CESifo Working Papers

www.cesifo.org/wp

Geography, Institutions, and the Making of Comparative Development

Raphael Auer

CESifo Working Paper No. 3874

Category 6: Fiscal Policy, Macroeconomics and Growth

July 2012

CESifo *Working Paper No. 3874*

Geography, Institutions, and the Making of Comparative Development

Abstract

While the direct impact of geographic endowments on prosperity is present in all countries, in former colonies, geography has also affected colonization policies and, therefore, institutional outcomes. Using non-colonized countries as a control group, I re-examine the theories put forward by La Porta et al. (1999) and by Acemoglu et al. (2001), finding strong support for both theories, but also evidence that the authors' estimates are mildly biased since they confound the effect of the historical determinants of institutions with the sizeable direct impact of geographic endowments on development.

JEL-Code: O110, P160, P510, R110, N500, F540.

Keywords: growth, institutions, geography, comparative development, colonialism.

Raphael Auer
Swiss National Bank
raphael.auer@snb.ch

26.09.2011
This is a substantially revised chapter of my 2006 PhD thesis. I thank Daron Acemoglu and Xavier Gabaix for discussions, guidance and support; Josh Angrist, Martin Brown, Sylvain Chassang, Emmanuel Farhi, Nicola Gennaioli, Simon Johnson, Mark Melitz, Marcel Peter, Philip Saure, Andrei Shleifer, David Weil, an anonymous referee at the Swiss National Bank's working paper series, as well as seminar participants as ETH Zurich and the annual conferences of the American Economic Association, the European Economic Association, the European Development Network, the Econometric Society, and especially Gerard Padró i Miquel for helpful comments and discussions; and Domagoj Arapovic for excellent research assistance.

Is the substantial inequality in the wealth of nations a result of man-made history or the inevitable consequence of nature?

Two rivaling schools of thought emphasize geographic endowments or institutions, respectively, as the main determinant of comparative development. The "endowments" view, developed by Diamond (1997), Bloom and Sachs (1998), Gallup et al. (1998), and Frankel and Romer (1999), among others, argues that climate, topography, location, and other geographic features directly affect the prevalence of disease, the productivity of labor, and prosperity.

In contrast, the "institutions" view, pioneered in its modern form by North (1981), argues that the organization of society is the basic force of comparative development. This hypothesis has received strong support from the empirical work of Mauro (1995), La Porta et al. (1997, 1998, and 1999), Hall and Jones (1999), Acemoglu et al. (2001 and 2002), and Feyrer and Sacerdote (2008). These authors argue that the endogenous quality of institutions is induced by the course of history.

Although it is fair to say that the "institutions" view is currently the dominant view of development,[1] it is not free from criticism. A major concern is that the instrumental variables used to establish the effect of institutions have been affected by endowments and early economic development and that the instrumentation strategies are, therefore, invalid. For example, the correlation between disease environment and income per capita can be attributed to either the indirect effect of settler mortality rates on colonization policies, in accordance with the theory of Acemoglu et al. (2001), or to the direct impact of disease on income (see, for example, Mc Arthur and Sachs (2000)). Similarly, the finding that legal systems based on British common law are generally associated with higher income compared with systems based on civil law could reflect the causal impact of the legal system on economic performance, but it could also reflect the fact that the British tended to colonize countries with more favorable endowments.[2]

This study contributes to the understanding of the partial effects of institutions and of en-

[1] A frequent finding in the literature is that, once the quality of institutions is accounted for, endowments matter only marginally for development. See also Easterly and Levine (2003) and Rodrik et al. (2004).

[2] As I argue below, adding controls for the potential direct effect of endowments to the empirical estimations does not alleviate this concern because these controls potentially also affect development both directly and indirectly via colonization policies. What is missing in the current literature is a clear control group that distinguishes the direct effect of endowments from the impact of colonization policies. I set out to build such a control group below.

dowments on comparative development. The key insight is that one can utilize the interaction of colonial history and geography to identify the partial effects of institutions and endowments. In countries that have been colonized, geographic location has affected colonization policies.

The indirect effects of endowments on colonization policies are only present in former colonies. In contrast, the direct impact of endowments on development is also present in countries that have not been colonized (noncolonies). Because endowments had both a direct and an indirect effect on colonization policies in former colonies but shaped development only directly in noncolonies, the difference in how geography has affected economic outcomes in these two groups can identify the determinants of development. In this respect, the current study draws on the hypothesis of Acemoglu et al. (2002), who argue that the effect of geography on economic development was reversed during colonization: endowments that were favorable for development early on led later to unfavorable colonization policies.[3]

Using this basic insight, I first highlight the role of disease on development throughout history. To that end, I construct a measure of the geographic potential for disease, i.e., the level of disease that would prevail if a country was untouched by Western civilization. For a former colony, a 1% higher level of potential for disease is associated with a roughly 1.2% lower level of income per capita. In a baseline estimation of this paper, about three-quarters of the total effect of disease is attributed to the institution-building channel, i.e., to the impact of settler mortality rates on colonization policies and institutions, hence confirming the theory of Acemoglu et al. (2001). The remaining one-quarter of the effect is attributed to the direct impact of disease on development.

I next examine the theory put forward by La Porta et al. (1999), i.e., the importance of legal origin for institutional outcomes, considering that location and transportation costs could have mattered for colonizer identity and, consequently, for legal origin. Indeed, I document that the location of former colonies can predict legal origin very well. Controlling for this relation, I find that the causal effect of adopting a common law system is even more substantial than the estimations of La Porta et al. suggest. The reason for this is the following. Countries with a location that made them likely to be colonized by Britain are, on average, remote from export

[3] This basic insight is also related to the work of Nunn and Puga (2010), who demonstrate that the slave trade reversed the impact of internal transportation cost through the protection from slave traders provided by rugged terrain.

markets, which is detrimental to growth. Consequently, the positive effect of adopting a British legal system is partly obscured by the negative effect of remoteness.

Overall, this leads me to conclude that both the direct and the indirect institution-building effects during colonization matter for development. This finding reconciles the contrasting findings of the two rivaling views in the literature. In the studies arguing for the importance of institutions, in identifying the relation between institutions and income, these studies attribute the correlation between endowments and income entirely to the impact of institutions. Similarly, the literature arguing for the importance of geography attributes this correlation entirely to the direct impact of endowments. Because both channels matter, existing studies overestimate their starting hypothesis.

The structure of this paper is as follows. Section 1 lays down the criteria used to classify countries into "former colonies" and "noncolonies". Section 2 documents that, for the group of former colonies, geographic endowments had a marked effect on colonization policies. Section 3 discusses the econometric framework and discusses the assumptions under which the comparison of the correlations between endowments and outcomes in colonies and noncolonies is a meaningful way to identify the determinants of development. Section 4 presents the results and Section 5 the robustness analysis. Section 6 concludes.

1 Defining Colonies and Noncolonies

The sample criterion for this study is that information on the country's level of economic development (GDP per capita in 2003 from the World Bank development indicators) and its quality of institutions (the score for the "Rule of the Law" from Kaufmann et al. (2005)) is available. Countries with less than half a million inhabitants in 2004 are excluded from the analysis, yielding a sample size of 151 countries.

To discern the direct effect of endowments on growth from the institutional channel, it is expedient to construct a valid control group in which only the direct effect of endowments is present. This raises the question of how to find a consistent classification scheme to allocate countries into either "colonies" or "noncolonies".

Colonization is clearly defined in biology, where it describes the process by which a species spreads into new areas to form a new settlement separated by open ground from the original colony (see, for example, Encyclopedia Britannica (2010)). In the social sciences, the definition is less concise; colonialism "normally refers to a period of history from the 15th to the late 19th centuries when nations of Europe established colonies on other continents". Colonization refers to the building and maintaining of colonies in one territory by people from another territory and is thus a process "whereby sovereignty over the colony is claimed by the metropolis and social structure, government and economics within the territory of the colony are changed by the colonists."[4]

Following these generally accepted concepts, the history of a country that is classified as a "former colony" should thus satisfy two main criteria. The first criterion is that the colony must have been characterized by a substantial loss of sovereignty during the $15th$ to the late $19th$ century, where I define a "substantial loss of sovereignty" to be the case if the country was an official colony, was under the control of an empire-affiliated organization, such as the Dutch and British East Indies Companies, had the status of protectorate, or lost sovereignty over its foreign policy following a military conflict with an imperialist empire. The second criterion is that the colony and colonizer are physically separate (i.e., colonies and their colonizer empire do not form a contingent empire), which is makes it possible to unambiguously distinguish colony and colonizer (otherwise, the question arises of whether Hungary was a colony or a reigning part of the Austro-Hungarian empire) and constitutes a prerequisite to set up completely independent institutional arrangements in each colony and its colonizing nation.

With this definition of former colonies, 56 countries are classified as noncolonized nations, and 95 are classified as former colonies. Table A1 lists all 151 countries included in this study, specifies whether each is classified as a former colony or a noncolony, and provides further information on endowments and geography. Arapovic and Auer (2011) extensively discuss the detailed historical information that has been used to classify countries into these two groups.[5]

[4] The above quotes are taken from http://en.wikipedia.org/wiki/Colonialism. The site was accessed on August 15 2010.

[5] Table 9 in the Appendix of the working paper version of this paper presents various alternative classifications of former colonies, demonstrating that none of the results of this paper is dependent on the precise way of defining colonies assumed below.

2 Geography and Colonization Policies

With a clear definition of control and treatment group at hand, one needs to construct a measure for early disease environment and for the likelihood of adopting a certain legal origin for the entire sample. This section thus quantifies the extent to which variation in colonization policies can be attributed to variation in geographic endowments. The estimated models are then used to generate two relevant measures summarizing all geographic characteristics that can explain the mortality rates of settlers during colonization and the extent to which the spread of the British and French colonial empires was shaped by the colony's location.

2.1 Climate Endowments and Settler Mortality Rates

To gauge whether the correlation between disease and income can be attributed to the direct importance of germs for prosperity or to the indirect effect of settler mortality on institutional development during colonization, one must first understand the extent to which historical mortality rates were driven by geographic endowments that might affect development today.

In Columns (1) to (4) of Table 1, the dependent variable is the natural logarithm of the settler mortality rate collected by Acemoglu et al. (2001).[6] In Column 1 of Table 1, Malaria Ecology—a variable measuring the geographic potential for malaria, taken from Kiszewski et al. (2004)—is the only regressor. For better interpretability, all variables in Table 1 are standardized. The correlation between Malaria Ecology and Settler Mortality is highly significant. More important, the R-square of this single equation is 0.47—in other words, nearly half of the variation in mortality rates can be explained by variation in the geographic potential for malaria, thus reinforcing doubts that settler "mortality [merely] provides an exogenous indicator of 'germs'" (Easterly and Levine (2003), p. 12).

Although Malaria Ecology is arguably one of the most direct gauges of the disease environment during colonization, this variable displays extremely limited variation in the group of noncolonies (see Table 2). Malaria Ecology is thus not well suited for the analysis that follows because the

[6] In Table 5, Malta and the Bahamas are missing from the sample of Acemoglu et al. (2001) because the respective populations of these two countries are smaller than 500,000 (see sample criterion). None of the results of this paper is dependent on the exclusion of these two countries.

latter relies on within-group variation of endowments and economic outcomes. Columns (2) to (4) thus use other information on geography that is available for all countries. In Column (2), distance from the equator—which is often used as a proxy for endowments because it correlates with average temperature, rainfall, drought, and humidity, all of which in turn affect the prevalence of disease—is included in the regression, again resulting in a highly statistically significant relation and a high R-square.

The estimation in Column (3) refines this approach and includes direct measures of average temperature, rainfall, drought, and humidity as independent variables. Average annual temperature, minimum and maximum monthly rainfall, and temperature at maximum humidity are taken from Parker (1997). Warmer climate and pronounced dry (low minimum monthly rain) or wet seasons (high maximum monthly rain) are characterized by high mortality rates. Three of the regressors are significant on their own (although this finding is not particularly significant since the regressors are highly collinear). The bottom of Table 1 reports a p-value corresponding to the joint null hypothesis that the included geographic variables together did not affect settler mortality rates. This hypothesis is rejected at the 0.1% significance level.

Column (4) adds four dummies that, respectively, equal one if a country is characterized by natural incidence of savanna, natural incidence of temperate grassland or forest, or Mediterranean climate, or if it has mountains. All dummies are taken from Parker (1997).[7]

Columns 1 to 4 of Table 1 document the (somewhat unsurprising) fact that settler mortality rates are the result of a country's endowments. Thus, it is straightforward to enlarge the sample collected by Acemoglu et al. (2001) by following the strategy of Kiszewski et al. (2004): the natural prevalence of germs is determined by a country's climate and landscape. One can estimate this relation between climate and disease by using the mortality rates collected from historical sources and a set of geographic variables. The estimated relation between germs and geography can then be extrapolated to construct a measure of early disease environment using widely available geographic information.

[7]Column 4 also controls for the sampling population and adds three dummies that, respectively, equal one if the mortality rate was sampled from soldiers in campaign, from bishops, or from forced laborers. Indeed, the sampling population has a sizeable influence on mortality. For example, compared with soldiers stationed in barracks, soldiers in a campaign are twice as likely to die from disease.

Using the estimated coefficient relating geography and settler mortality in Column 4 of Table 1, I next predict several measures of the geographic potential for disease in 151 countries. In the analysis below, I refer to this measure as "Early Disease Environment," or EDE.

Paralleling the definition of "settler mortality" in Acemoglu et al. (2001), EDE refers to the logarithm of the annualized probability of death for European males in the age cohort of soldiers. It is important to note that the use of EDE—measuring the hypothetical mortality rate rather than the actual one—is in accordance with the institution-building hypothesis of Acemoglu et al. (2001), who provide evidence that knowledge about the widespread prevalence of disease alone was sufficient to deter migration to a colony.

2.2 Location and Legal Origin

Did endowments also affect the legal origins of former colonies? The two world maps in Figure 1 document the maximum spread of the British (upper map) and French (lower map) empires.[8]

Geography's role in shaping the colonial empire of the European imperialists is especially evident in Africa. The French laid the foundation for their African possessions in 1830 with the invasion of Algeria, one of the closest landing points on the African continent from the Mediterranean coast of France. From this base, starting with the establishment of a protectorate in Tunisia in 1881, French influence gradually expanded into neighboring North Africa and later into West Africa. France's African empire eventually covered the territory of the modern nations of Benin, Chad, Central African Republic, Côte d'Ivoire, Guinea, Mali, Mauritania, Niger, Senegal, and the Republic of Congo.

In contrast, British colonization of the African continent initially served to ensure supremacy over strategic shipping routes, foremost the India trade. British colonization in Africa started with the occupation of the Cape Colony in 1795 (before the Suez Canal had been built), which served as a supply point to restock food and water. The occupation of Egypt in 1882 was influenced by the country's geographic location as the invasion served mostly to secure control of the Suez Canal, again with the objective to guarantee access to Britain's possessions in the Indian and

[8] Figure 1 is adapted from Wikipedia Commons (http://en.wikipedia.org/wiki/Colonial_empire, accessed on 28.08.2011)

Pacific Oceans.

In the British case, the early colonies of Cape Colony and Egypt served as bases to gradually expand into adjacent territories that led to the formation of British East Africa (spanning the territory of today's Kenya, Uganda, and Tanzania) and British South Africa (spanning South Africa, Zambia, Zimbabwe, Malawi, Lesotho, Botswana, and Swaziland).

Because these two empires started their expansions into Africa at very different locations and then expanded their empires to adjacent territories, French colonies in Africa were on average geographically much closer to Europe than Britain's possessions were on this continent. Moreover, the relative distance from the colonies to France versus to Britain affected colonial origin: Algiers is 412 nautical miles away from Marseille, as opposed to 1734 from London. Given that most French African colonies are, in turn, located close to Algeria, French African colonies are much closer to France than they are to the UK.

Geography also markedly shaped the identity of the colonizers on other continents. For example, although the Australian coast had previously been visited by Dutch, Portuguese, and French expeditions, the country was colonized only in 1788, when British naval dominance made a considerable colonization effort worthwhile: at the time, only the British fleet was able to send out and support a sizeable expedition to this remote location. This British naval dominance is easily documented by the number of ships in commission. For example, in 1775, the Royal Navy had 117 ships of the line and 82 cruisers; France had 59 and 37, respectively; and Spain 64 and 28, respectively. By 1809, Britain possessed more than half of the world's warships (see Black and Woodfine (1988)).[9]

To formalize these regularities, Columns (5) to (10) in Table 1 estimate a model of the geographic determinants of colonial and legal origin. For these specifications, I have collected sailing distances from Netpas.net, which is a software used to calculate shipping routes. For each country in the sample, I collect the distance to the country's capital via its main harbor from London, Le Havre, Marseille, Cadiz, or Lisbon. The distances are defined as the closest possible sailing

[9] Of all nations, why did Britain emerge as the dominant nation on the seas? Acemoglu et al. (2005) argue that geography itself—combined with initial institutions—explains the rise of Britain. Britain's geographic location predisposes it to becoming a naval power. Its coastline measures 12,429 km (8,982 kilometers when counting only mainland England), as opposed to France's coast, measuring 3,427 km in length (all coast lengths are measured on a 20-km grid).

route that does not use either the Panama or Suez Canals (i.e., to go from Europe to Asia would require either shipping around the Cape of Good Hope or westward around South America via Ushuaia).[10]

Using these sailing distances, I then construct measures of average distance from the colonizers, defined as the logarithm of the average nautical distance from the colonizer's main port. For France, the smaller of the distances from Le Havre and Marseille is used. I also define measures of relative sailing distances. For example the relative distance to France compared with Britain is calculated as Ln(sailing distance to France)-Ln(sailing distance to Britain).

The specifications in Columns 5 to 10 of Table 1 relate the relative distances from the colonizers and the absolute distance from Europe to the probability of being colonized by or adopting the legal system of a given colonizer. In all specifications, the probability of adopting a particular legal system, conditional on having been colonized, is estimated. This conditionality is appropriate because, in the analysis below, I want to establish the effect of adopting a particular legal system conditional on the fact that the legal system has been imposed by a foreign power.

In Columns 5 and 6 of Table 1, the dependent variable is a dummy that equals one for a former British colony, and the sample includes all former French and British colonies (58 countries). The probit estimation in Column 5 relates the British colony dummy to the logarithm of the average (ln) distances from France and Britain. Column 6, in turn, relates the same dummy to the logarithm of the country's relative distance from France, defined as Ln(distance from France) – Ln(distance from Britain). Compared with French colonies, former British colonies tend to be located closer to Britain. The order of magnitude of the coefficients suggests that location was a major determinant of colonizer identity. For example, in Column 5, a one-percent increase in the absolute distance from Europe is associated with a 0.886-percent increase in the predicted z-score for the country's probability of becoming a British colony.[11]

Not all countries have adopted the legal system of their colonizer. For example, Egypt was a British protectorate, but its legal system is nevertheless based on the Napoleonic Code. To

[10] If a colony is landlocked, the distance to the nearest major port is used, and the shortest distance from the port to the colony via a major river, or, if this is not available, via roads is added to the pure sailing distance.

[11] Further examination, undertaken in the working paper version of this paper, demonstrates that the latter nations are also more distant from Europe in terms of geographic (instead of sailing) distance, geographically less open to trade, and located further away from the equator.

demonstrate that location can explain the legal as well as the colonial origin, Columns 7 and 8 next relate the legal origin of former colonies to geography. In these two columns, the dependent variable is a dummy equal to one if a country has a legal system of British origin. The sample is restricted to all former colonies with either British or French legal origins. As is to be expected from the previous analysis, countries that adopted the British legal system are relatively closer to Britain than to France and are relatively more distant from Europe. The estimations of Columns 9 and 10 in Table 1 repeat the legal origin specifications of Columns 7 and 8 but include a larger sample consisting of all former colonies. The dependent variable is again a dummy equal to one if a country has a legal system of British origin.[12]

In parallel with the construction of EDE, I also construct estimates of the geographic proximity to the colonies. These predicted variables measure the estimated probability that a country—had it been colonized—would have adopted a British, French, or other legal system. First, the probability that a colony adopts a particular legal system given the country's geographic location is estimated using the model of Column 10 in Table 1. Using this model, I then predict measures of "relative proximity" to Britain, France, and other nations for 149 countries (Gaza and the West Bank are not included in the Netpass data and, since the distance from London to London is 0, the proximity variable cannot be constructed for the UK).

3 Are Colonies and Noncolonies "Comparable"? - An Econometric Framework

This section discusses the econometric framework emphasizing whether the comparison of the correlation between endowments and outcomes in colonies and noncolonies is a meaningful exercise to identify the determinants of development. With this emphasis in mind, this section proceeds in two steps. First, it describes the econometric framework that are estimated below in Section 4

[12] In the working paper version of this paper, I also estimate richer geographic models and multinomial Probit estimation allowing for the colonial or legal origin to be French, British, Spanish, or "other", where the "other" group includes countries with German, Scandinavian, or communist legal origins. Moreover, in these specifications, absolute and relative sailing distances are shown to have substantial explanatory power for the legal origin.

. Second, it discusses which deviations from the underlying assumptions of the framework would invalidate the identification methodology.

Throughout the analysis, let Y_i denote the logarithm of GDP per capita and R_i the measure of institutional quality in country i. Denote geographic endowments by E_i and the measure summarizing European colonization policies by P_i. Last, the dummy C_i equals 1 for former colonies and 0 otherwise. Abstracting from covariates, the joint model of colonization, institutions, and income is given by:

$$Y_i = \widetilde{\lambda}_Y + \widetilde{\lambda}'_Y C_i + \widetilde{\alpha} R_i + \widetilde{\eta}_Y E_i + \widetilde{\nu}_{Y,i} \quad (1)$$

$$R_i = \widetilde{\lambda}_R + \widetilde{\lambda}'_R C_i + \widetilde{\eta}_R E_i + \widetilde{\beta} Y_i + C_i \widetilde{\theta}_R P_i + \widetilde{\nu}_{R,i} \quad (2)$$

$$P_i = \widetilde{\lambda}_P + \widetilde{\theta}_P E_i + \widetilde{\nu}_{P,i}, \quad (3)$$

where (3) applies only to former colonies.

A country's institutions and income level depend on endowments through three potential channels. First, endowments may directly affect technology and income, measured by $\widetilde{\eta}_Y$ in Equation (1). Second, the analysis allows for a potential direct effect of endowments on institutions, measured by $\widetilde{\eta}_R$ in Equation (2). The latter channel accounts for the possibility that the organization of society and the quality of institutions depends directly on climate, disease, and other endowments. For example, terrain ruggedness may affect the fractionalization of the population along ethnic lines, thereby influencing the accountability of the local political elite, which in turn affected postcolonial institutions (see Gennaioli and Rainer (2007)). Third, the theories relating institutional origin to colonial experience predict that endowments affected colonization policies and institutional outcomes in former colonies, measured by $\widetilde{\theta}_P$ in Equation (3).

With these three distinct effects in mind, consider an estimation of the reduced form of Equations (1), (2), and (3) in a sample composed of former colonies such that $C_i = 1$ for all observations. In an instrumental variable estimation using this sample, the first-stage coefficient of endowments could be significant because colonization policies were affected by endowments ($\widetilde{\theta}_P \widetilde{\theta}_R$), because endowments have a direct effect on institutions ($\widetilde{\eta}_R$), or because endowments directly impact income, which in turn affects institutions ($\widetilde{\beta}\widetilde{\eta}_Y$). In the second-stage estimation of Equation (1), the effect of institutions on income could be overstated because the restriction that endowments do not directly affect development ($\widetilde{\eta}_Y = 0$) is needed to identify the system. Because of this restriction, the correlation between endowments and income is entirely attributed to the institutional channel, and the coefficient of instrumented institutional quality in (1) is biased if geography also has a direct effect on income.

In contrast, consider an estimation of the reduced form of Equations (1), (2), and (3) in a sample also including non-colonized nations.[13]

$$Y_i = \lambda_Y + \lambda_Y' C_i + \alpha \overrightarrow{R_i} + \eta_Y E_i + \nu_{Y,i} \tag{4}$$

$$\overrightarrow{R_i} = \lambda_R + \lambda_R' C_i + \eta_R E_i + \theta_R (E_i C_i) + \nu_{R,i} \tag{5}$$

where $\overrightarrow{R_i}$ is the first-stage projection of R_i. The first-stage estimation of the reduced-form model in Equation (5) includes the main effect of endowments, a colony dummy, and the interaction of these two variables. Because the additional variation in the group of noncolonized countries determines the coefficient of the direct impact of endowments on income (η_Y), the estimation can disentangle the true relation between institutions and income. η_R captures the direct effect of geography on institutional development while θ_R captures the institution-building effect of endowments during colonization.

There are two potential concerns with the identification strategy proposed in (5) and (4). A first concern is that initially richer countries with better institutions were more likely to become colonizers whereas initially poorer countries tended to be colonized. Because both institutional and income differences may persist until today, colonization in the past is likely correlated with current economic outcomes even if colonization itself had no effect on development. As is demonstrated in the Appendix 7, this endogeneity problem does not invalidate the identification strategy: the point coefficients of interest (θ_R and a) can still be identified because the identification strategy utilizes only the variation of endowments and institutional outcomes *within* each of these two groups, which is not affected by the endogenous selection *across* the two groups.

[13] When comparing the coefficients in Equations (1), (2), and (3) to the ones in (5) and (4), $\theta_R = \widetilde{\theta}_R \widetilde{\theta}_P / \left(1 - \widetilde{\alpha}\widetilde{\beta}\right)$ and $\nu_{R,i} = \left(\widetilde{\nu}_{Ri} + C_i \widetilde{\theta}_R \widetilde{\nu}_{Pi}\right) / \left(1 - \widetilde{\alpha}\widetilde{\beta}\right)$, demonstrating that there may be heteroskedasticity between the two groups of countries. All results presented below are thus estimated with heteroskedasticity-robust standard errors.

The second concern regards the "comparability" of colonies and noncolonies: are these two groups of nations intrinsically too different to be compared using the linear relations between endowments, institutions, and income implicitly assumed in Equations (3), (1), ad (2)? For example, Table 2 presents summary statistics for each of these two groups, listing the measures of economic outcomes, endowments, and the constructed measures of EDE and proximity to the UK. There are pronounced differences between those two groups, so some further examination is expedient.

The first facet of the comparability concern is that the relation between endowments and development is likely nonlinear. For example, a warmer climate is good for economic outcomes in the mostly temperate group of noncolonies whereas the reverse holds true in the mostly tropical group of former colonies. In the empirical analysis below, this concern is tackled by relying on nonparametric estimation techniques that account for the potentially nonlinear overall effect of an endowment on development and then examine whether an effect is present in the former colonies in addition to the common nonlinear effect.

A second facet of the comparability issue regards this study's assumption that the direct effects of geographic endowments on prosperity are equal across all countries. These direct effects of endowments may affect the richer noncolonies less than the poorer group of former colonies because richer countries are intrinsically better at dealing with unfavorable endowments, for example, by finding remedies against disease. Then, if intrinsically more capable societies both did not suffer much from disease and selected into colonizing other nations, the identification strategy proposed in this paper would be invalid.

The basic underlying assumption of this paper is thus that there are some general relations between geography and development that are common to all nations. It may be noteworthy that the limited power of European imperialists to deal with the forces of nature is actually demonstrated by the above-documented sizeable effects of endowments on colonization policies. If European settlers originated from intrinsically more capable nations, they should also have been able to deal with disease abroad. If European fleets were not hindered by high transportation costs, location should not have had an effect on legal origin. Thus, the very fact that geography had such a marked impact on the way in which nations were colonized serves as evidence that the colonizers were subject to the same forces of nature as were the colonies.

A note of caution is in order regarding the comparability of former colonies and the rest of the world and, therefore, the generality of the results presented below. The analysis of this study is based on the premise that the direct effects of geographic endowments on prosperity are equal across all countries. Nevertheless, the analysis does not assume that the effect of institutions on income is the same in these two groups of countries: the instrument employed utilizes the interaction of endowments and the colony dummy. It consequently varies only within the group of former colonies. The estimation results presented below thus measure the effect of institutions on income in the group of former colonies, but not necessarily in the rest of the sample.

4 The Partial Effects of Endowments and Institutions

4.1 Disease, Institutions, and Prosperity

This section applies the above-developed methodology to examine the theory of the colonial origins of institutions developed by Acemoglu et al. (2001), i.e., it examines whether the correlation between disease and income can be attributed to the direct importance of germs for prosperity or to the indirect effect of settler mortality on institutional development during colonization.

Table 3 displays the relation between EDE, institutions, and income differences. The upper Panel B presents the second stage, relating disease and institutional outcomes to income. The lower Panel A presents the relation between disease and colonial history with institutional outcomes. Column 1 regresses the settler mortality rate on income for the sample of former colonies in Acemoglu et al. (2001) (excluding Malta and the Bahamas, which do not satisfy the sample criterion of this study). Column 2 repeats this specification using EDE instead of the settler mortality rate. Because EDE is available for all countries, the sample includes all 95 colonies examined in this paper.

Columns 2 to 4 highlight the methodology of this paper. Columns 2 and 3 document that EDE is strongly correlated with development in former colonies, but this is not the case in the sample of noncolonies. In the estimation of Column 2 restricted to former colonies, a 1% lower level of early disease environment is associated with a 1.174% *higher* income per capita and a 0.566 percentage points *higher* score of the rule of law. In a non-colonized nation, the same difference is associated with a 0.292% *lower* income per capita and a 0.022 percentage points *lower* score of the rule of law (see Column 3).

Column 4 disentangles the direct and indirect institution-building effects of disease on prosperity. The assumption identifying the relation between institutions and income is that the additional impact of disease in former colonies is the exclusive result of the adopted colonization policies and, thus, institutions. A one-standard-deviation difference in institutional quality is estimated to result in a difference in income per capita of 1.62 ($\approx (1.17 - 0.29) / (0.566 - 0.022)$) log points. Column 4 also documents that disease environment has a substantial direct effect on income. For a given institutional quality, a one-standard-deviation difference in the level of EDE is associated with a -0.292 ($\approx -0.256 - 0.022 * 1.62$) log points difference in level of income per capita.

Although these findings highlight the importance of colonization policies for growth, they also document that the point estimates of Acemoglu et al. (2001) are mildly inflated because they attribute all of the correlation between disease and development to the institutional channel. Column 5 documents the magnitude of this overestimation. Consider again a 1% lower EDE, but this time in the estimation of Column 5, which includes only former colonies. In this group, a 0.01 lower EDE is associated with an increase in the score for the rule of law of 0.566 percentage points. Because the direct effect of mortality is restricted to equal zero, the estimation attributes all of the difference in income levels to institutional quality. The coefficient of the rule of law in Column 5 is estimated to be 2.077, which - up to rounding - satisfies $0.566 * 2.077 \approx 0.566 * 1.624 + 0.256$. The importance of institutions is overstated by about 27% in the sample restricted to former colonies.

Although the direct effect of disease on growth is thus nonnegligible, it should also be noted that the results of this paper still indicate that, for former colonies, the effect of early disease environment on economic growth works mostly through the indirect effects on settler mortality, colonization policies, and institutional outcomes.

It is worthwhile to examine how these uncovered magnitudes compare with the results of alternative approaches. In particular, a recent literature estimates bounds on IV estimation coefficients if the exclusion restriction is not fully satisfied. Columns 6 and 7 use the "plausibly exogenous" IV estimator of Conley et al. (2007), which relaxes the exclusion restriction and performs sensitivity analysis with respect to the degree to which this restriction is violated.

For the sensitivity analysis, the plausibly exogenous IV estimator requires the use of prior information on the distribution of potential values in the degree to which the exclusion restriction is violated. In the context of the analysis at hand, this corresponds to an a-priori estimate of the direct effect of health on growth. Column 6 uses point estimates from Weil (2007), who uses microeconomic estimates of health's effect on individual income to construct macroeconomic estimates of health's effect on income per capita. Because Weil (2007) in large part focuses on the effect of adult survival rates (ASR, defined as the fraction of fifteen-year-olds who will survive to age sixty), his point estimates are particularly comparable to the settler mortality measure of Acemoglu et al., which is defined as the annual mortality rate of adults ($ln(ASR) = -45 * ln(SettlerMortalityRate)$). Weil reports a variety of estimates for the effect of ASR on income per capita, finding that a one-standard-deviation difference in ASR is associated with a difference in GDP of between 7.1% and 44.7%. The estimation in Column 7 uses bounds on the direct effect of health on income derived from Bloom and Canning (2005), who assume that health directly multiplies human capital in the production function. These authors also use adult survival rates to measure health, and their results imply that a one-standard-deviation difference in output is associated with a difference in income per capita of between 15.6% and 77.4%.[14]

The estimations in Columns 6 and 7 show that using this alternative approach results in very wide confidence intervals for the effect of institutions on growth. When using the estimates from Weil as bounds for the direct effect of health on income, the 99% confidence interval estimated for the coefficient of the rule of the law on GDP per capita is estimated from 0.472 to 1.42. In Column 7, when using the bounds from Bloom and Canning (2005), the 99% CI is estimated from

[14] These bounds are computed in the following way. Weil surveys the literature on the effect of height on earnings, finding that three different "methods for estimating the return to health characteristics produce a total of six estimates for the return to height: three from variation in childhood inputs (0.080, 0.094, and 0.078), two from twins (0.033 and 0.035), and one from long-run historical data (0.073)" (see page 1288). The return to height is expressed in the percentage increase of an individual's wage per 1 cm of height. Second, Weil regresses the effect of ASR on height, finding a coefficient of between 16.6 and 26.4 (see his Table 2 on page 1291). Comparing these estimates reveals that a change of the ASR from 0 to 1 implies an effect of income ranging from $0.033 * 16.6 = 54.78\%$ to $0.094 * 26.4 = 248.2\%$. Last, taking into account that the standard deviation of ASR in Weil's sample ranges from 0.13 to 0.18 yields the numbers stated in the main text: from $0.033 * 16.6 * 0.13 \approx 7.1\%$ to $0.094 * 26.4 * 0.18 \approx 44.7\%$. The relevant figures from Bloom and Canning (2005) result in a 95% percent confidence interval of $120\% - 430\%$ for the coefficient of ASR on a worker's income. Using this range and the standard deviation of ASR yields a range from 15.6% to 77.4%.

-0.13 to 0.911, i.e., the coefficient of the rule of law is not significant at the 1% level.

In contrast, the estimation in Column 4, which uses the interaction methodology developed in this paper, yields a 99% CI ranging from 1.1 to 2.15. Thus, the methodology proposed in this paper yields an effect of colonization-induced institutions that results in both economically and statistically more significant support for the theory of Acemoglu et al. (2001) than would be yielded by alternative methodologies using additional information.

The estimated direct effect of EDE on growth of -0.292 ln points per standard deviation difference in EDE is also consistent with findings from the literature directly estimating this effect from microeconomic information: taking the average of Weil's (2007) estimates (6.55% effect of one cm of height on income, 20.825 of ARS on height, and an ASR standard deviation of 0.15) results in an average estimate of the difference in health of about -0.200 ln points per standard deviation. Taking the same average from the point coefficients constructed in Bloom and Canning (2005) yields an average estimate of -0.412. Thus, the point estimate determined here corresponds roughly to the averages reported in Weil (2007) and Bloom and Canning (2005).

The presented findings are not the result of latent nonlinearities, which is documented in Figure 2 and in Columns 8 and 9 of Table 3. Figure 2 presents a scatter plot relating EDE to the country's score for the rule of law. Former colonies are represented by an "x," and noncolonies are represented by a circle. For each of these groups, a linear regression fit is displayed. As shown in Figure 2, former colonies are, on average, characterized by high levels of EDE. Correspondingly, a positive interaction could also have resulted from the stronger relation between disease and prosperity for higher values of EDE.

To document that this is not the case, Column 8 of Table 4 adds "EDE Square" to the estimation. EDE Square is equal to the square of the difference between EDE and the minimum observed value of EDE. The addition of EDE Square to the estimation in Column 8 has no substantial impact on the interaction coefficient. Moreover, the coefficient of EDE Square is positive, implying that the relation between disease environment and institutions becomes weaker for high values of mortality: even if the data exhibit nonlinearity, this would bias the interaction term of mortality in a purely linear estimation towards being positive, not negative.

Column 9 deals more directly with any potential nonlinearities. This estimation proceeds in two steps, the first of which is to allow for a nonparametric general relation of the form rule of law=f(EDE), where f(...) is estimated using the entire sample. Using this nonparametric estimation, "nonparametrically smoothed EDE" is predicted, which is equal to the smoothed value of EDE, i.e., nonparametrically smoothed EDE=f(EDE). Given that the smoothed EDE variable, by construction, has a linear effect on the rule of law, the interaction coefficient only picks up differences in how EDE affects the rule of law in colonies and noncolonies. Again, I find that the effect of institutions on outcome in the linear specification of Column 4 underestimates the effect of the rule of law.

4.2 Location, Legal Origin, and Prosperity

La Porta et al. (1997, 1998, and 1999) argue that differences in the historical origins of the legal system—most notably in whether the country has adopted a system based on common or civil law—resulted in considerable differences in economic outcomes. The authors are well aware that countries that were initially more successful could have adopted better legal systems, and they argue that the relation is causal because legal institutions were often superimposed by a foreign colonizer. They further argue that the random variation in legal systems that was induced by colonization can be utilized to establish the effect of legal origin on prosperity.

However, the analysis in Section 2 demonstrated that countries with different legal origins also tend to have very different locations. Are coastal nations richer on average as a result of the generally positive impact of access to the open sea and geographic proximity to trade partners (See Frankel and Romer (1999), Rodriguez and Rodrik (2000), Rodrick et al. (2004), Dollar and Kray (2003), and Feyrer (2009A and 2009 B))? Or is it the case that Britain, with its substantial naval power, tended to colonize such nations more often? To answer this question, this section uses the above-constructed estimates of relative proximity to Britain, France, and other nations.

Table 4 examines whether relative proximity to Britain, France, and the other colonizers also influenced economic outcomes directly. Columns 1 to 3 serve to compare the empirical approach of this study to empirical specifications in the spirit of La Porta et al. In Column 1, the sample includes all former colonies, and the dependent variable includes only the measure of "Proximity to Britain," hence comparing a British legal origin to all other legal origins. In the upper Panel

B, the dependent variable is the logarithm of GDP per capita, and the sample includes 95 former colonies. In Column 2, the same specification is estimated in the sample of 54 noncolonies.

Although the probability of being colonized by the UK, conditional on being colonized, is strongly positively associated with development in the group of former colonies, the opposite is the case in the rest of the world. In the estimation in Column 1 of Panel B in Table 4, the sample includes only former colonies, and the coefficient of proximity to Britain is estimated at 0.99; i.e., for former colonies, a location relatively closer to Britain such that the country is 10% more likely to be colonized by the UK (where proximity takes values between 0 and 1) leads to a 0.099 ln-points increase in predicted income per capita. In contrast, the same estimation in the sample of noncolonies results in a negative coefficient of -1.67. A similar pattern of a reversed effect of remoteness can be found when evaluating a simple regression of proximity to the UK on the rule of law in Panel A: the estimated coefficient is positive and significant for the sample of former colonies but negative and significant in the sample of noncolonies.

Next, consider the IV estimation in Column 3. In this estimation, the first-stage estimation is presented in Panel A and the corresponding second-stage estimation in Panel B. In Column 3, the rule of law is instrumented with the geographic prediction of the British colony dummy interacted with the colony dummy. The interpretation of this interaction is thus that whereas remoteness from Europe may directly affect growth, the additional effect present in former colonies is driven by the institution-building channel, i.e., the fact that remote nations tended to be colonized by the British. In this estimation, the first-stage interaction coefficient is estimated at 2.22, i.e., for a given direct impact of location, a country that is 10% more likely to adopt a British legal origin has a higher predicted score for the rule of law by 0.222 points. In the second-stage estimation in Panel B, a change in the rule of law of one standard deviation is associated with a change in income per capita of 1.2 log points.

On the contrary, the direct effect on growth of location such that a country is more likely to be colonized by the British is negative. For a noncolony, a location such that the country is 10% more likely to be colonized by the British is associated with a 0.129 lower score of the rule of law. In addition, a direct estimated effect of remoteness on income of -0.12 is observed, which, in turn, is associated with a 0.275 ($\approx 0.129 * 1.2 + 0.12$) lower ln income.

The methodology of this paper yields point estimates for the effect of institutions on income that are slightly higher than those obtained when using an identification strategy based on La Porta et al. To compare the results presented in Column 3 to the existing literature, Column 4 presents an estimation in the spirit of La Porta et al. In this IV estimation, the sample includes only colonies, and the instrument is the measure of being colonized by the British. The identifying assumption made in this specification is equivalent to that of La Porta et al., i.e., that a location making the country likely to be colonized by the British affects development via legal origin but that such "proximity" to Britain itself has no impact on prosperity.

The coefficients for the rule of law in Columns 3 and 4 compare as follows. In a former colony, a change of 0.1 in the score for "proximity to Britain" is associated with a difference in the score for the rule of law of 0.93 and a difference in the logarithm of GDP per capita of 0.99, hence resulting in a coefficient of $0.99/0.93 \approx 1.06$ for the rule of law. In the specification of Column 3, which also allows for proximity to Britain to affect income directly, the additional direct effect of the difference in GDP of -0.12 log points is attributed to the direct impact of location on income. Consequently, the coefficient for the rule of law is estimated at $(0.99 + 0.12)/0.93 \approx 1.20$.

Nevertheless, the importance of legal origin for development is much higher in the estimation using noncolonies as a control group in Column 3 than when only using the group of former colonies in Column 4.

The most important pattern uncovered in Column 2 is that location has a sizeable direct negative impact on institutional outcomes. For noncolonies, higher proximity to Britain is associated with worse institutional outcomes. This correlation, reflecting the detrimental effect of the lack of possibilities for trade and the associated effect on the local political economy, leads to an underestimation of the causal impact of legal origin on economic development. For example, the effect of a higher likelihood of adopting a legal system based on the British system is estimated to be about 50% larger in Column 2 than in Column 1, which neglects the direct effect of access to trade on economic outcomes.[15]

[15] Segura-Cayuela (2006) argues that openness to trade can hurt a country's institutional performance because liberalization removes general equilibrium price effects of expropriation. Heavily expropriated goods are expensive in a closed economy whereas the world market price is given in a small open economy. Liberalization thus enables groups with political power to engage in more costly expropriation.

The estimation of Column 4 underestimates the importance of legal origin for institutional development because remoteness from Europe had two effects on development that work in opposite directions. On the one hand, remote nations tended to be colonized by the British, hence resulting in better institutional outcomes. On the other hand, remoteness itself is detrimental for growth. These opposing effects suggest that Frankel and Romer (1999) underestimate the direct effect of access to trade (although the measures of proximity constructed in this section are arguably only crude measures for the geographic potential for trade because other centers of economic activity, such as East Asia, have arisen since the World Wars).

How do the results shown in Columns 3 and 4 compare to results from an approach using exogenous information on the effect of trade on growth? Column 5 repeats the specification of Column 4 using the Plausibly Exogenous IV estimator of Conley et al. (2007). To perform the sensitivity analysis with respect to the degree to which the exclusion restriction in the specification of Column 4 is violated, I use bounds from Feyrer's (2009 A and B) work on the effect of trade on growth. Both papers use time series variation in how geography has affected trade differently during different periods: the shift toward air transportation and the $1967-75$ closing of the Suez Canal, respectively. Because the identification strategy in Feyrer (2009 A and B) relies on time-series variation (instead of on the cross section, as Frankel and Romer's (1999) analysis does), the results of these papers can be utilized to augment the cross-sectional analysis of this paper.

Column 5 thus uses the plausible lower bound of -0.753 for the effect of the conditional probability of being colonized by the UK, imputed from the IV estimations in Feyrer (2009 A). It uses an upper bound of -0.157 imputed from the IV estimations in Feyrer (2009 B).[16] Using these bounds, the 99% CI for the coefficient of the rule of law is found to lie between 0.835 and 2.91. The plausibly exogenous IV approach is much less precise than the methodology proposed in this paper (the same 99% CI in Column 3 ranges from 0.84 to 1.56).

The estimations in Columns 1 to 4 in Table 1 differentiate only between UK and non-UK

[16] These numbers are calculated in the following way. Feyrer's (2009 A) highest estimate for the effect on growth of a 1% higher growth rate of trade is equal to 0.754. Feyrer's (2009 B) lowest estimate is equal to 0.157. A regression of the logarithm of trade on the logarithm of sailing distance to Europe, as defined above in Table 1, yields a coefficient of -3.01; i.e., a 1% increase in sailing distance to Europe lowers trade by 3.01%. In turn, the coefficient of Prop UK on Distance to Europe in a simple regression is equal to 0.332. Taking the bounds from Feyrer's works and these coefficients yields bounds of -0.753 $(0.754 * 0.332 * (-3.01))$ and -0.157 $(0.157 * 0.332 * (-3.01))$.

colonies, but there may also be differences in economic outcomes between the other colonizers. The estimation in Column 6 adds relative proximity to France, leaving Spanish, Portuguese, Austrian, Italian, Dutch, and Belgian colonies in the omitted group.[17] In this specification, proximities to both France and Britain are significant direct determinants of GDP, and the estimated coefficients are nonnegligible in magnitude. Proximity also has a sizeable and significant direct effect on the rule of law, where the direct effect is of opposite sign as the indirect effect working via colonization policies. Hence, this specification again confirms that conventional specifications underestimate the importance of colonization to institutional development and, in turn, of institutional development to income.

Columns 7 to 9 analyze the effect of legal rather than colonial origin on institutional outcomes and long-term growth. As demonstrated in Table 1 above, countries that have adopted the British legal system are relatively closer to Britain than to France and are relatively more distant from Europe. Column 7 estimates baseline comparisons between economic outcomes in countries with British vs. non-British legal origin. Column 8 also adds "other," defined as German, Scandinavian or Russian legal origin, leaving French legal origin as the omitted group. The first-stage interaction effect of probability to adopt a British legal origin, conditional on being colonized, is somewhat larger in magnitude than the comparable first-stage interaction in Column 3 investigating the effect of colonial origin.

An important robustness test is presented in Column 9, which addresses the fact that the direct effects of proximity could be nonlinear. Because noncolonies include many European nations, they are, on average, close to Britain or France (see Figure 3). Thus, a positive interaction of "proximity to the UK" and the colony dummy could also have resulted from the fact that the relation between distance and prosperity is stronger for longer distances. Following the approach developed in the previous sections, I thus again estimate a nonparametric specification to deal directly with any potential nonlinearities. This estimation proceeds in two steps, the first of which is to allow for a nonparametric general relation rule, f(proximity UK), that is estimated using the entire sample. Using this nonparametric estimation, I predict "nonparametrically smoothed

[17] The two measures of proximity to either the UK or France have been derived from a multinomial Probit estimation allowing for the relating absolute and relative sailing distances to colonial origin. See the working paper version for the estimation results.

Proximity UK," which is equal to the smoothed value of Proximity UK, i.e., smoothed Proximity UK=f(Proximity UK).

Given that the smoothed proximity variable, by construction, has a linear effect on the rule of law, the interaction coefficient only picks up differences in how distance affects the rule of law in colonies versus noncolonies. Again, this difference is highly significant, and the estimated effect of institutions on income per capita is similar to the baseline estimation of Column 3.

5 Robustness Analysis

Table 5 examines the robustness of the findings presented in the previous section with respect to changes in the sample, the addition of further controls, and the use of an alternative measure of institutional outcomes. The structure of Table 5 mirrors that of Tables 3 and 4.

The IV estimation of Column 1 presents a joint estimation including EDE and the measure of geographic likelihood to adopt a British or other legal origin as controls (French legal origin is the omitted group). It includes the interactions of these three variables with the colony dummy as instruments for institutions. Moreover, in this joint estimation, both sets of instruments (EDE and the legal origin probabilities, both interacted with the colony dummy) are statistically and economically significant. This exercise of using both the Acemoglu et al. and the La Porta et al. instruments allows us to test the overidentified system; i.e., it tests whether the interaction of EDE with the colony dummy yields the same prediction with regard to the rule of law as does the interaction of the proximity variables. This overidentification test, reported at the bottom of Panel C, cannot be rejected at the 10% level of significance.

Columns 2 to 5 check whether the results presented so far are driven by the inclusion of specific groups of countries. A first concern is that most countries in Africa have been colonized and are poor and characterized by adverse endowments. If African countries are poor for reasons other than colonization, the inclusion of this group could be the sole driver of the presented relations. In Column 2, the sample thus excludes all 47 countries on the African tectonic plate.

A second concern is that the group of oil-rich nations on the Arabian Peninsula, with a dry and hot climate, are not representative of the theories of development examined in this study because their wealth from oil has overshadowed all other forces of development. The estimation in Column 3 thus excludes 16 Middle Eastern nations (Bahrain, Cyprus, Egypt, Arab Republic, Iran, Islamic Republic, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Turkey, United Arab Emirates, Republic of Yemen; United Nations definition according to Wikipedia).

On the European continent, southern countries are significantly poorer, perhaps because of specific historical developments over the last 500 years (see Acemoglu et al. (2005)) rather than to the pure impacts of climate and transportation costs. To investigate whether this is the case, the estimation in Column 4 thus excludes 10 southern European nations (Albania, Bosnia and Herzegovina, Croatia, Greece, Italy, Macedonia, Portugal, Slovenia, Serbia, and Spain; United Nations definition). Among the group of former colonies, the "neo-Europes," Australia, Canada, New Zealand, and the USA, stand out in that they are rich and endowed with rather mild climates. The estimation in Column 5 thus excludes these four "neo-Europes". Last, one could argue that former Soviet countries were in fact not truly independent nations, and a similar case could be made for all former members of the Warsaw Pact. To address this potential concern, the estimation in Column 6 excludes all former members of the Warsaw pact except Russia itself.

Inclusion or exclusion of each of the above-mentioned groups of countries only has a limited impact on the estimated coefficient of the rule of law, the interaction coefficients in the first-stage estimation, and the direct impact of endowments in the second-stage estimation.

Rather than changing the composition of the sample, Columns 7 to 10 add different sets of controls to the estimation. Both economic outcomes and geographic endowments vary considerably across the continents, but to a much lesser extent within each continent. Are the results presented so far driven by cross-continent differences, or can endowments and colonial history also explain differences within continents? To answer this question, the estimation in Column 7 includes continent dummies for Africa, Asia, Europe, and Oceania, thus making the Americas the omitted group. In this estimation, owing to the relatively small within-continent variation of GDP per capita, the coefficient for the rule of law is estimated somewhat lower, at 1.2. Nevertheless, the coefficient is significant at high levels, the first stage is well identified, and the overidentification test cannot be rejected.

Column 8 adds ethnic fractionalization from Alesina et al. (2004) to the estimation. This variable takes values between 0 and 1 and is higher for societies that are ethno-linguistically more fractionalized. Such fractionalization could be detrimental for institutional outcomes because internal conflict arises more often, thereby making it easier for the ruling elite to play groups against each other, as highlighted, for example, by Padro-I-Miquel (2007). Confirming the identification assumption made in the empirical analysis of Mauro (1995), fractionalization indeed influences development mostly through its impact on institutional outcomes. However, the addition of this variable has no impact on the main and interaction effects of endowments or the estimated coefficient for the rule of law.

As Glaeser et al. (2004) have argued, colonization policies may also have affected development directly via the human capital that European settlers "embodied" when moving to the colonies. Because the offspring of these settlers continued to invest heavily in human capital, colonization and settlement policies could have affected growth via the level of human capital rather than via the quality of institutions. Column 9 thus adds "Literacy Rate in the Adult Population" from Parker (1997) as a control to the estimation. Also conditional on the level of human capital, institutions are a major force of development. While it is highly likely that colonization policies also influenced the rate of human capital accumulation (see, for example, Gallego (2010) for an in-depth analysis of the effect of colonization on schooling policies), the effect on growth via institutional features other than schooling is also present conditional on the effect on human capital accumulation.

In Column 10, to control for a richer set of geographic information, the specification adds seven geographic variables to the estimation. The logarithm of elevation, a dummy equal to one for landlocked countries, distance from the equator, the length of the country's coastline, the percentage of a country's surface area that is arable, and the "Total Sum of Minerals"—all from Parker (1997)—are included in the estimation. Total Sum of Minerals is equal to the sum of the country's share of world reserves in the 20 most important minerals (excluding oil). The estimation also adds Malaria Ecology from Kiszewski et al. (2004).

As has been argued by Glaeser et al. (2004), colonization policies may also have affected development directly via the human capital that European settlers "embodied" when moving to the colonies. Since also the offspring of these settlers continued to invest heavily in human capital, colonization and settlement policies could have affected growth via the level of human capital rather than via the quality of institutions. Column 9 thus adds "Literacy Rate in the Adult Population" from Parker (1997) as a control to the estimation. Also conditional on the level of human capital, institutions are a major force of development. While it is plausible that colonization policies did also influence the rate of human capital accumulation (see, for example, Gallego (2010) for an in-depth analysis of the effect of colonisation on schooling policies), the effect on growth via other institutional features other than schoolling is also present conditional on the effect on human capital accumulation.

In Column 10, to control for a richer set of geographic information, the specification adds seven geographic variables to the estimation. The logarithm of elevation, a dummy equal to one for landlocked countries, distance from the equator, the length of the country's coastline, the percentage of a country's surface that is arable, and the "Total Sum of Minerals" – all from Parker (1997) – are included to the estimation. Total sum of minerals is equal to the sum of the country's share in world reserves in the 20 most important minerals (excluding oil). The estimation also adds Malaria Ecology from Kiszewski et al. (2004).

The last robustness check of Table 4 examines whether the importance of institutions hinges on the use of the score for the rule of the law to measure institutional outcomes. Column 11 instead uses the 1996 to 2004 average for "Control of Corruption" from Kaufmann et al. (2005), measuring the extent to which public power is exercised for private gain. Control of Corruption is standardized in the same fashion as is the score for the rule of law, and the second-stage coefficients for these two (instrumented) measures of institutional outcomes hence can easily be compared. Indeed, the coefficient is very similar; it is estimated at 1.16, as compared with 1.36 when using the score for the rule of law.

6 Conclusion

This paper estimates the partial effects of geographic endowments and institutions on income. The existing literature partly fails to distinguish between these two channels of development because endowments have influenced colonization policies and institutions, but they may also affect prosperity directly.

The paper's basic insight is that one can utilize the interaction of history and geography to distinguish the effects of institutions and geographic endowments on comparative development. Historical events—such as colonization or the rise of trade with the New World—have influenced how climate, transportation costs, and disease have affected development. For example, during colonization, the mortality rates of European settlers affected colonization policies, which in turn determined the quality of institutions in the respective colonies. Disease environment may, however, also directly affect economic outcomes.

Whereas endowments directly affect income in all countries, the institutional channel only applies to a subset of countries, namely former colonies. On the basis of this insight, I develop an instrumental variable framework that identifies the relation between income and institutions while also allowing geographic endowments to directly affect growth.

I apply the developed methodology to examine the theories of Acemoglu et al. (2001) and La Porta et al. (1997) that relate settler mortality rates or the historical origin of the legal system to institutional outcomes. Although I confirm both of these theories, I also demonstrate that their empirical evidence somewhat overstates the importance of their starting hypotheses. For the case of settler mortality rates, I document that about one-quarter of the correlation between disease and income can indeed be attributed to the direct effect of the disease rather than the indirect effect of settler mortality rates on colonization policies. For the case of legal origins, I document that the causal effect of having a common law origin is in fact somewhat larger than the existing empirical literature suggests. The reason for this is that Britain, with its strong navy, tended to colonize nations that are remote from Europe. This remoteness had a detrimental direct effect on development, hence partly masking the positive impact of an efficient legal system on economic development.

These two examples highlight the main conclusion of this study: although endowments do matter directly for income differences today, they mattered even more in the past. Because the same variables have affected development through different channels at different stages in history, only the interaction of history and geography can clearly identify the forces of development.

References

[1] **Acemoglu, Daron and Simon Johnson.** 2005. "Unbundling Institutions." *Journal of Political Economy*, 113(5), pp. 949-995.

[2] **Acemoglu, Daron, Simon Johnson, and James A. Robinson**. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review*, 91(5), pp. 1369-1401.

[3] ——. 2002. "Reversal of fortune: Geography and institutions in the making of the modern world income distribution." *The Quarterly Journal of Economics*, November 2002, 117 (4), pp. 1231-1294

[4] ——. 2003 "Disease and Development in Historical Perspective." *Journal of the European Economic Association, Papers and Proceedings,* 1(2-3), pp. 397-405.

[5] **Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg**. 2003 "Fractionalization." *Journal of Economic Growth*, 8 (2), pp. 155-94.

[6] **Arapovic,** Domagoj and **Auer,** Raphael (2010), "A Database of the Colonial History of Modern Nations During Imperialism", Mimeo, Princeton University (availalbe on request).

[7] **Barro, Robert J. and Lee Jong-Wha**. 2000 "International Data on Educational Attainment: Updates and Implications." Manuscript, Harvard University, Cambridge.

[8] **Black**, Jeremy and **Woodfine**, Philip Editors 1988 "The British navy and the use of naval power in the eighteenth century", Leicester University Press, 1988.

[9] **Bloom, David E. and David Canning**, "Health and Economic Growth: Reconciling the Micro and Macro Evidence," Mimeograph, Harvard School of Public Health, 2005.

[10] **Bloom, David E. and Jeffrey D. Sachs**. 1998 "Geography, Demography, and Economic Growth in Africa." *Brookings Papers on Economic Activity*, 1998 (2), pp. 207-73.

[11] **Encyclopedia Britannica** (2010), http://www.britannica.com/EBchecked/topic/126354/colonization, accessed on 15,August, 2010.

[12] **Dell, Melissa, Benjamin F. Jones and Benjamin Olken**. 2008. "Climate Shocks and Economic Growth: Evidence from the Last Half Century" November 2008. NBER Working Paper #14132.

[13] **Diamond, Jared M**. 1997 *Guns, germs and steel: The fate of human societies.* New York: W.W. Norton & Co..

[14] **Djankov, Simeon, Ed Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer**. 2003. "The New Comparative Economics." *Journal of Comparative Economics*, 31(4), pp. 595-619.

[15] **Dollar, David, and Aart Kraay**. 2003. "Institutions, Trade and Growth: Revisiting the Evidence," *Journal of Monetary Economics*, 50(1), pp. 133-162.

[16] **Easterly, William and Ross Levine**. 2003. "Tropics, germs, and crops: how endowments influence economic development," *Journal of Monetary Economics*, 50(1), pp. 3-39.

[17] **Engerman, Stanley L. and Kenneth L. Sokoloff**. 1997. "Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies." In: Stephen Haber, ed., *How Latin America fell behind*,pp. 260-304.Stanford, CA: Stanford University Press

[18] **Frankel, Jeffrey A. and David Romer**. 1999. "Does Trade Cause Growth?", *The American Economic Review*, Vol. 89 (3), pp. 379-399

[19] **Feyrer, James D. and Bruce Sacerdote**. 2008. "Colonialism and Modern Income – Islands as Natural Experiments" *The Review of Economics and Statistics.* Forthcoming.

[20] **Feyrer, James D.** (2009A) "Trade and Income – Exploiting Time Series in Geography" NBER Working Paper 14910, April 2009

[21] **___** (2009B) " Distance, Trade, and Income – The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment" NBER Working Paper 15557, December 2009

[22] **Gallego**, Francisco A. (2010). "Historical Origins of Schooling: The Role of Democracy and Political Decentralization", The Review of Economics and Statistics, Vol. 92, No. 2, Pages 228-243, 2010.

[23] **Gallup, John L., Andrew D. Mellinger, and Jeffrey D. Sachs**. 1998. "Geography and Economic Development," National Bureau of Economic Research, Working Paper No. 6849.

[24] **Ed Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer**. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9 (3), pp. 271 -304.

[25] **Gennaioli, Nicola and Ilia Rainer**. 2007. "The Modern Impact of Precolonial Centralization in Africa" The *Journal of Economic Growth*, 12(3), pp.185-234.

[26] **Hall, Robert E. and Charles I. Jones**. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *The Quarterly Journal of Economics*, 114 (1), pp. 83-116.

[27] **Kaufmann, Daniel; Aart Kraay and Massimo Mastruzzi**. 2005. "Governance Matters IV: Governance Indicators for 1996-2004" World Bank Policy Research Working Paper Series, No. 3630.

[28] **Kiszewski, Anthony; Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey D. Sachs**. 2004. "A Global Index of the Stability of Malaria Transmission." *American Journal of Tropical Medicine and Hygiene*, 70(5), pp. 486-498.

[29] **Knack, Stephen and Philip Keefer**. 1995 "Institutions and Economic Performance: Cross-Country Tests Using Alternative Measures." *Economics and Politics*, 7(3), pp. 207-227.

[30] **La Porta, Rafael; Florencio Lopez-de-Silanes, Andrei Shleifer and Robert W. Vishny**. 1997. "Legal determinants of external finance," *Journal of Finance*, 52(3), pp. 1131-50.

[31] ——. 1998. "Law and Finance," *The Journal of Political Economy*, 106(6), pp. 1113-1155.

[32] ——. 1999. "The Quality of Government," *Journal of Law, Economics and Organization*, 15(1), pp. 222-279.

[33] **Lipset, Seymour M**. 1960. *Political Man: the Social Bases of Politics*. Doubleday: Garden City, N.Y.

[34] **Mauro, Paolo**. 1995."Corruption and Growth." *The Quarterly Journal of Economics*, 110(3), pp. 681-712.

[35] **Mellinger, Andrew; John Gallup, and John Jeffrey D. Sachs**. 2000. "Climate, Coastal Proximity, and Development." in *Oxford Handbook of Economic Geography*, ed. Gordon L. Clark, Maryann P. Feldman, and Meric S. Gertler, Oxford University Press.

[36] **North, Douglass C**. 1981. *Structure and change in economic history*. New York: W. W. Norton & Co.

[37] **Nunn, Nathan and Diego Puga**. 2010. "Ruggedness: The Blessing of Bad Geography in Africa" Working Paper, Harvard University.

[38] **Parker, Philip M**. 1997. *National cultures of the world: A statistical reference*, *Cross Cultural Statistical Encyclopedia of the World, Vol. 4* Westport, CT: Greenwood Press.

[39] **Rodrik, Dani; Arvind Subramanian, and Francesco Trebbi**. 2004. "Institutions rule: The primacy of institutions over geography and integration in economic development", *Journal of Economic Growth*, 9 (2), pp. 131-165.

[40] **Sachs, Jeffrey D.** 2003. "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income". February 2003. NBER Working Paper #9490

[41] **Segura-Cayuela**, Ruben (2006). "Inefficient Policies, Inefficient Institutions and Trade," 2006 Meeting Papers 502, Society for Economic Dynamics

[42] **Weil, David N**. 2007."Accounting for The Effect of Health on Economic Growth," *The Quarterly Journal of Economics*, 122 (3), pp. 1265-1306.

7 Appendix: A Remark on the Endogeneity of Colonisation

This remark and its proof show that the identification does *not* assume that colonization is orthogonal to either income or institutions. If colonization is correlated with $\widetilde{\nu}_{Y,i}$ *or* $\widetilde{\nu}_{R,i}$, the colony dummies λ'_Y and λ'_R are biased, but the other coefficients are not affected.

Remark 1 *Assume that*

$$\widetilde{\nu}_{R,i} = \gamma_R C_i + \widetilde{\epsilon}_{R,i} \quad and \quad \widetilde{\nu}_{Y,i} = \gamma_Y C_i + \widetilde{\epsilon}_{Y,i} \ ,$$

where, by construction, $\widetilde{\epsilon}_{R,i}, \widetilde{\epsilon}_{Y,i} \perp C_i$*. Denote the expectation of the two-stage least square point estimates of* θ_R *and* α *in the estimation of (4) and (5) by* $E\left[\widehat{\theta}_R\right]$ *and* $E\left[\widehat{a}\right]$*. It is true that*

$$\begin{aligned} E\left[\widehat{\theta}_R\right]\Big|_{\gamma_{R\neq 0} \ or \ \gamma_{Y\neq 0}} &= E\left[\widehat{\theta}_R\right]\Big|_{\gamma_{R=0} \ and \ \gamma_{Y=0}} = \theta_R \\ E\left[\widehat{a}\right]\Big|_{\gamma_{R\neq 0} \ or \ \gamma_{Y\neq 0}} &= E\left[\widehat{a}\right]\Big|_{\gamma_{R=0} \ and \ \gamma_{Y=0}} \end{aligned}$$

Proof. *Consider first the structural model (1) and (2), with the impact of colonization policies (3) netted into the determinants of the rule of law.*

$$Y_i = \widetilde{\lambda}_Y + \widetilde{\delta}_Y C_i + \widetilde{\alpha} R_i + \widetilde{\eta}_Y E_i + \widetilde{\nu}_{Y,i} \tag{6}$$

$$R_i = \widetilde{\lambda}_R + \widetilde{\delta}_R C_i + \widetilde{\eta}_R E_i + \widetilde{\beta} Y_i + C_i \widetilde{\theta}_R E_i + \widetilde{\nu}_{R,i} \tag{7}$$

The reduced from of the first stage (7) is

$$R_i = \lambda_R + \lambda'_R C_i + \eta_R E_i + \theta_R C_i E_i + v_{R,i},$$

where $\lambda_R = \frac{\widetilde{\lambda}_R + \widetilde{\beta}\widetilde{\lambda}_Y}{1-\widetilde{\alpha}\widetilde{\beta}}$, $\lambda'_R = \frac{\widetilde{\delta}_R + \widetilde{\beta}\widetilde{\delta}_Y + \widetilde{\beta}\gamma_Y + \gamma_R}{1-\widetilde{\alpha}\widetilde{\beta}}$, $\eta_R = \frac{\widetilde{\eta}_R + \widetilde{\beta}\widetilde{\eta}_Y}{1-\widetilde{\alpha}\widetilde{\beta}}$, $\theta_R = \frac{\widetilde{\theta}_R}{1-\widetilde{\alpha}\widetilde{\beta}}$ *and* $v_{R,i} = \frac{\widetilde{\beta}\gamma_Y + \gamma_R}{1-\widetilde{\alpha}\widetilde{\beta}} C_i + \frac{\widetilde{\epsilon}_{R,i} + \widetilde{\beta}\widetilde{\epsilon_{Y,i}}}{1-\widetilde{\alpha}\widetilde{\beta}}$*. If either* $\gamma_Y \neq 0$ *or* $\gamma_R \neq 0$, $v_{R,i}$ *is correlated with the colonization dummy. Denote all estimated coefficients by a* $\widehat{}$ *superscript. The four FOCs of the OLS minimization problem yield* *the following point estimates for the coefficients*

$$\widehat{\lambda}'_R = \frac{\sum\limits_{i,D=1} (Y_i - (\eta+\theta)X_i)}{N_1} - \frac{\sum\limits_{i,D=0} (Y_i - \eta X_i)}{N - N_1} \quad and \quad \widehat{\lambda}_R = \frac{\sum\limits_{i,D=0} (Y_i - \eta X_i)}{N - N_1},$$

$$\widehat{\eta}_R = \frac{Cov(Y,X|D=0)}{Var(X|D=0)} \quad and \quad \widehat{\theta}_R = \frac{Cov(R,E|D=1)}{Var(E|D=1)} - \frac{Cov(R,E|D=0)}{Var(E|D=0)}.$$

Due to the endogeneity of colonization, $E\left[\widehat{\lambda}'\right] \neq \lambda'_R$*, but* $\widehat{\theta}_R$ *is an unbiased estimator of* θ :

$$E\left[\widehat{\theta}_R\right] = E\left[\frac{\sum\limits_{i,D=1}\left(Y_i - \overline{Y}_{D_i=1}\right)\left(E_i - \overline{E}_{D_i=1}\right)}{\sum\limits_{i,D=1}\left(E_i - \overline{E}_{D_i=1}\right)^2} - \frac{\sum\limits_{i,D=0}\left(Y_i - \overline{Y}_{D_i=0}\right)\left(E_i - \overline{E}_{D_i=0}\right)}{\sum\limits_{i,D=0}\left(E_i - \overline{E}_{D_i=0}\right)^2}\right]$$

where $\nu_{R,i} = \frac{\widetilde{\beta}\gamma_Y + \gamma_R}{1-\widetilde{\alpha}\widetilde{\beta}} C_i + \frac{\widetilde{\epsilon}_{R,i} + \widetilde{\beta}\widetilde{\epsilon_{Y,i}}}{1-\widetilde{\alpha}\widetilde{\beta}}$, $\sum\limits_{i,D=1} \frac{\nu_{R,i}}{N_1} = \frac{\widetilde{\beta}\gamma_Y + \gamma_R}{1-\widetilde{\alpha}\widetilde{\beta}} + \sum\limits_{i,D=1} \frac{\widetilde{\epsilon}_{R,i} + \widetilde{\beta}\widetilde{\epsilon_{Y,i}}}{1-\widetilde{\alpha}\widetilde{\beta}} \frac{1}{N_1}$*, and* $\sum\limits_{i,D=0} \frac{\nu_{R,i}}{N_1} = \sum\limits_{i,D=0} \frac{\widetilde{\epsilon}_{R,i} + \widetilde{\beta}\widetilde{\epsilon_{Y,i}}}{1-\widetilde{\alpha}\widetilde{\beta}} \frac{1}{N-N_1}$*. By construction,*

$$E\left[\left(\frac{\widetilde{\epsilon}_{R,i} + \widetilde{\beta}\widetilde{\epsilon_{Y,i}}}{1-\widetilde{\alpha}\widetilde{\beta}} - \sum_{i,D=1} \frac{\frac{\widetilde{\epsilon}_{R,i} + \widetilde{\beta}\widetilde{\epsilon_{Y,i}}}{1-\widetilde{\alpha}\widetilde{\beta}}}{N_1}\right)\left(E_i - \sum_{i,D=1} \frac{E_i}{N_1}\right)\right] = 0$$

$$E\left[\left(\frac{\widetilde{\epsilon}_{R,i} + \widetilde{\beta}\widetilde{\epsilon_{Y,i}}}{1-\widetilde{\alpha}\widetilde{\beta}} - \sum_{i,D=0} \frac{\frac{\widetilde{\epsilon}_{R,i} + \widetilde{\beta}\widetilde{\epsilon_{Y,i}}}{1-\widetilde{\alpha}\widetilde{\beta}}}{N - N_1}\right)\left(E_i - \sum_{i,D=0} \frac{E_i}{N - N_1}\right)\right] = 0$$

Therefore, $E\left[\widehat{\theta}_R\right] = \theta_R$ *holds for any combination of* γ_R *and* γ_Y*. Consequently, it is also true that* $\frac{\partial E\left[\widehat{\theta}_R\right]}{\partial \gamma_R} = \frac{\partial E\left[\widehat{\theta}_R\right]}{\partial \gamma_Y} = 0$*. Next, consider the second-stage estimate of* α, $\widehat{\alpha}$*. This coefficient for the rule of law is part of the solution to the second-stage least square minimization problem*

$$\min_{\widehat{\lambda_Y}, \widehat{\lambda'_Y}, \widehat{\alpha}, \widehat{\eta}_Y} \sum_i \left(Y_i - \left(\widehat{\lambda_Y} + \widehat{\lambda'_Y} C_i + \widehat{\alpha}\overrightarrow{R_i} + \widehat{\eta}_Y E_i\right)\right)^2 \tag{8}$$

Where $\overrightarrow{R_i}$ *is the projection of* R_i *obtained from the first stage. It is important to note that since the colony dummy* $\widehat{\lambda}'_R$ *in the first-stage estimation is biased, it is not true that* $E\left[\overrightarrow{R_i}\right] = E[R_i]$*. This has, however, no consequence for* $\widehat{\alpha}$*, which depends only on with-group variations and covariances.*

The FOCs of the minimization problem (8) yield

$$\widehat{\lambda'_R} = \frac{\sum_{i,D=1}\left(Y_i - \widehat{\alpha}\overrightarrow{R_i} - \widehat{\eta}_Y X_i\right)}{N_1} - \frac{\sum_{i,D=0}\left(Y_i - \widehat{\alpha}\widetilde{R_i} - \widehat{\eta}_Y X_i\right)}{N - N_1}, \tag{9}$$

$$\widehat{\lambda}_R = \frac{\sum_{i,D=0}\left(Y_i - \widehat{\alpha}\widetilde{R_i} - \widehat{\eta}_Y X_i\right)}{N - N_1}, \tag{10}$$

$$0 = \sum_i \overrightarrow{R_i}\left(Y_i - \left(\widehat{\lambda_Y} + \widehat{\lambda'_Y}C_i + \widehat{\alpha}\widetilde{R_i} + \widehat{\eta}_Y E_i\right)\right), \tag{11}$$

$$0 = \sum_i E_i\left(Y_i - \left(\widehat{\lambda_Y} + \widehat{\lambda'_Y}C_i + \widehat{\alpha}\widetilde{R_i} + \widehat{\eta}_Y E_i\right)\right). \tag{12}$$

Define the following average within-group covariances and average within-group variances.

$$\begin{aligned}
\widetilde{Cov}(Y,E) &\equiv (N-N_1)\left(Cov\left(Y,E|\,D=0\right)\right) + N_1\left(Cov\left(Y,E|\,D=1\right)\right)\\
\widetilde{Cov}\left(Y,\overrightarrow{R_i}\right) &\equiv (N-N_1)\left(Cov\left(Y,\overrightarrow{R_i}\middle|\,D=0\right)\right) + N_1\left(Cov\left(Y,\overrightarrow{R_i}\middle|\,D=1\right)\right)\\
\widetilde{Cov}\left(\overrightarrow{R_i},E\right) &\equiv (N-N_1)\,Cov\left(\overrightarrow{R_i},E\middle|\,D=0\right) + N_1\left(Cov\left(\overrightarrow{R_i},E\middle|\,D=1\right)\right)\\
\widetilde{Var}(E) &\equiv (N-N_1)\,Var\left(E|\,D=0\right) + N_1 Var\left(E|\,D=1\right)\\
\widetilde{Var}\left(\overrightarrow{R_i}\right) &\equiv (N-N_1)\,Var\left(\overrightarrow{R_i},E\middle|\,D=0\right) + N_1 Var\left(\overrightarrow{R_i},E\middle|\,D=1\right)
\end{aligned}$$

These variances and covariances equal the standard definitions, except that the across-group differences in the mean between noncolonies and colonies are netted out. For example, the average within-group variance of R_i is equal to the variance of R_i in the entire sample if the mean of R is equal in former colonies and in the noncolonies. With this notation, the point estimate of α equals

$$\widehat{\alpha} = \frac{\widetilde{Var}(E)\,\widetilde{Cov}(Y,R) - \widetilde{Cov}(Y,E)\,\widetilde{Cov}(R,E)}{\widetilde{Var}(E)\,\widetilde{Var}(R) - \left(\widetilde{Cov}(R,E)\right)^2} \tag{13}$$

Due to the presence of the standard small-sample instrumental variable bias, it is not generally true that $E[\widehat{\alpha}] = \alpha$. However, since all of the elements in (13) depend exclusively on the within-group variation, the small sample bias of $\widehat{\alpha}$ is not affected by the endogeneity of colonization; i.e., $\frac{\partial E[\widehat{\alpha}]}{\partial \gamma_R} = \frac{\partial E[\widehat{\alpha}]}{\partial \gamma_Y} = 0$. ■

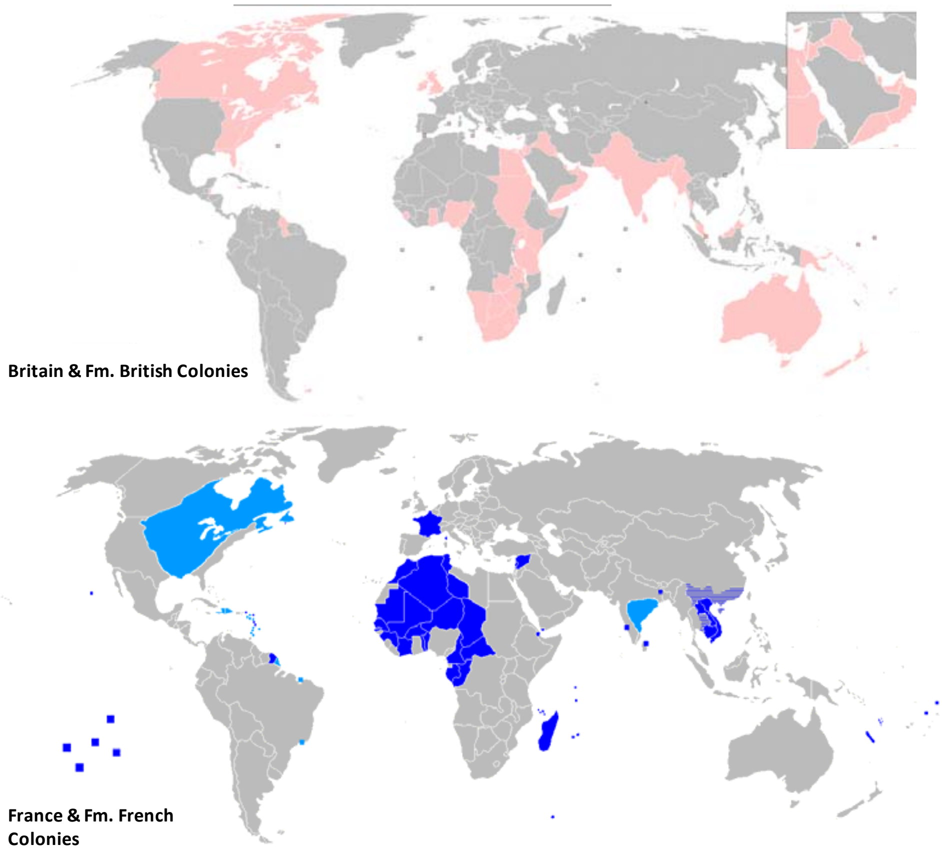

Figure 1

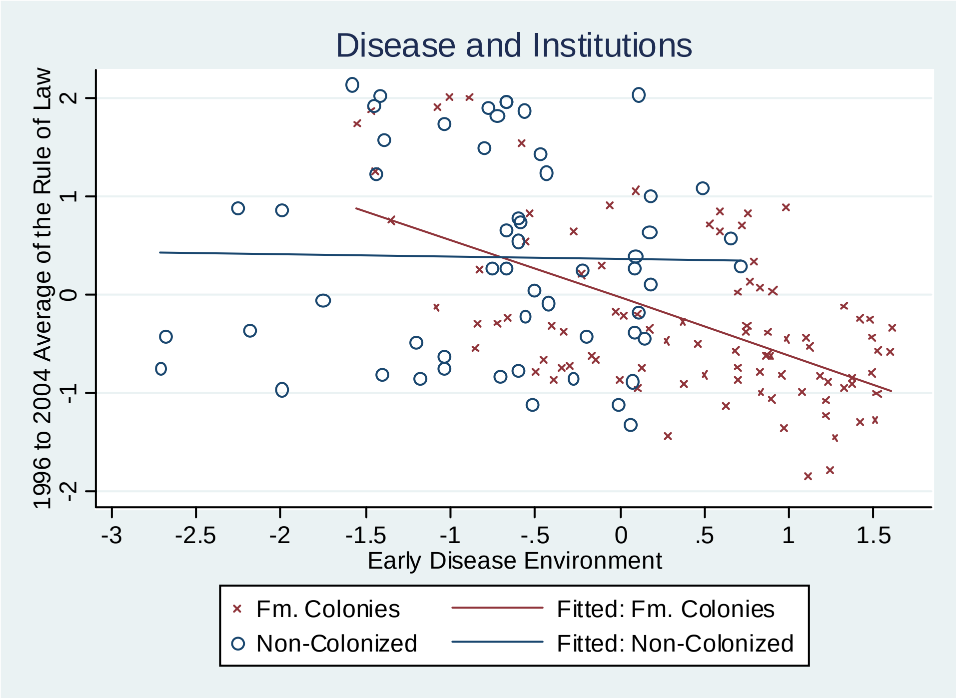

Figure 2

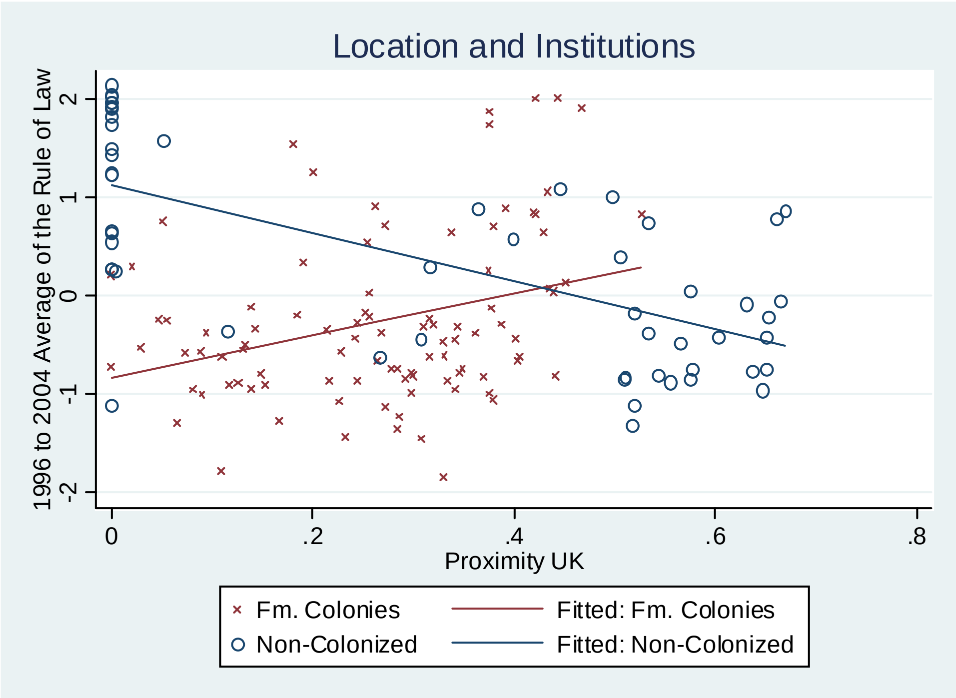

Figure 3

Table 1 - The Geographic Determinants of Settler Mortality Rates and Legal Origin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Description:	*Malaria & Mortality*	*Latitude*	*Rainfall & Temperature*	*Extensive Georg. Model*	*Colonial Org. and Sailing Dist.* *Avg. Sail. Dist.*	*Rel Sailing Dist.*	*UK and FRA Legal Org. and* *Avg. Sail. Dist.*	*Rel Sailing Dist.*	*Colonial Org. and Sailing Dist.* *Avg. Sail. Dist.*	*Rel Sailing Dist.*
Estimation:	(1)- (4) OLS w. Clustered Std. Errors						(5) - (10) Probit Estimation			
Sample:	*(1)- (4): 62 former colonies of Acemoglu et al. (2001)*				*(5)-(6) UK and FRA Colonies*		*UK and FRA LEGOR & Colonies*		*95 former colonies*	
Dependent Variable:	**Ln of the Settler Mortality Rate from Acemoglu et al. (2001)**				***UK Colony Dummy***		***(7)-(10) UK Legal Origin Dummy***			
Malaria Ecology	0.642									
(std. of Ln (ME+1))	[0.105]***									
Latitude		-0.664								
(std.)		[0.184]***								
Avg. Temperature			0.652	0.495						
(std.)			[0.258]**	[0.308]						
Min. of Monthly Rain			-0.322	-0.193						
(std.)			[0.065]***	[0.065]***						
Max. of Monthly Rain			0.217	0.144						
(std.)			[0.091]**	[0.089]						
Temp. at max Humidity			-0.038	-0.507						
(std.)			[0.211]	[0.291]*						
Average Sailing Dist. From					0.886		0.75			
from FRA + UK (log)					[0.277]***		[0.251]***			
Relative Sailing Dist. UK						-3.765		-5.138		
compared to Sail. Dist. FRA						[1.756]**		[2.628]*		
Average Sailing Dist. UK,									0.471	
FRA, DEU, ESP, PORT									[0.219]**	
Relative Sailing Dist. UK										-2.487
comp. to Avg. Sail. Dist. Others										[1.144]**
Terrain Dummies				y						
Sampling Population Dummies				y						
p-value: total model	<0.001	<0.001	<0.001	<0.001	0.0007	0.0175	0.0029	0.0506	0.0310	0.0297
Observations	62	62	62	62	58	58	58	58	95	95
Clusters	35	35	35	35	-	-	-	-	-	-
R-squared (pseudo for probit)	0.47	0.23	0.48	0.72	0.1495	0.0733	0.108	0.0996	0.0351	0.0294

Notes: Table 1 presents the relation between geography and the settler mortality estimates from Acemoglu et al. (2001) in Columns 1 to 5 and the relation between geography and colonial and legal origin (see La Porta et al. (1998) in Columns 5 to 10. Malaria Ecology is from Kiszewski et al. (2004). Other geographic variables are from Parker (1997). Both relative and absolute sailing distances are from Netpas.net and use information on the closest sailing distance that neither passes the Suez or the Panama Canal. All dependent variables except dummies are standardized. The bottom rows report two Wald tests corresponding to the joint null hypothesis that the geographic variables all equal 0 and that the three population dummies all equal 0 (Column 5 only). Heteroskedasticity robust and clustered standard errors in parentheses; **significant at 5%; ***significant at 1%.

Table 2 - Summary Statistics

	Number of Observations	Mean	Standard Deviation	Min Value	Max Value	Number of Observations	Mean	Standard Deviation	Min Value	Max Value
	Former Colonies					***Non-Colonies***				
Log (GDP per Capita 2003)	95	7.066	1.538	4.443	10.472	56	8.302	1.488	5.319	10.556
1996-2004 Avg. of "Rule of Law"	95	-0.264	0.846	- 1.842	2.003	56	0.385	1.026	- 1.316	2.137
Latitude (in Degrees)	95	-0.591	0.661	- 1.589	1.546	56	1.003	0.593	- 0.829	2.199
Malaria Ecology	93	0.426	1.026	- 0.783	2.377	54	-0.734	0.246	- 0.783	1.021
EDE	95	0.418	0.862	- 1.549	2.605	56	-0.710	0.801	- 2.712	0.710
Prob. Legor. UK	95	26.6%	12.6%	0.0%	52.6%	54	31.1%	27.7%	0.0%	67.0%

Notes: Table 2 displays summary statistics of geographic endowments and economic outcomes for the groups of former colonies and non-colonized nations. Rule of Law is from Kaufmann et al. (2005), Malaria Ecology is from Kiszewski (2004), EDE is constructed from Column 4 of Table 1, and "Prob. Legor. UK" is constructed from a model combining the dependent variables of Columns (5) and (6) of Table 1.

Table 3 - Estimating the Partial Effects of Disease and Institutional Quality

	(1)	(2)	(3)	(4)	(5)	(6)		(7)		(8)	(9)
	Sample of AJR (2001) OLS	*Former Colonies OLS*	*Not Colonized OLS*	*Full Sample IV*	*Former Colonies IV*	*Sample of AJR (2001) Plausibly Exogenous IV using Bounds for Direct Effect from*				*Full Sample IV*	
						Weil (2007)		*Bloom and Canning (2000)*		*Adding EDE^2 Term*	*Nonparametric EDE Main Effect*
Panel B: OLS or Second Stage Results - Dependent Variable is the Ln of GDP per Capita in 2003											
Rule of Law				1.624	2.077	0.472	1.42	-0.132	0.911	1.191	2.206
				[0.265]***	[0.233]***	*lower bound*	*upper bound*	*lower bound*	*upper bound*	[0.377]***	[0.272]***
EDE (std.)		-1.174	-0.292	-0.256						0.081	
		[0.133]***	[0.241]	[0.100]**						[0.259]	
Ln Mortality AJR	-1.002					-0.447	-0.071	-0.774	-0.156		
(std.)	[0.142]***					*(assumed bounds)*		*(assumed bounds)*			
EDE^2										-0.098	
										[0.072]	
Non-parametricaly smooted EDE											-1.789
											[1.040]*
Colony y/n				0.108							
				[0.180]							
R-Sq for OLS	0.474	0.433	0.025	-	-	-	-	-			-
Panel A: First Stage Estimation - Dependent Variable is the 96-04 Avg. of "Rule of Law"											
EDE		-0.566	-0.022	-0.022	-0.566	First Stage N/A				-0.145	
		[0.090]**	[0.158]	[0.157]	[0.090]**					[0.324]	
EDE* Colony y/n				-0.543						-0.627	-0.495
				[0.181]**						[0.263]**	[0.101]***
EDE^2										0.035	
										[0.082]	
Non-parametricly smooted EDE											1.722
											[0.646]***
Ln Mortality Acemoglu et al. (2001)											
Colony y/n				-0.397							
				[0.185]*							
Model Information and Hypothesis Tests											
Anderson Canonical Correlation LR Statistic (identification/IV relevance test all instrument)											
P Value:	-	-	-	0.0017	<0.0001	-	-	-	-	<0.001	<0.001
Observations	62	95	56	151	95	62		62		151	151

Notes: Table 3 presents the relation between EDE and institutional quality in Panel A and the relation between EDE and/or instrumented institutional quality and income in Panel B. The measure of early disease environment (EDE) is predicted from Table 1, Column 4. The variable "EDE * Colony y/n" is the interaction of the colony dummy and EDE. "EDE Square" equals (EDE+2.72)^2, where -2.72 is the minimum value of EDE in the sample. The variable "Non-parametrically smooted EDE" is the projection of a nonparametric estimation of EDE on Rule of Law in the entire sample. Heteroskedasticity robust standard errors in parentheses; * significant at 5%; ** significant at 1%

Table 4 - Proximity, Legal Origin, and Prosperity (Two-Stage Least Square Estimations)

	(1)	(2)	(3)	(4)	(5)		(6)	(7)	(8)	(9)
	Former Colonies	*Non-colonies*	*Full Sample*	*Former Colonies*	*Plausibly Exog. IV using bounds from Feyrer (2010 A and B)*		*Full Sample*	*Full Sample*	*Full Sample*	*Full Sample*
Panel B: OLS or Second Stage Results - Dependent Variable is the Ln of GDP per Capita in 2003										
Rule of Law			1.2	1.06	0.835	2.91	1.43	1.29	1.35	1.22
			[0.18]**	[0.42]*	*lower bound*	*upper bound*	[0.17]**	[0.17]**	[0.18]**	[0.20]**
Prob. Col. UK	0.99	-1.67	-0.12		-0.753	-0.157	-0.74			
	[0.62]	[0.34]**	[0.22]		*(assumed bounds)*		[0.27]**			
Prob. Col. FRA							-0.9			
							[0.29]**			
Prob. To Adopt UK Legal System								0.09	0.13	
								[0.27]	[0.28]	
Prob. To Adopt other Legal System									-0.03	
									[0.28]	
Non-parametrically smooted Proximity UK										-0.18
										[0.23]
Panel A: First Stage Estimation - Dependent Variable is the 96-04 Avg. of "Rule of Law"										
Prob. Col. UK	0.93	-1.29	-1.29	0.93	First Stage N/A		-1.22			
	[0.30]**	[0.25]**	[0.25]**	[0.30]**			[0.38]**			
Prob. Col. UK * Colony Y/N			2.22				1.9			
			[0.39]**				[0.52]**			
Prob. Col. FRA							0.08			
							[0.38]			
Prob. Col. FRA * Colony Y/N							-0.64			
							[0.50]			
Prob. To Adopt UK Legal System								-1.63	-1.61	
								[0.26]**	[0.31]**	
Prob. To Adop UK System * Colony y/n								3.05	3.09	2.56
								[0.53]**	[0.55]**	[0.49]**
Prob. To Adopt other Legal System									0.02	
									[0.31]	
Prob. To Adopt. Other System * Colony y/n									1.02	
									[0.77]	
Non-parametrically smooted Proximity UK										1.14
										[0.21]**
Model Information and Hypothesis Tests										
Colony y/n (both stages)	n	n	y	n	n	n	y	y	y	y
Joint Wald Test: Direct Effect of Endowments on Income Equal to 0 (Reduced Form Est. combining first- and second-stage effect)										
P Value:	-	-	0.2655	-	-	-	<0.001	<0.001	<0.001	<0.001
Anderson Canonical Correlation LR Statistic (identification/IV relevance test all instrument)										
P Value:			<0.001	0.003			<0.001	<0.001	<0.001	<0.001
Hansen J Test of Overidentication (all Instruments)										
P Value:	-	-	-	-	-	-	0.0739	-	0.2496	-
Observations	95	54	149	95	95	95	149	149	149	144
R-sq first stage	0.088	0.258	0.248	0.088	na	na	0.26	0.323	0.335	0.292

Notes: Table 4 presents the relation between relative proximity to the colonizers, institutional outcomes, and income. Panel A presents the first-stage estimations relating proximity to institutional outcomes and Panel B the second-stage estimations relating institutional outcomes and proximity to income per capita. In Columns 1-2, the sample includes only former colonies or only non-colonies, and the independent variable in Panel A is the relative proximity to the UK. In Column 1, the sample includes only former colonies and the independent variable in Panel A is the relative proximity to the UK. In Column 3, the first-stage estimation adds the colony dummy and the interaction of this dummy with relative proximity to the UK. The second-stage estimation adds relative proximity to the UK. Column 4 includes only former colonies. Column 5 employs the plausibly exogenous instrumental variable estimator from Conley et al. (2010) uses different measures of proximity (see text). Heteroskedasticity-robust standard errors in parentheses; * significant at 5%; ** significant at 1%.

Table 5 - Robustness Analysis (Two-Stage Least Squares Estimation Results)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Joint Estimaiton	*Excluding Africa*	*Excl. Middle East*	*Excl. South. Europe*	*Excl. AUS CAN, NZL, USA*	*Excl. Warsaw Pact*	*Continent Dummies*	*Ethnic Fract.*	*Human Capital*	*Geogr. Controls*	*Instrumenting for Corpt. Cont.*
Panel B: OLS or Second Stage Estimation - Dependent Variable is the Ln of GDP per Capita in 2003											
Rule of Law 1996 to 2004	**1.36**	**1.06**	**1.04**	**1.43**	**1.44**	**1.29**	**1.2**	**1.27**	**1.16**	**1.25**	
	[0.15]**	**[0.12]****	**[0.18]****	**[0.23]****	**[0.20]****	**[0.14]****	**[0.13]****	**[0.16]****	**[0.17]****	**[0.16]****	
Control of Corruption 1996 to 2004											**1.16**
											[0.17]**
EDE	-0.33	-0.27	-0.55	-0.33	-0.33	-0.37	-0.21	-0.33	-0.23	-0.32	-0.31
	[0.08]**	[0.10]**	[0.11]**	[0.12]**	[0.08]**	[0.09]**	[0.08]*	[0.08]**	[0.08]**	[0.11]**	[0.13]*
Prob. To Adopt UK Legal System	-0.14	-1.02	-0.84	-0.37	0.02	0.02	-0.5	-0.24	-0.76	0.01	0.09
	[0.41]	[0.44]*	[0.42]*	[0.56]	[0.45]	[0.40]	[0.43]	[0.41]	[0.42]	[0.39]	[0.41]
Prob. To Adopt other Legal System	-0.18	-0.39	0.02	-0.35	-0.2	-0.27	-0.31	-0.19	-0.36	-0.1	0
	[0.24]	[0.26]	[0.20]	[0.45]	[0.27]	[0.25]	[0.25]	[0.23]	[0.23]	[0.25]	[0.28]
Ethnic Fractionalization								-0.39			
								[0.30]			
Adult Literacy									0.016		
									[.004]**		
Panel A: First Stage Estimation - in (1) - (8) Dep. Var is the 1996 to 2004 Average of the Rule of Law											**Control of Corruption**
EDE	-0.07	-0.07	-0.25	0.06	-0.07	0.08	0.05	-0.06	-0.03	-0.04	-0.07
	[0.12]	[0.13]	[0.14]	[0.22]	[0.13]	[0.13]	[0.13]	[0.12]	[0.12]	[0.17]	[0.12]
EDE * Colony Y/N	-0.44	-0.63	-0.29	-0.57	-0.3	-0.59	-0.57	-0.35	-0.34	-0.59	-0.44
	[0.16]**	[0.19]**	[0.17]	[0.24]*	[0.15]	[0.16]**	[0.17]**	[0.16]*	[0.16]*	[0.23]*	[0.16]**
Prob. Legor. UK	-2.46	-2.46	-2.32	-0.73	-2.46	-2.23	-2.5	-2.24	-2.62	-1.68	-2.46
	[0.44]**	[0.45]**	[0.56]**	[2.45]	[0.44]**	[0.46]**	[0.43]**	[0.42]**	[0.41]**	[0.44]**	[0.44]**
Prob. Legor. UK * Colony Y/N	3.74	5.95	3.1	2.02	3.4	3.51	2.92	3.47	3.65	3.67	3.74
	[0.70]**	[0.99]**	[0.80]**	[2.51]	[0.70]**	[0.71]**	[0.77]**	[0.69]**	[0.69]**	[0.68]**	[0.70]**
Prob. Legor. Other	-0.02	-0.02	0.26	1.07	-0.02	0.35	-0.26	0	-0.17	0.02	-0.02
	[0.30]	[0.30]	[0.33]	[1.45]	[0.30]	[0.22]	[0.28]	[0.28]	[0.29]	[0.32]	[0.30]
Prob. Legor. other * Colony Y/N	0.33	0.62	-0.05	-0.76	0.61	-0.04	-0.38	-0.07	0.44	0.44	0.33
	[0.61]	[0.65]	[0.62]	[1.54]	[0.61]	[0.58]	[0.84]	[0.63]	[0.60]	[0.65]	[0.61]
Ethnic Fractionalization								-0.74			
								[0.30]*			
Adult Literacy									0.01		
									[0.00]**		
Panel C - Model Information and Hypothesis Tests											
Colony Dummy	y	y	y	y	y	y	y	y	y	y	y
Continent Dummies							y				
Further Geographic Controls										y	
Joint Wald Test: Direct Effect of Endowments Equal to 0 (first- and second-stage total effect from reduced form estimation)											
P Value	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Anderson Canonical Correlation LR Statistic (identification/IV relevance test all instrument)											
P Value	<0.001	<0.001	<0.001	0.0022	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Hansen J Test of Overidentication (all Instruments)											
P Value	0.1922	0.6544	0.3359	0.1656	0.172	0.128	0.762	0.271	0.346	0.623	0.1536
Observations	149	102	135	139	145	129	149	147	149	140	149
R2 First Stage	0.456	0.436	0.503	0.457	0.441	0.551	0.493	0.484	0.477	0.564	0.456

Notes: Table 5 presents robustness tests for two-stage least-squares relation between institutions, endowments, and income. In the second-stage estimation of Panel B, the dependent variable is the logarithm of 2003 per capita GDP. In Panel A, the dependent variable is a measure of institutional outcomes. In Columns 1 to 10, this measure is equal to the 1996 to 2004 average for the score of the rule of law. The estimation s, in turn, exclude 47 African countries (Column 2), 14 Middle Eastern countries (3), 10 countries in Southern Europe (4), the four neo-Europes (5), and all members of the Warsaw pact except Russia (6). The estimation in Column 7 adds four continent dummies for Africa, Asia, Oceania, and Asia (neither first- nor second- stage coefficients for the dummies are reported). Column 8 adds ethnic fractionalization from Alesina et al. (2004). Column 9 adds Malaria Ecology from Kiszewski et al. (2004) and elevation, a landlocked dummy, distance from the equator, the length of coastline, the percentage of a country's surface that is arable, and the "Total Sum of Minerals" from Parker (1997) to the estimation. Column 9 adds "Literacy Rate in the Adult Population" from Parker (1997) to the estimation. Column 11 repeats the baseline specification using the 1996 to 2004 average for control of corruption from Kaufmann et al. (2005). Control of Corruption is standardized, with higher values associated with more control of corruption. Heteroskedasticity-robust standard errors in brackets; * significant at 5%; ** significant at 1%.