NBER WORKING PAPER SERIES

THE EUROPEAN ORIGINS OF ECONOMIC DEVELOPMENT

William Easterly
Ross Levine

Working Paper 18162
http://www.nber.org/papers/w18162

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2012

Steven Pennings provided superb research assistance in the final stages of this paper. The data collection project also lasted across many generations of RAs and we have received excellent research assistance and heroic data collection efforts from Alejandro Corvalan, Tomislav Ladika, Alex Levkov, Julia Schwenkenberg, Tobias Pfutze, and Liz Potamites. We also received very helpful comments from Andrei Shleifer, from our discussant Enrico Spolaore and participants in the UCLA Long Term Persistence Conference in May 2012 including Romain Wacziarg and David Weil, and seminar participants at Brown University, Harvard University, Johns Hopkins University, University of California, Berkeley, and Yale University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The European Origins of Economic Development
William Easterly and Ross Levine
NBER Working Paper No. 18162
June 2012, Revised July 2014
JEL No. N10,O1,O4

ABSTRACT

Although a large literature argues that European settlement outside of Europe shaped institutional, educational, technological, cultural, and economic outcomes, researchers have been unable to directly assess these predictions because of an absence of data on colonial European settlement. In this paper, we construct a new database on the European share of the population during colonization and examine its association with the level of economic development today. We find: (1) a strong and uniformly positive relationship between colonial European settlement and development, (2) a stronger relationship between colonial European settlement and economic development today than between development today and the proportion of the population of European descent today; and (3) no evidence that the positive relationship between colonial European settlement and economic development diminishes or becomes negative at very low levels of colonial European settlement, contradicting a large literature that focuses on the enduring adverse effects of small European settlements creating extractive institutions. The most plausible explanation of our findings is that any adverse effect of extractive institutions associated with minority European settlement was more than offset by other things the European settlers brought with them, such as human capital and technology.

William Easterly
New York University
Department of Economics
19 W. 4th Street, 6th floor
New York NY 10012
and NBER
william.easterly@nyu.edu

Ross Levine
Haas School of Business
University of California at Berkeley
545 Student Services Building, #1900 (F685)
Berkeley, CA 94720-1900
and NBER
Ross_levine@haas.berkeley.edu

Countries have followed divergent paths of economic development since European colonization. Some former colonies, such as the Congo, Ethiopia, Guinea-Bissau, Malawi, and Tanzania, have experienced little economic development over the last few centuries, with current per capita Gross Domestic Product (GDP) of about $2 per day. Others are among the richest countries in the world today, including Australia, Canada, and the United States, all with per capita GDP levels of greater than $20,000 per annum. Others fall along the spectrum between these extremes.

To explain these divergent paths, many researchers emphasize that the European share of the population during colonization shaped national rates of economic growth through several mechanisms. Engerman and Sokoloff (1997) (ES) and Acemoglu, Johnson, and Robinson (2001) (AJR) stress that European colonization had enduring effects on political institutions and hence on economic development. They argue that when Europeans encountered natural resources with lucrative international markets and did not find the lands, climate, and disease environment suitable for large-scale settlement, only a few Europeans settled and created authoritarian political institutions to extract resources. The institutions created by Europeans in these "extractive colonies" impeded long-run development. But, when Europeans found land, climate, and disease environments that were suitable for smaller-scale agriculture, they settled, forming "settler colonies" with political institutions that fostered development. This perspective has three testable implications: (1) a large proportion of Europeans during colonization is a precursor to successful economic development, (2) countries that had a small colonial European settlement will have lower levels of economic development today than countries with essentially no colonial European settlement due to the extractive institutions created by small European regimes, and (3) colonial European settlement will have a stronger association with development today than current European settlement (the proportion of the population that is of European descent today) because of the enduring effect of institutions created during the colonial period.

Rather than focusing on political institutions, Glaesser, La Porta, Lopez-de-Silanes, and Shleifer (2004) (GLLS) emphasize that the European share of the population during colonization influenced the rate of human capital accumulation. They argue that Europeans brought human capital and human capital creating institutions, which shape long-run economic growth, as emphasized by Galor (2011). Since human capital disseminates slowly over generations, more Europeans during colonization expedited human capital accumulation across the entire population, not just among those of European descent. This human capital view also yields three testable implications: (1) the proportion of Europeans during colonization will be positively related to human capital development and hence economic development today, (2) countries that had a small colonial European settlement will have greater levels of development today than countries with essentially no colonial European settlement, which differs from the political institutions view, and (3) the proportion of Europeans during colonization will matter more for economic development than the proportion of the population of European descent today because of the slow dissemination of human capital.

Other researchers, either explicitly or implicitly, highlight additional mechanisms through which European migration had positive or negative effects on development. North (1990) argues that the British brought comparatively strong political and legal institutions that were more conducive to economic development than the institutions brought by other European nations. This view stresses the need for a sufficiently strong European presence to instill those institutions, but does not necessarily suggest that the proportion of Europeans during colonization will affect economic development today beyond some initial threshold level. More recently, Spolaore and Wacziarg (2009) stress that the degree to which the genetic heritage of a colonial population was similar to that of the economies at the technological frontier positively affected the diffusion of technology and thus economic development, where European migration materially affected the genetic composition of economies. Putterman and Weil (2010) emphasize that the experiences with statehood and agriculture

of the ancestors of people currently living within countries help explain cross-country differences in economic success. And, Comin, Easterly, and Gong (2011) likewise find that the ancient technologies of the ancestors of populations today help predict per capita income of those populations. In both of these papers, the ancestral nature of a population helps account for cross-country differences in economic development, where European colonization materially shaped the composition of national populations. As in the human capital view, the emphasis is on things that Europeans brought with them, such as technology.

Although this considerable body of research emphasizes the role of European settlement during colonization on subsequent rates of economic development, what has been missing in the empirical literature is the key intermediating variable: colonial European settlement. Researchers have not—to the best of our knowledge—directly measured colonial European settlement and examined its association with economic development.[1]

This paper's two purposes are (1) to construct a new database on the European share of the population during colonization and (2) to evaluate the relationship between colonial European settlement and the level of economic development today. Although we do not isolate examine the specific mechanisms linking colonial European settlement with current levels of economic development as emphasized in each of the individual theories discussed above, we do assess the core empirical predictions emerging from the literature on the relationship between European settlement and economic development. In particular, we assess (1) whether the proportion of Europeans during colonization is positively related to economic development today, (2) whether the proportion of Europeans during colonization is more important in accounting for cross-country differences in current economic development than the proportion of the population of European descent today, and (3) whether countries that had a small colonial European settlement have lower levels of economic

[1] AJR looked at a variable that was the European share in 1900, but this was well after the end of the colonial period for Latin America. Our measures for Latin America are much earlier, as we explain below.

development today than countries with essentially no colonial European settlement, i.e., is there any cross-country evidence that under some initial conditions, countries with greater levels of colonial European settlement systematically have lower levels of economic development today.

To accomplish these goals, we first compile a new database on the European share of the population during colonization from an assortment of historical sources and assess different theories about the determinants of European colonization. We examine the historical determinants of European colonization both to check the credibility of our new data on colonial European settlement and to examine differing views about the factors shaping European colonization. As a guide, we employ a very simple model of the costs and benefits of European settlement. Some determinants have already been discussed in the literature, such as pre-colonial population density, latitude, and the disease environment facing Europeans. Pre-colonial population density raises the costs to Europeans of obtaining and securing land for new settlers. Latitude raises the benefits of simply transferring European technologies (such as for agriculture) to the newly settled areas. A harsh disease environment facing Europeans raises the expected costs of settlement.

To this list of common determinants of European settlement, we add one very important new variable: indigenous mortality from European diseases. Indigenous mortality from European diseases is a tragic natural experiment that is a very good predictor of European settlement, since it removed or weakened indigenous resistance to Europeans invading new lands, and made plenty of fertile land available to settlers. The phenomenon is limited to lands that had essentially zero contact with Eurasia for thousands of years, since even a small amount of previous contact was enough to share diseases and develop some resistance to them. For example, trans-Sahara and trans-Indian Ocean contacts were enough to make Africa part of the Eurasian disease pool (McNeil 1976, Karlen 1995, Oldstone 1998). Historical studies and population figures show that only the New World (the

Americas and Caribbean) and Oceania (including Australia and New Zealand) suffered large-scale indigenous mortality due to a lack of resistance to European diseases (McEvedy and Jones 1978).

This first part of the analyses provides new evidence about the factors shaping European colonization. First, we find that colonial European settlement tends to be smaller in areas where there was a highly concentrated population of indigenous people, and where the population did not die in large numbers from diseases brought by Europeans. This finding provides the first direct empirical support for AJR's (2002) hypothesis that in areas with a high concentration of indigenous people, Europeans did not settle in large numbers and instead established extractive regimes. This finding is a key building block in AJR's (2002) theory of a "reversal of fortunes," in which formerly successful areas, i.e., areas with a high concentration of indigenous people, became comparatively poorer due to the enduring effects of extractive political regimes. Second, Europeans tended to settle in large concentrations in lands further from the equator. Third, although biogeography—a measure of the degree to which an area is conducive to the domestication of animals and plants—explains human population density before the era of European colonization, it does not account for colonial European settlement. Finally, many factors—including AJR's (2000) commonly used measure of the degree to which European settlers died of disease—do not explain European settlement.

In the second part of the analyses, we assess the relationship between the European share of the population during colonization and economic development today and discover three key findings. First, colonial European settlement is strongly, positively associated with economic development today. This relationship holds after accounting for (i) British legal heritage, (ii) the percentage of years the country has been independent since 1776, (iii) the ethnic diversity of the current population, and (iv) current institutions. The relationship between economic development today and the proportion of Europeans during colonization vanishes when controlling for a measure of current human capital or a measure of government quality, which are consistent with the views that human

capital and political institutions are intermediating channels through which colonial European settlement shaped current levels of economic development.

Second, the European share of the population during colonization is more strongly associated with economic development today than the percentage of the population today that is of European descent. Europeans during the colonization era seem to matter more for economic development today than Europeans today. This finding is consistent with the view that Europeans brought growth-promoting characteristics—such as institutions, human capital, technology, connections with international markets, and cultural norms—that diffused to the rest of the population over generations. This result de-emphasizes the importance of Europeans per se and instead emphasizes the impact of what Europeans brought to economies during colonization.

The third—and perhaps most novel—result is that the positive marginal association between the European share of the population during colonization and economic development today is larger—not smaller or negative—when examining only former colonies with very few European settlers. This result does not necessarily conflict with the ES and AJR view that small European settlements fostered the creation of extractive political regimes that slowed long-run economic development. However, this result does suggest that other countervailing forces, such as the transmission and dissemination of human capital skills and technology, overcame any adverse effects from small European settlements. Even when considering countries that had zero or low European share of the population during colonization, we find no evidence that a marginal increase of Europeans during the colonial period was ever associated with lower levels of economic development today. If anything, we find that the estimated relationship between colonial European settlement and current economic development is larger when considering only countries that had low levels of European settlement. Thus, the positive relationship between colonial European settlement and economic development today is not just about the difference between settler and extractive colonies.

Ample qualifications temper our conclusions. First, we do not separately identify each potential channel through which the European share of the population during colonization shaped long-run economic development. Rather, we simply provide the first assessment of the net relationship between colonial European settlement and comparative economic development. Second, we do not assess the welfare implications of European settlement during colonization. Europeans often cruelly oppressed, murdered, and even committed genocide against indigenous populations, as well as the people that they brought as slaves (see Acemoglu and Robinson 2012 for compelling examples). Some of these victims disappeared from the population and left no descendants today. Thus, GDP per capita today cannot measure the welfare effects of European settlement; it can only measure economic activity today within a particular geographical area. Although there is no question about European oppression and cruelty, there are questions about the net effect of European colonization on the level of economic development today. In this paper, we examine the relationship between European settlement during the colonial period and economic development to inform debates about the sources of the divergent paths of economic development taken by countries around the world since the colonial period.

Our work relates also to an extensive and growing body of research on the historical determinants of economic development. Spolaore and Wacziarg (2013) summarize the first wave of studies, where in their words "history casts a very long shadow," in contrast to the previous emphasis on current policies in economic development. For example, in addition to the research discussed above, other papers address the deep historical roots of modern-day levels of social capital, civic capital, or democracy (Stephen Haber, 2014, Luigi Guiso, Paola Sapienza, Luigi Zingales, 2013, Torsten Persson and Guido Tabellini 2010, Tabellini 2010.)

The remainder of the paper is organized as follows. Section 1 defines and discusses the data, while Section 2 provides preliminary evidence on the determinants of human settlement prior to

European colonization and the factors shaping European settlement. Section 3 presents the paper's core results on the relationship between colonial European share and economic development. Section 4 reports an exercise in development accounting to calculate what share of global development can be attributed to Europeans. Section 5 concludes.

1. Data

This section describes only the two data series that we construct: (1) the European share of the population during colonization and (2) the degree to which a region experienced large scale indigenous mortality due to the diseases brought by European explorers in the 15th and 16th centuries. Since the other data that we employ in our analyses are taken from readily available sources, we define those variables when we present the analyses below.

1.1 Euro share

We compile data on the European share of the population during colonization (*Euro share*) from several sources. Since colonial administrators were concerned about documenting the size and composition of colonial populations, there are abundant—albeit disparate—sources of data. Of course, there was hardly anything like a modern statistical service in colonial times, so that different administrators across different colonies in different time periods used different and often undocumented methods for assembling population statistics. Thus, we use a large variety of primary and secondary sources on colonial history to piece together data on the European share of the population.

Although the Data Appendix provides detailed information on our sources, the years for which we compiled data on each country, and discussions about the quality of the data, it is worth emphasizing a few points here. First, we face the tricky issue of choosing a date to measure European

share. We would like a date as early as possible after initial European contact to use European settlement as an initial condition affecting subsequent developments. At the same time, we do not want to pick a date that is too early after European contact since it is only after some process of conquest, disease control, and building of a rudimentary colonial infrastructure that it became possible to speak of a European settlement. Given these considerations, we try to choose a date at least a century after initial European contact, but at least 50 years before independence. This means that for conceptual reasons we do not seek to use a uniform date across all colonies. For example, Europeans were colonizing and settling Latin America long before colonizing Africa. We also lack a continuous time series for each country; rather, the data reflect dates when colonial administrators in particular locales happened to measure or estimate populations. So we cannot always adhere strictly to our guidelines and our dates of measurement vary from the early part of the colonial period to around the date of independence.

Second, we adopt a "dog did not bark" strategy for recording zero European settlement. If we find no historical sources documenting any European settlement in a particular colony, we assume that there were no such settlers. This procedure runs the risk of biasing downward European settlement. However, we believe colonial histories (which are virtually all written by European historians) are extremely unlikely to fail to mention significant European settlements. We checked and confirmed the validity of this procedure using the Acemoglu et al. (2001) data appendix, which gives the share of Europeans in the population in 1900.

1.2 Indigenous mortality

We examine several predetermined factors that potentially influenced European settlement—including the degree to which Europeans brought diseases that wiped out the indigenous population. Others have carefully documented this tragic experience, but we believe that we are the first to use it to explain the nature of colonization and its effect on subsequent economic development.

Although Europeans established at least a minimal level of contact with virtually all populations in the world during the colonial period, this contact had truly devastating effects on indigenous populations in some regions of the world but not in others. Some regions had been completely isolated from Eurasia for thousands of years, and thus had no previous exposure or resistance to Eurasian diseases. When Europeans then made contact with these populations—which typically occurred during the initial stages of global European exploration and hence long before anything resembling "European settlements," European diseases such as smallpox and measles spread quickly through the indigenous population, decimating the indigenous people. For example, when the Pilgrims arrived in New England in 1620, they found the indigenous population already very sparse because European fisherman had occasionally landed along the coast of New England in the previous decades. Similarly, De Soto's expedition through the American South in 1542 spread smallpox and wiped out large numbers of indigenous people long before British settlers arrived.

Thus, we construct a dummy variable, *Indigenous mortality*, which equals one when a region experienced large-scale indigenous mortality due to the spread of European diseases during the initial stages of European exploration. To identify where Europeans brought diseases that caused widespread fatalities, we use the population data of McEvedy and Jones (1978) and three epidemiological world histories (McNeil 1976, Karlen 1995, Oldstone 1998). Diseases had circulated enough across Eurasia, Africa and the sub-continent, so that indigenous mortality did not shoot up with increased exposure to European explorers, traders, and slavers during European colonization.

The New World (Americas and Caribbean) and Oceania (the Pacific Islands, Australia, and New Zealand) were different. When European explorers and traders arrived, the microbes that they brought triggered extremely high mortality rates, which accords with their previous isolation from European diseases. The evidence suggests that mortality rates of 90 percent of the indigenous population after European contact were not unusual. Although we compiled a country-by-country variable for large-scale indigenous mortality, our review of the evidence and the historical narrative indicates little measurable variation within the New World and Oceania. As a result, *Indigenous mortality* wound up being a simple dummy for countries in the New World and Oceania. This suggests caution in interpreting the results on *Indigenous mortality*. Although the data indicate that large-scale indigenous mortality occurred in the New World and Oceania but not elsewhere (McEvedy and Jones, 1978, McNeil, 1976, Karlen, 1995, Oldstone, 1998), *Indigenous mortality* is ultimately a dummy variable for these regions of the world. So it also might proxy for other features of these regions, such as geographic isolation, rather than European-induced mortality.

2 Preliminaries

2.1 Where Did People Settle?

European settlers confronted a non-European world of very uneven population density. The pre-existing density had at least two material— but opposing—effects on European settlement. First, indigenous population density probably reflected the attractiveness of the land for human settlement, including Europeans. Second, indigenous population density probably reflected the potential for the indigenous people to resist European settlers.

Table 1 examines the determinants of population density in 1500, drawing on a rich and multidisciplinary literature. The dependent variable is the logarithm of population density in 1500, which we call *Population density 1500* and is taken from Acemoglu et al (2002).

We examine four potential determinants of population density in 1500. First, *Biogeography* is an index of the prehistoric (about 12,000 years ago) availability of storable crops and domesticable animals, where large values signify more mammalian herbivores and omnivores weighing greater than 45 kilograms and more storable annual or perennial wilds grasses, which are the ancestors of staple cereals (e.g., wheat, rice, corn, and barley).[2] We expect that *Biogeography* is positively associated with *Population density 1500*. Second, *Latitude* measures the absolute value of the distance of the colony from the equator. Third, *Malaria ecology* is an ecologically-based spatial index of the stability of malaria transmission in a region, where larger values signify a greater propensity for malaria transmission.[3] We do not have strong priors on the relationship between population density in 1500 and either *Latitude* or *Malaria ecology*. While *Latitude* or *Malaria ecology* might influence the suitability of a region to European settlement, it is not clear that they will shape population density before European colonization. Fourth, *Indigenous mortality* is a dummy variable that equals one if the region experienced a large drop in the indigenous population from diseases brought by Europeans. As defined above, we constructed this variable from historical sources.

We find that population density in 1500 was greater in environments that were more conducive to the domestication of animals and the cultivation of storable plants as measured by *Biogeography*. These results are robust to including the other explanatory variables. Unsurprisingly, human settlement was denser in areas where it was easier to produce food.

Malaria ecology does not have a robust, independent link with population density in 1500. This is consistent with the view that although characteristics like the prevalence of malaria might have shaped European settlement during colonization, they did not affect the population density of

[2] Taken from Hibbs and Olsson (2004), *Biogeography* equals the first principal component of (a) the number of annual perennial wild grasses known to exist in the region in prehistoric times with mean kernel weight of greater than ten milligrams and (b) the number of domesticable large mammals known to exist in the region in prehistoric times with a mean weight of more than 45 kilos.

[3] The Malaria ecology index is from Kiszewski et al (2004) and captures of the stability of malaria transmission based biological characteristics of mosquitoes such the proportion of blood meals taken from human hosts, daily survival of the mosquito, and duration of the transmission season and of extrinsic incubation.

former colonies *before* Europeans arrived. In turn, *Latitude* is negatively correlated with pre-Colombian population density, suggesting that human first settled in warmer climates.

Note that *Indigenous mortality*, which occurred after colonization, is negatively associated with population density in 1500, when considering the simple, unconditional correlation. This finding indicates that areas that were isolated from Europeans prior to colonization—and hence more susceptible to European-borne diseases—had lower population densities in 1500 AD. This may be related to Spolaore and Wacziarg's (2009) result on diffusion of technology as a function of when different branches of humanity became separated. Populations in Oceania and the Western Hemisphere had been isolated from the rest for a very long time, and hence they did not get either (1) the more advanced technology originating in the Old World that would have helped support a larger population or (2) the exposure to European diseases before colonization that would have them more resistant to European diseases and hence to European settlement. We will see that this combination of low indigenous population density and vulnerability to European diseases plays a large role in accounting for where Europeans settled.

2.2 Where Did Europeans Settle?

We now turn from the question of what shaped the settlement of humans before 1500 to the question of what shaped the settlement of Europeans during colonization. Table 2 provides regression results concerning which factors shaped European settlement during colonization, where the dependent variable is the proportion of Europeans in the colonial population (*Euro share*).

The regressors in Table 2 are as follows. First, we include *Population density 1500*. Since the regressions also include other variables to control for the attractiveness of the land for settlement, we examine the relationship between *Euro share* and population density in 1500 conditional on the generalized attractiveness of the land for human settlement. A plausible interpretation of the

conditional impact of *Population density 1500* on *Euro share* is that it gauges the ability of the indigenous population to resist European settlement. Second, *Indigenous mortality* provides additional information on the inability of the indigenous population to resist European settlers. If European diseases eliminate much of the indigenous population, this would clearly reduce their ability to oppose European settlement. Third, *Latitude* might have special relevance for European settlers to the extent that they are attracted to lands with the same temperate climate as in Europe. Fourth, *Precious Metals* is an indicator of whether the region has valuable minerals since this might have affected European settlement. Fifth, one cost of settling in a particular country might be its distance from Europe, so we use the distance from London to assess this view (*London*). Finally, we examine other possible determinants of the attractiveness of the land for settlement, including *Biogeography, Maria ecology*, and *Settler Mortality*, where *Settler mortality* equals historical deaths per annum per 1,000 European settlers (generally soldiers, or bishops in Latin America) and is taken from AJR's (2001) highly influential study of comparative economic development.

The results show that three factors account for the bulk of cross-country variation in European settlement. First, the density of the indigenous population matters. In regions with a high concentration of indigenous people who could resist European occupation, Europeans comprised a much smaller fraction of the colonial population than in other lands. Second, indigenous mortality matters. Where the indigenous population fell drastically because of European diseases, European settlers were more likely to settle. Third, there is a positive relationship between *Euro share* and *Latitude*, even when conditioning on *Population density 1500* and *Indigenous mortality*. Europeans were a larger proportion of the colonial population in higher (more temperate) latitudes, plausibly because of the similarity with the climate conditions in their home region.

These three characteristics, *Population density 1500*, *Indigenous mortality*, and *Latitude* help explain in a simple way the big picture associated with European settlements, or the lack thereof, in

regions around the world. Where all three factors were favorable for European settlement, such as Australia, Canada, New Zealand, and the United States, the European share of the colonial population was very high. When only some of the three factors were favorable, there tended to be a minority share of European settlers. Latin America suffered large-scale indigenous mortality, but only some regions were temperate, and most regions had relatively high pre-Columbian population density (which is why more people of indigenous origin survived in Latin America compared to North America, even though both regions experience high indigenous mortality rates when exposed to European diseases). Southern Africa was temperate and had low population density, but did not experience large-scale indigenous mortality. These factors can also explain where Europeans did not settle. The rest of sub-Saharan Africa was tropical and again did not experience much indigenous mortality from exposure to the microbes brought by Europeans during colonization. And, most of Asia had high population density, did not suffer much indigenous mortality from European borne diseases, and is in or near the tropics, all of which combine to explain the low values of *Euro share* across much of Asia.

Note that none of the other possible determinants that we consider are significant after controlling for these three determinants. Indeed, European colonial settlement, unlike pre-Columbian population, was NOT driven by the intrinsic, long-run potential of the land—as measured especially by *Biogeography*.

One of the most famous variables in the literature on explaining European settlement is the *Settler mortality* measure calculated by AJR. Our collection of actual data on colonial settlement allows the first assessment of the ability of this variable to explain European settlement. As shown in Table 2, *Settler mortality* does have a significant simple correlation with European settlement, confirming the prediction in AJR. But, when it is included with the three variables that we found most robust in accounting for European settlement, the *Settler mortality* becomes insignificant and

does not materially alter the statistical significance of the other variables. Apparently, *Settler mortality* does not exert an independent effect on *Euro share*, but *Indigenous mortality* does.

3. Results: Do Europeans Matter?

3.1 Do Europeans Matter? Simple graphical analyses

We begin by using graphs to illustrate the relationship between *Euro share* and the current level of economic development as measured by the average of the log of real per capita GDP over the decade from 1995 to 2005 (*Current income*). Using data averaged over a decade reduces the influences of business cycle fluctuations on our measure of current economic development.

Figure 1 shows (1) the number of countries with values of *Euro share* within particular ranges, (2) the actual countries with these particular values of *Euro share*, and (3) the corresponding median level of *Current income* for countries with values of *Euro share* within the listed ranges. Several patterns emerge. First, median *Current income* is positively associated with *Euro share*. Second, very few countries have Euro share greater than 0.125. While ES and AJR do not provide an empirical definition of settler colonies, Figure 1 suggests very few countries fall into this category. This raises questions about the relationship between *Current income* and fluctuations in *Euro share* among apparently "extractive colonies."

Figures 2a and 2b illustrate the relationship between *Current income* and *Euro share* using *Lowess*, which is a nonparametric regression method that fits simple models to localized subsets of the data and then smoothes these localized estimates into the curves provided in Figures 2a and 2b. Figure 2a illustrates the relationship for the full sample of non-European countries, while Figure 2b provides the curve for the sub-sample of countries with measured (i.e., not imputed) values of *Euro share* greater than zero and with *Euro share* less than 0.125.

As shown, there is no apparent region in which an increase in *Euro share* is associated with a reduction in *Current income*. The relationship is positive, especially for low levels of *Euro share*. This runs counter to the ES and AJR prediction that a small European settlement will hurt economic development, relative to a situation with no European settlement, because a small minority European settlement will establish extractive political regimes that stymie economic growth.

3.2 Euro share and economic development today

In this section, we use regressions to condition on a range of national characteristics and assess the independent relationship between *Current income* and *Euro share*.

We consider the following cross-country regression:

$$\textit{Current income} = \alpha * \textit{Euro share} + \beta' \boldsymbol{X} + u, \quad (1)$$

where $\boldsymbol{X}$ is a matrix of national characteristics that we define below, and u is an error term, potentially reflecting economic growth factors that are idiosyncratic to particular countries, as well as omitted variables, and mis-specification of the functional form. Different theories provide distinct predictions about (a) the coefficient on *Euro share* (α), (b) whether α changes when conditioning on particular national characteristics, and (c) how α changes across sub-samples of countries.

We get some insight into the channels connecting *Euro share* and *Current income* by examining how α changes when controlling for the different potential channels discussed above: political institutions and human capital. Thus, if *Euro share* is related to current levels of economic development through its effect on the formation of enduring political institutions, then *Euro share* will not enjoy an independent relationship with economic development today when controlling for political institutions. And, if *Euro share* is related to economic development today through its effect on the spread of human capital, then *Euro share* will not enter significantly when controlling for educational attainment today.

We begin by evaluating equation (1) while conditioning on an array of national characteristics (***X***). *British Legal origin* is dummy variable that equals one if the country has a common law (British) legal tradition. This dummy variable both captures the argument by North (1990) that the United Kingdom instilled better growth-promoting institutions than other European powers and the view advanced by La Porta et al (2008) that the British legal tradition was more conducive to the development of growth-enhancing financial systems than other legal origins, such as the Napoleonic Code passed on by French and other European colonizers. *Education* equals the average gross rate of secondary school enrollment from 1995 to 2005 and is taken from the World Development Indicators. *Independence* equals the fraction of years since 1776 that a country has been independent. As in Beck, Demirguc-Kunt, and Levine (2003) and Easterly and Levine (2003), we use this to measure the degree to which a country has had the time to develop its own economic institutions. *Government quality* is an index of current level of government accountability and effectiveness and is taken from Kaufman et at. (2002). *Ethnicity* is from Easterly and Levine (1997) and measures each country's degree of ethnic diversity. In particular, it measures the probability that two randomly selected individuals from a country are from different ethnolinguistic groups. Since the purpose of our research is to examine the impact of European settlement outside of Europe, all of the regressions exclude European countries.

Using ordinary least squares (OLS), Table 3 shows that there is—with a few notable exceptions—a positive, significant relation between *Current income* and *Euro share*. For example, regression (1) indicates that an increase in *Euro share* of 0.1 (where the mean value of *Euro share* is 0.07 and the standard deviation is 0.07) is associated with an increase in *Current income* of 0.36 (where the mean value of *Current income* is 8.2 and the standard deviation is 1.3). The strong positive link between the European share of the population during colonization and current economic development holds when conditioning on different national characteristics, with two key exceptions.

The coefficient on *Euro share* falls drastically and becomes insignificant when conditioning on either *Government quality* or *Education*. These findings are consistent with—though by no means a definitive demonstration of—the views that the share of Europeans in the population during colonization shaped long-run economic development by affecting political institutions and human capital accumulation.

These results could be driven by a few former colonies in which Europeans were a large fraction of the population during economic development and that just happen to be well-developed former colonies today. Thus, we conduct the analyses for a sample of countries in which *Euro share* was less than 12.5 percent. The goal of restricting the sample to only those countries where Europeans account for a small proportion of the population is to assess whether the relation between *Euro share* and *Current income* holds when there is only a small minority of Europeans. While there is no formal definition of what constitutes a "minority European colony," we use less than 12.5 percent European as a conservative benchmark of a non-settler colony and because there is a natural break in the distribution of *Euro share* across countries at this level.

As shown in Table 2b, however, the coefficient on *Euro share* actually becomes larger when restricting the sample to those countries in which *Euro share* is less than 12.5 percent. The increase in the coefficient on *Euro share* when restricting the sample to former colonies with small values of *Euro share* suggests that the relationship between the European share of the population during colonization and the level of economic development does not simply represent the economic success of "settler colonies." Rather, a marginal increase in *Euro share* has a bigger effect on subsequent economic development in colonies with only a few Europeans—there seems to be diminishing marginal long-run development product to *Euro share*. A marginal increase in *Euro share* is associated with an especially large boost to long-run economic growth in former colonies with only a small share of Europeans.

The relationship between *Current income* and *Euro share* is sensitive to controlling for political institutions and human capital accumulation. As shown in Table 2b, the size of the economic association between *Current income* and *Euro share* shrinks and becomes insignificant when conditioning on *Education* (regression 3) or *Government Quality* (regression 5).

The coefficient on British legal origin is never significant (nor will it be in the rest of the paper). It is also of interest that many of the colonies with Euro Share < 0.125 were Spanish colonies. Hence we find no evidence for the popular view that British colonization or legal origin led to more development than Spanish colonization or legal origin.

3.3 Is it Europeans during Colonization or Europeans today?

Euro share might proxy for the proportion of the current population that is of European descent. Figure 3 shows there is indeed a positive association between colonial *Euro share* and European share in modern times (measured in 2000 from Putterman and Weil, 2010). Consequently, it may be inappropriate to interpret the results on *Euro share* as reflecting the enduring impact of Europeans during the colonization period on economic development. Rather, Europeans might have simply migrated to economically successful countries after colonization.

To assess the strength of the independent relationship between the level of economic development today and the European share of the population during the colonial era, we therefore control for the proportion of the population today that is of European descent.

In Tables 3a and 3b, we find a positive relationship between *Current income* and *Euro share* even when controlling for the current proportion of the population of European descent, when not controlling for *Education* or *Government Quality*. *Euro 2000 P-W* is usually significant, but *Euro share* remains significant in the same regressions as in earlier results. That is, the significance (of both old and current Euro share) vanishes when we control for the channels of human capital or

political institutions, but is significant in other regressions. These results are robust to limiting the sample to having colonial *Euro share* <0.125, as reported in Table 3b.

A graph helps understand whether the proportion of Europeans during the colonization period is more strongly associated with current economic development than the proportion of the population today that is of European descent. Examining the scatter plot in Figure 3, consider three groups of countries: (1) countries in which Euro share was high both in colonial times and today (e.g. North America), (2) countries in which Euro share was low both in colonial times and today (e.g. South Africa), and (3) countries in which Euro share today is much higher than it was in colonial times (e.g. some Central and South American countries). If colonial Euro share did not matter independently for income today, then we would expect group (3)'s income to be more like group (1)'s income. But, this is not what we find. In contrast, if colonial Euro share does matter independently for income today, then we would expect group (3)'s income to have lower income than group (1) and to have similar income to group (2). This is what we observe. The proportion of Europeans during the colonization period is independently associated with economic development today, and the results are not driven by the proportion of Europeans today.

3.4 Explaining the Reversal of Fortune

In a widely cited article, Acemoglu, Johnson, and Robinson (2002) documented a "reversal of fortune:" "Among countries colonized by European powers during the past 500 years, those that were relatively rich in 1500 are now relatively poor." (p. 1231) They proxied prosperity in 1500 with two indicators -- urbanization rates and population density – and both gave the result of a strong negative correlation with per capita income today.

Acemoglu et al. (2002) argued that densely populated areas were more likely to induce Europeans to adopt extractive institutions, and these extractive institutions in turn stymied economic

development, leading to a reversal of fortune. Thus, they hypothesized that successful areas before European colonization, as measured by population density, would experience comparatively slow growth after colonization because the growth-retarding effects of the extractive political institutions created by Europeans. Acemoglu et al (2002) argued that Europeans were less likely to settle in large numbers in densely populated areas, and they associated small (or zero) European settlement with extractive institutions. They also suggested a direct positive effect of indigenous population density on the productivity of extractive institutions: there was more prosperity for Europeans to tax away for themselves, and there was a large labor force to exploit in European-owned plantations and mines.

This paper provides the first confirmation of AJR's prediction that population density in 1500 was inversely associated with the share of European settlers in the colonial population using actual data on colonial European settlement (Regressions in Table 2). We then verify that this colonial share of European settlers has a strong positive association with per capita income today (Regressions in Table 3). Hence, the results in Table 2 and 3 taken together provide an empirical explanation of the Reversal of Fortune. We have identified the key intermediating variable as colonial European settlement – lower pre-colonial population density implies more European settlers. We are unable, however, to confirm that institutions were the principle channel through which European settlement shaped the reversal of fortunes. Also in contrast to the AJR analysis, our explanation of the reversal of fortune does not require or feature a negative effect on development of a small minority of European settlers.

4. How much development is attributable to Europeans?

In this section, we do some global development accounting to illustrate how much of development might be associated with European settlers. This exercise uses the estimated equation for *Euro share* with no controls

(1) $\ln(CurrentIncome_i) = \alpha + \beta EuroShare_i + \varepsilon_i$

Next, define the counterfactual $CurrentIncome^{CF}$ for every country outside of Europe by removing the European effect:

(2) $CurrentIncome_i^{CF} = CurentIncome_i \,.\, e^{-\beta EuroShare_i}$

Of course, $CurentIncome_i = CurrentIncome_i^{CF}$ for any country i where $Euroshare_i$=0.

The counterfactual population-weighted global mean is then simply the weighted mean across all non-European countries of $CurrentIncome_i^{CF}$, where P_i is population in country *i*, and *P* is total global population.

(3) $\tilde{y}^{CF} = \sum_i \left(\frac{P_i}{P}\right) CurentIncome_i^{CF}$

The global population-weighted per capita income $\tilde{y}$ is

(4) $\tilde{y} = \sum_i \left(\frac{P_i}{P}\right) CurentIncome_i$

The share of development attributed to European settlement is then $s_e = \left(\frac{\tilde{y} - \tilde{y}^{CF}}{\tilde{y}}\right)$.

As an illustrative exercise, we use the sample and the coefficient from regression (1) of Table 2a, which is the simplest regression for the full sample of all countries outside of Europe. The coefficient estimate is β = 3.622.

Using the 2000 population weights, the data and estimated coefficients indicate that 40% of the development outside of Europe is associated with the share of European settlers during

colonization $\left(\frac{\tilde{y}-\tilde{y}^{CF}}{\tilde{y}}\right)$. We repeat our frequent caveat that global per capita income is not a welfare measure, especially in light of the history of European exploitation of non-Europeans.

As an exercise in positive analysis, however, it is striking how much of global development today is associated with the migration and settlement of Europeans during the colonial era (not even considering the development of Europe itself). It is even more striking that this large average income outcome in a non-European world today of over five billion is associated with the migration of only six million European settlers in colonial times.

5. Conclusions

The previous literature was correct to focus on colonial settlement by Europeans as one of the pivotal events in the history of economic development. In this paper, we provide the first direct evidence that the proportion of Europeans during colonization is strongly and positively associated with the level of economic development today. These findings hold when restricting the sample to non-majority-settler colonies and conditioning on the current proportion of the population of European descent.

These results relate to theories of the origins of the divergent paths of economic development followed since Europeans colonization. ES and AJR stress that when endowments lead to the formation of settler colonies, this produced more egalitarian, enduring political institutions that fostered long-run economic development. And, GLLS emphasize that Europeans brought human capital that slowly disseminated to the population at large and boosted economic development. Our results are consistent with both of these effects: former colonies with larger number of Europeans have much higher levels of economic development today than former colonies with a smaller proportion of Europeans. ES and AJR also suggest a negative effect of minority European settlement on economic development relative to a situation with no European settlement because minority

European settlements would establish extractive political institutions with enduring adverse effects on economic development. Our results, however, suggest that, on net, any negative extractive effects from minority European settlements on economic development today are dominated by other things Europeans brought with them. We find the positive effect of Europeans during colonization on economic development today becomes larger—not smaller or negative—when examining only former colonies in which the European share of the population during the colonial period was small or zero.

Bibliography

Acemoglu, D., S. Johnson and J. Robinson (2001), "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91 (5): 1369-1401.

Acemoglu, D., S. Johnson and J. Robinson (2002), "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution", *Quarterly Journal of Economics* 117(4): 1231-1294.

Acemoglu, D., S. Johnson and J. Robinson (2005), "Institutions as the Fundamental Cause of Long-Run Growth", in Aghion and Durlauf, editors, *Handbook of Economic Growth*, Amsterdam: North Holland .

Acemoglu, D. and J. Robinson, (2008), "Persistence of Elites, Power and Institutions", *American Economic Review*, 98 (1): 267-293.

Acemoglu, D. and J. Robinson (2012), *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* Crown Publishers: New York,

Albouy, D. (2012), "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment", *American Economic Review* 102(6): 3059-3076.

Baron, Castro (1942) "La poblacion de El Salvador: estudio acerca de su desenvolvimiento desde la epoca prehispanica hasta nuestros dias" (The population of El Salvador: study of their development from pre-Hispanic era to the present day). Spain: Consejo Superior de Investigaciones Cientificas.

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

Beck, T., A. Demirguc-Kunt, and R. Levine (2003),"Law, Endowments, and Finance." *Journal of Financial Economics*, 70: 137-181.

Bisin, A. G. Topa and T. Verdier (2004), "An Empirical Analysis of Religious Homogamy and Socialization in the U.S, *Journal of Political Economy*, 112(3): 615-64.

Bisin, A. and T. Verdier, "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits", *Quarterly Journal of Economics*, 115(3): 955-988.

Bockstette, V., A. Chanda, and L. Putterman (2002), "States and Markets: the Advantage of an Early Start," *Journal of Economic Growth*, 7: 347-369.

Comin, D., W. Easterly, and E. Gong (2010), "Was the Wealth of Nations Determined in 1000 B.C.?" *American Economic Journal: Macroeconomics*, 2(3): 65–97.

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

Easterly, W. and R. Levine (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics*, 112 (4): 1203-1250.

Easterly, W. and R. Levine (2003), “Tropics, germs, and crops: the role of endowments in economic development.” *Journal of Monetary Economics*, 50 (1): 3-40.

Easterly, W. (2007), "Inequality does cause underdevelopment." *Journal of Development Economics*, 84 (2): 755-776.

Engerman, S. and K. Sokoloff (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,* in Haber, editor, Stanford: Stanford University Press: 260–304

Fernandez, R. and A. Fogli (2009), “Culture: An Empirical Investigation of Beliefs, Work, and Fertility”, *American Economic Journal: Macroeconomics* 1 (1): 146–77.

Galor, G. (2011), *Unified Growth Theory*, Princeton NJ: Princeton University Press.

Glaeser, E., R. La Porta, F. Lopez-de-Silanes and A. Shleifer (2004), "Do Institutions Cause Growth?", *Journal of Economic Growth*, 9 (3): 271-303.

Guiso, L. P. Sapienza, L. Zingales (2010), “Civic Capital as the Missing Link”, in Jess Benhabib, Alberto Bisin and Matthew Jackson (editors) *Handbook of Social Economics*, Amsterdam: North Holland.

Guiso, L. P. Sapienza, L. Zingales (2013), Long-term Persistence, Einaudi Institute for Economics and Finance Working Paper 23/13, September 2013

Haber, S., (2014) “Where does democracy thrive? Climate, technology, and the evolution of economic and political institutions,” Stanford University mimeo.

Hibbs, D.A. Jr. and O. Olsson , (2004) “Geography, biogeography, and why some countries are rich and others are poor” Proceedings of the National Academy of Sciences 101 (10): 3715-372.

Hofstede, G. (2001), *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*, 2nd Edition Thousand Oaks CA: Sage Publications.

Hofstede, G and G. Hofstede (2005), “Cultures and Organizations: Software of the Mind” 2nd Edition, New York: McGraw-Hill.

Karlen, A. (1995), *Man and Microbes: Disease and Plagues in History and Modern Times* New York: Simon & Schuster.

Kaufman, D. A. Kraay, and P. Zoido-Lobaton (2002) “Governance Matters II: Updated Indicators for 2000-01”, World Bank Policy Research Working Paper 2772.

Kleibergen, F., and R. Paap (2006), “Generalized reduced rank tests using the singular value decomposition”. *Journal of Econometrics* 133: 97-126.

Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S.E. Sachs and J. Sachs (2004), "A Global Index Representing the Stability of Malaria Transmission", *American Journal of Tropical Medicine and Hygiene* 70 (5): 486-498.

La Porta, R., F. Lopez-de-Silanes, and A. Shleifer (2008), "The Economic Consequences of Legal Origins," *Journal of Economic Literature*, 46: 285–332.

Licht, A., C. Goldschmidt, and S. Schwartz (2007), "Culture rules: The foundations of the rule of law and other norms of governance", *Journal of Comparative Economics,* 35: 659–688.

Madison A. (2006), "The World Economy", Organization for Economic Co-operation and Development. Development Centre, Vol. 1-2.

McAlister L. (1984), "Spain and Portugal in the New World, 1492-1700", USA: U of Minnesota Press.

McNeil W. (1976), *Plagues and Peoples* New York: Anchor Press.

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

Miguel, E, and R. Fisman (2007), "Corruption, Norms and Legal Enforcement: Evidence from Diplomatic Parking Tickets," *Journal of Political Economy*, 115(6): 1020–1048.

Miguel, E., S. Saiegh, and S. Satyanath (2011), "Civil War Exposure and Violence," Economics and Politics 23 (1), March 2011.

North D. (1990), "Institutions, Institutional Change, and Economic Performance" UK: Cambridge University Press.

Nunn, N. (2009), "The Importance of History for Economic Development," *Annual Review of Economics,* 1(1): 65-92.

Nunn, N. (2008), "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, 123 (1): 139–176.

Oldstone M. (1998), *Viruses, Plagues, and History*, USA: Oxford University Press.

Persson, T. and G. Tabellini (2010), "Democratic Capital: The Nexus of Political and Economic Change", *American Economic Journal: Macroeconomics,* 1(2): 88–126.

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

Rosentblat A. (1954), "La poblaci n i nd gena y el mestiza e en Am rica" (The indigenous population and racial mixing in America), Buenos Aires: Editorial Nova

Spolaore E. and R. Wacziarg, R. (2009), "The Diffusion of Development", *Quarterly Journal of Economics,* 124 (2): 469-529.

Spolaore E. and R. Wacziarg, R. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature* 2013, 51(2): 325–369

Tabellini, G. 2010. "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association* 8 (4): 677–716.

Tabellini, G. (2008), "The Scope of Cooperation: Values and Incentives," *Quarterly Journal of Economics*, 123: 905–950.

Table A: Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	Obs.	Mean	Std. Dev.	Min	Max	Median
Euro Share	129	0.07	0.17	0.00	0.90	0.00
Euro 2000 Putterman Weil	115	0.13	0.24	0.00	0.90	0.00
Current Income	123	8.18	1.25	5.48	11.04	8.13
Population density 1500	94	0.50	1.52	-3.83	4.61	0.42
Indigenous Mortality	127	0.29	0.46	0.00	1.00	0.00
Latitude	129	0.20	0.12	0.01	0.67	0.18
Malaria Ecology	114	5.13	7.28	0.00	31.55	1.44
Settler Mortality	80	4.70	1.20	2.15	7.99	4.51
Biogeography	82	0.01	1.31	-1.02	3.79	-0.65
Legal Origin	129	0.40	0.49	0.00	1.00	0.00
Education	122	57.50	30.67	5.60	152.84	60.30
Independence	89	0.31	0.34	0.00	1.00	0.10
Government Quality	128	-0.47	1.93	-4.91	4.62	-0.62
Ethnicity	115	0.38	0.32	0.00	1.00	0.33

Table B: Variable Definitions

	Definition	Source
Euro Share	Proportion of Europeans in colonial population	Constructed. See Appendix for details.
Euro 2000 P-W	Proportion of Europeans in 2000 population. Constructed from Putterman and Weil's (2010) migration database by (for each country in the sample) adding the proportion of ancestors coming from each European country.	Putterman and Weil (2010)
Current Income	Ln average of GDP per capita over 1995-2005 (PPP, Constant 2005 International $)	World Bank World Development Indicators
Population density 1500	Log Population per square km in 1500	AJR (2002)
Indigenous Mortality	Dummy variable reflecting high rates of indigenous mortality from European diseases.	McEvedy and Jones (1978), McNeil (1976), Karlen (1995), Oldstone (1998)
Latitude	The absolute value of latitude in degrees, divided by 90 to be between 0 and 1	CIA World Factbook
Malaria Ecology	An index of the stability of malaria transmission based biological characteristics of mosquitoes such the proportion of blood meals taken from human hosts, daily survival of the mosquito, and duration of the transmission season and of extrinsic incubation.	Kiszewski et al (2004)
Settler Mortality	Log of potential settler mortality, measured in terms of deaths per annum per 1,000 "mean strength" (constant population)	AJR (2001)
Biogeography	The first principal component of log of number of native plants species and log number of native animals specifics, where plants are defined as " storable annual or perennial wild grasses with a mean kernel weight exceeding 10 mg (ancestors of domestic cereals such as wheat, rice, corn, and barley)" and animals are defined denotes the number of species of wild terrestrial mammalian herbivores and omnivores weighing >45 kg that are believed to have been domesticated prehistorically in various regions of the world." Hibbs and Olsson (2004) p2.	Hibbs and Olsson (2004)
British Legal Origin	A dummy variable indicating British legal origin.	La Porta et al (1999)
Education	Average rate of gross secondary school enrollment from 1995-2005	World Bank World Development Indicators
Independence	The fraction of years since 1776 that a country has been independent	Easterly and Levine (1997)
Government Quality	The first principal component of the six governance indicators from the 2002 vintage of Kaufman et al	Kaufman et al (2002)
Ethnicity	An index of ethnic diversity (updated).	Easterly and Levine (1997)

Table 1: Human Settlement before European Colonization

The sample is non-European countries. The dependent variable is the log of population density in 1500. Biogeography is an index of domesticable animals and plants existing prior to colonization. Indigenous mortality is a dummy variable which is positive if a substantial number of natives died due to initial contact with Europeans. Maria ecology is an ecologically-based spatial index of the stability of malaria transmission. All specifications are estimated using OLS with heteroskedasticity-consistent standard errors. The null hypothesis of the F test is that the coefficients on all the explanatory variables equal zero. P values are reported in parentheses. ***, ** and * represent significance at 1, 5 and 10% level respectively. More detailed variable definitions and sources are provided in Table B and the Data Appendix.

	(1)	(2)	(3)	(4)	(5)
	Population density 1500	Population density 1500	Population density 1500	Population density 1500	Population density 1500
Biogeography	0.622***				0.726***
	(0.00)				(0.00)
Latitude		-3.166*			-5.656***
		(0.10)			(0.00)
Malaria Ecology			0.0279		0.0101
			(0.10)		(0.61)
Indigenous Mortality				-1.370***	-0.509
				(0.00)	(0.19)
Observations	72	95	89	95	72
R-squared	0.17	0.055	0.018	0.183	0.409
Prob>F	0.00	0.10	0.10	0.00	0.00
F test:	19.96	2.819	2.743	22.17	7.743

Table 2: What Determined the Degree of European Settlement?

The sample is non-European countries. The dependent variable Euro share is the proportion of Europeans in the colonial population. Population density 1500 is the log of population density in 1500. Indigenous mortality is a dummy variable which is positive if a substantial proportion of natives died due to initial contact with Europeans. Latitude is the absolute value of distance from the equator. P values are reported in parentheses. ***, ** and * represent significance at 1, 5 and 10% level respectively. More detailed variable definitions and sources are provided in Table B and the Data Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Euro share	*Euro share*	*Euro share*	*Euro share*	*Euro share*	*Euro share*	*Euro share*
Population density 1500	-0.0358**	-0.0271***	-0.0272***	-0.0265**	-0.0231**	-0.0269***	-0.0318**
	(0.02)	(0.01)	(0.01)	(0.01)	(0.04)	(0.01)	(0.02)
Indigenous Mortality	0.155***	0.142***	0.142***	0.137***	0.110***	0.145***	0.0879**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
Latitude		0.673***	0.675***	0.698***	0.716***	0.691***	0.684***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Precious Metals			-0.00303				
			(0.92)				
London				0.000			
				(0.98)			
Biogeography					-0.0169		
					(0.15)		
Malaria Ecology						0.000945	
						(0.42)	
Settler Mortality							-0.0151
							(0.19)
Observations	94	94	94	90	71	88	72
R-squared	0.374	0.546	0.546	0.549	0.587	0.543	0.584
Prob>F	0.00	0.00	0.00	0.00	0.00	0.00	0.00
F test:	12.68	12.67	9.411	9.449	6.226	8.339	8.871

Table 3a: Does the degree of European settlement explain per capita income today?

The sample is non-European countries. Current income is the log of average of per capita income over 1995-2005 Euro share is proportion of Europeans in the colonial population. Legal origin is a dummy variable which is positive if a country's laws are based on the United Kingdom's legal system. Current education is the average rate of secondary school enrollment from 1998 to 2002. Independence is the fraction of years since 1776 that a country has been independent. Government quality is an index of measures of current government accountability and effectiveness. Ethnicity is a measure of a country's ethnic diversity. All specifications are estimated using OLS with heteroskedasticity-consistent standard errors. The null hypothesis of the F test is that the coefficients on all the explanatory variables equal zero. P values are reported in parentheses. ***, ** and * represent significance at 1, 5 and 10% level respectively. More detailed variable definitions and sources are provided in Table B and the Data Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)
	Current Income	Current Income	Current Income	Current Income	Current Income	Current Income
Euro Share	3.623***	3.626***	0.621	3.344***	0.511	3.437***
	(0.00)	(0.00)	(0.23)	(0.00)	(0.31)	(0.00)
British Legal Origin		-0.0024				
		(0.99)				
Education			0.0309***			
			(0.00)			
Independence				0.836**		
				(0.02)		
Government Quality					0.429***	
					(0.00)	
Ethnicity						-1.341***
						(0.00)
Observations	123	123	119	88	123	111
R-squared	0.166	0.166	0.638	0.303	0.449	0.374
Prob>F	0.00	0.00	0.00	0.00	0.00	0.00
F test:	63.09	31.32	105.4	37.12	81.85	85.57

Table 3b: Does the degree of European settlement explain per capita income today?

The sample is countries with *Euro share* values of less than 0.125. Current income is the log of average of per capita income over 1995-2005. Euro share is proportion of Europeans in the colonial population. Legal origin is a dummy variable which is positive if a country's laws are based on the United Kingdom's legal system. Current education is the average rate of secondary school enrollment from 1998 to 2002. Independence is the fraction of years since 1776 that a country has been independent. Government quality is an index of measures of current government accountability and effectiveness. Ethnicity is a measure of a country's ethnic diversity. All specifications are estimated using OLS with heteroskedasticity-consistent standard errors. The null hypothesis of the F test is that the coefficients on all the explanatory variables equal zero. P values are reported in parentheses. ***, ** and * represent significance at 1, 5 and 10% level respectively. More detailed variable definitions and sources are provided in Table B and the Data Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)
	Current Income	Current Income	Current Income	Current Income	Current Income	Current Income
Euro Share	8.378***	8.401***	-0.904	10.65***	3.612	9.846***
	(0.00)	(0.00)	(0.69)	(0.00)	(0.14)	(0.00)
British Legal Origin		-0.0365				
		(0.88)				
Education			0.0326***			
			(0.00)			
Independence				0.822*		
				(0.05)		
Government Quality					0.427***	
					(0.00)	
Ethnicity						-1.212***
						(0.00)
Observations	110	110	108	78	110	98
R-squared	0.047	0.047	0.6	0.182	0.361	0.244
Prob>F	0.00	0.00	0.00	0.00	0.00	0.00
F test:	12.72	6.328	95.92	14.68	27.53	36.09

Table 4a: Which has more of an effect on per capita income today, colonial or recent European settlement?

The sample is non-European countries Current income is the log of average of per capita income over 1995-2005. Euro share is proportion of Europeans in the colonial population. Euro 2000 P-W is the proportion of Europeans in the 2000 population (using Putterman and Weil's (2010) migration database). Legal origin is a dummy variable which is positive if a country's laws are based on the United Kingdom's legal system. Current education is the average rate of secondary school enrollment from 1998 to 2002. Independence is the fraction of years since 1776 that a country has been independent. Government quality is an index of measures of current government accountability and effectiveness. Ethnicity is a measure of a country's ethnic diversity. All regressions are OLS; P-values are reported in parentheses. ***, ** and * represent significance at 1, 5 and 10% level respectively. More detailed variable definitions and sources are provided in Table B and the Data Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)
	Current Income	Current Income	Current Income	Current Income	Current Income	Current Income
Euro Share	1.905***	1.677***	0.412	2.105***	-0.380	2.104***
	(0.00)	(0.00)	(0.48)	(0.00)	(0.55)	(0.00)
Euro 2000 Putterman-Weil	1.358***	1.476***	0.162	1.061	0.904**	1.115***
	(0.00)	(0.00)	(0.66)	(0.11)	(0.02)	(0.01)
British Legal Origin		0.129				
		(0.62)				
Education			0.0316***			
			(0.00)			
Independence				0.635		
				(0.22)		
Government Quality					0.402***	
					(0.00)	
Ethnicity						-1.120***
						(0.00)
Observations	112	112	110	85	112	102
R-squared	0.187	0.19	0.638	0.322	0.435	0.374
Prob>F	0.00	0.00	0.00	0.00	0.00	0.00
F test:	68.9	47.04	62	52.28	51.94	98.81

Table 4b: Which has more of an effect on per capita income today, colonial or recent European settlement?

The sample is non-European countries with *Euro Share*<.125. Current income is the log of average of per capita income over 1995-2005. Euro share is proportion of Europeans in the colonial population Euro 2000 P-W is the proportion of Europeans in the 2000 population (using Putterman and Weil's (2010) migration database). Legal origin is a dummy variable which is positive if a country's laws are based on the United Kingdom's legal system. Current education is the average rate of secondary school enrollment from 1998 to 2002. Independence is the fraction of years since 1776 that a country has been independent. Government quality is an index of measures of current government accountability and effectiveness. Ethnicity is a measure of a country's ethnic diversity. All regressions are OLS. P values are reported in parentheses. ***, ** and * represent significance at 1, 5 and 10% level respectively. More detailed variable definitions and sources are provided in Table B and the Data Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)
	Current Income	Current Income	Current Income	Current Income	Current Income	Current Income
Euro Share	6.455**	6.087**	-1.709	9.618***	2.329	8.702***
	(0.02)	(0.04)	(0.51)	(0.00)	(0.46)	(0.00)
Euro 2000 Putterman-Weil	1.026**	1.136**	0.287	0.313	0.757	0.701
	(0.02)	(0.03)	(0.54)	(0.73)	(0.14)	(0.13)
British Legal Origin		0.102				
		(0.71)				
Education			0.0333***			
			(0.00)			
Independence				0.771		
				(0.19)		
Government Quality					0.408***	
					(0.00)	
Ethnicity						-1.094***
						(0.00)
Observations	104	104	102	77	104	94
R-squared	0.07	0.072	0.599	0.184	0.356	0.251
Prob>F	0.00	0.00	0.00	0.00	0.00	0.00
F test:	9.794	6.648	59.24	10.72	17.72	25.22

Figure 1: Distribution of European Share at Colonization and Median Current Income

This figure shows the number of countries classified in groups according to their European shares at colonization (left axis). The median current income (in logs) for each group is also reported (right axis).

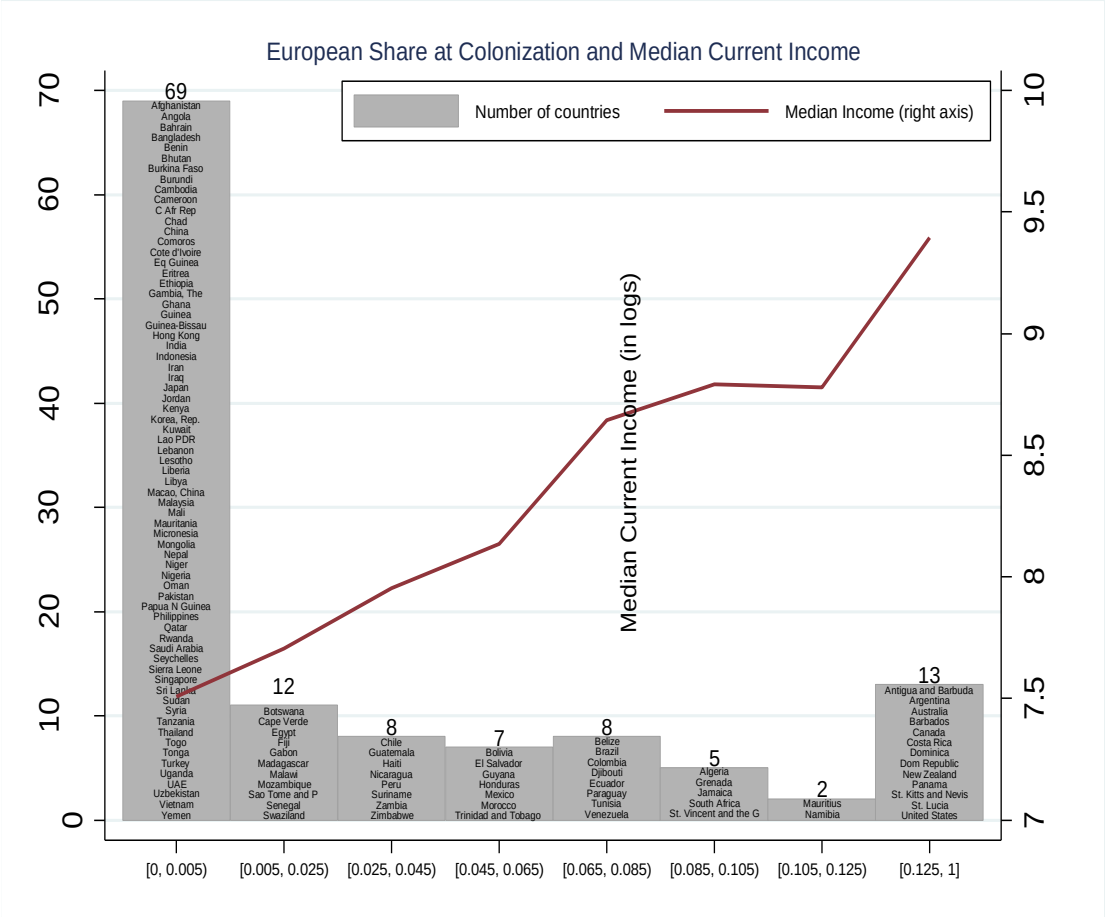
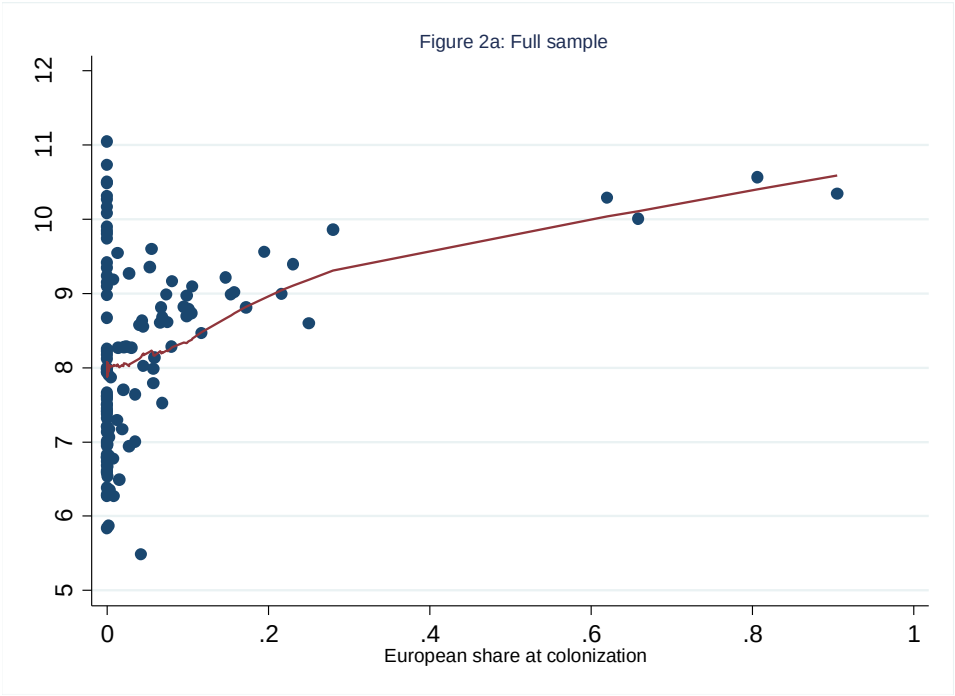


Figure 2: Current Income and European Share at Colonization

This figure presents data on current income (measured by average log of GDP per capita from 1995 to 2005) and European share at colonization. Both figures 2a and 2b present basic scatter plots with non-linear fitted values (using Stata's *lowess*). Figure 2a uses the full sample, whereas figure 2b considers only countries with 0<*Euro Share*<.125.



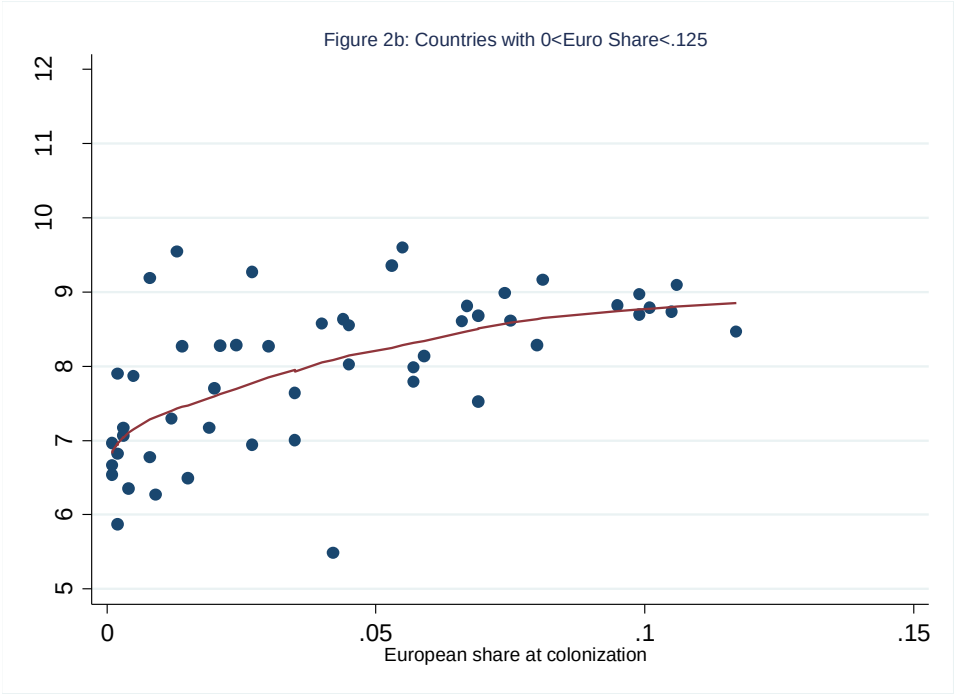
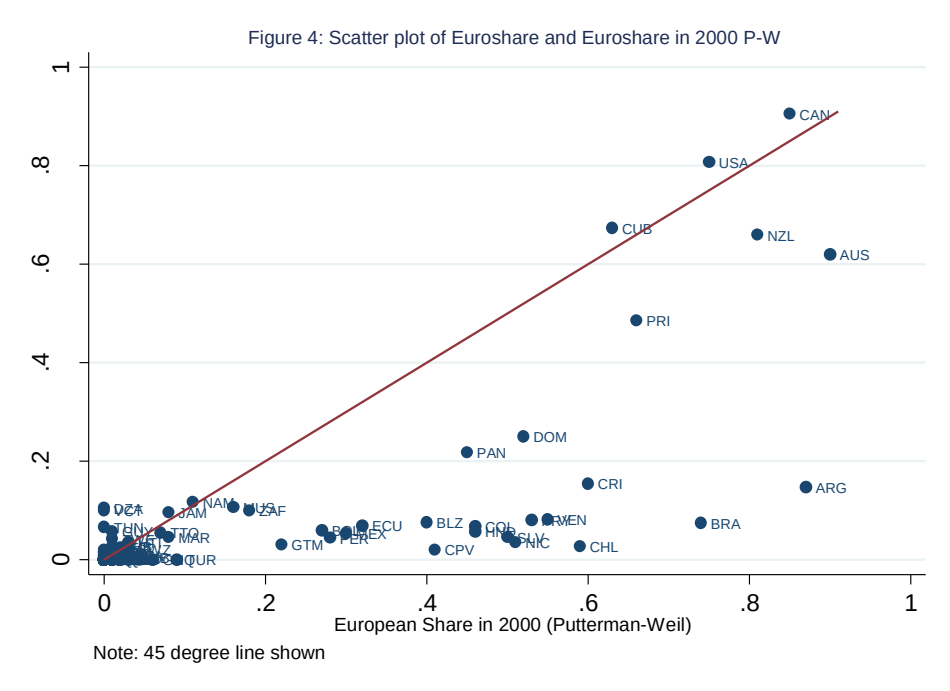


Figure 3: Colonial European Share and European Share Today

This figure shows a simple scatter plot comparing the proportion of Europeans in the Colonial population with the same proportion in 2000.



Data Appendix

This appendix describes the construction of the dataset on the European share of the population in countries around the world during colonization. The primary goal is to define what we did, so that the numbers are transparent and replicable. At the end of the appendix, we list some of the problems that we encountered in constructing the database, and the difficulties that we faced in choosing which years to use in defining a country's "European share of the population during colonization." The dataset and other key information are contained in the excel workbook titled "Appendix Europeans," which is available on request. In this Appendix when discussing details of the dataset, we refer to specific worksheets within this workbook. We also have created Stata do files that replicate the results in the tables; these files and the full dataset is also available on request.

Data sources and definitions

We primarily rely on 46 sources, which are listed in the worksheet titled "bibliography" and the worksheet titled "web." Many of these are scholarly books about particular regions or countries and some are atlases. As a few examples, Robert Wells wrote *The Population of the British Colonies in America before 1776*; Simeon Ominde wrote *The Population of Kenya, Tanzania, and Uganda*; and, McEvedy and Jones assembled *Atlas of World Population History*. We also use primary sources (such as national and colonial censuses) to both check these sources and to expand the number of countries and data points. Besides the books and official documents listed in the worksheet "bibliography," some datasets are provided online. We list these in the worksheet "web."

In terms of defining "European settlers," we strive in collecting the data to identify Europeans as people from the geographic region of Europe; it is NOT a racial or ethnic description. So, some observers might consider the populations of some countries outside of Europe as racially or ethnically equivalent to Europeans, but that is irrelevant to us. We are only concerned with resettlement from

Europe to outside Europe, defined geographically. Furthermore, in assembling the data on settlers, we strive to exclude colonial officials or business people that are temporarily stationed abroad; we strive to only include Europeans who permanently resettle outside of Europe.

Data: By country, year, and source

For each country, we provide an entry for each year for which we found information on the European share of the population. For each data point, we provide the source of the information (including the page number) and brief notes about the data, whenever relevant. Some of these notes are important. For example, the 1744 and 1778 values for Argentina are based only on the population around Buenos Aires, for the country as a whole. Similarly, one of the values for Ecuador in 1781 measures only the population around Quito. These notes, the data, and the sources of each data point are listed in the worksheet, "country_year_source."

For example, Lyle N. McAlister, in Spain and Portugal in the New World, 1492 -1700, (University of Minnesota Press, 1984) provides data on page 131 on the population of Argentina in 1570. He notes that there are 2000 whites, 4000 blacks, 300,000 "others" in Argentina. Since whites are typically used to describe people of European descent, we calculate European share in Argentina in 1570 as 0.0065.

In some cases, the data sources provide a range of years (rather than a single year) corresponding to data on the share of the population that is European. For example, one of the observations on Mexico is listed as 1568-1570 in the underlying data. In these cases, we choose the average of the range and enter this as the year for the observation. Thus, for the Mexico example, we choose the year 1569 when entering this data. All of these cases are separately identified in the worksheet "periods." This has no bearing on our analyses, but might be relevant for others that use these data.

In a few cases, we found two data sources that provide information on the same year (or range of years) for the same country. In some cases, these two data sources agree, in which case we simply report both observations within the worksheet "country_year_sources." In a few cases, the data sources give different numbers for the share of Europeans in a country in a given year. In this case, we report both numbers in the worksheet "country_year_sources" and use the average of the two observations when constructing the data on which we conduct our analyses.

Data: Share of Europeans used in the analyses

From these data scattered over many years since the 16th century, we construct several measures of the share of Europeans during colonization for each country, where we have one measure per country. To do this, we arrange the data in a manner that is amenable for the construction of a single measure of the share of Europeans during colonization for each country. In the worksheet "euro share," each row is a country. The columns provide possible data entries for many years running from 1540 (which is our first observation, for Chile) through to the late 20th century. We do not include all years as column headings; rather, we only include years for which we have at least one non-missing value for one country.

First, we construct simple, objective measures that average the value of each country over particular periods. For example, we average each country's entries over the period from 1500 to 1801; and, we average values over the period from 1700 to 1950. These measures are provided in the worksheet titled "euro share." Other researchers can obviously take these data and use whichever periods they find most appropriate.

These simple, objective measures for computing the share of Europeans during colonization, however, have some limitations. Specifically, averaging over uniform time periods for all countries might not create accurate measures for each particular country of the proportion of the population that

is European during a colony's formative period—the period when a colony was creating an initial set of (potentially enduring) political, educational, and cultural institutions. We fully recognize that there is not a precise definition of "the" formative period of colonization. Nevertheless, influential studies of comparative economic development emphasize the potential role of Europeans during a colony's history when it establishes major institutional norms. This motivates our efforts to give empirical substance to this amorphous notion. From this perspective, using the European share of the population of Mexico in 1650 might be more appropriate than using the share in 1850, but using the European share of the population in 1650 in some parts of Sub-Saharan Africa (or other parts of the world) might be inappropriate because European colonization evolved differently there. Thus, we face a challenging goal: account for these historical differences in the timing and process of colonization to construct a more conceptually useful measure of "euro share" for each country. We face this challenge while operating under severe data constraints.

Thus, the second method for constructing a measure of each country's European share of the population during the formative years of colonization attempts to select the best year, or range of years, given the particulars of the country and data availability. We would like a date as early as possible after initial European colonization to use European settlement as an initial historical condition affecting subsequent developments. At the same time, we can't pick a date that is too early after the start of European colonization. It was only after some process of conquest, disease control, and building rudimentary colonial infrastructure that it becomes possible to speak of a European community that might influence economic, political, and cultural conditions. Given these considerations, it would not make sense to use a uniform date across all colonies.

Thus, we formulated the following "guidelines." Sub ect to data limitations, we tried to constrain the timing of the European measure to be at least a century after initial European contact.

Furthermore, we tried to choose a date that was at least 50 years before independence to measure the colonial period. Finally, if there were a few measures close together, we took the average.

In the worksheet "euro share," we provide a measure of each country's European share of the population – *euro share*—that represents our assessment of the best year, or range of years, for measuring European share during the formative years of colonization for each particular country, where this assessment is almost always done subject to extreme data constraints. When using a range of years, we average to compute *euro share*. The worksheet also provides the year, or range of years, used to compute *euro share*. This second method is neither simple nor fully objective. Though we do our best to follow the guidelines outlined above, data limitations and the idiosyncrasies of colonial histories make things complex and subjective. Nevertheless, we believe *euro share* is a more accurate representation of the share of Europeans during the formative years of colonization that simply averaging over a uniform time period for all former colonies.

Though the excel spreadsheet "euro share" provides the details for each country, it is valuable to illustrate some of the constraints that we face and the choices that we made. For much of Latin America, Angel Rosentblat (1954) provides detailed estimates of the composition of the population in 1650. We have used these estimates when available. In many countries, the next available observation is not until a century (or more) later. For example, after 1650, the next observation is not until 1798 in Brazil, 1940 in Bolivia, 1777 in Mexico, and 1744 in Argentina. For some of these countries, we have earlier population estimates that are reported in the excel file. For example, we have observations in 1570 for Argentina, Bolivia, Brazil, and Mexico. But, following the guidelines sketched above, we determined that this was too early after Europeans first arrived.

There are other problems, some of which force us to break with the "guidelines" sketched above. For example, the first number that we have for the United States is the 1790 census, which is obviously not fifty years before the country became independent. Similarly, we do not have data on

the composition of the population before 1950 for several countries in Africa, including the Democratic Republic of Congo, Djibouti, Gabon, Sao Tome and Principe, Senegal, Tunisia, and Rwanda. Jamaica and El Salvador provide some particular challenges. For Jamaica, the numbers on European share of the population show considerable variability over the period from 1570 to 1673; but then, the numbers (provided by various sources) are quite consistent from 1700 through 1943. So, to compute *euro share*, we take the average over the period from 1700-1750. For El Salvador, there is one extremely large observation in 1796 (0.48) that deviates from other estimates in nearby years (e.g., 0.03 in 1807) provided by the same source (Baron Castro, 1942). Since (1) the estimates for *euro share* over the entire period with available data from 1551 to 1950 are reasonably constant except for this one observation 1796 and (2) there seems to be a change in the definition of "white" for this particular year, we decided not to include this observation in the "euro share" worksheet. For El Salvador, therefore, we compute the average over the period from 1551-1807, excluding this 1796 outlier because of the change in definition.

Data: Countries in which Europeans did not settle

There are many countries in which Europeans did not settle to any appreciable degree. In these countries, unsurprisingly, we do not find historical sources documenting the share of Europeans during the colonial period. Thus, we face a problem: We do not want eliminate these countries from the sample when we know that European settlers were not a material part of their history, but we do not have documentation to that effect. Thus, although we cannot strictly prove that there were no Europeans, available evidence suggests that Europeans did not settle everywhere and we can incorporate this information into our analyses.

We follow the following procedure. We conduct a worldwide search for colonial data on European settlement. Besides the sources listed in the workbook, we examine many additional sources in search of data. When we fail to find any mention of European settlement in any of these

sources for a given nation, we coded that country as having zero European settlement. This procedure runs the risk of biasing downward European settlement for these countries. But, colonial histories seem unlikely to ignore material European settlement. We confirm our data using information from Acemoglu et al. 2001 on the European share of the population in 1900.

Problems: A brief discussion of a few of the problems

There are many challenges associated with constructing this database on the European share of the population during the period of colonization. First, although colonists did document the number of Europeans in the total population at various points during colonization, the processes and periodicities were not standardized. We just do not know if the same colonial power used the same methods in different countries in different years, not to mention differences across European powers. For example, different colonial powers probably used different methods to estimate population numbers. In a census-based method, there could be an undercount of non-European populations, which would bias population numbers downward and European shares upward. In a sampling methodology, there is zero expected bias only if the samplings were random. Unfortunately, we have almost no information on methods followed to get these population numbers. Put simply, we do not have a continuous time series of data collected using similar measures by a centralized coordinating entity.

Second, most of the cells in the data running from the 16th century to the 20th century are empty. This means that for many countries we cannot get measures of the share of Europeans in the population within decades of the years that we would ideally want to measure *euro share*. Although most countries do not experience huge changes in the share of Europeans, some experience substantial changes, so this is another challenge facing the construction of the database.

Third, the basic conceptual predictions about the role of Europeans during colonization and their enduring influence on economic development do not provide a concrete definition of when to measure the share of Europeans during colonization. We try to measure the share of Europeans about a century after the start of colonization and 50 years before independence, but this simply represents the articulation of a hopefully helpful empirical guideline. Without ignoring these—and other—challenges, we constructed the database and use it to provide empirical evidence on the relationship between the share of Europeans during colonization and comparative economic development.