Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality

Louis Putterman
Brown University

David N. Weil*
Brown University and NBER

September, 2008

Abstract

We construct a matrix showing the share of the year 2000 population in every country that is descended from people in different source countries in the year 1500. Using this matrix, we analyze how post-1500 migration has influenced the level of GDP per capita and within-country income inequality in the world today. Indicators of early development such as early state history and the timing of transition to agriculture have much better predictive power for current GDP when one looks at the ancestors of the people who currently live in a country than when one considers the history on that country's territory, without adjusting for migration. Measures of the ethnic or linguistic heterogeneity of a country's current population do not predict income inequality as well as measures of the ethnic or linguistic heterogeneity of the current population's ancestors. An even better predictor of current inequality in a country is the variance of early development history of the country's inhabitants, with ethnic groups originating in regions having longer histories of agriculture and organized states tending to be at the upper end of a country's income distribution. However, high within-country variance of early development also predicts higher income per capita, holding constant the average level of early development.

Keywords: Economic Growth, Migration, Income Inequality, State History, Linguistic Distance
JEL Codes: F22, N30, O40

---

* E-mail: Louis_Putterman@Brown.Edu and David_Weil@Brown.Edu. We thank participants at the Brown Macro Lunch and the NBER Summer Institute for helpful comments and Bryce Millett, Momotazur Rahman, Isabel Tecu, Ishani Tewari, Yaheng Wang, and Joshua Wilde for valuable research assistance.

Economists studying income differences among countries in the world today have been increasingly drawn to examine the influence of long-term historical factors. While the theories underlying these analyses vary, the general finding is that things that were happening 500 or more years ago matter for economic outcomes today. Hibbs and Olsson (2004, 2005), for example, find that a region-level indicator of timing of the Neolithic revolution explains differences among countries in incomes and the quality of institutions in 1997. Comin, Easterly, and Gong (2006) show that the state of technology in a country 500, 1500, or even 3000 years ago has predictive power for the level of output today. Bockstette, Chanda and Putterman (2002) find that an index of the presence of state-level political institutions from year 1 to 1950 has positive correlations, significant at the 1% level, with both 1995 income and 1960-95 income growth. And Galor and Moav (2007) provide empirical evidence for a link from the timing of the transition to agriculture to current variations in life expectancy.

Examining this sort of historical data immediately raises a problem, however: the further back into the past one looks, the more the economic history of a given *place* tends to diverge from the economic history of the *people* who currently live there. For example, the territory that is now the United States was inhabited in 1500 largely by hunting, fishing, and horticultural communities with pre-iron technology, organized into relatively small, pre-state political units.[1] By contrast, a large fraction of current U.S. population is descended from people who in 1500 lived in settled agricultural societies with advanced metallurgy, organized into large states. The example of the United States also makes it clear that, because of migration, the long-historical background of the people living in a given country can be quite heterogeneous. This observation, combined with the finding that long-history of a country's residents affects the average level of income, naturally raises the question of whether heterogeneity in background of a country's residents is a determinant of income inequality within the country.

Previous attempts to deal with the impact of migration in modifying the influence of long-term historical factors have been somewhat *ad hoc*. Hibbs and Olsson, for example, acknowledge the need to account for the movement of peoples and their technologies, but do so only by treating four non European countries (Australia, Canada, New Zealand and the U.S.) as if they were in Europe. Comin, Easterly, and Gong similarly add dummy variables to their regression model for countries with "major" European migration (the four mentioned above) and "minor" European migration (mostly in Latin America). In other cases, variables meant to measure other things may in fact be proxying for migration. For example, the measure of the origin of a country's legal systems examined by La Porta *et al*. (1998) may be proxying for the origins of countries' people. This is also true of Hall and Jones's (1999) proportion speaking European languages measure. The apparent effect of institutions that were either brought along by European settlers or imposed by non-settling colonial powers, as found in Acemoglu, Johnson, and Robinson (2001, 2002), may be proxying for population shifts themselves,

[1] Anthropologists subscribing to cultural evolutionary models speak of political institutions evolving from the band to the tribe to the chiefdom and finally the state (see, for instance, Johnson & Earle, 1987). There were no pre-Columbian states north of the Rio Grande, according to such schema.

despite their attempt (discussed below) to control for the European-descended population share.

In this paper we pursue the issue of migration's role in shaping the current economic landscape in a much more systematic fashion than previous literature. We construct a matrix detailing the year-1500 origins of the current population of almost every country in the world. (Throughout the paper, we use the term "migration" to refer to any movement of population across current nation borders, although we are cognizant that these movements included transport of slaves and forced relocation as well as voluntary migration.) We then use this matrix as a tool to examine how early development and the pattern of population movements across borders have impacted current income and inequality.

The most thorough previous work along these lines is in the papers by Acemoglu, Johnson, and Robinson (AJR) mentioned above, where they calculate the share of the population that is of European descent for 1900 and 1975. There are a number of conceptual and operational differences between our approach and theirs. Our estimates break down ancestor populations much more finely than "European" and "non-European." This distinction is important both in the Americas, where there is great variation in the fraction of the population descended from Amerindians vs. Africans, and also in other regions, where important non-native populations are not descended from Europeans (consider the large Chinese-descended populations in Singapore and Malaysia, or Indian descendants in South Africa, Malaysia, and Fiji). Even when we use our matrix to construct a measure of the European population fraction, there are considerable differences between our data and AJR's. They use as their measure of the European population the fraction of people who are "white," while we also include an estimate of the fraction of European ancestors among mestizo populations, which is generally between 30 and 50%. In Mexico, for example, AJR estimate the European population in 1975 to be 15%, even though (in their data) there is an additional 55% of the population that is mestizo. Our estimate of the European share of ancestors for today's Mexicans is 29%. The AJR estimates are primarily based on data in McEvedy and Jones (1978), which sometimes apply to whole regions, and occasionally involve extrapolation from as far in the past as 1800. Our data are based on a broader selection of more recent sources, including encyclopedias, government reports, and compilations by religious groups, which are summarized in Appendix A and Putterman (2006). The correlation between our measure of the European fraction and the AJR measure is 0.89.[2]

The rest of this paper is structured as follows. Section 1 describes the construction of our migration matrix, and then uses the matrix to lay out some of the

[2] The largest differences occur in the Americas. For example, for the five Central American countries of El Salvador, Nicaragua, Panama, Costa Rica, and Honduras, AJR use a uniform value of 20% European; our estimates range from 45% in Panama to 87% in Costa Rica. The largest outlier in the other direction is Trinidad and Tobago, which they list as 40% European and is only 7% in our measure. Here they seem to have erroneously counted all non-Africans as European, despite the presence of a large Asian population.

important facts regarding the population movements that have reshaped genetic and cultural landscapes in the world since 1500. We find that a significant minority of the world's countries have populations mainly descended from the people of other continents and that these countries themselves are quite demographically heterogeneous. In Section 2, we apply our migration matrix to analyze the determinants of current income. Using several measures of early development, we show that adjusting the data to reflect where people's ancestors came from improves the ability of measures of early social and technological development to predict current levels of income. In Section 3, we turn to the issue of inequality. We use the migration matrix to construct various measures of the heterogeneity of countries' populations in terms of the early development of the countries their ancestors came from, and show that these heterogeneity measures predict income inequality. We also show that ethnic groups originating in regions with higher levels of early development tend to be placed higher in a recipient country's income distribution. In Section 4, we examine the effect on current income of heterogeneity in early development. We find that, holding constant the average level of early development, heterogeneity in early development raises current income, a finding which we interpret as indicating spillovers of growth-promoting traits among national origin groups. Section 5 concludes.

1. Large-scale population movements since 1500

We use the year 1500 as a rough starting point for the era of European colonization of the other continents. It is well known that most contemporary residents of countries such as Australia and the United States are not descendants of their territory's inhabitants circa 1500, but of people who arrived subsequently from Europe, Africa, and other regions. But exactly what proportions of the ancestors of today's inhabitants of each country derive from what regions and from the territories of which present-day countries has not been systematically studied. Detailed genetic studies are thus far too sparse to provide the required data. Accordingly, we examined a wide array of secondary compilations to form the best available estimates of where the ancestors of the long-term residents of today's countries were living in 1500. Generally, these estimates have to work back from information presented in terms of ethnic groupings. For example, sources roughly agree on the proportion of Mexico's population considered to be mestizo, that is having both Spanish and indigenous ancestors, on the proportion having exclusively Spanish ancestors, on the proportion exclusively indigenous, and on the proportion descended from migrants from other countries. There is similar agreement about the proportion of Haitians descended from African slaves, the proportion of people of (east) Indian origin in Guyana, the proportion of "mixed" and "Asian" people in South Africa, and so on. Such information plus information helpful to the decomposition of mixed categories—for instance, an archive on the slave trade permitting estimates to be made of the proportion of slaves in a given region who originated from a certain part of Africa identifiable with certain present-day countries—makes possible estimates of the proportion of a current population's ancestors likely to have lived in 1500 in the territory of other specific contemporary countries.[3]

Using these methods, a matrix of migration since 1500 was constructed. It has 165 complete rows, each for a present-day country, the same number of complete columns, representing the same countries, and its entries are the proportion of long-term residents' ancestors estimated to have lived in each source country in 1500, summing to one. Appendix A briefly describes our sources and methods, with the appendices of Putterman (2006) providing further details including written summaries of the factors behind the estimate for each row.

The principal diagonal of the matrix provides a quick indication of differences in the degree to which countries are now populated by the ancestors of their historical populations. The diagonal entries for China and Ethiopia (with shares below half a percent being ignored) are 1.0, while the corresponding entries for Cuba, the Dominican Republic and Haiti are 0.0 and that of Paraguay is close to 0.5. In some cases, the diagonal entry may give a misleading impression without further analysis; for example, the diagonal entry for Botswana is 0.31 because only 31% of Botswanans' ancestors are estimated to have lived in present-day Botswana in 1500, but another 67% were Africans who migrated to Botswana from what is now neighboring South Africa in the 17th and 18th centuries. (Note that the use of 1500 as our starting point means that the ancestors of white South Africans are attributed to Netherlands and other European countries, not South Africa.)

Figures 1a and 1b are histograms of the proportion of countries and people, respectively, falling into decile bands with respect to the proportion of the current people's ancestors residing in the same or an immediate neighboring country in 1500. The figures show bimodal distributions, with 11.5% of countries having 0 to 10% indigenous or near-indigenous ancestry and 67.9% of countries having 90 to 100% such ancestry. Altogether, 80.1% of the world's people (excluding those in the smallest countries, which are not covered) live in countries that are more than 90% indigenous in population, while 10.1% live in countries that are less than 30% indigenous, with the rest (dominated by Central America, the Andes, and Malaysia) falling in between.

The compositions of non-indigenous populations are also of interest. The populations of Australia, New Zealand and Canada are overwhelmingly of European origin, while Central American and Andean countries have both large Amerindian and substantial European-descended populations, and Caribbean countries and Brazil have substantial African-descended populations. Guyana, Fiji, Malaysia and Singapore are among those countries with substantial minorities descended from South Asians, while Malaysia and Singapore also have large Chinese-descended populations. We illustrate differences both in the proportions of people of non-local descent and in the composition of those people by means of Map 1. Country shading indicates the proportion of the

[3] The use of categories such as "mestizo" and "colored" in the sources from which the matrix was constructed necessitated guesses as to the proportion of ancestors of such groups belonging to each of two or more contributing populations. See the Appendix and Putterman (2006).

population not descended from residents of the same or immediate neighboring countries. Pie charts, drawn for thirteen macro-regions, show the average proportions descended from European migrants, from migrants (or slaves) from Africa, and from migrants from other regions, as well as the proportion descended from people of the same region.[4] In terms of territory, about half the world's land mass (excluding Greenland and Antarctica), comprising almost all of Africa, Europe and Asia, is in countries with almost entirely indigenous populations (shown in black), while about a third has less than 20% indigenous inhabitants, and the remainder, dominated by Central America, the Andes and Malaysia, falls somewhere in between. The heterogeneity of regions in the Americas and Australia/New Zealand is highlighted by the pie charts, showing strong European dominance in Australia/New Zealand, the U.S., Canada, and eastern South America, stronger indigenous presence in the Andes, and strong African representation in the Caribbean. We consider the effects of this heterogeneity in Section 3.

While we are mostly interested in using the migration matrix to better understand the determinants of long-run economic performance in countries as presently populated, the versatility of the data can be illustrated by using it to calculate the number of descendants of populations that lived five centuries ago and to see how they've fared. Given data on country populations in 2000, the matrix will tell total number of people today who are descended from each 1500 source country, and where on the globe they are to be found. For instance, using 2000 population figures from Penn World Tables 6.2 for the matrix's 165 countries, we find that there were 31.6 million descendants of 1500's Irish alive at the turn of the millennium, of whom 11.8% lived in Ireland itself, 76.2% in the U.S., 1.0% in the U.K. and 5.2% in Australia. According to the matrix and the sources it's based on, there are essentially no descendants of the indigenous population of Hispaniola (today's Haiti and Dominican Republic), since the Arawak people who lived there died out during the early decades of colonial rule due to disease and the effects of enslavement.

Combining the information in the matrix with population data for the years 1500 and 2000 yields a number of interesting insights. Because population data for 1500 are very noisy, particularly at the country level, we confine our analysis to looking at 11 large regions (data are from McEvedy and Jones, 1978). The first two columns of Table 1 list the estimated population of each region in 1500 and 2000. The third column shows the increase in total population over the 500 year period. The primary determinant of this

---

[4] Regions were defined with the aim of keeping their number small enough for purposes of display and grouping countries with similar population profiles. The Caribbean includes Cuba, Dominican Republic, Haiti, Jamaica, Puerto Rico, and Trinidad and Tobago. Europe is inclusive of the Russian Republic. North Africa, West and Central Asia includes all African and Asian countries bordering the Mediterranean, including Turkey, the traditional Middle East, Afghanistan, and former Soviet republics in the Caucasus and Central Asia. South Asia includes Pakistan, India, Bangladesh, Sri Lanka, Nepal and Bhutan. East Asia includes Mongolia, China, Hong Kong, North and South Korea, Japan, and Taiwan. Southeast Asia includes the remainder of Asia plus New Guinea and Fiji. Note that for calculation of the pie chart shares, ancestors are assumed to be from "the same region" if they are from countries in the regions as thus indicated. This assumption means that Europeans are left out of the "European migrant" category of the pie charts if they live in Europe, even if they've migrated within the continent, and likewise for sub-Saharan Africans in SSA.

increase in density is the level of economic development in 1500. Europe, East Asia, and South Asia, which were highly developed, had the smallest increases in density. The U.S. and Canada, Australia and New Zealand, and the Caribbean, which were relatively lightly populated, lacked urban centers, and were still home to many pre-agricultural societies in 1500, had the largest increases.[5] The next three columns of the table use the matrix to track the relationship between ancestor and descendant populations. In column 4, we calculate the number of descendants per capita for each region in 1500, which can be thought of as a kind of "genetic success" quotient. The lowest values of this measure are in the Caribbean and Australia and New Zealand, where native populations were largely displaced. by European colonizers. Among the regions that were relatively developed in 1500, Europe, not surprisingly, has the highest number of descendants per capita. However, Europe's figure is exceeded by those for Mexico and Central America, sub-Saharan Africa, and above all Southeast Asia, which were regions that were relatively poor (and thus somewhat less densely populated) in 1500 but in which the native population was either not entirely or hardly at all displaced by migrants. Column 5 calculates the fraction of the current regional population that is descended from the region's own 1500 ancestors. This ranges from a zero for the Caribbean to almost one for South Asia and East Asia. Finally, column 6 shows the fraction of descendants of the 1500 population that still live in the same region. Disregarding the case of the Caribbean, this is lowest for Europe, which has retained only 66 percent of its descendants. The second largest exporter of people, by this measure, is sub-Saharan Africa, followed by Mexico and Central America. In every other region, there has been little significant export of population.

2. *Reassessing the Effects of Early Economic Development*

In the introduction, we noted that studies including Hibbs and Olsson (2004, 2005), Comin, Easterly, and Gong (2006) and Chanda and Putterman (2006) find strong correlations between measures of early agricultural, technological, or political development and current levels of economic development, but that these studies make relatively *ad hoc* adjustments, if any, to account for the large population movements on which this paper focuses. The new migration matrix puts us in a position to remedy these shortcomings and thereby put the theory that very early development persists in its effects on economic outcomes to a more stringent test.

We use two measures of early development. The first is an index of state history called *statehist*. The index takes into account whether what is now a country had present a supra-tribal government, the geographic scope of that government, and whether that government was indigenous or by an outside power. The version used by us, as in

[5] Estimates of pre-Columbian population in the Americas are highly controversial due to considerable uncertainty about the death rates in epidemics that followed European contact. Since McEvedy and Jones's estimates fall toward the low end of some more recent appraisals, the resulting estimates of the increase in population density since 1500 could be overstated.

Chanda and Putterman (2006, 2007), considers state history for the fifteen centuries to 1500, and discounts the past, reducing the weight on each half century before 1451-1500 by an additional 5%. Let s_{it} be the state history variable in country *i* for the 50 year period *t*. s_{it} ranges between 0 and 50 by definition, being 0 if there was no supra-tribal state, 50 if there was a home-based supra-tribal state covering most of the present-day country's territory, 25 if there was supra-tribal rule over that territory by a foreign power, and taking values ranging from 15 (7.5) to 37.5 (18.75) for home- (foreign-) based states covering between 10 and 50% of the present-day territory or when several small states co-exist on that territory. *statehist* is computed by taking the discounted sum of the state history variables over the thirty half centuries and normalizing it to be between 0 and 1 (by dividing it by the maximum achievable, i.e. the *statehist* value of a country that had s_{it} = 50 in each period). In a formula:

$$statehist = \frac{\sum_{t=0}^{29} (1.05)^{-t} s_{i,t}}{\sum_{t=0}^{29} (1.05)^{-t} 50}$$

For illustration, Ethiopia has the maximum value of 1, China's *statehist* value is 0.906 (due to periods of political disunity), Egypt's value is 0.760, Spain's 0.562, Mexico's 0.533, Senegal's 0.398, and Canada, the U.S. Australia and New Guinea have *statehist* values of 0.[6]

Our second measure of early development, *agyears*, is the number of millennia since a country transitioned from hunting and gathering to agriculture. Unlike a similar measure used by Hibbs and Olsson, which had values for nine macro regions, these data are based on individual country information augmented by extrapolation to fill gaps within regions. The data were assembled by Putterman with Trainor (2006) by consulting region- and country-specific as well as wider-ranging studies on the transition to agriculture, such as MacNeish (1991) and Smith (1995). The variable *agyears* is simply the number of years prior to 2000, in thousands, since a significant number of people in an area within the country's present borders are believed to have met most of their food needs from cultivated foods. The highest value, 10.5, occurs for four Fertile Crescent countries (Israel, Jordan, Lebanon and Syria) followed closely by Iraq and Turkey (10), Iran (9.5), China (9) and India (8.5). Near the middle of the pack are countries like Belarus (4.5), Ecuador (4), Ivory Coast (3.5) and Congo (3). At the bottom are countries like Haiti and Jamaica (1) which received crop-growing immigrants from the American mainland only a few hundred years before Columbus, New Zealand (0.8), which obtained agriculture late in the Austronesian expansion, and Cape Verde (0.5), Australia (0.4) and others in which agriculture arrived for the first time with

[6] Bockstette *et al*. (2002) and Chanda and Putterman (2006) also use versions of *statehist* that include data for the years between 1501 and 1950. The variable that we call *statehist* in this paper is the same as what Chanda and Putterman (2006, 2007) call *statehist1500*. Details on the construction of the state history index, and the data itself, can be found in Putterman (2004). Note that by beginning with 1 C.E., *statehist* ignores some difference in the onset of state-level society, i.e. those between the most ancient states like Mesopotamia and Egypt (third millennium, B.C.E.), and more recent ones like Rome and pre-Colombian Mesoamerica (first millennium, B.C.E.).

European colonists.[7] It is worth noting that while *statehist* measures a stock of experience with state-level organization that takes into account, for example, set-backs like the disappearance, break-up, or annexation of an existing state by a neighboring empire, *agyears* simply measures the time elapsed since agriculture's founding in the country, with no attempt to gauge temporal changes in the kind, intensity, or prevalence of farming within the country's territory.[8]

For each of these explanatory variables, we conduct a series of tests both with the variable in its original form and with a version adjusted to account for migration. Supposing the "early developmental advantages" proxied by *statehist* and *agyears* to be something that migrants bring with them to their new country, the adjusted variables measure the average level of such advantages in a present-day country as the weighted average of *statehist* or *agyears* in the countries of ancestry, with weights equal to population shares. For instance, ancestry-adjusted *statehist* for Botswana is simply 0.312 times the *statehist* value for Botswana plus 0.673 times *statehist* for South Africa (referring to the people in South Africa in *that* year, not those there presently) plus weights of 0.005 each times the *statehist* values of France, Germany and the Netherlands (the ancestral homes of Botswana's small Afrikaner population). Our dependent variable is the log of year 2000 per capita income.

Table 2 shows our results. Each regression includes the unadjusted form of our early development measure, the adjusted form, or both. Not surprisingly, given previous work, the tests suggest significant predictive power for the unadjusted variables. However, for both measures of early development, adjusting for migration produces a very large increase in explanatory power. In the case of *statehist*, the R^2 goes from .06 to .22, while in the case of *agyears* it goes from .09 to .24. The coefficients on the measures of early development are also much larger using the adjusted than the unadjusted values. In the third and sixth columns of the table we run "horse race" regressions including both the adjusted and unadjusted measures of early development. We find that the coefficients on the adjusted measures retain their significance and become larger while the coefficients on the unadjusted measures become negative and significant.

In the remainder of Table 2 we present tests of the robustness of our findings. We start by constructing measures of *statehist* and *agyears* that are adjusted in the spirit of Hibbs and Olsson (2004, 2005) by simply assigning to four countries (the United States, Canada, New Zealand, and Australia) the *statehist* and *agyears* values of the United Kingdom.[9] As the table shows, these adjusted versions perform better than the unadjusted ones, but not nearly as well as the versions we construct using the migration matrix. When we run "horserace" regressions including *statehist* and *agyears* adjusted using both our matrix and the Hibbs-Olsson method (columns 8 and 10), the coefficients

[7] For further description, see Putterman with Trainor (2006).

[8] The difference is primarily due to data availability. Accounts of the histories of kingdoms, dynasties, and empires are considerably easier to come by than are detailed agricultural histories.

[9] Hibbs and Olsson actually assign these countries the values for the region treated as inheriting the Mesopotamian agrarian tradition, which includes all of North Africa, the Middle East and Europe.

on the matrix-adjusted measures rise in size and significance, while the coefficient on the Hibb-Olsson adjusted measures become negative and significant.

We then construct a series of other measures from out matrix. The first is the fraction of the population made up of "natives" (that is, people whose ancestors lived there in 1500). We include this alongside our measures of adjusted *statehist* and *agyears* in order to check that we are not just picking up the fact that there is a correlation between the share of immigrants in a country and the source of those immigrants. In a similar spirit we construct a measure of the fraction of the descendants of each country's people in 1500 who live in that country today, which we call "retained population." For example, only 40.3% of those descended from the 1500 population of what's now the United Kingdom live there today, whereas 97.4% of Indian descendants still live in India.[10] Neither of these measures eliminates the statistical significance of our adjusted history measures. Retained population enters our regression with a negative sign and is marginally significant, suggesting either that the venting of surplus population may have aided growth or that characteristics that led to countries being able to implant their population abroad also led them to be richer today.

Our second set of robustness checks examines whether our adjusted measures of *statehist* and *agyears* are simply proxying for a large European population or for speaking a European language. In columns 13-16 we include the fraction of the population descended from 1500 inhabitants of European countries, a variable that we create using the matrix. Not surprisingly, given that most of the world's highest income countries are either in Europe or mainly populated by persons of European descent, the European descent variable comes in very significantly. By itself, it explains 41 percent of the variance in the log of GDP per capita. However, even controlling for this variable, our adjusted measures of state history and agriculture are quite significant (t-statistics above 4) and their magnitude falls by only a quarter in comparison with the regressions that don't control for the fraction European. In columns 16-18, we include the fraction of the population speaking one of five European languages (English, French, German, Spanish, and Italian), which is used by Hall and Jones (1999) as an instrument for "social infrastructure." This variable explains only 14 percent of the variation in log of income per capita by itself, and has a negligible effect on the magnitude and significance of our measures of early development.

The finding that adjusting for migration improves the predictive power of measures of early development is consistent with the hypothesis that early technological and social development conferred human capabilities that continued to affect economic performance into the industrial era. These findings suggest that especially Europeans and to some extent East and South Asians carried their historically-bequeathed human capital with them to the Americas, Australia, Malaysia, and elsewhere. They are also consistent

[10] Note that the migration matrix is a rather blunt tool to use for this sort of exercise, because (even with the added population data) it doesn't tell us how many people left the country in question but only how many descendants they have today and where the descendants live. A small number of émigrés may have produced a large number of descendants (for example, the French Canadians) or a large number of émigrés may have produced relatively few (for example, African slaves shipped to the Caribbean).

with the possibility that the historically-bequeathed human capital disadvantage of Africans has played out in new homes such as Jamaica and Haiti, although not ruling out the possibility that their arrival in these places as slaves rather than as migrants may also have played a role. By contrast, the findings of Table 2 cast doubt on a more geographically oriented theory of the importance of early development, which would hold that places that developed early did so because they had favorable climates, and that these favorable climates are responsible for their economic advantage today.

The other finding of Table 2 that is worth pointing out is that, once one adjusts for migration, the explanatory power of measures of early development is relatively high. Even in their unadjusted form, regressions like these suggested that long-term factors play a surprisingly large role in current economic outcomes. The results using adjusted early development suggest that this is all the more true.

2.1 Source Region and Current Region Regressions

Although our interest in most of this paper is in how the migration matrix can be used to map data on place-specific early development into a measure of early development appropriate to a country's current population, the matrix can also be used to infer characteristics of the source countries based only on current data. More specifically, if we assume that emigrants from a particular region share some characteristics that affect the income of countries to which they have migrated, then we can back out these characteristics by looking at data on current outcomes and migration patterns.

To pursue this idea we regress log GDP per capita in 2000 on the fraction of the current population that comes from each of the 11 regions defined previously for the exercises of Table 1. We call the coefficients from this regression, shown in column (1) of Table 3, "source region coefficients." Loosely speaking, they measure how having a country's population composed of people from a particular region can be expected to affect GDP per capita. For example, the source region coefficient for Europe is 2.34, while that for sub Saharan Africa is zero, since this is the omitted category. Thus these coefficients say that moving 10% of a country's population from European to African origin would be expected to lower ln(GDP) by .234 points.[11]

The second column of Table 3 shows a more conventional regression of the log of GDP per capita in the year 2000 on dummies for the region in which the country is located (as in the first column, sub-Saharan Africa is the omitted region). We call these "current region coefficients." The R^2 of the regression with current region dummies is

[11] There are three surprisingly high coefficients in this column: US and Canada, the Caribbean, and Australia and New Zealand. In all three cases the explanation is that the source populations in question contributed a small share of the population to only a few current countries. For example, descendants of people living in the US and Canada as of 1500 contribute only 3.1% and 3.3% of the populations of those two countries, and are found nowhere else in the world. Thus, because the US and Canada are wealthy, this source population gets assigned a high coefficient in the regression. For this reason, we focus our attention on source region coefficients for populations that account for larger population shares in more countries.

about .05 lower than the R^2 of the regression with source region shares. It is also interesting to compare the coefficients on the source and current regions. There is a strong tendency for regions that are rich to also have large values for their source region coefficients. For example, among the six source regions that account for 97% of the world's population (in size order: East Asia, South Asia, Europe, sub-Saharan Africa, Southeast Asia, and North Africa/West and Central Asia) the magnitudes of the coefficients are very similar, with the single exception of South Asia. This similarity of coefficients in the two regressions is not much of a surprise, given the fact, discussed above, that most countries are populated primarily by people whose ancestors lived in that same country 500 years ago. In column (3) of Table 3, we regress log income in 2000 on *both* the source region and current region measures. The R^2 is somewhat higher than in the first two columns, indicating that source regions are not simply proxying for current regions, or vice versa. F-tests easily reject the null hypotheses that either the coefficients on source region or on current region are zero. Interestingly, the source region coefficients on Europe and East Asia remain positive, while the current region coefficients become negative, suggesting that having population from these regions, rather than being located in them, is what tends to make countries rich.

3. Population Heterogeneity and Income Inequality

The finding that current income is influenced by the early development of a country's people, rather than of the place itself, provides evidence against some theories of why early development is important, but leaves many others viable. Early development may matter for income today because of the persistence of institutions (among people, rather than places), because of cultural factors that migrants brought with them, because of long-term persistence in human capital, or because of genetic factors that are related to the amount of time since a population group began its transition to agriculture (Galor and Moav, 2007).

Many of the theories that explain the importance of early development in determining the level of income at the national level would also support the implication that heterogeneity in the early development of a country's population should raise the level of income inequality. For example, if experience living in settled agricultural societies conveys to individuals some cultural characteristics that are economically advantageous in the context of an industrial society, and if these characteristics have a high degree of persistence over generations, then a society in which individuals come from heterogeneous backgrounds in terms their families' economic history should *ceteris paribus* be more unequal. (Following this logic, a country's heterogeneity in early development might also affect the country's average level of income. We examine this question further below).

We pursue three different approaches to examining the determinants of within-country income inequality. We begin by showing that heterogeneity in the historical level of development of country's residents predicts the level of income inequality in a cross-country regression. Second, we construct measures of population heterogeneity

based both on the current ethnic and linguistic groupings and on the ethnic and linguistic differences among the sources of a country's current population. We show that allowing for these other measures of heterogeneity does not reduce the importance of heterogeneity in historical development as a predictor of current inequality.
Finally, we pursue an implication of these findings by asking whether, within a country, people originating from countries that had characteristics predictive of low national income are in fact found to be lower in the income distribution.

3.1 Historical Determinants of Current Inequality

We create two measures of the heterogeneity of the early development of a country's population, using the same state history and history of agriculture variables examined above. The first is the weighted variance of the state history of the countries that contributed to a given country's current population, where the weights are the fractions of that source country's descendants in current population. The second is a similar construction for the years of agricultural history. There is a broad range in the heterogeneity of *agyears* and *statehist*. The mean of the within-country standard deviation of *agyears* is .756, and the standard deviation across countries is .705. The mean within-country standard deviation of *statehist* is .095, and the standard deviation across countries is .088.

In this exercise our dependent variable is the gini coefficient in 1991 or the closest year available (using the high quality sample of Deininger and Squire, 1996). We experiment with including as additional controls the level of the adjusted early development measure as well as the log of current income. The results are show in the first four columns of Table 4.

Our finding is that heterogeneity in the early development experience of the ancestors of a country's population is significantly related to current inequality. To give a feel for the size of the coefficients, we look at the case of *agyears*. The standard deviation of *agyears* in Brazil is 1.894 millennia. By contrast, in countries which have essentially no in-migration, such as Japan, the standard deviation is zero. Applying the regression coefficient of .0656 from the fourth column of Table 3, this would say that variation in early development in Brazil would be expected to raise the gini there by .12, which is certainly an economically significant amount. Since Brazil's gini was .60 and Japan's .35, the exercise suggests that about half of the difference in inequality between the two countries may be attributable to the difference in the heterogeneity of their populations' early development experiences.

We can perform a similar exercise using the source region coefficients estimated in Table 3. Recall that, unlike the exercise just conducted using *statehist* and *agyears*, the estimation of source region coefficients does not require us to know anything about technology or institutions several millennia into the past. The estimates in column (1) of Table 3 simply say that on average, countries with populations originating in certain regions are richer than those with populations originating in other regions. Now we ask whether variation in the source regions of a country's population predicts within-country inequality. Specifically, we create a measure of the weighted standard deviation of the source region coefficients of a country's population, where the weights are the fractions of the population originating in each of the 11 regions. This is a measure of the heterogeneity of a country's population in terms of the source region coefficients of its people's ancestors. For example the Philippines and Mexico have very similar average source region coefficients (.803 and .851, respectively), but differ in the standard deviations of their source region coefficients. In the Philippines, the standard deviation is zero, since the population is entirely composed of people from the Southeast Asia region. In Mexico, the standard deviation of the source region coefficients is 0.958, reflecting a composition of 70% people from the region of Mexico and Central America (source region coefficient 0.234) and 29% from Europe (source region coefficient 2.34). The highest values of the standard deviation of the source region coefficient are found in Canada (5.94), the U.S. (5.90), Belize (2.93), Guatemala (1.82), New Zealand (1.76), and Cape Verde (1.13).

In columns (5) and (6) of Table 4 we present regressions of the gini coefficient on the standard deviation of the source region coefficients, with and without controlling for the average level of the source region coefficient. As expected, the standard deviation of source region significantly positively affects the gini. For example, using the coefficient in column (6) of the table, .0309, the variation in source region coefficients among the population of Brazil would raise the gini coefficient .031 points relative to a country with completely homogenous population in terms of source region coefficient.

3.2 Other Measures of Heterogeneity

Our finding in the last section was that heterogeneity of a country's migrants with respect to measures of early development (*statehist* and *agyears*) contributes to current income inequality. Similarly, heterogeneity of migrants with respect to source region coefficients, which we interpret as an indirect measure of early development, contributes to current inequality. We now pursue the question of whether heterogeneity in the background of migrants more generally may affect the level of income inequality in a country. If this were the case, then in our previous findings early development might simply be proxying for more general heterogeneity. To address this issue, we examine two standard measures of heterogeneity as well as two new measures created using the matrix, and we compare the predictive power of these measures to each other and to the measures that incorporate early development.

Our theory is that a country made up of people who are similar in terms of culture, language, religion, skin color, or similar attributes will *ceteris paribus* have lower inequality. This could take place through a number of different channels. Populations that are similar in the dimensions just listed may be more likely to intermarry and mix socially than populations that are diverse. This mixing could by itself reduce any inequality in the groups' initial endowments, and would also likely to be associated with an absence of institutions that might magnify ethnic, racial, or economic distinctions.

Countries in which people feel a strong sense of kinship with other citizens might also be expected to more actively redistribute income or promote economic mobility.

The first heterogeneity measure we use is ethnic fractionalization from Alesina *et al* (2003). This is the probability that two randomly selected individuals will belong to the same ethnic group. Alesina *et al*. find that higher ethnic fractionalization is robustly correlated with poor government performance on a variety of dimensions.

We create a second measure of fractionalization using the data in the matrix, which we call "historic fractionalization." This is

$$1-\sum_i w_i^2 ,$$

where w_i is the fraction of a country's ancestors coming from country i. Unlike the ethnic fractionalization index, the historic fractionalization index does not take into account ethnic groups composed of people who came from several source countries, such as African Americans, but instead differentiates among, for example, Ghanaian, Senegalese, Angolan, and other ancestors of current residents of the United States (The historical fractionalization index also has the odd property that it includes heterogeneity within individuals. For example a country composed entirely of people who are each half Italian and half Irish will have a fractionalization value of 0.5). As Alesina *et al*. point out, individual self-identification with ethnic groups can change as result of economic, social or political forces. Thus ethnicity has a significant endogenous component, when one looks over spans of centuries, that is absent in the case of historical fractionalization. Factors such as institutions may directly affect the perception of the degree of ethnic heterogeneity within a county.

Ethnic and historical fractionalization are almost uncorrelated (correlation coefficient .16). In particular, a large number of African countries have values of ethnic fractionalization near one but historical fractionalization near zero. The reason is that in these countries there is fractionalization based on tribal affiliation that is unrelated to the movement of people over current international borders over the last 500 years. There are also several countries populated by immigrants (Haiti, Jamaica, Argentina, Israel, the United States) that have a high historic fractionalization because they contain immigrants from many different countries, but a low level of ethnic fractionalization because immigrant groups from similar countries are viewed has having a single ethnicity.

The third measure of heterogeneity we use is "cultural diversity" as constructed by Fearon (2003). Fearon's measure is similar in spirit to the ethnic heterogeneity measure described above, but goes further in making an additional adjustment for different degrees of dissimilarity among the ethnic groups in a country's population. The specific measure of dissimilarity used is based on the language that people speak. Fearon constructs measures of linguistic distance among all currently spoken languages (we describe a very similar methodology below). His measure of cultural diversity is then one minus the average degree of linguistic proximity among two randomly drawn individuals in the population. Desmet, Ortuño-Ortín, and Weber (forthcoming), using a similar measure, find that higher linguistic heterogeneity predicts a lower degree of government income redistribution.

Our final measure of heterogeneity is similar in approach to Fearon's, but instead of using the language that a country's residents speak *today*, we use data on the languages spoken in the countries inhabited by their ancestors in 1500, according to our matrix. Differences in language may directly impede mixing of people from different source countries. In addition, linguistic closeness may well be proxying for other dimensions of culture (such as religion) that could have similar impacts on the degree of mixing among a country's constituent populations and/or the openness of institutions.[12] For these reasons, historical diversity in languages of a country's ancestors may have an impact on inequality that lasts long after the residents of a country have come to speak the same language.

Our starting point in creating a measure of linguistic heterogeneity among a country's ancestor population is to determine the language spoken by people arriving from each potential source country. For each of the 165 countries in our matrix, we do our best to choose the dominant or most prevalent language of the country's population in the year 1500 (see Appendix B at http://www.econ.brown.edu/fac/Louis_Putterman/). Whenever possible, we use historical summaries to determine what was the largest ethnic group in the year 1500, and the language spoken by that group at that time. In some cases where historical information was not available, we use the current day (indigenous) language of the largest current indigenous group. Obviously our method is flawed in ignoring any heterogeneity of languages spoken within a source country. This problem is especially acute because our definition of "country" uses current borders, which are often unrelated to linguistic or cultural fault lines at the time that people emigrated. Thus, for example, immigrants from Sicily and Venice, who would not have been able to understand each other, are treated at having spoken the same language. However, as much of the heterogeneity that we measure relates to gross differences in language among the sources of a country's population (such as Amerindians vs. Europeans), our hope is this mis-measurement will not be too severe.

We then construct a matrix of linguistic distance among each pair of source country languages. The starting point for linguistic distance is a tree showing the relations among all current and known past languages (Gordon, 2005). Every language can be characterized by its family (such as Indo-European or Uralic), and then a series of "nodes," representing the branching points of the language tree, ending in the language itself. For example, the full tree of Spanish is Indo-European, Italic, Romance, Italo-Western, Western, Gallo-Iberian, Ibero-Romance, West Iberian, Castilian, Spanish. Any two languages in the same family can be connected by going up and then down a certain number of nodes. For example, the tree for Italian is common with Spanish through Italo-Western, and is then followed by Italo-Dalmation, Italian. Italian and Spanish thus

[12] Spolaore and Wacziarg (2008) use genetic distance, a measure of the time since two populations shared a common ancestor, as an indicator of cultural similarity between countries. They argue that genetic distance determines the ability of countries to learn from each other, and show that it predicts income gaps among pairs of countries.

have four nodes in common. We measure the distance between any pair of languages as[13]

$$d_{i,j} = 1 - \left(\frac{\text{\# of common nodes between } i \text{ and } j}{\frac{1}{2} \times \left(\text{\# of nodes for langauge } i + \text{\# of nodes for language } j \right)} \right)^{\lambda} .$$

Languages from different families have no nodes in common, and so the distance between them is one. The parameter λ is assumed to be between zero and one, implying that earlier common nodes have a larger weight in the distance function than later ones. In practice, we follow Fearon in assuming $\lambda = 0.5$.[14]

Finally, we combine our linguistic distance measure with the information on source countries in the matrix. Let ***L*** be the matrix of linguistic distances and ***A*** be the matrix with current countries as rows and source countries as columns. Our new measure, which we call "historical linguistic heterogeneity," is the diagonal of ***ALA'***.

Table 5 presents regressions of income inequality, as measured by the gini coefficient, on our various measures of heterogeneity. The first four columns compare the four measures of heterogeneity described above. Controlling for current income, neither of the two measures based on the current population, ethnic fractionalization and cultural (linguistic) diversity, is statistically significant. By contrast, the two variables that use the matrix to measure historical heterogeneity, historical fractionalization and historical linguistic fractionalization, enter very significantly with the expected positive sign. Of the two, the latter, which takes into account degrees of dissimilarly based on linguistic distance, does a better job; along with income, it explains 25% of the variation in the gini coefficient. Much of the superior predictive power of the measures based on historical variation as compared to current variation is driven by Latin America which in terms of language currently spoken does not look very heterogeneous, but does look heterogeneous in terms of historic languages. It is remarkable to see how much better distance among the languages spoken by people's ancestors predicts inequalities today than does distance among the languages spoken by those people themselves. Patterns of social differentiation which arose during the encounters of people from different continents appear to show persistence even after extensive intermixing and linguistic homogenization.

The next four columns of Table 5 repeat these regressions, controlling for the mean and standard deviation of the state history measures, as in columns 1 and 2 of Table 4.[15] The somewhat surprising finding here is that variation in terms of state history

[13] The only difference between our method and Fearon's is that in the denominator he uses 15, which is the maximum number of nodes for any language.

[14] Experimenting with value of lambda in the range 0.25-0.75 had very little effect on the results shown below.

[15] To save space, we don't report parallel exercises using the standard deviation of *agyears*. In Section 3.3, we also focus on *statehist.*. Tables 3 and 4 show that *statehist* and *agyears* have similar explanatory power,

dominates the other forms of heterogeneity that we examine. None of the other four measures of heterogeneity comes close to statistical significance. Variation in early development among a country's people is far more important than more standard forms of heterogeneity (in language or ethnicity) as an explanation for inequality. Similarly, variation in the linguistic background of a country's ancestors, despite its surprising predictive power relative to that of present languages spoken, is not important once one controls for variation in early development.

3.3 Source Country Early Development as a Determinant of Relative Income

The results in Tables 4 and 5 show that heterogeneity in the historical background of a country's residents is correlated with income inequality today. A number of mechanisms could produce such a correlation. One simple theory is that when people with high and low *statehist* are mixed together, the high *statehist* people have some advantage which leads them to percolate up to the top of the income distribution, and then there is enough persistence that their descendants are still there hundreds of years later. A second theory is that situations in which high and low *statehist* people are mixed together tended to occur in cases of colonialization and/or slavery, and that in these circumstances high *statehist* people were able to create institutions that led groups at the top of the income distribution to remain there. We do not propose to test these theories against each other. Instead we test an auxiliary prediction that follows from either of them: specifically, in countries with a high standard deviation of *statehist*, it is the ethnic groups that come from high *statehist* countries that tend to be at the top of the income distribution. Confirming this prediction would give us additional confidence that the link between the standard deviations of *statehist* and the current level of inequality is not spurious.

To test this prediction, we looked for accounts of socio-economic heterogeneity by country or region of ancestral origin in the ten countries in our sample having the highest standard deviation of *statehist*. It is in countries where *statehist* is highly variable where we would be most likely to find differences in outcomes among nationality groups with different values of *statehist*. The countries are listed in Table 6. Not surprisingly, all are former colonies, seven of them in the Americas. Of the latter, three are in Central America, three in South America, and one in the Caribbean. We also list in Table 6 the United States, which has the 16^{th} highest standard deviation of *statehist* in the sample, and is of particular interest due to its size, economic importance, and good data availability.

For each country in the table we first show the breakdown of the population in terms of origin countries or groups of similar countries, according to the matrix. We then show the weighted average value of *statehist* for each origin country or group. The next three columns are based on information about the *current* ethnic breakdown in the country. Ethnic groups as currently identified sometimes correspond to individual origin

---

and we accord slight priority to *statehist* because of its more nuanced tracking of 1500 years of social history (see our discussion comparing the two measures in Section 2).

groups, but are often combinations, frequently labeled mestizo, mulatto or Creole. For each current ethnic group, we then present estimates of average *statehist* and the relative value of current income, listed as high, middle and low or high, upper middle, lower middle, and low. To estimate *statehist* for a mixed ethnic group we use the assumptions underlying the matrix that relate mixed groups to source populations. For example, the group termed "colored" in South Africa is assumed to have half of its ancestors coming in equal proportions from five European countries (England, Portugal, and Afrikaner source countries Netherlands, France and Germany) and the other half in unequal proportions from South Africa itself (35%), India (10%) and Indonesia (5%). These assumptions are reported in the region appendices describing the construction of the matrix.

Leaving details to Appendix C[16], we note immediately that the ordering of *statehist* values and the ordering of socio-economic status in Table 6 has at least some correspondence in every country. For nine of the eleven countries listed—Fiji, Cape Verde, Guyana, Paraguay, Panama, South Africa, El Salvador, Nicaragua and Venezuela—the socio-economic ordering perfectly dovetails with that of *statehist* values. In two countries—Trinidad and Tobago and the United States—there are discrepancies in the orderings of Asians and "Whites," with Chinese and (S. Asian) Indians having lower incomes than Whites in the first country despite having higher *statehist*, while Asians in general have higher incomes than Whites in the U.S. despite lower average *statehist*. For the U.S., there is a further discrepancy in that "Black" Americans have lower average incomes than American Indians and Alaska Natives, despite having somewhat higher average *statehist* values. While no statistical significance should be attached to the counts just mentioned, since the categorizations are quite broad and require some judgments to be made, the general pattern clearly supports the expectation.

A few patterns are noteworthy. Paraguay and El Salvador are representative of the many Latin American countries in which the main identifiable groupings, listed in order of both socio-economic status and of average *statehist*, are European, mestizo, and Amerindian. Like other countries in or bordering the Caribbean, three of the represented countries—Panama, Nicaragua and Venezuela—add a group of largely African descent to this tri-partite pattern. In each of the latter countries, the White group remains on top and the Amerindian group on the bottom. The Black group, with higher *statehist* than the Amerindians,[17] is variously found on approximate par with the mestizos (Nicaragua), between the mestizo and Amerindian groups (Panama), or sharing the bottom rung with the Amerindians (Venezuela).

In two of the other represented countries of the Americas—Guyana and Trinidad and Tobago—there are substantial populations of South Asian origin. The socio-economic positioning of this group is lower than predicted by their average *statehist*. This result, contradicting our general hypothesis, seems related to the economic hard

[16] http://www.econ.brown.edu/fac/Louis_Putterman/ .

[17] This is due to the existence of some states in Africa before 1500 but their absence in the Americas outside of Mexico, Guatemala and the Andes. Note that the situation is reversed in some cases, for instance the indigenous people of South Africa have a lower *statehist* value than do those of Mexico and Peru.

times on which South Asia itself had fallen by the 19th Century and the manner in which millions were brought from that region to the Caribbean to work in indentured servitude after Britain outlawed slavery.

Of the two African countries represented in Table 6, Cape Verde began as a Portuguese plantation economy employing slaves brought from the African mainland. At the time of the country's independence from Portugal, in 1975, the society was described as being stratified along color lines, with people of darker complexion usually found in the lower class and people of lighter complexion constituting the "bourgeoisie" (Meintel, 1984; Lobban, 1995). The correlation between complexion and socioeconomic class is consistent with our proposed explanation of the correlation between standard deviation of *statehist* and the gini coefficient seen in Table 4. In South Africa, the major population categories are Black African, White, "colored" (with both European and either African, Indian, or Malay ancestors), and Indian or Asian. The socio-economic standings of these groups today remains heavily influenced by the history of European settlement and subordination of the local population, and partly as a result, the average incomes for those in the four groupings are ordered exactly in accord with the ordering of average *statehist*.

The only case in Table 6 not located in the Americas or Africa is Fiji, whose population is classified by government statisticians as indigenous (55.0%), Indian (41.0%) and Other (mainly European and Chinese, 4.0%). Average household incomes per adult in the three groups are ordered identically to average *statehist* values. Although the reported income gap between the Indian and native Fijian populations is far smaller than the difference in *statehist*, the government statisticians comment that the incomes of Indo-Fijians are probably undercounted, since much of it comes from private business activities likely to be underreported.

Turning finally to the U.S., the Census Bureau reports a breakdown of the population into White non-Hispanic, Hispanic any race, Black, Asian, American Indian and Alaska Native, and other small categories. These groups' reported median incomes have the same ordering as their average *statehist* values, with the exception of the higher Asian than White income and the higher American Indian than Black income. The simple correlation between the five *statehist* and the five income values, with equal weighting on all observations, is 0.747.

On balance, the evidence from the ten countries with the highest internal variation of *statehist* and from the sixteenth-ranking United States appears to support the idea that correlation between within-country differences in income and corresponding differences in the early development indicator *statehist* at least partially account for the predictive power of the standard deviation of *statehist* in the Table 4 regressions. Indeed, in this section we have found within countries (as the previous section found between countries) that there is considerable persistence and reproduction of income differences which appears to reflect social differences dating back up to half a millennium. To be sure, in the majority of cases just discussed differences in societal capabilities during the era of European expansion played themselves out to a considerable degree in the form of outright dominance of some over others, including appropriation of land, control of government and monopoly of armed force, and involuntary movement of millions of people between macro-regions to meet the conquering population's labor demands. How persistent early differences would have proven to be in the absence of the exercise of raw power is a question that goes beyond the scope of our paper. The point for present purposes is that as history has in fact unfolded, such differences have been remarkably persistent.

4. Population Heterogeneity and Income Levels

As a final exercise, we discuss the relationship between heterogeneity in the early development history of a country's population and the average level of GDP per capita. We showed in section 2 that a higher average level of early development in a country was robustly correlated with higher current income. The explanation for this finding is that people whose ancestors were living in countries that developed earlier brought with them some advantage—such as human capital, knowledge, culture, or institutions—which raises the level of income in their country up until today.

Depending on what exact advantage is conferred by earlier development, there might also be implications for how the variance of early development among a country's contributing populations would affect output. For example, if early development conferred some cultural attribute that was good for growth, then in a population containing some people with a long history of development and some with a short history, this growth-promoting cultural trait might simply be transferred from the long history group to the short history group. Similarly, growth-promoting institutions brought along by people with a long history of development could be extended to benefit people with short histories of development. An obvious model for such transfer is language: in many parts of the world, descendents of people with short histories of development speak languages that come from Europe, which has a long history of development. If growth-promoting characteristics also transfer in this fashion, then a country with half its population coming from areas with high *statehist* and half from areas with low *statehist* might be richer than a country with the same average *statehist* but no heterogeneity.

The above logic would tend to predict that, holding average history of early development constant, a higher variance of early development would raise a country's level of income. However, there are channels that work in the opposite direction. As shown in the previous section, higher variance of early development predicts higher inequality. Inequality is often found to negatively impact growth (see, for example, Easterly 2007), and one could easily imagine that the inequality generated by heterogeneity in early development history would lead to the inefficient struggles over income redistribution or the creation of growth-impeding institutions. This is certainly the flavor of the story told by Sokoloff and Engermann (2000). Similarly, the ethnic diversity that comes along with a population that is heterogeneous in its early development history could hinder the creation of growth-promoting institutions. In 1986, Japan's Prime Minister Yasuhiro Nakasone, commenting on the skills of the

American labor force, said that "there are things the Americans have not been able to do because of multiple nationalities there … things are easier in Japan because we are a monoracial society."[18]

In Table 7 we present regressions of the log of current income per capita on the standard deviation of each of our three measures of early development (*statehist*, *agyears*, and source region coefficients), with and without controls for the mean of each of the variables as well as the current gini coefficient. Once the mean level of *statehist* is controlled for, the standard deviation of *statehist* has a positive and significant effect on current income. In the case of the regression using *agyears* the coefficient is similarly positive, but is only significant at the 7% level. Interestingly, the coefficient on the standard deviation of the source region coefficient is not significant at all once the mean of the source region coefficients is included.

Columns 3, 6, and 9 include the value of the gini coefficient on the right hand side of the regression. Obviously, since we just presented a series of regressions in which it was the dependent variable, we consider the gini coefficient to be endogenous. Nevertheless, controlling for the gini is a way to test the theory that one effect of heterogeneity in early development is to reduce current income by raising inequality. The gini coefficient enters the regressions that include either *statehist* or *agyears* negatively. More important, including the gini raises the coefficient on the standard deviation of either *statehist* or *agyears* (and makes the latter statistically significant). This is what we would expect in the case where one of the channels by which heterogeneity in development affected income was inequality. In the case of the regressions including the source region coefficients, the coefficient on gini is positive and significant, and adding the gini to the right hand side lowers the coefficient on the standard deviation of the source region coefficients.

The positive coefficients on the standard deviations of *statehist* and *agyears* imply, as discussed above, that a heterogeneous population will be better off than a homogeneous population with the same average level of early development. For example, using the coefficients in Column 2 of Table 7, a country with a population composed of 50% people with a *statehist* of 0.4 and 50% with a *statehist* of 0.6 will be 25 percent richer than a homogenous country with *statehist* of 0.5. A country with 50% of the population having *statehist* of 1.0 and 50% with *statehist* of zero would be 3.1 times as rich as a homogenous country with the same average *statehist*. (This latter example is quite outside the range of the data, however. The highest values of the standard deviation of *statehist* in our data set are Fiji (0.346), Cape Verde (0.294) and Guyana (0.293). In the example, the standard deviation is 0.5).

The coefficients also have the unpalatable property that a country's predicted income can sometimes be raised by replacing high *statehist* people with low *statehist* people, since the decline in the average level of *statehist* will be more than balanced by the increase in the standard deviation. For example, the coefficients just discussed imply that combining populations with *statehist* of 1 and 0, the optimal mix is 83% *statehist*=1

[18] *Time Magazine*, "Nakasone's World-Class Blunder," October 6, 1986.

and 17% *statehist*=0. A country with such a mix would be 66% richer than a country with 100% of the population having a *statehist* of 1.[19]

One explanation for this somewhat counterintuitive finding is that during the long era of European expansion spanning the 15th to early 20th centuries, European-settled countries like the United States, Chile, Mexico and Brazil having substantial African and/or Amerindian minorities attained considerably higher incomes than many homogenously populated Asian countries with relatively long state histories, including Bangladesh, Pakistan, India, Sri Lanka, Indonesia and China. Chanda and Putterman (2007) argue that the underperformance of the latter group of countries during the 1500 – 1960 period is an exception to the rule (which they find to have held up to 1500 and again since 1960) that earlier development of agriculture and states has been associated with greater economic development during most of world history. While our regression result reflects the fact that population heterogeneity has not detracted from economic development in the first group of countries, it may be well to treat it as a by-product of specific historical contingencies, and not to infer from it that "catch up" by the latter countries would be speeded up by infusions of low *statehist* populations into existing high *statehist* countries.

5. Conclusion

Conquest, colonialism, migration, slavery, and epidemic disease reshaped the world that existed before the era of European expansion. Over the last 500 years, there have been dramatic movements of people, institutions, cultures, and languages among the world's major regions. These movements clearly have implications for the course of economic development. Existing literature has already made a good start at examining how institutions were transferred between regions and the long lasting economic effects of these transfers. However the human side of the story – the relationship between where the ancestors of a country's current population lived and current outcomes – has received relatively little attention, in part due to the absence of suitable data. In this paper, we introduce a "world migration matrix" to account for international movements of people since the year 1500. We use the matrix to document some major features of world migration history such as the bi-modality of the distribution of indigenous and non-indigenous people by country and the variations in the primary source regions for immigrant-populated countries.

In the second part of the paper, we demonstrate the utility of the migration data by using it to re-visit the hypothesis that early development of agrarian societies and their sociopolitical correlates—states—conferred developmental advantages that remain relevant today. We confirm that in a global sample, countries on whose territories agriculture and states developed earlier have higher incomes. But we conjecture that

[19] The specification that we use implies that this property must hold as long as the coefficients on both the mean and standard deviation are positive. However, when we use variance on the right had side, in which case the property does not automatically hold, it is nonetheless implied by the estimates.

people who moved from one region to another carried the human capabilities built up in that area with them. We find that re-calculating state history and agriculture measures for each country as weighted averages by place of origin of their people's ancestors considerably improved the fit of these regressions.

In Part 3, we show that the heterogeneity of a country's population in terms of the early development of its ancestors as of 1500 was strongly correlated with income inequality. We also show that heterogeneity with respect to country of ancestry or with respect to the ancestral language does a better job than does current linguistic or ethnic heterogeneity in predicting income inequalities today. As an additional test of the theory that early development conferred lasting advantage, we show that the rankings of ethnic or racial groups within a country's income distribution are strongly correlated with the average levels of groups' early development indicators. Finally, in Part 4, we find that heterogeneity of early development, holding the mean level constant, is associated with higher, per capita income. We interpret this finding as indicating that the effect of spillovers of growth-promoting characteristics between groups having different early development histories more than compensated for any negative effect on growth of higher inequality due to heterogeneity.

The overall finding of our paper is that the origins of a country's population – more specifically, where the ancestors of the current population lived some 500 years ago – matters for economic outcomes today. Having ancestors who lived in places with early agricultural and political development is good for income today, both at the level of country averages and in terms of an individual's position within a country's income distribution. Exactly *why* the origins of the current population matter is a question on which we can only speculate at this point. People who moved across borders brought with them human capital, cultures, genes, institutions, and languages. People who came from areas which developed early evidently brought with them versions of one or more of these things that were conducive to higher income. Future research will have to sort out which ones were the most significant. The fact that early development explains an ethnic group's position within a country's income distribution suggests that "good institutions" coming from regions of early development cannot be the whole story, although it does not prove that institutions are not of enormous importance. More research is also needed to understand how early development led to the creation of growth promoting characteristics (whatever these turn out to be) as well as the process by which these characteristics are transferred between populations of high and low early development. Our hope is that the availability of a compilation of data on the reconfiguration of country populations since 1500 will make it easier to address such issues in future research.

References

Acemoglu, D., Johnson, S., Robinson, J. (2001). "The colonial origins of comparative development: An empirical investigation," *American Economic Review* 91 (5), 1369–1401.

Acemoglu, D., Johnson, S., Robinson, J. (2002). "Reversal of fortunes: Geography and institutions in the makingof the modern world income distribution," *Quarterly Journal of Economics* 117 (4), 1231–1294.

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

Bockstette, Valerie, Areendam Chanda and Louis Putterman (2002). "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth* 7: 347-69.

Burkett, John P., Catherine Humblet and Louis Putterman (1999). "Pre-Industrial and Post-War Economic Development: Is There a Link?" *Economic Development and Cultural Change,* 47 (3): 471-95.

Chanda, Areendam and Louis Putterman, 2006, "State Effectiveness, Economic Growth, and the Age of States," pp. 69-91 in Matthew Lange and Dietrich Rueschemeyer, eds., *States and Development: Historical Antecedents of Stagnation and Advance* Basingstoke, England, Palgrave MacMillan, 2005.

Chanda, Areendam and Louis Putterman (2007). "Early Starts, Reversals and Catch-up in the Process of Economic Development," *Scandinavian Journal of Economics* 109, 387-413.

Comin, Diego, William Easterly, and Erick Gong, "Was the Wealth of Nations determined in 1000 BC?" mimeo, 2006.

Deininger, Klaus, and Lyn Squire, "A New Data Set and Measure of Income Inequality," *World Bank Economic Review,* 10 (September) 1996, 565-591.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber, "Linguistic Diversity and Redistribution," *Journal of the European Economic Association*, forthcoming.

Easterly, William, "Inequality Does Cause Underdevelopment: Insights from a New Instrument," *Journal of Development Economics*, Volume 84, Issue 2, November 2007, Pages 755-776.

Fearon, James, D., "Ethnic Structure and Cultural Diversity by Country," *Journal of Economic Growth*, 8:2, June 2003, 195-222.

Galor, Oded, and Omer Moav, “The Neolithic Origins of Contemporary Variation in Life Expectancy,” Mimeo, November 2007

Gordon, Raymond G., Jr. (ed.), 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/

Hall, R., and C. Jones, (1999). “Why do some countries produce so much more output than others?” *Quarterly Journal of Economics* 114 (1), 83–116.

Hibbs, Douglas A., and Ola Olsson, 2004, “Geography, biogeography, and why some countries are rich and others are poor,” *Proceedings of the National Academy of Sciences* 101: 3715-3720.

Hibbs, Douglas A., and Ola Olsson (2005). “Biogeography and Long-Run Economic Development,” *European Economic Review,* 49: 909-938.

Johnson, Allen and Timothy Earl, 1987*, The Evolution of Human Societies: From Foraging Groups to Agrarian State*. Stanford, CA: Stanford University Press.

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

MacNeish, Richard (1991). *The Origins of Agriculture and Settled Life*. Norman, OK: University of Oklahoma Press.

McEvedy, C. and R. Jones (1978). Atlas of World Population History, Viking Press.

Putterman, Louis (2004). “State Antiquity Index Version 3”, available at http://www.econ.brown.edu/fac/Louis_Putterman/.

Putterman, Louis (2006) “World Migration Matrix, 1500 – 2000”, available at http://www.econ.brown.edu/fac/Louis_Putterman/.

Putterman, Louis, with Cary Anne Trainor (2006). “Agricultural Transition Year Country Data Set”, available at http://www.econ.brown.edu/fac/Louis_Putterman.

Smith, Bruce (1995). *The Emergence of Agriculture*. New York: Scientific American Library.

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

Spolaore, Enrico, and Romain Wacziarg, 2008, “The Diffusion of Development.” Mimeo.

Table 1. Current population and descendants, by region.

Region	Population in 1500 (millions)	Population in 2000 (millions)	Population Growth Factor	Descendants per person of 1500	Fraction of current population descended from region's 1500 ancestors	Fraction of descendants of 1500 population who live in same region
U.S. and Canada	1.12	315	281.1	8.8	0.031	0.997
Mexico and Central America	5.8	137	23.6	18.2	0.646	0.850
The Caribbean	186	34.4	184.8	0.7	0.000	0.000
South America	7.65	349	45.6	9.2	0.200	0.979
Europe	77.7	680	8.8	16.0	0.942	0.657
North Africa/West and Central Asia	35.5	530	14.9	14.6	0.887	0.911
South Asia	103	1,320	12.8	12.9	0.971	0.963
East Asia	132	1,490	11.3	11.6	0.980	0.939
Southeast Asia	18.7	555	29.7	28.5	0.940	0.972
Australia and New Zealand	0.2	22.9	114.4	3.7	0.032	1.000
Sub Saharan Africa	38.3	656	17.1	19.5	0.881	0.843

Table 2: Historical Determinants of Current Income

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Indep. Var.	Dependent Variable: ln(GDP per capita 2000)									
Statehist	.892 (.330)		-1.38 (.32)							
Ancestry Adjusted *Statehist*		2.02 (.38)	3.33 (.41)					2.72 (.46)		
Agyears				.134 (.035)		-.188 (.044)				
Ancestry Adjusted *Agyears*					.269 (.040)	.450 (.053)				.388 (.050)
HO adjusted *statehist*							1.27 (.32)	-.698 (.349)		
HO adjusted *agyears*									.173 (.034)	-.122 (.039)
Constant	8.17 (.14)	7.60 (.17)	7.50 (.17)	7.87 (.21)	7.05 (.23)	6.96 (.22)	8.02 (.14)	7.55 (.17)	7.66 (.19)	7.00 (.22)
No. obs.	136	136	136	147	147	147	136	136	147	147
R-squared	0.060	0.221	0.272	0.080	0.241	0.290	0.122	0.231	0.127	0.257

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Indep. Var.	Dependent Variable: ln(GDP per capita 2000)							
Ancestry Adjusted *Statehist*	2.11 (.37)			1.48 (.32)			2.10 (.39)	
Ancestry Adjusted *Agyears*		.270 (.041)			.182 (.034)			.283 (.042)
Native	-.973 (.290)	-.816 (.292)						
Retained	-1.05 (.39)	-.781 (.382)						
Fraction European Descent			1.77 (.16)	1.60 (.16)	1.49 (.16)			
Fraction European Languages						1.19 (.21)	1.01 (.19)	.945 (.188)
Constant	9.15 (.51)	8.29 (.49)	7.89 (.11)	7.27 (.13)	7.01 (.19)	8.20 (.13)	7.28 (.17)	6.78 (.23)
No. obs.	126	136	152	139	149	126	114	123
R-squared	0.295	0.284	0.413	0.566	0.513	0.144	0.413	0.390

Table 3: Source Regions and Current Regions as Determinants of Current Income

Regression number	(1)	(2)	(3)	
Independent Variables	Source Regions	Current Regions	Source Regions	Current Regions
U.S. and Canada	35.5 (6.9)	3.03 (0.16)	-313 (45)	10.7 (1.4)
Mexico and Central America	0.234 (0.600)	1.10 (0.24)	2.31 (0.92)	-1.27 (0.53)
The Caribbean	16.5 (9.5)	1.33 (0.30)	24.1 (9.2)	0.198 (0.236)
South America	0.327 (0.213)	1.35 (0.20)	0.952 (0.426)	-0.455 (0.363)
Europe	2.34 (0.16)	2.23 (0.18)	2.72 (0.39)	-0.323 (0.409)
North Africa/West and Central Asia	1.29 (0.21)	1.28 (0.21)	0.709 (1.330)	0.558 (1.229)
South Asia	0.869 (0.265)	0.388 (0.175)	3.09 (0.42)	-2.57 (0.42)
East Asia	2.14 (0.54)	1.81 (0.56)	4.82 (0.60)	-2.86 (0.88)
Southeast Asia	0.803 (0.242)	1.07 (0.32)	1.64 (0.65)	-0.962 (0.526)
Australia and New Zealand	8.13 (2.09)	2.72 (0.17)	-1.53 (0.77)	0.378 (0.380)
Constant	7.27 (0.11)	7.34 (0.13)	7.22 (0.12)	
No. obs.	152	152	152	
R-squared	0.630	0.584	0.685	

Note: Sub Saharan Africa is the omitted region in all regressions. In regression 1, the independent variables are the shares of the population in each country originating in each region. In regression 2, the independent variables are dummies for a country being located in a particular region. In regression 3 the independent variables are both of the above.

Table 4: Historical Determinants of Current Inequality

	(1)	(2)	(3)	(4)	(5)	(6)
Indep. Var.	Dependent variable: Gini Coefficient					
Standard Deviation of *Statehist*	.517 (.088)	.498 (.092)				
Ancestry Adjusted *Statehist*		-.0340 (.0367)				
Standard Deviation of *Agyears*			.0654 (.0114)	.0656 (.0107)		
Ancestry Adjusted *Agyears*				-.0134 (.0049)		
Standard Deviation of Source Region Coefficients					.0254 (.0171)	.0309 (.0145)
Mean Source Region Coefficient						-.0830 (.0147)
Ln(y2000)	-.0314 (.0069)	-.0276 (.0074)	-.0357 (.0065)	-.0250 (.0065)	-.0328 (.0070)	.0227 (.0120)
Constant	.619 (.065)	.604 (.063)	.653 (.059)	.632 (.055)	.667 (.062)	.899 (.074)
No. obs.	97	97	100	100	100	100
R-squared	0.312	0.319	0.350	0.407	0.157	0.341

Table 5: Ethnic, Linguistic, and Historical Determinants of Current Inequality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Indep. Var.	Dependent Variable: Gini Coefficient							
Ethnic Fractionalization	.0802 (.0484)				-.00563 (.04652)			
Historical Fractionalization		.137 (.033)				-.0373 (.0620)		
Cultural Diversity			-.0179 (.0516)				-.0327 (.0449)	
Historical Linguistic Fractionalization				.178 (.038)				.00246 (.08834)
Standard Deviation of *Statehist*					.502 (.093)	.585 (.177)	.498 (.090)	.493 (.217)
Ancestry Adjusted *Statehist*					-.0346 (.0381)	-.0380 (.0373)	-.0378 (.0362)	-.0341 (.0369)
Ln(y2000)	-.0153 (.0116)	-.0291 (.0073)	-.0289 (.0089)	-.0307 (.0075)	-.0284 (.0104)	-.0273 (.0075)	-.0306 (.0085)	-.0276 (.0075)
Constant	.492 (.118)	.610 (.067)	.649 (.088)	.625 (.068)	.613 (.107)	.605 (.063)	.640 (.081)	.604 (.064)
No. obs.	97	97	96	97	97	97	96	97
R-squared	0.118	0.222	0.099	0.252	0.319	0.322	0.331	0.319

Table 6: ***Statehist*** **and relative income for ancestry groups and current ethnic groups**

#	Country	Standard Dev. of *Statehist*	Gini[20]	Component groups (region)	Percent population [21]	*Statehist* (average)	Component groups (ethnic)	Percent population [22]	*Statehist* (average)	Relative Income
1	Fiji	.346	.425	European	2.2	0.693	Other[23]	4	0.745	High
				Indian	45.0	0.688	Indo-Fijian	41	0.688	Middle
				Fijian	52.1	0.000	Fijian	55	0.000	Low
2	Cape Verde	.295	.51	Portuguese	36.5	0.723	White	1.0	0.723	High
				African	63.5	0.178	Creole	71.0	0.450	Middle
							Black	28.0	0.178	Low
3	Guyana	.293	.402	Chinese	0.7	0.906	Chinese	0.3	0.906	High
				Portuguese	1.3	0.723	Portuguese	0.4	0.723	Middle
							Mixed[24]	11.2	0.410	Middle
				S. Asian	54.0	0.677	East-Indian	51.9	0.677	Middle
				African	39.0	0.142	Black	30.8	0.142	Middle
				Guyanese	5.0	0.000	Amerindian	5.3	0.000	Low

[20] Source: Deininger and Squire (1996), except: Fiji – 1977, Deininger, Squire, "Measuring Income Inequality Database", World Bank; Cape Verde – World Development Indicators, 2001; Paraguay – World Development Indicators, 2003; El Salvador - World Development Indicators, 2002

[21] Computed from Matrix.

[22] Based on: Fiji – Household Survey 2002-03; Cape Verde – Census 1950 (quoted in Lobban, R., "Cape Verde: Crioulo Colony to Independent Nation", 199); Guyana – Census 1980; Paraguay – Census 2002; Panama – Fearon, J. D. Data set described in "Ethnic and Cultural Diversity by Country". Journal of Economic Growth 8, 2 (June 2003): 195-222.; South Africa – Household Survey 2005; Trinidad and Tobago - Continuous Sample Survey of Population; El Salvador – CIA Factbook; Nicaragua – CIA Factbook; Venezuela – CIA Factbook,; United States - U.S. Census, Vintage 2004

[23] Europeans, Chinese

[24] ½ East Indian, ½ African

4	Paraguay	.291	.58	European, non-Spanish	6.0	0.705	European (incl. Spanish)	3.8	0.578	High
				Spanish	47.0	0.562	Mestizo	94.7	0.281	Middle
				Paraguayan / Brazilian	46.0	0.000	Amerindian	1.1	0.000	Low
5	Panama	.291	.565	Chinese	1.5	0.906	Chinese	2.0	0.906	High
				S. Asian	4.0	0.677	White	10.0	0.578	High
				European	45.2	0.578	Mestizo	68.0	0.281	Upper Middle
				African	13.0	0.142	mixed West-Indian	13.0	0.142	Lower Middle
				Panamanian	35.7	0.000	Amerindian	6.0	0.000	Low
6	South Africa	.289	.623	European	18.0	0.710	White	9.2	0.710	High
				Indian / S. Asian	3.4	0.670	Indian / Asian	2.5	0.670	Upper Middle
				South-African	78.7	0.000	Colored (mixed)[25]	8.9	0.452	Lower Middle
							Black African	79.4	0.000	Low
7	Trinidad and Tobago	.284	.417	Chinese	1.5	0.906	Chinese	0.2	0.906	Upper Middle
				European	7.1	0.671	White / Caucasian	0.7	0.671	High
				S. Asian	45.4	0.677	Indian	40.5	0.677	Low
				African	46.0	0.166	Mixed[26]	14.9	0.504	Lower Middle

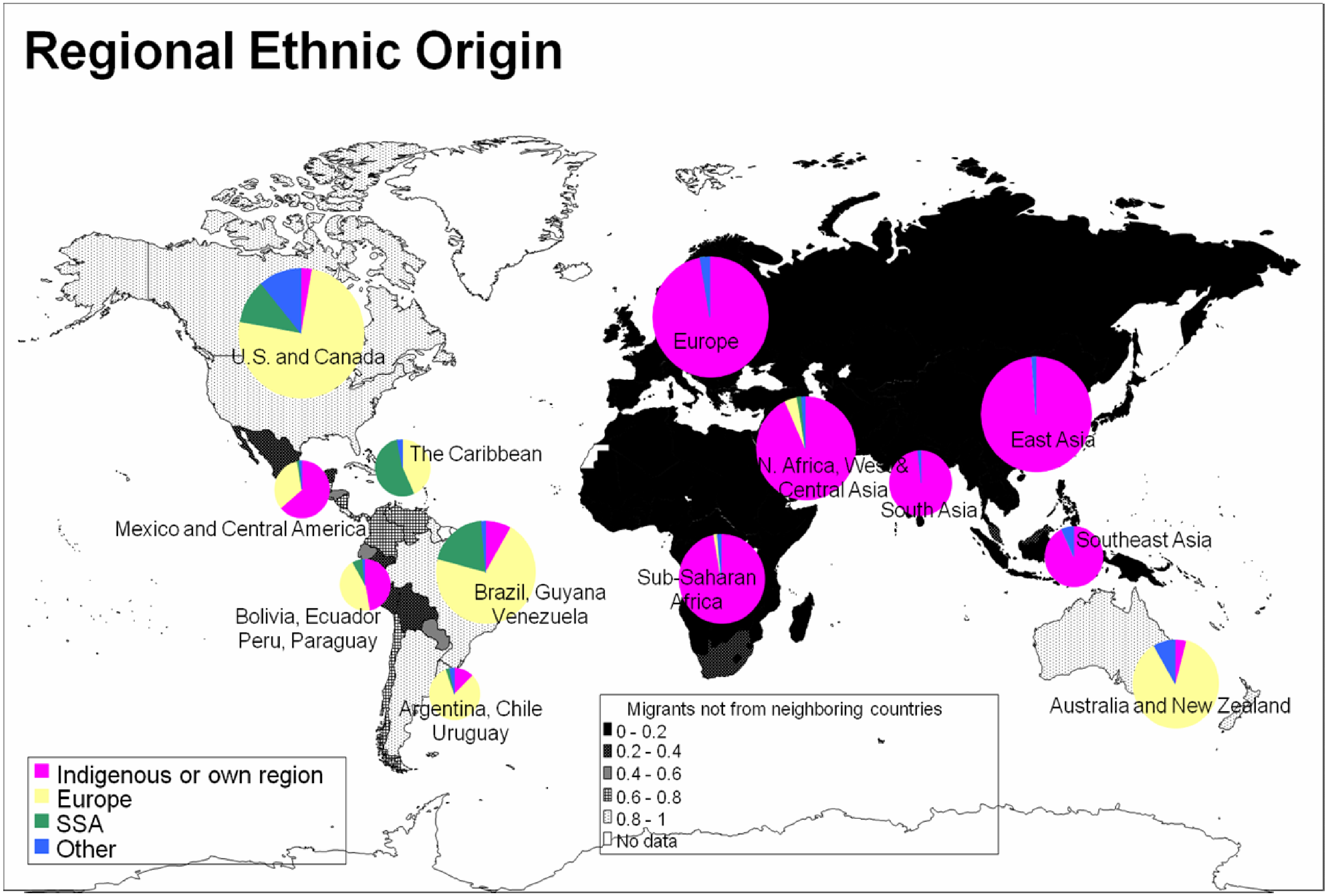
[25] .35 African, .1 S. Asian, .05 Indonesian, .1 UK, .1 Netherlands, .1 France, .1 Germany and .1 Portugal
[26] 1/3 African, 1/3 S. Asian, 1/3 European.

							African	43.5	0.166	Low
8	El Salvador	.281	.52	Spanish	50.0	0.562	White	9.0	0.562	High
				Salvadoran	50.0	0.000	Mestizo	90.0	0.281	Middle
							Amerindian	1.0	0.000	Low
9	Nicaragua	.277	.503	European	51.0	0.568	White	17.0	0.568	High
				African	9.0	0.142	African (Creole)	9.0	0.142	Middle
				Nicaraguan	40.0	0.000	Mestizo	69.0	0.281	Middle
							Amerindian	5.0	0.000	Low
10	Venezuela	.273	.538	European	53.5	0.565	White	21.0	0.565	High
				African	9.9	0.142	Mestizo	68.0	0.281	Middle
				Venezuelan	36.0	0.000	Black	10.0	0.142	Low
							Amerindian	1.0	0.000	Low
16	United States	.234	79	European	74.3	0.648	White not Hispanic	67.4	0.654	Upper Middle
				Asian	4.2	0.597	Asian	4.2	0.597	High
				Central and South American	6.3	0.490	Hispanic of any race	14.1	0.528	Lower Middle
				Sub-Saharan African	11.4	0.157	Black	12.8	0.157	Low
				North-American[27]	3.1	0.000	American Indian and Alaska Native	1.0	0.000	Lower Middle

[27] Includes Hawaii and Alaska

Table 7: The Effect of Heterogeneity in Early Development on Current Income

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Indep. Var.	Dependent variable: Ln(GDP per capita 2000)								
Standard Deviation of *Statehist*	1.47 (1.14)	2.24 (.98)	3.83 (1.05)						
Ancestry Adjusted *Statehist*		1.97 (.43)	1.65 (.44)						
Standard Deviation of *Agyears*				.283 (.133)	.217 (.120)	.457 (.136)			
Ancestry Adjusted *Agyears*					.265 (.052)	.186 (.056)			
Standard Deviation of Source Region Coefficients							.327 (.056)	.0348 (.0546)	-.00844 (.04701)
Mean Source Region Coefficient								1.02 (.07)	1.11 (.08)
Gini			-3.64 (1.01)			-3.99 (1.08)			1.36 (.68)
Constant	8.49 (.17)	7.52 (.23)	8.96 (.50)	8.39 (.18)	7.03 (.30)	8.85 (.62)	8.46 (.13)	-.198 (.620)	-1.42 (.81)
No. obs.	97	97	97	100	100	100	100	100	100
R-squared	0.014	0.229	0.306	0.034	0.240	0.316	0.063	0.705	0.714


Regional Ethnic Origin
U.S. and Canada
The Caribbean
Mexico and Central America
Bolivia, Ecuador
Peru, Paraguay
Brazil, Guyana
Venezuela
Argentina, Chile
Uruguay
Europe
N. Africa, West &
Central Asia
South Asia
East Asia
Southeast Asia
Sub-Saharan
Africa
Australia and New Zealand
Indigenous or own region
Europe
SSA
Other
Migrants not from neighboring countries
0 - 0.2
0.2 - 0.4
0.4 - 0.6
0.6 - 0.8
0.8 - 1
No data

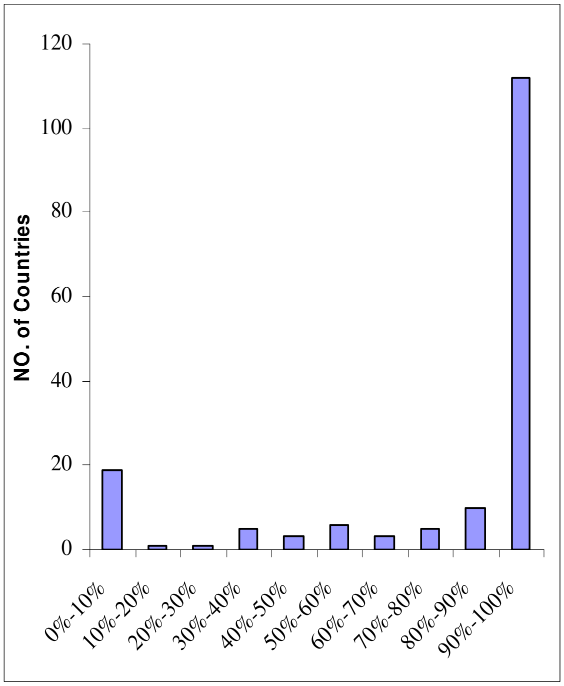


Figure 1a Distribution of countries by proportion of ancestors from own or immediate neighboring country.

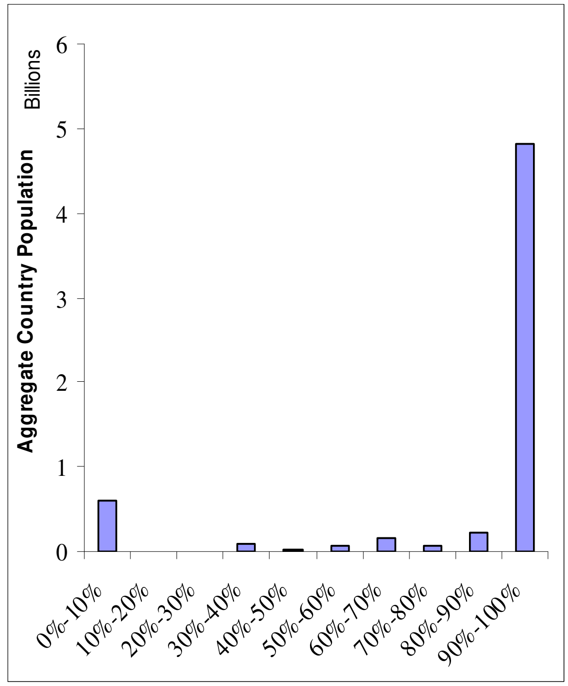


Figure 1b Distribution of world population by proportion of ancestors from own or immediate neighboring country.

Appendix A: World Migration Matrix, 1500 – 2000*

The goal of the matrix is to identify where the ancestors of the permanent residents of today's countries were living in 1500 C.E. In this abbreviated description, we address some major conceptual issues relevant to the construction of the matrix and identify some of the main sources of information consulted.

The migration matrix is a table in which both row and column headings are the names of presently existing countries, and cell entries are estimates of the proportion of the ancestors of those now permanently residing in the country identified in the row heading who lived in the country identified by the column heading in 1500. The country designations of the column headings are identical to those of the row headings even though borders and country names tended to be quite different in 1500 than they are today. An ancestor is treated as having lived in what is now, say, Indonesia, if the place they resided in that year is within the borders of Indonesia today.

When ancestors could be identified only as part of an ethnic group that lived in a region now straddling the borders of two or more present-day countries, we try to estimate the proportion of that group living in each country and then allocate ancestry accordingly. For example, if a given ancestor is known to have been a "Gypsy" but if we have no information on which country he or she lived in in the year 1500, we apply an assumption (see the main appendix in Putterman, 2006) regarding the proportion of Gypsies who lived in Greece, Romania, Turkey, etc., as of 1500. The Gypsy example is one of many illustrating the fact that most of our data sources organize their information around ethnic groups rather than territory of origin. While the use of information on ethnicity was unavoidable in the process of constructing the matrix, it was not a focus of attention in its own right.

In cases in which ancestors are known to have migrated more than once between 1500 and 2000, countries of intervening residence are not indicated in the matrix. For example, an Israeli whose parents lived in Argentina but whose grandparents arrived in Argentina from Ukraine, is listed as having had ancestors in Ukraine.

People of mixed ancestry are common in many countries, for example people of mixed Amerindian and Spanish ancestry in Mexico. Such an individual is treated as having a certain proportion of ancestors living in Mexico and a certain proportion living in Spain in 1500, the proportion being determined by estimates consulted during construction of the matrix.

---

* This is an abbreviated version of the general appendix which is linked to region summaries and the data set itself in Putterman (2006), which can be viewed at http://www.econ.brown.edu/fac/Louis_Putterman/ .

Because our interest is in the possible impact of its people's origins on each country's economic performance, we try to identify the origins of long-term residents only, thus leaving out guest or temporary workers. Very little data is available about the duration of stay of most temporary workers, so we made educated guesses as to what portion of the originally temporary residents have become permanent.

The matrix includes entries on all countries existing in 2000 having populations of one half million or larger. A country is included as a source country for ancestors of the people of another country if at least 0.5% of all ancestors alive in 1500 are estimated to have lived there. Some entries smaller than 0.5% are found in the matrix, but these occur as a result of special decompositions applied to populations that our sources identify by ethnic group rather than by country of origin—e.g. Gypsies, Africans (descended from slaves, especially in the Americas), and Ashkenazi Jews. The full appendix details the method of assigning fractions of these populations to individual source countries.

Some of the more important sources from which data were drawn for the construction of the matrix are listed below. See the full appendix and region notes for other sources and details.

Columbia Encyclopedia (online edition)
CIA World Factbook
Countriesquest.com
Encyclopædia Britannica (online edition)
Everyculture.com
Library of Congress, Federal Research Division, Country Studies
MSN Encarta Encyclopedia (online edition)
Nationsencyclopedia.com
World Christian Database (Original source for WCE)
World Christian Encyclopedia