Title	Alternative instruments for institutional quality and the effect of european settlements on economic development
Authors(s)	Silva, Luis Vaz
Publication date	2004-06
Publication information	Silva, Luis Vaz. "Alternative Instruments for Institutional Quality and the Effect of European Settlements on Economic Development" (June, 2004).
Series	UCD Centre for Economic Research Working Paper Series, WP04/19
Publisher	University College Dublin. School of Economics
Item record/more information	http://hdl.handle.net/10197/1297

CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2004

Alternative Instruments for Institutional Quality and the Effect of European Settlements on Economic Development

Luis Vaz Silva, University College Dublin

WP04/19

June 2004

DEPARTMENT OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Alternative instruments for institutional quality and the effect of European settlements on economic development

ABSTRACT

The study of the effect of institutional quality on economic performance has a long tradition in economic development. Considering the problems of reverse causality in the link between economic development and institutions, most recent research employs instrumental variables for the measurement of this effect. The present paper explores the impact of European settlements on economic development. These settlements are explained as a function of climate, disease environment, and availability of land. Here these variables are found to determine to a great extent European settlements in Africa and the Americas. Consequently, these variables are used as instruments for institutional quality and the large effect of institutions on income per capita documented by previous studies is through them confirmed. This study finds evidence that limitations of other instruments are overcome using these variables as instruments.

1. INTRODUCTION

Singapore had in 2001 a GDP per capita (PPP US$) of 22,680 according to the Human Development Report 2001. The same report estimates a GDP per capita for India of 2840 US$. These two countries exhibit at best the very large disparities in economic performance between the worlds richest and the poorest nations.

Nevertheless, some of the developments that resulted in these substantial differences are relatively recent. Just two centuries ago, India was considered the crown jewel of the British Empire, while Singapore was poor and virtually uninhabited. Despite this initial disadvantage, the latter is at present much wealthier than the former. The reasons for these cross-country differences in economic performance have attracted considerable interest in recent years. Researchers are engaged in trying to answer one fundamental question: what allowed some countries to enjoy high growth rates over sustained periods, while (most) other countries remained poor?

There is still little consensus on which the most relevant determinants are, but the literature on this subject tends to be divided into three different groups:

- Geographical factors
- Institutions
- Policies

All three hypotheses are well established in literature on economic development, but some recent literature on economic development (Easterly and Levine, 2002; Rodrik et al., 2002) suggests that institutional quality is the most important determinant and has the primacy over geographical factors and the effect of policies in explaining disparate levels of development and performance. These papers claim that once institutions are controlled for, geographic endowments and the effect of policies have no significant effect on incomes or growth. However, these authors concede that geography may affect income levels through institutions, by determining the countries that had in the past the best prospects for developing high quality institutions (Acemoglu et al., 2001a; Engerman and Sokoloff, 1997).

A problem involved in the use of institutions variables (such as protection against expropriation risk), as a measure of institutional quality, is that of endogeneity and reverse causality. Wealthier countries can afford, and probably prefer to have, better institutions. This problem does not affect most geographical variables, because climate zone, being landlocked, or distance from the coast, do not vary with different levels of economic performance.

The preferred method for addressing problems of endogeneity consists in employing instrumental variables (IV). This requires variables that have no direct effect on the dependent variable but are correlated, either positively or negatively, with the endogenous explanatory variable. It is in this context that the work by Acemoglu, Johnson and Robinson (henceforth AJR, 2001a) becomes particularly relevant, because they use European settler mortality rates (ESM) as an instrument for the institutions variable. Although other instruments have been tested, including the fraction of the population speaking English or other Western European languages (Hall and Jones, 1999), considerable attention has been given to ESM in the recent literature. In at least two papers it was considered superior to alternative instruments available (Rodrik et al., 2002; Dollar and Kraay, 2003).

However, AJR's paper is affected by limitations resulting mostly from the non-existence of reliable data on ESM.

A total of 71 countries are referred to, but direct estimates of settler mortality in the first half of the nineteenth century are available for only 20 countries.[1]

The data for the remaining 51 countries are estimates extrapolated from special groups, e.g. mortality rates of bishops, small samples of soldiers (sometimes measured during outbreaks of yellow fever), or from a different country for which similar mortality rates are assumed.

The following problems result from this lack of direct ESM estimates:

- Omission of many countries because of the unavailability of any plausible ESM estimates. This is particularly a problem for most Southern African countries.

- The construction of estimates for Latin America relies largely on data from Gutierrez (1986) for mortality rates of bishops in Latin America, and Curtin (1989) for direct estimates on mortality rates facing Europeans in the 1860's in Mexico. As a result, more than half of all Latin American observations are very similar (out of 24 countries, two have ESM rates of 68.9, eight countries have 71 and another five countries have 78.1). It is relevant to note that this variable is statistically insignificant when the sample is restricted to Latin America (see panel C in table B1). The three alternative geographical variables are all statistically significant.

- There is evidence of measurement error in the estimates used: According to AJR, the healthiest environment for Europeans in

[1] These data come from Philip Curtin (1989).

South America is found in the Guyanas. Guyana and Suriname have ESM rates which are less than half those of any other South American country. This is surprising because this region has a long history for being hazardous to Europeans. For instance, AJR (2001a) note that the Pilgrims decided to settle in North America instead of Guyana partially for the reason that they already were aware of the high mortality rates in Guyana. The penal colony in French Guyana became famous for the same unfortunate reason. Even today, the fraction of the total population that is of European descent is much lower there than anywhere else in South America (2% for Guyana and 1% for Surinam).

Another example of possible measurement error is the very high estimate of 280 deaths per thousand settlers for Angola. The estimate for French *Soudan* (in North and Central Africa) was used in the absence of direct estimates. However, Angola has a long history of European immigration and, with 8% of the total population being European in 1975, had the second highest fraction of population from European descent in Sub-Saharan Africa (second to South Africa). José C. Curto (1999) finds a death rate of 35.5 per thousand in Luanda for the period 1797 to 1832 (a total of 30 censuses were undertaken there between 1773 and 1844). Throughout this period, the city of Luanda had a sizeable European population, reaching 28.6% of the total population in the census of 1844 (José J. Lopes de Lima, 1846).

The ESM variable suffers from two additional problems not related to poor measurement:

- The model is not entirely convincing in predicting where Europeans settled. One of the premises of the theory is that

European settlements are a function of ESM in the colonies. Nevertheless, the regression results show that ESM accounts for 31% of the variation in the dependent variable European settlements in 1900. This leaves room for most of the variation to be explained by factors other than settler mortality rates.

- One final problem concerns the inclusion of Asian countries in the sample. These countries were never candidates to receive Europeans in significant numbers.
  Clearly, institutional quality in Asian countries is independent of European settlements, and therefore cannot be explained through the channels theorized in AJR's paper, so the inclusion of Asian countries in the base sample is inappropriate. The suspicion is that their inclusion just adds confusion to the estimation and, therefore the results should improve with their exclusion from the base sample. This argument will be examined in section 3.

The objective of the present paper is to review alternative instruments to ESM for institutional quality. For that purpose, European settlements are made a function of climate, availability of land, and disease environment. The argument is that potential land ownership was the most important driving force in European overseas emigration. The critical factors allowing for European settlement were (i) the relative abundance of land (strongly related to local population densities), (ii) a temperate climate (essential for the crops and livestock brought from Europe), and (iii) a benign disease environment.

Other factors, and particularly high incomes in the receiving incomes, also contributed to European emigration. Alan Taylor (2002) documents that over 2/3 of English emigration between 1640 and 1660 were directed to the

West Indies. However, this emigration never resulted in the establishment of European populations in significant numbers because of its temporary nature and lower share of families and females.

These explanations of European settlements will be discussed in section 2. Section 3 discusses estimation results, with section 4 concluding the present paper.

2. THE DATA AND DESCRIPTIVE STATISTICS

Appendix A contains the data on the variables used in this paper. Table A1 shows data on incidence rates from the tropical diseases malaria, yellow fever and dengue. The data on dengue and yellow fever are from The American Geographical Society (1952), whereas the data on malaria are from Pampana and Russell (1955). For all three diseases shown, the numbers correspond to the fraction of the population living in areas with the respective disease. The data for malaria refers to 1946 and the data for the other two diseases refers to 1951, these being the oldest estimates available. The incidence of these pathologies is endogenous, but little progress had been made in their eradication as late as the 1950's. Table A6 shows that dengue fever is only lethal in tropical Asia, therefore it was only included for this region. The last column, which corresponds to the disease environment index used as explanatory variable, is the average of the three diseases.

A value above 0.8 corresponds to high exposure to malaria and yellow fever and is a synonym of high-risk environment for Europeans. All the countries with very high values for this variable are situated either in West and Central Africa, or in the part of tropical South America that corresponds roughly to an axis stretching from the Guyanas to Panama.

It is interesting to note how the values for the disease environment (DE) index drop as the distance to the equator increases. Southern Africa provides a good illustration of this. This index decreases successively from 0.98 in Congo (Kinshasa), to 0.69 in Angola, 0.41 in Namibia and 0.1 in South Africa. The east coast shows a similar trend: 1.00 in Tanzania, 0.5 in Mozambique and 0.1 in South Africa.
In the case of very large countries embracing both tropical and temperate regions, the national average is likely to conceal large regional disparities. This is certainly the case of Brazil, where the value of 0.53 results from the average of high-risk environments in the Amazon Basin and low risk environments in the South and Southeast (where most Europeans settled).

Table A2 includes data on the suitability of climate (CS). This data is available on the website of CID-Harvard, and is used to test the hypothesis that similarity to European climate is a determinant of European settlements. The values correspond to the fraction of land area that is situated within a particular type of the Koeppen-Geiger climate zones. The most desirable is undoubtedly the C-type (mild, humid), which in table A2 corresponds to the first three columns. Cf (the first column) indicates absence of dry season with adequate precipitation throughout the year, Cs represents Mediterranean climate (dry summer) and Cw a dry winter season. The climate zones D (snow, forest) and H (highland) were also included in the eligible climate zones, even if less favourable than C-type. D-type zones are relevant for North America where European settlement was widespread, and H-type zones constitute frequently the only regions in the tropics where Europeans could settle (for example Colombia). Accordingly, highlands were only considered for tropical countries and were not considered eligible outside the tropics (thereby excluding H-type zones in Canada, the USA, Argentina, Chile and Morocco). Considering that highlands are sub optimal for agricultural use when compared with the fertile plains of C-type regions, a specific weight of 0.5 was imposed on them. This still allows highlands to be considered suitable for European settlement, but results in these regions being considered less attractive than C-type areas.

The results are similar to those obtained in table A1. Countries with 0% of their land area situated in temperate regions are either in West or Central Africa, North-eastern South America, Central America or the Caribbean. The highest observations are recorded for North America, southern parts of South America, Australia, and the northern and southern tips of Africa. The only country in this sample with 100% of its landmass within temperate climate zones is Uruguay.

Table A3 presents estimates for the availability of land that European settlers could encounter in the mid-19th century. The data for this variable (LA) is a direct result of the division of total land area considered by FAO to be suitable for agricultural use, by the estimated population around 1850. The first column shows the land area suitable for agriculture in square kilometres, using data from FAO. The data on estimated population in 1850 comes from McEvedy and Jones (1978) and Banks (1976).

The last column shows the number of square kilometres of suitable land per inhabitant in 1850, thus presenting a good estimate for the quantity of land available to European settlers in a particular region. The hypothesis is that Europeans did not settle in significant numbers in already densely populated regions (particularly Asia).

As a result, the countries with most land suitable for agricultural use per capita in 1850 were: Australia (8.12 square kilometres per inhabitant), Botswana (2.77), Namibia (2.48), Argentina (1.71), New Zealand (1.38) and Uruguay (1.16). It is interesting to note that all these countries, except Namibia and Botswana, which register zero in the climate variable, received European settlers in significant numbers. On the other hand, the countries with least land available in 1850 are situated in densely populated regions in tropical Asia, Africa or the Caribbean.

Table A4 provides the data for the institutions index variable. This is a composite indicator that results from three different measures of institutional quality. The use of a composite indicator has the advantage of capturing more elements that might determine overall institutional quality. The first column contains data from the corruption index compiled by Transparency International. The second column adds data from the rule of law index included in the International Country Risk Guide and the third column reports data on a measure of political freedom from Freedom House. This last data was considered relevant, since AJR argue that countries, which received fewer settlers, developed more authoritarian institutions, while settler colonies were more likely to set up democratic institutions. The last column shows the average per country after standardizing the freedom and rule of law data using the standard deviation from the corruption index. Countries for which only two observations exist were included in the base sample, while countries with only one observation were not considered.
The amplitude extends from 2.8 (Zimbabwe) to 9.8 (Canada), the mean is 5.8 but the average for African countries falls to 5.0, while the Latin American average is 6.3. According to this institutions index, the countries with the best institutional quality, apart from already mentioned Canada, are New Zealand (9.6), Australia and Malta (9.5).

The last table in Appendix A (A5) documents the results obtained with factor analysis, a technique that explores the existence of clusters of interrelated variables. It may be used to examine whether a large number of variables can be reduced to a smaller number of composite variables. In this particular case, the intention was to examine the feasibility of reducing the variables used to measure availability of land, climate, and disease environment to only one composite variable, and, if so, what the appropriate weights are.

The following table shows that the variables are statistically related.

Table 2.1: Correlation Matrix

	Climate	Availability Of Land	Disease Environment
Climate		0.171	-0.602
Availability of Land			-0.292

As expected, the relationship between the variable disease environment and the other two variables is negative. Equally unsurprising is that the climate and disease variables have a significantly stronger association between them than with the measure of land abundance for European settlers.

A principal component extraction was selected using the Statistical Package for the Social Sciences (SPSS), and the two resulting hypothetical variables are visible in table A5. The variable CSLADE results from reducing all three variables to one composite, whereas CSDE is the product of only two variables (Climate Suitability and Disease Environment).

CSLADE measures the attractiveness for European settlers of a particular country taking into account the three above-mentioned variables. The highest observations are obtained for Australia (3.64), New Zealand (2.32), Uruguay (2.31) and Argentina (1.98). Equatorial Guinea has the lowest observation with –1.14.

Dropping LA results in the composite variable CSDE. The option for this last variable results from the suspicion that CSLADE might be unsuitable due to insufficient correlation of the variable LA with the other two variables. This problem is overcome with the variable CSDE since the two remaining determinants are highly correlated (coefficient of correlation is

-0.602). The hypothesis that CSDE and LA, as autonomous determinants perform better than CSLADE, is confirmed in the next section using ordinary least-squares estimates.

CSDE varies between a minimum of –1.19 (Equatorial Guinea) and a maximum of 2.39 (Uruguay). Countries in the Gulf of Guinea and West Africa have generally the lowest observations, while the most attractive countries are now Uruguay, New Zealand (2.32), Algeria (1.91) and the USA (1.89).

3. EUROPEAN SETTLEMENTS AND INSTITUTIONS

3.1 Endogeneity of European Settlements

In this section I address one first question: is the fraction of Europeans endogenous or exogenous to income levels? In practical terms, what we try to find out is whether European settlements were significantly affected by the economic performance of the colony or if other reasons determined where Europeans settled. In order to examine this topic, it was decided to conduct a Durbin-Wu-Hausman test of exogeneity of instruments. The basic principle is to use the residuals (ν_i) from the equation

(1) $ES_i = \alpha + \beta_1 CSDE_i + \beta_2 LA_i + \nu_i$,

as an autonomous regressor in the structural equation. Two different structural equations, with and without institutions as explanatory variable, were examined:

(2) $ly99_i = \alpha + \beta_1 ES_i + \beta_2 \nu_i + \mu_i$,

(3) $ly99_i = \alpha + \beta_1 ES_i + \beta_2 INS_i + \beta_3 \nu_i + \mu_i$,

where ES is the fraction of European population and ly99 is the log GDP per capita in 1999.

We test the null hypothesis that the coefficients of the residuals ν_i are not statistically different from zero using a simple t test. If we reject the null hypothesis at a small significance level, we conclude that Europeans is endogenous because ν_i and μ_i are correlated.

In this particular case, the t test statistic for β_2 in equation (3) is –1.21 (p-value of 0.229), while the t test value for β_3 in equation (4) is –1.59 (p-value of 0.116). Thus, we can conclude that there is no strong evidence of endogeneity in the variable European settlement.

Several factors, other than the relative prosperity of the colony, influenced settlements. The hypothesis here discussed is that Europeans settled preferentially in temperate regions (allowing for the cultivation of crops from native Europe) with a favourable disease environment and abundant land for agricultural use. The most prosperous regions outside Europe in the period preceding massive European emigration, for instance India and China (AJR, 2001b), did not receive significant European settlement. Instead, most Europeans preferred to settle in regions with low income and population densities (particularly Australia, North America and southern tip of South America). In accordance with this hypothesis and evidence from the endogeneity test, European settlements will be examined as exogenous source for variation in the institutions variables.

3.2 Ordinary Least-Squares Estimates

In the present section alternative instruments for institutional quality will be examined, considering the non-availability of more reliable estimates for ESM in overseas territories throughout the 19^{th} century. Appendix C reports the results of the ordinary least-squares regressions. The first table in this appendix reviews the effect of institutions on economic performance and the subsequent tables scrutinize potential determinants of European settlements and institutional quality.

Table C1 presents the estimates of regressions with income levels as dependent variables. Ly99 was included as dependent variable, due to some missing observations in the Log GDP per capita in 1995 (Ly95) data used by AJR. Both income levels are calculated in PPP basis at current prices using World Bank data. Although the estimates remain quite similar (as expected), this new data has the benefit of allowing for larger sample sizes. It seems reasonable to assume that eventual small differences result mostly from the extended sample size.

Columns (1) to (5) include estimates with Ly95 as dependent variable, while the last five columns have Ly99 fulfilling the same role. We will focus on columns (6) to (10) only, because the first five regressions are closely related and show the same trends. On the whole, institutions are highly correlated with economic development. They explain more than 50 % of the variation in the dependent variable Ly99, whether average protection against expropriation risk (54%) in column (6) or the institutions index (56%) in column (7) are used as measure of institutions. As pointed out by AJR, these results must be interpreted with care due to problems of reverse causality. It is possible, and even likely, that wealthier countries opt to have better institutions. Moreover, institutions variables might be capturing the effect of other determinants of income levels that are correlated with institutions.

Column (8) shows that the fraction of Europeans in total population also has a strong correlation with economic performance. It is important to point out that the highest fraction attained by Europeans in total population during the 20th century was considered, using AJR data, instead of measuring these fractions in one particular year (the authors suggest 1900 and 1975). Two reasons concur for this approach:

- It is relevant for measuring accurately the feasibility of European settlement in the region. In some cases Europeans settled in significant numbers but left later for reasons other than high disease-related mortality rates (for instance Algeria, Angola or Zimbabwe).
- It is relevant to measure accurately the impact of European settlement on institutions and economic performance. Even if European populations left the ex-colony, it is likely that their presence had some long-lasting effects in the country.

Do these estimates make economic sense? If we take one African country whose institutions rank below the African average of 5.0, for instance Kenya which is credited with 4.0, and estimate its potential GDP per capita if it had institutions ranked among the best in the continent (for example Botswana has 8.1), then we obtain very large differences in income levels. The GDP per capita could increase by 4.1 x 0.46 = 1.89, to a new level of 6.93 + 1.89 = 8.82. In fact, Botswana has a GDP per capita of 8.84. The very small difference between the fitted value and the actual observation suggests that both countries might be situated particularly close to the regression line.

Columns (9) and (10) add the fraction of Europeans as additional independent variable once institutions are controlled for. Column (10) shows that R-squared is increased from 0.56 (without ES) to 0.63 (with ES). Both variables have high t test statistics and the "correct" sign. An F test for the inclusion of the new variable results in the null hypothesis being easily rejected at the 1% significance level (F test statistic of 16.26 against critical value of 7.08). Using the average protection against expropriation risk, as in column (9), produces a marginally better fit, which is certainly due to lower correlation of this measure of institutions with ES (0.34 versus 0.57 for institutions index).

Overall, these results support the idea that Europeans influenced economic performance otherwise than just through better institutions. This might have been the case of a considerable and enduring impact that technology and human capital had on productivity. Diamond (1997), in particular, argues that local populations had a profound knowledge of the local suites of wild plants and animals, and therefore only failed to create large agricultural surpluses due to the inexistence of sufficient plants and animals suitable for domestication. It was only with the age of discoveries that new crops and livestock reached temperate regions of the southern hemisphere, thereby allowing for dramatic increases in local agricultural productivity.
Panel B illustrates estimates for regressions with the same dependent and independent variables, but with smaller samples due to the exclusion of Asian countries. The purpose is to test the hypothesis that the inclusion of Asian countries is inappropriate in estimations involving European populations or their mortality rates, considering that tropical Asia was never a candidate for extensive settlement by Europeans.
Columns (6) and (7) show that there is little difference in the effect of institutions on income levels when Asian countries are excluded. However, the effect of Europeans on GDP per capita, captured in column (8), is considerably enhanced. These results support the assumption that the inclusion of Asian countries in the base sample just adds confusion. The inclusion of Europeans as additional control in column (10) was again tested with an F test and the null hypothesis was once more rejected at the 1% significance level (F test statistic of 24.14 versus critical value of 7.08).
Table C2 examines the determinants of institutions and European settlements. Panel A uses protection against expropriation risk as measure of institutional quality, which is successively regressed on ESM, fraction of Europeans, and the three new variables climate, availability of land and disease environment. All variables are found to be individually significant and have the "correct" sign. The relationships between settler mortality on

one hand, and disease environment on the other hand, and institutions is negative, as anticipated. Nevertheless, the quality of institutions is not explained to a large extent by these explanatory variables. The best fit is obtained in column (2), although even in this case no more than 27% of the variation in the dependent variable is explained by the fraction of Europeans. Columns (6) to (9) use different combinations of the new variables. Column (9) uses all three variables as regressors but this estimate adds little explanatory power to the estimate obtained in column (6) and disease environment is statistically not different from zero.

Panel B exploits the composite institutions index as measure of institutional quality and, in spite of results that are generally similar to those in the previous panel, it is interesting for the significantly higher correlation between the dependent variable and the determinants ESM and fraction of Europeans.

Finally, Panel C looks at possible determinants of European settlements. AJR document that ESM explains 31% of the variation in European settlements. The results in column (1) are similar, with the small differences being certainly due to the different measurement of European settlements already explained in the previous section.

As an alternative, European settlements are modelled as

$$(4) \qquad ES_i = \alpha + \beta_1 CS_i + \beta_2 LA_i + \beta_3 DE_i + \nu_i,$$

where CS_i is the suitability of climate, LA_i the availability of agriculturally suitable land and DE_i the disease environment. Columns (2) to (4) show that these new variables have all individually significant explanatory power. Column (5) shows that between the suitability of climate and the availability of appropriate land it is possible to explain nearly half of the variation in European settlements. However, the variable LA is only at the 10% level statistically significant. The inclusion of all three variables in the last column, increases R-squared to 0.55, while all remain significant (though one of them only at the 10% level), and all have the "correct" sign. The inclusion of DE was tested and the null hypothesis rejected at the 1% significance level (F test statistic of 9.52 versus critical value of 7.08). These results demonstrate that it is possible to by-pass the problems involved in the use of ESM as explanatory variable for the settlement of Europeans because the settlement was a function of the local disease ecology, the suitability of the climate and the availability of agricultural land in quantity and quality. It is worth mentioning that the variables disease environment, and to a lesser extent climate suitability, are statistically highly related to Log European settler mortality.

Table C3 in Appendix C reports the estimates of regressions using the composite variables resulting from factor analysis. As some suspicions about the validity of CSLADE persisted, it was decided to examine CSDE and LA as autonomous independent variables. In general, the use of CSLADE results less satisfactory than CSDE and LA as autonomous regressors due to somewhat higher residuals. This supports the assumption that LA is insufficiently correlated with CS and DE to allow for only one composite variable to summarize efficiently all three determinants. The combined use of CSDE and LA explains 22% of the variation in institutional quality, independently of the measure employed, and 55% of the variation in European settlements. The variables are found to be statistically different from zero and their coefficients have the "correct" sign. Nonetheless, F tests were performed to evaluate the effectiveness of the inclusion of LA in the model. We fail to reject the null hypothesis at the 10% significance level for the regression in column (2) (F test statistic of 2.23 versus critical value of 2.79), we reject it at the 5% significance level for the estimation in column (5) (F test value of 5.92 against critical value of 4.00) and we easily reject the null hypothesis at the 1% significance level for the regression in column (8) (F test statistic of 15.26 versus critical value of 7.08).

This table reconfirms that there is a lot to be explained when institutions are regressed on the new variables, a similar limitation to the one encountered when ESM was used as determinant of institutional quality. However, CSDE and LA result more powerful in predicting European settlements.

Table C4 reports OLS estimates for the determinants of institutions excluding Asian countries. The impact of ESM on institutional quality as measured with the institutions index is –0.81 and larger than the impact on protection against expropriation risk. An even larger improvement is observable in column (7), considering that ESM now accounts for 43% of the variation in European settlements. The relationships between CSDE/LA and the dependent variables are also stronger with the exclusion of Asia from the sample. It is now possible to explain 57% of the variation in European settlements with the composite variable CSDE and the availability of land.

3.3 Instrumental Variable Estimates

The purpose of instrumental variables regressions is to tackle those situations, in which OLS is inconsistent:

1. Measurement error in the explanatory variable
2. Lagged dependent variable model with AR(1) error
3. Simultaneity (or endogeneity)

In this particular case, IV estimation is useful because institutions variables are possibly endogenous. Additionally, IV might be helpful in eliminating (or at least mitigating) eventual bias created by error in the measurement of institutional quality. IV estimates require one or more variables that are exogenous in the structural equation, i.e. have no partial effect on income levels, and must be related, either positively or negatively, to the endogenous explanatory variable. Several possible instruments will be examined in this section.

Appendix D documents the estimates from IV regressions. Table D1 uses the average protection against expropriation risk as a measure of institutions, whereas table D2 has the institutions index fulfilling the same rule. Panel A of table D1 reports 2SLS estimates of the coefficient of institutions and Panel B gives the corresponding first stages. Column (1) reconfirms AJR's finding that the 2SLS estimates are larger than the OLS estimates. The coefficient of average protection against expropriation risk is now 1.03 with a standard error of 0.18 when ESM is used as proxy. The small differences to AJR's results (coefficient of 0.94 and standard error of 0.16) are certainly a consequence of using GDP per capita in 1999, instead of GDP per capita in 1995, allowing for the base sample to be slightly increased from 64 to 66 observations. According to AJR (2001a), the larger coefficients of 2SLS estimates can be explained with the attenuation bias resulting from measurement error in the institution variable being probably more important than reverse causality and omitted variables bias.
Columns (2) to (5) examine the effect of using alternative instruments. Particularly noteworthy is the estimate in column (2) (using CSDE as instrument for the institutions variable) with a coefficient of 1.10. Column (5) shows that the estimate with highest precision is obtained with the fraction of Europeans as instrument, due to a relatively small standard error of 0.16. Overall, these results corroborate that it is possible to replicate the

highly significant estimates attained with ESM as instrument using the alternative instruments here proposed.

Columns (6) to (10) replicate the same regressions as in the first five columns, in a base sample where the observations for Asian countries were dropped. As a result, the coefficient of institutions using ESM as instrument is now higher at 1.12, suggesting once again that estimations using settler mortality benefit from the exclusion of Asian countries. The coefficients obtained with the alternative instruments in columns (7) to (10) record small decreases. This is particularly evident in the last three columns, where the fraction of Europeans and the availability of land are used as instrumental variables. These two variables are especially sensitive to the high population densities in tropical Asia, and therefore the omission of Asian countries renders these variables less relevant.

Table D2 repeats the estimation strategy illustrated in table D1 with the average protection against expropriation risk being now replaced by the institutions index as the institutions variable.

The alternative instruments (with the exception of Europeans) result in larger coefficients for the explanatory variable. The estimated coefficient is highest in columns (2) and (7) with both regressions using CSDE as sole instrument. Contrary to the previous table, the exclusion of Asian countries does not reduce the estimates in columns (7) to (10), which in fact increase a little. This suggests that the relationship between geographical variables (climate, disease) and institutions, as measured by the institutions index, is weaker in tropical Asia. The relatively small sample of Asian countries (10 observations) is more likely to be highly influenced by outliers with very good country rankings in the institutions index (in particular Singapore and Hong Kong).

In general, the results from IV estimation show a large effect of institutions on economic development. The exclusion of Asian countries is beneficial to estimation strategies using settler mortality rates as instrument, in accordance with the channels theorized by AJR. Furthermore, the IV results prove that the alternative instruments here reviewed are capable of satisfactorily replacing ESM as exogenous source of variation in the institutions variable.

4. CONCLUSIONS

In this paper the basic hypothesis advanced by AJR is reviewed and refined. An alternative model designed to overcome the limitations of the ESM variable used in other studies is tested. European settlement is explained as a function of climate, availability of land, and disease environment. The last of these variables is used as a proxy for the mortality rates Europeans could have faced in the mid-19th century, while the climate variable serves to test the hypothesis that Europeans preferred to settle in more familiar climate zones, suitable for the crops and livestock from native Europe. The availability of land variable was included to reflect the assumption that the very high population densities of tropical Asia prevented Europeans from settling in the region.

Bearing in mind the strong relationship between climate and disease environment, factor analysis was used to create the new composite variable CSDE. This variable proofed highly successful in the instrumental variable estimation.

The estimation results show that my model explains more than 50% of the variation in European settlement, and it remains highly significant in sub-samples restricted to countries from Africa or South America, while the ESM variable is much less successful in predicting European settlements and results statistically insignificant when the sample is restricted to Latin America.

The present study did not find strong evidence for endogeneity in the variable European settlements. One hypothesis here discussed is that Europeans avoided the establishment of settlements in the most prosperous overseas regions due to high local population densities. Instead, they preferred regions with benign disease ecology, a familiar climate and abundant land suitable for farming.
The fraction of Europeans was tested as an additional regressor. The results reveal that European settlements affect income levels after controlling for the quality of institutions, suggesting additional channels of causality such as the introduction of new crops and livestock leading to increased agricultural productivity.

The results from instrumental variable estimation reconfirm the large effect of institutions on economic development documented in previous studies. In addition, the IV results prove that the alternative model tested in this study is capable of satisfactorily replacing ESM in explaining the variation in institutions in the 20th century.

Although in this paper new light has been shed on the determinants of European settlements, and the important relationship between these and the quality of institutions confirmed, a lot remains to be explained on the reasons for such diverging institutional quality around the world. One hint for further research in this area, that results directly from this study, is that different regions should be examined individually, considering that the determinants for Africa, South America, and Asia, are possibly not the same.

BIBLIOGRAPHY

Acemoglu, Daron; Johnson, Simon and Robinson, James A. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review*, December 2001a. 91(5), 1369-401.

Acemoglu, Daron; Johnson, Simon and Robinson, James A. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *National Bureau of Economic Research* (Cambridge, MA) Working Paper No. 8460, April 2001b.

American Geographical Society, The. "Distribution of Dengue and Yellow Fever: Atlas of Diseases – Plate 5.", *The Geographic Review*, Vol. 42, No. 4, 1952.

Banks, Arthur S. *Cross-National Time Series 1815-1973.* Inter-University Consortium for Political and Social Research, Ann Arbor, MI, 1976.

Curtin, Philip D. *Death by migration: Europe´s encounter with the tropical world in the 19th Century.* New York: Cambridge University Press, 1989.

Curto, José C. "The Anatomy of a Demographic Explosion: Luanda, 1844-1850." *The International Journal of African Historical Studies*, 1999. 32(2/3), 381-405.

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

Dollar, David and Kraay, Aart. "Institutions, Trade, and Growth." *Journal of Monetary Economics*, January 2003. 50(1), 133-162.

Easterly, William and Levine, Ross. "Tropics, Germs, and Crops: How Endowments Influence Economic Development." *National Bureau of Economic Research* (Cambridge, MA) Working Paper No. 9106, August 2002.

Engerman, S. and Sokoloff, K. "*Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies."* In: Haber, S.H., ed., How Latin America Fell Behind. Stanford, CA: Stanford University Press, 1997. 260-304.

Gutierrez, Hector. "La Mortalité des Eveques Latino-Americains aux XVIIe et XVIIIe Siecles." *Annales de Demographie Historique,* 1986. 83-116.

Hall, Robert and Jones, Chad I. "Why do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics,* February 1999. 114(1), 83-116.

Lopes de Lima, José J. *Ensaios Sobre a Statistica das Possessões Portuguezas.* Lisbon, 1846.

Machiavelli, N., 1519. *Discourses on Livy.* New York: Oxford University Press, 1987.

McEvedy, Colin and Jones, R. *Atlas of World Population History.* New York: Facts on File, 1978.

Montesquieu, C. L., 1750. *The Spirit of Laws.* New York: Cambridge University Press, 1989.

Pampana, E. J. and Russell, P.F. *Malaria: A World Problem*. Geneva: WHO, 1955.

Rodrik, Dani; Subramanian, Arvind and Trebbi, Francesco. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *National Bureau of Economic Research* (Cambridge, MA) Working Paper No. 9305, October 2002.

Smith, Adam, 1776. *Wealth of Nations.* Edinburgh: Adam and Charles Black, 1889.

Taylor, Alan. *American Colonies: The Settling of North America.* New York: Penguin Books, 2002.

Appendix A: Data sources

Table A1 **Data on tropical diseases**

country	Yellow fever	Malaria	Dengue	total	country	Yellow fever	Malaria	Dengue	total
Afghanistan	0	0.85		0.43	Liberia	0.98	1		0.99
Angola	0.38	1		0.69	Libya	0	0.15		0.08
United Arab Emirates	0	0.24		0.12	Sri Lanka	0	1	0.99	0.66
Argentina	0	0.08		0.04	Lesotho	0	0		0.00
Australia	0	0.12		0.06	Morocco	0	0.42		0.21
Burundi	1.01	1		1.01	Madagascar	0	1		0.50
Benin	1	1		1.00	Mexico	0	0.66		0.33
Burkina Faso	1	1		1.00	Mali	0.91	0.49		0.70
Bangladesh	0	1	1	0.67	Myanmar (Burma)	0	1	0.45	0.48
Belize	0	1		0.50	Mongolia	0	0		0.00
Bolivia	0.57	0.73		0.65	Mozambique	0	1		0.50
Brazil	0.07	0.99		0.53	Mauritania	0.01	0.17		0.09
Brunei	0	1		0.50	Malawi	0	1		0.50
Bhutan	0	0.81		0.41	Malaysia	0	1	0.8	0.60
Botswana	0	1		0.50	Namibia	0	0.81		0.41
Central African Republic	0.99	1		1.00	Niger	0.94	0.53		0.74
Canada	0	0		0.00	Nigeria	1	1		1.00
Chile	0	0.09		0.05	Nicaragua	0	0.63		0.32
China	0	0.35		0.18	Nepal	0	0.42		0.21
Ivory Coast	0.98	1		0.99	New Zealand	0	0		0.00
Cameroon	1	1		1.00	Oman	0	0.14		0.07
Congo	0.94	1		0.97	Pakistan	0	0.79	0.63	0.47
Colombia	0.8	0.88		0.84	Panama	0.99	1		1.00
Costa Rica	0	1		0.50	Peru	0.06	0.7		0.38
Cuba	0	1		0.50	Philippines	0	1		0.50
Cyprus	0	0		0.00	Papua New Guinea	0	1		0.50
Djibouti	0.01	1		0.51	Puerto Rico	0	1		0.50
Dominican Republic	0	1		0.50	Korea, Democratic People's Republic of	0	1		0.50
Algeria	0	0.06		0.03	Paraguay	0	1		0.50
Ecuador	0.01	0.5		0.26	Qatar	0	1		0.50
Egypt	0	0.47		0.24	Rwanda	1.01	1		1.01
Eritrea	0.39	1		0.70	Saudi Arabia	0	0.06		0.03
Ethiopia	1	0.84		0.92	Sudan	0.87	0.31		0.59
Gabon	0.99	1		1.00	Senegal	0.99	1		1.00
Ghana	1	1		1.00	Sierra Leone	1	1		1.00
Guinea	0.99	1		1.00	El Salvador	0	1		0.50
Gambia, The	1.01	1		1.01	Somalia	0.7	1		0.85

Guinea-Bissau	0.74	1		0.87	Suriname	0.36	1		0.68
Equatorial Guinea	1.12	1		1.06	Swaziland	0	1		0.50
Greenland	0	0		0.00	Syria	0	0.34		0.17
Guatemala	0	0.57		0.29	Chad	0.97	0.67		0.82
French Guiana	0.97	1		0.99	Togo	0.99	1		1.00
Guyana	0.7	1		0.85	Thailand	0	1		0.50
Honduras	0	0.65		0.33	Trinidad	0	0		0.00
Haiti	0	1		0.50	Tunisia	0	0.76		0.38
Indonesia	0	1	0.86	0.62	Taiwan	0	1		0.50
India	0	0.97	0.98	0.65	Tanzania, United Republic of	1	1		1.00
Iran	0	0.88		0.44	Uganda	1	1		1.00
Iraq	0	0.98		0.49	Uruguay	0	0		0.00
Israel	0	0.98		0.49	United States	0	0.25		0.13
Jamaica	0	1		0.50	Venezuela	0.88	1		0.94
Jordan	0	0.44		0.22	Vietnam	0	1	1	0.67
Japan	0	0.38		0.19	Western Sahara	0	0.44		0.22
Kenya	1.01	1		1.01	Yemen	0	0.13		0.07
Cambodia	0	1		0.50	South Africa	0	0.19		0.10
Korea, Republic of	0	1		0.50	Zaire	0.96	1		0.98
Kuwait	0	0		0.00	Zambia	0.07	1		0.54
Laos	0	1		0.50	Zimbabwe	0	1		0.50
Lebanon	0	0.55		0.28					

Notes: Dengue fever is lethal only in Tropical Asia.

Hong Kong	0	1	0.5	Hong Kong and Mauritius were credited with 0 for
Mauritius	0	0	0	yellow fever due to total absence in their geographical area.
Singapore			0.6	Singapore has the same as Malaysia.
Malta	0	0.75	0.38	Malta was credited with exactly the same as Italy and Tunisia (both have the same numbers).

Source Column C — Pampana, E. J., and P. F. Russell. 1955. *Malaria: A World Problem*. WHO, Geneva. Page 4.

Source Column B — Distribution of Dengue and Yellow Fever: Atlas of Diseases – Plate 5." The American Geographical Society, The Geographic Review, Vol. 42, No. 4, 1952.

Source Column D — Distribution of Dengue and Yellow Fever: Atlas of Diseases – Plate 5." The American Geographical Society, The Geographic Review, Vol. 42, No. 4, 1952.

Appendix A: Data sources

Table A2 **Data on suitability of climate**

country	wbcode	cultccf	cultccs	cultccw	cultcdf	cultcdw	cultch		C Type	H Type/2	Total
Angola	AGO	0	0	0.36	0	0	0		0.36	0	0.36
Argentina	ARG	0.64	0	0.07	0	0	0.07		0.71	0	0.71
Australia	AUS	0.35	0.16	0.00	0	0	0		0.52	0	0.52
Burundi	BDI	0	0	0	0	0	0.03		0	0.02	0.02
Benin	BEN	0	0	0	0	0	0		0	0	0
Burkina Faso	BFA	0	0	0	0	0	0		0	0	0
Bangladesh	BGD	0	0	0.09	0	0	0		0.09	0	0.09
Belize	BLZ	0	0	0	0	0	0		0	0	0
Bolivia	BOL	0	0	0.17	0	0	0.36		0.17	0.18	0.35
Brazil	BRA	0.09	0	0.13	0	0	0		0.22	0	0.22
Botswana	BWA	0	0	0	0	0	0		0	0	0
Central African Republic	CAF	0	0	0	0	0	0		0	0	0
Canada	CAN	0.004	0.003	0	0.67	0	0.04		0.68	0	0.68
Chile	CHL	0.23	0.30	0	0	0	0.42		0.53	0	0.53
Ivory Coast	CIV	0	0	0	0	0	0		0	0	0
Cameroon	CMR	0	0	0	0	0	0		0	0	0
Congo	COG	0	0	0	0	0	0		0	0	0
Colombia	COL	0	0	0	0	0	0.23		0	0.12	0.12
Costa Rica	CRI	0	0	0	0	0	0		0	0	0
Djibouti	DJI	0	0	0	0	0	0		0	0	0
Dominican Republic	DOM	0	0	0	0	0	0		0	0	0
Algeria	DZA	0	0.76	0	0	0	0		0.76	0	0.76
Ecuador	ECU	0	0	0	0	0	0.69		0	0.35	0.35
Egypt	EGY	0	0	0	0	0	0		0	0	0
Eritrea	ERI	0	0	0	0	0	0.32		0	0.16	0.16
Ethiopia	ETH	0	0	0	0	0	0.59		0	0.30	0.30
Gabon	GAB	0	0	0	0	0	0		0	0	0
Ghana	GHA	0	0	0	0	0	0		0	0	0
Guinea	GIN	0	0	0	0	0	0		0	0	0
Gambia, The	GMB	0	0	0	0	0	0		0	0	0
Guinea-Bissau	GNB	0	0	0	0	0	0		0	0	0
Equatorial Guinea	GNQ	0	0	0	0	0	0		0	0	0
Guatemala	GTM	0	0	0	0	0	0.44		0	0.22	0.22
French Guiana	GUF	0	0	0	0	0	0		0	0	0
Guyana	GUY	0	0	0	0	0	0		0	0	0

Honduras	HND	0	0	0	0	0	0.45		0	0.23	0.23
Haiti	HTI	0	0	0	0	0	0		0	0	0
Indonesia	IDN	0	0	0	0	0	0.18		0	0.09	0.09
India	IND	0	0	0.27	0	0	0.01		0.27	0	0.27
Jamaica	JAM	0	0	0	0	0	0		0	0	0
Kenya	KEN	0	0	0	0	0	0.25		0	0.13	0.13
Liberia	LBR	0	0	0	0	0	0		0	0	0
Libya	LBY	0	0.32	0	0	0	0		0.32	0	0.32
Sri Lanka	LKA	0	0	0	0	0	0		0	0	0
Lesotho	LSO	0.95	0	0.05	0	0	0		1	0	1
Morocco	MAR	0	0.74	0	0	0	0.08		0.74	0	0.74
Madagascar	MDG	0	0	0.36	0	0	0		0.36	0	0.36
Mexico	MEX	0.02	0	0	0	0	0.59		0.02	0.30	0.32
Mali	MLI	0	0	0	0	0	0		0	0	0
Myanmar	MMR	0	0	0.22	0	0	0.09		0.22	0.04	0.26
Mozambique	MOZ	0.005	0	0.19	0	0	0		0.20	0	0.20
Mauritania	MRT	0	0	0	0	0	0		0	0	0
Malawi	MWI	0	0	0.99	0	0	0		0.99	0	0.99
Malaysia	MYS	0	0	0	0	0	0.14		0	0.07	0.07
Namibia	NAM	0	0	0	0	0	0		0	0	0
Niger	NER	0	0	0	0	0	0		0	0	0
Nigeria	NGA	0	0	0	0	0	0		0	0	0
Nicaragua	NIC	0	0	0	0	0	0.53		0	0.26	0.26
New Zealand	NZL	0.96	0	0	0	0	0.04		0.96	0	0.96
Pakistan	PAK	0	0	0.01	0	0	0.03		0.01	0	0.01
Panama	PAN	0	0	0	0	0	0		0	0	0
Peru	PER	0	0	0.10	0	0	0.38		0.10	0.19	0.29
Paraguay	PRY	0.03	0	0.74	0	0	0		0.77	0	0.77
Rwanda	RWA	0	0	0	0	0	0.24		0	0.12	0.12
Sudan	SDN	0	0	0	0	0	0.04		0	0.02	0.02
Senegal	SEN	0	0	0	0	0	0		0	0	0
Sierra Leone	SLE	0	0	0	0	0	0		0	0	0
El Salvador	SLV	0	0	0	0	0	0.06		0	0.03	0.03
Somalia	SOM	0	0	0	0	0	0		0	0	0
Suriname	SUR	0	0	0	0	0	0		0	0	0
Swaziland	SWZ	0.72	0	0.21	0	0	0		0.93	0	0.93
Chad	TCD	0	0	0	0	0	0		0	0	0
Togo	TGO	0	0	0	0	0	0		0	0	0
Trinidad	TTO	0	0	0	0	0	0		0	0	0
Tunisia	TUN	0	0.62	0	0	0	0		0.62	0	0.62
Tanzania, United Republic of	TZA	0	0	0.14	0	0	0		0.14	0	0.14

Uganda	UGA	0	0	0	0	0	0.03		0	0.02	0.02
Uruguay	URY	1	0	0	0	0	0		1	0	1
United States	USA	0.50	0.03	0	0.31	0	0.02		0.84	0	0.84
Venezuela	VEN	0	0	0	0	0	0.04		0	0.02	0.02
Vietnam	VNM	0	0	0.46	0	0	0		0.46	0	0.46
South Africa	ZAF	0.14	0.05	0.38	0	0	0		0.58	0	0.58
Zaire	ZAR	0	0	0.10	0	0	0.03		0.10	0.02	0.11
Zambia	ZMB	0	0	0.77	0	0	0		0.77	0	0.77
Zimbabwe	ZWE	0	0	0.48	0	0	0		0.48	0	0.48

Climate:

A	Tropical, rainy
B	Dry
C	Mild, humid
D	Snow, forest
E	Polar
H	Highland

Appendix A: Data sources

Table A3 **Data on availability of land**

Country	Agricultural Area	Estimated Population in 1850	Land availability	Country	Agricultural Area	Estimated Population in 1850	Land availability
Angola	574000	2400000	0.24	Liberia	25710	312000	0.08
Argentina	1707950	1000000	1.71	Libya	136450	600000	0.23
Australia	4870170	600000	8.12	Morocco	264250	3000000	0.09
Burkina Faso	83730	1576000	0.05	Malawi	32000	600000	0.05
Bangladesh	97160	23000000	0.00	Madagascar	264400	2000000	0.13
Bahamas	100			Mexico	980590	7750000	0.13
Belize	840			Mali	317500	1998000	0.16
Benin	18270	630000	0.03	Malta	140	130000	0.00
Bolivia	313070	1250000	0.25	Myanmar	103220	8000000	0.01
Brazil	1982200	7250000	0.27	Mauritania	394800	545000	0.72
Barbados	190	140000	0.00	Mauritius	1130	180000	0.01
Botswana	260020	94000	2.77	Malaysia	48250	1000000	0.05
Cape Verde	650	100000	0.01	Mozambique	470800	2250000	0.21
Central African Republic	48600	530000	0.09	Namibia	386530	156000	2.48
Canada	671500	2500000	0.27	Niger	125140	1345000	0.09
Chile	158380	1500000	0.11	Nigeria	699000	12292000	0.06
Cote d'Ivoire	164400	1261000	0.13	Nicaragua	57200	300000	0.19
Cameroon	81600	2057000	0.04	New Zealand	165800	120000	1.38
Congo	224000	6400000	0.04	Pakistan	243810	11000000	0.02
Congo (french)	101540	323000	0.31	Panama	17630	100000	0.18
Colombia	450840	2000000	0.23	Peru	303100	2000000	0.15
Costa Rica	20480	100000	0.20	Paraguay	157660	500000	0.32
Dominican Republic	32970	200000	0.16	Rwanda	15050	1016000	0.01
Algeria	443040	3000000	0.15	Sudan	1100480	5400000	0.20
Ecuador	50350	800000	0.06	Senegal	80500	1009000	0.08
Egypt	28550	5500000	0.01	Sierra Leone	26640	1072000	0.02
Equatorial Guinea	3340	89000	0.04	El Salvador	12880	400000	0.03
Ethiopia	594800	3800000	0.16	Singapore	110	10000	0.01
Gabon	51350	175000	0.29	Suriname	560	94000	0.01
Ghana	117000	1576000	0.07	Chad	479100	1413000	0.34
Guinea-Conakry	118150	1261000	0.09	Togo	28800	504000	0.06
Guinea-Bissau	13600	252000	0.05	Trinidad and Tobago	1250	80000	0.02
The Gambia	5430	126000	0.04	Tunisia	81800	1000000	0.08
Guatemala	28430	900000	0.03	Tanzania	380500	3400000	0.11
Guyana	13760	194000	0.07	Uganda	69800	2700000	0.03
Hong Kong	0		0.00	Uruguay	150560	130000	1.16
Honduras	30900	400000	0.08	USA	4312000	24000000	0.18

Haiti	14550	900000	0.02	Venezuela	201790	1500000	0.13
India	1792400	189000000	0.01	Vietnam	64820	6000000	0.01
Indonesia	383000	16000000	0.02	South Africa	955500	1863000	0.51
Jamaica	5070	400000	0.01	Zambia	349850	600000	0.58
Kenya	252600	2900000	0.09	Zimbabwe	196150	350000	0.56
Sri Lanka	23510	2250000	0.01				

Note: Agricultural Area in square kilometres

In most cases McEvedy and Jones (1978) present direct estimates of the population per country in 1850. However, for seven regions (six in Africa plus the three Guyanas) estimates are only available for a group of countries. For these cases, it was decided to calculate the fraction of the total population per country within the region using 1950 data from the U.S. Bureau of the Census. Assuming that the proportions remained generally unchanged during the previous century, these fractions were extrapolated to the 1850 McEvedy and Jones data for the region, thereby obtaining estimates for each country. Direct estimates for Liberia (312000) are available from Banks (1976), and were therefore preferred to estimates through extrapolation (441000). No direct estimates exist for Singapore in 1850, but McEvedy and Jones (1978) refer a population of 22000 in 1900. Considering that Singapore was uninhabited in 1800, a total population of 10000 was assumed for the city-state.

Source
Column B FAO
Column C McEvedy and Jones, Atlas of World population History
Banks Arthur S., Cross-National Time Series 1815-1973

Appendix A: Data sources

Table A4 **Data for institutions index**

Country	Corruption Index	Rule of law index	Freedom	Total	Country	Corruption Index	Rule of law index	Freedom	Total
Algeria	2.7	61.0	4.5	4.3	Jamaica		69.0	8.8	8.1
Angola	1.7	51.8	4.4	3.3	Kenya	1.9	60.8	4.5	4.0
Argentina	2.8	52.5	7.7	4.8	Liberia	2.7	48.3	4.5	3.5
Australia	8.6	82.5	9.9	9.5	Libya	2.0	68.8	3.3	4.2
Bahamas		76.0	9.9	9.3	Madagascar	1.7	60.3	8.6	5.3
Bangladesh	1.2	60.8	7.6	4.8	Malawi	2.9	59.5	6.7	5.0
Barbados			9.9		Malaysia	4.9	75.8	5.5	6.4
Belize	4.5		9.9	7.2	Mali	3.4	57.8	8.7	5.7
Benin	4.1		7.8	6.0	Malta		77.8	9.9	9.5
Bolivia	2.2	67.0	9.7	6.3	Mauritania	1.7		5.5	3.6
Botswana	6.4	77.8	8.8	8.1	Mauritius	4.5		9.0	6.8
Brazil	4.0	63.8	7.7	6.0	Mexico	3.6	70.5	8.7	6.6
Burkina Faso	2.3	59.5	6.6	4.8	Morocco	3.7	71.3	5.5	5.7
Myanmar	1.8	60.8	3.3	3.6	Mozambique	3.8	59.5	7.6	5.6
Cameroon	2.2	63.3	4.4	4.3	Namibia	5.7	75.5	8.7	7.7
Canada	9.0	85.5	9.9	9.8	New Zealand	9.5	79.5	9.9	9.6
Cape Verde			9.8		Nicaragua	2.5	58.0	7.7	5.1
Central African Republic			5.5		Niger	5.2	59.0	6.6	5.7
Chad			4.4		Nigeria	1.6	52.8	6.5	4.1
Chile	7.5	75.5	8.8	8.3	Pakistan	2.6	57.0	4.5	4.0
Colombia	3.6	61.3	6.6	5.3	Panama	3.0	71.8	9.8	6.9
Congo	2.3	43.5	4.4	3.0	Paraguay	1.7	62.3	6.7	4.8
Congo (french)	1.7	59.8	5.6	4.3	Peru	4.0	69.5	9.7	7.0
Costa Rica	4.5	72.0	9.8	7.4	Rwanda	5.2		3.4	4.3
Côte d'Ivoire	2.7	59.0	5.6	4.5	Senegal	3.1	66.5	7.6	5.8
Dominican Republic	3.5	69.5	9.9	6.9	Sierra Leone	2.6	48.3	6.5	4.1
Ecuador	2.2	59.8	7.7	5.1	Singapore	9.3	91.5	5.5	8.9
Egypt	3.4	68.0	4.4	5.0	South Africa	4.8	67.3	9.8	7.2
El Salvador	3.4	72.0	8.7	6.7	Sri Lanka	3.7	63.8	7.6	5.9
Equatorial Guinea			4.4		Sudan	1.8	54.0	3.3	3.1
Ethiopia	3.5	59.5	5.5	4.8	Surinam		63.0	9.8	8.0
Gabon	2.6	66.3	5.6	5.0	Tanzania	2.7	57.3	6.6	4.8
Gambia		68.0	5.5	6.3	Togo		60.3	5.5	5.6
Ghana	3.9	61.3	8.7	6.1	Trinidad and Tobago	4.9	72.5	7.7	6.9
Guatemala	2.5	67.3	7.6	5.7	Tunisia	4.8	72.3	4.5	5.8
Guinea		61.8	4.5	5.2	Uganda	2.1	62.5	4.5	4.2
Guinea-Bissau	5.2	48.3	6.5	5.0	United States	7.7	79.8	9.9	9.0

Guyana	2.6	64.8	8.8	6.0	Uruguay	5.1	71.8	9.9	7.6
Haiti	2.2	55.8	4.4	3.8	Venezuela	2.5	57.5	7.5	5.0
Honduras	2.7	65.3	7.7	5.7	Vietnam	2.4	69.3	3.4	4.4
Hong Kong	8.2	84.5		9.3	Zambia	2.6	49.0	5.6	3.8
India	2.7	65.3	8.7	6.0	Zimbabwe	2.7	38.5	4.4	2.8
Indonesia	1.9	58.5	7.6	4.9					

Source Freedom 2001-2002 — Freedom House
Source Corruption Index 2002: — Transparency International
Source Rule of law index 2002: — International Country Risk Guide

Appendix A: Data sources

Table A5 **Data from factor analysis**

Country	*CSLADE*	*CSDE*	*Country*	*CSLADE*	*CSDE*
Angola	0.02	0.08	Libya	0.88	1.03
Argentina	1.98	1.80	Morocco	1.28	1.57
Australia	3.64	1.43	Malawi	1.21	1.53
Burkina Faso	-1.05	-1.09	Madagascar	0.28	0.40
Bangladesh	-0.43	-0.38	Mexico	0.47	0.61
Belize		-0.25	Mali	-0.56	-0.59
Benin	-1.05	-1.09	Malta	1.39	1.75
Bolivia	0.07	0.13	Myanmar	0.12	0.25
Brazil	0.06	0.09	Mauritania	0.53	0.44
Botswana	0.54	-0.25	Malaysia	-0.34	-0.29
Central African Republic	-1.03	-1.09	Mozambique	0.06	0.11
Canada	1.56	1.82	Namibia	0.59	-0.10
Chile	1.20	1.46	Niger	-0.64	-0.66
Cote d'Ivoire	-1.01	-1.08	Nigeria	-1.04	-1.09
Cameroon	-1.05	-1.09	Nicaragua	0.41	0.52
Congo	-0.85	-0.86	New Zealand	2.32	2.32
Congo-Brazzaville	-0.92	-1.04	Pakistan	-0.24	-0.18
Colombia	-0.57	-0.61	Panama	-1.01	-1.09
Costa Rica	-0.25	-0.25	Peru	0.36	0.47
Dominican Republic	-0.26	-0.25	Paraguay	0.95	1.14
Algeria	1.59	1.91	Rwanda	-0.89	-0.89
Ecuador	0.60	0.78	Sudan	-0.35	-0.37
Egypt	0.08	0.19	Senegal	-1.04	-1.09
Equatorial Guinea	-1.14	-1.19	Sierra Leone	-1.05	-1.09
Ethiopia	-0.44	-0.42	El Salvador	-0.26	-0.20
Gabon	-0.97	-1.09	Singapore	-0.46	-0.42
Ghana	-1.04	-1.09	Suriname	-0.58	-0.56
Guinea-Conakry	-1.03	-1.09	Chad	-0.69	-0.79
Guinea-Bissau	-0.85	-0.88	Togo	-1.04	-1.09
The Gambia	-1.06	-1.11	Trinidad and Tobago	0.45	0.59
Guatemala	0.35	0.50	Tunisia	0.84	1.07
Guyana	-0.81	-0.84	Tanzania	-0.82	-0.84
Hong Kong	1.21	1.55	Uganda	-1.02	-1.06
Honduras	0.32	0.45	Uruguay	2.31	2.39
Haiti	-0.31	-0.25	USA	1.58	1.89
India	-0.12	-0.02	Venezuela	-0.90	-0.96
Indonesia	-0.35	-0.29	Vietnam	0.14	0.29
Jamaica	-0.31	-0.25	South Africa	1.33	1.47
Kenya	-0.85	-0.88	Zambia	0.98	1.07
Sri Lanka	-0.55	-0.52	Zimbabwe	0.59	0.61
Liberia	-1.02	-1.08			

Appendix A: Data sources

Table A6 **Data on impact of tropical diseases**

Number of Disability-Adjusted Life Years lost in 1990 per 1000 inhabitants	Latin America and Caribbean	Sub-Saharan Africa	China	India	Other Asia and Islands
Malaria	229	13545	29	598	1265
Tropical Cluster Diseases	391	2778	121	1573	298
Dengue		10	15	222	128
Deaths per 1000 inhabitants	Latin America and Caribbean	Sub-Saharan Africa	China	India	Other Asia and Islands
Malaria	7	366		13	39
Tropical Cluster Diseases	10	31	1	18	2
Dengue				7	4

Note: Tropical Cluster Diseases include Trypanosomiasis, Chagas, Schistosomiasis, Leishmaniasis, Lymphatic filariasis and Onchocerciasis.

Source: World Bank

Appendix B: Regression Results by Continent

Table B1 **Determinants of Institutions in Latin America**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A	Dependent variable is Average Protection Against Expropriation Risk 1985-1995							
ESM	-0.16							
	(0.72)							
Climate		1.08						
		(0.58)						
Land availability			0.50					
			(0.55)					
Disease environment				-0.67				
				(0.67)				
Europeans					2.08			
					(1.08)			
R-squared	0.003	0.07	0.03	0.03	0.19			
p-value	0.822	0.074	0.365	0.326	0.068			
Number of observations	23	23	23	23	23			
Panel B	Dependent variable is Institutions Index							
ESM	-0.47							
	(0.72)							
Climate		-0.42						
		(1.23)						
Land availability			-0.40					
			(0.88)					
Disease environment				-0.54				
				(0.88)				
Europeans					0.28			
					(1.54)			
R-squared	0.02	0.01	0.02	0.02	0.003			
p-value	0.522	0.736	0.653	0.545	0.855			
Number of observations	23	24	23	24	24			
Panel C	Dependent variable is European Settlement							
ESM	-0.02							
	(0.14)							
Climate		0.61			0.26	0.48		0.13
		(0.18)			(0.18)	(0.20)		(0.16)
Land availability			0.49		0.37		0.42	0.37
			(0.10)		(0.10)		(0.10)	(0.11)
Disease environment				-0.50		-0.23	-0.27	-0.23
				(0.18)		(0.15)	(0.09)	(0.11)
R-squared	0.001	0.56	0.70	0.37	0.75	0.61	0.79	0.80
Number of observations	24	24	24	24	23	24	23	23

Note: heteroskedastic-consistent standard errors are in parentheses.

Appendix B: Regression Results by Continent

Table B2 **Determinants of Institutions in Africa**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A	Dependent variable is Average Protection Against Expropriation Risk 1985-1995							
ESM	-0.12							
	(0.20)							
Climate		0.85						
		(0.66)						
Land availability			0.53					
			(2.34)					
Disease environment				-0.35				
				(0.60)				
Europeans					5.15			
					(2.04)			
R-squared	0.01	0.03	0.002	0.01	0.05			
p-value	0.555	0.213	0.824	0.564	0.018			
Number of observations	27	27	27	27	27			
Panel B	Dependent variable is Institutions Index							
ESM	-0.15							
	(0.21)							
Climate		-0.28						
		(0.66)						
Land availability			1.00					
			(0.25)					
Disease environment				-0.82				
				(0.62)				
Europeans					7.31			
					(4.39)			
R-squared	0.03	0.004	0.21	0.05	0.12			
p-value	0.497	0.671	0.000	0.194	0.105			
Number of observations	30	39	40	40	38			
Panel C	Dependent variable is European Settlement							
ESM	-0.03							
	(0.01)							
Climate		0.10			0.09	0.07		0.07
		(0.04)			(0.03)	(0.03)		(0.04)
Land availability			0.08		0.06		-0.003	0.03
			(0.07)		(0.06)		(0.07)	(0.06)
Disease environment				-0.09		-0.04	-0.09	-0.03
				(0.03)		(0.03)	(0.03)	(0.03)
R-squared	0.39	0.41	0.07	0.43	0.46	0.47	0.43	0.49
Number of observations	32	40	41	41	40	40	41	40

Note: heteroskedastic-consistent standard errors are in parentheses.

Appendix C: OLS Regression Results

Table C1 **Determinants of income per capita**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A	Dependent variable is log GDP per capita in 1995					Dependent variable is log GDP per capita in 1999				
Average protection against	0.52			0.36		0.54			0.38	
expropriation risk 1985-1995	(0.05)			(0.08)		(0.05)			(0.07)	
Institutions Index		0.47			0.34		0.46			0.33
		(0.04)			(0.07)		(0.04)			(0.06)
Europeans			2.51	1.51	1.19			2.70	1.56	1.4
			(0.25)	(0.32)	(0.38)			(0.24)	(0.29)	(0.36)
R-squared	0.52	0.59	0.46	0.66	0.66	0.54	0.56	0.45	0.67	0.63
Number of observations	64	67	70	64	67	66	79	81	66	77
Panel B	without Asia	without Asia	without Asia	without Asia	without Asia	without Asia	without Asia	without Asia	without Asia	without Asia
	Dependent variable is log GDP per capita in 1995					Dependent variable is log GDP per capita in 1999				
Average protection against	0.51			0.29		0.54			0.3	
expropriation risk 1985-1995	(0.05)			(0.06)		(0.05)			(0.07)	
Institutions Index		0.44			0.24		0.45			0.25
		(0.04)			(0.05)		(0.05)			(0.06)
Europeans			2.64	1.72	1.67			2.85	1.82	1.86
			(0.25)	(0.30)	(0.31)			(0.25)	(0.31)	(0.34)
R-squared	0.54	0.55	0.58	0.71	0.67	0.53	0.52	0.54	0.69	0.64
Number of observations	55	58	61	55	58	56	69	71	56	67

Note: heteroskedastic-consistent standard errors are in parentheses.

Appendix C: OLS Regression Results

Table C2 **Determinants of institutions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A	Dependent Variable is Average Protection Against Expropriation Risk in 1985-1995								
ESM	-0.57								
	(0.16)								
Europeans		2.62							
		(0.60)							
Climate			2.03			1.80	1.53		1.46
			(0.53)			(0.54)	(0.61)		(0.62)
Availability of Land				0.42		0.28		0.28	0.26
				(0.17)		(0.08)		(0.07)	(0.10)
Disease environment					-1.64		-0.69	-1.38	-0.50
					(0.51)		(0.58)	(0.52)	(0.60)
R-squared	0.24	0.27	0.18	0.09	0.14	0.22	0.20	0.17	0.23
Number of observations	66	66	65	65	65	65	65	65	65
Panel B	Dependent Variable is Institutions Index								
ESM	-0.78								
	(0.17)								
Europeans		3.87							
		(0.63)							
Climate			1.70			1.50	0.53		0.64
			(0.74)			(0.76)	(0.96)		(0.98)
Availability of Land				0.54		0.47		0.36	0.37
				(0.12)		(0.12)		(0.09)	(0.10)
Disease environment					-2.18		-1.87	-1.85	-1.44
					(0.54)		(0.74)	(0.55)	(0.74)
R-squared	0.31	0.37	0.09	0.10	0.18	0.18	0.18	0.22	0.22
Number of observations	69	78	78	78	79	77	78	78	77
Panel C	Dependent Variable is Europeans								
ESM	-0.12								
	(0.03)								
Climate		0.53			0.46	0.31		0.30	
		(0.13)			(0.13)	(0.14)		(0.15)	
Availability of Land			0.14		0.11		0.10	0.09	
			(0.11)		(0.06)		(0.07)	(0.05)	
Disease environment				-0.49		-0.33	-0.41	-0.26	
				(0.09)		(0.10)	(0.09)	(0.09)	
R-squared	0.28	0.36	0.24	0.37	0.49	0.46	0.48	0.55	
Number of observations	72	79	80	80	78	79	79	78	

Note: heteroskedastic-consistent standard errors are in parentheses.

Appendix C: OLS Regression Results

Table C3 **Determinants of institutions using factor analysis**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable is Average Protection Against Expropriation Risk in 1985-1995			Dependent Variable is Institutions Index			Dependent Variable is Europeans		
CSDE	0.63	0.56		0.70	0.60		0.18	0.16	
	(0.17)	(0.17)		(0.20)	(0.21)		(0.03)	(0.03)	
LA		0.25			0.39			0.09	
		(0.08)			(0.11)			(0.05)	
CSLADE			0.67			0.79			0.2
			(0.15)			(0.18)			(0.03)
R-squared	0.19	0.22	0.22	0.16	0.22	0.21	0.46	0.55	0.54
Number of observations	65	65	65	78	77	77	79	78	78

Note: heteroskedastic-consistent standard errors are in parentheses.

Appendix C: OLS Regression Results

Table C4 **Determinants of institutions without Asia**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable is Average Protection Against Expropriation Risk in 1985-1995			Dependent Variable is Institutions Index			Dependent Variable is Europeans		
ESM	-0.54			-0.81			-0.16		
	(0.17)			(0.17)			(0.03)		
CSDE		0.65	0.56		0.68	0.57		0.19	0.17
		(0.18)	(0.18)		(0.20)	(0.21)		(0.03)	(0.03)
LA			0.28			0.41			0.08
			(0.06)			(0.11)			(0.04)
R-squared	0.22	0.23	0.27	0.34	0.17	0.24	0.43	0.5	0.57
Number of observations	56	55	55	59	68	67	62	69	68

Note: heteroskedastic-consistent standard errors are in parentheses.

Appendix D: IV Regression Results

Table D1 **Regressions of Log GDP per capita in 1999**

						without Asia	without Asia	without Asia	without Asia	without Asia
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A	Two-Stage Least-Squares									
Average protection against expropriation risk 1985-1995	1.03	1.10	0.98	1.04	0.97	1.12	1.08	0.93	1.00	0.91
	(0.18)	(0.22)	(0.18)	(0.19)	(0.16)	(0.27)	(0.22)	(0.16)	(0.18)	(0.14)
Panel B	First Stage									
ESM	-0.57					-0.54				
	(0.16)					(0.17)				
CSDE		0.63		0.56			0.65		0.56	
		(0.17)		(0.17)			(0.18)		(0.18)	
Climate			1.46					1.7		
			(0.62)					(0.76)		
Land availability			0.26	0.25				0.29	0.28	
			(0.10)	(0.08)				(0.07)	(0.06)	
Disease environment			-0.50					-0.35		
			(0.60)					(0.67)		
Europeans					2.62					3.00
					(0.60)					(0.61)
R-squared	0.24	0.19	0.23	0.22	0.27	0.22	0.23	0.28	0.27	0.39
Number of observations	66	65	65	65	66	56	55	55	55	56

Note: heteroskedastic-consistent standard errors are in parentheses.

Appendix D: IV Regression Results

Table D2 **Regressions of Log GDP per capita in 1999**

	(1)	(2)	(3)	(4)	(5)	without Asia (6)	without Asia (7)	without Asia (8)	without Asia (9)	without Asia (10)
Panel A	Two-Stage Least-Squares									
Institutions Index	0.76	0.92	0.84	0.82	0.70	0.76	0.96	0.85	0.83	0.72
	(0.10)	(0.18)	(0.14)	(0.14)	(0.09)	(0.11)	(0.22)	(0.16)	(0.16)	(0.10)
Panel B	First Stage									
ESM	-0.78					-0.81				
	(0.17)					(0.17)				
CSDE		0.70		0.60			0.68		0.57	
		(0.20)		(0.21)			(0.20)		(0.21)	
Climate			0.64					0.26		
			(0.98)					(1.00)		
Land availability			0.37	0.39				0.38	0.41	
			(0.10)	(0.11)				(0.10)	(0.11)	
Disease environment			-1.44					-1.66		
			(0.74)					(0.75)		
Europeans					3.87					4.01
					(0.63)					(0.63)
R-squared	0.31	0.16	0.22	0.22	0.37	0.34	0.17	0.26	0.24	0.45
Number of observations	69	78	77	77	78	59	68	67	67	68

Note: heteroskedastic-consistent standard errors are in parentheses.