Institutions, Human Capital and Development: Online Appendix

Daron Acemoglu **Francisco A. Gallego** **James A. Robinson**

February 26, 2014

Abstract

This online appendix presents (i) a detailed description of the construction of the proxies used in this paper for population density before colonization at the regional level, (ii) Tables including the coefficients estimated for the control variables in the first and second stages in Tables 4 to 8 (these tables present the same numbers as in the main text in order to make the comparison easier), and (iii) a table including estimates for exactly-identifies models.

1 Pre-Colonial Population Density before Colonization

We use data from Bruhn and Gallego (2012) for population density for Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Salvador, Guatemala, Honduras, Mexico, Panama, Paraguay, Peru, the US, Uruguay, and Venezuela. For the other countries included in the sample we use country-specific sources or the HYDE 3.1 Dataset from Goldewijk et al. (2010).

1.1 Country-Specific Sources

Australia

We construct estimates from National Population Inquiry (1975). The estimates correspond to 1788, the same year Britain established a penal colony in Sydney, New Southern Wales.

Benin, Cameroon, Ghana, Nigeria and Democratic Republic of the Congo

We constructed data for these countries using African historical population estimates from http://www.dataverse.pitt.edu/ (described in Manning, 2013). In short, this source presents estimates for roughly 70 African regions by decades from 1850 to 1950. The regions are selected considering the main precolonial slave trade regions and colonial and post-colonial frontiers, among other factors. Because of this, sometimes a single country –considered with its present frontier- will cross over two or more Manning regions. In these cases we where able to generate within country population variability. The variability was generated taking the relevant administrative unit –region, province, etc- and imputing the population density of the corresponding Manning region. If a single administrative unit was included in two or more Manning regions, we averaged the population density of the corresponding Manning regions weighted by the area of the administrative. The main assumption here is that the population is distributed uniformly in each Manning region; while this is clearly a very strong assumption, the estimates probably still give us a pretty good sense of the order of magnitude of the population with variability within the country.

Egypt

We take data from Jan Lahmayer's World population estimates (http://www.populstat.info/.)

Indonesia

We take data from Nitisastro (2006), who provides a extensive demographic study of Indonesia. The historical estimates are for the six main islands of the country, which is not the ideal

disaggregation (because the relevant administrative unit is smaller) but provides within country variability. We impute the value of the whole island to all the provinces that belong to that island. Once again, the main assumption here is that the population distributes uniformly in each of the six islands.

Sri Lanka

The source is de Zwart (2011), that shows population estimates for two regions (Maritime and Kandy) in 1789. This data can be interpreted as the pre-British population and enable us to generate within country variability. As before, we imputed the population density of Maritime and Kandy to the modern provinces that are contained in them, and we use a weighted average when part of a province is in Kandy and part in Maritime.

New Zealand

We take data from Cumberland (1950) on pre-contact population for the Maori people for the whole country and combined it with a pre-contact estimate of the distribution of the population to construct population density estimates for three large areas and then impute those values to the modern regions as in previous countries.

Vietnam

Similar to New Zealand, for Vietnam we took estimates for the whole country in 1850 from Jan Lahmayer's World population estimates and combined them with a proxy of the distribution of the population in Vietnam in 1847 to construct the population density for five areas. After that, we matched these five areas with the eight modern Vietnam regions and imputed the corresponding value. The proxy was taken from Langlet-Quach Thanh Tâm (2000) and shows the spatial repartition of male registered voters ages 18 to 65. Thus, the main assumption for this to be a good proxy for population distribution is that the distribution of male registered voters ages 18 to 65 is similar to the one of the whole population.

1.2 HYDE 3.1 Database (Goldewijk et al., 2010)

For countries not included in Bruhn and Gallego (2012) and for which we did not find country-specific sources of information, we use data from Goldewijk et al. (2010). The HYDE 3.1 database provides population estimates for the Holocene (approximately the period 10000 BC to AD 2000) for the whole World presented in a raster format with a resolution of 5 min longitude/latitude. Relevant to this paper, the database presents estimates every hundred years

until 1700, where they start to be every 10 years. The estimates are constructed using several sources, some of which where used by Bruhn and Gallego (2012)–e.g. Denevan, 1992. A more detailed explanation of the methodology and sources can be found in Goldewijk et al. (2010).

As the resolution of the grid that contains the data is smaller than every administrative unit that we where interested in, we aggregate the data at the desired level using GIS software. As every grid is of the same size, the aggregation is simply an average of the value of the grid cells that are contained in a administrative unit.

As before, for each country we chose a single precolonial date for each country, sufficiently close to the colonization date. When the colonization occurred after 1700, the date we chose was at the most 9 years before the colonization date; for example, if the colonization was in 1769 the chosen date was 1760. Nevertheless, in this same example, we also compared the population of 1770 (that is closer to 1769) with the population from 1760 to see if using one or the other could cause different results: it was never the case that populations 10 years apart where significantly different from each other.

When the colonization was before 1700 we also used the closest date before the colonization. As the estimates where more years apart, using an alternative dates sometimes did change significantly the estimates. Nevertheless, conceptually is important to choose a single date for the whole country that is close but previous to the colonization date of any if the country's administrative units.

Finally, following Bruhn and Gallego (2012) for unpopulated regions we use an estimate of 0.000069, which is consistent with the minimum population density of the populated regions in the HYDE dataset.

References

[1] Cumberland, Kenneth (1950), A Land Despoiled: New Zealand about 1838, Whitcombe and Tombs.

[2] De Zwart, Pim (2011), "Labour Relations in Ceylon, ca. 1650",Global Collaboratory on the History of Labour Relations, 1500- 2000: collab.iisg.nl/web/labourrelations/ceylon.

[3] Goldewijk, Klein, Beusen and Janssen (2010) "Long-term dynamic modeling of population and built-up area in a spatially explicit way: HYDE 3.1", The Holocene June 2010 20: 565-573.

[4] Langlet-Quach Thanh Tâm (2000), “La répartition spatiale de la population”, Population et développement au Viêt-nam http://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_7/b_fdi_03_05/010023011.pdf)

[5] Manning, Patrick (2009), "Replication data for: African Population Estimates, 1850-1960", http://hdl.handle.net/1902.1/15281 World-Historical Dataverse [Distributor] V1 [Version]

[6] National Population Inquiry (1975), *Population and Australia: A Demographic Analysis and Projection*, Australian Government Publishing Service, Canberra

[7] Nitisastro, Widjojo (2006), Population trenes in Indonesia, Equinox Publishing; 1st Equinox Ed Edition

2 Complete Estimates of Tables 4 to 8

Table 4: Semi-structural regressions, years of schooling, cross-country sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Panel A: Second-stage regressions											
	Dependent variable is log GDP per capita in 2005											
Estimation Method	2SLS								LIML			
Years of schooling	0,314	0,305	0,274	0,317	0,177	0,171	0,131	0,178	0,177	0,171	0,122	0,170
	(0,054)	(0,054)	(0,101)	(0,116)	(0,106)	(0,106)	(0,128)	(0,134)	(0,106)	(0,106)	(0,135)	(0,144)
A. R. confidence intervals	[.17, .44]	[.16, .44]	[-.00, .48]	[-.01, .56]	[-.15, .41]	[-.16, .40]	[-.34, .43]	[-∞, .55]	[-.16, .41]	[-.16, .40]	[-.35, .43]	[-.36, .54]
Dummy for different source of protestant missions	0,257	-0,102	-0,189	-0,254	-0,127	-0,293	-0,198	-0,247	-0,127	-0,293	-0,164	-0,223
	(0,374)	(0,354)	(0,519)	(0,506)	(0,490)	(0,507)	(0,594)	(0,560)	(0,491)	(0,508)	(0,620)	(0,582)
Latitude		1,433	1,363	1,317		0,829	0,752	0,517		0,829	0,737	0,508
		(0,812)	(0,817)	(0,786)		(0,813)	(0,864)	(0,849)		(0,813)	(0,873)	(0,855)
Africa			-0,585	-0,487			-0,480	-0,369			-0,487	-0,381
			(0,356)	(0,448)			(0,453)	(0,571)			(0,459)	(0,584)
America			-0,220	-0,275			-0,067	-0,024			-0,054	-0,015
			(0,243)	(0,253)			(0,388)	(0,362)			(0,396)	(0,368)
Asia			-0,215	-0,133			-0,259	-0,179			-0,253	-0,181
			(0,395)	(0,430)			(0,375)	(0,437)			(0,377)	(0,439)
French colony				0,048				0,209				0,212
				(0,291)				(0,350)				(0,352)
British colony				-0,176				-0,092				-0,083
				(0,293)				(0,344)				(0,350)
log capped potential settler mortality					-0,475	-0,450	-0,427	-0,449	-0,475	-0,450	-0,435	-0,454
					(0,181)	(0,189)	(0,209)	(0,199)	(0,181)	(0,189)	(0,217)	(0,204)
log population density in 1500					-0,114	-0,121	-0,107	-0,085	-0,114	-0,121	-0,109	-0,087
					(0,062)	(0,062)	(0,060)	(0,065)	(0,062)	(0,062)	(0,061)	(0,067)
Kleibergen and Paap (2006) test (p-value)	0,00	0,00	0,00	0,00	0,03	0,03	0,02	0,01	0,03	0,03	0,02	0,01
Over-identification test (p-value)	0,99	0,76	0,43	0,44	0,87	0,94	0,48	0,50	0,87	0,94	0,48	0,50

Table 4, Panel B: First-stage regressions
Dependent variable is years of schooling

Primary enrollment in 1900	0,088	0,088	0,051	0,051	0,069	0,072	0,046	0,047	0,069	0,072	0,046	0,047
	(0,016)	(0,016)	(0,017)	(0,018)	(0,016)	(0,017)	(0,018)	(0,021)	(0,016)	(0,017)	(0,018)	(0,021)
Protestant missionaries in early 2Oth century	0,938	0,958	1,173	1,168	0,657	0,577	0,935	0,938	0,657	0,577	0,935	0,938
	(0,423)	(0,425)	(0,318)	(0,362)	(0,444)	(0,462)	(0,406)	(0,431)	(0,444)	(0,462)	(0,406)	(0,431)
Dummy for different source of protestant missions	0,552	0,493	2,394	2,40	-0,708	-0,592	1,859	1,834	-0,708	-0,592	1,859	1,834
	(1,390)	(1,420)	(1,403)	(1,448)	(1,530)	(1,519)	(1,465)	(1,519)	(1,530)	(1,519)	(1,465)	(1,519)
Latitude		0,438	0,238	0,283		-1,564	-0,869	-1,051		-1,564	-0,869	-1,051
		(1,752)	(1,833)	(1,931)		(2,075)	(2,089)	(2,260)		(2,075)	(2,089)	(2,260)
Africa			-1,842	-1,824			-1,182	-1,194			-1,182	-1,194
			(0,612)	(0,650)			(0,963)	(1,175)			(0,963)	(1,175)
America			0,289	0,291			0,618	0,656			0,618	0,656
			(0,533)	(0,547)			(0,774)	(0,803)			(0,774)	(0,803)
Asia			0,701	0,697			0,635	0,640			0,635	0,640
			(0,854)	(0,880)			(0,894)	(1,091)			(0,894)	(1,091)
British colony				0,001				0,028				0,028
				(0,642)				(0,797)				(0,797)
French Colony				-0,063				0,168				0,168
				(0,616)				(0,640)				(0,640)
log capped potential settler mortality					-1,042	-1,104	-0,602	-0,629	-1,042	-1,104	-0,602	-0,629
					(0,359)	(0,403)	(0,461)	(0,502)	(0,359)	(0,403)	(0,461)	(0,502)
log population density in 1500					-0,131	-0,120	-0,067	-0,061	-0,131	-0,120	-0,067	-0,061
					(0,139)	(0,145)	(0,148)	(0,180)	(0,139)	(0,145)	(0,148)	(0,180)
R-squared	0,599	0,599	0,718	0,718	0,677	0,68	0,734	0,734	0,677	0,68	0,734	0,734
F-stat excluded IVs	25,94	25,49	18,84	12,02	12,62	12,42	8,7	5,58	12,62	12,42	8,7	5,58
Observations	62	62	62	62	62	62	62	62	62	62	62	62

Notes: Panel A presents second-stage regressions with years of schooling instrumented using protestant missionaries and primary enrollment in 1900 and Panel B presents the corresponding first-stage regressions, with one observation per country. All variables described in the main text. Standard errors robust against heterocedasticity in parentheses. A.R. confidence intervals correspond to the 95% Anderson-Rubin confidence intervals robust against weak instruments and heterocedasticity. The p-values of the Kleibergen and Paap (2006) test correspond to a test in which the null hypothesis is that the equation is under-identified and, under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions+1). The p-values of the over-identification test correspond to a Hansen over-identification test and under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions).

Table 5: Semi-structural regressions, rule of law, cross-country sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Panel A: Second-stage regressions											
	Dependent variable is log GDP per capita in 2005											
	2SLS								LIML			
Rule of law	1,413	1,634	1,346	1,361	1,705	1,791	1,519	1,424	1,714	1,794	1,592	1,579
	(0,177)	(0,274)	(0,194)	(0,212)	(0,378)	(0,450)	(0,298)	(0,275)	(0,383)	(0,453)	(0,337)	(0,353)
A. R. confidence intervals	*[1.15, 1.92]*	*[1.21, 2.66]*	*[.86, 1.86]*	*[.88, 1.84]*	*[1.06, ∞]*	*[1.06, ∞]*	*[.75, ∞]*	*[.67, ∞]*	*[1.06, ∞]*	*[1.06, ∞]*	*[0.75, ∞]*	*[.67, 2.57]*
Latitude		-1,976	-1,358	-1,269		-1,056	-0,741	-0,597		-1,062	-0,855	-0,835
		(1,332)	(0,987)	(1,014)		(1,313)	(1,080)	(1,113)		(1,317)	(1,130)	(1,230)
Africa			0,202	0,110			-0,434	-0,423			-0,404	-0,356
			(0,446)	(0,447)			(0,375)	(0,371)			(0,384)	(0,388)
America			1,287	1,148			0,795	0,679			0,833	0,757
			(0,406)	(0,362)			(0,383)	(0,388)			(0,398)	(0,421)
Asia			0,280	0,273			-0,476	-0,288			-0,499	-0,333
			(0,297)	(0,305)			(0,398)	(0,391)			(0,414)	(0,412)
French colony				-0,010				-0,013				-0,017
				(0,349)				(0,341)				(0,349)
British colony				-0,236				-0,258				-0,267
				(0,311)				(0,335)				(0,349)
Dummy for different source of protestant missions					-2,791	-2,810	-0,543	-0,409	-2,807	-2,817	-0,621	-0,572
					(1,116)	(1,156)	(0,928)	(0,866)	(1,125)	(1,159)	(0,963)	(0,922)
Primary enrollment in 1900					0,018	0,020	-0,009	-0,005	0,018	0,020	-0,009	-0,006
					(0,009)	(0,009)	(0,010)	(0,010)	(0,009)	(0,009)	(0,011)	(0,011)
Protestant missionaries in early 2Oth century					0,059	-0,001	0,184	0,261	0,058	-0,002	0,170	0,233
					(0,212)	(0,222)	(0,170)	(0,180)	(0,213)	(0,222)	(0,177)	(0,194)
Kleibergen and Paap (2006) test (p-value)	0,0	0,030	0,020	0,020	0,030	0,120	0,060	0,060	0,030	0,120	0,060	0,060
Over-identification test (p-value)	0,230	0,390	0,250	0,140	0,690	0,810	0,340	0,170	0,690	0,810	0,350	0,190

	Table 5, Panel B: First-stage regressions Dependent Variable is rule of Law											
log capped potential settler mortality	-0,597	-0,476	-0,292	-0,231	-0,402	-0,366	-0,235	-0,226	-0,402	-0,366	-0,235	-0,226
	(0,098)	(0,111)	(0,109)	(0,109)	(0,113)	(0,122)	(0,103)	(0,107)	(0,113)	(0,122)	(0,103)	(0,107)
log population density in 1500	-0,081	-0,083	-0,152	-0,160	-0,062	-0,069	-0,111	-0,126	-0,062	-0,069	-0,111	-0,126
	(0,058)	(0,054)	(0,051)	(0,050)	(0,056)	(0,057)	(0,057)	(0,060)	(0,056)	(0,057)	(0,057)	(0,060)
Latitude		1,620	1,831	1,724		0,885	1,252	1,212		0,885	1,252	1,212
		(0,743)	(0,697)	(0,771)		(0,773)	(0,853)	(0,950)		(0,773)	(0,853)	(0,950)
Africa			-0,964	-0,982			-0,306	-0,422			-0,306	-0,422
			(0,378)	(0,377)			(0,387)	(0,394)			(0,387)	(0,394)
America			-1,238	-1,112			-0,621	-0,618			-0,621	-0,618
			(0,342)	(0,344)			(0,348)	(0,364)			(0,348)	(0,364)
Asia			-0,586	-0,629			0,146	-0,008			0,146	-0,008
			(0,384)	(0,389)			(0,392)	(0,407)			(0,392)	(0,407)
British colony				0,350				0,218				0,218
				(0,179)				(0,237)				(0,237)
French colony				0,146				0,126				0,126
				(0,214)				(0,211)				(0,211)
Primary Enrollment in 1900					-0,002	-0,004	0,006	0,003	-0,002	-0,004	0,006	0,003
					(0,007)	(0,007)	(0,008)	(0,009)	(0,007)	(0,007)	(0,008)	(0,009)
Protestant missionaries in early 2Oth century					0,021	0,066	0,049	0,004	0,021	0,066	0,049	0,004
					(0,171)	(0,173)	(0,165)	(0,176)	(0,171)	(0,173)	(0,165)	(0,176)
Dummy for different source of protestant missions					1,410	1,344	0,661	0,593	1,410	1,344	0,661	0,593
					(0,483)	(0,506)	(0,70)	(0,693)	(0,483)	(0,506)	(0,70)	(0,693)
R-squared	0,508	0,551	0,638	0,656	0,603	0,612	0,664	0,669	0,603	0,612	0,664	0,669
F-stat excluded instruments	31,820	13,480	9,550	8,320	6,270	4,580	5,070	5,450	6,270	4,580	5,070	5,450
Observations	62	62	62	62	62	62	62	62	62	62	62	62

Notes: Panel A presents second-stage regressions with rule of law instrumented using log capped potential settler mortality and log population density in 1500 and Panel B presents the corresponding first-stage regressions, with one observation per country. All variables described in the main text. Standard errors robust against heterocedasticity in parentheses. A.R. confidence intervals correspond to the 95% Anderson-Rubin confidence intervals robust against weak instruments and heterocedasticity. The p-values of the Kleibergen and Paap (2006) test correspond to a test in which the null hypothesis is that the equation is under-identified and, under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions+1). The p-values of the over-identification test correspond to a Hansen over-identification test and under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions).

Table 6: Full 2SLS and LIML estimates, cross-country sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Second-stage regressions							
	Dependent variable is log GDP per capita in 2005							
	2SLS				LIML			
Years of schooling	0,223	0,224	0,069	0,186	0,217	0,218	-0,019	0,094
	(0,073)	(0,074)	(0,129)	(0,142)	(0,077)	(0,078)	(0,194)	(0,244)
Rule of law	1,126	1,123	1,324	1,062	1,168	1,170	1,701	1,464
	(0,355)	(0,378)	(0,390)	(0,374)	(0,387)	(0,415)	(0,674)	(0,730)
Dummy for different source of protestant missions	-1,666	-1,655	-1,238	-1,193	-1,720	-1,702	-1,367	-1,355
	(0,587)	(0,560)	(0,638)	(0,527)	(0,629)	(0,599)	(0,810)	(0,674)
Latitude		-0,040	-0,647	-0,276		-0,096	-1,230	-0,888
		(0,950)	(1,078)	(1,021)		(0,998)	(1,463)	(1,518)
Africa			-0,421	-0,265			-0,423	-0,262
			(0,340)	(0,359)			(0,390)	(0,395)
America			0,483	0,237			0,729	0,494
			(0,367)	(0,397)			(0,522)	(0,624)
Asia			-0,511	-0,329			-0,568	-0,398
			(0,342)	(0,350)			(0,410)	(0,386)
British colony				-0,291				-0,284
				(0,301)				(0,344)
French colony				0,016				0,003
				(0,296)				(0,334)
Kleibergen and Paap (2006) test (p-value)	0,10	0,260	0,110	0,070	0,10	0,260	0,110	0,070
Over-identification test (p-value)	0,620	0,60	0,20	0,120	0,630	0,620	0,340	0,220

Table 6, Panel B: First stage regressions

Dependent variable is	Years of schooling				Rule of law			
Primary enrollment in 1870	0,069	0,072	0,046	0,047	-0,002	-0,004	0,006	0,003
	(0,016)	(0,017)	(0,018)	(0,021)	(0,007)	(0,007)	(0,008)	(0,009)
Protestant missionaries in early 2Oth century	0,657	0,577	0,935	0,938	0,021	0,066	0,049	0,004
	(0,444)	(0,462)	(0,406)	(0,431)	(0,171)	(0,173)	(0,165)	(0,176)
log capped potential settler mortality	-1,042	-1,104	-0,602	-0,629	-0,402	-0,366	-0,235	-0,226
	(0,359)	(0,403)	(0,461)	(0,502)	(0,113)	(0,122)	(0,103)	(0,107)
log population density in 1500	-0,131	-0,120	-0,067	-0,061	-0,062	-0,069	-0,111	-0,126
	(0,139)	(0,145)	(0,148)	(0,180)	(0,056)	(0,057)	(0,057)	(0,060)
Dummy for different source of protestant missions	-0,708	-0,592	1,859	1,834	1,410	1,344	0,661	0,593
	(1,530)	(1,519)	(1,465)	(1,519)				
Latitude		-1,564	-0,869	-1,051		0,885	1,252	1,212
		(2,075)	(2,089)	(2,260)				
America			0,618	0,656			-0,621	-0,618
			(0,774)	(0,803)			(0,348)	(0,364)
Asia			0,635	0,640			0,146	-0,008
			(0,894)	(1,091)			(0,392)	(0,407)
Africa			-1,182	-1,194			-0,306	-0,422
			(0,963)	(1,175)			(0,387)	(0,394)
French colony				0,168				0,126
				(0,640)				(0,211)
British colony				0,028				0,218
				(0,797)				(0,237)
R-squared	0,677	0,68	0,734	0,734	0,603	0,612	0,664	0,669
Observations	62	62	62	62	62	62	62	62
	F statistic for excluded instruments in relevant equation							
Institutions	6.44	6.37	1.22	1.25	6.27	4.58	5.07	5.45
Schooling	12.62	12.42	8.70	5.58	0.05	0.16	0.79	0.10

Notes: Panel A presents second stage regressions with rule of law and years of education instrumented using the log capped potential settler mortality, log population density in 1500, protestant missionaries and primary enrollment in 1900 and Panel B presents the corresponding first stage regressions, with one observation per country. All variables described in the main text. Standard errors robust to heterocedasticity in parentheses. The p-values of the Kleibergen and Paap (2006) test correspond to a test in which the null hypothesis is that the equation is under-identified and, under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions+1). The p-values of the over-identification test correspond to a Hansen overidentification test and under the null the statistic is distributed as chi-squared with degrees of freedom=(number of overidentifying restrictions).

Table 7: Robustness exercises. Full LIML models, second-stage regressions, cross-country sample

Sample	Excluding Neo-Europes		All					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years of schooling	0,034	0,174	-0,043	0,041	0,009	0,120	0,119	0,197
	(0,199)	(0,196)	(0,191)	(0,198)	(0,136)	(0,196)	(0,186)	(0,187)
Rule of law	1,432	1,149	1,382	1,218	1,193	1,181	1,487	1,336
	(0,679)	(0,595)	(0,591)	(0,501)	(0,386)	(0,389)	(0,687)	(0,662)
Dummy for different source of protestant missions	-0,577	-0,638	-0,314	-0,362	-0,573	-0,979	-1,055	-1,041
	(0,968)	(0,850)	(1,106)	(1,015)	(0,823)	(0,956)	(0,908)	(0,847)
Latitude	-0,711	-0,310	-2,061	-1,912	-1,942	-2,044	-2,045	-1,846
	(1,465)	(1,284)	(1,478)	(1,434)	(1,288)	(1,316)	(1,793)	(1,70)
Africa	0,084	0,099	0,294	0,345	0,124	0,148	0,743	0,570
	(0,424)	(0,374)	(0,865)	(0,785)	(0,699)	(0,701)	(1,313)	(1,237)
America	1,086	0,622	0,895	0,786	1,130	0,838	0,754	0,569
	(0,639)	(0,630)	(0,871)	(0,798)	(0,756)	(0,836)	(0,90)	(0,858)
Asia			-0,068	0,031	0,220	-0,013	0,428	0,305
			(0,863)	(0,783)	(0,766)	(0,804)	(1,326)	(1,249)
British colony		-0,302		-0,098		-0,284		-0,318
		(0,30)		(0,324)		(0,391)		(0,332)
French colony		0,006		0,210		0,109		0,006
		(0,284)		(0,297)		(0,272)		(0,301)
Falciparum malaria index 1994			-1,042	-0,998				
			(0,542)	(0,556)				
Wald test for humidity (p-value)					0,08	0,66		
Wald test for temperature (p-value)					0,59	0,08		
Wald test for religion affiliation in 1900 (p-value)							0,33	0,31
Kleibergen and Paap (2006) test (p-value)	0,09	0,08	0,08	0,11	0,01	0,05	0,14	0,13
Over-identification test (p-value)	0,33	0,24	0,52	0,57	0,85	0,81	0,66	0,47
Observations	58	58	61	61	62	62	62	62

Notes: Second-stage regressions with rule of law and years of schooling instrumented using the log capped potential settler mortality, log population density in 1500, protestant missionaries and primary enrollment in 1900, with one observation per country. Columns 1 and 2 present regressions excluding the Neo-Europes (Australia, Canada, New Zealand, and the US). All the remaining columns present estimates for the complete sample. In columns 4 and 5 we add the following controls for humidity and temperature: average, minimum, and maximum monthly high temperatures, and minimum and maximum monthly low temperatures, and morning minimum and maximum humidity, and afternoon minimum and maximum humidity (from Parker, 1997). In columns 7 and 8 we add controls for the share of the catholic, Muslim, and protestant population in 1900 from the World Christian Encyclopedia. All variables described in the main text. Standard errors robust against heterocedasticity in parentheses. The p-values of the Kleibergen and Paap (2006) test correspond to a test in which the null hypothesis is that the equation is under-identified and, under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions+1). The p-values of the over-identification test correspond to a Hansen over-identification test and under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions).

Table 8: Effects of years of schooling on institutions, second-stage regression, cross-country sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Dependent Variable is rule of law											
	2SLS								LIML			
Years of schooling	0,081	0,073	0,173	0,163	-0,021	-0,028	0,086	0,039	-0,022	-0,031	0,086	0,038
	(0,034)	(0,034)	(0,062)	(0,077)	(0,072)	(0,072)	(0,062)	(0,087)	(0,072)	(0,073)	(0,063)	(0,088)
A. R. confidence intervals	[-.01, .16]	[-.02, .15]	[-.05, .32]	[-.02, .34]	[-.27, .13]	[-.28, .12]	[-.09, .25]	[-∞ , .23]	[-.27, .13]	[-.27, .12]	[-.09, .25]	[-∞ , .23]
Dummy for different source protestant missions	1,706	1,376	0,691	0,703	1,317	1,177	0,616	0,640	1,318	1,182	0,617	0,642
	(0,240)	(0,309)	(0,396)	(0,407)	(0,295)	(0,346)	(0,451)	(0,459)	(0,296)	(0,348)	(0,452)	(0,460)
Latitude		1,309	1,525	1,508		0,724	1,369	1,301		0,723	1,369	1,30
		(0,796)	(0,715)	(0,754)		(0,828)	(0,755)	(0,841)		(0,830)	(0,755)	(0,841)
Africa			-0,095	-0,133			-0,204	-0,375			-0,204	-0,376
			(0,234)	(0,244)			(0,328)	(0,370)			(0,328)	(0,371)
America			-0,558	-0,541			-0,649	-0,619			-0,649	-0,618
			(0,128)	(0,144)			(0,281)	(0,306)			(0,281)	(0,306)
Asia			0,208	0,179			0,099	-0,026			0,099	-0,026
			(0,285)	(0,274)			(0,314)	(0,322)			(0,314)	(0,322)
French colony				0,030				0,116				0,116
				(0,172)				(0,187)				(0,187)
British colony				0,061				0,217				0,218
				(0,167)				(0,201)				(0,201)
log capped potential settler mortality					-0,425	-0,406	-0,180	-0,197	-0,426	-0,409	-0,180	-0,198
					(0,163)	(0,179)	(0,093)	(0,103)	(0,163)	(0,180)	(0,093)	(0,103)
log population density in 1500					-0,065	-0,072	-0,105	-0,124	-0,065	-0,073	-0,105	-0,124
					(0,061)	(0,061)	(0,049)	(0,052)	(0,061)	(0,062)	(0,049)	(0,052)
Kleibergen and Paap (2006) test (p-value)	0,0	0,0	0,0	0,0	0,030	0,030	0,020	0,010	0,030	0,030	0,020	0,010
Over-identification test (p-value)	0,810	0,570	0,930	0,930	0,840	0,630	0,810	0,80	0,840	0,630	0,80	0,80
Observations	62	62	62	62	62	62	62	62	62	62	62	62

Notes: Second stage regressions with years of schooling instrumented using protestant missionaries and primary enrollment in 1900. All variables described in the main text. Standard errors robust against heterocedasticity in parentheses. A.R. confidence intervals correspond to the 95% Anderson-Rubin confidence intervals robust to weak instruments and heterocedasticity. The p-values of the Kleibergen and Paap (2006) test correspond to a test in which the null hypothesis is that the equation is under-identified and, under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions+1). The p-values of the over-identification test correspond to a Hansen over-identification test and under the null the statistic is distributed as chi-squared with degrees of freedom=(number of over-identifying restrictions).

3 Exactly Identified Models

Table Appendix 1: IV estimates, cross-country sample

	Panel A: Second-stage regressions			
	Dependent variable is log GDP per capita in 2005			
Years of schooling	0.176	0.167	0.036	0.124
	(0.125)	(0.158)	(0.233)	(0.246)
Rule of law	1,252	1.303	1,759	1,705
	(0.612)	(0.811)	(0.980)	(0.952)
Dummy for different source of protestant missions	-1,657	-1,653	-1,781	-1,841
	(0.752)	(0.767)	(0.864)	(0.785)
Latitude		-0.208	-1.295	-1.239
		(1.287)	(1.907)	(2.017)
Africa			-0.310	-0.098
			(0.407)	(0.422)
America			0.662	0.521
			(0.672)	(0.713)
Asia			-0.641	-0.449
			(0.433)	(0.433)
British colony				-0.382
				(0.372)
French colony				-0.004
				(0.341)
Kleibergen and Paap (2006) test (p-value)	0,110	0,180	0,110	0,10
Observations	62	62	62	62

Table Appendix 1, Panel B: First-stage regressions								
Dependent variable is	Years of schooling				Rule of law			
Protestant missionaries in early 2Oth century	1,277	1,277	1,287	1,211	0.053	0.086	0.150	0.133
	(0.534)	(0.577)	(0.357)	(0.363)	(0.156)	(0.154)	(0.141)	(0.164)
log capped potential settler mortality	-1,329	-1,329	-0,712	-0.698	-0,415	-0,382	-0,296	-0,306
	(0.379)	(0.415)	(0.430)	(0.431)	(0.117)	(0.127)	(0.111)	(0.115)
Dummy for different source of protestant missions	3,958	3,959	5,201	4,964	1,361	1,223	1,403	1,349
	(1.105)	(1.101)	(0.929)	(0.972)	(0.305)	(0.331)	(0.366)	(0.392)
Latitude		-0.008	-0.566	-0.670		0.767	1.059	0.951
		(2.484)	(1.977)	(2.036)		(0.776)	(0.824)	(0.899)
America			-0.772	-0.901			-0.008	-0.051
			(0.986)	(1.037)			(0.266)	(0.260)
Asia			1,794	1,807			-0.121	-0.099
			(0.719)	(0.723)			(0.158)	(0.173)
Africa			0.991	0.791			0.390	0.338
			(0.990)	(0.973)			(0.297)	(0.300)
French colony				0.319				0.095
				(0.609)				(0.201)
British colony				0.151				0.122
				(0.614)				(0.197)
R-squared	0,57	0,57	0,709	0,71	0,593	0,601	0,63	0,632
Observations	62	62	62	62	62	62	62	62
	F statistic for excluded IVs for relevant equation							
Institutions	5,340	4,510	11,310	9,330	0,110	0,290	0,980	0,550
Schooling	11,490	9,40	2,380	2,20	11,860	8,320	6,170	5,880

Notes: Panel A presents just identified models with rule of law instrumented using the ln of potential settler mortality and years of schooling instrumented using Protestant missionaries and Panel D presents the corresponding first stage regressions. All variables described in the main text. Standard errors robust to heterocedasticity in parentheses. Coefficients and standard errors for the constants are not reported to save space. The p-values of the Kleibergen and Paap (2006) test correspond to a test in which the null hypothesis is that the equation is underidentified and, under the null the statistic is distributed as chi-squared with degrees of freedom=(number of overidentifying restrictions+1).