AGRICULTURAL SUITABILITY AND COLONIZATION CHOICES

James B. Ang* Per G. Fredriksson† Satyendra Kumar Gupta‡

Abstract

Historical evidence indicates that the quest for new territories by the Western European colonizers was driven by strong incentives to produce sugar, tea, timber, and other commodities. Once arrived, settlers also had a strong interest in becoming self-sufficient in food. This paper establishes that the timing and duration of colonization by the Western European powers were strongly influenced by regional agro-ecological factors. In particular, a recently constructed measure of the potential average crop yield is shown to have a robust effect on both the timing and duration of colonization. Areas with higher agricultural productivity tended to be colonized earlier and were subject to colonization for a longer duration of time. The findings are robust to controlling for multiple indicators of early development, geographical and ecological conditions, selection issues, and to the utilization of the Cox stratified proportional hazard model.

Key words: Colonization; duration; agriculture; productivity; crop yield; Columbian exchange; food.

JEL classification: N00; O10; O57

*Corresponding author. Department of Economics, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332. Phone +65 65927534, Fax +65 67946303, E-mail: James.Ang@ntu.edu.sg.

†Department of Economics, University of Louisville, Louisville, KY 40292, USA. Phone +1 502-852 4858, Fax +1502-8527672, E-mail: per.fredriksson@louisville.edu.

‡Department of Economics, Indian Institute of Management Rohtak, India. Phone: +91 1262-215559, Email: satyendra.gupta@iimrohtak.ac.in.

Acknowledgements: The authors thank Angeliki Kourelis, Olivier Hart-Vilain, Ed McInnis, Richard Rogerson, Laura Tetreault, and the participants at a University of Louisville Economics Department brown bag seminar for helpful discussions and comments. The usual disclaimers apply.

1 Introduction

Adam Smith considered the discoveries in the 1490s of America and the passage to the East Indies by the rounding of the Cape of Good Hope to be "the two greatest and most important events in recorded history of mankind" (Smith, 1776, p.626). The consequence was the subsequent colonization by Europeans of a large part of the world along these sea routes. A consensus has emerged that the colonial legacies play an important role in contemporary economic and political development (see, e.g., North, 1990; Hall and Jones, 1999; Sokoloff and Engerman, 2000; Acemoglu et al., 2001; Acemoglu et al., 2002). However, our understanding of the determinants of the timing and duration of colonization remains limited. Potential determinants such as factor endowments and geography have received insufficient attention, despite the fact that they may have affected subsequent economic, political, and institutional development (Acemoglu et al., 2001).

The present paper contributes to the emerging literature that seeks to identify the deep roots of contemporary economic development. By studying the determinants of the timing and duration of colonization, we facilitate an understanding of the long term influence of colonization. Our paper complements Ertan et al. (2016), who also study the timing (but not the duration) of European colonization by focusing on factors that *repelled* the colonizers. In particular, they examine the effects of the timing of the agricultural transition (the Neolithic revolution), statehood experience ("state antiquity"), and the level of technological development in the potential colonies. While we agree that these are plausible determinants of the timing of colonization, we focus instead on what *attracted* colonizers to a particular area. We argue that *agricultural productivity* was a crucial consideration by the colonizers, both in terms of the selection of sites suitable for colonization and the feasibility to remain in an area. We provide robust evidence that agricultural productivity is a major determinant of both the timing and duration of colonization.

The European colonizers were attracted by the economic opportunities offered by under-developed and under-defended areas. They competed for claims on natural resources and other advantages gained (such as opportunities for international trade) through the establishment of permanent settlements.[1] The economic returns from large-scale plantations in the colonies were sufficiently large for millions of African slaves to be brought to the Americas to work as slave laborers in agricultural activities (see, e.g., Govan, 1942; Fogel and Engerman, 1974).[2]

In the 15$^{\text{th}}$-18$^{\text{th}}$ centuries, before the Industrial Revolution, resources were limited and further European economic growth faced natural constraints (Hansen and Prescott, 2002).[3] The European population increased from 30 million in 1000 AD to 74 million in 1340 AD (Livi-Bacci, 2012).[4]

[1]The Dutch and the English East India Companies were set up during the 17$^{\text{th}}$ century to exploit trading opportunities in the East.

[2]The British plantation economy was so profitable that on the eve of the American revolution, the wealth generated from Jamaica was comparable to the wealth produced in a large English county such as Lancashire or Sussex (Burnard, 2001; Higman, 2005).

[3]Emphasizing the important role of land, Adam Smith (1776) argued that the growth of every species happens in proportion to the means of their subsistence, and the species can never multiply beyond their means.

[4]However, the population size was essentially stagnant for the next 100 years and beyond due to high mortality rates associated with the Black Death in Europe (Encyclopedia Britannica, 2017).

Population growth appears to have overtaken the capacity of farming to feed much of Western Europe. A growing number of people lived in tumultuous and insalubrious cities that were prone to riots (Bayly, 2004, p.93). With the discovery of the Americas, vast quantities of relatively unpopulated land became available. Crops such as sugar, coffee, soybeans, oranges, and bananas from the Old World were all introduced to the New World, and the Americas quickly started supplying vast quantities of these crops globally. Crosby (1989, p.666) emphasizes the importance of the Columbian exchange across the Atlantic, noting that: "The coming together of the continents was a prerequisite for the population explosion of the past two centuries, and certainly played an important role in the Industrial Revolution. The transfer across the ocean of the staple food crops of the Old and New Worlds made possible the former".[5,6]

What guided the European colonization efforts? Europe was predominantly an agricultural society in the 15th and 16th centuries and land played a key role in the economy (Broadberry et al., 2015). Agriculture was naturally seen as an important economic activity, suggesting that colonizers would favor agriculturally productive areas. Moreover, any long-term settlers would have to be largely self-sufficient in terms of food production and timber, as it would be costly and uncertain (particularly considering the food shortages in Europe) to transport supplies from elsewhere. Productive agricultural activities and the construction of housing would have to commence almost immediately in order for settlers to survive harsh winters and defend against the hostile indigenous population. This consideration appears to be particularly crucial early on in the colonization process, when food imports were not available from, e.g., other nearby earlier-colonized areas with potential surpluses. For example, the historian de Herrera y Tordesillas (1601-1615) reports that during Columbus' second voyage to Hispaniola (in present-day Dominican Republic) in 1493, hunger and disease led to mutiny by his men; this led to punishment, disillusion, and more hunger and disease. Moreover, (Master) George Percy was onboard one of three ships chartered by King James I that sailed with 144 men and boys in December 1606 to establish a colony in present-day Virginia. Percy wrote in June-July 1607 that the area had a fruitful soil, with abundant trees (mulberry, cherry, walnut, cedar, sassafras, and cypress trees, as well as vines). However, by August-September 1607, the men were suffering from a lack of fresh water, and were dying of diseases and conflicts with the native population, but mostly from famine (Percy and Quinn, 1967).

Based on the discussion above, we argue that agricultural productivity was a main determinant of the decision to colonize an area, both in the short and long run.[7] Thus, greater land productivity

[5] While sea routes to Africa were known to European sailors since the 14th century, there was no major colonization in Africa until the 'Scramble for Africa' conference took place in Berlin in 1884-1885. Meanwhile, Spanish sailors established colonies in Americas within a few years of their discovery, and were subsequently followed by other Europeans.

[6] We recognize that the colonization of many islands occurred because they were near important trade routes, as the winds may simply have brought ships to new locations. Feyrer and Sacerdote (2009) used wind patterns to study the colonization of islands.

[7] We recognize that the exploitation of agricultural opportunities was not the sole objective behind early colonists' exploration of sea routes through the Atlantic and the rounding of the Cape of Good Hope. Trade (controlled by traders and protected by the Crown) and religion (controlled by the Church) were interlinked motivations. Traders were looking for an alternative route to India to bring spices at low cost to Europe. It was a disappointment when Christopher Columbus returned from his first journey to the Americas (believed to be India) without any spices.

should have influenced the *timing* of an area's colonization in a positive fashion. Moreover, agricultural productivity should also have affected the *duration* of an area's colonization, as an uncertain supply of crops should have made an area less attractive to hold onto. To our knowledge, this is the first paper that studies the association between the timing and duration of post-1500 European colonization and the productivity of agriculture across non-European territories.[8]

The previous literature studying the contemporary effects of the interactions between Europe and the rest of the world includes a branch exploring the colonizers' influence abroad, e.g., on the institutions left behind (see Engerman and Sokoloff, 1997; La Porta et al., 1998; Acemoglu et al., 2001). The length of colonization has been used to explain economic growth and income levels by, for instance, Grier (1999), Bertocchi and Canova (2002), Price (2003) and Feyrer and Sacerdote (2009). Another strand of literature explores the impact of the discovery of the New World on Europe.[9] We believe our work may inform future research on these topics.[10]

Our measure of land productivity is an index of the potential average crop yield across different regions in each country, constructed by Galor and Özak (2016). We provide results using OLS, the Heckman (1979) two-step model, and the Cox stratified hazard model (Cox and Oakes, 1984).

Our cross-country analysis provides robust evidence that a greater potential crop yield was associated with an earlier and more extended colonization. The results survive a number of robustness checks, including tests for unobserved omitted variable bias, the exclusion of some colonizers from the sample, the inclusion of early development indicators, and controlling for several geographical and ecological conditions. Moreover, we find consistent estimates when the analysis is performed at the grid-cell level, which increases the credence of our main findings obtained at the country

However, subsequent Spanish settlers brought gold and silver from the Americas which led to excess supply of precious metals. This resulted in a prolonged period of high inflation in Western Europe that is also known as *The Price Revolution* (e.g., Hamilton, 1934). Others suggest that the main cause of *The Price Revolution* was not the excess supply of money in terms of gold and silver, but rather increasing population and low agricultural productivity (e.g., Gould, 1964). A large majority of Europeans lived close to starvation as the population expansion appears to have outstripped the growth of agricultural production.

Evangelists wanted to spread Christianity to new territories. In the early period of the Spanish empire, the Catholic Church worked to protect the colonial interests of the Spanish empire. However, we argue that the economic profits from agricultural activities played a major role in when and how long colonization occurred.

[8]The term colonization is quite broad, as earlier expansions by empires may be categorized as colonization. We focus on the post-1482 colonization by the Western European powers of the rest of the world (as in Olsson, 2009). Earlier imperial expansions such as the Roman and Ottoman empires, and the expansion by Imperial Japan in Asia in the early twentieth century, are thus excluded.

We make no distinction between settler colonization (direct administration, or direct rule, mainly in the New World) and exploitative colonization (indirect administration, or indirect rule, mainly in the Old World). Settler colonies were established in unpopulated or thinly-populated territories, and involved large number of emigrants from western or central Europe. These emigrants created institutions and cultures based on their experiences from the motherlands. The geographical factors identified to play key roles in the determination of the location of settler colonies include malaria ecology and disease rates (Acemoglu et al., 2001), and climate similar to Europe (Crosby, 1986). Exploitative colonialism involved a small number of colonists, whose main objective was to profit from the colony's resources, to exploit indigenous labor, or to use African slaves in the plantation economy.

[9]Acemoglu et al. (2005) argue that profits from the three-corner Atlantic trade strengthened the merchant class, which resulted in stronger pro-business institutions and increased economic growth in Europe. Moreover, the introduction of potatoes to the Old World strongly influenced its population growth and urbanization rate (Mokyr, 1981; Nunn and Qian, 2011).

[10]For example, one may speculate that the incentive to invest in quality institutions may have depended on the long term viability of local crop production.

level.[11]

The paper is organized as follows. The next section gives a background on the agricultural activities in former colonies. Section 3 outlines our model specification, and describes the main variables used in the analysis. Section 4 discusses the findings of our cross-section analysis and presents multiple sensitivity tests. Section 5 presents the findings of our grid-cell level analysis. Section 6 concludes.

2 Background: Sugar, Tea, and Timber

Before the advent of European colonies, most sedentary civilizations were built on the cultivation of a complex carbohydrate, such as maize, potatoes, rice, millet, or wheat (Mintz, 1985). In order to emphasize the role of agricultural productivity in the decision to colonize, in this section we discuss the importance of three key commodities: sugar, tea, and timber. First, historical evidence suggests that the consumption of sugar (imported from the Caribbean and the Americas) mixed with tea (imported from China and India) and other hot drinks became a cheap and easy source of calories for a growing urban working class in England (Burbank and Cooper, 2010).[12] Large scale organized agricultural production in the colonies indirectly supported the Industrial Revolution. Mintz (1985, p.180) argues that sugar helped create an industrial working class in the United Kingdom, suggesting that sugar "by provisioning, sating - and, indeed drugging - farm and factory workers, sharply reduced the overall cost of creating and reproducing the metropolitan proletariat."[13]

Second, consumption of tea in England increased exponentially during the 18th and 19th centuries. Customs records (i.e., legal imports) show that England received 20,000 pounds of green tea in the year 1700; the quantity increased to five million pounds by the year 1760, and to twenty million pounds by the year 1800 (Mintz, 1985, p.113). In addition, large amounts of tea entered England illegally to avoid taxes.[14]

[11] Since the grid-cell level analysis does not allow for within country variation, caution should be exercised in the interpretation of these results.

[12] While sugar does not provide protein, it offers calories in the absence of meat (Hersh and Voth, 2009). For the poor, a cup of sugary tea could reduce the feelings of hunger, provide energy for a short time, and substitute for a hot meal when the price of heating fuel was high (Mintz, 1985). The per capita consumption of sugar increased 20-fold from 1663 to 1775, and a further five-fold increase occurred between 1835 and 1935 (Nunn and Qian, 2010). By 1900, sugar provided almost 20 percent of the calorie intake of the average Englishman (Mintz, 1985). Hersh and Voth (2009) estimate that the increase in sugar availability between years 1600 and 1850 increased English welfare by eight percent.

[13] As a result of the outbreak of war in 1689, there was a ban on trade between England and France. This promoted the consumption of rum, a by-product of sugar, in Britain and Ireland. The consumption of rum increased manifold. Rum became part of many areas of English social life. It was even distributed as bribery during elections. Import of rum from the New World into England and Wales increased from a low of 2000 gallons a year in 1700 to two million gallons a year in 1773 (Parker, 2011).

[14] According to the government estimates, in 1766 the quantity of illegally imported tea was at least as large as the legal quantity (Mintz, 1985, p.113). James Cooper succinctly highlights the importance of trade for the availability of tea and sugar in *The Sea Lions*; *Or, The Lost Sealers*. The novel's heroine, Mary Pratt, criticizes the idea of trade: "Ah! why cannot men be content with the blessings that Providence places within our immediate reach, that they must make distant voyages to accumulate others!". The deacon responds: "You like your tea, Mary Pratt – and the

Third, timber from large and dense forests in the colonies was important for Europe. Demand for wood—required to build houses, for heating, and for ship building—increased continuously. The first English energy crisis in the 16th century pushed up the price of wood, and forced people to use coal.[15] Timber was important for ship building, but the scarcity of native forests in England constrained the supply of timber. Before the discovery of the New World, the Baltic countries were the main source of timber. However, the supply was prone to be cut off by hostile naval powers.[16] Subsequently, England was supplied with timber from the two Americas, Asia, Africa, and the South Seas.[17]

Agricultural produce and final product created a triangular trading market between the colonizers and the colonized (Acemoglu et al., 2005). The discovery of the Americas opened opportunities for the English population to consume cheap foods such as fish and dried fruits, as well as luxury drinks (e.g., wine) from the Americas (previously imported from France and the Mediterranean). Tobacco, an American product, gained popularity in Europe. At later stages of colonization, the colonies provided new markets for English products such as cloth manufactured in mills. The colonies were also helpful in tackling the unemployment problems resulting from changes in agrarian organization, with the poor becoming laborers or small farm holders in the New World (Harrison and Parry, 1968).

3 Empirical Approach and Data

3.1 Regression Model

We estimate the following regression model to evaluate the influence of potential crop yield on the timing of colonization:

$$col_i = \alpha + \beta cropyield_i + \gamma controls_i + \epsilon_i \tag{1}$$

where col_i is either years since colonization or the duration of colonization for geographical entity i (where i represents a country or a grid-cell), $cropyield_i$ is the mean of potential (average) crop yield in i, and $controls_i$ is a vector of control variables included in the regression model to account for the possible influence of other geographical features on the timing of colonization. We hypothesize that a geographical entity with higher potential crop yield should face a greater likelihood of early

sugar in it, and your silks and ribbons that I've seen you wear; how are you to get such matters if there's to be no going on v'y'ges? Tea and sugar, and silks and satins don't grow along with the clams on 'Yster Pond" (Cooper, 1860, p. 285) (the deacon uniformly pronounces the word 'oyster' as 'yster'). Mary changes the subject.

[15] However, the population was neither enthusiastic to use this 'dirty source' of energy, nor were they equipped to use it properly (Nef, 1977).

[16] This scarcity severely affected the performance of the English naval forces in several important wars: three Anglo-Dutch wars, the American Revolution, and contests with Napoleon (Albion, 1926). Napoleon's influence excluded England from practically every port in Europe which could furnish masts and ship timber.

[17] The English Navy appears to have sustained its long struggle with the Napoleonic Empire due to the uninterrupted supply of Canadian pines and oaks (Albion, 1926, p.346). It was subsequently discovered that Indian teak was superior for ship building. Consequently, foreign shipbuilding facilities opened in India, producing ships that played key roles in the English expeditions to Ceylon, Java, Manila, the Moluccas, and the Red Sea between 1795 and 1800 (Albion, 1926).

colonization by Europeans, and a longer period of colonization. Our baseline sample for the cross-country analysis includes data for up to 85 countries.

Following the standard practice in the long-run comparative development literature, Eq. (1) will be estimated mainly using the OLS estimator. When estimating the equation by OLS, our main concern is omitted variable bias. We will address this concern by performing the coefficient stability test developed by Altonji et al. (2005). Additionally, the analysis is also performed at the grid-cell level, in order to check if the country-level estimations are subject to endogeneity bias. Moreover, since the timing of colonization is positively skewed by construction, a negative binomial estimator is used as a robustness check. The results indicate that our OLS estimates are unlikely to be driven by some unobserved influences and that the negative binomial estimates are consistent with the OLS results (see section 4 for details), thus providing credence to the use of a simple estimator.

3.2 Data

A. Outcome measures

Years since colonization. The first dependent variable is the number of years elapsed, in 2000 AD, since European colonization is estimated to have occurred. The measure is expressed in hundreds of years. Table A1 in Appendix I provides the data. Identifying a single date for colonialism is difficult for several reasons. The initiation process of colonization was a gradual and informal process in many countries. For example, in the case of America, European settlers reached the New World and started settlements. Gradually, the populations of these settlements increased to hundreds of thousands of individuals of European descent. In the case of India, the colony was for all practical purposes under the control of the British East India Company after the battle of Plassey in 1757. However, it came under the direct control of the British crown as a colony only in 1858. To overcome such challenges, Ertan et al. (2016, p.170) use the following criteria (for their sample of 92 countries): "the process of colonization is considered to have started in a country when 20% or more of a country's territory was largely under the control of the colonizing power, provided that the majority of the territory would eventually be controlled by the same colonizer or another succeeding colonizer."

As a robustness check, we also utilize the data from Olsson (2009) on the timing of colonization. This dataset provides the date of colonization, date of independence, and a colonizing event for 143 former colonies which are currently independent countries.[18] Olsson (2009) identifies the date of colonization according to whether the Western colonizers became the major political power in the region concerned. His definition differs from the criteria used by Ertan et al. (2016). Olsson (2009, p.536) defines a Western colony as "a new and lasting political organization created outside Europe by Western countries....from the 15th to the 20th centuries through either invasion and conquest, and/or settlement colonization. Its rulers are in sustained dependence on a geographically remote

[18] A colonizing event refers to an historical event, battle or treaty between the colonizers and the natives that significantly tilted the power and control in favor of the colonialists.

mother country or imperial center that claims exclusive rights of possession of the colony or in other ways strongly dominates politics in the country." The correlation between the Olsson (2009) and Ertan et al. (2016, p.170) data on the timing of colonization equals 0.88 for 79 observations.

Duration of colonization. Our second outcome variable is the number of years that a country was under colonial rule. This is constructed by taking the difference between the year of independence, and the year of colonization identified by Olsson (2009). In the robustness checks, we use the year of independence from Olsson (2009) and the year of colonization from Ertan et al. (2016) to construct an alternative measure of the duration of colonization. The correlation between the two resulting measures of duration equals 0.79 for the 79 observations. Table A1 in Appendix I presents Olsson (2009) data on the duration of colonization (measured in hundreds of years) across countries.

B. Main explanatory variable: Potential average crop yield

As discussed above, our main independent variable is the potential average crop yield. This index captures the potential crop yield, measured in millions of kilocalories per hectare per year, across different regions in each country. It addresses the limitations of the available weight-based agricultural yield indices. For example, the land productivity index constructed by Ramankutty et al. (2002) does not capture the fact that equally suitable land may have a large variation in potential crop yield; caloric-intensive crops may not be cultivated in some regions (see, Galor and Özak, 2015).

The data on potential crop yield are taken from Galor and Özak (2016), who construct their dataset using the Global Agro-Ecological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). GAEZ supplies global yield and growth cycle estimates for 48 crops in grids with a cell size of 5'x5' (approximately 100 km^2).[19] GAEZ supplies potential crop yield estimates for each crop based on three alternative levels of input—high, medium and low—and two feasible sources of water supply—rain-fed water supply and irrigation. The FAO dataset supplies potential yield estimates for each crop in each agro-climatic grid while accounting for the effect of temperature and moisture on the growth of the crop.

The potential crop yield index is constructed using low level of inputs combined with rain-fed agricultural cultivation methods in order to account for the conditions during the pre-industrial era. This reflects farming practices prevalent during the early stages of development. Furthermore, the potential crop yield data are based on agro-climatic conditions that are independent of any human intervention. This mitigates any potential endogeneity pertaining to agricultural inputs, irrigation method, and soil quality.

Galor and Özak (2016) convert each potential crop yield in the GAEZ data (measured in tons per hectare per year) into caloric yield (millions of kilocalories per hectare per year), which

[19] These 48 crops are alfalfa, banana, barley, buckwheat, cabbage, cacao, carrot, cassava, chickpea, citrus, coconut, coffee, cotton, cowpea, dry pea, flax, foxtail millet, green gram, groundnuts, indigo rice, maize, oat, oil palm, olive, onion, palm heart, pearl millet, phaseolus bean, pigeon pea, rye, sorghum, soybean, sunflower, sweet potato, tea, tomato, wetland rice, wheat, spring wheat, winter wheat, white potato, yams, giant yams, subtropical sorghum, tropical highland sorghum, tropical lowland sorghum, and white yams.

represents the nutritional variation across crops. Comparisons across potential crop yields thus become feasible. The US Department of Agriculture (USDA) National Nutrient Database for Standard Reference provides data on the caloric content of various crops. Using these measures, the estimated average regional potential crop yield reflects the average regional levels of two variables – potential crop yield and crop cycle – for crops that provide the average caloric yield in each cell. We take the mean of the potential average crop yield for all grids in a country, which becomes our measure of potential average crop yield (see Table A1 in Appendix I).

C. Control variables

Several standard geographic influences which may potentially confound the relationship between potential average crop yield and the timing and duration of colonization are included in the regressions. They include terrain ruggedness, absolute latitude, elevation (mean and standard deviation), navigational distance from Europe to the colony (or nearest port), distance to nearest waterway, and landlockedness. Detailed sources are given in Appendix II. The summary statistics and the correlations of the variables are given in Table 1.

[Table 1]

4 Empirical Results and Discussion

4.1 Cross-Country Evidence

Columns (1a)-(1d) of Table 2 present our findings using a cross-section of 85 countries. Column (1a) includes no control variables, while column (1b) adds absolute latitude, and column (1c) in addition controls for mean elevation, variation in elevation, ruggedness, and landlockedness. Finally, the effect of geographic distance, measured via sea routes and land routes, respectively, is controlled for in column (1d), following Ertan et al. (2016). The influence of potential average crop yield on the timing of colonization is positive and significant at the 1% level in all three models. The estimate in column (1d) suggests that a one standard deviation (i.e., 944 units) increase in the potential average crop yield results in a territory's colonization occurring 90 years earlier, equivalent to 0.6 standard deviations. This supports our hypothesis that the European colonizers prioritized control of territories with high agricultural productivity. Two control variables exhibit significant coefficients: mean elevation has a negative effect on the speed of colonization, whereas increased variation in elevation has a positive effect.

[Table 2]

The partial regression plot in Figure 1 shows the relationship between the timing of colonization and potential average crop yield. The partial regression line displays a positive and statistically significant influence of potential average crop yield on the timing of colonization, after accounting for the geographic controls included in column (1d) of Table 2.

[Figure 1]

In the last four columns of Table 2, we repeat the same exercises but focus instead on the duration of colonization. We now control for the share of the population with European ancestors.[20] Colonization may have lasted longer in countries with a relatively greater share of the population being of European descent. The association with the colonizer may have been stronger and therefore sustained for a longer time period. To test this hypothesis, we regress the duration of colonization on crop yield and obtain the residuals. We then regress these residuals on the share of the population of European descent. This share explains these residuals at the 5 percent significance level.[21]

The results reported in last four columns of Table 2 are qualitatively and quantitatively similar to those discussed for the timing of colonization. A greater potential average crop yield is associated with a longer duration of colonization. Overall, our findings indicate that the colonizers sought to control more agriculturally productive regions longer than other regions. The partial regression plot in Figure 2 shows a consistent relationship between the duration of colonization and potential average crop yield.

[Figure 2]

4.2 Robustness to Bias from Unobservables

Although we have included some geographical controls in the estimations, concerns remain that some unobservables may be correlated with the potential average crop yield and the timing or duration of colonization. To address this concern, we use the approach suggested by Altonji et al. (2005). The basic idea is that the selection of observables may be used to estimate possible bias from unobservables. To assess the bias from unobservables, they suggest calculating the ratio $\frac{\widehat{\beta}^F}{\widehat{\beta}^R - \widehat{\beta}^F}$, where $\widehat{\beta}^{\mathrm{F}}$ and $\widehat{\beta}^{\mathrm{R}}$ are estimated coefficients of the full and restricted model, respectively. This ratio explains how much stronger the selection on unobservables should be, in comparison with the selection on observables, for the estimated effect using observables to be explained away. For the influence of potential average crop yield on the timing of colonization, columns (1a) to (1c) in Table 2 have ratios equal to 10.53, 8.63, and 53.28, respectively (note that column (1d) is the unrestricted model). The corresponding ratios are 10.58, 9.89, and 67.00 for columns (2a) to (2c), respectively. Since all ratios equal at least 8.63, the selection of unobservables must be much larger than the selection on observables in order to attribute the entire estimation effects to unobservables. Hence, our estimates are unlikely to be driven by unobservables.

4.3 Robustness to the Colonizer

We cannot rule out the possibility that the strategies and considerations of different colonizers were dissimilar. In particular, they may have viewed the importance of agricultural productivity

[20] All results are robust to dropping the share of the population of European descent.

[21] The standardized beta coefficient is 0.20 (one std. dev. in the share of the population of European descent explains 0.20 std. dev. in the residuals obtained in the previous step). However, the R^2 is only 0.04.

in a different light. For example, the first major export products from Spanish America were silver and gold, initially looted from natives and later mined at a heavy cost.[22] However, the main objective soon shifted to obtaining crops such as tea, sugar, cotton, sugar, and timber, which were needed in the motherland.[23] In contrast, the British and the French came late to America. They colonized land that was sparsely inhabited by the native population, and concentrated on agricultural activities (Denevan, 1976; Engerman et al., 2002). Our results may therefore be driven by the identity of some colonizers.

[Table 3]

Table 3 presents standardized beta coefficients for samples focusing on countries colonized by Britain, France and Other Western European countries (Spain, Portugal, Italy, and The Netherlands). In the latter group, combining several colonizers is necessary since the number of observations for each colonizer is insufficient to directly test our hypothesis. Our findings suggest that the potential average crop yield coefficient sizes are substantially higher in the British and French samples. While the sample sizes are small and the colonizers were to some extent active during different time periods, these results nevertheless indicate that British and French localization- and exit-decisions may have been relatively more concerned with agricultural productivity than other nations. This supports the view that the Spanish, Portuguese, Italian, and Dutch colonizers were relatively more interested in finding precious metals, although potential average crop yield still played a role. Similar evidence is reported in Table 4. Here, the colonizing nations Spain, Britain, France and Portugal are instead dropped one at a time. The potential average crop yield coefficients remain stable here, except in columns (1a) and (2a) where Spain is dropped. The greater coefficient sizes suggest that Spain may have been less concerned with agricultural productivity compared to its competitors.

[Table 4]

4.4 Robustness to Early Development Indicators

In this sub-section, we study the potentially confounding influence of various early development indicators, the majority of which are also utilized by Ertan et al. (2016). These include the timing of the transition to agriculture (Putterman, 2006), the history of statehood experience (Bockstette et al., 2002), an index of the level of technology (Comin et al., 2010), and population density (McEvedy and Jones, 1978; WDI, 2012), the early disease environment (Acemoglu et al., 2001), biogeography

[22] An example is the large quantity of gold and silver taken in ransom after capturing the Inca emperor Atahualpa in 1531 (Diamond, 1997).

[23] On June 7, 1494, Spain and Portugal came to an agreement (the Treaty of Tordesillas) to settle conflicts over lands newly discovered or explored by Columbus and other late 15th-century explorers. Spain was given exclusive rights to all newly discovered and undiscovered lands in the region west of a line of demarcation from pole to pole, approximately 320 miles west of the Cape Verde Islands, and Portugal received areas east of the line. Neither power was to occupy territory that was already ruled by a Christian. Other European powers did not accept this or a subsequent agreement, which moved the line to 1,185 miles west of the Cape Verde Islands and enabled Portugal to claim the coast of Brazil (see https://www.britannica.com/event/Treaty-of-Tordesillas).

(Diamond, 1997; Olsson and Hibbs, 2005), and settlement history (Ahlerup and Olsson, 2012). Panel A of Table 5 reports the results using the timing of colonization as the dependent variable. Panel B reports the estimates for the duration of colonization. Note that our measure of potential average crop yield is consistently significant in all models presented in Table 5.

[Table 5]

The transition from hunter-gatherer activities to agricultural practices induced the formation of early stationary societies, which should have encouraged the formation of culture and early institutions.[24] These developments may have had a confounding influence on the decision to colonize newfound territories. The presence of strong pre-colonial institutions may have been impediments to colonization (Hariri, 2012).

Column (1) indicates that the timing of the agricultural transition (the Neolithic Revolution) has a statistically significant negative influence on the timing of colonization (Panel A). Column (2) studies the role of accumulated statehood experience at the time that the first colonization wave began around 1500 AD. We find that states with more extended statehood experiences tended to be colonized at a later time. Both these findings are consistent with Ertan et al. (2016). The corresponding models in Panel B, however, suggest that the timing of the agricultural transition and statehood experience had no bearing on the duration of colonization.

Superior technologies (including horses and gunpowder) allowed fewer than six hundred Spanish men, led by Hernando Cortez, to conquer the Aztec empire with millions of subjects (McNeill, 1998, p.19).[25] Column (3) in Panel A indicates that technologically advanced states tended to be colonized later, consistent with Ertan et al. (2016). In contrast, Panel B indicates that the level of technology in 1500AD had no bearing on the duration of colonization.

Population density may have mattered. The colonizers were in direct competition with the native population for the control of natural resources. The colonizers would also have sought access to labor. Moreover, high population density may indicate a favorable location for economic activity. For example, the early Spanish colonizers of the Americas may have gravitated to places where large populations implied the potential availability of precious metals and a labor force to enslave.[26] Accounting for population density in year 1500AD in column (4) reveals that it reduced the probability of early colonization. However, duration was unaffected.

Next, we account for the early disease environment. Africa is geographically relatively close to Europe, and was known to the Europeans centuries before the start of African colonization in the 18th century. Even then, most outposts in Africa were established to support the spice trade with South and East Asia. One plausible reason for this reluctance to colonize African countries

[24] Note that the potential average crop yield and the timing of the Neolithic Revolution exhibit a relatively weak and negative correlation (-0.24).

[25] The effect of disease also played a role, however (Diamond, 1997; McNeill, 1998).

[26] However, note that the correlation between population density (log) and crop yield equals -0.13. For the subsample of Spanish colonies (19 observations), the correlation is -0.297.

may have been the presence of an adverse disease environment (Acemoglu et al., 2001).[27] This suggests a negative association between the early disease environment and the timing and duration of colonization. The results reported in column (5) support this hypothesis.

Column (6) accounts for the effect of biogeographic endowment. Olsson and Hibbs (2005) argue that the initial geographical and biological conditions in the prehistoric era influenced agricultural development. Similarly, Diamond (1997) proposes that the geographical endowment of the Fertile Crescent explains why the agricultural revolution originated there, and then spread to other regions. The endowment includes the large size of the Eurasian continent, the large Mediterranean zone in the western part, and the East–West orientation of its major axis. These factors contributed to a disproportionately-sized endowment of plants suited for cultivation and animals suited for domestication. The findings in column (6) support the argument that biogeographically more endowed and therefore more advanced countries were colonized later; the duration of colonization was unaffected, however.

Ahlerup and Olsson (2012) argue that a longer settlement history is associated with more ethnic and ethnolinguistic diversity. Internal social diversity among the native population may have played a role in the decision regarding when, and for how long, to colonize a region.[28] We use Ahlerup and Olsson's (2012) measure of duration of human settlement to account for the effect of internal social diversity. The results reported in column (7) indicate that a history of human settlement has a negative influence on both the timing and duration of colonization. Finally, in column (8) we run a horse race between our early development indicators. Among the control variables, only early disease environment (Panel A only), biogeography, and human settlement remain significant at conventional levels. Note that these findings should be interpreted with caution, however, as some of the measures exhibit high correlations.

We note that in all models reported in Panels A and B, the influence of potential average crop yield remains robust and statistically significant at least at the 5% level, suggesting a relatively robust influence.

4.5 Robustness to the Inclusion of Other Covariates

While it appears that agricultural productivity (potential average crop yield) was an important determinant of the timing and duration of colonization, we recognize that land productivity was perhaps not immediately or easily detectable to many sea captains, noblemen, and other decision makers (or their advisors) scouting for new lands. Instead, they may have used proxies such as temperature, precipitation, and the availability of fresh water as inputs into their decisions. In

[27]For example, yellow fever and malaria were widespread across the African continent. Conversely, in the Americas, the Europeans affected the native population severely and negatively by bringing new diseases. A large population of aborigines died in the Americas as a result.

[28]For instance, British colonial rule in India began after the decisive victory of the East India Company over the ruler of a large Indian state in the Battle of Plassey. Internal social diversity and mistrust among the native rulers of neighboring states was one reason for this victory.

Table 6, columns (1a) and (2a) control for the influence of temperature and precipitation.[29] While the level of precipitation appears to have had a positive effect on the timing and duration of colonization, temperature apparently did not impact these decisions.

[Table 6]

Human survival and agricultural productivity depend critically on the availability of fresh water. For example, the men who arrived in present-day Virginia in 1607 soon (by the end of the following summer) suffered from a shortage of high quality water (Percy and Quinn, 1967). Data on the volume of fresh water available at the beginning of colonization are unavailable. We therefore use the total surface of all inland water bodies (including lakes, reservoirs, and rivers) as a fraction of the total land area to serve as a proxy for the availability of fresh water. The results provided in columns (1b) and (2b) do not provide support for this hypothesis, perhaps due to the inexact measure available.

Finally, a measure of the variation of potential crop yield across all regions within a country attempts to control for colonization decisions possibly being affected by the degree of inequality in agricultural productivity. However, the variation in potential crop yield is insignificant in columns (1c) and (2c). Importantly, our main results remain robust throughout all models in Table 6.

4.6 Other Robustness Checks

This sub-section presents several additional findings using alternative samples, a different measure of agricultural productivity, an alternative estimation method, and alternative measures of the year of colonization.

The Spanish and Portuguese early efforts during the 1500s were motivated by finding precious metals in America and gaining favorable trading positions in the Asian spice trade (Landes, 1999; Lange et al., 2006). Moreover, the British and Dutch colonizers imported slaves from Africa for work on sugar plantations in the Caribbean and South America (Osterhammel, 2005). The early approach thus emphasized extraction. Olsson (2009) divides colonization into 'mercantilist' and 'imperialist' eras, based on evidence from, e.g., Fieldhouse (1984), Pakenham (1991), Osterhammel (2005) and Lange et al. (2006). The mercantilist colonization activities during the Age of Discovery ended around 1700, and the associated colonization lasted until approximately 1820.

The general thinking in Europe changed during the imperialist period. This era included the French and American revolutions, the independence of most South American nations around 1820, the Age of Enlightenment associated with scientific thinking and greater equality, the abolishment of British slave trade in 1807, the profound impact of the Industrial Revolution on European societies, and increasing nationalistic rivalries (Olsson, 2009). Importantly, the decline of the Ottoman Empire started after 1800 AD. Many territories previously under the control of the Ottoman Empire

[29] For most of the countries, no official records of temperature and precipitation exist before 1900. We assume that the average temperature and precipitation did not change significantly over the last five centuries, and use average temperature and precipitation for years 1961-1990.

were available for conquest by the Western Europeans powers. The imperialist era crested around 1880-1900. After the 1880s, the European colonial powers, in particular the British and the French, divided unexplored territories in Africa between themselves (the "Scramble for Africa") for the next 40 years. We therefore differentiate between the early wave of colonization until 1881 AD and the subsequent period, dividing our sample into two subgroups.

In Table 7, the results presented in columns (1) and (2) of Panel A indicate that our results for potential average crop yield on the timing of colonization are driven primarily by the period prior to 1881 AD. The coefficient loses significance in column (2). Note that this finding should be interpreted with caution due to the low number of observations. However, this lends some support to the view that the colonizers' demand for fertile land was greater in the earlier time period. The ability to transport large amounts of food on their own ships was limited, and to continuously import food from the motherland was not viable. Post-1881 food imports from colonies located relatively nearby were likely more feasible. Moreover, the coefficient on the share of European population becomes statistically significant at the 1% level in column (2), Panel B. This supports the notion that the colonizers remained longer in colonies with a greater share of the population of European descent.

[Table 7]

In order to mitigate any concern that our results are influenced by the choice of land productivity measure, in column (3) we use the land suitability index produced by Ramankutty et al. (2002). This index reflects how suitable a land is for farming based on soil quality and climate, and ranges from 0 to 1. The results indicate a robust and statistically significant association between land suitability and the timing and duration of colonization, reinforcing our earlier findings.

Our measure of the timing of colonization is positively skewed by definition. The negative binomial estimator is consequently an appropriate estimation method. In column (4) the estimated effect of potential average crop yield on the timing and duration of colonization remains statistically significant at the 1% level when this alternative approach is used. Next, we log transform the outcome variables to get a more normal distribution (column (5)); the results remain intact. Next, we use an alternative dataset constructed by Olsson (2009) for our outcome measures.Reassuringly, the results presented in column (6) remain consistent with our earlier findings.

Finally, some crops (such as cassava, sweet potato, white potato and maize) were only introduced to the Old World during the Columbian exchange. Hence, using crop yield data before 1500 AD may prevent us from identifying the historical effects of these crops on colonization. Consequently, we check the sensitivity of our results by using post-1500 crop yield data in column (7); but doing so still delivers consistent results.

4.7 Robustness to Selection Bias

Since some countries were never colonized, selection bias may potentially have affected our OLS results presented above. It would be of major concern if this were the case. In further robustness analysis, we address such concerns by using Heckman's (1979) two-step regression approach. In the first step, we find the determinants of colonization. In the second step, the regressions presented in Table 2 are repeated.

Ideally, all countries should be included in the first step regression model. However, a number of small countries have miniscule population sizes. Following Ertan et al. (2016), we focus on 111 non-European countries for which navigational distance data is available, and which either were or could have been colonized by a Western European power. These countries accounted for 95.4% of the world's population outside of Western Europe. However, data for all control variables are unavailable, and we are left with 103 observations; 85 former colonies, and 18 countries that were never colonized.

[Table 8]

In the first step of the model, we include the full set of control variables: potential average crop yield, latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigational distance, and land distance from the nearest coast. The results of the second step are reported in Table 8. The findings are qualitatively and quantitatively similar to those shown in Table 2. These findings suggest that the hypothesis of 'no selection bias' cannot be rejected in any of the specifications. We may therefore assume that our main results are not affected by selection bias.[30]

4.8 Further Analysis of Duration of Colonization

In this section, we utilize a stratified semi-parametric Cox proportional hazard model, which allows the baseline hazard of an end to colonization to differ across colonizing countries (Cox and Oakes, 1984). The null hypothesis is that individual and all covariates have zero slope, meaning that the log hazard-ratio function is constant over time. Rejection of the null hypothesis indicates deviation from the proportional-hazards assumption. The proportional hazards test of Grambsch and Therneau (1994) provides evidence in favor of stratifying our estimations according to the colonizer.[31] Accordingly, we divide the sample into five groups: Britain, France, Portugal, Spain, and Others (Italy, The Netherlands). Stratification may potentially be important if the different colonizers were looking for fundamentally different features in the areas to be colonized, and during different

[30] The results for the two-step Heckman's (1979) model are robust when extending the analysis to 143 countries, which necessitates dropping the navigation distance variable.

[31] The Cox Proportional Hazards Model for a single binary covariate, x, with coefficient β_1 describes the hazard of an event at time $t>0$ or time origin $t=0$ such that $\frac{\lambda(t;1)}{\lambda(t;0)} = \frac{\lambda_0(t)\exp(\beta_1 \times 1)}{\lambda_0(t)\exp(\beta_1 \times 0)}=\exp(\beta_1)$, where $\frac{\lambda(t;1)}{\lambda(t;0)}$ is the hazard ratio of $x=1$ compared to $x=0$ (Der and Everitt, 2005). The key assumption of the model is that the Hazard Ratio is constant over time. To test the validity of this assumption, Grambsch and Therneau (1994) propose a proportional hazard test of non-zero slope in a weighted regression, where rejection of the null hypothesis of a zero slope for individual and all covariates indicates that the proportional-hazards assumption is invalid. Accordingly, the test results for the individual covariates of the full specification (column (4), Table 8) are (p-values of the Chi-square test are given in parenthesis): potential average crop yield (0.899), absolute latitude (0.782), mean elevation (0.786), variation in elevation (0.859), ruggedness (0.732), landlockedness (0.659), navigation distance (0.948) and distance to coast (0.967). Hence, we can stratify our estimations according to the colonizer, given that the proportional hazards assumption is not violated.

eras (as discussed above). These considerations may in turn affect the duration of colonization periods.

[Table 9]

Our findings are presented in Table 9. The negative coefficient on potential average crop yield in all models indicates that higher agricultural productivity is associated with a lower probability of an end to colonization (i.e., a lower hazard rate of an event). In other words, a higher potential crop yield prolonged the duration of colonization. This analysis suggests that our earlier results are robust. It lends support to the hypothesis that potential average crop yield has a positive association with the duration of colonization (columns (1) to (4)). The results are similar if a basic Cox hazard model is used. Figure 3 suggests two distinct phases of independence in the former colonies (see Figure 3). The first phase of independence took place before 1850 and the second phase occurred after 1850. Hence, we separately estimate the duration model for countries that gained independence in the second phase. The results reported in column (5) indicate that the influence of potential crop yield on colonization duration remains consistent for this sub-sample.

[Figure 3]

5 Grid Cell Level Analysis

This section presents results obtained using grid cell level data. This method of estimation helps mitigate the concern of potential endogeneity problems that may arise in cross-section studies(see also, e.g., Michalopoulos (2012)). The world is divided into one-by-one degree grid cells (around 111 by 111 square kilometers at the equator). We then retrieve data for potential average crop yield and the control variables at the grid cell level (see Appendix II for the sources).

[Table 10]

We assume that a grid cell was under colonial rule if it is part of a present-day country that was once colonized. We are aware that this assumption may be of some concern. For example, the year of colonization is estimated to be 1758 for India. This is the year when the British colonists, under the label of the East India Company, became a dominant political force in India and controlled over 20% of the territory or population (this qualifies India as a British colony according to the criteria used by Ertan et al. (2016)). However, Portuguese and Dutch colonists had established trading posts in India almost one hundred years earlier, although their influence remained concentrated to some small coastal areas (Encyclopedia Britannica, 2017; Government of Goa India, 2017). Ideally we should take the separate years of colonization by different colonists into account when performing the grid cell level analysis. However, such a task is infeasible considering the scant information available and the scope of our study. Caution should be exercised when interpreting the results.

Table 10 presents our estimation results utilizing the models from Table 2 after clustering the standard errors at the country level. Column (1a) shows that the unconditional influence of potential average crop yield on the timing of colonization is statistically significant. The influence of potential average crop yield on the timing of colonization remains robust throughout.

Another concern is that the results are influenced by a few large countries in our sample. The largest number of grid cells is located inside the current borders of the United States and Canada, and both countries were colonized. To address this concern, we remove all observations belonging to these two countries. Column (1e) suggests that our main result is not vulnerable to this concern.

Next, we repeat the analysis for the duration of colonization. The results reported in columns (2a)-(2d) indicate that the effect of potential average crop yield remains robust. It remains significant when the grid-cell observations from Canada and the United States are removed (column (2e)). Note that both the adjusted R^2 and the coefficient size increase substantially. The inflated coefficient size may be due to greater variation (many observations with the same year of independence are removed). Another possibility is that Canada and the United States were predominantly settled by Europeans. Their descendants inherited values of independence and democracy from their ancestors, which may have yielded a lower probability of prolonged subjugation.[32]

6 Summary and Conclusions

The long term effects of agroecology, in particular geography and climate, have become important research topics in economics. This is part of a broader agenda to understand the deep roots of comparative development. The recent literature has identified the persistent influence of historic events and how they shape the characteristics of institutions and culture. Colonization is one such major historic event. This study contributes to this literature by identifying the role of agricultural productivity (potential average crop yield) as an important factor determining both the timing and duration of colonization.

In particular, we hypothesize that colonizers preferred to settle and control regions with higher potential crop yield. This improved their own chances of survival, and offered the opportunity to export agricultural products back to the mother country. Areas with higher potential crop yield therefore tended to have a higher likelihood of early colonization and a longer duration of colonization. The notion that potential crop yield was an important consideration for the selection of suitable areas for colonization is consistent with the historical evidence derived from the periods of colonization around the world. The demand for commodities such as sugar, tea and timber, created an incentive for colonizers to find new territories.

Our empirical cross-country results indicate that areas with higher potential crop yield were colonized at an earlier point in time, and experiences a greater duration of colonization. The results are robust to a number of considerations, thus providing strong support to the view that agricultural productivity played a key role in colonial history.

[32] Our results also hold when Australia and Brazil (with relatively fewer grid cells) are excluded from the estimations, along with the United States and Canada (results available upon request).

References

ACEMOGLU, D., JOHNSON, S. & ROBINSON, J. (2005). 'The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth', American Economic Review 95, 546-579.

ACEMOGLU, D., JOHNSON, S. & ROBINSON, J.A. (2001). 'The Colonial Origins of Comparative Development: An Empirical Investigation', American Economic Review 91, 1369-1401.

_____ (2002). 'Reversal Of Fortune: Geography And Institutions In The Making Of The Modern World Income Distribution', Quarterly Journal of Economics 117, 1231-1294.

AHLERUP, P. & OLSSON, O. (2012). 'The Roots of Ethnic Diversity', Journal of Economic Growth 17, 71-102.

ALBION, R.G. (1926) *Forests and Sea Power: The Timber Problem of the Royal Navy, 1652-1862* (Cambridge: Harvard University Press).

ALTONJI, J.G., ELDER, T. & TABER, C. (2005). 'Selection on Observed and Unobserved variables: Assessing the Effectiveness of Catholic Schools', Journal of Political Economy 113, 151-184.

AUER, R. (2013). 'Geography, Institutions, and the Making of Comparative Development', Journal of Economic Growth 7, 347-369.

BAMFORD, P. (1956) *Forests and French Sea Power 1660-1789* (Toronto: University of Toronto Press).

BAYLY, C.A. (2004) *The Birth of the Modern World, 1780-1914: Global Connections and Comparisons* (Oxford: Blackwell).

BERTOCCHI, G. & CANOVA, F. (2002). 'Did Colonization Matter for Growth?: An Empirical Exploration into the Historical Causes of Africa's Underdevelopment', European Economic Review 46, 1851-1871.

BOCKSTETTE, V., CHANDA, A. & PUTTERMAN, L. (2002). 'States and Markets: The Advantage of an Early Start', Journal of Economic Growth 7, 347-369.

BROADBERRY, S., CAMPBELL, B.M., KLEIN, A., OVERTON, M. & VAN LEEUWEN, B. (2015) *British Economic Growth, 1270–1870* (Cambridge: Cambridge University Press).

BURBANK, J. & COOPER, F. (2010) *Empires in World History: Power and the Politics of Difference* (Princeton and Oxford: Princeton University Press).

BURNARD, T.G. (2001). 'Prodigious Riches: The Wealth of Jamaica Before the American Revolution', Economic History Review 506-524.

CHANDA, A., COOK, C.J. & PUTTERMAN, L. (2014). 'Persistence of Fortune: Accounting for Population Movements, There Was No Post-Columbian Reversal', American Economic Journal: Macroeconomics 6, 1-28.

COMIN, D., EASTERLY, W. & GONG, E. (2010). 'Was the Wealth of Nations Determined in 1000 BC?' American Economic Journal: Macroeconomics 2, 65-97.

COOPER, J.F. (1860) *The Sea Lions, or, The Lost Sealers* (New York: Appleton).

COX, D.R. & OAKES, D. (1984) *Analysis of Survival Data* (Boca Raton: CRC Press).

CROSBY, A.W. (1986) *Ecological Imperialism: The Biological Expansion of Europe, 900-1900*

(Cambridge: Cambridge University Press).

_____ (1989). 'Reassessing 1492', American Quarterly 41, 661-669.

DENEVAN, W.M. (1976) *The Native Population of the Americas in 1492* (Madison: University of Wisconsin Press).

DER, G. & EVERITT, B.S. (2005) *Statistical Analysis of Medical Data Using SAS* (Boca Raton: Taylor & Francis).

DIAMOND, J. (1997) *Guns, Germs and Steel: The Fates of Human Societies* (New York: Norton).

EASTERLY, W. & LEVINE, R. (2016). 'The European Origins of Economic Development', Journal of Economic Growth 21, 225-257.

ENCYCLOPEDIA BRITANNICA (2017a). 'Black Death', Retrieved from https://global.britannica.com/event/Black-Death (Accessed: April, 2017).

_____ (2017b). 'Dutch East India Company', Retrieved from https://global.britannica.com/topic/Dutch-East-India-Company (Accessed: April, 2017).

ENGERMAN, S.L. & SOKOLOFF, K.L. (1997) Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies: A View from Economic Historians of the United States, in *How Latin America Fell Behind*, ed. S. Haber (Stanford: Stanford University Press), 260-304.

ENGERMAN, S.L., SOKOLOFF, K.L., URQUIOLA, M. & ACEMOGLU, D. (2002). 'Factor Endowments, Inequality, and Paths of Development among New World Economies [with Comments]', Economia 3, 41-109.

ERTAN, A., FISZBEIN, M. & PUTTERMAN, L. (2016). 'Who was Colonized and When? A Cross-country Analysis of Determinants', European Economic Review 83, 165-184.

FEYRER, J. & SACERDOTE, B. (2009). 'Colonialism and Modern Income: Islands as Natural Experiments', Review of Economics and Statistics 91, 245-262.

FIELDHOUSE, D.K. (1984) *Economics and Empire, 1830-1914* (London: MacMillan Publishers).

FOGEL, R.W. & ENGERMAN, S.L. (1974) *Time on the Cross: The Economics of American Negro Slavery* (New York: W.W. Norton & Company).

GALLUP, J.L., MELLINGER, A.D. & SACHS, J.D. (2010). 'Geography Datasets', Centre for International Development, Harvard University.

GALOR, O. & ÖZAK, Ö. (2015). 'Land Productivity and Economic Development: Caloric Suitability vs. Agricultural Suitability', Brown University, Department of Economics Working Papers 2015-5.

_____ (2016). 'The Agricultural Origins of Time Preference', American Economic Review 106, 3064-3103.

GOULD, J.D. (1964). 'The Price Revolution Reconsidered', Economic History Review 17, 249-266.

GOVAN, T.P. (1942). 'Was Plantation Slavery Profitable?' Journal of Southern History 8, 513-535.

GOVERNMENT OF GOA INDIA (2017). 'Department of Tourism, Portuguese Conquests Retrieved from http://www.goatourism.gov.in/history/portuguese-conquests (Accessed: April, 2017).'

GRAMBSCH, P.M. & THERNEAU, T.M. (1994). 'Proportional Hazards Tests and Diagnostics Based on Weighted Residuals', Biometrika 81, 515–526.

GRIER, R.M. (1999). 'Colonial Legacies and Economic Growth', Public Choice 98, 317-335.

HALL, R.E. & JONES, C.I. (1999). 'Why Do Some Countries Produce So Much More Output Per Worker Than Others?' Quarterly Journal of Economics 114, 83-116.

HAMILTON, E.J. (1934) *American Treasure and the Price Revolution in Spain, 1501-1650* (Cambridge: Harvard University Press).

HANSEN, G.D. & PRESCOTT, E.C. (2002). 'Malthus to Solow', American Economic Review 92, 1205-1217.

HARIRI, J.G. (2012). 'The Autocratic Legacy of Early Statehood', American Political Science Review 106, 471-494.

HARRISON, J. & PARRY, J. (1968). 'Colonial Development and International Rivalries Outside Europe', The New Cambridge Modern History 3, 1559-1609.

HECKMAN, J.J. (1979). 'Sample Selection Bias as a Specification Error', Econometrica 47, 153-161.

HERSH, J. & VOTH, H.-J. (2009). 'Sweet Diversity: Colonial Goods and the Rise of European Living Standards after 1492 (July 4, 2009)', Available at https://ssrn.com/abstract=1443730.

HIGMAN, B.W. (2005) *Plantation Jamaica, 1750-1850: Capital and Control in a Colonial Economy* (Kingston: University of the West Indies Press).

LA PORTA, R., LOPEZ-DE-SILANES, F., SHLEIFER, A. & VISHNY, R.W. (1998). 'Law and Finance', Journal of Political Economy 106, 1113-1155.

LANDES, D.S. (1999) *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor* (New York: W.W. Norton).

LANGE, M., MAHONEY, J. & VOM HAU, M. (2006). 'Colonialism and Development: A Comparative Analysis of Spanish and British Colonies', American Journal of Sociology 111, 1412-1462.

LIVI-BACCI, M. (2012) *A Concise History of World Population* (Chichester: John Wiley & Sons).

MCEVEDY, C. & JONES, R. (1978) *Atlas of World Population History* (New York: Facts on File).

MCNEILL, W. (1998) *Plagues and Peoples* (New York: Anchor).

MICHALOPOULOS, S. (2012). 'The Origins of Ethnolinguistic Diversity', American Economic Review 102, 1508.

MINTZ, S. (1985) *Sweetness and Power: the Place of Sugar in World History* (New York: Viking).

MOKYR, J. (1981). 'Irish History with the Potato', Irish Economic and Social History 8, 8-29.

NEF, J.U. (1977). 'Early Energy Crisis and its Consequences', Scientific American 237, 140-150.

NORTH, D.C. (1990) *Institutions, Institutional Change and Economic Performance* (Cambridge: Cambridge University Press).

NUNN, N. & PUGA, D. (2012). 'Ruggedness: The Blessing of Bad Geography in Africa', Review of Economics and Statistics 94, 20-36.

NUNN, N. & QIAN, N. (2010). 'The Columbian Exchange: A History of Disease, Food, and Ideas', Journal of Economic Perspectives 24, 163-188.

_____ (2011). 'The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment', Quarterly Journal of Economics 126, 593-650.

NYU DEVELOPMENT RESEARCH INSTITUTE (2009). 'Global Development Network Growth Database'.

OLSSON, O. (2009). 'On the Democratic Legacy of Colonialism', Journal of Comparative Economics 37, 534-551.

OLSSON, O. & HIBBS, D.A., JR. (2005). 'Biogeography and Long-Run Economic Development', European Economic Review 49, 909-938.

OSTERHAMMEL, J. (2005) *Colonialism: A Theoretical Overview* (Princeton: Markus Wiener Publishers).

PAKENHAM, T. (1991) *The Scramble for Africa* (London: Weidenfeld & Nicolson).

PARKER, M. (2011) *The Sugar Barons: Family, Corruption, Empire and War* (London: Hutchinson).

PERCY, G. & QUINN, D.B. (1967) *Observations gathered out of "A discourse on the plantation of the southern colony in Virginia by the English, 1606"* (Charlottesville: University Press of Virginia).

PRICE, G.N. (2003). 'Economic Growth in a Cross-section of Nonindustrial Countries: Does Colonial Heritage Matter for Africa?' Review of Development Economics 7, 478-495.

PUTTERMAN, L. (2006). 'Agricultural Transition Year Country Data Set', Brown University.

RAMANKUTTY, N., FOLEY, J.A., NORMAN, J. & MCSWEENEY, K. (2002). 'The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change', Global Ecology and Biogeography 11, 377-392.

SMITH, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations* (London: Methuen & Co., Ltd.).

SOKOLOFF, K.L. & ENGERMAN, S.L. (2000). 'History Lessons: Institutions, Factors Endowments, and Paths of Development in the New World', Journal of Economic Perspectives 14, 217-232.

WDI (2012) *World Development Indicators 2012 database* (Washington DC: World Bank).

Table 1: Summary statistics and correlations

Panel A: Summary Statistics					
	Obs	Mean	Standard deviation	Minimum	Maximum
Years since colonization (in 100s)	85	2.18	1.51	0.64	4.95
Duration of colonization	79	1.48	1.02	0.38	3.87
Potential average crop yield	85	1375	944	7	4478
Absolute Latitude	85	17.2	11.1	0.2	43.7
Mean elevation	85	0.5	0.4	0.0	1.7
Variation in elevation	85	0.4	0.3	0.0	1.6
Ruggedness	85	1.1	1.0	0.1	6.2
Landlockedness	85	0.2	0.4	0.0	1.0
Navigational distance	85	5.8	3.2	1.0	13.0
Distance to coast	85	0.1	0.2	0.0	0.9

Panel B: Correlations										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Years since colonization	1									
(2) Duration of colonization	0.95	1								
(3) Potential average crop yield	0.65	0.62	1							
(4) Absolute Latitude	0.08	-0.02	-0.09	1						
(5) Mean elevation	-0.08	-0.10	-0.10	-0.01	1					
(6) Variation in elevation	0.43	0.35	0.04	0.17	0.47	1				
(7) Ruggedness	0.13	0.10	0.14	0.21	0.53	0.32	1			
(8) Landlockedness	-0.29	-0.29	-0.04	-0.10	0.48	-0.14	0.17	1		
(9) Navigational distance	-0.08	-0.05	-0.05	-0.03	0.19	0.23	0.29	0.10	1	
(10) Distance to coast	-0.30	-0.30	-0.13	-0.19	0.36	-0.24	-0.01	0.73	-0.02	1

Notes: This table presents summary statistics and correlations of variables used in the baseline estimation. The number of observations is 79.

Table 2: Potential Crop Yield and Colonization

	(1a)	(1b)	(1c)	(1d)	(2a)	(2b)	(2c)	(2d)
Dependent Variable	Timing of Colonization				Duration of Colonization			
Potential average crop yield	0.65***	0.67***	0.61***	0.60***	0.61***	0.61***	0.56***	0.55***
	(6.94)	(6.56)	(6.31)	(6.29)	(5.17)	(5.02)	(4.55)	(4.47)
European population share					0.16**	0.19*	0.15	0.16
					(2.22)	(1.81)	(1.44)	(1.44)
Absolute Latitude		0.09	0.02	0.01		-0.05	-0.13	-0.14
		(0.92)	(0.20)	(0.06)		(-0.42)	(-1.04)	(-1.13)
Mean elevation			-0.22**	-0.26**			-0.15	-0.18
			(-2.00)	(-2.37)			(-1.21)	(-1.37)
Variation in elevation			0.49***	0.53***			0.38***	0.41***
			(5.07)	(5.69)			(4.05)	(4.23)
Ruggedness			-0.03	0.01			0.03	0.06
			(-0.32)	(0.07)			(0.25)	(0.54)
Landlockedness			-0.06	-0.08			-0.13	-0.13
			(-0.90)	(-0.99)			(-1.67)	(-1.53)
Navigation distance				-0.11				-0.08
				(-1.32)				(-0.88)
Distance to coast				0.07				0.02
				(0.90)				(0.31)
Adj. R^2	0.42	0.42	0.61	0.61	0.39	0.38	0.52	0.51
Observations	85	85	85	85	79	79	79	79

Notes: This table presents standardized beta coefficients. The dependent variable in the first four columns is the colonization timing whereas the next four columns use the duration of colonization as the outcome variable. The constant term is not reported for brevity. t statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level.

Table 3: Potential crop yield and colonization by colonizer

	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
	Timing of Colonization			Duration of Colonization		
Potential average crop yield	0.80***	1.04***	0.47**	0.83***	0.85***	0.37
	(4.34)	(3.32)	(2.59)	(3.50)	(3.08)	(1.46)
European population share				-0.12	0.38	0.31**
				(-0.62)	(1.50)	(2.40)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Colonizer	Britain	France	Others	Britain	France	Others
Adj. R^2	0.35	0.60	0.66	0.32	0.59	0.52
Observations	33	24	28	30	22	27

Notes: This table presents standardized beta coefficients by sequentially keeping observations of countries colonized by Britain, France, and other Western European countries (i.e. Spain, Portugal, Italy, and The Netherlands). The dependent variable in the first three columns is the colonization timing whereas the next three columns use the duration of colonization as the outcome variable. The constant term is not reported for brevity. t statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level. The baseline controls are absolute latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigational distance from Europe, and distance to coast.

Table 4: Robustness to the omission of each colonizer

	(1a)	(1b)	(1c)	(1d)	(2a)	(2b)	(2c)	(2d)
	Timing of colonization				Duration of colonization			
Potential average crop yield	0.77***	0.58***	0.54***	0.58***	0.70***	0.42***	0.51***	0.53***
	(5.84)	(5.89)	(4.28)	(6.17)	(4.58)	(3.20)	(3.18)	(6.01)
European population share					0.03	0.27***	0.13	0.17*
					(0.28)	(3.52)	(1.02)	(1.85)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Omitted colonizer	Spain	Britain	France	Portugal	Spain	Britain	France	Portugal
Adj. R^2	0.46	0.71	0.60	0.61	0.40	0.62	0.46	0.50
Observations	66	52	61	81	60	49	57	75

Notes: This table presents standardized beta coefficients by sequentially omitting observations of countries colonized by Spain, Britain, France, and Portugal. The dependent variable in the first four columns is the colonization timing whereas the next four columns use the duration of colonization as the outcome variable. The constant term is not reported for brevity. t statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level. The baseline controls are absolute latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigational distance from Europe, and distance to coast.

Table 5: Robustness to controlling for the effects of early development

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Crop yield and Timing of Colonization							
Dep. Var. =	*Timing of Colonization*							
Potential average crop yield	0.56***	0.53***	0.50***	0.55***	0.50***	0.53***	0.42***	0.27**
	(5.75)	(5.31)	(4.80)	(5.66)	(5.36)	(5.36)	(4.53)	(2.51)
Years since agricultural transition	-0.17***							-0.08
	(-2.91)							(-0.86)
State history in 1500AD		-0.18*						-0.03
		(-1.94)						(-0.31)
Technology in 1500AD			-0.19**					0.12
			(-2.40)					(0.77)
Population density in 1500AD				-0.21***				0.11
				(-2.99)				(1.24)
Early disease environment					-0.39***			-0.20*
					(-4.59)			(-1.73)
Biogeography						-0.22***		-0.40***
						(-3.02)		(-2.74)
Human Settlement							-0.49***	-0.57***
							(-4.77)	(-3.54)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.63	0.63	0.63	0.64	0.68	0.64	0.72	0.79
Observations	85	85	85	85	85	85	82	82
	Panel B: Crop yield and Duration of Colonization							
Dep. Var. =	*Duration of Colonization*							
Potential average crop yield	0.54***	0.55***	0.56***	0.54***	0.50***	0.53***	0.37***	0.31*
	(4.23)	(4.19)	(4.42)	(4.42)	(4.01)	(4.28)	(2.84)	(1.96)
European population share	0.14	0.16	0.16	0.13	0.08	0.13	0.09	-0.04
	(0.97)	(1.15)	(1.17)	(1.00)	(0.62)	(1.01)	(0.77)	(-0.23)
Years since agricultural transition	-0.03							-0.10
	(-0.30)							(-0.58)
State history in 1500AD		-0.00						0.01
		(-0.00)						(0.07)
Technology in 1500AD			0.02					0.19
			(0.16)					(1.00)
Population density in 1500AD				-0.07				0.12
				(-0.58)				(0.88)
Early disease environment					-0.23*			-0.13
					(-1.87)			(-1.16)
Biogeography						-0.05		-0.33**
						(-0.42)		(-2.36)
Human Settlement							-0.48***	-0.55***
							(-3.46)	(-2.73)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.51	0.51	0.51	0.51	0.53	0.51	0.61	0.61
Observations	79	79	79	79	79	79	76	76

Notes: This table presents standardized beta coefficients by controlling for some effects of early development. The dependent variable in panel A is the colonization timing whereas panel B uses the duration of colonization as the outcome variable. t statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level. The baseline controls are absolute latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigational distance from Europe, and distance to coast.

Table 6: Robustness to the inclusion of other covariates

	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
	Timing of Colonization			Duration of Colonization		
Potential average crop yield	0.52***	0.60***	0.58***	0.46***	0.55***	0.56***
	(4.94)	(6.18)	(5.95)	(3.63)	(4.37)	(4.15)
European population share				0.24	0.15	0.16*
				(1.65)	(1.39)	(1.71)
Temperature	-0.09			0.26		
	(-1.05)			(1.30)		
Precipitation	0.30**			0.30*		
	(2.17)			(1.95)		
Water/total area		-0.03			-0.03	
		(-0.59)			(-0.54)	
Variation in potential crop yield			0.09			-0.06
			(0.62)			(-0.37)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.65	0.61	0.61	0.54	0.51	0.51
Observations	84	84	85	78	78	79

Notes: This table presents standardized beta coefficients by controlling for other covariates, including temperature, precipitation, the availability of fresh water and variation of potential crop yield. The dependent variable in the first three columns is the colonization timing whereas the next three columns use the duration of colonization as the outcome variable. The constant term is not reported for brevity. t statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level. The baseline controls are absolute latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigational distance from Europe, and distance to coast.

Table 7: Other robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Potential crop yield and Timing of Colonization						
	Using Pre-1881 colonization period	Using Post-1881 colonization period	Using land suitability index	Using a negative binomial estimator	Log-transforming the dependent variable	Using Olsson's (2009) colonization data	Using post-1500 crop yield data
Average crop yield	0.58***	0.16		0.23***	0.58***	0.47***	
	(3.66)	(0.72)		(6.66)	(7.00)	(5.36)	
Average crop yield (Post 1500)							0.30***
							(3.83)
Land suitability			0.43***				
			(3.96)				
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.61	0.16	0.42	-	0.57	0.46	0.36
Observations	44	41	85	85	85	79	85
	Panel B: Potential crop yield and Duration of Colonization						
	Using Pre-1881 colonization period	Using Post-1881 colonization period	Using land suitability index	Using a negative binomial estimator	Log-transforming the dependent variable	Using Olsson's (2009) colonization data	Using post-1500 crop yield data
Average crop yield (pre-1500)	0.56***	0.44***		0.34***	0.60***	0.41***	
	(3.04)	(2.97)		(5.31)	(5.52)	(4.18)	
Average crop yield (post-1500)							0.20**
							(2.02)
Land suitability			0.42***				
			(4.14)				
European population share	-0.02	0.44**	0.19*	0.08**	0.18**	0.13*	0.20*
	(-0.19)	(2.30)	(1.97)	(2.09)	(2.57)	(1.73)	(1.70)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.28	0.33	0.18	-	0.38	0.16	0.05
Observations	44	35	79	79	79	79	79

Notes: This table presents standardized beta coefficients by considering alternative samples, approaches and outcome measures. The dependent variable in panel A is the colonization timing whereas panel B uses the duration of colonization I don't think these abbreviations are used in the paper as the outcome variable. The constant term is not reported for brevity. t statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level. The baseline controls are absolute latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigational distance from Europe, and distance to coast. Column (1) uses pre-1700 colonization data only whereas column (2) uses data for the post-1700 period only. Column (3) replaces potential crop yield with land suitability. Column (4) uses a negative binomial estimator. In column (5), the dependent variables are log-transformed. Column (6) uses the colonization data of Olsson (2009) to construct the outcome measures. Column (7) uses post-1500 potential crop yield data.

Table 8: Potential crop yield and colonization: Heckman Selection Model

	(1a)	(1b)	(1c)	(1d)	(2a)	(2b)	(2c)	(2d)
Dependent variable		Timing of colonization				Duration of colonization		
Potential average crop yield	0.64***	0.64***	0.59***	0.59***	0.56***	0.56***	0.52***	0.51***
	(7.89)	(7.84)	(8.14)	(6.55)	(6.33)	(6.30)	(6.43)	(6.20)
European population share					0.18**	0.17**	0.14*	0.15**
					(2.12)	(2.02)	(1.89)	(1.97)
Absolute Latitude		0.15	0.08	0.12		0.00	-0.09	-0.07
		(1.28)	(0.80)	(0.60)		(0.03)	(-0.64)	(-0.32)
Mean elevation			-0.25	-0.29			-0.14	-0.16
			(-1.63)	(-1.53)			(-0.77)	(-0.76)
Variation in elevation			0.57***	0.62***			0.41***	0.44***
			(5.53)	(4.78)			(3.39)	(3.22)
Ruggedness			-0.03	0.00			0.04	0.08
			(-0.33)	(0.04)			(0.37)	(0.68)
Landlockedness			-0.07	-0.17			-0.16	-0.23
			(-0.75)	(-0.92)			(-1.44)	(-1.02)
Navigation distance				-0.04				-0.06
				(-0.26)				(-0.46)
Distance to coast				0.23				0.13
				(0.87)				(0.52)
Mills/lambda	-0.01	-0.66	-0.85	-1.19	-0.04	-0.26	-0.14	-0.01
	(-0.02)	(-0.77)	(-1.16)	(-0.71)	(-0.96)	(-0.78)	(-0.64)	(-0.53)
Observations	103	103	103	103	103	103	103	103

Notes: This table presents standardized beta coefficients from two-step Heckman selection model. The independent variables used in the first-step probit regression are potential average crop yield, latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigational distance, and land distance from nearest coast. The dependent variable in the first four columns is the colonization timing whereas the next four columns use the duration of colonization as the outcome variable. The constant term is not reported for brevity. t statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level.

Table 9: Duration of Colonization Analysis using Cox Stratified Proportional Hazard Model

	(1)	(2)	(3)	(4)	(5)
					Independence since 1850 only
Potential average crop yield	-0.46***	-0.51***	-0.56***	-0.54***	-0.67**
	(-2.96)	(-2.98)	(-3.11)	(-2.94)	(-2.43)
European population share	-0.20	-0.14	-0.09	-0.13	-0.02
	(-1.46)	(-0.94)	(-0.59)	(-0.77)	(-0.09)
Absolute Latitude		-0.12	-0.14	-0.12	-0.41
		(-0.77)	(-0.79)	(-0.69)	(-1.49)
Mean elevation			-0.00	-0.14	-0.27
			(-0.01)	(-0.60)	(-0.84)
Variation in elevation			-0.03	0.09	0.11
			(-0.17)	(0.48)	(0.48)
Ruggedness			-0.05	0.05	0.31
			(-0.34)	(0.30)	(1.22)
Landlockedness			0.45***	0.17	0.14
			(2.70)	(0.70)	(0.40)
Navigation distance				-0.11	-0.09
				(-0.71)	(-0.37)
Distance to coast				0.51**	0.71*
				(2.01)	(1.94)
Pseudo R^2	0.04	0.04	0.07	0.08	0.08
Maximum likelihood	-165.43	-164.28	-158.78	-156.95	-113.35
Proportional Hazards test (global Chi-square test for all covariates)	0.27 (p=0.606)	0.33 (p=0.846)	0.79 (p=0.992)	1.46 (p=0.993)	-
Observations	79	79	79	79	59

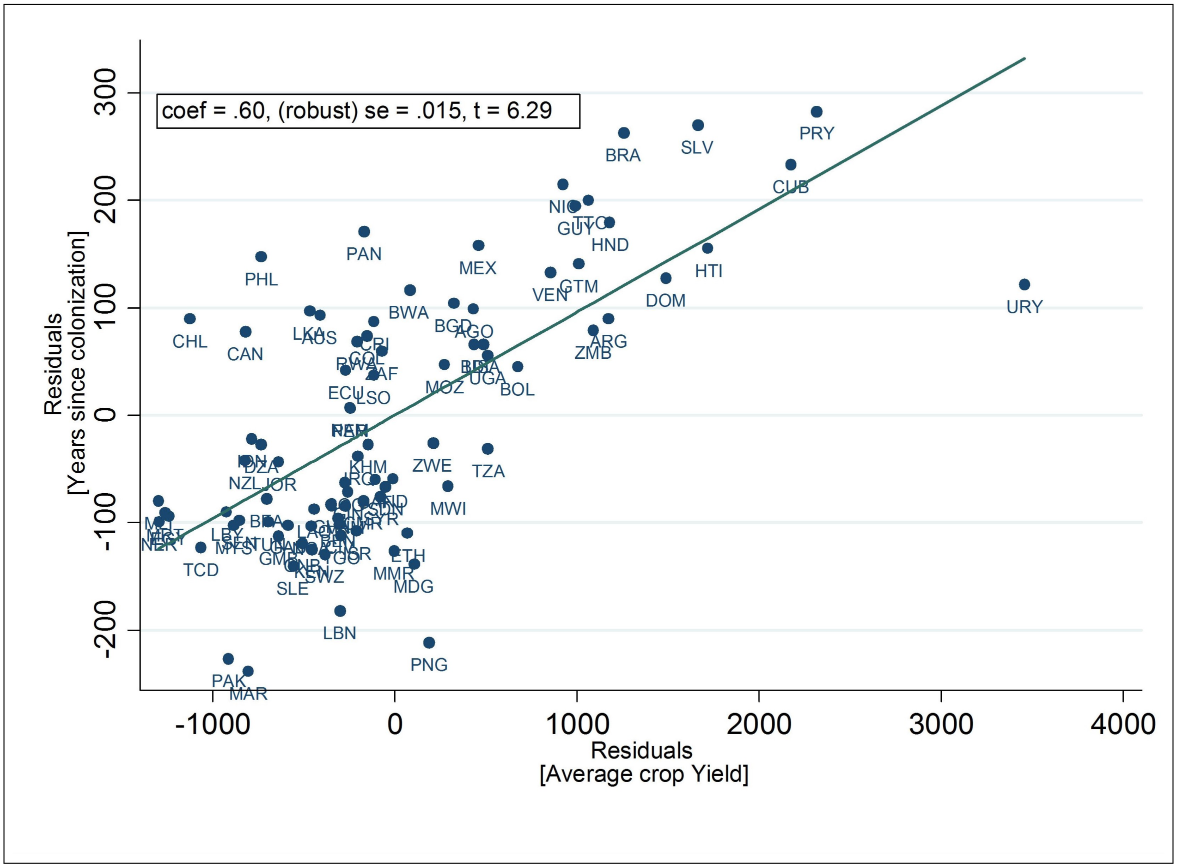
Notes: We utilize a stratified semi-parametric Cox proportional hazard model to perform an analysis on the duration of colonization. This approach allows the baseline hazard of an end to colonization to differ across colonizing countries. Since the Grambsch and Therneau (1994) global tests indicate that the proportional hazards assumption is not violated, as indicated by p>0.10 in all cases, we can stratify our estimation according to the colonizer. The negative coefficient estimates of potential average crop yield imply that higher agricultural productivity is associated with a lower probability of an end to colonization, consistent with the earlier finding that higher potential crop yield potential tends to increase the duration of colonization. The constant term is not reported for brevity. *t* statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level.

Table 10: Potential average crop yield and colonization (grid size 1 by 1 decimal degree)

	(1a)	(1b)	(1c)	(1d)	(1e)	(2a)	(2b)	(2c)	(2d)	(2e)
	Timing of Colonization					Duration of Colonization				
Potential average crop yield	0.39***	0.45***	0.45***	0.42***	0.48***	0.41**	0.41**	0.41**	0.39**	0.49***
	(2.64)	(3.36)	(3.66)	(3.46)	(4.96)	(2.13)	(2.05)	(2.13)	(2.20)	(4.47)
European population share						0.04	0.05	0.07	0.01	-0.11
						(0.24)	(0.26)	(0.37)	(0.07)	(-0.94)
Latitude (absolute)		0.18	0.20	-0.06	-0.29***		-0.02	-0.02	-0.19	-0.20
		(1.33)	(1.45)	(-0.34)	(-2.67)		(-0.07)	(-0.08)	(-1.00)	(-1.58)
Mean elevation			0.01	0.01	0.06			-0.01	-0.02	0.09
			(0.10)	(0.11)	(0.85)			(-0.07)	(-0.23)	(1.12)
Variation in elevation			0.23***	0.21***	0.20***			0.26***	0.25***	0.18***
			(4.04)	(5.14)	(4.46)			(5.26)	(5.48)	(4.33)
Ruggedness			-0.09	-0.05	-0.10			-0.09	-0.05	-0.09
			(-1.30)	(-0.73)	(-1.54)			(-1.50)	(-0.76)	(-1.34)
Landlocked				-0.32***	-0.24***				-0.26**	-0.23**
				(-3.82)	(-2.79)				(-2.60)	(-2.39)
Distance via sea				-0.23*	-0.18**				-0.16	-0.10
				(-1.96)	(-2.21)				(-1.23)	(-0.97)
Distance from coast				0.12	-0.02				0.13	-0.04
				(1.46)	(-0.17)				(1.53)	(-0.42)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Global	Global	Global	Global	Excl. USA & Canada	Global	Global	Global	Global	Excl. USA & Canada
Adj. R^2	0.15	0.18	0.21	0.33	0.47	0.16	0.16	0.20	0.27	0.43
Observations	11214	11214	11214	11214	7679	10966	10966	10966	10966	7431

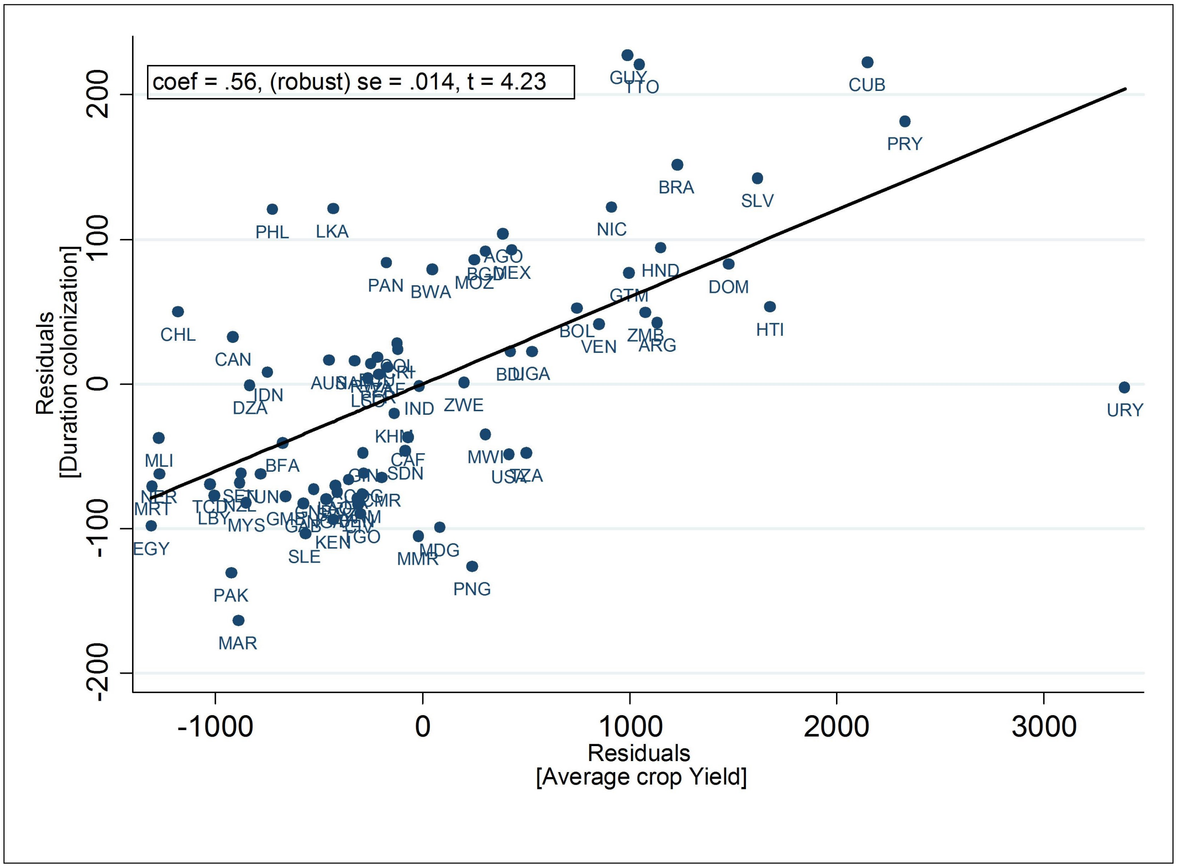
Notes: This table presents standardized beta coefficients of potential average crop yield on years since colonization before 2000 AD. All the control variables – absolute latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigational distance from Europe, and distance to coast – are computed at grid level of size 1 by 1 degree decimal. Constant term is included in all the regression, but not reported for brevity. t statistics are given in parentheses; *, **, and *** represent statistical significance at 10%, 5%, and 1% level.

Figure 1: The association between colonization timing and potential average crop yield



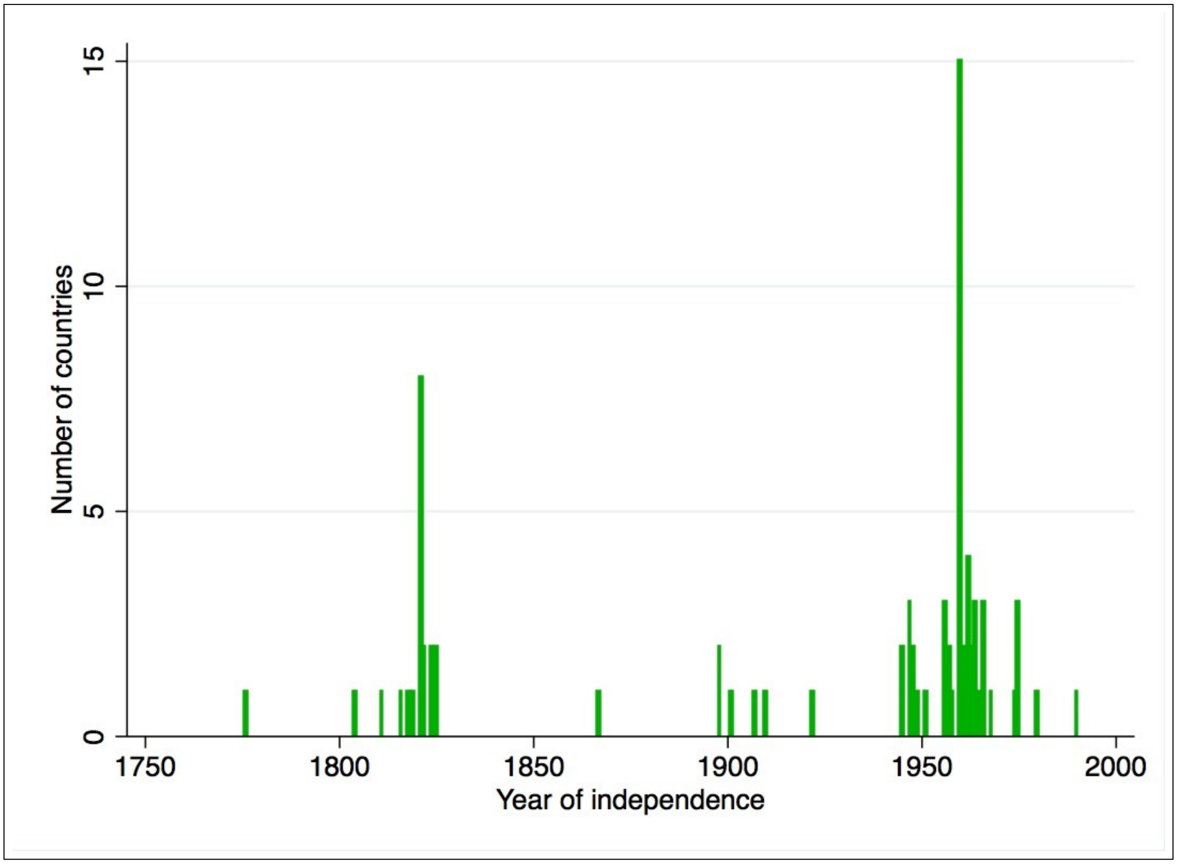
Notes: The partial regression line diagram illustrates the influence of potential average crop yield on the timing of colonization while partialing out the effects of other control variables, including latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigation distance from Europe, and distance from coast (see column (1d) of Table 2). The number of observations is 85.

Figure 2: The association between colonization duration and potential average crop yield



Notes: The partial regression line diagram illustrates the influence of potential average crop yield on the duration of colonization while partialing out the effects of other control variables, including latitude, mean elevation, variation in elevation, ruggedness, landlockedness, navigation distance from Europe, and distance from coast (see column (2d) of Table 2). The number of observations is 79.

Figure 3: Histogram for the year of independence among former colonies



Notes: The histogram illustrates the year of independence among former colonies in the sample. The number of observations is 79. The first country to obtain independence in the sample is the USA (1776), and the last is Namibia (1990).

Appendix I

Table A1. The Timing and Duration of Colonization, and Potential Crop Yield, by Country

Country	Years since colonization (in 100s of years)	Duration of colonization (in 100s of years)	Potential Crop Yield
Algeria	1.53	1.15	127
Angola	2.5	2.25	1468
Argentina	4.2	2.36	2192
Australia	1.8	0.81	180
Bangladesh	2.43	1.9	1301
Benin	1.06	0.66	1178
Bolivia	4.62	2.87	2362
Botswana	1.15	0.81	598
Brazil	4.67	2.89	2478
Burkina Faso	1.04	0.64	869
Burundi	0.97	0.59	1288
Cambodia	1.37	1.01	1132
Cameroon	1.16	0.76	1107
Canada	3	1.67	155
Central African Republic	0.97	0.57	1261
Chad	0.9	0.5	372
Chile	4.6	2.78	162
Colombia	4.51	2.7	1731
Congo, Rep.	1.09	0.69	1071
Costa Rica	4.36	2.57	1984
Cote d'Ivoire	1.07	0.67	1216
Cuba	4.89	3.87	3668
Dominican Republic	4.95	3.16	3428
Ecuador	4.66	2.88	1637
Egypt	1.18	0.4	7
El Salvador	4.72	2.93	3182
Ethiopia	0.64	-	919
Gabon	1.15	0.75	957
Gambia	1.11	0.76	994
Ghana	1.26	0.83	1207
Guatemala	4.76	2.97	2810
Guinea	1.51	1.09	1263
Guinea Bissau	1.14	0.88	1185
Guyana	4.2	3.86	2542
Haiti	4.5	2.54	3669
Honduras	4.76	2.97	3080
India	2.35	1.82	1100
Indonesia	2.45	1.9	856
Iraq	0.8	-	496
Israel	0.83	-	1070
Jordan	0.82	-	162
Kenya	1.12	0.75	879
Laos	1.07	0.56	1408
Lebanon	0.8	-	1508
Lesotho	1.32	0.98	1642
Libya	0.89	0.4	65
Madagascar	1.05	0.65	1514
Malawi	1.09	0.73	1700
Malaysia	1.26	0.83	770
Mali	1.13	0.73	281
Mauritania	1.02	0.62	30

Mexico	4.79	3.03	2012
Morocco	0.88	0.44	893
Mozambique	2.5	2.25	1501
Myanmar	1.15	0.63	1392
Namibia	1.16	1.06	462
New Zealand	1.6	0.67	361
Nicaragua	4.76	3	2624
Niger	0.78	0.38	99
Nigeria	1.15	0.75	1028
Pakistan	1.51	0.98	384
Panama	4.62	2.83	1790
Papua New Guinea	1.16	0.91	1968
Paraguay	4.63	2.74	3637
Peru	4.67	2.88	1239
Philippines	4	2.98	1013
Rwanda	1.01	0.64	936
Senegal	1.35	0.95	726
Sierra Leone	1.04	0.65	1109
South Africa	2.2	1.3	874
Sri Lanka	3.42	2.9	1166
Sudan	1.02	0.58	496
Swaziland	1.07	0.75	1663
Syria	0.8	-	822
Tanzania	1.09	0.7	1512
Togo	1.03	0.63	1265
Trinidad and Tobago	4.08	3.7	2695
Tunisia	1.19	0.75	501
Uganda	1.06	0.68	1465
United States	3.5	1.26	1658
Uruguay	2.74	0.99	4478
Venezuela	4	2.21	2421
Vietnam	1.33	0.78	1204
Zambia	1.1	0.74	1780
Zimbabwe	1.03	0.83	1473

Notes: Years since colonization measures the number of years (in hundreds) elapsed, in 2000 AD, since colonization by Europeans was estimated to occur. The data are taken from Ertan et al. (2016). *Duration of colonization* is the number of years (in hundreds) that a country was under colonial rule. The data are taken from Ertan et al. (2016). It is measured as the difference between the year of independence and the first year of colonization estimated by Olsson (2009). Data for some countries are not provided by Olsson (2009). Potential Crop Yield comes from Galor and Özak (2016).

Appendix II: Data Sources

A. Outcome variables

Timing of colonization. This outcome variable is measured as the number of years elapsed, in 2000 AD, since colonization by Europeans was estimated to occur. The data are compiled by Ertan et al. (2016), who use some quantitative and objective criteria to identify the date of colonization. According to their criteria, the process of colonization is considered to have started in a country when "20% or more of a country's territory was largely under the control of the colonizing power, provided that the majority of the territory would eventually be controlled either by the same authority or by another colonizer."

Duration of colonization. This variable is measured as the number of years that a country was under colonial rule. It is constructed by taking the difference between the year of independence and the first year of colonization identified by Olsson (2009).

B. Explanatory variable

Potential Average Crop Yield. This index captures the potential average crop yield (measured in millions of kilocalories per hectare per year) across different regions in a country. Crop yield index is constructed using crop yield (measured in tons per hectare per year) from the Global Agro-Ecological Zones (GAEZ) project of Food and Agriculture Organization (FAO) and caloric content of various crops from US Department of Agriculture Nutrient Database for Standard Reference (source: Galor and Özak, 2016).

Land suitability. This index, which ranges from 0 and 1, reflects how suitable a land is for farming based on soil quality and climate. This variable is used as a robustness check. The data are taken from Ramankutty et al. (2002).

C. Baseline geographic controls

Ruggedness. The calculation for ruggedness takes a point on the earth's surface and measures the difference in elevation between this point and each of the eight major directions of the compass (north, northeast, east, southeast, south, southwest, west, and northwest). The index at the central point is given by the square root of the sum of the squared differences in elevation between the central point and the eight adjacent points (source: Nunn and Puga, 2012).

Distance to waterways. The distance, in thousands of kilometres, from a GIS grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells of a country (source: Gallup et al., 2010).

Elevation. The average elevation of a country, in thousands of kilometers above sea level, calculated using geospatial data at a 1-degree resolution from the Geographically based Economic data (G-ECON) project.

Landlockedness. This dummy variable indicates whether a country is landlocked or not (source: NYU Development Research Institute, 2009).

Latitude. The absolute value of latitude of a country's centroid (CIA, 2015).

Navigational distance. Distance between Camaret-sur-mer and the nearest port of historical significance in each country (source: Ertan et al., 2016).

Share of European population. The fraction of a country's population having European ancestors during colonization (source: Easterly and Levine, 2016).

Variation in elevation. This measure indicates the standard deviation of elevation within a country (source: Michalopoulos, 2012).

D. Other controls

Years since agricultural transition. The number of years elapsed, in 2000 AD, since the transition to agriculture was estimated to occur (in thousand years) (source: Putterman, 2006).

Biogeography. The first principal component of the standardized numbers of domesticable wild plants and animals (source: Olsson and Hibbs, 2005).

Early Disease Environment. The annualized probability of death for European males in the age cohort of soldiers (source: Ertan et al., 2016, derived from Auer, 2013).

Duration of human settlement. The historical duration of human settlement (in million years) (source: Ahlerup and Olsson, 2012).

Mean land quality (soil-based). This variable measures the average agricultural suitability of land across regions based on soil properties (source: Michalopoulos, 2012).

Population density in 1500 AD (log). The population in 1500 AD divided by land area (source: McEvedy and Jones, 1978; World Development Indicators, 2012).

Precipitation. The average precipitation of a country, in a unit of 10 millimeters per month over the period from 1961-1990, calculated using geospatial data at a 1-degree resolution from the Geographically based Economic data (G-ECON) project.

State history up to 1500 AD. An index of state history covering the period from 1 AD to 1500 AD, scaled to take values between 0 and 1. The dataset was originally introduced by Bockstette et al. 2002, but the current paper uses its latest version, v3.1. (source: Chanda et al., 2014).

Temperature. The average temperature of a country, in a unit of 10 degree Celsius per month over the period from 1961-1990, calculated using geospatial data at a 1-degree resolution from the Geographically based Economic data (G-ECON) project.

Proportion of water area. Total surface of all inland water bodies (including lakes, reservoirs, and rivers) as a fraction of the total land area (source: CIA, 2015).