

Web Appendix

Web Appendix A1.

Data description of measures of culture and institutions

Variable name	Description	Source	Download from
Culture			
PDI	Power Distances focuses on the degree to which inequality in the distribution of power is accepted and expected in a society.	Hofstede (2001) ^a	Geert Hofstede's website: http://www.geert-hofstede.com/
IDV	Individualism concerns the extent to which a society reinforces individual or collective achievement.	-	-
MAS	Masculinity measures the extent to which a society reinforces the traditional work role model of male achievement, control and power.	-	-
UAI	Uncertainty Avoidance focuses on the level of tolerance for uncertainty and ambiguity within a society.	-	-
LTO	In a later survey, Hofstede identified a fifth dimension of culture: Long Term Orientation. Values associated with Long Term Orientation are thrift and perseverance; values associated with Short Term Orientation are respect for tradition, fulfilling social obligations, and protecting one's face.	-	-
al_ethnic	Ethnic fractionalization	Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003) Fractionalization, <i>Journal of Economic Growth</i> , 8: 155-194.	QOG: http://www.qog.pol.gu.se/
al_language	Linguistic fractionalization	-	-
al_religion	Religious fractionalization	-	-
dpi_tf	Total fractionalization	Database of political	-

^a The Hofstede dataset from 1967 and 1973 contained 52 observations, including three regions. The regions are Arab World (Egypt, Iraq, Kuwait, Lebanon, Libya, Saudi Arabia, United Arab Emirates), East Africa (Ethiopia, Kenya, Tanzania, Zambia), and West Africa (Ghana, Nigeria, Sierra Leone). The IBM survey was conducted in all of these countries, but Hofstede gathered the observations into regions because of few observations and loss of data. I have split up the regions, obtaining a total of 63 observations. Last, the Hofstede dataset was later extended to include 16 more countries. This amounts to a total of 79 observations. Controlling for these extensions of the data (in addition to that already controlled for by including regional dummies) does not change the results much: I have run the regressions on the much smaller sample of 49 original countries, which still produces strong instruments in Table 1, except columns (4) and (7). The estimate on corruption increases in all columns on this smaller sample. Instead of reducing the sample, I have run the regressions including the dummy "extension" equal to zero, if the observation is among the original 49 single-country observations, one otherwise. This dummy should pick up additional effects from both the regional extension, but also the later extension of the dataset done by Hofstede. The regional dummies included in most of the regressions should pick up potential effects of the regional extensions. The same exercise was done for Table 2, where including the additional dummy "extension" does not change the results, but running the regressions on the smaller sample of 49 countries only produces strong instruments in columns (2), (3), (5), (6), and (10).

		institutions	
el_avef	Average Value of Ethnolinguistic Fractionalization	Easterly, W. and R. Levine (1997) Africa's Growth Tragedy: Policies and Ethnic Divisions, Quarterly Journal of Economics, 4: 1203-1250.	-
el_elf60	Commonly used ELF-index		-
fe_cultdiv	Cultural Diversity	Fearon, J.D. (2003) Ethnic and Cultural Diversity by Country, Journal of Economic Growth, 8: 195-222.	-
fe_efra	Ethnic Fractionalization	-	-
ht_colonial	Colonial Origin	Hadenius, A. and J. Teorell (2005) Assessing Alternative Indices of Democracy, C&M Working Papers 6, IPSA, August 2005	-
lp_catho80	Religion: Catholic	La Porta, R., F. López-de-Silanes, A. Shleifer, and R. Vishny (1999) The Quality of Government, Journal of Law, Economics and Organization, 15(1): 222-279.	-
lp_muslim80	Religion: Muslim	-	-
lp_no_cpm80	Religion: Other Denomination	-	-
lp_protmg80	Religion: Protestant	-	-
r_atlas	Ethnolinguistic Fractionalization	Roeder, P. G. (2001) Ethnolinguistic Fractionalization (ELF) Indices, 1961 and 1985.	-
r_elf61	Ethnolinguistic Fractionalization, 1961	-	-
r_elf85	Ethnolinguistic Fractionalization, 1985	-	-
r_muller	Ethnolinguistic Fractionalization	-	-
r_roberts	Ethnolinguistic Fractionalization	-	-
wvs_auth	Respect for authority	World Values Survey	-
wvs_auton	Autonomy index	World Values Survey	-
wvs_e114m	Having a strong leader, mean	World Values Survey	-
wvs_e114p	Having a strong leader, %	World Values Survey	-
wvs_e124m	Respect for individual human rights, mean	World Values Survey	-
wvs_e124p	Respect for individual human rights, %	World Values Survey	-
wvs_rel	Religiousness	World Values Survey	-
wvs_reln	Religiosity Scale (mean)	World Values Survey	-
Institutions			
rl2006	Rule of law, 2006	Kaufmann et al. (2009)	http://info.worldbank.org/governance/wgi/index.asp
rq2006	Regulatory quality, 2006	-	-
ge2006	Government Effectiveness, 2006	-	-
ps2006	Political Stability No Violence, 2006	-	-
va2006	Voice and Accountability, 2006	-	-
p_xconst	Constraint on executive	Polity IV dataset	QOG: http://www.qog.pol.gu.se/
bti_prp	Private property	Bertelsmann Transformation Index	-
bti_rol	Rule of law	-	-
ciri_speech	Freedom of speech	Cingranelli, D. L. and D. L. Richards (1999) Measuring the Level, Pattern, and Sequence of Government Respect for Physical Integrity Rights, International Studies Quarterly, 43(2): 407-418.	-
dlls1_fie, dlls1_fic, dlls1_tde, dlls1_tdc	Measures substantive and procedural statutory intervention in two	Djankov, S., R. La Porta, F. López-de-Silanes, and A. Shleifer (2003) Courts: The	-

	forms of judicial cases at lower-level civil trial courts: the eviction of a residential tenant for nonpayment of rent (formalism: dlls1_fie, total duration: dlls1_tde), and the collection of a check returned for nonpayment (formalism: dlls1_fic, total duration: dlls1_tdc).	Lex Mundi Project, Quarterly Journal of Economics, 118: 453-517.	
kk_gg	Index of Objective Indicators of Good Governance	Knack, S. and M. Kugler (2002) Constructing an Index of Objective Indicators of Good Governance, PREM Public Sector Group, World Bank.	-
fh_rol	Rule of law	Freedom House	-
fi_legprop	Legal system and property rights	Fraser Institute	-
gir_ga	Government accountability	Global Integrity (2007) The Global Integrity Report 2007. Methodology Whitepaper.	-
h_polcon3	Political constraint index	Henisz, W. J. (2002) The Institutional Environment for Infrastructure Investment, Industrial and Corporate Change, 11(2): 355-389	-
h_polcon5	Political constraint index	Henisz, W. J. (2000) The Institutional Environment for Economic Growth, Economics and Politics, 12(1): 1-31.	-
hf_prights	Heritage Foundation property rights	Heritage Foundation	-
icrg_qog	ICRG indicator of quality of government	International Country Risk Guide	-
irai_prrg	Property rights and rule based government	IDA Resource Allocation Index	-
irai_qpa	Quality of public administration	-	-
irai_tac	Transparency, Accountability, and Corruption in the Public Sector	-	-
p_polity	Combined polity score	Polity IV	-
p_polity2	Revised combined polity score	-	-
no_procedure	Number of procedures of resolving a court case involving nonpayment of a commercial debt.	World Bank (2004) Doing Business in 2004: Understanding Regulation. Washington, DC: Oxford University Press (for World Bank).	http://www.doingbusiness.org/
proc_compl	Procedural complexity of resolving a court case involving nonpayment of a commercial debt.	-	-

Web Appendix A2.

Principal component analysis

Principal component analysis (pca) allows me to replace a set of highly correlated regressors by their principal components (pc). This eliminates potential multicollinearity bias and also increases the precision of the corruption estimate by reducing the number of control variables.

Using principal components analysis produces principal components that span the entire included dataset and that are orthogonal to one another. In the present problem, I include all 13 control variables from column (7) of Table 1 (7 regional dummies and 6 variables: *logdistcr*, *logelev*, *fh_rol*, *logh_polcon5*, *logproc_compl*, and *lp_protmg80*), producing 12 principal components. Table A1 shows the variance contribution of the principal components and Table A2 shows the correlation between the pc's and the 13 variables (this is the output of the simple pca procedure in STATA).

Table A1 and A2 here

I choose which principal components to include in the regression analysis using two different strategies (Jolliffe, 2002). These are both represented in Table A3 that includes OLS regressions of corruption on GDP per capita, including the principal components as control variables. In column (1) of Table A3, I follow a standard rule and include all the principal components with an eigenvalue above 1. These 7 principal components span 81% of the total variation in the included 13 variables. This method of choosing the principal components makes sure that we get as much of the variance in the data, which is less important in the present analysis. Instead, in columns (2) through (11) of Table A3, I pick the principal components based on their explanatory power vis a vis economic development. Column (2) includes all 12 components, column (3) excludes the single component with the lowest level of significance and so on. I continue excluding components until all the included components are significant. This latter strategy suits the present analysis better, as the purpose is to include the variables that best span the entire set of deep determinants of GDP per capita. Column (11) of Table A3 is the same regression as that in column (9), panel A of Table 1. The three

significant components mainly capture variation in geography and institutions (see Table A2). More specifically, the regional dummies (pc3, pc5), elevation (pc5), and property rights institutions (pc8).

Table A3 here

Table A4 shows the corresponding TSLS regressions. Column (1) includes the 7 principal components with eigenvalue above 1. Columns (2) through (8) exclude the insignificant principal components one by one. I end up with six significant principal components, compared to only three, when the exclusion was based on the OLS regressions. Column (8) of Table A4 is the same regression as column (9) of panel B in Table 1. These six components capture variation in all deep determinants (see Table A2). Specifically the regional dummies (pc3, pc4), rule of law (pc1), property rights institutions (pc1, pc8), contracting institutions (pc10), and culture (pc9, pc10). The estimate of corruption is again larger than the estimate in Table 1, column (2). The instruments are fairly strong with TSLS size distortions below 15% and TSLS bias below 5%. The corresponding LIML estimates are very similar (not shown). For comparison, column (9) of Table A4 includes the three significant components from the OLS analysis.

Table A4 here

Table A1. Pca, the variance contributions of the principal components

Component	Eigenvalue	Difference	Proportion	Cumulative
pc1	2.87	0.97	0.22	0.22
pc2	1.90	0.57	0.15	0.37
pc3	1.32	0.09	0.10	0.47
pc4	1.24	0.10	0.10	0.56
pc5	1.14	0.06	0.09	0.65
pc6	1.08	0.04	0.08	0.73
pc7	1.04	0.28	0.08	0.81
pc8	0.76	0.12	0.06	0.87
pc9	0.64	0.14	0.05	0.92
pc10	0.49	0.16	0.04	0.96
pc11	0.33	0.13	0.03	0.98
pc12	0.20	0.20	0.02	1.00

Notes. Principal component analysis, unrotated. Included variables: ssa, soa, mena, eap, eca, lac, na, logdistcr, logelev, fh_rol, log_hpolcon5, logproc_compl, lp_protmg80. 119 observations.

Table A2. Principal components (eigenvectors), correlations

Variable	pc1	pc2	pc3	pc4	pc5	pc6	pc7	pc8	pc9	pc10	pc11	pc12
ssa	-0.25	0.42	0.20	-0.21	-0.42	-0.15	-0.02	0.35	-0.27	-0.02	-0.24	-0.02
soa	-0.09	-0.04	-0.23	0.04	0.04	0.75	-0.45	0.31	0.09	0.11	0.07	0.01
mena	-0.17	-0.15	-0.43	0.06	0.42	-0.53	-0.14	0.31	0.21	0.04	0.02	0.16
eap	0.06	0.04	-0.24	0.69	-0.23	0.10	0.46	-0.13	-0.05	0.20	0.08	0.04
na	0.10	0.28	0.21	0.29	0.46	-0.06	-0.42	-0.34	-0.44	0.18	-0.15	-0.08
eca	0.35	-0.03	-0.13	-0.56	0.21	0.16	0.29	-0.33	0.00	0.03	0.06	-0.12
lac	-0.05	-0.43	0.54	0.18	-0.13	-0.05	-0.23	-0.19	0.27	-0.33	0.12	0.04
logdistcr	-0.33	0.45	0.18	-0.05	0.21	0.05	0.12	-0.06	0.09	-0.02	0.73	0.21
logelev	-0.27	0.14	0.31	0.08	0.44	0.23	0.39	0.11	0.39	0.01	-0.50	-0.06
fh_rol	0.51	0.04	0.21	0.01	0.08	0.06	0.06	0.27	-0.06	0.01	-0.08	0.77
logh_polcon5	0.43	0.00	0.23	0.15	0.18	-0.05	0.12	0.56	-0.07	-0.08	0.29	-0.54
logproc_comp	-0.23	-0.45	0.29	-0.15	0.01	-0.02	0.11	0.11	-0.21	0.75	0.12	0.03
lp_protmg80	0.30	0.33	0.08	-0.01	-0.22	-0.17	-0.25	-0.09	0.63	0.49	-0.01	-0.11

Notes. Principal component analysis, unrotated. Included variables: ssa, soa, mena, eap, eca, lac, na, logdistcr, logelev, fh_rol, log_hpolcon5, logproc_compl, lp_protmg80. 119 observations.

Table A3. OLS regressions including principal components

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable: rgdpl22006											
cci2006	9.368***	11.062***	11.077***	11.022***	10.404***	10.238***	10.262***	10.136***	10.084***	10.187***	10.238***
	(1.128)	(1.147)	(1.131)	(1.137)	(0.553)	(0.547)	(0.537)	(0.568)	(0.574)	(0.572)	(0.567)
pc1	0.384	-0.406	-0.413	-0.388							
	(0.645)	(0.649)	(0.643)	(0.641)							
pc2	-0.468	-0.555	-0.556	-0.553	-0.521	-0.513	-0.514	-0.508			
	(0.458)	(0.417)	(0.415)	(0.415)	(0.389)	(0.397)	(0.394)	(0.388)			
pc3	-1.190*	-1.122**	-1.121**	-1.123**	-1.148**	-1.155**	-1.154**	-1.159**	-1.161**	-1.157*	-1.155*
	(0.608)	(0.541)	(0.539)	(0.537)	(0.560)	(0.554)	(0.562)	(0.572)	(0.575)	(0.584)	(0.590)
pc4	-0.204	-0.265	-0.266								
	(0.499)	(0.449)	(0.447)								
pc5	1.365*	1.009**	1.006**	1.018**	1.147**	1.182**	1.177**	1.203**	1.214**	1.193**	1.182**
	(0.710)	(0.505)	(0.503)	(0.504)	(0.483)	(0.485)	(0.475)	(0.507)	(0.494)	(0.473)	(0.499)
pc6	-0.916	-0.728	-0.727	-0.733	-0.801	-0.819	-0.817	-0.831	-0.837		
	(0.601)	(0.516)	(0.515)	(0.513)	(0.566)	(0.558)	(0.564)	(0.582)	(0.593)		
pc7	-0.499	-0.419	-0.418	-0.421	-0.450	-0.458					
	(0.582)	(0.477)	(0.476)	(0.476)	(0.485)	(0.484)					
pc8		-2.780***	-2.782***	-2.775***	-2.696***	-2.675***	-2.678***	-2.662***	-2.655***	-2.668***	-2.675***
		(0.632)	(0.630)	(0.631)	(0.619)	(0.627)	(0.631)	(0.653)	(0.662)	(0.673)	(0.688)
pc9		-1.164	-1.167	-1.157	-1.043	-1.012	-1.017				
		(0.883)	(0.882)	(0.885)	(0.809)	(0.812)	(0.803)				
pc10		1.044	1.042	1.047	1.102	1.117	1.115	1.126	1.131	1.121	
		(0.801)	(0.799)	(0.794)	(0.763)	(0.765)	(0.769)	(0.764)	(0.738)	(0.741)	
pc11		-0.176									
		(0.826)									
pc12		-1.608	-1.617	-1.583	-1.195						
		(1.280)	(1.269)	(1.263)	(1.307)						
Observations	119	119	119	119	119	119	119	119	119	119	119
R-squared	0.768	0.814	0.814	0.813	0.812	0.811	0.809	0.805	0.802	0.797	0.793
F-test cci2006 = 9.223, p	0.898	0.112	0.104	0.117	0.0348	0.0664	0.0555	0.111	0.136	0.0946	0.0759

Notes. OLS estimates. Dependent variable is real GDP per capita in 2006 from PWT. The principal components are those produced from the principal components analysis of Tables A1 and A2. Robust standard errors in paranthesis. Constant included in all regressions. Asterisks *, **, and *** indicate significance at the 10, 5, and 1% level, respectively.

Table A4. TSLS regressions (second stage, cci2006 endogenous) including principal components

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable: rgdpl22006									
cci2006	9.565*** (2.368)	15.062*** (5.239)	15.395*** (2.676)	15.411*** (2.670)	14.371*** (2.152)	13.636*** (2.060)	14.069*** (2.076)	14.606*** (1.968)	10.207*** (0.854)
pc1	0.450 (1.277)	-3.092 (2.922)	-3.260** (1.404)	-3.263** (1.404)	-2.804** (1.171)	-2.550** (1.131)	-2.763** (1.146)	-3.015*** (1.134)	
pc2	-0.810 (0.510)	-0.660 (0.548)	-0.680 (0.599)	-0.688 (0.597)	-0.524 (0.588)				
pc3	-1.690** (0.761)	-1.689*** (0.577)	-1.692*** (0.557)	-1.700*** (0.554)	-1.656*** (0.559)	-1.608*** (0.564)	-1.580*** (0.590)	-1.600** (0.623)	-1.799** (0.755)
pc4	0.065 (0.604)	-1.177 (0.855)	-1.203* (0.660)	-1.208* (0.659)	-1.151* (0.618)	-1.205* (0.620)	-1.271** (0.611)	-1.328** (0.622)	
pc5	2.279** (0.904)	0.020 (1.729)							2.186*** (0.689)
pc6	-1.295 (0.870)	-0.882 (0.811)	-0.844 (0.704)	-0.827 (0.672)	-0.986 (0.663)	-1.091 (0.704)	-1.053 (0.728)		
pc7	-0.660 (0.591)	-0.780 (0.594)	-0.792 (0.541)	-0.811 (0.548)	-0.753 (0.555)	-0.651 (0.558)			
pc8		-2.776*** (1.051)	-2.817*** (0.928)	-2.831*** (0.912)	-2.545*** (0.887)	-2.442*** (0.914)	-2.602*** (0.969)	-2.734*** (1.012)	-2.293** (0.968)
pc9		-3.076 (2.169)	-3.137** (1.371)	-3.111** (1.394)	-3.018** (1.358)	-2.947** (1.290)	-3.023** (1.311)	-3.018** (1.330)	
pc10		1.695 (1.034)	1.659 (1.129)	1.645 (1.117)	1.927* (1.078)	2.088** (1.002)	2.048** (1.017)	2.027* (1.041)	
pc11		0.246 (1.313)	0.236 (1.261)						
pc12		-2.760 (4.320)	-2.982 (2.761)	-3.030 (2.674)					
Observations	69	69	69	69	69	69	69	69	69
R-squared	0.782	0.816	0.815	0.814	0.814	0.815	0.810	0.800	0.783
OID p-value	0.126	0.350	0.336	0.343	0.179	0.228	0.261	0.248	0.216
Kleibergen-Paap F	16.22°	3.314	11.75°	11.83°	15.15°	15.14°	15.92°	17.02°	80.25°
Cragg-Donald F	10.72°	3.511	10.88°	11.09°	13.25°	13.44°	12.68°	13.82°	66.37°
F-test cci2006 = 8.706, p	0.717	0.225	0.012	0.012	0.008	0.017	0.009	0.003	0.079
Robust Hausman test, p	0.631	0.786	0.451	0.436	0.677	0.886	0.714	0.551	0.923

Notes. TSLS second stage estimates. Dependent variable is real GDP per capita in 2006 from PWT. The principal components are those produced from the principal components analysis of Tables A1 and A2. Robust standard errors in paranthesis. Corruption is instrumented with Hofstede s Power Distances (PDI) and Individualism (IDV). Constant included in all regressions. Asterisks *, **, and *** indicate significance at the 10, 5, and 1% level, respectively. Dots °, °°, and °°° indicate TSLS size distortions of a maximum of 20, 15, and 10%, respectively.