

# Determinants and Economic Consequences of Colonization: A Global Analysis

*Arhan Ertan<sup>+</sup>, Louis Putterman\* and Martin Fiszbein\**

---

## Abstract

Research on economic growth suggests that the era of colonization has had an impact on the levels of economic development of countries around the globe. However, why some countries were colonized early, some late, and others not at all, and what effect these differences have had on current income, has not been studied systematically. In the first part of this paper, we show that both the occurrence and the timing of colonization can be explained by (a) differences in levels of pre-1500 development, (b) proximity to the colonizing powers, (c) disease environment, and (d) latitude. In the second part, we analyze the developmental consequences of colonization while taking the endogeneity of colonization's occurrence and timing into account. Whereas naïve estimates can suggest large impacts, we find that neither the fact nor the timing of colonization affect income today once colonization's impact on the composition of the population and the quality of institutions is controlled for.

---

JEL Classification: O11, O13, O40, O57

Keywords: Colonization, Growth, Institutions, Pre-Modern Development, Migration.

---

<sup>+</sup> New York University

<sup>\*</sup> Brown University. Corresponding author: [Louis\\_Putterman@Brown.Edu](mailto:Louis_Putterman@Brown.Edu). Financial support from the William R. Rhodes Center for International Economics and Finance is gratefully acknowledged. We wish to thank Frank Kleibergen and Philipp Ketz for econometric advice, Dina Akbar, Brian Hanrahan, and Tanya Nguyen for helpful research assistance, and participants at presentations at Brown University, Bogazici University, and City University of New York for their helpful comments.

# **Determinants and Economic Consequences of Colonization: A Global Analysis**

*“The discovery of America and that of a passage to the East Indies by the Cape of Good Hope are the two greatest and most important events recorded in the history of mankind.”*

*Adam Smith, Wealth of Nations (1776)*

*“... the colonization of America, trade with the colonies, the increase in the means of exchange and in commodities generally, gave to commerce, to navigation, to industry an impulse never before known...”*

*Karl Marx and Friedrich Engels,  
Manifesto of the Communist Party (1848)*

## **1. Introduction**

There is little disagreement among historians that the process by which Western European nations set sail into the Atlantic and then Indian Oceans, began the conquest of their islands and coastlines, and eventually came to control vast swaths of territory in the Americas, Africa, Asia and the Pacific, is one of the most important factors that shaped the economic contours of the modern world. The age of colonialism began with the European discoveries of sea routes around Africa’s southern coast (1488) and to the Americas (1492), or perhaps a bit earlier with the settlement of previously uninhabited Atlantic islands like Cape Verde in 1462 (see Landes, 1998). Thereafter, by discovery, conquest, and settlement, the emerging nation-states of Portugal, Spain, the Dutch Republic, France, and England expanded their reach throughout the world, spreading European institutions, culture, and genes, and forcing or inducing massive cross-continental movements of Africans as well others. By the time that the era of colonization

ended in the decades after World War II, the populations of countries in the Americas, Australia, New Zealand, and elsewhere had been radically transformed. New nation-states had been brought into being on four continents—North and South America, Africa, and Australia—with borders bearing no relation to pre-colonial precedents.

Before World War II, two-fifths of the world's land area and a third of its population were in colonies, dependencies, or dominions. A further third, in terms of territory, had been colonized by Europeans some time between the 15<sup>th</sup> and 19<sup>th</sup> centuries and had already emerged as independent nations. In many cases, it was not the once-colonized peoples who became independent, but rather the descendants of the colonizers, so that the process of colonization was never really reversed. In other cases, post-colonial populations were mainly descended from those that the colonizers had imported as slave or indentured laborers, or by admixtures of indigenous, “imported” and colonizing populations. What is called “the Third World” or “the developing world” consists overwhelmingly of ex-colonies, including both ones that underwent dramatic changes in source population (such as those in the Americas) and ones that did not (such as most in Africa, India, and elsewhere in Asia). Fortunately for our purposes, however, not all of the non-European world was colonized, making it possible to try to measure colonization’s effects by studying the colonized and the never colonized side by side.

The impact of the colonial era is directly or indirectly recognized in some of the most influential papers on long run economic growth. La Porta *et al.* (1999) emphasize the importance of the European origins of legal systems. Hall and Jones (1999) attribute large cross-country differences in productivity to differences in “social infrastructure,” instrumented by the proportion speaking European languages. Sokoloff and Engermann (2000) argue that factor endowments were important in explaining long-run economic success in the Americas partly by determining the type of settlers and labor force drawn to different regions. Acemoglu *et al.* (2001, 2002) argue that differing types of institutions dating back to differing modes of European colonization account for much of the cross-country divergence in current incomes. More recent contributions, such as Putterman and Weil (2010) and Ashraf and Galor (2012), have emphasized the impact of

colonization on comparative development through the movement of people and resulting changes in population composition.<sup>1</sup>

However, neither the studies just mentioned nor others that focus more directly on the effects of colonization attempt to explain why some countries were colonized and others not, or why some became European colonies by as early as the 15<sup>th</sup> century and others were colonized only in the late 19<sup>th</sup> or early 20<sup>th</sup> centuries. The present paper shows that the occurrence and timing of colonization can be substantially explained by early history, geography, and disease environments, and that treating colonization as endogenous to these factors and controlling for its impact on source population and on institutions alters our assessment of what colonization's effect was.

Our research begins with the observation that the non-European world was extremely diverse on the eve of Europe's overseas expansion. Regions that would eventually be colonized included places like the American and Canadian plains and Australia which were relatively lightly populated by pre-literate tribes or bands making their living by foraging; other regions like central Mexico and parts of Nigeria that were more densely populated, had organized states, and relied mainly on agriculture; and places like India and Egypt with still denser populations, older civilizations, and access to most of the technologies used by contemporary Europeans. Colonial rule was established in places whose climates and soils were suitable for European-style agriculture and where European mortality and morbidity were similar to those at home, but also in places with tropical climates and with disease environments so daunting that Europeans did not see their interiors until centuries after the conquest of the Americas, often long after establishing outposts on their coasts. The non-European world in the 1490s also contained areas that were as or more densely populated, technologically advanced, politically organized, and literate as Europe itself: countries like Ottoman Turkey, Japan, Korea, and China. It seemed to us perhaps not coincidental that this last group of

---

<sup>1</sup> Putterman and Weil (2010) show that about 10% of the world's people today live in countries where the large majority of the population's ancestors lived in a different world region in 1500, with 11.5% of countries having less than 10% of the population's ancestry being indigenous (present in 1500). Another 80% of the world's people live in countries with 90% or more indigenous ancestry. Ashraf and Galor (2012) emphasize the genetic side of the increased diversity of populations in the U.S., Canada, Australia, New Zealand, and Latin America, while also emphasizing the colonial origins of that diversity. Both the genetic (Ashraf and Galor) and the early developmental (Putterman and Weil) diversity made higher by colonization are associated with significantly higher per capita incomes today.

countries managed to escape European annexation throughout the four plus centuries of Europe's colonial expansion. More generally, the great historical diversity of the non-European world led us to wonder whether both the occurrence and the timing of colonization could be explained by a few of the factors just mentioned, and whether accounting for their impacts might be important to assessing how colonization affected the subsequent economic development of the countries in question.

We posit three main sets of determinants of which countries got colonized and when. They are: (1) an historical determinant: the level of technological and social development of the country on the eve of the colonial era, measured by one or more of (a) its history of state-level polities, (b) the timing of its Neolithic revolution (transition to agriculture), and (c) a composite measure of locally-mastered technology in 1500; (2) geographic determinants: (a) the navigational distance of the country from the colonizing powers, (b) whether reaching the country also involved overland travel and the distance needing to be traveled overland, if so, and (c) latitude; and (3) a biological determinant: the degree to which endemic diseases posed a barrier to European control, proxied by the variable 'malaria ecology.'

Consistent with our conjecture, we find pre-modern development (factor (1) above) to be one of the most important determinants of colonization, one that both decreases the probability of being colonized and delays the date of colonization. Geographic proximity to Europe increases the likelihood and hastens the occurrence of colonization, while distance from the equator (absolute latitude) has the opposite effects. The role played by the disease environment is more complicated: a less favorable disease environment causes colonization to occur later in time, but its positive effect on the probability of being colonized is not significant. Together, pre-modern development, the geographic variables and malaria ecology explain over half of the variation in whether countries were colonized at all and almost half of the variation in the date of colonization for those countries that were colonized.

To investigate the impacts of colonization and of its timing while taking their causes into account, we compare naïve models that include them among exogenous determinants of current income to Two Stage Least Squares models in which they are treated as endogenous. We also estimate Three Stage Least Squares (3SLS) and GMM

models in which both colonization or its timing and the proportion of people descended from European migrants are treated as endogenous, and in which both the latter variables and the quality of institutions are included among the determinants of income today. We find that, without controlling for endogeneity, being colonized has an insignificant negative relationship to current income that becomes statistically significant when other controls are added. Using a parallel approach, being colonized later has a significant negative relationship to current income that becomes insignificant when other controls are added. In our multiple equation models, however, we find that neither colonization nor its timing have direct effects on income, although colonization and its timing may exert indirect effects on income through effects on quality of institutions and European-descended share of population.

European overseas colonization wasn't the first example of colonization in world history. In ancient times, the Assyrian, Persian, Hellenistic, Roman, and other empires conquered large parts of the Near East and the Mediterranean basin. The Mongol conquests of the 13th century, followed by the fall of the Byzantine Empire and the rise of the Ottoman and Mughal empires, among others, reshaped large parts of Eurasia. Russia, China, and the Ottomans all ruled large land-based empires for hundreds of years lasting into the 20<sup>th</sup> Century. But the empires of Portugal, Spain, the Netherlands, France, and Britain, and the more minor overseas efforts of Belgium, Germany, and Italy differed because they involved the conquest of non-contiguous territories, played a key role in the appearance of modern nation-states on every inhabited continent, and may have directly contributed to the emergence of industrial capitalism and the modern world economy. In current usage, the term "colonial era" typically refers to the age of overseas colonization by these European powers, and our study restricts itself to that period and its effects.

While ours is the first paper we are aware of that attempts to explain which countries were colonized and when in a systematic, statistical fashion, it is also one of the few to consider the impact of colonization at the most general level, without differentiating by the identity of the colonizer or qualitative dimensions such as direct or indirect rule, "extractive" or settler colony. It remains for future work to marry more discriminating treatments of the effects of colonization to appropriate recognition of the endogeneity of colonization's occurrence and timing.

The remainder of the paper proceeds as follows: Section 2 briefly discusses relevant literature. Section 3 sets out our hypotheses, empirical strategy, and the data to be used. In Section 4.1, we present the results of the Logit and Ordinary Least Squares (OLS) regressions that predict the occurrence and timing of colonization, respectively. Section 4.2 reports the results of Two-Stage Least Squares (2SLS), Three-Stage Least Squares (3SLS), and GMM models that analyze the effects of being colonized (and of being colonized later rather than earlier) on current level of income. Section 5 concludes by summarizing our main findings.

## **2. Literature**

We know of no previous attempt to use statistical methods to explain, for a large set of countries, which ones were colonized and when. Feyrer and Sacerdote (2009) use patterns of wind speed and direction to instrument for duration of colonization in 81 island nations, but their instruments are somewhat specialized to their sample and leave out factors we find important to predicting colonization and its timing in our broader sample. In contrast to the present paper, which studies 111 countries accounting for 95.4% of the world's population outside of Europe, the countries studied by Feyrer and Sacerdote account for only 1.5% of that population.

A major motivator of our paper is the desire to test the hypothesis that pre-modern economic and technological development is an important determinant of which countries were colonized and when. Diamond (1998) argues that differences in level of technological and social development associated with the timing of agricultural revolutions, as well as the diffusion of technical knowledge across landmasses at similar latitudes versus obstruction of such diffusion by latitudinal differences, deserts, and oceans, are the main factors explaining who was in a position to conquer whom beginning in the 15<sup>th</sup> century. He points out that advanced Eurasian societies, including China, Korea and Japan, Ottoman Turkey, Persia, and Moghul India, enjoyed similar levels of development around 1500, leading to our conjecture that those areas would have been much less likely to be colonized by Europeans before Europe obtained a decisive technological advantage over them. Angus Maddison (2001) finds that the major Eurasian civilizations had almost identical per capita incomes circa 1500. Our focus on the role of pre-modern development in accounting for the occurrence and timing of

colonization explains why it is occurrence and timing rather than duration of colonization to which we direct our attention.

Other major factors we hypothesize to affect the occurrence and timing of colonization include distance from Europe, presence of land barriers, and disease burdens. Landes (1998) discusses the role of geographic proximity and accessibility in the colonization first of islands off of West Africa, then those in the Caribbean, then the parts of the American mainland and stopping points on the ocean route from Western Europe to India, and so forth. The role of malaria and yellow fever in impeding European penetration of Africa and parts of Southeast Asia is also noted by Landes (1998) as well as Acemoglu *et al.* (2001).

Regarding the economic effects of colonization, discussions by social scientists and historians vary from those treating it as fundamentally exploitative and hence detrimental (consider Walter Rodney's *How Europe Underdeveloped Africa*, 1972, or the literature on the detrimental effects of British colonial policy on the textile industry in India, referenced in Clingingsmith and Williamson, 2008) to those viewing colonization as having modernizing effects likely to help the colonized country in the long run. Few studies systematically bring large multi-country data sets and statistical methods to bear on the question, however.

Numerous papers find evidence that who colonized a country and what institutions the colonizer planted there matter for subsequent economic performance (Sala-i-Martin, 1997, La Porta et al., 1997, 1998, Hall and Jones, 1999, Acemoglu, Johnson and Robinson, 2001, 2002, Bertocchi and Canova, 2002, Banerjee and Iyer, 2005, Easterly and Levine, 2012). While suggestive of the topic's importance, the impact of having been colonized per se seems rarely to be studied. One exception is Auer (2011), who finds that the manner in which variables such as rainfall and average elevation affect key institutional determinants of current income differs in ex-colonies and countries never colonized.

As mentioned, the other recent studies that look at the impact of colonization as such (rather than at the differential impacts of different colonizing powers) treat the *duration* of colonial rule rather than the occurrence or timing of colonization as the main



variable of interest.<sup>2</sup> Using a sample of 63 former colonies, Grier (1999) finds a positive and significant correlation between colonial duration and country economic growth during the 1961 – 1990 period. Grier also looks for causal channels and concludes that investments in education explain the positive effect of colonial duration, whereas investments in infrastructure lack comparable explanatory power.

Like Grier, Feyrer and Sacerdote (2009) study the impact of the colonial epoch's duration, but they take income level rather than growth rate as dependent variable and, as noted, they go beyond other studies by endogenizing the duration variable using a sophisticated instrument set focusing on wind speed and direction data. Similar to Grier, they find longer colonial duration to be beneficial, and more so when the duration comes after 1700 than before. This resembles a finding of Olsson (2009) that colonization has a more positive impact on the subsequent likelihood of democratic political institutions when experienced later in time, a finding which he suggests results from the improving of the colonizing powers' own institutions with time. The result might also be related to Grier's finding about education, since relatively modern forms of education also emerged in the colonizing countries only late in the colonial era.

While less focused on the effects of colonization than on its specific consequences, e.g. institutions or legal systems, the lively debate of the past decade and a half on the respective roles of institutions and geography in determining levels of development intersects extensively with the analysis of colonization's impact on income in Section 4.2. Acemoglu, Johnson and Robinson (2001, 2002) use early urbanization and settler mortality differences associated with extractive versus settlement-oriented colonies to attempt to identify decisive effects of institutions favorable to investment and human capital formation. Several authors, including Sachs (2003), criticize Acemoglu *et al.* for downplaying what they view as ongoing causal impacts of geography, climate and disease on income levels (see also Easterly and Levine, 2003). Glaeser *et al.* (2004) question whether it is the institutions Europeans imposed, or rather the attitudes and human capital they brought with them when they settled in colonies and former colonies,

---

<sup>2</sup> A partial exception is Price (2003) who includes a colonial heritage variable in his study of determinants of economic growth rates in 78 non-industrialized countries. However, in addition to focusing on growth (during 1960 – 85) rather than income level, Price defines colonial heritage as having been a colony during the 20<sup>th</sup> century, thus treating Latin American countries as non-colonies.

that account for the greater development of settler than of non-settler colonies. Both Putterman and Weil's (2010) finding of persistence of economic advantage among populations as they moved around the world after 1500, and Easterly and Levine's (2012) finding of a strong link between European settlement during the colonial era and income today, might be interpreted in the spirit of Glaeser *et al.*<sup>3</sup>

### **3. Hypotheses, Data, and Empirical Strategy**

#### *Who was colonized and when?*

Our starting point is the desire to test the conjecture that those non-European regions that were most similar to Europe with respect to technology and political organization in the 15<sup>th</sup> century were the ones least likely to fall to European colonial expansion, and to do so later in the colonial era, if at all. This would have been the case, we think, thanks to defensive capabilities rooted in state-level organization and technology. Societies with coherent states and armies commanding technologies close to the Eurasian technological frontier of the early modern period, for instance the Ottoman, Safavid, Mughal, and Ming empires, were not ones that Europeans could easily dominate in 1500.<sup>4</sup> Some would fall later due in part to a growing technological gap with Europe, in part to waning organizational strength. Quite different were state-level societies lacking steel weapons and other key technologies, like the Incas and Aztecs, as well as stateless societies in other parts of the Americas and Oceania, which were easy for Europeans to dominate given the technological gaps at the time of European contact.

Relative level of development in 1500 can be described by a continuum ranging from groups that relied on hunting and gathering, lacked state-level polities, and exhibited low population densities and absence of cities (Australia, parts of Southeast Asia, southern Africa, and the upper Amazon river basin) to societies subsisting principally on settled agriculture and animal husbandry, manifesting state-level polities, and having higher population densities and cities (much of Europe, north Africa, the Middle East, Iran, South Asia, and East Asia). As indicators of developmental status, we use four previously-studied measures: centuries since transition to reliance on agriculture

---

<sup>3</sup> See also the discussion in Spolaore and Wacziarg (forthcoming).

<sup>4</sup> Indeed, would-be colonizers were in a number of cases expelled by Asian powers. Examples include the reclaiming of Oman and Zanzibar from the Portuguese by the Sultan of Oman in 1650.

(dubbed *ageyears*), as compiled by Putterman and Trainor (2006); the index of state presence, scale, and indigenoussness since 1 CE (dubbed *statehist*) compiled by Putterman (2007);<sup>5</sup> and the composite index of technologies including writing, plough cultivation, firearms and steel in use by the population in 1500, assembled by Comin, Easterly and Gong (2010) (dubbed *tech1500*).

Although societies less technologically advanced than Europe's was in the 15<sup>th</sup> and ensuing centuries were more easily subdued by Europeans, other factors also appear to have influenced who was colonized and when. The preoccupation of the early explorers with the spices of the semi-tropical "indies," especially today's India and Indonesia, and the positions of the navigation routes pioneered by Columbus and Da Gama in the late 15<sup>th</sup> century caused some lands (Hispaniola, the southern tip of Africa, India itself) to be reached and explored earlier than others (Australia, New Guinea, Fiji). For simplicity, we use navigation distance from a port centrally located among those used by the main colonizing powers—Camaret-sur-mer, located at the northwestern tip of France—to control for effective distance from Western Europe at the time of colonization. Distances are calculated using routes appropriate to the era prior to opening of the Suez and Panama canals.

Areas deep in the hinterlands of continents, for instance Afghanistan or Mongolia, were not directly encountered by naval exploration and were far more costly to reach with armed personnel and equipment, given the greater cost of overland travel. Because even some lands having sea outlets tended to be reached by land crossings, we control for being accessed at least in part over land by defining the variable *landroute* as 1 for entirely landlocked countries and as 0.5 for non-landlocked countries that were in practice accessed via land routes.<sup>6</sup> We control both for the categorical distinction of *landroute* and, separately, for the specific distance travelled over land.<sup>7</sup>

---

<sup>5</sup> We consider state history up to 1500 only. Data for two additional countries, Iraq and Saudi Arabia, were developed for use in this paper and are added to the state history data set available at [http://www.econ.brown.edu/fac/Louis\\_Putterman/antiquity%20index.htm](http://www.econ.brown.edu/fac/Louis_Putterman/antiquity%20index.htm).

<sup>6</sup> Specifically, we assign *landroute* a value of 0.5 for several countries of the Americas having Pacific coastlines (El Salvador, Bolivia [which earlier in its history included a small coastline], Ecuador, Peru, Chile) that could be reached from Europe by sailing around the southern tip of South America, but was more often reached by disembarking on the Atlantic coast of Panama and setting sail in other ships on the isthmus's Pacific side. The three other cases in which we set *landroute* at 0.5 are Sudan, which has access to the Red Sea but was reached by prospective European colonizers overwhelmingly by way of the Nile (French forces also approached what is now Sudan from the west); Jordan, also having a Red Sea outlet

That the “scramble for Africa” did not take place until more than three centuries after the colonial powers competed for control of the New World and parts of Asia is widely attributed to the hazards posed by Africa’s disease environment, although other factors may also have been at work. Malaria and yellow fever have also been credited with discouraging European settlement elsewhere, such as New Guinea.<sup>8</sup> When controlling for disease, it is important to avoid reverse causality. Because this is made possible by the now widely used malaria ecology variable, we adopt that to control for the disease environment, although there is a risk of overlooking other relevant diseases.

Finally, we control for latitude due to the possibility that Europeans sought out specific climates either for their own physical comfort or familiarity or due to the kinds of crops that might be grown. While there is abundant evidence that Europeans preferred temperate climates for purposes of settlement, Landes, Sokoloff and Engermann, and others suggest that the initial pattern of conquest was driven more by the potential of land for growing high value crops like sugarcane, as well as by the availabilities of silver (Bolivia) and gold (West Africa). Nevertheless, absolute latitude delivers a number of statistically significant results in our analysis, apparently controlling for factors of considerable importance.

Since there is no obvious way to treat the status of having been colonized and the date of colonization as a single variable, we conduct parallel analyses of the two, using the dichotomous *col* as dependent variable in a set of Logit regressions and  $\text{Ln}(\text{colyr})$  (where *colyr* stands for the year of colonization) as dependent variable in a set of OLS

---

today but overwhelmingly approached overland from the Mediterranean during both its Ottoman and its British colonial periods; and Georgia, in principle accessible through the Mediterranean and Black Seas, but in practice reachable by Europeans only over land during most of the age of colonization, due to Ottoman control of the passage between the two seas. We assign fully landlocked countries like Afghanistan *landroute* values of 1.

<sup>7</sup> Higher coefficients on land than on navigation distance will confirm the expectation that land transportation was more costly, and hence our rationale for not combining land and sea distance in a single measure. Using the more categorical *landroute* in addition to the land distance is motivated by concern for the fixed costs of shifting from sea to land transport and may help to control for differences a land barrier made to the colonizers’ calculations.

<sup>8</sup> Marcus (2009, p. 41) writes, “Malaria ... interfered with European colonization in parts of Southeast Asia. For example, malaria was well established in New Guinea, especially in the lowland areas. It inhibited European settlement there.” New Guinea resembles Africa in that it took centuries after landings on its coasts before Europeans saw areas further inland.

regressions that use the same sets of explanatory variables.<sup>9</sup> Regarding which countries should be treated as having been colonized and when, judgments are unavoidable due to the existence of gray areas such as whether being indirectly ruled or being deemed a protectorate constitutes colonization, and whether the country is a colony as soon as the eventual colonizer has a coastal toehold. We developed our own data for both *col* and *colyr* from various sources. Table A.1 lists the countries in our sample by year of colonization or never-colonized status, as per our determination. The bases of these determinations are explained in our online appendix.<sup>10</sup> Table 1 provides descriptive statistics for all of the variables, and Table A.2 gives brief descriptions and source notes on each variable.

*Do having been colonized, and being colonized early vs. late, affect current income?*

Although what determined the occurrence and timing of colonization is of interest in its own right, it may be of still more interest to economists insofar as understanding it helps to sort out questions about how and whether colonization has lingering effects on incomes today. In Section 4.2, we demonstrate the usefulness of treating colonization and its timing as endogenous for determining their impact on average incomes today in a broad cross-section of countries, doing so by estimating and presenting models of gradually increasing complexity culminating in pairs of 3SLS and GMM models.

In the simplest approach, one could predict, say, *col* using the variables discussed above, then include *col* in a regression for income, with any of the determinants of *col* that can be treated as having no further effect on income serving as instruments. It is helpful for this purpose that the variables used to represent the pre-modern development of the societies in question (*statehist*, *agyears*, and *tech1500*) should in principle no longer apply directly to the countries of today, since it is the pre-modern development levels of the ancestors of those *now* inhabiting each country that are more relevant, following the logic of Putterman and Weil (2010) (see also Comin *et al.*, 2010). Instead, each pre-modern development measure has a “migration adjusted” counterpart—*adjstate*

---

<sup>9</sup> We chose the natural log of *colyr* because we conjectured that factors temporally retarding (hastening) colonization would operate on a slower time scale in earlier than in later years thanks to the accelerating pace of technological change as the colonizing countries industrialized.

<sup>10</sup> Located at [to be supplied at the time of working paper release].

for *statehist*, *adjagyears* for *agyears*, *adjtech1500* for *tech1500*—where *adjstate*, for instance, is the average *statehist* value of the countries of origin of the current population’s ancestors, weighted by the ancestry population shares in Putterman and Weil (2010). The correlations between the adjusted and unadjusted series are weak, and inclusion of the adjusted variables in the second stage regressions helps to assure the unadjusted versions’ excludability from those stages and their ability to serve as instruments.

Although variables such as latitude are unchanging and might prove difficult to exclude from the second stage equation, our distance measures are amenable to a maneuver similar to that just described for the early development indicators. Specifically, we expect colonization and its timing to be determined by the distances corresponding to navigation routes before the completion of the Suez and Panama canals, but current income would be more influenced, if at all, by shipping distances of the post-canal era, which differ from their pre-canal counterparts for many countries. In the second stage regressions, we thus use post-canal navigation distance to Camaret, with its pre-canal counterpart as instrument. Trade and diffusion of ideas today are also influenced by air distances and by proximity to world trading hubs outside of Europe, which were largely irrelevant in the era of colonization. We include the smallest of the air distances of each country to one of three centers of the late 20<sup>th</sup> century global economy (New York, Berlin, and Tokyo), called *min-air*, as an additional geographic determinant of current income. We also use post-canal navigation distance, called *min-nav*, to the closer of New York or Tokyo as an additional control when predicting incomes today.

The remaining question is whether there are other determinants of current income that ought to be accounted for in the regressions discussed above. It seems likely to us on *a priori* grounds, backed also by various recent studies, that one of the main ways that colonization has affected the country incomes of today is by influencing the quality of their institutions (Acemoglu et al., 2001, 2002; Acemoglu and Robinson, 2012), and that one of the main channels for that influence has been the settlement in the colony or ex-colony of substantial numbers of people from Europe (Glaeser et al., 2004; Putterman and Weil, 2010; Easterly and Levine, 2012; Spolaore and Wacziarg, forthcoming). Yet settlement of Europeans is likely to have been influenced by many of the same factors

that influenced the occurrence and timing of colonization. In particular, Europeans preferred settlement in places with temperate climates (mainly in high latitudes), with less hostile disease environments, and perhaps also closer to Europe. On the other hand, early development of indigenous populations is likely to have affected the level of European settlement negatively, since the lands that had hosted old agrarian civilizations, like Egypt and India, lacked open land suitable for European settlement, while areas with less or no agrarian development, like Canada, Argentina, and Australia, had plenty of such space; and regions of intermediate development, like Mexico and Peru, fell in between.

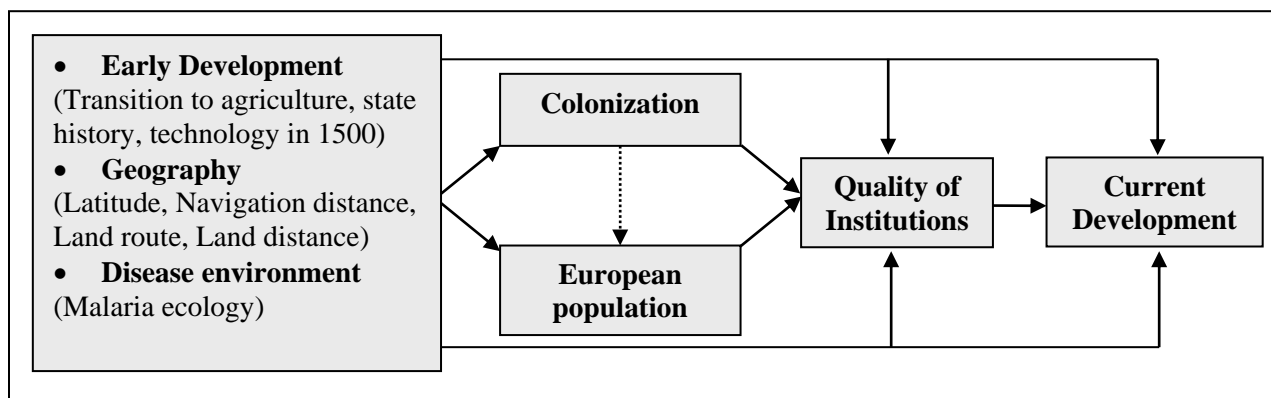
These considerations led us to adopt as a more plausible model of colonization's impact on income, one that treats both colonization (*col*) (or its timing,  $\text{Ln}(\text{colyr})$ ) and European settlement (*eu*) as endogenous determinants of institutional quality (*institutions*), and *col* (or  $\text{Ln}(\text{colyr})$ ), *eu* and *institutions* together as determinants of GDP per capita, or more precisely  $\text{Ln}(\text{gdp p.c.})$ .<sup>11</sup>

A schematic representation of the relationships between the determinants of colonization's occurrence and timing, European settlement, quality of institutions, and current income is shown in Figure 1.

---

<sup>11</sup> Use of the natural log form is standard in the literature, among other reasons because it eliminates dependence of the coefficient estimates on the units in which GDP is measured. Note that whereas Easterly and Levine (2012) use newly assembled data on the European share of population at the roughly half-way point in each country's history as a colony, our *eu* variable, taken from Putterman and Weil (2010), indicates the estimated share of Europeans among the ancestors of each country's year 2000 population. These shares can differ, for instance for countries such as the United States and Argentina which experienced appreciable migration of Europeans after the end of their years as colonies. It is also not clear that Easterly and Levine's sources attempt to attribute fractional European ancestry to "mixed race" populations (mestizo, mulatto), as do Putterman and Weil. Nevertheless, the simple correlation between the EL and PW estimates of European share of population for the 100 countries having data on both variables is a relatively high .7306.

**Figure 1: Determinants and Consequences of Colonization**



The causal links mapped in Figure 1 will be assessed through a series of estimations. To start with, we study the determinants of colonization and its timing through the following estimating regressions:

$$col_i = a_0 + a_1 X_{1i} + u_i \quad (1)$$

$$\ln(col_{yr})_i = b_0 + b_1 X_{1i} + v_i \quad (2)$$

where  $i$  indexes countries and  $X_1$  is the vector of determinants of (occurrence or timing of) colonization. In different specifications,  $X_1$  will include alternative combinations of variables capturing early development, geography and disease environment.

Next, we estimate the effects of colonization on contemporary economic and institutional development. The estimating equations to assess the effects of the occurrence of colonization and the share of Europeans in the current population ( $eu$ ) are the following:

$$\ln(gdp\ p.c.)_i = c_0 + c_1 institutions_i + c_2 col_i + c_3 eu_i + c_4 X_{2i} + \varepsilon_i \quad (3)$$

$$institutions_i = d_0 + d_1 col_i + d_2 eu_i + d_3 X_{2i} + \epsilon_i \quad (4)$$

where  $X_{2i}$  is a vector of controls.  $X_{1i}$  and  $X_{2i}$  have *some* variables in common, but not all – which is crucial in our identification strategy.



OLS estimates of equations (3) and (4) would be biased because of the endogeneity of *col* and *eu*. But with suitable instrumental variables, we can obtain reliable estimates by exploiting exogenous variation in *col* and *eu*. As explained before, the early development variables included in  $X_{1i}$  are replaced in  $X_{2i}$  by their migration-adjusted versions, while pre-canal navigation distance to Camaret is replaced by a set of post-canal navigation distances and air distances that capture the current effects of geographic location on development. The variables included in  $X_{1i}$  that are excluded in  $X_{2i}$  can be used as instruments for *col* and *eu*.

Equations (3) and (4) will be estimated separately by 2SLS and then as a system of two simultaneous equations using 3SLS and GMM. These latter two procedures exploit all of the information to estimate both equations and are thus more efficient than estimating 2SLS equation by equation. In turn, GMM is more robust than 3SLS because it does not assume homoskedasticity.

To assess the effects of the timing of colonization, the same procedure used for the estimation of equations (3) and (4) is repeated with the only difference being that the equations include as regressor  $\text{Ln}(\text{colyr})$  instead of *col*.

We assembled data for all non-European countries with populations of over one-half million for which information on the variables of interest is available. In our sample, we have a total of 111 countries, consisting of 92 non-European countries which were for some time colonized by Western European countries and 19 that were not colonized by those countries (Afghanistan, Armenia, Azerbaijan, China, Georgia, Iran, Japan, Kazakhstan, South Korea, Kyrgyzstan, Liberia, Mongolia, Nepal, Saudi Arabia, Taiwan, Thailand, Turkey, Turkmenistan, and Uzbekistan).

Note that countries in Central Asia and the Caucasus that were colonized by Russia and later incorporated into the Soviet Union are not European colonies in the sense of this paper, and we likewise treat countries that emerged from the Ottoman Empire as if they were not colonies until ruled by France or Britain. Since we cannot rule out the possibility that some of these countries might have been colonized (or colonized earlier) by the European powers had the Ottoman and Russian empires not existed, it seems important to ask whether results obtained for our full sample hold also in a sample of countries that excludes them. We accordingly perform the robustness exercises of

estimating our models on restricted samples: without the non-European countries of the former Soviet Union (FSU); without the “Levant” countries Lebanon, Syria, Israel, Jordan and Iraq, which were among the last countries to pass into Western European hands; and without countries from either group. We also experiment both with dropping and with treating as never colonized Ethiopia, a country colonized exceptionally late (1936) and for an exceptionally brief period (five years). And we experiment both with dropping and with treating as colonized Taiwan, an island that had limited Spanish and Dutch settlements in the mid-1600s before the arrival of most of its current population’s ancestors from mainland China, and that we accordingly consider never colonized in our main analysis.<sup>12</sup> We find qualitatively similar results when these countries are excluded from the analysis or included under alternative assumptions, as noted later.<sup>13</sup>

Although the results generated by our investigations of colonization’s impact on contemporary incomes and institutions are of interest in their own right, we emphasize that we view them as illustrative rather than definitive, since there are undoubtedly other compelling specifications that could be explored. Our primary goal is to demonstrate that the occurrence and timing of colonization are indeed explicable by factors including but—crucially, from an econometric standpoint—not limited to ones that impact contemporary economic performance, and to show that taking their endogeneity into account is important for assessing what past colonization has meant to today’s economies.

## 4. Results

### 4.1. *Determinants of Colonization and Its Timing*

Table 2 reports five Logit regressions predicting whether each of the countries in our sample was ever colonized ( $col = 1$ ) or never colonized ( $col = 0$ ) by a Western European colonial power. Each specification—with one exception—includes malaria ecology, pre-canal navigation distance, *landroute*, land distance, absolute latitude, and at least one measure of pre-modern development. The regressions of columns (1) – (3) each

---

<sup>12</sup> Taiwan’s colonization by Japan during 1895 to 1945 is not pertinent, given our paper’s focus on colonization by Western European powers.

<sup>13</sup> We decided not to drop the North African countries also controlled at times by the Ottoman Empire due to weaker Ottoman control and earlier European colonization than in the Levant.

contain only one pre-modern development indicator, while that of column (4) includes all three of those indicators with the accompanying variables, column (5) includes only pre-modern development indicators, and column (6) only other explanatory variables. The coefficients on latitude and pre-canal navigation distance are always significant at the 1% level, indicating that areas further from Europe and further from the equator were less likely to be colonized. Land distance is significant at either the 10% or the 5% level, depending on the specification, and it obtains coefficients ranging from about eleven to about sixteen times the size of those on navigation distance, a ratio similar to some estimates of the difference in cost of transportation by land versus sea prior to the railroad and the internal combustion engine. The coefficients on malaria ecology and *landroute* are insignificant.

When entered individually in columns (1) – (3), each of the pre-modern development indicators obtains a negative and significant coefficient, supporting our conjecture that having experienced greater pre-modern development tended to ward off colonization. The coefficients for *statehist* and *tech1500* are significant at the 10% level, that for *agyears* at 1%. When all three measures are included, only *agyears* is individually significant. The Pseudo  $R^2$  of 0.564 in column (4) suggests that our explanatory variables explain well over half of the variation of colonization outcomes. With the Pseudo  $R^2$  of column (5) standing at 0.247, it appears that greater pre-modern development alone explains about a quarter of the variation, or a little less than half of what the complete regression explains. Together, however, the remaining variables can explain about 50% of the variation with no control for early development, making their joint importance clearly greater than that of the early development indicators alone.

In regressions shown in tables S.1-S.3 of the supplementary online appendix, we checked whether the same basic results hold for samples that leave out the countries in the late-colonized “Levant” group, those of the former Soviet republics (FSU countries, never overseas colonies of Western Europeans), or both. Results are qualitatively similar in these restricted samples. The explanatory power of the full regression as indicated by Pseudo  $R^2$  reaches 0.635 when the “Levant” countries are dropped, with the early development measures alone achieving a Pseudo  $R^2$  of 0.451. Dropping the FSU countries lowers the full regression’s Pseudo  $R^2$  to 0.498, though the early development

variables alone still explain a similar share of variance as in the full sample. When both “Levant” and FSU are dropped, the full regression’s Pseudo  $R^2$  is 0.561 and that of the early development variables alone is 0.353. Table S.4 shows that if either Ethiopia or Taiwan is dropped from the sample or if Ethiopia is treated as never colonized or Taiwan is treated as colonized in 1624, overall results remain largely the same.<sup>14</sup> In another exercise (see Table S.5), we included only the 81 sample countries located in Asia and Africa (the non-European Old World), also finding qualitatively similar results (with full regression Pseudo  $R^2$  of 0.553.).

Table 3 parallels Table 1 but includes observations only for the 92 countries of our sample that were subjected to Western European colonization at some time between 1463 and 1937. These regressions attempt to explain *when* rather than whether colonization occurred, so OLS is appropriate. Columns (1) – (3) show that when any one of the three early development variables is used with the geography measures, it obtains a positive coefficient that is significant at the 1% level, supporting our expectation that earlier developed areas were colonized later, controlling for other factors. When all three variables are included in column (4), only *tech1500* obtains a statistically significant coefficient, at the 5% level, although it seems worth noting that the *p*-value for joint significance of the three early development variables in column (5) is 0.0001. In all columns, we find significant effects of pre-canal navigation distance and latitude on timing resembling those for occurrence of colonization (thus, each makes colonization both less likely and, if it occurs, later). One variable not significant in explaining colonization’s occurrence does have a significant impact on timing: a more unhealthy malaria ecology is highly significantly associated with later colonization, as anticipated. The coefficients on *landroute* and land distance have the predicted signs but are in all cases statistically insignificant, although the point estimates on land distance are again consistent with much higher overland than sea transportation costs.

Overall, the regressions appear to explain a little under half of the variance in colonization’s timing, with a maximum  $R^2$  of 0.465. The column (5) regression suggests that the pre-modern development variables alone can again explain a little under a quarter

---

<sup>14</sup> Although no individual coefficient on the three early development variables is significant in some of the estimates of Table S.4, the three variables retain their expected signs and are jointly significant.

of the variance, or about half of what the full regression explains, while the other variables alone appear to explain 34.2% of overall variance, or about three-quarters of the variance the full equation explains.

Estimating the regressions of Table 3 for restricted samples again finds similar results. There is a small improvement in  $R^2$  for the full regression—from .465 to .486—when the “Levant” countries are dropped (Table S.6). For the fully specified regression corresponding to column (4), we also check estimates imposing alternative treatments of Ethiopia and Taiwan, and find the results to be qualitatively similar (see Table S.7). (There is no FSU country to be dropped from the colonized countries sample.) In the Old World subsample of 62 colonized countries, the  $R^2$  for the full regression falls to .393, but results remain similar and the three early development measures remain individually significant at the 5% level or better, although as a group explaining only 6.8% of the overall variance (Table S.8).

The general goal of this section has been to test our conjectures that which non-European countries became overseas colonies of Western European powers during the long era of European territorial expansion from the 1460s to the 1930s is predictable on the basis of three main sets of determinants: level of economic and social development in the centuries before 1500, geography as reflected in latitude and proximity to Western Europe (mainly by naval routes), and disease environment. Our results met our expectations well, showing that about half of the variation in having been colonized or not and half of the variation in the date of colonization, for those countries that were colonized, is predicted by these set of factors. Countries that had more advanced levels of development, as measured by an index of their history of state-level polities between 1 and 1500 C.E., by the number of years since first reliance upon agriculture in their territory, and by an index of technology in 1500, were less likely to be colonized at all and, if colonized, were likely to have entered into that status at a later point in time. Countries further from Western Europe by sea using routes available before the Suez and Panama canals, and countries located further from the equator were also less likely to be colonized and were likely to be colonized later if colonized. Countries with adverse disease environments as measured by malaria ecology were neither more nor less likely to be colonized than others, but tend to have been colonized significantly later if they

were colonized. These results are not especially sensitive to whether we include, exclude, or (where applicable) apply alternative treatments to the countries about which doubts about colonization status were mentioned: “Levant” states, the non-European former Soviet republics, Ethiopia, and Taiwan.

## ***4.2 Have Colonization and Its Timing Influenced Current Income?***

### *4.2.1 OLS estimates*

As noted earlier, understanding the determinants of colonization and its timing could be important to economists who want to explain the wealth or poverty of nations, if colonization is a potential determinant of subsequent income and if conclusions about colonization’s impact on current income depend on the proper treatment of colonization as endogenous to geographic and historical factors. Consider first the regressions in tables 4 and 5 that explore the possible impact on current income of *col* and of  $\text{Ln}(\text{colyr})$  without accounting for their endogeneity, but with increasingly inclusive sets of controls. (The choice of control variables is informed by subsequent models that provide guidance on what affects income directly and what factors are important only for their indirect effects.)

The univariate regressions of the first column in each table suggest that having been colonized has no significant impact on current income but that among colonized countries, being colonized later has a highly significant negative effect on income. The addition of the geographic controls *landlocked* and *malaria* in each second column fails to qualitatively alter these conclusions, although each displays a highly significant negative effect in its own right. Column (3) adds a demographic factor that we suggested may be a major channel through which colonization has affected income, the fraction of the ancestors of the present-day population that lived in Europe in 1500 (*eu*). It obtains a significant positive coefficient in both tables and only slightly alters the point estimates on *landlocked* and *malaria*, but its inclusion qualitatively changes the observed effects of both *col* and  $\text{Ln}(\text{colyr})$ , doubling the magnitude of the coefficient on the former and making it significant at the 5% level, while changing the coefficient on the latter from negative and significant to positive and insignificant. The coefficients on *col* and  $\text{Ln}(\text{colyr})$  exhibit no further qualitative changes when our quality of institutions measure

is added in column (4), while *institutions* itself obtains significant positive coefficients and *eu* remains significant and positive as well. The remaining specifications are added for comparison with later IV models, and we defer discussion.

#### 4.2.2 2SLS, 3SLS and GMM models

We can summarize the previous subsection by saying that OLS regressions can either support or fail to support the idea that having been colonized or having been colonized later rather than earlier was harmful to a country's economy, the conclusions being sensitive to what other factors are controlled for. We now turn to employing the insights from Section 4.1 to model *col* and  $\text{Ln}(\text{colyr})$  as endogenous, and we more carefully account for the likely structure of interactions between them and *eu*, *institutions*, and  $\text{Ln}(\text{gdp } p.c.)$ . As discussed in Section 3, our approach involves the estimation of simultaneous equations for  $\text{Ln}(\text{gdp } p.c.)$  and *institutions*, where *institutions* appears as a determinant of  $\text{Ln}(\text{gdp } p.c.)$  but not the converse.<sup>15</sup> The *gdp p.c.* and the *institutions* equations each include both *col* or  $\text{Ln}(\text{colyr})$  and *eu* among their explanatory variables, and in each equation, *col* or  $\text{Ln}(\text{colyr})$  and *eu* are instrumented using variables with no direct impact on  $\text{Ln}(\text{gdp } p.c.)$  and *institutions*. The strategy is aided by the use of related but only partially correlated variables as instruments versus controls—e.g. *statehist* vs. *adjstate*, *pre-canal* versus *post-canal* navigation distance—as discussed in Section 3.

The four right-hand columns of tables 6 and 7 display our complete 3SLS and GMM models. To the left of those columns we include for purposes of comparison in columns (3) and (4) regressions for *gdp p.c* and *institutions* in which *col* (or  $\text{Ln}(\text{colyr})$ ) and *eu* are instrumented but each regression is separately estimated (hence, each column is the second stage of a separate 2SLS equation system). The estimates in columns (1) and (2) provide further comparison by paralleling those in columns (3) and (4) but neglecting the possible impact of *eu* on either *institutions* or *gdp p.c.* and also neglecting the possible impact of *institutions* on *gdp p.c.* (as in column (3) of tables 4 and 5). Each

---

<sup>15</sup> Although causality may well run both from institutional quality to income and from income to institutional quality in the short to medium run, our specifications assume that over the course of the centuries from the colonial era to the present it is unnecessary to account for an income channel in the process by which geographic, historical and demographic factors have influenced the quality of institutions, because quality of institutions is determined mainly by factors other than income. In contrast, we believe that it is a matter of first-order importance that we allow for an institutional quality channel in the process by which those same factors have influenced income.

of the column (1) and (3) second-stage regressions also has an OLS counterpart in Table 4 or 5. All regressions are for the full sample of 111 countries for *col* and 92 countries for  $\text{Ln}(\text{colyr})$ .

Although we had clear prior beliefs regarding which variables could serve as instruments and which could not be excluded from directly affecting *institutions* or  $\text{Ln}(\text{gdp } p.c.)$ , we were careful to seek confirmation of the validity of our instruments by weak instruments tests, in particular the Kleibergen-Paap test, and of their excludability from the second stage regressions as indicated by Overidentifying Restrictions (OIR) tests, specifically the Hansen test. Varying which of the three early development indicators we included as instruments, we found that we require at least two of the three, among which must be *tech1500*, to satisfy the restrictions fully, and that including *statehist* and *tech1500* performed best, among the satisfactory sets. Our tests also delivered one mild surprise: we found that *latitude* cannot be excluded from the second stage regressions for *institutions*, despite our prior conjecture that it would affect *institutions* only through its influence on European settlement.<sup>16</sup> After shifting *latitude* from an instrument to a control role in the regressions for *institutions*, the weak instruments and OIR tests return satisfactory test statistics in the GMM models for both *col* and  $\text{Ln}(\text{colyr})$  and in the 3SLS model for *col*, with OIR test *p*-value acceptable at the 5% but not the 10% level in the 3SLS model for  $\text{Ln}(\text{colyr})$ .

Before further inspecting the full GMM and 3SLS models, we compare the 2SLS estimates of the impacts of *col* and  $\text{Ln}(\text{colyr})$  on current income in tables 6 and 7 to the corresponding OLS estimates in tables 4 and 5. These comparisons constitute the most direct tests of whether treating *col* and  $\text{Ln}(\text{colyr})$  as endogenous alters conclusions about their impacts on  $\text{Ln}(\text{gdp } p.c.)$  relative to conclusions deriving from OLS models. Specification (1) of Table 6 amounts to an IV version of Table 4's column (5), with *col* instrumented and with additional controls *adjstate*, *adjtech1500*, *post-canal*, *min-nav* and *min-air*, but with no controls for the effects of *institutions* and *eu*. The coefficient on *col*

---

<sup>16</sup> A possible explanation may be that settlement by some Europeans (say, from northern countries) raised institutional quality more than that by others (say, southern ones), and that the more institution-benefiting groups were disproportionately drawn to high latitudes. Latitude may remain excludable from the *income* regressions because the factor in question impacts it only through *institutions*, which appears as an endogenous variable there.



has different signs in the IV than in the OLS variant, but it remains far from significant in both regressions.

Adding *eu* and *institutions* to our OLS model made the negative coefficient on *col* highly significant in Table 4, columns (3) and (4). But the controls added in column (6) (and in Table 6) return *col* to insignificance. This is how it remains in that specification's 2SLS counterpart, Table 6's column (3). Treating *col* as endogenous does cause one important difference between the two regression estimates: whereas *eu* has a highly significant positive effect on  $\text{Ln}(gdp\ p.c.)$  in the OLS estimate (Table 4, col. (6)), the corresponding coefficient falls short of significance in the IV counterpart (Table 6, col. (3)).

The corresponding results for  $\text{Ln}(colyr)$  are shown in columns (4) and (5) of Table 5, and columns (1) and (3) of Table 7. When we leave out controls for *eu* and *institutions*, in Table 5's column (5) and Table 7's column (1), we estimate an insignificant negative effect of  $\text{Ln}(colyr)$  on  $\text{Ln}(gdp\ p.c.)$  using OLS, but a highly significant negative effect using 2SLS. So the indication regarding the effect of colonization's timing on *gdp p.c.*, without explicit controls for *eu* and *institutions*, goes from a highly significant negative effect with few controls in columns (1) and (2) of Table 5 to an insignificant effect with added controls in column (5) of that table and back to a highly significant negative effect in the 2SLS version of the regression in Table 7. This is evidence that accounting for the endogeneity of colonization's timing can be crucial to an assessment of whether or not that timing may be a significant determinant of current income.

Once controls for *eu* and *institutions* are added, however,  $\text{Ln}(colyr)$ 's effect on  $\text{Ln}(gdp\ p.c.)$  is insignificant in both OLS (Table 5, column (6)) and 2SLS (Table 7, column (3)) models. As with the corresponding *col* regressions, *eu* is significant in the OLS version but insignificant in the 2SLS counterpart. Taken together, these estimates suggest that among colonized countries, those colonized later have significantly lower incomes, but that the effect works via late colonization's impact on *institutions*, which is in turn influenced by  $\text{Ln}(colyr)$ 's impact on the European share of population.<sup>17</sup>

---

<sup>17</sup> While not explicitly modeling *eu* as a function of *colyr* in the analysis in our paper, we demonstrate the relationship statistically in Table S.9 of the Online Appendix by estimating a regression paralleling that of

Turning now to the full models in the four right-hand columns of Table 6, we find that having been colonized shows no significant direct effect on  $\text{Ln}(gdp\ p.c.)$ , whereas *institutions* shows a significant positive effect on that dependent variable in both the 3SLS and the GMM models. According to columns (6) and (8), however, being colonized has a significant positive effect on *institutions*, so colonization can be considered as having a positive *indirect* effect on *gdp p.c.* through this channel. According to the GMM regressions of Table 6, *eu* has a marginally significant positive effect on *institutions* and a more significant positive effect on  $\text{Ln}(gdp\ p.c.)$ —although neither is significant according to the 3SLS estimates—so *col* can also be seen as indirectly increasing *gdp p.c.* through the channel of European migration. Of the other controls, *landlocked* and *malaria* each have highly significant negative effects on  $\text{Ln}(gdp\ p.c.)$ , consistent with Gallup, Sachs and Mellinger (1999), and *min-air* (being further away from New York, Berlin, or Tokyo) displays a significant negative effect on  $\text{Ln}(gdp\ p.c.)$ <sup>18</sup>

The 3SLS and GMM results for  $\text{Ln}(colyr)$ , in Table 7, resemble those for *col* in that  $\text{Ln}(colyr)$  shows no direct effect on  $\text{Ln}(gdp\ p.c.)$  in the 3SLS and GMM models, with its only significant effect showing up in an *institutions* equation. The effect of  $\text{Ln}(colyr)$  on *institutions* is much weaker than that of *col*, however, being marginally significant in the 3SLS model only. Whether the difference is due to the smaller sample or the different nature of  $\text{Ln}(colyr)$  is unclear. The simpler 2SLS regressions of columns (1) and (2) however, without *eu* and *institutions* as controls, show strong negative effects of lateness of colonization on both *institutions* and  $\text{Ln}(gdp\ p.c.)$ .<sup>19</sup> To check the robustness of our estimates, we also estimated all models of tables 6 and 7 for the restricted samples discussed earlier. Here, we focus on the GMM models, available in tables S.9 and S.10 of

---

the full regression in Table 3 but with dependent variable *eu* and with  $\text{Ln}(colyr)$  added to the set of independent variables. We also estimate a two equation seemingly-unrelated-regressions model that includes both the equation just described and the full Table 3 model for  $\text{Ln}(colyr)$ . In both equations for *eu*,  $\text{Ln}(colyr)$  has a highly significant negative coefficient.  $\text{Ln}(colyr)$  is also a highly significant predictor of *eu* if all other controls are excluded.

<sup>18</sup> A small surprise is that greater post-canal navigation distance from Western Europe has a highly significant positive effect on *institutions*. High scores for countries including Australia, New Zealand and Japan versus low scores for countries in sub-Saharan and North Africa presumably contribute to this result. Note that the main results of the 3SLS and GMM estimates are qualitatively like those of 2SLS in columns (3) and (4). Also, the finding that *col* significantly affects *institutions* but has no direct effect on *gdp p.c.* holds even in the pared down models of columns (1) and (2).

<sup>19</sup> Note the contrast with the findings of Feyrer and Sacerdote (2008) and of Olsson (2009) discussed in Section 2.

the online appendix. The absence of a direct effect of colonization on  $\text{Ln}(gdp\ p.c.)$  is confirmed for all samples, while Table 6's indications that *col* may have an indirect effect on *income* via a significant positive effect on *institutions* become more tenuous. Specifically, the positive sign and magnitude of *col*'s effect on *institutions* hold for all of the restricted samples, but the coefficient loses significance once the FSU countries are dropped. A number of other coefficients show sensitivity to sample composition, probably the most important being that the significant positive effect of *col* on *institutions* disappears when both "Levant" and FSU are dropped from the sample.

Corresponding GMM regression estimates of the  $\text{Ln}(colyr)$  models for restricted samples are shown in Table S.11. The most noteworthy result is that the negative coefficients on  $\text{Ln}(colyr)$  in the  $\text{Ln}(gdp\ p.c.)$  equation becomes significant at the 5% level when the "Levant" countries are dropped. A perspective on these results is that the significant negative impact of late colonization in the initial naïve regressions of Table 5 resurfaces in the GMM models when "Levant" countries are dropped because they are exceptions to a rule that if a country was colonized, it is less damaging for it to have been colonized early, as were the countries of the Americas, than late, like most of sub-Saharan Africa. Arguably, the "Levant" countries, as well as Ethiopia, were colonized by Europeans too briefly for that experience to have had much effect on their economies. The result that the fraction of European-descended population impacts *institutions* continues to hold for all samples, except that the coefficient's significance level falls short of 10% in the Old World only sample.<sup>21</sup>

Finally, we note that the most closely related past studies, as reviewed in Section 2, showed more interest in colonization's duration than in its occurrence or initial timing. We chose, in contrast, to focus on occurrence and timing because we had a set of factors

---

<sup>21</sup> As an additional robustness check we obtained DFBETA statistics for the OLS estimations of the effects of the occurrence and timing of colonization. DFBETA statistics indicate the difference in the estimated coefficient of a given regressor when the  $i^{\text{th}}$  observation is excluded from the sample, scaled by the estimated standard error of the full-sample coefficient. Observations for which the DFBETAs are high in absolute value are considered outliers. A commonly used threshold, suggested by Bollen and Jackman (1990), considers as an outlier an observation for which the absolute value of the DFBETA is higher than one. We considered the effects of *col* and  $\text{Ln}(colyr)$  on both current income and *institutions*, and also explored the effects of *eu* on both these outcomes, as well as the effects of *institutions* on  $\text{Ln}(gdp\ p.c.)$  in the *col* and the *colyr* regression. In all cases, we found no outliers in the sample. There are no observations that change the estimates of interest by more than one standard deviation.

in mind with which to explain them, but not colonization's duration.<sup>22</sup> We nevertheless undertook one set of exercises to seek assurance that our findings about the effect of colonization's timing on institutions and income is not proxying for an effect of duration and would also survive the addition of a control for duration: we estimated versions of the models in Table 7 in which we added *duration* as an additional control (see Table S.12). That variable obtains a significant negative coefficient only in the simpler 2SLS equation for  $\text{Ln}(gdp\ p.c.)$  corresponding to column (3), but its inclusion leaves  $\text{Ln}(colyr)$  negative and highly significant in that regression, while lowering the significance level of  $\text{Ln}(colyr)$  in column (4)'s *institutions* regression.  $\text{Ln}(colyr)$  has no significant coefficients in the modified  $\text{Ln}(gdp\ p.c.)$  equations, as in Table 7, and its already low significance in the 3SLS equation for *institutions* drops below the 10% level when *duration* is added. Coefficients on a few of the other variables lose significance when *duration* is added; most notably the significant impact of *eu* on *institutions* disappears in the GMM estimates, though not in their 3SLS counterpart. We can summarize by saying that our key finding that late colonization seems detrimental to current income if we fail to control for *eu* and *institutions* but not otherwise is robust to inclusion of an extra control for length of the colonial period.

## 5. Conclusion

We set out to study both the determinants and the effects of the colonization of non-European countries by Western European countries between the 15<sup>th</sup> and the 20<sup>th</sup> centuries, as well as the determinants and effects of colonization's timing. Our first important result was confirming the conjecture that which countries were colonized and when are to a significant degree explained by the level of development of the potential targets of colonization on the eve of the colonial era, as measured by years elapsed since transition to agriculture, experience of indigenous state-level polities, and level of technology adoption as of 1500 C.E. Our second result was the finding that roughly another quarter of the variation in whether countries were colonized and when is explicable by navigation distance from northwest Europe, latitude, and disease

---

<sup>22</sup> The timing of decolonization appears to be explained by a quite different set of factors, the study of which is beyond this paper's scope.

environment, bringing explained variance to the 50% neighborhood for both dependent variables.

Given our focus on the impact of early or pre-modern development, it is interesting to compare simple averages of our early development indicators for different categories of country. In Table 8, we divide our sample into 39 “early colonized” countries that fell under European rule during the 380 years prior to the 1842 takeover of Hong Kong by Britain, 54 “later colonized” countries having colonization dates in the 95-year interval from the latter event to Ethiopia’s 1937 takeover by Italy, and 18 “never colonized” countries. As expected, we find the average values of *tech1500*, *statehist*, and *agyears* to be considerably higher for the later- than for the earlier-colonized countries. Moreover, the never-colonized countries have average values of *tech1500*, *statehist* and *agyears* much higher than do the later-colonized countries. Distributions differ among groups in a statistically significant manner ( $p < .01$ ) with the exception of the distribution of *agyears* in early vs. late colonized countries.<sup>23</sup>

In the remainder of our investigation, we undertook to examine whether knowing the determinants of colonization and of its timing matters for judging the impact of colonization on the level of development of non-European countries today. To obtain a benchmark for comparison, we first estimated naïve regressions in which the fact and timing of colonization are permitted to impact (or show an association with) current income either alone or in concert with other factors, such as being landlocked, high or low malaria ecology, absolute latitude, size of European-descended population share, and quality of institutions. These regressions suggested negative relationships of both the fact of colonization and the lateness of colonization with the level of current income, but the effect for *col* (being colonized) is significant only with the addition of some and not other controls, while the effect for  $\text{Ln}(\text{colyr})$  (being colonized late) becomes insignificant once controls additional to latitude and malaria ecology are added.

---

<sup>23</sup> We performed Mann-Whitney U tests for difference of distribution of each of the three variables as between each pair of country sets.  $p$ -values are below 0.001 for each pair of country groups in the case of *tech1500*, below 0.01 for early vs. late and for late vs. never colonized and below 0.001 for early vs. never colonized in the case of *statehist*, and below 0.001 for early vs. never and for late vs. never in the case of *agyears*, but  $p = 0.1568$  in the comparison of early to late developed countries using *agyears*. All reported  $p$ -values are for two-tailed tests.

We then illustrated the impact of treating *col* and  $\text{Ln}(\text{colyr})$  as endogenous by estimating 2SLS models directly comparable to some of the OLS models just described. The coefficient on *col* remained insignificant when instrumented. However the coefficient on the instrumented  $\text{Ln}(\text{colyr})$  became highly significant, unlike that on  $\text{Ln}(\text{colyr})$  in the OLS counterpart, when controls for *eu* and *institutions* were omitted. Thus, accounting for *colyr*'s endogeneity is important in the reduced form setting in which *colyr* is allowed to pick up the effects that otherwise operate via its impacts on *eu* and *institutions*.

Our 3SLS and GMM models made more explicit the roles of European settlement and institutions, simultaneously estimating equations for income and institutional quality while instrumenting for *col* or  $\text{Ln}(\text{colyr})$  and for *eu*. These models deliver results in which neither the fact nor the timing of colonization is in the last analysis a significant determination of income in its own right. However, the models provide support for the idea that *col* and *colyr* have indirect effects on income by way of *eu* and *institutions*. Early colonization is indirectly associated with higher current incomes because it is associated with larger European-descended shares of current population (compare again Easterly and Levine, 2012) and with higher quality of institutions. It is worth remembering that this does not mean that any given country's people saw better outcomes for their own descendants thanks to having been colonized. Benefits may, rather, have accrued principally to descendants of the colonizers and of other European settlers who followed in their wake.

While we believe that our investigation pushes forward understanding by showing that colonization and its timing can be effectively explained and that there are benefits to treating them as endogenous variables insofar as they play a role in explaining long-run economic growth, we present our full multi-equation models for the determination of income and institutions more in the spirit of illustration than as a definitive entry into the important debates regarding the role of institutions, geography, genes, and culture in explaining comparative development. Among other limitations, we have limited our analysis to considering colonization as an undifferentiated phenomenon, rather than considering potentially distinguishing dimensions of colonization such as direct or indirect rule, identity of the colonizer, and differences in institutional regimes. Future

research may make further advances by taking such distinctions into account while also accounting, as our study has begun to do, for the endogenous nature of colonization itself.

## **References**

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

Acemoglu, Daron, Simon Johnson, and James Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117 (4): 1231–1294.

Acemoglu, Daron and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity and Poverty*. New York: Crown Publishers.

Ashraf, Quamrul and Oded Galor. 2012. "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review* (in press).

Auer, Raphael. 2011. "Geography, Institutions, and the Making of Comparative Development," unpublished paper, Swiss National Bank.

Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, Institutions and Economic Performance: The Legacy of Colonial Tenure Systems in India." *American Economic Review* 95 (4): 1190-1213.

Bertocchi, Graziella and Fabio Canova. 2002. "Did Colonization Matter for Growth? An Empirical Exploration into the Historical Causes of Africa's Underdevelopment." *European Economic Review* 46 (10): 1851-1871.

Bollen, Kenneth A. and Robert Jackman. 1990. "Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases." Pages 257-291 in John Fox and J. Scott Long (eds.) *Modern Methods of Data Analysis*. Newbury Park, CA: Sage.

Clingingsmith, David and Jeffrey Williamson. 2008. "Deindustrialization in 18<sup>th</sup> and 19<sup>th</sup> Century India: Mughal Decline, Climate Shocks and British Industrial Ascent." *Explorations in Economic History* 45 (3): 209-234.



Comin, Diego, William Easterly and Erick Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics* 2 (3): 65-97.

Diamond, Jared. 1998. *Guns, Germs and Steel: The Fate of Human Societies*. New York: Norton & Co.

Easterly, William and Ross Levine. 2003. "Tropics, Germs, and Crops: How Endowments Influence Economic Development." *Journal of Monetary Economics* 50 (1): 3-39.

Easterly, William and Ross Levine. 2012. "The European Origins of Economic Development." Unpublished Paper, New York University and Brown University.

Feyrer, James and Bruce Sacerdote. 2009. "Colonialism and Modern Income: Islands and Natural Experiments." *Review of Economics and Statistics* 91 (2): 245-262.

Gallup, John Luke and Jeffrey Sachs with Andrew Mellinger. 1999. "Geography and Economic Growth." Pages 127-178 in Boris Pleskovic and Joseph Stiglitz (eds.) *Annual World Bank Conference on Development Economics 1998*. Washington: The World Bank.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9 (3): 271-303.

Grier, Robin. 1999. "Colonial Legacies and Economic Growth." *Public Choice* 98 (3-4): 317-335.

Hall, Robert and Charles Jones. 1999. "Why do some countries produce so much more output than others?" *Quarterly Journal of Economics* 114 (1): 83-116.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1997. "Legal Determinants of External Finance." *Journal of Finance* 52 (3):1131-1150.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1998. "Law and Finance." *Journal of Political Economy* 106 (6): 1113-1155.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1999. "The Quality of Government." *Journal of Law, Economics and Organization* 15 (1): 222-279.

Landes, David S. 1998. *The Wealth and Poverty of Nations: Why Some are so Rich and Some so Poor*. New York: Norton & Co.

Maddison, Angus. 2001. *The World Economy: A Millennial Perspective*. Paris: Development Center of the OECD.

Marcus, Bernard. 2009. *Deadly Diseases and Epidemics: Malaria, 2<sup>nd</sup> Edition*. New York: Chelsea House.

Olsson, Ola. 2009. "On the Democratic Legacy of Colonialism." *Journal of Comparative Economics* 37 (4): 534–551.

Price, Gregory. 2003. "Economic Growth in a Cross-section of Nonindustrial Countries: Does Colonial Heritage Matter for Africa?" *Review of Development Economics* 7 (3): 478-495.

Putterman, Louis. 2007. "State Antiquity Index (Statehist)." Data set posted at [http://www.econ.brown.edu/fac/Louis\\_Putterman/antiquity%20index.htm](http://www.econ.brown.edu/fac/Louis_Putterman/antiquity%20index.htm).

Putterman, Louis, with Cary Anne Trainor. 2006. "Agricultural Transition Year Country Data Set." Data set posted at [http://www.econ.brown.edu/fac/Louis\\_Putterman/agricultural%20data%20page.htm](http://www.econ.brown.edu/fac/Louis_Putterman/agricultural%20data%20page.htm).

Putterman, Louis and David Weil. 2010. "Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality." *Quarterly Journal of Economics* 125 (4): 1627-1682.

Rodney, Walter. 1972. *How Europe Underdeveloped Africa*. London: Bogle-L'Ouverture Publications.

Sachs, Jeffrey. 2003. "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income." National Bureau of Economic Research Working Paper No 9490.

Sala-i-Martin, Xavier. 1997. "I Just Ran Two Million Regressions." *American Economic Review* 87 (2): 178–183.

Sokoloff, Kenneth and Stanley Engermann. 2000. "Institutions, Factor Endowments and Paths to Development in the New World." *Journal of Economic Perspectives* 14 (3): 217-232.

Spolaore, Enrico and Romain Wacziarg. Forthcoming. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature* (in press).

Weil, David. 2009. *Economic Growth (2nd Edition)*. Boston: Addison Wesley.

**Table 1. Descriptive statistics** (in order of appearance of the variables)

Variable	Mean	St. Dev.	Minimum	Maximum
<i>col</i>	0.8288288	0.3783667	0	1
<i>Ln(colyr)</i>	7.479084	0.0889553	7.28756	7.56838
<i>statehist</i>	0.2855204	0.3236524	0	1
<i>adjstate</i>	0.4044595	0.2847889	0	1
<i>agyears</i>	41.92342	25.39964	4	105
<i>adjagyears</i>	50.58251	22.526	13.5699	104
<i>tech1500</i>	0.3802222	0.2708923	0	0.883
<i>adjtech1500</i>	0.5353006	0.2388352	0.133	0.929
<i>malaria</i>	5.560935	8.188799	0	32.203
<i>pre-canal</i>	6.528928	3.731555	0.965	14.054
<i>landroute</i>	0.2702703	0.4252474	0	1
<i>landdist</i>	0.1362883	0.2702144	0	1.209
<i>latitude</i>	19.90155	12.67551	0.228	48.19
<i>Ln(gdp p.c.)</i>	8.059607	1.228591	5.37935	10.592
<i>institutions</i>	37.85311	22.03788	2.97727	96.2985
<i>eu</i>	13.47721	24.8645	0	91.89
<i>landlocked</i>	0.2432432	0.4309865	0	1
<i>post-canal</i>	5.31482	2.414368	0.965	11.141
<i>min-air</i>	2.962685	1.333445	0	6.694
<i>min-nav</i>	4.280487	1.964481	0	7.57
<i>duration</i>	148.7391	113.7534	5	513

See Table A.2 for variable definitions and sources.

**Table 2. Determinants of being colonized (Logit regressions)**

	Dependent variable: <i>col</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
<i>statehist</i>	-2.121* (1.209)			-1.262 (1.839)	-1.075 (1.356)	
<i>tech1500</i>		-2.205* (1.245)		1.167 (2.327)	-0.286 (1.875)	
<i>ageyears</i>			-0.0350*** (0.0127)	-0.0350* (0.0195)	-0.0397** (0.0173)	
<i>malaria</i>	-0.0584 (0.0757)	-0.0408 (0.0677)	-0.0903 (0.0609)	-0.0994 (0.0717)		-0.0417 (0.0747)
<i>pre-canal</i>	-0.287*** (0.0924)	-0.297*** (0.0986)	(0.0609) (0.119)	-0.342*** (0.119)		-0.319*** (0.105)
<i>landroute</i>	0.414 (1.048)	0.547 (0.985)	0.349 (0.977)	0.229 (1.022)		0.903 (0.911)
<i>landdist</i>	-4.032* (2.144)	-4.588** (1.972)	-4.100** (2.060)	-3.704* (2.133)		-5.056** (2.092)
<i>latitude</i>	-0.148*** (0.0457)	-0.136*** (0.0387)	-0.138*** (0.0480)	-0.143*** (0.0517)		-0.148*** (0.0407)
Constant	9.235*** (2.200)	9.119*** (2.016)	10.66*** (2.611)	10.67*** (2.538)	4.172*** -0.762	8.472*** (1.797)
N	111	111	111	111	111	111
Pseudo R <sup>2</sup>	0.5342	0.5306	0.5602	0.5639	0.2473	0.5088

Standard errors in parentheses

\*\*\* significant at the 1% level

\*\* significant at the 5% level

\* significant at the 10% level

**Table 3. Determinants of timing of colonization (OLS regressions)**

	Dependent variable: Ln( <i>colyr</i> )					
	(1)	(2)	(3)	(4)	(5)	(6)
<i>statehist</i>	0.0695*** (0.0261)			-0.0305 (0.0465)	-0.0679 (0.0501)	
<i>tech1500</i>		0.130*** (0.0266)		0.153** (0.0583)	0.250*** (0.0608)	
<i>ageyears</i>			0.00106*** (0.000264)	0.0000967 (0.000415)	-0.000507 (0.000392)	
<i>malaria</i>	0.00641*** (0.000990)	0.00563*** (0.000997)	0.00655*** (0.000960)	0.00551*** (0.00100)		0.00643*** (0.000971)
<i>pre-canal</i>	0.00722*** (0.00223)	0.00518** (0.00228)	0.00763*** (0.00210)	0.00510** (0.00232)		0.00857*** (0.00224)
<i>landroute</i>	0.0158 (0.0328)	0.0197 (0.0289)	0.0180 (0.0311)	0.0187 (0.0287)		0.00533 (0.0281)
<i>landdist</i>	0.0458 (0.0759)	0.0488 (0.0598)	0.0430 (0.0665)	0.0522 (0.0570)		0.0642 (0.0614)
<i>latitude</i>	0.00208** (0.000791)	0.00173** (0.000723)	0.00194** (0.000799)	0.00174** (0.000730)		0.00258*** (0.000807)
Constant	7.336*** (0.0258)	7.331*** (0.0255)	7.312*** (0.0267)	7.328*** (0.0277)	7.430*** (0.0166)	7.337*** (0.0263)
N	92	92	92	92	92	92
R <sup>2</sup>	0.393	0.461	0.409	0.465	0.221	0.342

Standard errors in parentheses

\*\*\* significant at the 1% level

\*\* significant at the 5% level

\* significant at the 10% level

**Table 4. Explorations of the impact of colonization on current income (OLS regressions)**

	Dependent variable: $\ln(gdp\ p.c.)$					
	(1)	(2)	(3)	(4)	(5)	(6)
<i>col</i>	-0.429 (0.301)	-0.352 (0.251)	-0.661** (0.277)	-0.600*** (0.203)	-0.0621 (0.281)	-0.286 (0.214)
<i>landlocked</i>		-1.033*** (0.228)	-0.900*** (0.214)	-0.657*** (0.172)	-0.582*** (0.218)	-0.459** (0.180)
<i>malaria</i>		-0.0730*** (0.0115)	-0.0538*** (0.0120)	-0.0422*** (0.00986)	-0.0485*** (0.00995)	-0.0317*** (0.00903)
<i>eu</i>			0.0171*** (0.00343)	0.00918*** (0.00273)		0.00809*** (0.00303)
<i>institutions</i>				0.0266*** (0.00374)		0.0261*** (0.00404)
<i>adjstatehist</i>					-0.305 (0.470)	0.167 (0.406)
<i>adjtech1500</i>					2.080*** (0.482)	0.648 (0.518)
<i>post-canal</i>					0.0122 (0.0432)	-0.00675 (0.0315)
<i>min-nav</i>					-0.0816 (0.0539)	0.0294 (0.0473)
<i>min-air</i>					-0.0143 (0.0872)	-0.165** (0.0763)
Constant	8.415*** (0.273)	9.009*** (0.231)	8.895*** (0.243)	7.820*** (0.247)	7.859*** (0.552)	7.472*** (0.365)
N	111	111	111	111	111	111
R <sup>2</sup>	0.017	0.405	0.500	0.676	0.531	0.723

Standard errors in parentheses

\*\*\* significant at the 1% level

\*\* significant at the 5% level

\* significant at the 10% level

**Table 5. Explorations of the impact of the timing of colonization on current income (OLS regressions)**

	Dependent variable: $\ln(\text{gdp p.c.})$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\text{colyr})$	-5.692*** (1.050)	-2.265** (1.108)	1.017 (1.172)	0.244 (1.014)	-1.106 (1.130)	0.506 (1.072)
<i>landlocked</i>		-0.845*** (0.259)	-0.793*** (0.248)	-0.671*** (0.200)	-0.476* (0.253)	-0.431** (0.207)
<i>malaria</i>		-0.0600*** (0.0122)	-0.0505*** (0.0112)	-0.0386*** (0.0104)	-0.0436*** (0.0116)	-0.0308*** (0.0110)
<i>eu</i>			0.0199*** (0.00379)	0.0104*** (0.00303)		0.00965*** (0.00354)
<i>institutions</i>				0.0256*** (0.00456)		0.0254*** (0.00458)
<i>adjstatehist</i>					-0.238 (0.586)	0.413 (0.495)
<i>adjtech1500</i>					2.050*** (0.592)	0.239 (0.640)
<i>post-canal</i>					-0.00440 (0.0576)	0.00134 (0.0389)
<i>min-nav</i>					-0.0691 (0.0708)	-0.0261 (0.0572)
<i>min-air</i>					0.0163 (0.101)	-0.131 (0.0799)
Constant	50.56*** (7.809)	25.47*** (8.196)	0.545 (8.744)	5.399 (7.485)	15.94* (8.401)	3.595 (7.840)
N	92	92	92	92	92	92
R <sup>2</sup>	0.171	0.400	0.510	0.668	0.513	0.716

Standard errors in parentheses

\*\*\* significant at the 1% level

\*\* significant at the 5% level

\* significant at the 10% level



**Table 6. Impact of colonization on institutions and income (2SLS, 3SLS, and GMM regressions)**

Estimation method: Dependent variable:	IV eq by eq		IV eq by eq		3SLS		GMM	
	Ln (gdp p.c.)	<i>institutions</i>	Ln (gdp p.c.)	<i>institutions</i>	Ln (gdp p.c.)	<i>institutions</i>	Ln (gdp p.c.)	<i>institutions</i>
<i>institutions</i>			0.0317*** -0.00923		0.0339*** -0.00963		0.0276*** -0.00815	
<i>eu</i>			0.00838 (0.00520)	0.233 (0.164)	0.00686 (0.00539)	0.205 (0.155)	0.0106** (0.00465)	0.253* (0.143)
<i>col</i>	(0.143) (0.383)	46.80*** (13.15)	-0.282 (0.374)	37.39** (14.56)	-0.305 -0.378	32.99** (14.65)	-0.367 (0.348)	31.72*** (12.13)
<i>landlocked</i>	-0.565*** (0.216)	-0.657 (5.689)	-0.386** (0.183)	-2.696 (5.582)	-0.416** (0.188)	-2.674 (5.668)	-0.423*** (0.163)	-3.558 (4.754)
<i>malaria</i>	(4.754) (0.00920)	0.0655 (0.243)	-0.0325*** (0.00898)	0.234 (0.268)	-0.0302*** (0.00934)	0.141 (0.269)	-0.0311*** (0.00862)	0.233 (0.257)
<i>adjstatehist</i>	0.247 (0.460)	-1.936 (9.035)	0.299 (0.390)	4.963 (9.323)	0.180 (0.398)	6.351 (9.089)	0.313 (0.359)	8.642 (8.048)
<i>adjtech1500</i>	1.810*** (0.448)	12.38 (13.76)	0.393 (0.508)	-3.446 (16.06)	0.481 (0.510)	2.156 (15.63)	0.337 (0.484)	-7.007 (14.04)
<i>post-canal</i>	0.038 (0.0409)	3.165** (1.313)	-0.0148 (0.0363)	3.392*** (1.203)	-0.0206 (0.0370)	3.292*** (1.148)	-0.00547 (0.0352)	3.133*** (1.096)
<i>min-nav</i>	-0.0384 (0.0501)	-2.726* (1.620)	0.0535 (0.0523)	-1.584 (1.608)	0.0464 (0.0537)	-1.962 (1.626)	0.0521 (0.0516)	-1.463 (1.461)
<i>min-air</i>	-0.0734 (0.0850)	-0.123 (2.773)	-0.187** (0.0918)	-1.230 (2.532)	-0.183** (0.0927)	-0.536 (2.519)	-0.194** (0.0916)	-0.951 (2.345)
<i>latitude</i>		1.002*** (0.280)		0.909*** (0.282)		0.722** -0.294		0.826*** (0.257)
Constant	7.463*** (0.554)	-31.76* (18.68)	7.324*** (0.493)	-23.16 (19.07)	7.328*** (0.503)	-17.98 (18.99)	7.514*** (0.442)	-16.34 (16.76)
N	111	111	111	111	111	111	111	111
Kleibergen-Paap test p-value	0.004525	0.017737	0.002281	0.012828	0.002281	0.012828	0.002281	0.012828
OIR test p-value	0.0018	0.3344	0.4216	0.3925	0.115		0.6125	

Standard errors in parentheses

\*\*\* significant at the 1% level

\*\* significant at the 5% level

\* significant at the 10% level

**Table 7. Impact of timing of colonization on institutions and income (2SLS, 3SLS, and GMM regressions)**

Estimation method: Dependent variable:	IV eq by eq		IV eq by eq		3SLS		GMM	
	Ln (gdp p.c.)	institutions	Ln (gdp p.c.)	institutions	Ln (gdp p.c.)	institutions	Ln (gdp p.c.)	institutions
<i>institutions</i>			0.0558*** (0.0153)		0.0821*** (0.0181)		0.0719*** (0.0163)	
<i>eu</i>			-0.00678 (0.0136)	0.487 (0.432)	-0.0254 (0.0182)	0.775*** (0.285)	-0.0187 (0.0158)	0.562** (0.281)
Ln( <i>colyr</i> )	-4.622*** (1.717)	-148.5*** (54.72)	-4.276 (4.518)	77.47 (186.0)	-8.973 (5.650)	223.4* (123.6)	-7.579 (5.088)	106.2 (114.1)
<i>landlocked</i>	-0.468** (0.214)	-2.967 (4.640)	-0.282 (0.242)	-2.725 (4.413)	-0.290 (0.356)	-2.527 (5.675)	-0.160 (0.294)	-2.374 (4.555)
<i>malaria</i>	-0.0222* (0.0133)	0.622 (0.381)	-0.0121 (0.0203)	-0.314 (0.769)	0.00727 (0.0264)	-0.853 (0.570)	-0.000929 (0.0238)	-0.408 (0.517)
<i>adjstatehist</i>	0.997* (0.519)	0.0604 (10.44)	0.492 (0.482)	-2.666 (10.37)	0.485 (0.723)	0.169 (13.01)	0.757 (0.595)	4.796 (10.44)
<i>adjtech1500</i>	1.055* (0.544)	3.307 (17.80)	0.142 (0.788)	-0.398 (16.90)	0.289 (1.195)	-6.756 (18.40)	-0.101 (1.033)	-4.791 (16.66)
<i>post-canal</i>	0.0201 (0.0592)	2.285* (1.237)	0.00398 (0.0494)	0.622 (1.647)	-0.0180 (0.0775)	0.0973 (1.516)	0.00559 (0.0670)	0.137 (1.340)
<i>min-nav</i>	0.0728 (0.0683)	1.287 (2.050)	0.130 (0.144)	-3.215 (3.759)	0.268 (0.190)	-6.066* (3.263)	0.224 (0.173)	-3.420 (2.779)
<i>min-air</i>	-0.121 (0.0893)	0.590 (2.225)	-0.282* (0.164)	3.662 (3.245)	-0.390* (0.225)	4.828 (3.348)	-0.385* (0.206)	4.209 (2.885)
<i>latitude</i>		0.997*** (0.314)		0.317 (0.609)		-0.151 (0.351)		0.163 (0.340)
Constant	41.81*** (12.65)	1107.1*** (400.9)	38.12 (33.11)	-551.5 (1365.5)	72.20* (41.46)	-1622.4* (908.1)	62.23* (37.32)	-762.8 (838.3)
N	92	92	92	92	92	92	92	92
Kleibergen-Paap test p-value	0.006443	0.012474	0.036776	0.007126				
OIR test p-value	0.0017	0.4254	0.2534	0.2711	0.0699		0.1829	

Standard errors in parentheses

\*\*\* significant at the 1% level

\*\* significant at the 5% level

\* significant at the 10% level

**Table 8. Averages of early development indicators for different categories of country**

	Colonized before 1842	Colonized in 1842 or later*	Never colonized	Overall average
average <i>tech1500</i>	0.230436	0.414667	0.636037	0.3802222
average <i>statehist</i>	0.127168	0.298348	0.568469	0.2855204
average <i>agyears</i>	30.52564	40.91667	69.63889	41.92342

\*1842 is the year in which Hong Kong was colonized.

**Table A.1. List of sample countries with colonization year and status**

Country	<i>colyr</i>	Country	<i>colyr</i>	Country	<i>colyr</i>
Afghanistan	n.c.	Ghana	1874	Nepal	n.c.
Algeria	1847	Guatemala	1524	New Zealand	1840
Angola	1750	Guinea	1849	Nicaragua	1524
Argentina	1580	Guinea Bissau	1886	Niger	1922
Armenia	n.c.	Guyana	1580	Nigeria	1885
Australia	1820	Haiti	1550	Pakistan	1849
Azerbaijan	n.c.	Honduras	1524	Panama	1538
Bangladesh	1757	Hong Kong	1842	Papua New Guinea	1884
Benin	1894	India	1765	Paraguay	1537
Bolivia	1538	Indonesia	1755	Peru	1533
Botswana	1885	Iran	n.c.	Philippines	1600
Brazil	1533	Iraq	1920	Rwanda	1899
Burkina Faso	1896	Israel	1917	Saudi Arabia	n.c.
Burundi	1903	Jamaica	1580	Senegal	1865
Cambodia	1863	Japan	n.c.	Sierra Leone	1896
Cameroon	1884	Jordan	1918	Singapore	1824
Canada	1700	Kazakhstan	n.c.	South Africa	1780
Cape Verde	1462	Kenya	1888	Sri Lanka	1658
Central African Republic	1903	Korea	n.c.	Sudan	1898
Chad	1910	Kyrgyzstan	n.c.	Swaziland	1893
Chile	1540	Laos	1893	Syria	1920
China	n.c.	Lebanon	1920	Taiwan	n.c.
Colombia	1549	Lesotho	1868	Tanzania	1891
Congo	1891	Liberia	n.c.	Thailand	n.c.
Costa Rica	1564	Libya	1911	Togo	1897
Cote d'Ivoire	1893	Madagascar	1895	Trinidad and Tobago	1592
Cuba	1511	Malawi	1891	Tunisia	1881
Democratic Congo	1885	Malaysia	1874	Turkey	n.c.
Dominican Republic	1505	Mali	1887	Turkmenistan	n.c.
Ecuador	1534	Mauritania	1898	Uganda	1894
Egypt	1882	Mauritius	1638	United States	1650
El Salvador	1528	Mexico	1521	Uruguay	1726
Ethiopia	1936	Mongolia	n.c.	Uzbekistan	n.c.
Fiji	1874	Morocco	1912	Venezuela	1600
Gabon	1885	Mozambique	1750	Vietnam	1867
Gambia	1889	Myanmar	1885	Zambia	1890
Georgia	n.c.	Namibia	1884	Zimbabwe	1897

n.c. = not colonized

**Table A.2. Variable definitions and sources** (in order of appearance)

---

*col* Dummy variable set to 1 if most of the country's territory was colonized by Belgium, England, France, Germany, Italy, Netherlands, Portugal or Spain during the period between 1462 and 1945, otherwise 0. Judgment on whether foreign involvement meets the standard of colonization is our own and is explained for each country in the Supplementary Online Appendix. Colonies include cases of indirect rule as well as League of Nations protectorates but exclude cases where sources speak merely of a foreign "sphere of influence."

*colyr* First year in which colonial rule by one of the powers mentioned is considered to have been effective over 20% or more of the present-day country's territory. Determination of colonization year is our own and is explained for each country in the Supplementary Online Appendix.

*statehist* Average value of index for presence of supra-tribal government on territory constituting the present-day country, covering years 1 C.E.-1500 C.E. In a given year, the index value is the product of three indices covering the unit interval: (1) an index for existence of a state, (2) an index for that state being domestically based (= 1 if so, 0.5 if imposed by an external power), and (3) an index for territorial extent and unity (states ruling small shares of the country's current territory and multiple simultaneously extant states get lower values). Values are aggregated into 50-year periods, the period  $x$  half centuries prior to 1500 is discounted by  $(1.05)^x$ , the resulting numbers are summed, and the sum is normalized to the 0, 1 interval by dividing by the hypothetical maximum value. Data are from Putterman (2004, revised 2012).

*adjstate* Adjusted *statehist*, equal to average *statehist* value of countries in which the ancestors of the country's year 2000 population lived in year 1500, according to Putterman and Weil's World Migration Matrix, 1500 - 2000 (2010). That is, the *statehist* value for each country accounting for a non-zero fraction of the country's year 2000 population is weighted by its proportion of ancestors as indicated in that source and summed. Apart from inclusion of new countries, this variable is identical to the one called "ancestry-adjusted *statehist*" by Putterman and Weil.

*ageyears* Number of years before year 2000, in hundreds, that a substantial population living within what are the present country's borders began to obtain most of their calories from agriculture. Data are from Putterman and Trainor (2007), which in turn details its sources.

*adjageyears* Adjusted *ageyears*, obtained from *ageyears* in a manner paralleling the calculation of *adjstate* from *statehist*. In other words, a weighted average of *ageyears* values for source countries of country's year 2000 population, where weights are shares of year 1500 ancestors as indicated in Putterman and Weil (2010). Equivalent to "ancestry-adjusted *ageyears*" in the latter paper.

*tech1500* An index of the adoption of agricultural, military, communications, and other technologies, from Comin, Easterly and Gong (2010).

*adjtech1500* Adjusted *tech1500*, obtained from *tech1500* in a manner paralleling the calculation of *adjstate* from *statehist* or *adjagyears* from *agyears*. In other words, a weighted average of *tech1500* values for source countries of country's year 2000 population, where weights are shares of year 1500 ancestors as indicated in Putterman and Weil (2010).

*malaria* The malaria ecology index measures the suitability of a country's climate to mosquito breeding as well as the prevalence of mosquito species that feed only on humans. The source is Kiszewski et al. (2004).

*pre-canal* Distance between Camaret-sur-mer and the closest port of historical significance (usually the main port) in each country, without considering routes going through the Suez or Panama canals. Source: AXSMarine distance table ([www.axsmarine.com](http://www.axsmarine.com)). Unit of measure: thousands of nautical miles.

*landroute* This variable indicates whether there was land in the colonizers' path to the potentially colonized country. Landlocked countries have *landroute* values of 1. Countries that were reached through their coastlines have values of 0. Countries conventionally reached over land, such as countries located on Latin America's Pacific coast typically reached by crossing the Isthmus of Panama, have *landroute* = 0.5. Further details are in footnote 6 of the paper.

*landdist* The distance that the colonizers had to traverse through ground transportation. For landlocked countries, it reflects the distance from the country's historically most important city (usually but not always the current capital) to the closest oceanic port. For El Salvador, Ecuador, Peru, and Chile, *landdist* is the distance between Panama City and Balboa—the Atlantic and Pacific ports that are now joined by the Panama Canal. All *landdist* values and the cities and ports used to calculate them are available in the Distances\_notes.xls file posted with our working paper. Source: Google Earth. Unit of measure: thousands of nautical miles.

*latitude* Absolute value of latitude. Source: Weil (2009). Unit of measure: degrees.

$\ln(\text{gdp } p.c.)$  Natural logarithm of gross domestic product per capita averaged over 1996-2009. Source: Penn World Tables 7.0 (PPP Converted GDP Per Capita (Chain Series), at 2005 constant prices). Unit of measure: 2005 International dollar per person (2005 I\$/person).

*institutions* Quality of formal institutions. We consider an index that averages six Indicators from the Worldwide Governance Indicators (WGI), averaged over 1996-2009: 1) Voice and Accountability; 2) Political Stability and Absence of Violence; 3) Government Effectiveness; 4)

Regulatory Quality; 5) Rule of Law; and 6) Control of Corruption. The WGI are produced by Daniel Kaufmann (Brookings Institution), Aart Kraay (World Bank Development Research Group) and Massimo Mastruzzi (World Bank Institute). Source: <http://info.worldbank.org/governance/wgi/index.asp>. Unit of measure: Percentile rank among all countries (ranges from 0 (lowest) to 100 (highest) rank).

*eu* Share of year 1500 ancestors of country's year 2000 population who resided in countries of Europe, defined as including Russia and excluding Turkey. Calculated from the World Migration Index, 1500-2000 by Putterman and Weil (2010).

*landlocked* Dummy variable that indicates if a country has direct access to the ocean—if it does not, *landlocked* takes a value of 1.

*post-canal* Distance between Camaret-sur-mer and the country's main present-day port, considering routes going through the Suez or Panama canals. Source: AXSMarine distance table ([www.axsmarine.com](http://www.axsmarine.com)). Unit of measure: thousands of nautical miles.

*min-air* Minimum air distance from the most important airport in each country to either New York, Tokyo or Berlin, measured as the shortest distance along a path on the surface of the world's sphere. Source: Inter-Airport Distances Data, Research and Innovative Technology Administration (RITA), U.S. Department of Transportation (US DOT) (<http://www.transtats.bts.gov/distance.asp>). Unit of measure: thousands of nautical miles.

*min-nav* Minimum distance between the country's main present-day port and either New York or Tokyo, considering routes going through the Suez or Panama canals. Source: AXSMarine distance table ([www.axsmarine.com](http://www.axsmarine.com)). Unit of measure: thousands of nautical miles.

*duration* Years of colonial rule beginning with designated *colyr* and ending with year of independence, based on *Encyclopedia Britannica*.