Online Appendix to:

Growth Regressions for Country Analysis:

A Practitioner's Guide to Data and Methodology

By Wacker, K.M., R.C.M. Beyer, and L.C. Moller

A.1 Construction of infrastructure index

Our approach assembles truncated expectation maximization.

We first take the logs of phone lines, mobile, internet, electricity, and secure internet, all taken from WDI. Since those variables have descending order of availability, we first use the first one (log phone lines) to predict the second one (mobile) with a simple regression model with country fixed effects and country-specific time trends. Where original data on the second variable is missing, those missings are replaced by the regression predictions. The first two variables (consisting of original observations and predicted values where observations were missing) are then used to predict the third one (internet), and so on. Insignificant variables in each step are dropped from the regression. In a final step, missing values of log phone lines are predicted. Then, the procedure is repeated one more time to further increase the number of observations and to update predictions from the first iteration.

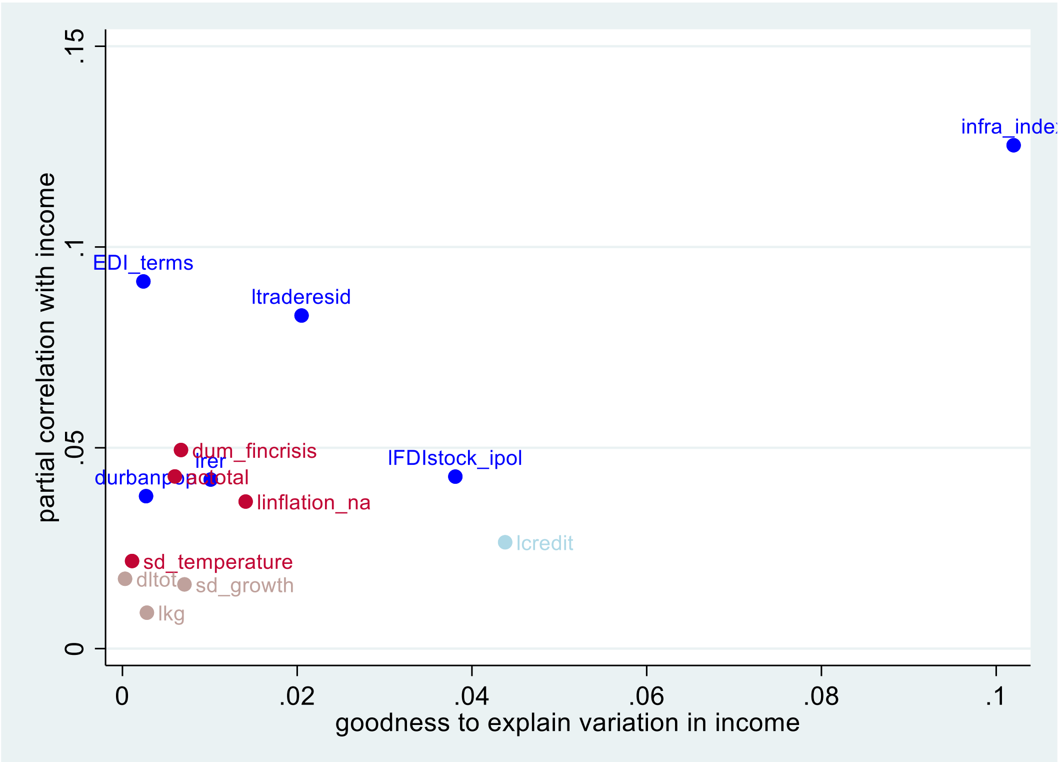
Each of those five updated infrastructure variables is then Pearson standardized and weighted. The first step standardizes all variables to a sample mean of 0 and a sample standard deviation of 1. Their weights in the final infrastructure index are the number of available original observations at each year, divided by the total number of original observations from all five series in the same year. For example, if 150 original observations for log phone lines were available in 1970, and 50 observations were available for electricity, the former gets a weight of 3/4 , and the latter gets a weight of 1/4, while the remaining three infrastructure variables are not part of the composite index in 1970. This weighting reflects two aspects: statistical reliability and economic relevance. Less observations of a variable for a given year mean more predicted values for that variable, which are less reliable than the original observations. Statistically, one hence wants to attribute less weight to this 'less reliable' series. Less observations also reflect less relevance of the indicator. Secure internet connections or mobile phones were simply not available or not as important in the 1970s, their availability was hence not recorded in the data. This constitutes an economic reason for attributing less weight to series with less original observations.

A.2 Detailed results about economic relevance

Table A.2 Standardized beta coefficients

	(1) Medium model (2)	(2) Large model (3)
linflation_na	-0.037*	-0.049***
lrer	0.042**	0.103***
ltraderesid	0.083***	0.062***
infra_index	0.125***	0.152***
dum_fincrisis	-0.049***	-0.040***
sd_temperature	-0.022*	-0.019
dltot	-0.017	-0.018
lkg	-0.009	0.029
sd_growth	-0.016	-0.027
durbanpop	0.038*	-0.027
lcredit	0.026	0.026
lFDIstock_ipol	0.043**	0.031
lEDI_ipol	-0.085*	-0.105**
lEDI_ipol_sq	0.098*	0.107**
actotal	-0.043**	-0.070***
lhc		0.063
dgini		-0.040**
FE	Country & period	Country & period
N	967	635

Figure A.1: Measures of economic relevance

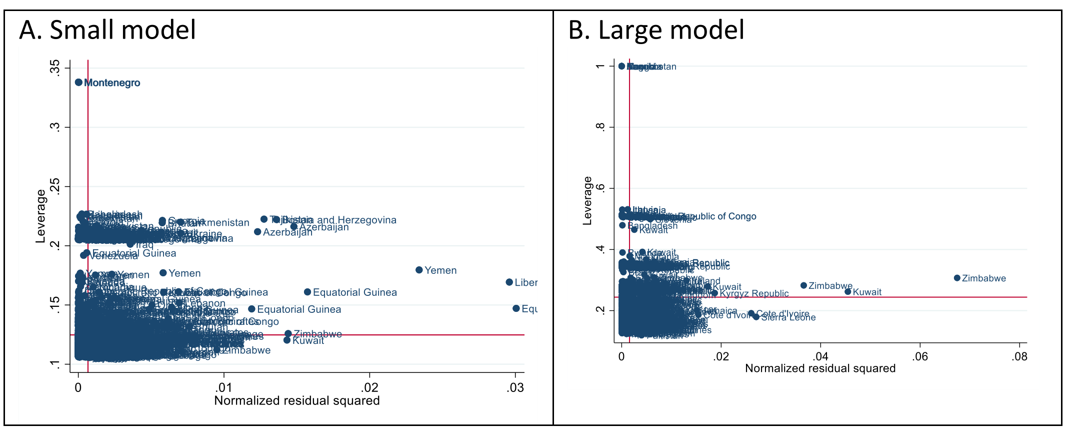


Note: The figure shows standardized beta coefficients (vertical axis) and standardized dominance statistics (horizontal axis) for the medium model. Red indicates a negative coefficient estimate. Light color indicates that the estimate is not statistically different from 0 on a 10% significance level. For export diversification (EDI), which enters with a linear and quadratic term, (absolute) beta coefficients of both terms are averaged and treated as a "set" in the dominance statistic.

A.3 Residual Diagnostics and Model Performance over Periods

Observations that give rise to a large residual ("outliers") and exercise high leverage are influential in linear regression analysis like ours in the sense that they can drive results. This is particularly worrisome if they are not considered representative of the sample (e.g., observations for countries after an outstanding episode of conflict). In Figure A.2, we hence provide leverage-vs.-residual plots of our estimates for the small and large model. Visual inspection suggests that Azerbaijan, Bosnia and Herzegovina, Equatorial Guinea, Liberia, Tajikistan, and Yemen could be potentially influential outliers in the small model, while Zimbabwe may be an influential outlier in the large model. We hence re-estimate both models without those potentially influential countries included. Table A.3 provides the results in columns (1) and (4), which suggest that there are only minor quantitative differences for the small model. Qualitatively, none of the estimates changes sign or its statistical level of significance.[1] Overall, we hence conclude that our main results are not driven by influential outliers.

Figure A.2: Leverage vs. Residual plots



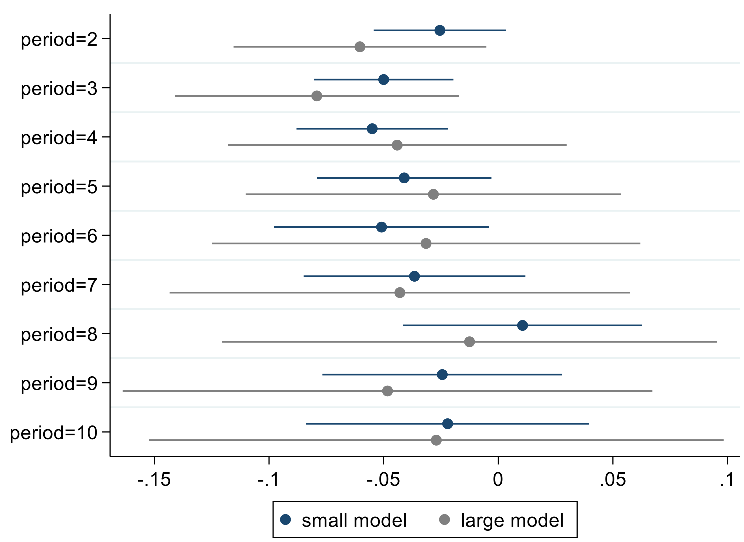
[1] For the large model in column (4), the estimated partial correlation for growth volatility is larger in absolute terms when Zimbabwe is excluded and turns statistically significant at the 10% level.

Table A.3: Results with outliers and different periods excluded

	(1)	(2)	(3)	(4)	(5)	(6)
		small model			*large model*	
VARIABLES	no outliers	earlier smpl	later smpl	no outliers	earlier smpl	later smpl
lagdependent	0.828***	0.766***	0.759***	0.750***	0.711***	0.709***
	(0.0210)	(0.0289)	(0.0296)	(0.0367)	(0.0437)	(0.0401)
linflation_na	-0.216***	-0.203***	-0.226***	-0.143***	-0.110***	-0.114***
	(0.0392)	(0.0422)	(0.0481)	(0.0280)	(0.0297)	(0.0424)
lrer	0.00166	0.00405	0.00438	0.0120***	0.0105***	0.0136***
	(0.00259)	(0.00310)	(0.00368)	(0.00326)	(0.00318)	(0.00358)
ltraderesid	0.0626***	0.132***	0.0813***	0.110***	0.135***	0.0866**
	(0.0201)	(0.0388)	(0.0287)	(0.0322)	(0.0354)	(0.0350)
infra_index	0.0740***	0.0564***	0.104***	0.0974***	0.0613***	0.115***
	(0.0143)	(0.0186)	(0.0187)	(0.0199)	(0.0225)	(0.0196)
dum_fincrisis	-0.0435***	-0.0494***	-0.0398***	-0.0252***	-0.0334***	-0.0141*
	(0.00848)	(0.0101)	(0.00885)	(0.00796)	(0.00994)	(0.00848)
sd_temperature	-0.0316	-0.0369	0.00173	-0.0216	-0.0324	-0.0299
	(0.0196)	(0.0242)	(0.0231)	(0.0231)	(0.0339)	(0.0254)
dltot	-0.0175	-0.0197	-0.0206	-0.0491	-0.0510	-0.145**
	(0.0383)	(0.0408)	(0.0592)	(0.0466)	(0.0396)	(0.0607)
lkg	-0.0348**	-0.0452*	-0.0424**	0.0285	0.0520**	0.0178
	(0.0165)	(0.0230)	(0.0209)	(0.0224)	(0.0248)	(0.0223)
sd_growth	-0.661***	-0.597***	-0.703***	-0.461*	-0.407*	-0.302
	(0.201)	(0.221)	(0.259)	(0.251)	(0.209)	(0.316)
durbanpop	0.00902***	0.0103***	0.00995**	-0.00643	-0.00928	-0.00761
	(0.00285)	(0.00387)	(0.00495)	(0.00563)	(0.00634)	(0.00525)
lcredit				0.0140	0.0259*	0.00381
				(0.0140)	(0.0145)	(0.0113)
lFDIstock_ipol				0.0140	0.0310**	0.0115
				(0.0120)	(0.0140)	(0.0128)
lEDI_ipol				-0.299**	0.00189	-0.370**
				(0.144)	(0.155)	(0.162)
lEDI_ipol_sq				0.131**	-0.0136	0.171**
				(0.0630)	(0.0681)	(0.0679)
actotal				-0.0180***	-0.0171***	-0.0200***
				(0.00445)	(0.00549)	(0.00427)
lhc				0.122	0.298**	0.141*
				(0.0859)	(0.130)	(0.0835)
dgini				-0.0112**	-0.00209	-0.0162***
				(0.00535)	(0.00564)	(0.00491)
Constant	1.733***	2.213***	2.128***	2.507***	2.608***	2.859***
	(0.182)	(0.246)	(0.264)	(0.312)	(0.335)	(0.353)
Observations	1,470	1,171	1,276	631	427	596
R-squared	0.913	0.856	0.874	0.932	0.915	0.912
Number of geo	162	168	168	127	112	128
Estimation	FE	FE	FE	FE	FE	FE
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes

Another concern could be changes in parameters or model performance over time. To gauge the relevance and potential impact of this concern, we first plot the estimates of the time dummies δ_t for the small and large model in Figure 4. If our model, for example, starts to perform increasingly worse in recent decades, e.g., due to structural shifts in the world economy, this would be visible in a time dummy estimate that is increasing over time (in absolute terms). Figure 4 suggests that this is not the case. If anything, we rather observe the time dummies, which can be interpreted as average model errors for all countries in a given period, to increasingly move towards 0 starting from period 8 (2005-2009). We consider this re-assuring for using the model for analyzing more recent growth episodes. Moreover, Table A.3 reports the estimates for the small and large model when dropping the last and first period ("earlier" and "later sample"), respectively. There are some quantitative differences across time-specific sub-samples. Notably, the estimate for infrastructure is larger in the later sample. However, it is reassuring that there are no qualitative differences across time-specific sub-samples, except for export diversification in the larger sample, which only shows the U-shaped pattern in the later sample and is statistically not different from 0 in the earlier sample.

Figure 4: Plot of time dummies



A.4 Additional insights about temperature variability

Given the imminent societal challenge of climate change, this section elaborates on our finding about the negative relationship between temperature variability and growth. First, we provide a discussion of the economic magnitude of this correlation. Remember that the estimated coefficient for *sd_temperature* was -0.0473 (-0.0337) in the medium (large) model, implying that a one unit increase in temperature variability is associated with a 4.73% (3.37%) decrease in income per capita in the short run and 23.9% (12.9%) in the long run (see section 2.1). The standard deviation of temperature variability within countries is around 0.16, implying that typical changes in temperature variability are usually related to income per capita losses of 3.8% (2.1%). To gauge what this could imply going forward, remember that much higher changes in temperature variability can be expected. For example, increases of temperature variability above 0.25 were observed over the last decade in several Asian countries such as Nepal (+0.35), South Korea (+0.29), Bangladesh (+0.27), or Bhutan (+0.26), which implies a much larger magnitude of income losses (more than double in case of Nepal).

Since Ortiz-Bobea et al. (2021) suggest that effects of temperature may be particularly large for agricultural productivity, we interact our *sd_temperature* variable with agricultural value added. At least for the small and medium model, a negative interaction is confirmed: the relationship between temperature variability and income per capita is more negative for countries that rely more heavily on agriculture (see table 3). What do those results imply for a scenario where temperature variability increases by 0.25 units going forward (a modest scenario)? For a country where the agricultural sector accounts for 4.4% of value added (the 0.25 percentile, e.g., Brazil or Zambia in the last period) the medium (small) model implies a long-run income per capita decline of 7.5% (3.5%). For a country where the agricultural sector accounts for 37% of value added (the 0.9 percentile, e.g., Mali or Niger in the last period), the medium (small) model implies a long-run income per capita decline of 10.7% (7.8%).

Table 3: Interaction of temperature variability with agriculture value added (other coefficients omitted)

VARIABLES	(1) small model	(2) medium model
sd_temperature	-0.0253	-0.0576**
	(0.0199)	(0.0238)
c.sd_temperature#c.agri_valadded	-0.00117***	-0.000798***
	(0.000254)	(0.000224)
Observations	1,279	878
R-squared	0.907	0.906
Number of geo	168	149
Estimation	FE	FE
Period FEs	Yes	Yes

Those estimates are no assessments of the causal effects of climate change on growth and should be taken with an additional grain of salt in view of the GMM results in the next subsection. However, they do illustrate that many countries experience an economically relevant negative association between temperature variability and income per capita. It hence appears important further investigate this relationship in future research and to prudently take it into account in the formulation of policy options and scenarios, as the World Bank does in the Country Climate and Development Reports.

A.5 System-GMM results

Our paper is mainly concerned about highlighting typical correlates of economic growth across countries. To meet the profession's ambition of identifying causal effects, it is common in other studies to use instrumental variable strategies for parameter identification. In the absence of plausible external instruments for many applications, the use of internal instruments is a possible alternative (i.e., lags and leads of potentially endogenous variables), which has also been used in growth econometrics from Caselli et al. (1996) to Berg et al. (2018). The most common approach in growth regressions, put forward by Levine et al. (2000) and Bond et al. (2001) is the use of the System-GMM estimator of Arellano and Bover (1995) and Blundell and Bond (1998), which combines internal instruments in a system of level and difference specifications.

While System-GMM offers flexible ways to address potential endogeneity, it is known to be highly susceptible to either weak identification (e.g., Bun and Windmeijer, 2010; Bazzi and Clemens, 2012; Kraay, 2015) or instrument proliferation (e.g., Roodman, 2009). Kraay (2015) illustrates possibilities to address the first problem when one is concerned about single (or a few) variables. Unfortunately, in-depth testing of all instruments (and culling the weak from strong instruments) or providing weak-identification robust inference for all variables appears futile.

To estimate our three main models (small, medium, large) with System-GMM and avoid at least instrument proliferation we follow the general rule of thumb that the set of instruments (which change with variables included) should not outnumber the cross sections (i.e., countries in the sample). We hence proceed as follows. First, we refrain from creating separate instruments by period (i.e., we use the "collapse" option of xtabond2). Second, for the lagged dependent variable, which is endogenous by construction, we always use lags 1 and 2 as instruments in the levels equation. Third, we always treat time dummies as strictly exogenous ("ivstyle" option). Fourth, we separate the remaining variables into those that are likely to be persistent and trending[2] and those which are plausibly not[3]. Our key concern relates to the former because they are likely to develop in tandem with income p.c. levels, giving rise to concerns about simultaneity bias. We hence instrument the former in both equations, while the latter are instrumented only in the levels equation. Finally, we chose the instrument lag dept for both sets of variables from lags 1 to *b*, where *b* ensures that the overall number of instruments is maximized under the condition that it does not outnumber the number of countries in the estimation sample. Since this strategy does not leave us with many instruments for the small model, we decided not to collapse the instrument set for the last group of (non-persistent/non-trending) variables. This leaves us with *b* = 3, 8, and 6 for the small, medium, and large sample, respectively. A generalized inverse is used to calculate the optimal weighting matrix for two-step estimation. Results, which we consider a rough cross-check on key model parameters, are reported in Table A.4.

The Hansen test of all three model specifications indicates that we cannot reject the null hypothesis that the over-identifying restrictions of the instrument set are valid, although the p-value of 0.15 in the small model is rather modes. At the same time, the fact that p-values are not too high suggests that proliferation of instruments is not a key concern.

[2] lrer, ltraderesid, infra_index, sd_temperature, lcredit, lFDIstock_ipol, lEDI_ipol, lEDI_ipol_sq, lhc.

[3] linflation, dum_fincrisis, dltot, lkg, sd_growth, durbanpop, actotal, dgini.

Table 4: System-GMM results

VARIABLES	(1) lrgdpna_pc	(2) lrgdpna_pc	(3) lrgdpna_pc
lagdependent	0.905***	0.896***	0.874***
	(0.0158)	(0.0248)	(0.0193)
linflation_na	-0.132*	-0.0783	-0.122**
	(0.0724)	(0.0742)	(0.0556)
lrer	-0.000782	0.000359	0.000998
	(0.00259)	(0.00275)	(0.00412)
ltraderesid	0.0534	0.108*	0.0187
	(0.0332)	(0.0617)	(0.0403)
infra_index	0.109***	0.105***	0.132***
	(0.0193)	(0.0263)	(0.0367)
dum_fincrisis	-0.116***	-0.0460**	-0.0307
	(0.0286)	(0.0220)	(0.0219)
sd_temperature	0.0112**	0.0157***	0.0125**
	(0.00500)	(0.00519)	(0.00562)
dltot	0.107	-0.297**	-0.0575
	(0.106)	(0.129)	(0.0717)
lkg	0.00204	-0.0131	0.00921
	(0.0266)	(0.0265)	(0.0263)
sd_growth	-0.740**	-0.441	-0.923
	(0.317)	(0.555)	(0.571)
durbanpop	0.00911*	0.0165*	0.00308
	(0.00543)	(0.00891)	(0.00687)
lcredit		0.0141	0.0234
		(0.0156)	(0.0164)
lFDIstock_ipol		0.00880	0.00375
		(0.00648)	(0.0124)
lEDI_ipol		0.0284	0.119
		(0.267)	(0.251)
lEDI_ipol_sq		0.0149	-0.0593
		(0.113)	(0.109)
actotal		-0.00459	-0.00455
		(0.00996)	(0.00547)
lhc			0.0679
			(0.102)
dgini			-0.00519
			(0.00658)
Observations	1,507	967	635
Number of geo	168	149	128
# of instruments	166	140	123
Hansen p value	0.335	0.430	0.150
AR(2) AB z value	0.001	0.018	0.050
Estimation	GMM	GMM	GMM
Sample	small model	medium model	large model

Table A.4 shows that the persistence in income p.c. increases, compared to the fixed effects estimate, which is consistent with the expected downward bias of the autoregressive parameter estimate under fixed effect estimation ("Nickell bias"). The most striking difference to the fixed effect specification is the positive correlation between income p.c. and temperature variability, which appears to be a rather robust finding under different System-GMM specifications we considered and which we put up for further investigation in future research.

Another result emerging from the System-GMM results in Table 4 is that, except for infrastructure, few parameter estimates remain statistically different from 0 across model specifications.

Overall, we do not think that the identification potential of System-GMM outweighs its well-known problems for broad descriptive policy analysis. While endogeneity concerns in ordinary fixed effects estimation are well-studied, the same cannot be said for GMM estimation under problems like weak identification (despite recent progress; e.g., by Andrews and Mikusheva, 2023).