particularly for the weak instruments test statistics, are often left unstated.

The first diagnostic tool for assessing the strength of identification is based on a Langrange-Multiplier (LM) test for underidentification using the Kleibergen and Paap (2006) rk statistic. This test, readily implemented in Stata using the `ranktest` package, allows researchers to determine whether the minimal canonical correlation between the endogenous variables and the instruments is statistically different from zero. Another way of framing the test is by asking whether, after partialling out exogenous covariates and cross-correlations with the other endogenous variables and instruments, does the weakest correlation between an instrument and one of the endogenous variables suffice to contribute enough independent variation to add to the empirical rank of the instrument matrix? The p-values for this test are readily available after running the 2SLS estimation using the `ivreg2` package for Stata. The LM test for underidentification provides a lower hurdle than the tests for weak instruments.

The second set of diagnostics are based on the Stock and Yogo (2005) characterization of weak instruments using the first-stage F statistic and its multivariate analogue, the Cragg-Donald Wald statistic or its robust counterpart, the Kleibergen-Paap Wald statistic. The usual approach in the applied literature is to conclude that instruments are weak if these test statistics exceed the critical values tabulated by Stock and Yogo. Much less common is the full use of the testing procedures detailed in Section 4 of Stock and Yogo (2005), which provides richer probabilistic tools for characterizing weak instruments. Here, we provide a few practical details on how to construct p-values for the weak-instruments tests introduced in Section 3.3 and used throughout our paper. Adapting the empirical procedures in Gauss deployed in Yogo (2004), the formulation of p-values in Stata proceeds as follows:

1. Obtain the asymptotic threshold values for the concentration parameter Λ corresponding to the weak instruments test (relative OLS bias or t-test size). These values are contained in a number of Gauss matrices on Motohiro Yogo's website.[1] We have converted these into Stata datasets (`lambfitBias.dta` and `lambfitSize.dta`) and provided them in supplementary material available online.

[1] Available WWW: https://sites.google.com/site/motohiroyogo/home/publications/TestingWI_Programs.zip?attredirects=0.

2. Select the relevant value $\widehat{\Lambda}$ from the appropriate column and row of the lambfitBias or lambfitSize matrices based on the number of endogenous variables, the number of instruments K, and the level of bias or size distortion of interest. The relative OLS bias test is based on the finite sample distribution of the 2SLS estimator and hence critical values can only be calculated for cases where there are at least two more overidentifying restritions than the number of endogenous variables. In all specifications where this condition is not met, we report the weak instruments test based on the size distortion of the t-test. Critical values for this test, however, are not tabulated for cases with more than two endogenous variables. Thus, in cases with more endogenous variables and/or instruments than available in the Stock-Yogo tabulations, we take the penultimate available critical value in the given row and column of the table.

3. Obtain the Cragg-Donald ($\widehat{CD}$) and Kleibergen-Paap ($\widehat{KP}$) Wald test statistics after estimating the given 2SLS growth regression using `ivreg2` in Stata.

4. Calculate the p-value for the given null hypothesis using the formula: $p = 1-\texttt{nchi2}(K, K \times \widehat{\Lambda}, K \times \widehat{CD})$, where `nchi2` is the noncentral χ^2 distribution with degrees of freedom K and noncentrality parameter $K \times \widehat{\Lambda}$. The p-value is valid for the Cragg-Donald statistic, which assumes homoskedastic error terms. While the $\widehat{KP}$ is robust to non-i.i.d. errors, its insertion in the p-value formula does not immediately follow since the Stock-Yogo diagnostics were not originally formulated for the non-i.i.d. case. Nevertheless, in characterizing weak instruments, we follow others in the literature and report $\widehat{CD}$ as well as $\widehat{KP}$ for each specification. Thus, while acknowledging that the p-values using the $\widehat{KP}$ statistic are not asymptotically correct, we report it along with that for the $\widehat{CD}$ statistic for each of the given bias or size tests.

B Weak-instrument robust inference

In Section 4.5, we employ the weak-instrument robust testing procedure of Kleibergen (2002) to examine 2SLS dynamic panel equations in levels and first differences. This procedure has been introduced as a higher power alternative to the Anderson-Rubin statistic. Here, we describe the steps for applying this method. Suppose that the dynamic panel growth equation is given by equation (3):

$$g_{i,t} = \beta \ln y_{i,t-1} + \mathbf{x}'_{i,t}\boldsymbol{\gamma}_1 + \widetilde{\mathbf{x}}'_{i,t}\boldsymbol{\gamma}_2 + \psi_i + \nu_{i,t}.$$

Suppose that $\boldsymbol{x}$ is a j-dimensional vector of endogenous growth determinants and $\widetilde{\boldsymbol{x}}$ is a k-dimensional vector of exogenous growth determinants including indicators for the period t. After constructing the appropriate instrument matrices for this equation in levels (LEV) and first differences (DIF), the method proceeds as follows:

1. Define the $j+1$ dimensional grid of possible values for the joint confidence region of β and $\boldsymbol{\gamma}_1$. In Figure 3, we restrict attention to a relatively narrow range of parameter values. Our principle was simply to start from the 2SLS point estimates and ensure that we chose a sufficiently wide range of values on both sides of that point estimate to encompass many values above and below zero. In the most general albeit infeasible case, one would want to examine the whole real line for each of the $j+1$ parameters. Lastly, one defines the increments over which to step along the range of values for a given parameter.

2. For the m-th $j+1$-tuple $(\beta^m, \boldsymbol{\gamma}_1^m)$, define $\widehat{g}_{i,t} = g_{i,t} - \beta^m \ln y_{i,t-1} - \mathbf{x}'_{i,t}\boldsymbol{\gamma}_1^m$ for the LEV equation and $\widehat{\Delta g}_{i,t} = \Delta g_{i,t} - \beta^m \Delta \ln y_{i,t-1} - \Delta \mathbf{x}'_{i,t}\boldsymbol{\gamma}_1^m$ for the DIF equation.

3. Regress $\beta^m \ln y_{i,t-1}$ ($\beta^m \Delta \ln y_{i,t-1}$) on all LEV (DIF) equation instruments and the exogenous covariates $(\Delta)\widetilde{\mathbf{x}}'_{i,t}$. Obtain the predicted values $\beta^m \widehat{\ln y_{i,t-1}}$ ($\beta^m \widehat{\Delta \ln y_{i,t-1}}$). Repeat the procedure for each of the j endogenous covariates in $\mathbf{x}'_{i,t}$.

4. Regress $\widehat{\Delta g}_{i,t}$ on $\beta^m \widehat{\Delta \ln y_{i,t-1}}$ and $\widehat{\Delta \mathbf{x}'_{i,t} \boldsymbol{\gamma}^m_1}$. Do the same for the LEV equation.

5. Test the joint significance of the right-hand side variables and store the associated p-values based on the large-sample $\chi^2(j+1)$ statistic for the given $j+1$-tuple $(\beta^m, \boldsymbol{\gamma}^m_1)$.

6. Using the resulting dataset comprised of p-values and $j+1$-tuples $(\beta^m, \boldsymbol{\gamma}^m_1)$ for the DIF and LEV equations, plot two-dimensional joint confidence ellipses (using the user-written `ellip` in Stata) for those values of $j+1$-tuples such that the p-value is greater than 0.05. More complex three-dimensional ellipsoids can be plotted in Matlab.

C Weak identification of nonlinear effects in Rajan & Subramanian (2008)

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 (presuming the full parabola is observed). 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 et al., 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.

Columns 1, 4, and 7 of Table C.1 show the underidentification and weak-instrument test statistics (p-values) for three regressions in Table 4 of Rajan and Subramanian (2008), where the aid effect is assumed linear. Instrumentation is strong. Columns 2, 5, and 8 show 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 columns 3, 6, and 9 , no useful degree of instrument strength is present regardless of periodization. In fact, we cannot reject that the structural equation is underidentified. 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.[2]

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, as we demonstrate in Section 3.3 of the paper. A more fruitful way forward is to find new instruments—better natural experiments to isolate the true effect of aid.

[2]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 instrument strength.

Table C.1: Weak instruments in nonlinear specifications of Rajan and Subramanian (2008, Tables 4A and 7A)

Period	1970-2000			1980-2000			1990-2000		
Aid Specification	Linear	Quadratic	Quadratic sans GNB[∓]	Linear	Quadratic	Quadratic sans GNB[∓]	Linear	Quadratic	Quadratic sans GNB[∓]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Kleibergen-Paap LM test p-value[†]	0.0004	< 0.0001	0.397	0.0002	0.837	0.837	0.014	0.363	0.363
Cragg-Donald Waldstat[‡]	31.63	13.70	0.412	29.37	0.012	0.012	8.52	0.14	0.14
H_0: t-test size>10% (p-value)	0.058	0.085	0.999	0.085	1.000	1.000	0.871	1.000	1.000
H_0: t-test size>25% (p-value)	0.001	0.008	0.983	0.001	1.000	1.000	0.285	0.996	0.996
H_0: relative OLS bias>10% (p-value)	0.005	0.109	0.999	0.008	1.000	1.000	0.538	1.000	1.000
H_0: relative OLS bias>30% (p-value)	0.001	0.0210	0.993	0.001	1.000	1.000	0.275	0.998	0.998
Kleibergen-Paap Wald stat[‡]	36.12	13.10	0.279	31.26	0.017	0.017	6.952	0.314	0.314
H_0: t-test size>10% (p-value)	0.025	0.105	1.000	0.061	1.000	1.000	0.921	0.999	0.999
H_0: t-test size>25% (p-value)	0.0001	0.011	0.990	0.001	1.000	1.000	0.388	0.988	0.988
H_0: relative OLS bias>10% (p-value)	0.001	0.132	1.000	0.005	1.000	1.000	0.647	1.000	1.000
H_0: relative OLS bias>30% (p-value)	0.0001	0.028	0.996	0.001	1.000	1.000	0.376	0.995	0.995

Notes: The estimates in columns 1, 3 and 7 are exact replications of columns 2, 3, and 4 in Table 4A of Rajan and Subramanian (2008). The estimates in columns 2, 4 and 8 are exact replications of columns 2, 3, and 4 in Table 7A of Rajan and Subramanian (2008). [∓] Guinea-Bissau is only included in the 1970-2000 regressions in the original Table 7A. [†] The null hypothesis of the Kleibergen-Paap LM test is that the structural equation is underidentified (i.e., the rank condition fails). The test uses a rank test procedure from Kleibergen and Paap (2006). [‡] The Cragg-Donald and Kleibergen-Paap Wald statistics correspond respectively to non-robust and heteroskedasticity-robust multivariate analogues to the first-stage F statistics. Below each test statistic, we report the p-values from tests of whether (i) the actual size of the t-test(s) that $\beta_{aid} = 0$ (and $\beta_{aid^2} = 0$) at the 5% significance level is greater than 10 or 25%, and (ii) the bias of the IV estimates of β_{aid} (and β_{aid^2}) are greater than 10 or 30% of the OLS bias. In both cases, the critical values are obtained from Stock and Yogo (2005). Although critical values do not exist for the Kleibergen-Paap statistic, we follow the approach suggested in Baum et al. (2007) and apply the Stock and Yogo critical values initially tabulated for the Cragg-Donald statistic. The critical values for (ii) are (less conservatively) based on three instruments since one cannot calculate critical values in the (finite-sample)bias tests for the case of one endogenous variable and fewer than three instruments.

D Further empirical and simulation results

D.1 Rajan & Subramanian Cross-Section Regressions (for 1990-2000)

Table D.1 reproduces the Rajan and Subramanian (2008) results from Table 2 in the paper with an additional three columns covering their period 1990-2000. The results are similar to those for the longer periods (1970- and 1980-2000) as discussed in the paper.

D.2 Sources of identification in the Hausmann et al (2007) five-year panel

Using the Hausmann et al. (2007) panel data for the five-year periodization, Table D.2 reports the same set of specification tests as in Table 4 based on their ten-year periodization. As noted in Section 3.3, the key result that export diversity (EXPY) increases growth does not hinge on the excludability of population size in the same restrictive manner that it did in the shorter panel with ten-year periodization. Although the result becomes null in column 6 when controlling for country size directly in the second stage, the effect of EXPY on growth is relatively robust to increasingly relaxing the excludability of the country size instruments in the levels and difference equation instrument matrices in columns 2-5. Nevertheless, there still remain concerns about the validity of the size instruments. While we cannot reject the validity of the size instruments on the basis of the difference-in-Hansen statistic in the specifications of columns 2-4, further unpacking of the levels and difference equation moment conditions in Section 4.4 revealed the validity of the size instruments could not be rejected for the levels equation according to the Hahn et al. (2011) test, the details of which are reported in the notes to Table 8.

D.3 Other measures of financial intermediation in Levine et al (2000)

Tables D.3 and D.4 estimate the same specifications as Table 6 using the two other measures of financial depth in Levine et al. (2000): private credit/GDP and the ratio of commercial to central bank credit, respectively. As noted in Section Section 4.2, the weak instruments problem of the system GMM estimator holds for these additional measures of financial depth. This can be seen most readily from the p-values for the weak-instruments tests reported in columns 6-9 of each table. The one slight difference with the liquid liabilities results in Table 6 is that we can reject the null of underidentification for the levels equation estimated using the collapsed instrument matrix in column 9 of Tables D.3 and D.4. Although these instruments pass the lower hurdle of underidentification, they remain weak.

D.4 Simulation results for a larger autoregressive parameter $\beta = 0.8$

Using the simulation procedure described in the paper, Figures D.1 and D.2 demonstrate the performance of the difference and system GMM estimators of γ (the coefficient on the endogenous growth determinant) when the persistence of the autoregressive parameter β increases from the baseline value of 0.2 to 0.8 (see equation (4)). The results are qualitatively unchanged from the baseline presented in Figures 1 and 2.

Table D.1: Instrumentation strength in Rajan and Subramanian (2008) cross-section regressions

Period	1970–2000 ($N = 78$)			1980–2000 ($N = 75$)			1990–2000 ($N = 70$)		
"Zero-Stage" Specification	Replication	Colonial vars. only	Population vars. only	Replication	Colonial vars. only	Population vars. only	Replication	Colonial vars. only	Population vars. only
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Point estimate: Aid/GDP	0.096	–15.944	0.078	–0.004	–0.308	–0.028	–0.389	–0.035	–0.294
	(0.070)	(633.474)	(0.067)	(0.095)	(0.389)	(0.084)	(0.194)	(0.442)	(0.144)
CLR confidence set*: Aid/GDP	[–0.027,0.292]	$(-\infty, \infty)$	[–0.039,0.254]	[–0.186,0.232]	$(-\infty, \infty)$	[–0.194,0.170]	[–1.463,–0.063]	$(-\infty, 7.860] \cup [8.071, \infty)$	[–0.874,–0.021]
Kleibergen-Paap LM test (p-value)†	0.0004	0.978	0.0001	0.0002	0.282	0.0001	0.014	0.288	0.004
Cragg-Donald Wald stat‡	31.63	0.001	35.90	29.37	1.41	40.54	8.52	1.69	12.86
H_0: t-test size>10% (p-value)	< 0.001	0.999	< 0.001	0.001	0.888	< 0.001	0.303	0.865	0.118
H_0: t-test size>25% (p-value)	< 0.001	0.980	< 0.001	< 0.001	0.341	< 0.001	0.013	0.295	0.002
H_0: relative OLS bias>10% (p-value)	< 0.001	0.996	< 0.001	< 0.001	0.772	< 0.001	0.161	0.735	0.049
H_0: relative OLS bias>30% (p-value)	< 0.001	0.987	< 0.001	< 0.001	0.503	< 0.001	0.040	0.455	0.008
Kleibergen-Paap Wald stat‡	36.12	0.001	31.62	31.26	1.41	39.65	6.95	1.18	9.00
H_0: t-test size>10% (p-value)	< 0.001	0.999	< 0.001	0.001	0.888	< 0.001	0.407	0.906	0.275
H_0: t-test size>25% (p-value)	< 0.001	0.984	< 0.001	< 0.001	0.340	< 0.001	0.026	0.385	0.011
H_0: relative OLS bias>10% (p-value)	< 0.001	0.997	< 0.001	< 0.001	0.770	< 0.001	0.239	0.801	0.142
H_0: relative OLS bias>30% (p-value)	< 0.001	0.990	< 0.001	< 0.001	0.502	< 0.001	0.071	0.546	0.034

Notes: In all specifications, the instrumental variable is aid/GDP predicted from the zero-stage regression. The dependent variable in all specifications is average annual growth in GDP per capita over the period. Heteroskedasticity-robust standard errors in parentheses. Following the original paper, we retain the degrees-of-freedom adjustment to the Kleibergen-Paap F and LM statistics based on robust standard errors. For each of the three periods, the first column is based on exact replication of the baseline result in Rajan and Subramanian (2008, Table 4); the second column removes donor and recipient population terms from the zero-th stage specification used to estimate the predicted aid/GDP instrument $\overline{a}_r$, retaining only the colonial ties indicators; the third column retains only the population terms in the zero-th stage. All specifications include dummies for sub-Saharan Africa and East Asia. *The CLR confidence set corresponds to the weak-instrument robust confidence set obtained using the conditional likelihood ratio test in Moreira (2003). †The null hypothesis of the Kleibergen-Paap LM test is that the structural equation is underidentified (i.e., the rank condition fails). The test uses a rank test procedure from Kleibergen and Paap (2006). ‡In this special case of a single endogenous regressor, the Cragg-Donald and Kleibergen-Paap Wald statistics reduce respectively to the standard non-robust and heteroskedasticity-robust first-stage F statistics. Below each, we report the p-values from tests of whether (i) the actual size of the t-test that $\beta_{aid} = 0$ at the 5% significance level is greater than 10 or 25%, and (ii) the bias of the IV estimates of β_{aid} reported in the table are greater than 10 or 30% of the OLS bias. In both cases, the critical values are obtained from Stock and Yogo (2005). Although critical values do not exist for the Kleibergen-Paap statistic, we follow the approach suggested in Baum et al. (2007) and apply the Stock and Yogo critical values initially tabulated for the Cragg-Donald statistic. The critical values for (ii) are (less conservatively) based on three instruments since Stock and Yogo do not tabulate critical values in the bias tests for the case of one endogenous variable and fewer than three instruments.

Table D.2: The (non-?)excludability of country size in 5-year panels of Hausmann et al. (2007)

Dependent variable	Growth	Growth	Growth	Growth	Growth	Growth
Estimator	IV$^{\mp}$	GMM-SYS$^{\mp}$	GMM-SYS	GMM-SYS	GMM-SYS	GMM-SYS
Size Instruments?	Yes	Yes	Yes, lev. Eq.	Yes, diff. eq.	No	Yes
Size Excluded?	Yes	Yes	Yes	Yes	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)
log initial GDP/capita	-0.030	-0.014	-0.015	-0.014	-0.008	-0.005
	(4.820)	(2.655)	(2.764)	(2.139)	(1.394)	(0.748)
log initial EXPY	**0.074**	**0.045**	**0.046**	**0.046**	**0.036**	**0.016**
	(5.105)	**(4.097)**	**(4.204)**	**(3.828)**	**(3.006)**	**(1.112)**
log human capital	0.004	0.004	0.003	-0.000	0.000	0.001
	(1.781)	(0.920)	(0.904)	(0.088)	(0.067)	(0.207)
log area						0.014
						(3.979)
log population						-0.009
						(3.233)
Observations	604	604	604	604	604	604
Number of Countries	79	79	79	79	79	79
Number of Periods	8	8	8	8	8	8
Number of Instruments	2	75	75	75	73	75
Hansen J test (p-value)	< 0.0001	0.507	0.502	0.467	0.623	0.267
Hansen J test excluding size instruments (p-value)		0.562	0.552	0.537	—	—
Difference-in-Hansen test or C statistic (p-value)$^{\pm}$		0.173	0.184	0.184	—	—
Kleibergen-Paap LM test (p-value)†	< 0.001	—	—	—	—	—
Cragg-Donald Waldstat‡	39.09	—	—	—	—	—
H_0: t-test size>25% (p-value)	< 0.001	—	—	—	—	—
H_0: relative OLS bias>30% (p-value)	< 0.001	—	—	—	—	—
Kleibergen-Paap Wald stat‡	34.25	—	—	—	—	—
H_0: t-test size>25% (p-value)	< 0.001	—	—	—	—	—
H_0: relative OLS bias>30% (p-value)	< 0.001	—	—	—	—	—

Notes: The dependent variable in all specifications is average annual growth over the period. The size instruments include log population and log area. The internal instruments refer to the lagged levels and lagged differences of endogenous right-hand side variables in the respective difference and levels equations of the dynamic panel GMM system of equations. $^{\mp}$ Columns 1 and 2 are based on Hausmann et al. (2007, Table 9, Columns 2 and 4). The null hypothesis of the difference-in-Hansen test (or C statistic, see Hayashi, 2000) is that the size instruments are valid. Following the original paper, we report heteroskedasticity-robust standard errors in parentheses and retain associated degrees of freedom adjustments for the first-stage test statistics. See the notes to Table D.1 for more details on the Kleibergen-Paap and Cragg-Donald tests, which apply in column 1 to the endogenous EXPY variable.

Table D.3: Weak instruments in dynamic panel regressions using private credit in Levine et al. (2000)

						Difference Equation		Levels Equation	
Estimator	GMM-SYS∓	GMM-SYS∓	OLS	OLS-FD	OLS-FE	2SLS	2SLS	2SLS	2SLS
Collapsed IV matrix	No	No	—	—	—	No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Private credit	1.522	1.494	0.823	0.807	0.945	0.284	9.291	1.451	2.320
	(0.001)	(0.001)	(0.004)	(0.046)	(0.040)	(0.826)	(0.864)	(0.033)	(0.109)
Log initial GDP/capita	-0.364	-0.398	-0.315	-14.016	-7.832	-12.420	0.552	0.593	1.109
	(0.001)	(0.001)	(0.088)	(0.000)	(0.000)	(0.000)	(0.989)	(0.448)	(0.614)
					Other parameter estimates omitted				
N	359	359	345	323	345	323	323	345	345
Number of countries	74	74	74	74	74	74	74	74	74
Number of instruments	75	75	—	—	—	40	12	40	12
IV: Lagged levels	Yes	Yes	—	—	—	Yes	Yes	No	No
IV: Lagged differences	Yes	Yes	—	—	—	No	No	Yes	Yes
Kleibergen-Paap LM test (p-value)	—	—	—	—	—	0.249	0.879	0.635	0.069
Cragg-Donald Waldstat	—	—	—	—	—	0.73	0.004	0.67	0.51
H_0: relative OLS bias>10% (p-value)	—	—	—	—	—	1.000	1.000	1.000	1.000
H_0: relative OLS bias>30% (p-value)	—	—	—	—	—	1.000	1.000	1.000	0.987
Kleibergen-Paap Wald stat	—	—	—	—	—	1.08	0.001	1.16	0.47
H_0: relative OLS bias>10% (p-value)	—	—	—	—	—	1.000	1.000	1.000	1.000
H_0: relative OLS bias>30% (p-value)	—	—	—	—	—	1.000	1.000	1.000	0.983

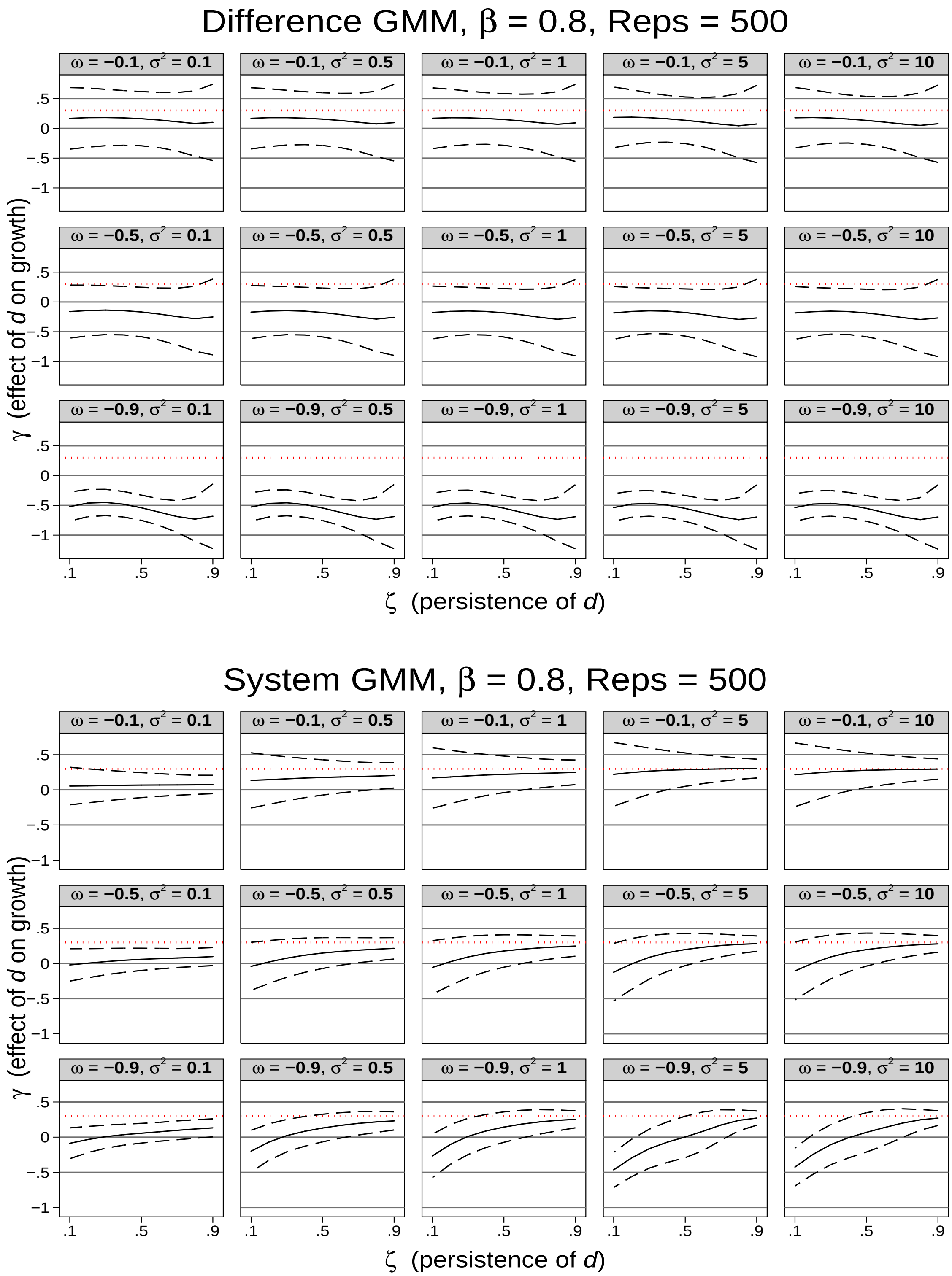
Notes: The dependent variable in all specifications is average annual growth in GDP per capita each period. ∓ Column 1 reproduces the published version of Levine et al. (2000, Table 5, Column 2), and column 2 reports our best attempted replication using the `DPD96` program for Gauss, the publicly available dataset, and a Gauss program used to generate their results provided by Thorsten Beck. Further details on the difference in sample sizes across columns, our replication efforts, and the associated differences in the Gauss and Stata programs for dynamic panel GMM regressions can be found in Appendix E.1. The following variables are included in the regressions but suppressed in the table here for presentational purposes: government size, openness to trade, inflation, average years of secondary schooling, black market premium, time period dummies and a constant. The first five of these variables are treated as endogenous. Following the original paper, we report p-values in parentheses. See the notes to Table D.1 for more details on the Kleibergen-Paap and Cragg-Donald tests, which apply in columns 6-9 to the full set of endogenous right-hand-side variables.

Table D.4: Weak instruments in dynamic panel regressions using commercial vs. central bank credit in Levine et al. (2000)

						Difference Equation		Levels Equation	
Estimator	GMM-SYS$^{\mp}$	GMM-SYS$^{\mp}$	OLS	OLS-FD	OLS-FE	2SLS	2SLS	2SLS	2SLS
Collapsed IV matrix	No	No	—	—	—	No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Commercial vs. Central Bank Credit	2.437	2.293	1.243	2.526	3.397	3.649	-7.670	0.864	4.708
	(0.001)	(0.001)	(0.033)	(0.017)	(0.010)	(0.158)	(0.979)	(0.536)	(0.198)
Log initial GDP/capita	-0.117	-0.348	-0.138	-13.896	-7.912	-12.817	-8.985	0.568	1.470
	(0.223)	(0.015)	(0.419)	(0.000)	(0.000)	(0.000)	(0.963)	(0.347)	(0.483)
					Other parameter estimates omitted				
N	359	359	345	324	345	324	324	345	345
Number of countries	74	74	74	74	74	74	74	74	74
Number of instruments	75	75	—	—	—	40	12	40	12
IV: Lagged levels	Yes	Yes	—	—	—	Yes	Yes	No	No
IV: Lagged differences	Yes	Yes	—	—	—	No	No	Yes	Yes
Kleibergen-Paap LM test (p-value)	—	—	—	—	—	0.489	0.963	0.500	0.022
Cragg-Donald Waldstat	—	—	—	—	—	0.71	< 0.001	0.88	0.76
H_0: relative OLS bias>10% (p-value)	—	—	—	—	—	1.000	1.000	1.000	1.000
H_0: relative OLS bias>30% (p-value)	—	—	—	—	—	1.000	1.000	1.000	0.954
Kleibergen-Paap Wald stat	—	—	—	—	—	0.84	< 0.001	1.25	0.75
H_0: relative OLS bias>10% (p-value)	—	—	—	—	—	1.000	1.000	1.000	1.000
H_0: relative OLS bias>30% (p-value)	—	—	—	—	—	1.000	1.000	1.000	0.954

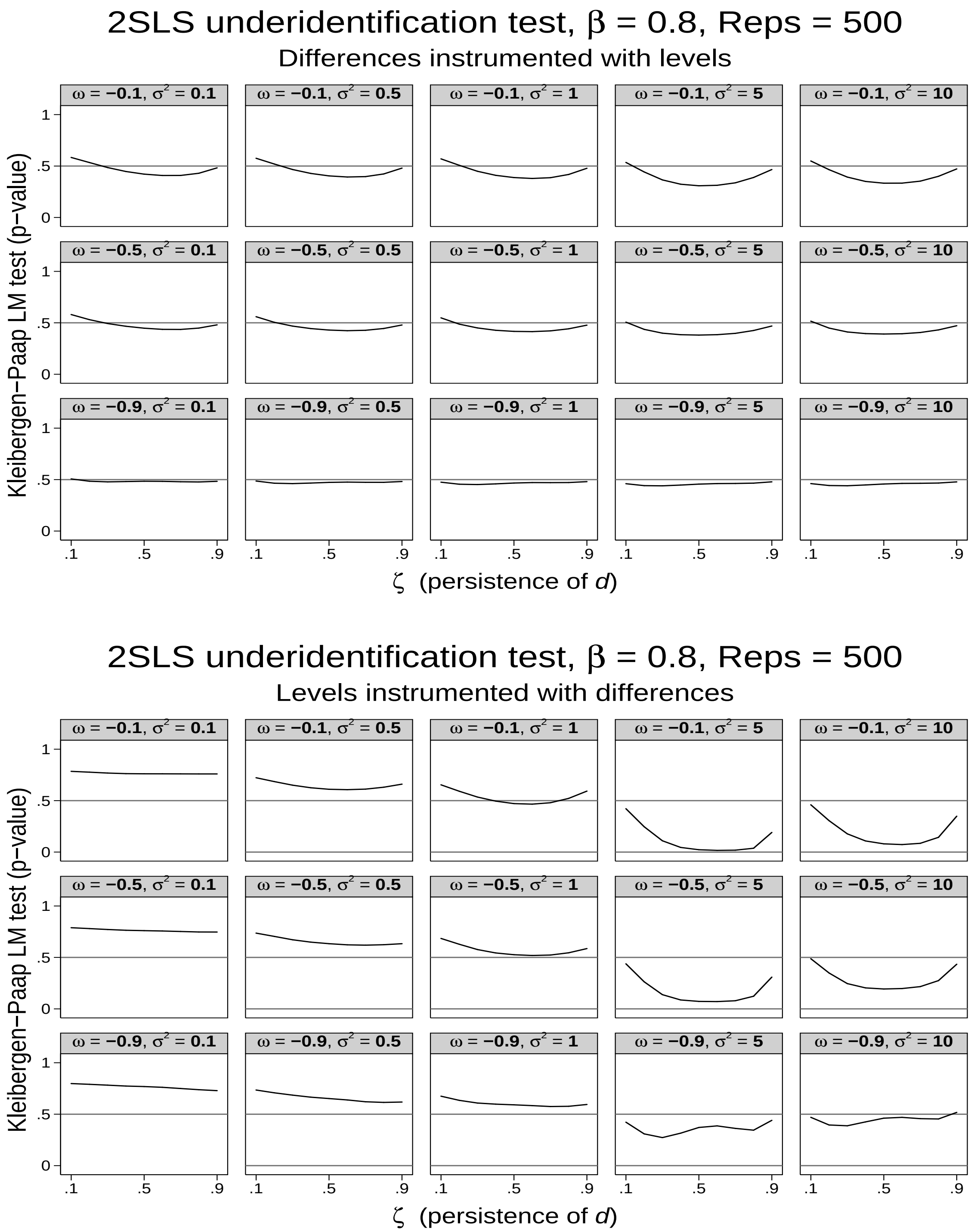
Notes: The dependent variable in all specifications is average annual growth in GDP per capita each period. $^{\mp}$ Column 1 reproduces the published version of Levine et al. (2000, Table 5, Column 3), and column 2 reports our best attempted replication using the `DPD96` program for Gauss, the publicly available dataset, and a Gauss program used to generate their results provided by Thorsten Beck. Further details on the difference in sample sizes across columns, our replication efforts, and the associated differences in the Gauss and Stata programs for dynamic panel GMM regressions can be found in Appendix E.1. The following variables are included in the regressions but suppressed in the table here for presentational purposes: government size, openness to trade, inflation, average years of secondary schooling, black market premium, time period dummies and a constant. The first five of these variables are treated as endogenous. Following the original paper, we report p-values in parentheses. See the notes to Table D.1 for more details on the Kleibergen-Paap and Cragg-Donald tests, which apply in columns 6-9 to the full set of endogenous right-hand-side variables.

Figure D.1: Power and size properties of GMM estimators in simulation results, $\beta = 0.8$



Notes: The graphs show parameter estimates and 95% confidence intervals from simulations of the model in equation (4) of the paper based on 500 draws of a sample size of 600 with 100 cross-sectional units and 6 time periods, fixed $\beta = 0.8$, varying $\zeta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, varying degrees of endogeneity $\omega \in \{-0.1, -0.5, -0.9\}$, and alternative variances of the idiosyncratic shock, $\sigma^2 \in \{0.1, 0.5, 1, 5, 10\}$, where the variance of cross-sectional heterogeneity is fixed at 1. The dashed red line shows the true value of $\gamma = 0.3$ in the simulations.

Figure D.2: Weak identification in simulation results, $\beta = 0.8$



Notes: The graphs show p-values from a Kleibergen-Paap LM test for (the null of) underidentification in the levels and differences equations from simulations of the model in equation (4) in the paper as detailed in the notes to Figure D.1. See the notes to Table D.1 for details on the Kleibergen-Paap test.

E Replicating growth studies

We describe here our replications of the empirical growth studies assessed in Sections 3 and 4.

E.1 Levine et al (2000)

Despite the provision by Levine et al. (hereafter, LLB) of a publicly available dataset (Financial_Intermediation_and_Growth_dataset.xls) on a World Bank website (`http://go.worldbank.org/40TPPEYOC0`), we faced a few difficulties in obtaining an exact replication of their dynamic panel GMM results. Nevertheless, on the basis of our replications efforts described here, we are highly confident that the subsequent OLS and 2SLS results that we report in Tables 6, D.3, and D.4 are those that LLB would have gotten at the time they wrote, with precisely the same data.

In the process of attempting to replicate the original LLB results using exactly the same version of their estimator in Gauss (`DPD96`), the same data, and the same program file provided by one of the LLB authors (Thorsten Beck), we discovered a bug in `DPD96`.[3] The bug produced different two-step GMM estimates across consecutive runs of the same program over the same data, even after reloading the data anew at each run. The result holds for the other two measures as well. While the estimates do not vary wildly, we believe that this sort of non-deterministic potential within this program for the deterministic dynamic panel GMM estimator could explain why the LLB result cannot be reproduced exactly within Gauss (or Stata).

Table E.1 below compares the published parameter estimates in Table 5 of LLB to replications using the original data and the `DPD96` program in Gauss and the `xtabond2` program in Stata. Columns 2, 5, and 8 correspond to the estimates in column 2 of Tables 6, D.3, and D.4, respectively. The replication based on `DPD96` is quite close to the original published estimates. In only one instance does the sign of the parameter estimate differ (inflation for the private credit outcome). Turning to the Stata replications in columns 3, 6, and 9, we find larger differences with the estimates obtained using Gauss despite setting all options in `xtabond2` to mimic the `DPD96` formulation (see Roodman, 2009a). Roodman (2009b) reports similar difficulties replicating their results.[4] The other point to notice is that the sample size apparently differs in the Gauss and Stata replications. This is actually not accurate, though. After inspection of the sample countries and years used in each, we find that the samples are identical and that `DPD96` output does not seem to be reporting the actual sample size.

E.2 Rajan & Subramanian (2008)

The original Rajan and Subramanian dataset and code were kindly provided by the authors. As noted in the paper, we exactly replicate their cross-section and dynamic panel results relevant to our discussion. The analysis in Tables 1, 2, and 5 meanwhile required us to supplement their original dataset 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

[3] Before proceeding to the replication, we removed three countries from the excel dataset, which were not listed as part of the 74 country panel in Table 9 of their published paper.

[4] Our initial efforts at replication were done in consultation with Roodman. Subsequently, after correspondence with Thorsten Beck, we obtained additional input into the Gauss replication. Our Stata replication for private credit slightly differs from that in Roodman (2009b) for two reasons. First, we do not use the Windmeijer (2005) two-step variance correction since this procedure was not available to LLB at the time of their study in the late 1990s. Second, we drop three countries from the publicly available excel dataset, which were not listed among the 74 countries in Table 9.

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 Table 6.1*, since real GDP/capita came from that source. The correlation between the two sources' population estimates is near unity.

In their dynamic panel GMM results, Rajan and Subramanian include the second through seventh lags as instruments for the difference equation in both specifications. They note that they are employing up to eight lags, but given that their panel consists of eight periods and only four of the five year periods since 1985 are actually used due to missing data on their institutional quality measure, their 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 equations regressing differenced endogenous variables on lagged levels. In Table 7 of the paper, 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 corresponding 2SLS regressions in subsequent columns.

E.3 Hausmann et al (2007)

The original Hausmann et al. dataset and code were kindly provided by the authors. In Table D.2, we exactly replicate their original pooled 2SLS and system GMM estimates for their panel based on a five-year periodization. In Table 4, despite applying their original code to the original data, we obtain slightly different estimates from those reported in their published paper for the system GMM specification on the panel with ten-year periodization. The pooled 2SLS estimates are identical. Nevertheless, the differences are trivial and in no way affect our main message in Table 4 (or the key findings in Hausmann et al.'s original paper for that matter).

E.4 DeJong & Ripoll (2006)

The original DeJong and Ripoll dataset and code were kindly provided by the authors. We are able to obtain exact replications of their dynamic panel GMM estimates in Table 2.

E.5 Hauk & Wacziarg (2009)

The original Hauk and Wacziarg dataset and code were kindly provided by the authors. We are able to obtain exact replications of their dynamic panel GMM estimates in Table 13.

E.6 Voitchovsky (2005)

The original Voitchovsky dataset was kindly provided by the author. Using the `DPD98` package (the successor to `DPD96`) for Gauss as originally deployed by the author, we are able to obtain a close replication of the system GMM estimates reported in Table 2 of the published paper. We could not obtain an exact replication of the published results likely due to the bug in the `DPD96` program noted above and inherited by the `DPD98`.

Voitchovsky (2005) constructs a non-standard set of instruments, motivated by arguments against using all the conventional Blundell-Bond moment conditions. For the DIF equation, the instruments include twice and thrice lagged income per capita, lagged investment, the twice lagged and difference in schooling rates, and the twice lagged difference in inequality measures. Anderson and Hsiao (1982) were the first to suggest using twice lagged differences as instruments for the lagged, differenced dependent variable in a dynamic panel setting (see also Arellano, 1989), though the typical Arellano and Bond (1991) or Blundell and Bond (1998) applications instrument contemporaneous differences with lagged levels, retaining the first lagged difference as an instrument for contemporaneous levels. For the LEV equation, the instruments include once lagged and differenced investment and schooling rates; the inequality measures in levels and lagged income per capita are treated (rather unconventionally) as exogenous in the levels equation.

Table E.1: Replicating Levine et al. (2000)

	Original	Replication	Replication	Original	Replication	Replication	Original	Replication	Replication
SYS-GMM Estimator	`DPD96`	`DPD96`	`xtabond2`	`DPD96`	`DPD96`	`xtabond2`	`DPD96`	`DPD96`	`xtabond2`
Collapsed IV matrix	No	No	No	No	No	No	No	No	No
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Liquid liabilities	2.952	2.834	3.176						
	(0.001)	(0.001)	(0.000)						
Private credit				1.522	1.494	1.451			
				(0.001)	(0.001)	(0.000)			
Commercial vs. Central Bank Credit							2.437	2.293	1.383
							(0.001)	(0.001)	(0.000)
Log initial GDP/capita	-0.742	-0.792	-0.525	-0.364	-0.398	-0.268	-0.117	-0.348	-0.225
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.020)	(0.223)	(0.015)	(0.008)
Government size	-1.341	-1.419	0.249	-1.987	-1.841	-0.195	-1.13	-1.088	0.555
	(0.001)	(0.001)	(0.481)	(0.001)	(0.001)	(0.576)	(0.001)	(0.001)	(0.000)
Openness to trade	0.325	0.372	-0.047	0.442	0.499	-0.016	0.497	0.620	0.646
	(0.169)	(0.124)	(0.847)	(0.010)	(0.021)	(0.929)	(0.002)	(0.001)	(0.000)
Inflation	1.748	1.675	1.074	-0.178	0.055	-0.598	-1.772	-2.413	-1.802
	(0.001)	(0.001)	(0.000)	(0.543)	(0.810)	(0.007)	(0.001)	(0.001)	(0.000)
Avg. yrs. secondary school	0.780	0.732	0.041	0.639	0.472	0.195	0.638	0.775	0.935
	(0.001)	(0.001)	(0.743)	(0.001)	(0.001)	(0.128)	(0.001)	(0.001)	(0.000)
Black market premium	-2.076	-2.014	-2.102	-1.027	-1.109	-1.062	-1.044	-1.395	-1.018
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Constant	0.06	1.061	-4.301	4.239	4.042	1.300	-5.677	-4.001	-5.713
	(0.954)	(0.195)	(0.002)	(0.001)	(0.001)	(0.119)	(0.001)	(0.001)	(0.277)
Observations	359	359	345	359	359	345	359	359	345
Number of countries	74	74	74	74	74	74	74	74	74
Number of instruments	75	75	75	75	75	75	75	75	75

Notes: The dependent variable in all specifications is average annual growth in GDP per capita each period. The following variables are included in the regressions but suppressed in the table here for presentational purposes: government size, openness to trade, inflation, average years of secondary schooling, black market premium, time period dummies and a constant. The first five of these variables are treated as endogenous. Following the original paper, we report p-values in parentheses based on two-step estimates without the Windmeijer (2005) correction, which became available after the LLB study.

References

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

Arellano, M., "A note on the Anderson-Hsiao estimator for panel data," *Economics Letters*, 1989, *31* (4), 337–341.

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

Baum, Christopher F., Mark E. Schaffer, and Stephen Stillman, "Enhanced Routines for Instrumental Variables/GMM Estimation and Testing," *Stata Journal*, 2007, *7* (4), 465–506.

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

Dalgaard, Carl-Johan, Henrik Hansen, and Finn Tarp, "On the Empirics of Foreign Aid and Growth," *Economic Journal*, 2004, *114* (496), 191–216.

DeJong, David N. and Marla Ripoll, "Tariffs and Growth: An Empirical Exploration of Contingent Relationships," *Review of Economics and Statistics*, 2006, *88* (4), 625–640.

Hahn, Jinyong, John Ham, and Roger Moon, "The Hausman test and weak instruments," *Journal of Econometrics*, 2011, *160* (2), 289–299.

Hansen, Henrik and Finn Tarp, "Aid and Growth Regressions," *Journal of Development Economics*, 2001, *64* (2), 547–570.

Hauk, William and Romain Wacziarg, "A Monte Carlo Study of Growth Regressions," *Journal of Economic Growth*, 2009, *14* (2), 1381–4338.

Hausmann, Ricardo, Jason Hwang, and Dani Rodrik, "What You Export Matters," *Journal of Economic Growth*, 2007, *12* (1), 1–25.

Hayashi, F., *Econometrics*, Princeton, New Jersey: Princeton University Press, 2000.

Kleibergen, Frank, "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica*, 2002, *70* (5), 1781–1803.

— **and Richard Paap**, "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 2006, *133* (1), 97–126.

Levine, Ross, Norman Loayza, and Thorsten Beck, "Financial Intermediation and Growth: Causality And Causes," *Journal of Monetary Economics*, 2000, *46* (1), 31–77.

Moreira, Marcelo J., "A conditional likelihood ratio test for structural models," *Econometrica*, 2003, *71*, 1027–1048.

Rajan, Raghuram and Arvind Subramanian, "Aid and Growth: What Does the Cross-country Evidence Really Show?," *Review of Economics and Statistics*, 2008, *90* (4), 643–665.

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

— , "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and Statistics*, 2009, *71* (1), 135–158.

Stock, James H. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression," in James H. Stock and Donald W. K. Andrews, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas J Rothenberg*, New York: Cambridge University Press, 2005.

Voitchovsky, Sarah, "Does the Profile of Income Inequality Matter for Economic Growth?," *Journal of Economic Growth*, 2005, *10* (3), 273–296.

Windmeijer, Frank, "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of Econometrics*, 2005, *126*, 25–51.

Yogo, Motohiro, "Estimating the elasticity of intertemporal substitution when instruments are weak," *Review of Economics and Statistics*, 2004, *86*, 797–810.