# THE DIFFUSION OF DEVELOPMENT*

ENRICO SPOLAORE AND ROMAIN WACZIARG

We find that genetic distance, a measure associated with the time elapsed since two populations' last common ancestors, has a statistically and economically significant effect on income differences across countries, even controlling for measures of geographical distance, climatic differences, transportation costs, and measures of historical, religious, and linguistic distance. We provide an economic interpretation of these findings in terms of barriers to the diffusion of development from the world technological frontier, implying that income differences should be a function of relative genetic distance from the frontier. The empirical evidence strongly supports this barriers interpretation.

## I. INTRODUCTION

What explains the vast differences in income per capita across countries? This paper provides new empirical evidence shedding light on this question.[1] At the center of our analysis is genetic distance, a measure based on aggregate differences in the distribution of gene variants across populations.[2] For the first time, we document and discuss the relationship between genetic distance and differences in income per capita across countries. We find that measures of genetic distance bear a statistically and economically significant relationship to income differences, and that this relationship is robust to controlling for a large number of measures of geographical distance, climatic differences, transportation costs, and measures of historical, linguistic, and religious distance. The effect of genetic distance holds not only for contemporary income differences, but also for income differences measured since 1500.

---

*We are grateful to Alberto Alesina, Robert Barro, Sam Bowles, Robert Boyd, Francesco Caselli, Dan Cox, Steven Durlauf, James Fearon, Oded Galor, Luigi Guiso, Peter Howitt, Yannis Ioannides, Larry Katz, Pete Klenow, Andros Kourtellos, Ed Kutsoati, Edward Leamer, John Londregan, Peter Lorentzen, Lisa Lynch, Paolo Mauro, Deborah Menegotto, Sharun Mukand, Louis Putterman, Gérard Roland, Fabio Schiantarelli, Antonio Spilimbergo, Susan Stokes, Chih Ming Tan, David Weil, Bruce Weinberg, Ivo Welch, several anonymous referees, and participants at numerous seminars and conferences for helpful comments. We gratefully acknowledge financial support from Stanford University's Presidential Fund for Innovation in International Studies.

[1] Contributions to the literature on the determinants of income per capita using cross-country regressions include Hall and Jones (1999), Acemoglu, Johnson, and Robinson (2001), Easterly and Levine (2003), Alcalá and Ciccone (2004), and Glaeser et al. (2004), among many others.

[2] Our source for genetic distances between human populations is Cavalli-Sforza, Menozzi, and Piazza (1994). Recent textbook references on human evolution are Boyd and Silk (2003) and Jobling, Hurles, and Tyler-Smith (2004). For a nontechnical discussion of these concepts see Dawkins (2004).

*The Quarterly Journal of Economics*, May 2009

Moreover, the effect of genetic distance on income differences is quantitatively large, statistically significant, and robust not only worldwide, but also within Europe, for which more precise measures of cross-country genetic distance are available.

In addition to establishing these facts, we provide an economic interpretation for our findings. What does genetic distance capture, and why is it correlated with income differences, even controlling for geographical distance and other factors? Technically, genetic distance measures the difference in gene distributions between two populations, where the genes under considerations are *neutral*—they change randomly and independent of selection pressure. The rationale for this approach is that divergence in neutral genes provides information about lines of descent. Most random genetic change takes place regularly over time, as in a molecular clock. Therefore, genetic distance measures the time since two populations have shared common ancestors—that is, the time since they have been the same population. In other words, genetic distance is a summary measure of general relatedness between populations. An intuitive analogue is the familiar concept of relatedness between individuals: two siblings are more closely related than two cousins because they share more recent common ancestors—their parents rather than their grandparents.

Because genetic distance is based on neutral change, it is not meant to capture differences in specific genetic traits that directly matter for survival and fitness. Hence, our results provide *no* evidence for a direct effect of specific genes on income or productivity. Our findings are not about some societies having some specific genes that make them directly richer. Instead, our results provide strong evidence that a general measure of genealogical relatedness between populations can explain income differences today, even though it reflects mostly neutral genetic variation. Why? Our interpretation is that genetic distance captures barriers to the diffusion of development. More closely related societies are more likely to learn from each other and adopt each other's innovations. It is easier for someone to learn from a sibling than from a cousin, and easier to learn from a cousin than from a stranger. Populations that share more recent common ancestors have had less time to diverge in a wide range of traits and characteristics that are transmitted across generations with variation. Of course, in human populations many of those traits are transmitted across generations culturally rather than biologically.[3] Similarity in such traits would tend to facilitate communication and understanding, and hence the diffusion and adaptation of complex technological and institutional innovations.

What traits are captured by genetic distance? We argue that, by its very definition, genetic distance is an excellent *summary statistic* capturing divergence in the whole set of implicit beliefs, customs, habits, biases, conventions, etc. that are transmitted across generations—biologically and/or culturally—with high persistence. In a nutshell, human genetic distance can be viewed as a summary measure of very long-term divergence in intergenerationally transmitted traits across populations. Our key hypothesis is that such long-term (and mainly random) divergence has created barriers to the diffusion of technological and institutional innovations across societies in more recent times. Although we provide a general economic interpretation of genetic distance in terms of barriers to the diffusion of development from the frontier, we remain largely agnostic about specific mechanisms of technology diffusion, as well as about the specific traits and characteristics that create the barriers.

If our interpretation is correct, the relevant measure of genetic distance associated with economic distance between two societies should not be the absolute genetic distance between them, but their *relative* distance from the world technological frontier. In our empirical analysis we test this central implication using Britain (in the nineteenth century) and the United States (in the twentieth century) as the world technological frontiers. Consistent with our hypothesis, we find that the effect of relative genetic distance on economic distance is positive, and larger than the effect of absolute genetic distance, itself only an imperfect proxy for relative genetic distance to the frontier. We view this as important evidence in support of a barriers effect. The historical evidence also suggests that the effect of genetic distance on income differences, although always positive and significant since 1500, increased considerably between 1820 and 1870, consistent with a salient role for relative genetic distance during the gradual

---

[3] Classic references on cultural transmission and evolution are Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985). See also Richerson and Boyd (2004). Economic analyses of cultural transmission across generations include Bisin and Verdier (2000, 2001) and others. This issue will be discussed in more detail in Section II of this paper.

spread of the Industrial Revolution. More broadly, our interpretation is consistent with the diffusion of economic development as emerging from the formation of a human web, gradually joined by different cultures and societies as a function of their relative distance from the technological and institutional frontier.[4]

This paper is part of a small but growing set of contributions that use human genetic data in empirical economic analyses.[5] Guiso, Sapienza, and Zingales (2004) used genetic distance between European populations as an instrument for trust in trade gravity regressions.[6] Giuliano, Spilimbergo, and Tonon (2006) directly study the effect of genetic distance on bilateral trade flows in gravity regressions and argue that the effect of genetic distance on trade volume is sensitive to controlling for geography (in contrast, in our analysis the effect of genetic distance on income differences is robust to geographical controls, including transportation costs). A major difference between our paper and these contributions is our focus on the determinants of income differences rather than trade flows. Moreover, we are the first economists to use worldwide measures of genetic distance across human populations, in addition to European data.

Desmet et al. (2007) document the close relationship between genetic distance and cultural differences and argue that genetic distance can be used to study nation formation in Europe. They show a strong and robust correlation between answers to the World Values Survey (WVS) and genetic distance, finding that European populations that are genetically closer give more similar answers to a set of 430 questions about norms, values, and cultural characteristics included in the 2005 WVS sections on perceptions of life, family, religion, and morals. They also find that the correlation between genetic distance and cultural values remains positive and significant after controlling for linguistic and geographic distances. Their empirical analysis supports our interpretation of genetic distance as a broad measure of differences in intergenerationally transmitted characteristics, including cultural values.

---

[4] For a historical overview of the formation of the complex web of exchanges and interactions across human communities going back to the Neolithic period, see McNeill and McNeill (2003).

[5] There also exists a different economic literature that uses genetic distances between species to evaluate biodiversity—for example, Weitzman (1992).

[6] In a more recent version of their paper, Guiso, Sapienza, and Zingales no longer use genetic distance, but rely on a new measure of somatic similarities (see our discussion in Section II.B).

Finally, Ashraf and Galor (2008) study the relationship between genetic diversity within a society and economic development in precolonial times, measured by population density, and find a nonmonotonic relation between genetic diversity and population density (a higher population density is associated with intermediate levels of genetic diversity). Their paper shares our focus on economic development, but considers different genetic data and effects (genetic heterogeneity within each society), whereas we study the effect of genetic distance between societies.

In Section II we present a simple framework in which genetic distance captures divergence in characteristics that are transmitted across generations within populations over the long run, and those differences act as barriers to the diffusion of development from the world technological frontier. The section also contains a general taxonomy of the mechanisms linking genetic distance and economic outcomes. Section III presents the data on genetic distance. In Section IV we present and discuss our empirical findings. Section V concludes.

## II. A CONCEPTUAL FRAMEWORK

In this section we present an analytic framework linking genetic distance, intergenerationally transmitted traits, and the diffusion of economic development from the technological frontier. Our analysis leads to a testable prediction: Income differences across societies should depend on their *relative* genetic distance from the technological frontier. In Section IV, we show that the empirical evidence strongly supports this prediction.

The main building block of our model is that genetic distance between populations captures the degree of genealogical relatedness of different populations over time. Thus, it can be interpreted as a general metric for average differences in characteristics transmitted across generations. In this paper we call *vertically transmitted characteristics* (or vertical characteristics) the set of characteristics passed on across generations within a population over the very long run—that is, over the time horizon along which populations have diverged.[7]

---

[7] This terminology is borrowed from the evolutionary literature on cultural transmission (Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985; Shennan 2002; Richerson and Boyd 2004). Such vertical transmission takes place across generations within a given population, and, in our definition, includes not only direct parent-to-child transmission, but also "oblique" transmission from

This leads to our second main idea: differences in vertical characteristics act as barriers to the diffusion of productivity-enhancing innovations across societies.[8] We argue that populations that share a more recent common history, and are therefore closer in terms of vertical characteristics, face lower costs and obstacles to adopting each other's innovations.[9] We are interested primarily in the diffusion of economic development in historical times, and especially after the Industrial Revolution. Thus, in our empirical analysis we focus on differences in vertical characteristics as barriers to the diffusion of development from the modern technological frontier.[10]

### II.A. A Simple Model

A stylized model illustrates our ideas in the simplest possible way.[11] Consider three periods ($o$ for "origin," $p$ for "prehistory," and $h$ for "history"). In period $o$ there exists only one population (population 0). In period $p$ the original population splits into two populations (1 and 2). In period $h$ each of the two populations splits in two separate populations again (population 1 into 1.1 and 1.2, and population 2 into 2.1 and 2.2), as displayed in Figure I. In this setting the genetic distance $d_g(i, j)$ between population $i$ and population $j$ can be simply measured by the number of periods

---

other genetically related people within the group. Hence, our definition is broader than usages of the term strictly limited to parent-to-child transmission. We thank Robert Boyd for pointing out this distinction to us.

[8] Policy-induced barriers to the diffusion of technology are analyzed by Parente and Prescott (1994, 2002). In our framework we interpret barriers more broadly to include all long-term societal differences that are obstacles to the diffusion of development.

[9] The idea that differences in long-term societal characteristics may act as barriers to the diffusion of development is stressed in a large literature on the diffusion of innovations, including the classic book by Rogers (1962). For a historical comparative analysis of managerial innovations and performance, see Clark and Wolcott (1999).

[10] World technological leadership since the British Industrial Revolution (1700s) has been predominantly associated with Britain and, by the late 1800s, the United States (Brezis, Krugman, and Tsiddon 1993). In the years right before the Industrial Revolution, the technological frontier was probably the Netherlands. According to Maddison (2003), in previous times the regions with the highest levels of income per capita were Italy (around 1500) and China (around 1000). We return to these issues in Section IV.

[11] In a previous version of this paper (available upon request), we presented a dynamic micro-founded extension of this model, built on Barro and Sala-i-Martin (1997). In that extension, imitation costs were a function of distance in vertical characteristics from the technological frontier, and hence income differences were a function of relative genetic distance in steady state.

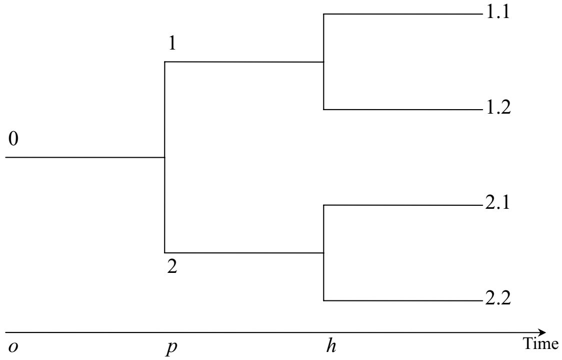

FIGURE I
Population Tree

since they were one population:

$$d_g(1.1, 1.2) = d_g(2.1, 2.2) = 1 \tag{1}$$

and

$$d_g(1.1, 2.1) = d_g(1.1, 2.2) = d_g(1.2, 2.1) = d_g(1.2, 2.2) = 2. \tag{2}$$

For simplicity, all vertical characteristics of a population are summarized as a point on the real line (i.e., a population $i$ has vertical characteristics $v_i$, where $v_i$ is a real number). Populations inherit characteristics from their ancestor populations with variations—a population $i'$ descending from a population $i$ will have characteristics

$$v_{i'} = v_i + \eta_{i'}. \tag{3}$$

Consider the simplest possible mechanism for variation: vertical change as a random walk—for every population $i', \eta_{i'}$ takes value $\eta > 0$ with probability 1/2 and $-\eta$ with probability 1/2.[12] Consequently, the distance in vertical characteristics between two populations $d_v(i, j) \equiv |v_j - v_i|$ is, on average, increasing in their genetic distance $d_g(i, j)$.[13] This captures our first main idea.

---

[12] This simplification is consistent with the molecular-clock interpretation of genetic distance itself. While more complex processes could be considered, this formalization has two advantages: it is economical ("Occam's razor"), and it illustrates how neutral random changes are sufficient to generate our theoretical predictions.

[13] Specifically, in period $h$ the expected difference in vertical characteristics between populations at a genetic distance equal to 2 and populations

Our second main idea can also be captured within this simplified setting. Assume that in periods $o$ and $p$ all populations produce output using the basic technology $Y = A_0L$, so that all populations have the same income per capita $y_0 = A_0$. In period $h$ a population happens to find a more productive technology $A_1 = A_0 + \Delta$, where $\Delta > 0$.[14] Denote this population, the technological frontier, as $f$.[15] We assume that populations farther from population $f$ in terms of vertical characteristics face higher barriers to adopting the new technology. To fix ideas, assume that a society $i$ at a vertical distance from the frontier equal to $d_v(i, f)$ can improve its technology only by

$$\Delta_i = [1 - \beta d_v(i, f)]\Delta, \tag{4}$$

where the parameter $\beta > 0$ captures the barriers to the horizontal diffusion of innovations due to distance in vertical characteristics.[16] Hence, income per capita in society $i$ is given by

$$y_i = A_0 + [1 - \beta d_v(i, f)]\Delta. \tag{5}$$

This implies that the economic distance between population $i$ and population $j$, measured by their income difference $d_e(i, j) \equiv |y_i - y_j|$, is a function of their *relative* vertical distance from the frontier $|d_v(i, f) - d_v(j, f)|$:

$$d_e(i, j) \equiv |y_j - y_i| = \beta\Delta|d_v(i, f) - d_v(j, f)|. \tag{6}$$

As we have shown, vertical difference $d_v(i, j)$ and genetic distance $d_g(i, j)$ are positively correlated. Therefore, on average, income differences across societies are increasing in their relative genetic

---

at a genetic distance equal to 1 is given by $E\{d_v(i, j)\ |d_g(i, j) = 2\} - E\{d_v(i, j)\ |d_g(i, j) = 1\} = \eta/2 > 0$. Of course, this is not a deterministic relationship. Some pairs of populations that are genealogically more distant may end up with more similar vertical characteristics than two more closely related populations, but that outcome is less likely to be observed than the opposite. On average, genetic distance and distance in vertical characteristics go hand in hand.

[14] We abstract from the possibility that the likelihood of finding the innovation may itself be a function of a society's vertical characteristics. Such direct effects of vertical characteristics would strengthen the links between genetic distance and economic outcomes, but are not necessary for our results.

[15] The model can be viewed as a very reduced form of a dynamic process in which the frontier economy produces several innovations, including improvements to the innovation process itself, in the spirit of the observation that "the greatest invention of the 19th century was the invention of the method of invention" (Alfred North Whitehead [1931, p. 38], quoted in Howitt and Mayer-Foulkes [2005]).

[16] Without loss of generality, we assume that $\beta$ is lower than 1/2. Alternatively, the formula could be rewritten as $\Delta_i = \max\{[1 - \beta d_v(i, f)]\Delta, 0\}$.

distance from the frontier society. Formally,

$$\begin{aligned}&E\{d_e(i,j)||d_g(i,f)-d_g(j,f)|=2\}\\ &-E\{d_e(i,j)||d_g(i,f)-d_g(j,f)|=1\}=\frac{\eta\beta\Delta}{3}>0.\end{aligned} \tag{7}$$

This result is intuitive. As we increase relative genetic distance from the frontier, the expected income gap increases. The size of the effect is a positive function of the extent of divergence in vertically transmitted characteristics ($\eta$), the extent to which this divergence constitutes a barrier to the horizontal diffusion of innovations ($\beta$), and the size of the improvement in productivity at the frontier ($\Delta$).

Our framework predicts a positive correlation between economic distance $|y_j - y_i|$ and relative genetic distance from the frontier $|d_g(i,f)-d_g(j,f)|$. It also accounts for a positive correlation between economic distance and simple genetic distance $d_g(i,j)$ as long as $|d_g(i,f)-d_g(j,f)|$ and $d_g(i,j)$ are positively correlated.[17] At the same time, our theory predicts that relative genetic distance from the frontier should have a stronger impact on economic distance than absolute genetic distance, because relative distance is a more accurate measure of relative distance from the frontier in terms of vertical characteristics. In fact, the expected economic distance associated with an absolute genetic distance $d_g(i,j)=1$ is $E\{d_e(i,j)|d_g(i,j)=1\}=\eta\beta\Delta$, whereas the expected economic distance associated with an equivalent level of relative genetic distance $|d_g(i,f)-d_g(j,f)|=1$ is higher:[18]

$$\begin{aligned}&E\{d_e(i,j)||d_g(i,f)-d_g(j,f)|=1\}\\ &=\frac{7\eta\beta\Delta}{6}>E\{d_e(i,j)||d_g(i,j)=1\}.\end{aligned} \tag{8}$$

In summary, our theory has the following testable implications:

IMPLICATION 1. Relative genetic distance from the frontier is positively correlated with differences in income per capita (economic distance).

---

[17] It is easy to verify that the two measures are positively correlated in our theoretical framework. More importantly, relative genetic distance from the frontier and absolute genetic distance are also positively correlated in the actual data, as we show in Section IV. Our framework provides an explanation for the observed positive correlation between economic distance and absolute genetic distance in the data: absolute genetic distance is an imperfect proxy of the economically relevant variable, relative genetic distance.

[18] An analogous relationship exists between $E\{d_e(i,j)\,||d_g(i,f)-d_g(j,f)|=2\}$ and $E\{d_e(i,j)\,||d_g(i,j)=2\}$.

IMPLICATION 2. The effect on income differences associated with relative genetic distance from the frontier is larger than the effect associated with absolute genetic distance.

As we will see in Section IV, both predictions are consistent with the empirical evidence.

### *II.B. A General Taxonomy*

To clarify the nature of the links between genetic distance and income differences, it is useful to introduce a broader classification of different mechanisms through which the transmission of characteristics across generations may in principle affect economic outcomes.

In general, traits can be transmitted across generations through DNA (call it "genetic transmission," or GT—e.g., eye color) or through pure cultural interactions (call it "cultural transmission," or CT—e.g., a specific language). Moreover, vertical characteristics, whether passed on through GT or CT, may affect income differences because of a direct (D) effect on productivity or because they constitute barriers (B) to the transmission of innovations across populations. There are four possible combinations of mechanisms through which intergenerationally transmitted characteristics may affect income differences. The following chart summarizes the four possibilities:

| | Direct effect (D) | Barrier effect (B) |
|---|---|---|
| Genetic transmission (GT) | Quadrant I | Quadrant II |
| Cultural transmission (CT) | Quadrant III | Quadrant IV |

For instance, genetic traits affecting the trade-off between quality and quantity of children in the theoretical framework proposed by Galor and Moav (2002) would be examples of GT direct effects (Quadrant I).[19] GT barrier effects (Quadrant II) could stem from visible characteristics (say, physical appearance) that do not affect productivity directly, but introduce barriers to the diffusion of innovations by reducing exchanges and learning across populations that perceive each other as different. This effect is related to the already cited study by Guiso, Sapienza, and Zingales (2004), who argue that differences in physical characteristics affect the extent of trust across populations, and that trust affects bilateral trade between different societies. Consistent with this view, in the

---

[19] For a discussion of related ideas, see also the recent book by Clark (2007) on the causes of the Industrial Revolution.

most recent version of their paper these authors use a measure of somatic similarity to instrument for trust, on the grounds that people tend to trust people who look similar to them physically. Direct economic effects of cultural characteristics have been emphasized in a vast sociological literature that goes back at least to Max Weber. A recent empirical study of the relationship between cultural values and economic outcomes that is consistent with the mechanisms of Quadrant III is provided by Tabellini (2005). Guiso, Sapienza, and Zingales (2006) define culture as "customary beliefs and values that ethnic, religious and social groups transmit fairly unchanged from generation to generation" and provide an extensive discussion of the links between cultural variables and economic outcomes. The link between differences in vertically transmitted characteristics—including cultural characteristics, as in Quadrant IV—is at the core of our own model. In the model presented in Section II.A, differences in neutral characteristics (traits that do not have a direct effect on productivity) explain income differences by acting as barriers to the diffusion of innovation across populations.

The distinction between GT and CT is useful to fix ideas, but is not a clear-cut dichotomy. In fact, this distinction (related to the distinction between nature and nurture), if taken too literally, may be misleading from an economic as well as from a biological perspective. Generally, the economic effects of human characteristics are likely to result from interactions of cultural and genetic factors, with the effects of genetic characteristics on economic outcomes changing over space and time depending on cultural characteristics, and vice versa. To illustrate this point, consider differences across individuals within a given population (say, the United States). Consider a clearly genetic characteristic, for instance having two X chromosomes, the purely genetic characteristic associated with the female sex. This characteristic is likely to have had very different effects on a person's income and other economic outcomes in the year 1900 and in the year 2000, because of changes in culturally transmitted characteristics over the century. This is a case where the impact of genes on outcomes varies with a change in cultural characteristics.[20] By the same token, one can think of the differential impact of a given cultural

---

[20] This is a variation on an example by Alison Gopnik in her comment to the Pinker vs. Spelke debate at http://www.edge.org/discourse/science-gender.html#ag. Pinker's response is also available at http://www.edge.org/discourse/science-gender.html.

characteristic (say, the habit of drinking alcohol) on individuals with different genetic characteristics (say, genetic variation in alcohol dehydrogenase, the alcohol-metabolizing enzyme). An example of a complex interaction in which culture affects genes is the spread of the gene for lactose tolerance in populations that domesticated cows and goats. Hence, in interpreting our empirical results we do not dwell much on the distinction between genetic and cultural transmission of traits, but instead interpret genetic distance as an overall measure of differences in the whole set of intergenerationally transmitted characteristics.[21]

## III. THE GENETIC DISTANCE DATA

### *III.A. Measuring Genetic Distance*

Because the data on genetic distance are not commonly used in the economics literature, we describe them in some detail. Genetic distance measures genetic differences between two populations. The basic unit of analysis is the *allele*, which is a particular form taken by a gene.[22] By sampling populations for specific genes that can take different forms, geneticists have compiled data on allele frequencies.[23] Differences in allele frequencies are the basis for computing summary measures of distance between populations. Following Cavalli-Sforza, Menozzi, and Piazza (1994), we use measures of $F_{\mathrm{ST}}$ distance, also known as "coancestor coefficients." $F_{\mathrm{ST}}$ distances, like most measures of genetic differences, are based on indices of heterozygosity, the probability that two

---

[21] That said, we do find clues pointing to cultural transmission, rather than purely biological transmission, as a likely mechanism behind our results. For instance, we find large effects of genetic distance on income differences within Europe, among populations that are geographically close, have shared very similar environments, and have had a very short time to diverge genetically. The view that cultural transmission trumps genetic transmission in explaining differences within human populations is standard among geneticists and anthropologists. For nontechnical discussions of these issues, see Diamond (1992), Cavalli-Sforza and Cavalli-Sforza (1995), Diamond (1997), and Richerson and Boyd (2004).

[22] A gene is commonly defined as a DNA sequence that encodes for a protein. The genetic data in Cavalli-Sforza, Menozzi, and Piazza (1994) have been obtained from "classical analysis," which focuses on protein polymorphism. More recent approaches look directly at the DNA. So far those studies, which include the Human Genome Diversity Project (http://www.stanford.edu/group/morrinst/hgdp.html) and the International HapMap Project (http://www.hapmap.org/), have confirmed the results from classical protein analysis, but are not yet available for extensive cross-regional analysis.

[23] Allele frequencies for various genes and for most populations in the world can be found at http://alfred.med.yale.edu/.

alleles at a given locus selected at random from two populations will be different. $F_{\mathrm{ST}}$ takes a value equal to zero if and only if the allele distributions are identical across the two populations, whereas it is positive when the allele distributions differ. A higher $F_{\mathrm{ST}}$ is associated with larger differences.[24]

Measures of genetic distance can be used to reconstruct phylogenies (or family trees) of human populations. $F_{\mathrm{ST}}$ is strongly related to how long two populations have been isolated from each other.[25] When two populations split apart, their genes can start to change as a result either of random genetic drift or natural selection. When calculating genetic distances to study population history and phylogenesis, geneticists concentrate on *neutral* characteristics that are not affected by strong directional selection, but only by random drift.[26] Importantly, our measures of genetic distance are based on such neutral markers only, and not on selected traits. When populations become separated, the process of random drift will take them in different directions, raising their genetic distance. The longer the period of separation, the greater the genetic distance becomes. If drift rates are constant, genetic distance can be used as a molecular clock—that is, the time elapsed since two populations separated can be measured by the genetic distance between them.[27] Consequently, $F_{\mathrm{ST}}$ is a measure of distance to the most recent common ancestors of two populations, or, equivalently, of their degree of genealogical relatedness.

To summarize, we use $F_{\mathrm{ST}}$ distance as a measure of genealogical relatedness between populations. A larger $F_{\mathrm{ST}}$ distance reflects a longer separation between populations, and hence, on average, a larger difference in vertical characteristics.

---

[24] Appendix I provides an illustration of the construction of $F_{\mathrm{ST}}$ for the simple case of two populations of equal size, and one gene that can take only two forms (i.e., two alleles).

[25] Isolation here refers to the bulk of the genetic heritage of a given population. As stressed by Cavalli-Sforza, Menozzi, and Piazza (1994), small amounts of intermixing between members of different populations do not affect measured genetic distance.

[26] Cavalli-Sforza, Menozzi, and Piazza (1994, p. 36). The classic reference for the neutral theory of molecular evolution is Kimura (1968). For more details on the neutral theory, the *molecular clock* hypothesis, and the construction and interpretation of measures of genetic distance, see Jobling, Hurles, and Tyler-Smith (2004).

[27] When genetic distance is based on neutral markers, and populations are sufficiently large, geneticists have shown that drift rates are indeed constant (very small populations are generally subject to faster random genetic drift).

### *III.B. The World Sample*

The genetic distance data are from Cavalli-Sforza, Menozzi, and Piazza (1994, pp. 75–76). Our main focus is on the set of 42 world populations for which they report all bilateral distances, computed from 120 alleles.[28] These populations are aggregated from subpopulations characterized by a high level of genetic similarity. However, measures of bilateral distance among these subpopulations are available only regionally, not for the world as a whole. Among the set of 42 world populations, the greatest genetic distance observed is between Mbuti Pygmies and Papua New Guineans, where the $F_{ST}$ distance is 0.4573, and the smallest is between the Danish and the English, where the genetic distance is 0.0021.[29] The mean genetic distance among the 861 available pairs is 0.1338. Figure II, from Cavalli-Sforza, Menozzi, and Piazza (1994, Figure 2.3.2B, p. 78), is a phylogenetic tree illustrating the process by which different human populations have split apart over time.[30] Such phylogenetic trees, constructed from genetic distance data, are the population analogs of family trees for individuals.

Genetic distance data are available at the population level, not at the country level. It was thus necessary to match

---

[28] Cavalli-Sforza, Menozzi, and Piazza (1994) also provide a different measure of genetic distance (Nei's distance). $F_{ST}$ and Nei's distance have slightly different theoretical properties, but their correlation (93.9%) is very high (Table I). We show below that the choice of measures does not affect our results.

[29] Among the more disaggregated data for Europe that we also gathered, the smallest genetic distance (equal to 0.0009) is between the Dutch and the Danish, and the largest (equal to 0.0667) is between the Lapps and the Sardinians. The mean genetic distance across European populations is 0.013. Genetic distances are roughly ten times smaller on average across populations of Europe than in the world data set.

[30] The figure was constructed to maximize the correlation between Euclidean distances to common nodes, measured along the branches, and the $F_{ST}$ genetic distance computed directly from allele frequencies. Hence, the tree diagram is a simplified summary of (but not a substitute for) the matrix of genetic distances between populations, organized by clusters. It is important to notice that the organization of populations by tree does not imply that genetic distance establishes a linear relation among all of them, either along the $x$-axis (abscissa) or along the $y$-axis (ordinate). The abscissa at the bottom of the diagram can be used to read the genetic distance between pairs of populations in the tree only when they share direct common ancestors. For example, the genetic distance between New Guineans and Australians can be calculated by reading the position of the node that separates the two populations, which is approximately at 0.1. It is also possible to measure average genetic distance between clusters of populations by reading the position of the node that separates two clusters on the abscissa. For example, the average genetic distance between African populations and the rest of the world is approximately 0.2. However, to read the genetic distance between any pair of populations, one should use (as we do) the complete matrix of genetic distances, which is provided in Cavalli-Sforza, Menozzi, and Piazza (1994, Table 2.3.1A, p. 75).

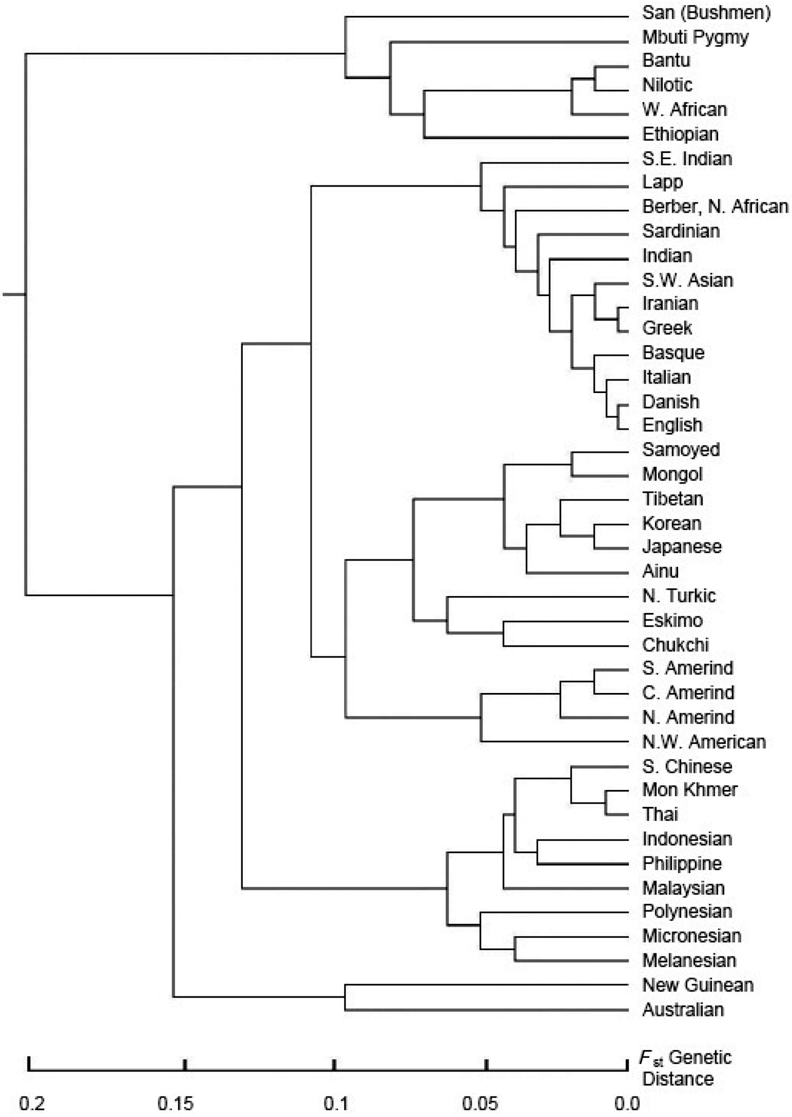

FIGURE II

Genetic Distance among 42 Populations

*Source.* Cavalli-Sforza, Menozzi, and Piazza (1994).

populations to countries. We did so using ethnic composition data by country from Alesina et al. (2003). It was possible to match ethnic group labels with population labels from Cavalli-Sforza, Menozzi, and Piazza (1994), using their Appendices 2 and 3 to identify the ethnic groups sampled to obtain genetic distances. Obviously, many countries feature several ethnic groups. Alesina et al. list 1,120 country-ethnic group categories. We matched virtually all of these categories to some genetic group. The only groups that were not matched were the ones that were not labeled in Alesina et al.—usually residual groups labeled "other" that represented a small share of a country's population.

As an example, the Alesina et al. (2003) data on ethnic groups has India composed 72% of "Indo-Aryans" and 25% of "Dravidians." These groups were matched, respectively, to the Cavalli-Sforza groups labeled "Indians" and "Dravidians" (i.e., S.E. Indian in Figure II). The residual category "India Other" (3% of the population) was not matched to any genetic group. Another example is Italy, where the ethnic groups labelled "Italian" and "Rhaetians" (a combined 95.4% of the population) were matched to the genetic category "Italian," whereas the "Sardinians" ethnic group (2.7% of the population) was matched to the "Sardinian" genetic group.[31]

This match served as the basis for constructing measures of genetic distance between countries, rather than groups. We constructed two such measures. The first was the distance between the plurality ethnic groups of each country in a pair, that is, the groups with the largest shares of each country's population. In the examples above, that means that the plurality genetic distance between India and Italy is the genetic distance between the Indian and the Italian genetic groups ($F_{\mathrm{ST}} = 0.026$). This resulted in a data set of 21,321 pairs of countries (207 underlying countries and dependencies) with available genetic distance data.[32] The second was a measure of weighted genetic distance. Many countries, such as the United States and Australia, are made up of subpopulations that are genetically distant, and for which both genetic distance data and data on the shares of each genetic group are

---

[31] The complete match of genetic groups to ethnic groups, and in turn to countries, is available upon request.

[32] For 27 countries, the data on group shares were missing from Alesina et al.'s (2003) database, but a match to genetic groups based on plurality groups was possible through information from the *Encyclopaedia Britannica*. Thus, the weighted measure of genetic distance covers 16,110 pairs, or 180 countries—27 fewer than the plurality match.

available. Assume that country 1 contains populations $i = 1, \ldots, I$ and country 2 contains populations $j = 1, \ldots, J$; denote by $s_{1i}$ the share of population $i$ in country 1 (similarly for country 2) and by $d_{ij}$ the genetic distance between populations $i$ and $j$. The weighted $F_{\mathrm{ST}}$ genetic distance between countries 1 and 2 is then

$$F_{\mathrm{ST}}^{W} = \sum_{i=1}^{I} \sum_{j=1}^{J} (s_{1i} \times s_{2j} \times d_{ij}), \tag{9}$$

where $s_{ki}$ is the share of group $i$ in country $k$ and $d_{ij}$ is the $F_{\mathrm{ST}}$ genetic distance between groups $i$ and $j$.[33] The interpretation of this measure is straightforward: it represents the expected genetic distance between two randomly selected individuals, one from each country. Weighted genetic distance is very highly correlated with genetic distance based on dominant groups (the correlation is 94%), so for practical purposes it does not make a big difference which one we use. We will use the weighted $F_{\mathrm{ST}}$ distance as the baseline measure throughout this study, as it is a more precise measure of average genetic distance between countries.

Error in the matching of populations to ethnic groups should lead us to understate the correlation between genetic distance and income differences. Several regions may be particularly prone to matching errors. One is Latin America, where it is sometimes difficult to identify whether populations are predominantly of European descent or of Amerindian descent. This is particularly problematic in countries with large proportions of Mestizos, that is, populations of mixed descent, such as Colombia (in this specific case the country's dominant group was matched to the South Amerindian category). Another is Europe, where countries can only be matched to one of four genetic groups (Danish, English, Greek, and Italian). As a strict rule, we matched countries to groups that were the closest genetically to that country's population, using data on regional genetic distance from Cavalli-Sforza, Menozzi, and Piazza (1994).

The ethnic composition in Alesina et al. (2003) refers to the 1990s. This is potentially endogenous with respect to current

---

[33] When some ethnic category was not matched to a genetic group due to a missing ethnic label in the Alesina et al. (2003) source data, the populations shares were rescaled to sum to 1 for the purpose of calculating weighted distances. Thus, for instance, the weighted genetic distance between India and Italy was calculated as $F_{\mathrm{ST}}^{W} = (0.972 \times 0.258 \times 0.0402) + (0.972 \times 0.742 \times 0.0261) + (0.028 \times 0.258 \times 0.0531) + (0.028 \times 0.742 \times 0.0449) = 0.0302$.

income differences if the latter are persistent and if areas with high income potential tended to attract European immigration since 1500. This would be the case, for example, under the view that the Europeans settled in the New World due to a favorable geographical environment.[34] To construct genetic distance between countries as of 1500, we also mapped populations to countries using their ethnic composition as of 1500, prior to the major colonizations of modern times. Thus, for instance, although the United States is classified as predominantly populated with English people for the current match, it is classified as being populated with North Amerindians for the 1500 match. This distinction affected mostly countries that were colonized by Europeans since 1500 to the point where the dominant ethnic group is now of European descent (New Zealand, Australia, North America, and some countries in Latin America). Because we do not have data on ethnic composition going back to 1500, the corresponding match refers only to plurality groups. Genetic distance in 1500 can be used as a convenient instrument for current genetic distance. The matching of countries to populations for 1500 is also more straightforward than for the current period, because Cavalli-Sforza, Menozzi, and Piazza (1994) attempted to sample populations as they were in 1500, likely reducing the extent of measurement error.

### *III.C. The European Sample*

Cavalli-Sforza, Menozzi, and Piazza (1994) also present matrices of genetic distance among populations within several regions. These submatrices cannot be merged with the world data, because they are based on sets of underlying genes distinct from the 120 genes used for the 42 populations in the world sample, and because the genetic distance between most groups in the regional samples and in the world sample are unavailable. They can, however, be used separately. We assembled a data set of genetic distances between 26 European populations, a much finer classification than the world sample, which only featured four distinct (nonminority) European populations (English, Danish,

---

[34] In fact, income differences are not very persistent for a long time horizon such as this—see Acemoglu, Johnson, and Robinson (2002). Our own data show that pairwise log income differences in 1500 are uncorrelated with the 1995 series in the common sample (Table II).

Italian, and Greek).[35] Matching populations to countries is more straightforward for the European sample than for the world sample, because the choice of sampled European populations generally corresponds to nation-state boundaries. This should reduce the incidence of measurement error. The populations were matched to 26 countries, resulting in 325 country pairs.[36] The largest $F_{\mathrm{ST}}$ genetic distance among those pairs was 0.032, between Iceland and Slovenia. The smallest, among countries matched to distinct genetic groups, was between Denmark and the Netherlands ($F_{\mathrm{ST}} = 0.0009$).

## IV. THE EMPIRICS OF INCOME DIFFERENCES

In this section we test the empirical implications of our model. We investigate the relationship between genetic distance and economic distance. Genetic distance is considered both relative to the technological frontier and in absolute terms. In line with our theory, we use log income per capita as a metric of economic performance. The data on per capita income are purchasing power-parity adjusted data from the World Bank, for the year 1995.[37]

### *IV.A. Genetic Distance to the Frontier*

We start with a simple descriptive approach. Does a country's genetic distance to the world technological frontier correlate with its income level? To investigate this hypothesis, we run income level regressions, for now confining our attention to the world sample, where we have data on all our variables for 137 countries. We consider the United States as the technological frontier in 1995. We measure distance to the United States using our weighted measure, which is more appropriate because the United

---

[35] Minority populations in the world sample also include Basque, Lapp, and Sardinian.

[36] These 26 countries are Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Macedonia, the Netherlands, Norway, Poland, Portugal, Russian Federation, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, the United Kingdom, and Serbia/Montenegro. The Basque, Lapp, and Sardinian populations were not matched to any country, and some countries were matched to the same groups (Croatia, Slovenia, Macedonia, and Serbia/Montenegro were all matched to the Yugoslavian population, whereas the Czech and Slovak Republics were both matched to the Czech population).

[37] We also used data from the Penn World Tables version 6.1 (Heston, Summers, and Aten 2002), which made little difference in the results. We focus on the World Bank data for 1995, as this allows us to maximize the number of countries in our sample.

TABLE I
INCOME LEVEL REGRESSIONS, WORLD DATA SET

| | (1) Univariate | (2) Add geographic distance | (3) Add linguistic and religious distance |
|---|---|---|---|
| $F_{ST}$ genetic distance to the United States, weighted | −12.906 | −12.523 | −10.245 |
| | (1.383)** | (1.558)** | (1.567)** |
| Absolute difference in latitude from the United States | | 1.970 | 1.518 |
| | | (0.868)** | (0.827)* |
| Absolute difference in longitude from the United States | | 0.438 | 0.786 |
| | | (0.454) | (0.401)* |
| Geodesic distance from the United States (1,000s of km) | | −0.179 | −0.191 |
| | | (0.075)** | (0.071)** |
| =1 for contiguity with the United States | | 1.055 | 0.452 |
| | | (0.300)** | (0.390) |
| =1 if the country is an island | | 0.505 | 0.362 |
| | | (0.397) | (0.483) |
| =1 if the country is landlocked | | −0.384 | −0.410 |
| | | (0.206)* | (0.198)** |
| =1 if the country shares at least one sea or ocean with the United States | | −0.201 | −0.080 |
| | | (0.197) | (0.171) |
| Freight rate to northeastern United States (surface transport) | | 3.460 | 5.794 |
| | | (2.507) | (2.816)** |
| Linguistic distance to the United States, weighted | | | −0.520 |
| | | | (0.648) |
| Religious distance to the United States, weighted | | | −2.875 |
| | | | (0.591)** |
| Constant | 9.421 | 8.876 | 10.499 |
| | (0.149)** | (0.536)** | (0.751)** |
| Observations | 137 | 137 | 137 |
| Adjusted $R^2$ | .39 | .46 | .53 |

*Note*. Dependent variable: log income per capita 1995.
Robust standard errors in parentheses.
*Significant at 10%.
**Significant at 5%.

States is a genetically diverse country (variation in this measure is dominated by distance to the English population).

Table I presents the results. In column (1), genetic distance to the United States is entered alone, and the coefficient has the expected negative sign and is highly significant statistically, with a $t$-statistic of about 9.3. In this specification, genetic distance entered alone accounts for 39% of the variation in log income levels. Figure III displays the univariate results of column (1) graphically. Columns (2) and (3) add several controls for geographic

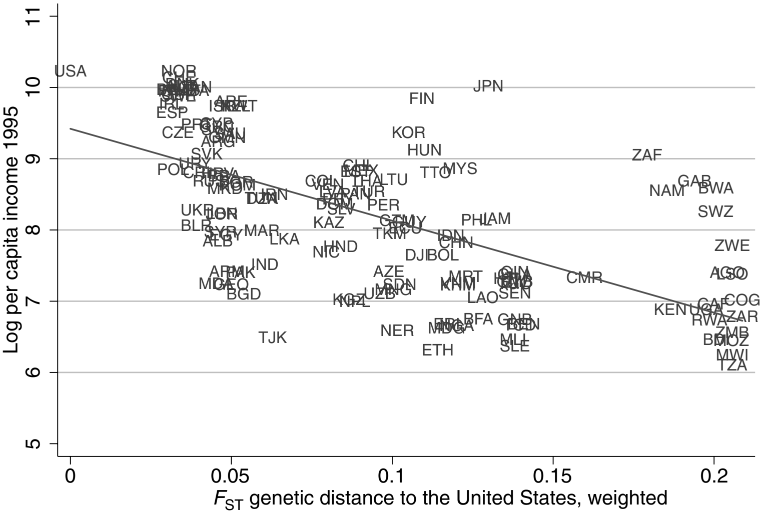

FIGURE III
Log Income in 1995 and Genetic Distance to the United States

distance from the United States and transport costs (column (2)), as well as linguistic and religious differences (column (3)). We will say a lot more about these control variables below, but for now it suffices to note that the coefficient on genetic distance is barely affected by the inclusion of the geographic distance controls, and that its magnitude is reduced by 20% by including linguistic and religious distance. The latter finding is consistent with our interpretation of genetic distance as capturing a broad set of vertical characteristics, including but not limited to language and religion. We return to this important topic in Section IV.F.

### *IV.B. Bilateral Approach*

To generalize the results of the previous section, we consider a specification in which the absolute difference in income between pairs of countries is regressed on measures of distance between the countries in this pair. In addition to being closer to our theoretical specification, this has two advantages. First, we can now investigate the correlation between absolute genetic distance and income differences. (Section II.A led to predictions about the sign of this correlation.) Second, we can make more efficient use of a wealth of bilateral distance data as regressors. We will use this bilateral approach for the rest of this paper.

We computed income differences between all pairs of countries for which income and other data were available, that is, 9,316 pairs (based on 137 underlying countries) in the world sample, and 325 pairs (based on 26 underlying countries) in the Europe sample. Define $G_{ij}^D$ as the *absolute* genetic distance between countries $i$ and $j$. Denote as $G_{ij}^R$ the genetic distance between $i$ and $j$ *relative* to the technological frontier (in most of what follows, the technological frontier is the United States). Then, by definition $G_{ij}^R = |G_{i,\mathrm{US}}^D - G_{j,\mathrm{US}}^D|$.[38] Our baseline specifications are

$$|\log y_i - \log y_j| = \beta_0 + \beta_1 G_{ij}^D + \beta_2' X_{ij} + \varepsilon_{ij} \tag{10}$$

and

$$|\log y_i - \log y_j| = \gamma_0 + \gamma_1 G_{ij}^R + \gamma_2' X_{ij} + \nu_{ij}, \tag{11}$$

where $X_{ij}$ is a set of measures of geographic and cultural distance and $\varepsilon_{ij}$ and $\nu_{ij}$ are disturbance terms.[39]

By using income differences rather than a single country's income level on the left-hand side, we can use bilateral measures of distance between countries on the right-hand side. Our regression is not directional: our specifications are not simply obtained by differencing level regressions across pairs of countries.[40] We should also stress that our specifications are reduced forms. Differences in income are presumably the result of differences in institutions, technologies, human capital, savings rates, etc., all of which are possibly endogenous with respect to income differences, and themselves a function of geographic and human barriers.

---

[38] Absolute and relative genetic distance are algebraically the same when one of the two countries in a pair is the frontier economy, and the two measures are also closely correlated when pairs involving one country that is very close genetically to the frontier economy are considered. The measures differ most for countries that are genetically far from each other (e.g., Ethiopia and Nigeria) but roughly equally distant from the frontier—in this case absolute distance is large, and relative distance is small. Relative genetic distance is meant to capture the fact that *per se* the distance between Nigeria and Ethiopia does not matter in explaining their income difference, because they are unlikely to learn frontier technologies from each other. Rather, what matters is their relative distance from the United States.

[39] We also estimated an alternative specification where the distance measures were all entered in logs. This did not lead to appreciable differences in the economic magnitude or statistical significance of any of the estimates. Because several countries were matched to the same genetic group, so that the corresponding pairs had a genetic distance of zero, taking logs resulted in the loss of valuable observations, so we omit these results here.

[40] Our methodology is as much akin to gravity regressions in the empirical trade literature as it is to levels or growth regressions in the literature on comparative development.

Before turning to the results, we must address a technical point regarding the disturbances $\varepsilon_{ij}$ and $\nu_{ij}$. In principle, if one is willing to assume that the measures of barriers are exogenous, equations (10) and (11) can be estimated using least squares. However, in this case the usual methods of inference will be problematic due to spatial correlation resulting from the construction of the dependent variable.[41] Appendix II illustrates why using the difference in log income as the dependent variable results in spatial correlation.

To address the problem of spatial correlation, we rely on two-way clustering of the standard errors, following the approach in Cameron, Gelbach, and Miller (2006). In our application, clustering arises at the level of country 1 and at the level of country 2, and is non-nested: each individual observation on income differences, say $|\log y_i - \log y_j|$ belongs to the group that includes country $i$ and the group that includes country $j$. The estimator in Cameron, Gelbach, and Miller (2006) allows for an arbitrary correlation between errors that belong to "the same group (along either dimension)" (p. 7). Their method is therefore directly applicable to the specific econometric issue we face (on p. 3 of their paper the authors specifically mention spatial correlation as a possible application of their estimator). Results obtained with this method feature standard errors that are an order of magnitude larger than those obtained with simple OLS with heteroscedasticity-robust standard errors, suggesting that spatial correlation was indeed an important issue. However, as we show below, genetic distance remains statistically significant even after correcting the standard errors for spatial correlation.[42]

---

[41] This, of course, was not a concern in the simple regressions presented in Section IV.A. These results featured $t$-statistics in excess of 9 in the world sample, much larger than the $t$-statistics that we find using the bilateral approach with two-way clustering. This reinforces our confidence that our results are not driven by standard errors that are too low due to spatial correlation.

[42] There are in principle several other ways to address the problem of spatial correlation. One approach would be to do feasible GLS by explicitly estimating the elements of the covariance matrix and introducing the estimated covariance matrix as a weighing matrix in the second stage of the GLS procedure. This is computationally very demanding, as the dimensionality of the matrix is large—in our application we have over 9,316 country pairs with available data on the variables of interest, and up to 137 covariance terms to estimate (for the same reason, it is difficult to implement tests of spatial correlation in our context). Another approach, which we pursued in a previous version of this paper, is to include in our regressions common country fixed effects, meant to soak up the spatial correlation. For this we relied on well-known results cited in Case (1991), showing that fixed effects soak up spatial correlation, though in a context quite different from ours. Following this insight, we modeled $\varepsilon_{ij} = \sum_{k=1}^{N} \gamma_k \delta_k + \eta_{ij}$, where

### *IV.C. Unconditional Results*

Table II presents some summary statistics for our variables. Throughout, we use a baseline sample of 9,316 country pair observations obtained from 137 underlying countries. We consider various measures of genetic distance. As already mentioned, our baseline measure is weighted $F_{\mathrm{ST}}$ genetic distance. We also used the weighted Nei genetic distance.[43] These measures bear a high correlation with each other (.939), and in practice it matters little which one we use. On the other hand, the theoretically more appropriate measure of relative distance to the United States bears a correlation of only .634 with absolute $F_{\mathrm{ST}}$ genetic distance. Finally, we considered $F_{\mathrm{ST}}$ genetic distance with countries matched to populations as they were in 1500. The correlation between this variable and the current measure is .827.

Our measure of absolute $F_{\mathrm{ST}}$ genetic distance, $G^D$, bears a positive correlations of .197 with the absolute value of log income differences in 1995. Genetic distance relative to the frontier, $G^R$, bears a higher correlation with income differences, equal to .337, which is directly in line with our model's prediction (these correlations are higher in the European sample, respectively .328 and .409).

Table III presents univariate regressions of income differences on various measures of genetic distance for the world sample. As a measure of the magnitude of the coefficients, we report the standardized beta coefficient on genetic distance for each regression.[44] Column (1) shows that, when entered alone in the regression, one standard deviation in $F_{\mathrm{ST}}$ genetic distance between plurality groups accounts for 16.79% of a standard deviation of income differences. This effect rises in magnitude to 26.98% when we consider genetic distance (also between plurality ethnic groups of each country in a pair) relative to the frontier (column (2)). This

---

$\delta_k = 1$ if $k = i$ or $k = j$, $\delta_k = 0$ otherwise, and $\eta_{ij}$ is a well-behaved disturbance term. We treated $\delta_k$ as fixed effects; that is, we introduced into the regression a set of $N$ dummy variables $\delta_k$, each taking a value of one $N - 1$ times—$\delta_j$ takes a value of 1 whenever country $j$ appears in a pair. This did not affect the qualitative and quantitative nature of our results compared to the solution we pursue here—but the standard errors were smaller than the ones we report in this version. Finally, we also included separate fixed effects for each country in a pair, with no significant changes in the estimated effects of genetic distance.

[43] In past work we also used both $F_{\mathrm{ST}}$ and Nei genetic distance based on plurality groups, with results very similar to those reported here.

[44] The standardized beta is defined as the effect of a one-standard-deviation change in the regressor, expressed as a percentage of one standard deviation of the dependent variable.

TABLE II

SUMMARY STATISTICS FOR THE MAIN VARIABLES (WORLD DATA SET)

Panel A. Simple correlations among genetic and economic distance measures

| | $F_{ST}$ gen. dist. weighted | $F_{ST}$ gen. dist., weighted, relative to United States | $F_{ST}$ gen. dist., 1500 match | Nei gen. dist., weighted | Abs. log income diff. 1995 | Abs. log income difference, 1870$^{a}$ |
|---|---|---|---|---|---|---|
| $F_{ST}$ genetic distance, weighted, relative to United States | .634 | 1 | | | | |
| $F_{ST}$ genetic distance, 1500 match | .827 | .544 | 1 | | | |
| Nei genetic distance, weighted | .939 | .742 | .782 | 1 | | |
| Abs. log income difference, 1995 | .197 | .337 | .203 | .231 | 1 | |
| Abs. log income difference, 1870$^{a}$ | .007 | .200 | .058 | .072 | .596 | 1 |
| Abs. log income difference, 1500$^{b}$ | −.088 | −.030 | .198 | −.068 | −.051 | .060 |

Panel B. Means and standard deviations

| Variable | # of obs. | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|---|
| $F_{ST}$ genetic distance weighted | 9,316 | 0.111 | 0.071 | 0.000 | 0.344 |
| $F_{ST}$ genetic distance, weighted, relative to United States | 9,316 | 0.062 | 0.048 | 0.000 | 0.213 |
| $F_{ST}$ genetic distance, 1500 match | 9,316 | 0.127 | 0.082 | 0.000 | 0.356 |
| Nei genetic distance, weighted | 9,316 | 0.018 | 0.013 | 0.000 | 0.059 |
| Abs. log income difference, 1995 | 9,316 | 1.290 | 0.912 | 0.000 | 4.133 |
| Abs. log income difference, 1870$^{b}$ | 1,326 | 0.658 | 0.488 | 0.000 | 2.110 |
| Abs. log income difference, 1500$^{a}$ | 325 | 0.327 | 0.237 | 0.000 | 1.012 |

*Note*. Number of observations: 9,316, except $^{a}$1,326 and $^{b}$325.

TABLE III
UNIVARIATE REGRESSIONS (TWO-WAY CLUSTERED STANDARD ERRORS)

| | (1) $F_{ST}$ gen. dist. | (2) $F_{ST}$ gen. dist. relative to United States | (3) Weighted $F_{ST}$ gen. dist. | (4) Weighted $F_{ST}$ gen. dist., relative to United States | (5) Weighted Nei gen. dist. | (6) Weighted regression |
|---|---|---|---|---|---|---|
| $F_{ST}$ genetic distance | 1.853 | | | | | 2.214 |
| | (0.508)** | | | | | (0.533)** |
| $F_{ST}$ genetic distance relative to the United States | | 3.541<br>(0.654)** | | | | |
| Weighted $F_{ST}$ genetic distance | | | 2.516<br>(0.630)** | | | |
| $F_{ST}$ gen. dist. relative to the United States, weighted | | | | 6.357<br>(0.996)** | | |
| Weighted Nei genetic distance | | | | | 16.868<br>(3.792)** | |
| Constant | 1.079 | 0.977 | 1.010 | 0.893 | 0.986 | 1.044 |
| | (0.051)** | (0.049)** | (0.059)** | (0.052)** | (0.057)** | (0.050)** |
| Standardized beta (%) | 16.79 | 26.98 | 19.71 | 33.65 | 27.01 | 20.07 |
| $R^2$ | .03 | .07 | .04 | .11 | .05 | .03 |

*Notes.* Dependent variable: absolute value of log income differences, 1995. Two-way clustered standard errors in parentheses; 9,316 observations from 137 countries.
*Significant at 10%.
**Significant at 5%.

means that the effect of genetic distance relative to the frontier is larger than the effect of bilateral genetic distance, exactly as predicted by Implication 2 of our model. Turning to the weighted measure, similar effects are found, with slightly larger magnitudes (columns (3) and (4)), respectively 19.71% and 33.65%. The larger magnitudes are consistent with the idea that weighted measures are better proxies for the expected genetic distance between countries. The effect is also larger when Nei genetic distance is used instead of $F_{\mathrm{ST}}$ (column (5)).[45]

We next make use of data from Cavalli-Sforza, Menozzi, and Piazza (1994) on the standard deviation of the genetic distance estimates. Because these data are based on allele frequencies collected from samples of different sizes, they are estimated more or less precisely depending on population pairs. We have data on the standard errors of each estimate of genetic distance, obtained from bootstrap analysis. In column (6), we linearly downweigh observations with higher standard errors on genetic distance. As expected, the magnitude of the resulting weighted least squares effect of $F_{\mathrm{ST}}$ genetic distance is larger than under simple OLS, consistent with the idea that measurement error is greater for pairs with high standard errors on genetic distance. Similar results are obtained using alternative measures of genetic distance.

While providing suggestive evidence in favor of Implications 1 and 2 of our theoretical model, these unconditional results may confound the effect of barriers linked to vertical characteristics with geographic barriers. In the next section, we control for a large number of measures of geographic distance. In everything that follows we will focus on the weighted relative genetic distance to the frontier as the baseline measure of genetic distance, that is, the measure used in column (4) of Table III, because it is theoretically more appropriate.

### *IV.D. Controlling for Geographic Factors*

Genetic distance and geographic isolation are likely to be highly correlated. The more isolated two groups become, the more they will drift apart genetically, since genetic admixture is made difficult by geographic barriers. It is therefore important to control adequately for geographic isolation: failing to do so would

---

[45] We found no evidence of nonlinear effects of genetic distance. A quadratic term in genetic distance bore an insignificant coefficient and the total effect of genetic distance in this quadratic specification, evaluated at the mean of genetic distance, was commensurate with the linear effect reported here.

ascribe to genetic distance an effect that should be attributed to geographic distance. In this section, we control for a vast array of measures of geographic isolation.

*Distance Metrics.* Our first set of measures of geographic isolation between countries consists of various measures of distance. We consider a measure of the great circle (geodesic) distance between the major cities of the countries in our sample, from a data set compiled by researchers at Centre d'Etudes Prospectives et d'Informations Internationales (CEPII).[46] We also include latitudinal distance—that is, simply the absolute value of the difference in latitude between the two countries $i$ and $j$ in each pair: $G_{ij}^{\mathrm{LA}} = |\mathrm{latitude}_i - \mathrm{latitude}_j|$. Latitude could be associated with climatic factors that affect income levels directly, as in Gallup, Mellinger, and Sachs (1998) and Sachs (2001). Latitude differences would also act as barriers to technological diffusion: Diamond (1997) suggests that barriers to the transmission of technology are greater along the North–South axis than along the East–West axis, because regions at the same latitude share similar climate, availability of domesticable species, soil conditions, etc. We should therefore expect countries at similar latitudes to also display similar levels of income. Third, we use a measure of longitudinal distance, $G_{ij}^{\mathrm{LO}} = |\mathrm{longitude}_i - \mathrm{longitude}_j|$, to capture possible geographic isolation along this alternative axis.

Table IV, column (2), includes these three measures jointly with $F_{\mathrm{ST}}$ genetic distance relative to the frontier. The effect of relative genetic distance barely changes at all compared to the baseline univariate regression replicated in column (1). We find evidence that latitudinal distance matters—the standardized beta on this variable is 11.97%, consistent with Jared Diamond's hypothesis.

*Microgeographic Factors.* In addition to these straightforward distance measures, we controlled for other measures of isolation between countries. In the context of gravity regressions for Europe, Giuliano, Spilimbergo, and Tonon (2006) argued that genetic distance was likely correlated with features of the terrain. These "microgeographic" features may not be well captured by

---

[46] The data are available at http://www.cepii.fr/anglaisgraph/bdd/distances.htm. The correlation between geodesic distance and weighted $F_{\mathrm{ST}}$ genetic distance is .349. This correlation rises to .486 if genetic distance is measured based on populations as they were in 1500, because the colonization era acted to weaken the link between genetic distance and geographic distance by shuffling populations across the globe. The correlations are lower in magnitude when considering distances relative to the frontier, although their relative magnitude is preserved.

TABLE IV
CONTROLLING FOR GEOGRAPHIC DISTANCE (TWO-WAY CLUSTERED STANDARD ERRORS)

| | (1) Baseline | (2) Distance metrics | (3) Add micro-geography controls | (4) Add transport costs | (5) Continent dummies | (6) Climatic difference control | (7) Tropical difference control |
|---|---|---|---|---|---|---|---|
| $F_{\mathrm{ST}}$ gen. dist. relative to the United States, weighted | 6.357 | 6.387 | 6.273 | 6.312 | 4.134 | 6.067 | 6.368 |
| | (0.996)$^{**}$ | (0.994)$^{**}$ | (0.989)$^{**}$ | (0.988)$^{**}$ | (1.046)$^{**}$ | (0.960)$^{**}$ | (1.003)$^{**}$ |
| Absolute difference in latitudes | | 0.523 | 0.494 | 0.494 | −0.228 | 0.254 | 0.497 |
| | | (0.241)$^{**}$ | (0.238)$^{**}$ | (0.237)$^{**}$ | (0.217) | (0.221) | (0.237)$^{**}$ |
| Absolute difference in longitudes | | 0.387 | 0.391 | 0.376 | 0.084 | 0.257 | 0.380 |
| | | (0.235)$^{*}$ | (0.226)$^{*}$ | (0.224)$^{*}$ | (0.162) | (0.224) | (0.224)$^{*}$ |
| Geodesic distance (1,000s of km) | | −0.050 | −0.057 | −0.081 | −0.008 | −0.062 | −0.081 |
| | | (0.028)$^{*}$ | (0.026)$^{**}$ | (0.039)$^{**}$ | (0.036) | (0.038)$^{*}$ | (0.039)$^{**}$ |
| =1 for contiguity | | | −0.456 | −0.462 | −0.284 | −0.328 | −0.464 |
| | | | (0.064)$^{**}$ | (0.064)$^{**}$ | (0.060)$^{**}$ | (0.061)$^{**}$ | (0.064)$^{**}$ |
| =1 if either country is an island | | | 0.178 | 0.180 | 0.119 | 0.162 | 0.181 |
| | | | (0.094)$^{*}$ | (0.094)$^{*}$ | (0.090) | (0.102) | (0.092)$^{*}$ |
| =1 if either country is landlocked | | | 0.071 | 0.078 | 0.110 | 0.084 | 0.075 |
| | | | (0.076) | (0.076) | (0.071) | (0.076) | (0.075) |
| =1 if pair shares at least one sea or ocean | | | −0.029 | −0.024 | 0.030 | 0.044 | −0.024 |
| | | | (0.062) | (0.062) | (0.050) | (0.059) | (0.062) |
| Freight rate (surface transport) | | | | 1.282 | −0.197 | 1.160 | 1.286 |
| | | | | (1.568) | (1.517) | (1.490) | (1.583) |
| Climatic difference of land areas, by 12 KG zones | | | | | | 0.032 | |
| | | | | | | (0.007)$^{**}$ | |
| Difference in % land area in KG tropical climates | | | | | | | −0.033 |
| | | | | | | | (0.083) |

TABLE IV
(CONTINUED)

| | (1) Baseline | (2) Distance metrics | (3) Add micro-geography controls | (4) Add transport costs | (5) Continent dummies | (6) Climatic difference control | (7) Tropical difference control |
|---|---|---|---|---|---|---|---|
| Constant | 0.893 | 0.866 | 0.889 | 0.675 | 1.919 | 0.284 | 0.681 |
| | (0.052)** | (0.066)** | (0.078)** | (0.263)** | (0.407)** | (0.263) | (0.264)** |
| Standardized beta (%) | 33.65 | 33.81 | 33.20 | 33.41 | 21.88 | 32.12 | 33.70 |
| $R^2$ | .11 | .12 | .13 | .13 | .22 | .15 | .13 |

*Notes.* Dependent variable: absolute value of log income differences, 1995.
Two-way clustered standard errors in parentheses. 9,316 observations from 137 countries in all columns. Column (5) includes two sets of continent dummies (estimates not reported): a set of dummies each equal to 1 if both countries in a pair are on the same given continent; and a set of dummies each equal to one if exactly one country belongs to a given continent, and the other not. Continents are defined as Europe, Africa, Latin America, North America, Asia, and Oceania.
*Significant at 10%.
**Significant at 5%.

simple metrics of distance. We included dummy variables taking a value of 1 if countries in a pair were contiguous, if they had access to a common sea or ocean, or if either country in a pair was an island or was landlocked.[47] These measures are meant to capture ease of communication and travel between countries, which may be associated both with barriers to technological diffusion and to population isolation (and thus genetic distance). Column (3) of Table IV shows that these variables have the expected signs, but their inclusion does not affect the coefficient on genetic distance. To summarize, although the additional controls have explanatory power for income differences, we found no evidence that the inclusion of microgeographic factors modifies the effect of genetic distance.

*Transportation Costs.* A good summary measure of geographic isolation is transportation costs. Giuliano, Spilimbergo, and Tonon (2006) use a new measure of transportation costs based directly on freight rates for surface transport (sea or land) between European countries. We have obtained the same data as those they used, for the world sample.[48] Column (4) of Table IV adds this measure of freight costs to our specification. We find that freight costs bear a positive relationship to income differences, as expected. However, this effect is not significant statistically and does not affect the signs or magnitudes of the other included variables, particularly genetic distance. We find no evidence that genetic distance captures the effect of geographic isolation or transportation costs in our application.[49]

*Continent Effects.* The largest genetic distances observed in our worldwide data set occur between populations that live on

---

[47] The common sea variable is the same as that used in Giuliano, Spilimbergo, and Tonon (2006). These authors also used a measure of the average elevation of the countries that lie between any two countries in a pair, as a measure of how hard it is to travel from one to the other. Although we calculate and use this variable for the Europe sample, where it is relatively straightforward to do so, there are simply too many possible paths between any two countries in the world for this to be practical in the broader sample of world countries.

[48] The data are available from http://www.importexportwizard.com/. The measure we used referred to 1,000 kg of unspecified freight transported over sea or land, with no special handling. This is the same definition used in Giuliano, Spilimbergo, and Tonon (2006). The data on 10,825 pairs of countries were downloaded from the website using a Perl script.

[49] In the previous version of this paper, we also used the approach in Limao and Venables (2001) and Hummels and Lugovskyy (2006) to measure trade costs indirectly through the matched partner technique, using the ratio of CIF to FOB exports. The measure of indirect trade costs is $\text{ITC}_{ij} = (\text{CIF}_{ij}/\text{FOB}_{ij}) - 1$. Results with this alternative measure of trade costs featured a much smaller sample, but were similar to those reported here.

TABLE V
ENDOGENEITY OF GENETIC DISTANCE AND DIAMOND GAP (TWO-WAY CLUSTERED STANDARD ERRORS)

| | (1) 2SLS with 1500 genetic distance | (2) Without New World | (3) Diamond gap, w/o New World | (4) Income 1500, Diamond gap |
|---|---|---|---|---|
| $F_{ST}$ genetic distance relative to the United States, weighted | 9.400 | 4.428 | 2.815 | |
| | (1.665)$^{**}$ | (1.252)$^{**}$ | (1.347)$^{**}$ | |
| $F_{ST}$ genetic distance relative to the English, 1500 match | | | | 1.737 |
| | | | | (0.427)$^{**}$ |
| Absolute difference in latitudes | 0.402 | 0.901 | 1.078 | 0.152 |
| | (0.293) | (0.420)$^{**}$ | (0.471)$^{**}$ | (0.138) |
| Absolute difference in longitudes | 0.601 | 0.349 | 0.781 | −0.007 |
| | (0.246)$^{**}$ | (0.258) | (0.333)$^{**}$ | (0.070) |
| Geodesic distance (1,000s of km) | −0.114 | −0.087 | −0.155 | −0.016 |
| | (0.039)$^{**}$ | (0.051)$^{*}$ | (0.055)$^{**}$ | (0.022) |
| =1 for contiguity | −0.381 | −0.471 | −0.461 | −0.048 |
| | (0.063)$^{**}$ | (0.069)$^{**}$ | (0.067)$^{**}$ | (0.040) |
| =1 if either country is an island | 0.209 | 0.134 | 0.176 | 0.004 |
| | (0.094)$^{**}$ | (0.113) | (0.115) | (0.053) |
| =1 if either country is landlocked | 0.052 | 0.016 | 0.022 | −0.059 |
| | (0.076) | (0.081) | (0.076) | (0.034)$^{*}$ |
| =1 if pair shares at least one sea or ocean | −0.043 | −0.060 | −0.071 | −0.068 |
| | (0.077) | (0.087) | (0.085) | (0.047) |
| Freight rate (surface transport) | 1.700 | 1.627 | 1.508 | −0.263 |
| | (1.341) | (1.809) | (1.781) | (0.847) |
| Diamond gap | | | 0.472 | 0.164 |
| | | | (0.137)$^{**}$ | (0.059)$^{**}$ |
| Constant | 0.488 | 0.701 | 0.760 | 0.338 |
| | (0.241)$^{**}$ | (0.312)$^{**}$ | (0.309)$^{**}$ | (0.144)$^{**}$ |
| # of observations | 9,316 | 6,105 | 6,105 | 325 |
| # of countries | 137 | 111 | 111 | 26 |
| Standardized beta (%) | 49.75 | 23.56 | 14.98 | 37.96 |
| $R^2$ | .10 | .11 | .13 | .22 |

*Notes*. Dependent variable: absolute value of log income differences, 1995 (columns (1)–(3)) or 1500 (column (4)).
Two-way clustered standard errors in parentheses. The Diamond gap is a dummy variable that takes on a value of 1 if one and only one of the countries in each pair is located on the Eurasian landmass, and 0 otherwise.
$^{*}$Significant at 10%.
$^{**}$Significant at 5%.

different continents. One concern is that genetic distance may simply be picking up the effect of cross-continental barriers to the diffusion of development, that is, continent effects. To test explicitly for this possibility, we added to our baseline specification two sets of continent dummies. We included one set of six dummies (one for each continent) taking on a value of one if the two countries in a pair were on the same continent. We also included a set of six dummies each equal to one if one country belonged to a given continent, and the other did not. The results are in column (5) of Table IV. The inclusion of continent dummies reduces by about one-third the magnitude of the genetic distance effect, but the latter remains statistically significant. Its magnitude is still large, with a standardized beta of 21.88%.

Figure III shows that many of the countries most genetically and economically distant from the United States are in sub-Saharan Africa. To examine whether this drives our results, we excluded any pair involving a sub-Saharan African country from our sample. In the resulting regression (available upon request), the standardized beta on genetic distance was 32.11% and was highly significant statistically. We therefore find no evidence that our results are driven by the inclusion of Africa in our sample. We will provide further evidence on the within-continent effects of genetic distance using the European data set in Section IV.H.

*Climatic Similarity.* Next, we constructed measures of climatic similarity based on 12 Koeppen–Geiger climate zones.[50] One measure is the average absolute value difference, between two countries, in the percentage of land area in each of the 12 climate zones. Countries have identical climates, under this measure, if they have identical shares of their land areas in the same climates. As a simpler alternative, we used the absolute difference in the percentage of land areas in tropical climates. As with latitude, climate may have direct effects on productivity, or barrier effects on technological diffusion: countries located in different climates may have experienced difficulties in adopting each other's modes of production, particularly in the agrarian era. Columns (6) and (7) of Table IV report the results. As expected, climatic differences are associated with greater income differences, even controlling for latitude differences. However, the inclusion of these variables hardly affects the coefficient on genetic distance.

---

[50] The 12 Koeppen–Geiger climate zones are tropical rain forest climate (Af), monsoon variety of Af (Am), tropical savannah climate (Aw), steppe climate (BS), desert climate (BW), mild humid climate with no dry season (Cf), mild humid climate with a dry summer (Cs), mild humid climate with a dry winter (Cw), snowy-forest climate with a dry winter (Dw), snowy-forest climate with a moist winter (Ds), tundra/polar ice climate (E), and highland climate (H). The data, compiled by Gallup, Mellinger, and Sachs, are available at http://www.ciesin.columbia.edu/eidata/.

We conclude that our results are robust to controlling for a wide variety of measures of geographic distance, microgeographic measures of isolation, continent effects, climatic differences, and transportation costs, whether entered in absolute terms or relative to the frontier.[51]

### IV.E. Endogeneity and the Diamond Gap

*Possible Endogeneity of Current Genetic Distance.* We attempted to control for the possible endogeneity of genetic distance with respect to income differences. Although differences in (neutral) allele frequencies between the populations of two countries do not result causally from income differences, migration could lead to a pattern of genetic distances today that is closely linked to current income differences. The issue arises from the pattern of colonization of the New World starting after 1500. Europeans tended to settle in larger numbers in the temperate climates of North America and Oceania. If geographic factors bear a direct effect on income levels, and were not properly accounted for in our regressions through included control variables, then genetic distance today could be positively related to income distance not because genetic distance precluded the diffusion of development, but because similar populations settled in regions prone to generating similar incomes.

To assess this possibility, we use our data on $F_{\mathrm{ST}}$ genetic distance as of 1500, relative to the English population, as an instrument for current genetic distance. This variable reflects genetic distance between populations as they were before the great migrations of the modern era, that is, as determined since the Neolithic era, and yet is highly correlated (.611) with current genetic distance relative to the United States, so it fulfills the conditions of a valid instrument. An added benefit of the IV approach is that it allows us to address in part possible measurement error in current genetic distance—the matching of populations to countries is much more straightforward for 1500, as explained above. In column (1) of Table V, the magnitude of the genetic distance effect is raised by one-third, with a standardized beta reaching 49.75%. As is usual in this type of application, the larger estimated effect

---

[51] In all these regressions, geographic distance was entered in absolute terms, not relative to the frontier. In results available upon request, all the measures of distance and transport costs used in Table IV were entered relative to the United States instead. If anything, the coefficient on genetic distance became larger.

may come from a lower incidence of measurement error under two-stage least squares.

To further assess whether our results are driven by endogeneity of the sort discussed above, column (2) of Table V excludes from the sample any pairs involving one or more countries from the New World (defined as countries in North America, Latin America, the Caribbean, and Oceania), where the endogeneity problem is likely to be most acute. The effect of genetic distance falls by about one-third, but remains both statistically and economically significant. The difference in latitudes becomes much larger, an observation to which we shall return below.

*The Diamond Gap.* Jared Diamond's (1997) influential book stressed that differences in latitude played an important role as barriers to the transfer of technological innovations in early human history, and later in the preindustrial era, an effect that could have persisted to this day. Our estimates of the effect of latitudinal distance provided evidence that this effect was still at play: in our regressions we found evidence that differences in latitudes help explain income differences across countries, and this effect was much larger when excluding the New World from our sample. However, Diamond took his argument one step further, and argued that Eurasia enjoyed major advantages in the development of agriculture and animal domestication because (a) it had the largest number of potentially domesticable plants and animals and (b) it had a predominantly East–West axis that allowed an easier and faster diffusion of domesticated species. By contrast, differences in latitudes in the Americas and Africa created major environmental barriers to the diffusion of species and innovations. More generally, Eurasia might have enjoyed additional benefits in the production and transfer of technological and institutional innovations because of its large size.[52] It is important to properly control for Diamond's geography story, as it is either a substitute or a complement to ours.

To test and control for a Eurasian effect, we constructed a dummy variable that takes on a value of 1 if one and only one of the countries in each pair is in Eurasia, and 0 otherwise (the "Diamond gap").[53] In order to test Diamond's hypothesis, we added

---

[52] This point is stressed in Kremer (1993). See also Masters and McMillan (2001).

[53] For further tests providing statistical support for Diamond's observations, see Olsson and Hibbs (2005).

the Diamond gap to regressions explaining income differences in 1995 (column (3) of Table V) and, using Maddison's historical income data, in 1500 (column (4)). For the former regression, we restricted our sample to the Old World.[54] As expected, in the regression for 1995 income differences, the Diamond gap enters with a positive and significant coefficient, and its inclusion reduces (but does not come close to eliminating) the effect of genetic distance. In column (4), using 1500 income differences as a dependent variable, the Diamond gap is also significant and large in magnitude, despite the paucity of observations. Again, the effect of genetic distance relative to the English population using the 1500 match remains large in magnitude, with a standardized beta of 37.96%. This provides suggestive quantitative evidence in favor of Diamond's observation that the diffusion of development was faster in Eurasia. We also conclude that genetic distance between populations plays an important role in explaining income differences even when controlling for the environmental advantages and disadvantages associated with Eurasia. Diamond's hypothesis on the long-term diffusion of development is complementary to ours.

### *IV.F. Controlling for Common History, Linguistic Distance, and Religious Distance*

In this section we control for additional possible determinants of income differences.[55] We consider common history variables (for example, whether countries shared a common colonial past), and variables capturing distances in specific cultural characteristics, such as language and religion. In principle, countries that are close in terms of genetic distance may also be close in terms of common history, language, and religion, so we check whether the effect of genetic distance on income differences is robust to controlling for these specific channels of historical and cultural similarities. In particular, in this section we discuss two related questions: (a) How correlated is genetic distance with measures

---

[54] It is appropriate to exclude the New World from the sample when using 1995 incomes because Diamond's theory is about the geographic advantages that allowed Eurasians to settle and dominate the New World. If we were to include the New World in a regression explaining income differences today, we would include the higher income per capita of nonaboriginal populations who are there because of guns, germs, and steel, that is, thanks to their ancestors' Eurasian advantage.

[55] Throughout this section we will use the specification in column (4) of Table IV as the baseline—that is, we include a large array of geographic isolation controls.

of linguistic and religious distances? (b) How is the effect of genetic distance on income differences affected by the introduction of these additional variables?

As we detail below, overall we find positive but relatively modest correlations between measures of linguistic or religious distances and genetic distance across populations. We also find that controlling for such distances reduces the effect of genetic distance on income, but only to a modest extent: the effect of genetic distance remains large and significant. These results confirm the robustness of the relationship between genetic distance and income differences, but may also be viewed as somewhat surprising. Because both language and religion tend to be transmitted across generations, one could have expected that measures of linguistic, religious, and genetic distance would capture similar patterns of genealogical relatedness, and that their joint inclusion would reduce the effect of genetic distance. However, as we discuss below, there are several reasons, related to the definition and measurement of these variables, that make them empirically distinct. This may shed light on why genetic distance plays such a predominant role in explaining income differences, and why its effect appears to be largely unaffected by the inclusion of linguistic and religious distances.

*Measures of Linguistic and Religious Distance.* We construct two measures of linguistic distance and one measure of religious distance. Our first approach to linguistic distance follows Fearon (2003). Fearon used data from Ethnologue to create linguistic trees, classifying languages into common families and displaying graphically the degree of relatedness of world languages. The linguistic tree in this data set contains up to 15 nested classifications. If two languages share many common nodes in the tree, these languages are more likely to trace their roots to a more recent common ancestor language. The number of common nodes in the linguistic tree, then, is a measure of linguistic similarity. For instance, according to this measure, French and Italian share four common nodes—both belong to the Indo-European/Italic/Romance/Italo-Western linguistic groupings. Using data on the linguistic composition of countries (also from Fearon [2003]), and matching languages to countries, we can construct indices of linguistic distance between countries. We did so, as for genetic distance, in two ways: first, we computed a measure of the number of common nodes shared by languages spoken by plurality groups within each country in a pair. Second, we computed a weighted measure of linguistic similarity, representing the expected number of common linguistic nodes between two randomly chosen individuals, one from each country in a pair (the formula is analogous to that of equation (9)).[56] Following Fearon (2003), we transformed each of these series so that they are increasing in linguistic distance (LD) and bounded by 0 and 1:

$$\text{LD} = \sqrt{\frac{(15 - \#\ \text{Common Nodes})}{15}}. \tag{12}$$

Our second measure of linguistic distance is based on work in the field of lexicostatistics (a branch of linguistics). We use data from Dyen, Kruskal, and Black (1992). They assembled data on 200 common "meanings" from all Indo-European languages. For each language, they compiled lists of words expressing these meanings. When words from two languages expressing a given meaning originated from a common source, these words were considered to be cognate. For instance, the words "table" in French and "tavola" in Italian are cognate because both stem from the word "tabula" in Latin. Aggregating over the 200 meanings, our measure of linguistic distance is the percentage of noncognate words, and as before we can compute an expected (weighted) measure and a measure based on the percentage of cognate words between the languages spoken by the plurality linguistic groups in each country in a pair. Again, the greater the percentage of cognate words, the more recently the languages shared a common ancestor language. In contrast to the linguistic trees data, this measure has the advantage of being a continuous measure of similarity. Its main drawback is that it is only available for Indo-European languages, so the geographic coverage is reduced to 62 countries (when considering the % cognate between plurality languages).[57]

To measure religious distance, we followed an approach similar to that used for linguistic distance. We relied on a nomenclature of world religions obtained from Mecham, Fearon, and Laitin

[56] Using the measure based on the plurality language or the weighted measure did not make any difference for the results. In keeping with what we did for genetic distance, we focus on weighted measures in our empirical work.

[57] In addition, when using the weighted measure of lexicostatistical distance, we lose further countries with sizable minorities of non-Indo-European speaking populations, such as India. For this variable, only 43 countries remain. These are mostly countries in Europe and the Americas.

(2006).[58] This nomenclature was broken down into religious families, first distinguishing between monotheistic religions of Middle-Eastern origin, Asian religions, and “others,” then subdividing each group into finer groups (such as Christians, Muslims, and Jews), and so on. The number of common classifications (up to five in this data set) is a measure of religious proximity. We matched religions to countries using Mecham, Fearon, and Laitin’s (2006) data on the prevalence of religions by country and transformed the data in a manner similar to that in equation (12).

Pairwise correlations between measures of genetic, linguistic, and religious distances are displayed in Table VI. These correlations are generally positive, as expected, but they are not very large in magnitude. For instance, the correlation between $F_{\mathrm{ST}}$ genetic distance and weighted linguistic distance is .227. The two alternative measures of linguistic distance bear a correlation of .745. Religious distance bears a correlation of .438 with linguistic distance and .171 with genetic distance.

As mentioned above, the relatively small magnitude of the correlation between linguistic distance and genetic distance may be viewed as somewhat surprising. Anthropologists and population geneticists, including Cavalli-Sforza, Menozzi, and Piazza (1994, pp. 98–105), have pointed out that there is usually little genetic admixture between populations that speak different languages and that linguistic trees often mirror genetic trees for aboriginal populations. However, these scholars have also stressed forces that may occasionally lead to dramatic divergence between genetic distance and linguistic distance. Historically, conquests and migrations have often been associated with language replacement as well as gene replacement, therefore creating a wedge between linguistic trees and genetic trees. An example is the very different relation between the Hungarian population and neighboring populations in terms of genetic distance vs. linguistic distance. The Hungarian language (Magyar) is a Uralic language, unrelated to most other languages in Europe, and was introduced into current Hungary by Magyar-speaking tribes around 900 A.D. However, modern Hungarians are genetically very close to their European neighbors, suggesting that the Magyars mixed with preexisting Slavic speaking populations when they arrived in modern-day Hungary and/or that a large

---

[58] An alternative classification obtained from the World Christian Database, with only three nested classifications, did not lead to appreciably different results.

TABLE VI
SUMMARY STATISTICS FOR GENETIC DISTANCE AND MEASURES OF LINGUISTIC AND RELIGIOUS DISTANCE

Panel A: Correlations between genetic distance and various measures of linguistic and religious distance

| | Weighted $F_{ST}$ genetic distance | Weighted linguistic distance | Weighted religious distance | Weighted $F_{ST}$ gen. dist. relative to United States | Weighted ling. dist. relative to United States | Weighted religious distance relative to United States | 1% cognate, relative to United States, plurality[a] |
|---|---|---|---|---|---|---|---|
| Weighted linguistic distance | .227 | 1 | | | | | |
| Weighted religious distance | .171 | .438 | 1 | | | | |
| Weighted $F_{ST}$ genetic distance relative to United States | .634 | .210 | .126 | 1 | | | |
| Weighted linguistic distance relative to United States | −.020 | .058 | .026 | .062 | 1 | | |
| Weighted religious distance relative to United States | .052 | .143 | .343 | .061 | .459 | 1 | |
| 1 − % cognate, plurality [a] | −.105 | .745 | .618 | .011 | .128 | .292 | 1 |
| 1 − % cognate, relative to United States, plurality [a] | .079 | .274 | .176 | .169 | .702 | .163 | .372 |

Panel B: Summary statistics for genetic distance and various measures of linguistic and religious distance

| Variable | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|
| Weighted $F_{ST}$ genetic distance | 0.111 | 0.071 | 0.000 | 0.344 |
| Weighted linguistic distance | 0.968 | 0.106 | 0.000 | 1.000 |
| Weighted religious distance | 0.841 | 0.151 | 0.089 | 1.000 |
| Weighted $F_{ST}$ genetic distance relative to United States | 0.062 | 0.048 | 0.000 | 0.213 |
| Weighted linguistic distance relative to United States | 0.088 | 0.169 | 0.000 | 1.000 |
| Weighted religious distance relative to United States | 0.149 | 0.134 | 0.000 | 0.999 |

*Note*. Number of observations = 9,316, except [a]number of obs. = 1,830.

number of non-Magyar-speaking individuals moved to Hungary and adopted Magyar as their language in following centuries. For example, the genetic distance between Hungarians and Poles is only 0.0025 (lower than the genetic distance between Swedes and Danes, who speak closely related languages). By contrast, Poles and Yugoslavs have a genetic distance equal to 0.0137, even though both populations speak Slavic languages (interestingly, the genetic distance between Hungarians and Yugoslavs is about the same, 0.0136). This example illustrates how populations who speak very different languages may be genetically close, whereas populations that speak more similar languages may be quite far apart genetically. Even more dramatic examples can be found among countries that were colonized by European powers and adopted the colonizers' language (English, French, Portuguese, or Spanish), while maintaining very distinct populations in terms of common ancestry. By the same token, conquests and conversions have led to the adoption of similar religions by genetically distinct populations, as well as of different religions by genetically close populations. In that respect, the relatively low correlation between genetic distance and linguistic or religious distance partly reflects the fact that these variables measure conceptually distinct relations between populations.

In addition, technical reasons related to the construction of the measures of linguistic and religious distances contribute to the low correlation between these measures and genetic distances. Genetic distance is a continuous measure, reflecting an objective molecular clock, and maps linearly into the time elapsed since different populations shared common ancestors. In contrast, the number of nodes is a discrete and imperfect measure of linguistic or religious distance, based on sometimes arbitrary classifications of languages into groups by linguists, and counting the discrete number of common nodes may not capture such distances appropriately.[59] For example, Fearon (2003) argues that the move from 0 to 1 common node is more important than the move from

---

[59] Populations may share few common nodes but linguistic splits may have occurred recently, in which case one is overestimating distance, or they may share lots of common nodes but the last split may have occurred a long time ago, in which case one is underestimating distance. The idea that these measures of linguistic and religious distance may include a lot of measurement error is confirmed by Fearon (2006) in a short unpublished comment on our work suggesting, for European countries, that genetic distance is robust to the inclusion of a variety of measures of linguistic distance in a regression seeking to explain income differences.

3 to 4 common nodes. The lexicostatistical measure may partly address this problem, because it is more continuous, but at the cost of losing all of the non-Indo-European-speaking countries.

Finally, a third reason for low correlations is that we consider genetic and cultural distances relative to the world technological frontier (the United States) in our regressions, in keeping with our theory. As shown in Table VI, the correlations between measures relative to the United States are lower than the correlations between simple distances. For instance, while the correlation between weighted linguistic and genetic distances is .227, once these variables are considered relative to the United States the correlation falls to .062.

*Regression Results.* Table VII presents results obtained when including measures of linguistic and religious distance as well as common history variables. We first control for variables representing a pair's common historical experience, obtained from CEPII. These are dummy variables for pairs that were ever part of the same country (for example, Austria and Hungary), were ever in a colonial relationship, have shared a common colonizer since 1945, and are currently in a colonial relationship (such as France and French Polynesia). These variables all bear the expected signs and have statistically significant coefficients (Table VII, column (2)). For instance, having had a common colonizer and having been part of the same country are associated with smaller income differences. The inclusion of these variables in the regression does not affect the magnitude of the genetic distance effect.

Turning to linguistic and religious distance, in Table VII, columns (3) and (4), both linguistic distance and religious distance enter with the expected positive signs and are statistically significant at the 5% level when entered individually. Their standardized betas are, respectively, 15.10% and 20.17%, so these variables can help account for a sizable fraction of the variation in income differences. When the two variables are entered jointly, only religious distance remains significant (column (5)). More importantly from our perspective, the inclusion of these variables, either alone or together, slightly reduces the effect of genetic distance on income differences. Comparing column (5) with column (2), the reduction in the coefficient on genetic distance (and in the standardized beta), is 11.5%. Column (7) shows the results obtained when including the measure of linguistic distance based on the percentage

## TABLE VII

### CONTROLLING FOR COMMON HISTORY, LINGUISTIC DISTANCE, AND RELIGIOUS DISTANCE (TWO-WAY CLUSTERED STANDARD ERRORS)

|  | (1) Baseline | (2) Colonial history controls | (3) Linguistic distance, weighted | (4) Religious distance, weighted | (5) Religious + linguistic, weighted | (6) Baseline (smaller sample) | (7) % cognate, plurality |
|---|---|---|---|---|---|---|---|
| $F_{ST}$ gen. dist. relative | 6.312 | 6.283 | 5.827 | 5.702 | 5.557 | 6.853 | 5.995 |
| to the United States, weighted | (0.988)$^{**}$ | (0.988)$^{**}$ | (0.944)$^{**}$ | (0.950)$^{**}$ | (0.940)$^{**}$ | (2.116)$^{**}$ | (2.109)$^{**}$ |
| =1 if countries were or |  | −0.217 | −0.223 | −0.209 | −0.214 | −0.262 | −0.196 |
| are the same country |  | (0.088)$^{**}$ | (0.087)$^{**}$ | (0.087)$^{**}$ | (0.087)$^{**}$ | (0.084)$^{**}$ | (0.090)$^{**}$ |
| =1 for pairs ever in |  | 0.304 | 0.255 | 0.307 | 0.282 | 0.109 | 0.167 |
| colonial relationship |  | (0.131)$^{**}$ | (0.134)$^{*}$ | (0.101)$^{**}$ | (0.106)$^{**}$ | (0.112) | (0.119) |
| =1 for common colonizer |  | −0.226 | −0.214 | −0.135 | −0.142 | −0.046 | −0.038 |
| post-1945 |  | (0.066)$^{**}$ | (0.063)$^{**}$ | (0.060)$^{**}$ | (0.059)$^{**}$ | (0.111) | (0.082) |
| =1 for pairs currently in |  | −1.033 | −0.823 | −0.969 | −0.873 | −0.720 | −0.444 |
| colonial relationship |  | (0.193)$^{**}$ | (0.200)$^{**}$ | (0.167)$^{**}$ | (0.176)$^{**}$ | (0.162)$^{**}$ | (0.179)$^{**}$ |
| Linguistic distance index, |  |  | 0.815 |  | 0.409 |  |  |
| relative to United States, weighted |  |  | (0.204)$^{**}$ |  | (0.292) |  |  |
| Religious distance index, |  |  |  | 1.373 | 1.172 |  |  |
| relative to United States, weighted |  |  |  | (0.266)$^{**}$ | (0.317)$^{**}$ |  |  |
| $1 -$ % cognate, relative to |  |  |  |  |  |  | 0.631 |
| United States, plurality |  |  |  |  |  |  | (0.185)$^{**}$ |
| Constant | 0.675 | 0.740 | 0.849 | 0.703 | 0.763 | 1.053 | 0.928 |
|  | (0.263)$^{**}$ | (0.257)$^{**}$ | (0.246)$^{**}$ | (0.250)$^{**}$ | (0.246)$^{**}$ | (0.179)$^{**}$ | (0.188)$^{**}$ |
| Standardized beta (%) | 33.41 | 33.26 | 30.84 | 30.18 | 29.42 | 23.58 | 20.63 |
| $R^2$ | .13 | .14 | .16 | .17 | .18 | .14 | .19 |

*Note*. Dependent variable: absolute value of log income differences, 1995. Two-way clustered standard errors in parentheses. 9,316 observations from 137 countries in columns (1)–(5), 1,830 observations from 61 countries in columns (6) and (7). All columns include geographic controls, that is, absolute difference in latitudes, absolute difference in longitudes, geodesic distance, dummy for contiguity, dummy = 1 if either country is an island, dummy = 1 if either country is landlocked, dummy = 1 if pair shares at least one sea or ocean, freight rate for surface transport (estimates not reported).

*Significant at 10%.

**Significant at 5%.

of cognate words between plurality languages.[60] To allow comparisons within a common sample, column (6) presents a baseline regression controlling for geographic distance, transport costs and common history variables, for the sample for which the lexicostatistical measure is available. We find results consistent with the ones obtained using Fearon's discrete measure of linguistic distance: comparing columns (6) and (7) of Table VII, the effect of genetic distance falls by 12.5% when controlling for lexicostatistical distance.

In summary, using the best available measures of linguistic and religious distance, the effect of genetic distance on income differences is reduced by about 12%, but the effect remains large and significant. Overall, these results are consistent with our interpretation: when we measure some specific differences in vertically transmitted traits, such as in language or religion, we obtain a reduction in the size of the coefficient on genetic distance, suggesting that genetic distance was capturing some of the barrier effects associated with differences in these vertical characteristics. However, the reduction is not large enough to suggest that genetic distance only captures the effect of linguistic and religious distance. On the contrary, the reduction is relatively modest, and the effect of genetic distance remains large and significant even when controlling for linguistic and religious distance. This suggests that language and religion are but two of the many vertical characteristics that differ across populations, and perhaps not the most important barriers to the diffusion of economic development.

As already mentioned, linguistic or religious distance and genetic distance do not necessarily capture the same long-term historical relations among populations. Societies with very different languages may share recent common ancestors, and therefore a large number of other important cultural and biological characteristics, whereas societies with different genetic histories and traits may share similar languages and religion because of more recent conquests or conversions. This opens up the very interesting question of what other vertical traits and characteristics are behind the large effects of genetic distance on income differences, besides language and religion. Although the identification of specific traits and characteristics is beyond the scope of this paper, and is left for further research, some further discussion is in order.

---

[60] Results using the weighted measure, which features far fewer observations, were similar and are available upon request.

There are several possible (and not mutually exclusive) channels through which relative genetic distance may operate as a barrier to the diffusion of innovations and development. A possibility is that genetic distance creates obstacles to interaction and communication that cannot be overcome through translation technologies (such as those readily available when people speak different languages). For example, genetic distance may reflect biological traits that, for cultural reasons—racism, discrimination, lack of trust—affect people's willingness to interact with each other. This would be consistent with work by Guiso, Sapienza, and Zingales (2004, revised in 2008) on cultural biases and trade.

Even when people are willing to interact with each other, communication and adaptation of each other's innovations may be hampered by deep cultural differences (norms, values, habits, etc.) that are not codifiable and translatable from one society to the other. This would be consistent with the evidence in Desmet et al. (2007), already mentioned in the Introduction, showing a strong correlation between genetic distance and answers to 430 questions about norms, values, and cultural characteristics in the World Values Survey, correlations that remain even after controlling for linguistic distance.[61] Such characteristics may facilitate the diffusion of innovations across cultures that share a set of common attitudes, while preventing or slowing down such diffusion when societies are more distant across a large range of values and norms.[62] More generally, our results suggest that (a) societies

---

[61] Desmet et al. (2007) compared the matrix of genetic distances between fourteen European countries from Cavalli-Sforza, Menozzi, and Piazza (1994) with the answers given in the World Values Survey (WVS) to 430 questions on "Perceptions of Life, Family and Religion and Morals" from the four WVS waves currently available online at http://www.worldvaluessurvey.org/. They construct a matrix of opinion poll distances across countries, such that each element of the matrix represents the average Manhattan distance between each pair of nations in their respective responses to the 430 questions (p. 25). They find that the WVS matrix of cultural distances and the matrix of genetic distances are strongly correlated, with a correlation coefficient equal to .64, and that "the hypothesis of non-positive correlation is strongly rejected based on a Mantel test with 100,000 replications ($p$-value of .00014)" (p. 27). They also find that the correlation between the matrix of WVS cultural distances and the matrix of genetic distances remains positive and significant even controlling for a matrix of geographical distances and a matrix of linguistic distances.

[62] For example, while the Germans and Austrians share the same language and many other characteristics, there also exist cultural dimensions where the distance between Germany and Britain is smaller than that between Austria and Britain. In the 1999 World Value Survey (http://www.worldvaluessurvey.org/), when asked about "important child quality that children can be encouraged to learn at home," "hard work" was listed as important by 38.7% of respondents in

differ in more respects than those captured by language and geography, (b) such differences go back to the distant historical and possibly prehistoric past, and (c) such differences still matter for differences in income per capita and the diffusion of modern economic development. Genetic distance represents a novel and useful way to summarize these important traits and characteristics, which are—almost by definition—difficult to codify and measure.

### *IV.G. Historical Income Data*

In this subsection we examine the time variation in the effect of genetic distance in the 500 years that surrounded the Industrial Revolution. We find a pattern of coefficients supportive of our model of diffusion. In Table VIII, we use income per capita data since 1500 from Maddison (2003), and repeat our basic reduced form regression for 1500, 1700, 1820, 1870, 1913, and 1960.[63] Our measure of genetic distance is now $F_{\mathrm{ST}}$ genetic distance between plurality groups, relative to the English population.[64] This is both the group to which the plurality genetic group in the United States is matched for the modern period, and (conveniently) the group located in the birthplace of the Industrial Revolution. For the 1500 and 1700 regressions, we use the early match for genetic distance, that is, genetic distance between populations as they were in 1492, prior to the discovery of the Americas and the great migrations

---

Britain, 23% in Germany, and only 9.9% in Austria, whereas "saving money" was mentioned as important by 32.8% of respondents in Britain, 35% in Germany, and 47.6% in Austria. These numbers are just quick examples of measured cultural characteristics where the British happen to be more similar to the Germans than to the Austrians. The genetic distance between the Germans and the English is less than half that between the English and the Austrians in our European sample.

[63] The data on income for 1960 and 1995 are from the Penn World Tables version 6.1. For the pre–Industrial Revolution periods (1500 and 1700), where the level of technology might be well captured by population density rather than per capita income, we also used the absolute difference in log population density as the dependent variable instead of the absolute difference in log per capita income. Relative genetic distance again came out positive and significant, with a standardized beta of about 30%, slightly smaller than but in line with what we find for income differences. For a recent analysis of economic development in precolonial, Malthusian times, using population density as the main dependent variable, see Ashraf and Galor (2008).

[64] We also used Italy as the reference point for the early periods, because there is evidence that Italy was the technological leader in Europe during the Renaissance. This led to no appreciable difference in the results. The Italians and the English are genetically very close relative to average worldwide genetic distance—the genetic distance between the English and the Italians is 0.0072 whereas the average genetic distance between world populations is 0.111.

## TABLE VIII
### REGRESSIONS USING HISTORICAL INCOME DATA (TWO-WAY CLUSTERED STANDARD ERRORS)

|  | (1) Income 1500 | (2) Income 1700 | (3) Income 1820 | (4) Income 1870 | (5) Income 1913 | (6) Income 1960 | (7) Income 1995 |
|---|---|---|---|---|---|---|---|
| Relative $F_{\mathrm{ST}}$ genetic distance to the English, 1500 match | 2.059 | 2.788 |  |  |  |  |  |
|  | (0.479)$^{**}$ | (0.515)$^{**}$ |  |  |  |  |  |
| $F_{\mathrm{ST}}$ genetic distance relative to the English, weighted |  |  | 0.671 | 1.684 | 1.967 | 3.503 | 4.948 |
|  |  |  | (0.338)$^{**}$ | (0.846)$^{**}$ | (0.924)$^{**}$ | (0.784)$^{**}$ | (0.785)$^{**}$ |
| Absolute difference in latitudes | 0.233 | 0.690 | 1.034 | 1.188 | 1.286 | 1.201 | 0.527 |
|  | (0.124)$^{*}$ | (0.180)$^{**}$ | (0.199)$^{**}$ | (0.274)$^{**}$ | (0.263)$^{**}$ | (0.244)$^{**}$ | (0.245)$^{**}$ |
| Absolute difference in longitudes | 0.040 | 0.164 | 0.525 | 0.692 | 0.950 | 0.742 | 0.420 |
|  | (0.075) | (0.085)$^{*}$ | (0.144)$^{**}$ | (0.267)$^{**}$ | (0.261)$^{**}$ | (0.248)$^{**}$ | (0.239)$^{*}$ |
| Geodesic distance (1,000s of km) | −0.015 | −0.082 | −0.096 | −0.124 | −0.175 | −0.171 | −0.087 |
|  | (0.023) | (0.025)$^{**}$ | (0.033)$^{**}$ | (0.064)$^{*}$ | (0.084)$^{**}$ | (0.065)$^{**}$ | (0.040)$^{**}$ |
| =1 for contiguity | −0.051 | −0.168 | −0.226 | −0.257 | −0.272 | −0.102 | −0.466 |
|  | (0.047) | (0.054)$^{**}$ | (0.053)$^{**}$ | (0.048)$^{**}$ | (0.049)$^{**}$ | (0.064) | (0.064)$^{**}$ |
| =1 if either country is an island | −0.069 | 0.003 | −0.067 | 0.070 | 0.062 | −0.010 | 0.177 |
|  | (0.031)$^{**}$ | (0.044) | (0.029)$^{**}$ | (0.099) | (0.088) | (0.073) | (0.095)$^{*}$ |
| =1 if either country is landlocked | −0.042 | −0.018 | 0.136 | 0.173 | 0.217 | 0.125 | 0.083 |
|  | (0.032) | (0.063) | (0.037)$^{**}$ | (0.078)$^{**}$ | (0.081)$^{**}$ | (0.089) | (0.075) |
| =1 if pair shares at least one sea or ocean | −0.027 | −0.045 | −0.009 | 0.070 | 0.118 | 0.082 | −0.011 |
|  | (0.045) | (0.067) | (0.042) | (0.049) | (0.047)$^{**}$ | (0.061) | (0.067) |
| Freight rate (surface transport) | −0.005 | 1.727 | 1.098 | 1.668 | 3.072 | 4.600 | 1.362 |
|  | (0.863) | (1.187) | (1.045) | (2.859) | (4.374) | (3.331) | (1.584) |
| Constant | 0.249 | 0.051 | 0.139 | 0.100 | −0.066 | −0.250 | 0.663 |
|  | (0.149)$^{*}$ | (0.170) | (0.192) | (0.471) | (0.689) | (0.526) | (0.267)$^{**}$ |

TABLE VIII
(CONTINUED)

| | (1)<br>Income<br>1500 | (2)<br>Income<br>1700 | (3)<br>Income<br>1820 | (4)<br>Income<br>1870 | (5)<br>Income<br>1913 | (6)<br>Income<br>1960 | (7)<br>Income<br>1995 |
|---|---|---|---|---|---|---|---|
| # observations | 325 | 406 | 1,035 | 1,485 | 1,653 | 4,753 | 9,316 |
| # countries | 26 | 29 | 46 | 55 | 58 | 98 | 137 |
| Standardized beta (%) | 45.01 | 40.48 | 8.74 | 14.95 | 14.89 | 29.07 | 32.84 |
| Standardized beta (%) (common sample of 325 obs.) | 45.01 | 40.58 | 10.20 | 32.63 | 29.71 | 26.24 | 25.35 |
| $R^2$ | .18 | .24 | .23 | .16 | .17 | .17 | .13 |

*Notes*. Dependent variable: difference in per capita income, in dates specified in row 2. Two-way clustered standard errors in parentheses.
*Significant at 10%.
**Significant at 5%.

of modern times.[65] For the subsequent periods we use the current match.

Table VIII shows that across all periods, the coefficient on relative genetic distance is statistically significant and positive. Moreover, the magnitudes are much larger than for the current period: in regressions obtained from a common sample of 26 countries (325 pairs) for which data are continuously available, standardized beta coefficients range from 25.35% (in 1995) to 45.01% (in 1500). Thus, genetic distance is strongly positively correlated with income differences throughout modern history. It is worth noting that genetic distance bears a large, positive, and significant effect on income differences for the past five centuries, even though income differences in 1500 and in 1995 are basically uncorrelated. (Table II shows this correlation to be −0.051 for the 325 country pairs for which data are available.) This noteworthy fact is highly consistent with our interpretation of genetic distance as a barrier to the diffusion of innovations across populations: genetic distance remains significant throughout the centuries despite significant reversals of fortune since 1500, and despite the fact the genetic distance itself remained highly persistent (composition effects related to the conquest of the New World being the only significant sources of change).

The time pattern of the effect in the common sample of 26 countries provides additional clues that support our interpretation. The standardized beta on genetic distance decreases from 1500 to 1820, then increases significantly in 1870 during the Industrial Revolution, and declines gradually thereafter. The shape of this time path during the 19th century is consistent with the view that the effect captures the diffusion of economic development from the world technological frontier—in particular, the gradual spread of the Industrial Revolution. A major shift in the growth regime (the Industrial Revolution) initially results in large income discrepancies. These discrepancies persist in proportion to genealogical relatedness. As more and more countries adopt the major innovation, the impact of genetic distance progressively

---

[65] Regressions for these early periods feature at most 29 countries. These countries are Australia, Austria, Belgium, Brazil, Canada, China, Denmark, Egypt, Finland, France, Germany, Greece, India, Indonesia, Ireland, Italy, Japan, Korea, Mexico, Morocco, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. There were 325 pairs (26 countries) with available data for 1500 income and 406 pairs (29 countries) for 1700.

declines.[66] At the same time, the slight decrease of the effect in recent times suggests that the impact of genetic distance may progressively decline, as more and more countries adopt the frontier innovations, and intersocietal barriers to the diffusion of development decrease through globalization and other forces.[67]

### *IV.H. Genetic Distance across European Countries*

As the last step in our empirical investigation, we provide a detailed analysis of the European sample. Analyzing the European data can be informative for several reasons. First, it constitutes a robustness check on the worldwide results. Second, matching populations to countries is much more straightforward for Europe than for the rest of the world, because the choice of sampled populations happens to match nation-state boundaries. This should reduce the incidence of measurement error. Third, genetic distances are orders of magnitude smaller across countries of Europe, and genetic specificities there have developed over the last few thousand years (and not tens of thousands of years). It is very unlikely that any genetic traits have risen to prominence within Europe as the result of strong natural selection over such a short period of time, so a finding that genetic distance based on neutral markers within Europe is associated with income differences would be evidence that barriers to the diffusion of development are primarily induced by differences in culturally transmitted traits.

We maintain the choice of the United States as the frontier country. This requires us to use measures of genetic distance based on plurality groups, because we lack the data to calculate weighted genetic distances from European countries to the United States.[68] To maintain consistency throughout, we also use measures of linguistic and religious distance based on plurality languages and plurality religions (this choice does not matter in terms of our

---

[66] In terms of the comparative statics of our simplified reduced-form model, an increase in the effect of genetic distance on income differences may be expected right after a big jump in the parameter $\Delta$ at the technological frontier. A more general interpretation is that the effect should increase after a series of big positive shocks to technology at the frontier, including possibly to the R&D technology itself. See the discussion in Howitt and Meyer-Foulkes (2005). An analytical formalization of these ideas within a dynamic extension of our framework is available upon request.

[67] Within our simplified model, globalization and other forces that reduce intersocietal barriers can be interpreted as a reduction in the parameter $\beta$.

[68] We do not have data on genetic distance between West Africans, Central Amerindians, and Chinese, on the one hand, and European populations at the level of precision of the European data set, on the other. These would be required to compute weighted genetic distance from European countries to the United States.

empirical results). Because the plurality population of the United States is matched to the English population, genetic distance is entered relative to the English. Similarly, linguistic distance is relative to the English language, and religious distance is relative to Protestants.[69]

*Univariate Regressions and Geographic Controls.* In Table IX we present univariate regressions and regressions controlling for geography and transport costs.[70] Genetic distance is again positively and significantly associated with income differences. Columns (1) and (2) confirm empirically Implication 2 of our model—the coefficient on relative genetic distance is about 30% larger than the coefficient on absolute genetic distance. These effects also are about 30% larger in magnitude than the corresponding effects found in the world sample (Table III, columns (1)–(4)). While genetic distances across European countries are smaller than in the world sample, so are the income differences to be explained.

We then add distance metrics and a large set of micro-geography controls as defined in Section IV.D.[71] The main direction of geographic inequality in Europe seems to be longitudinal—between the former Soviet Bloc countries and the West. The controls include the variables used in Giuliano, Spilimbergo, and Tonon (2006), namely a dummy variable taking a value of 1 if countries in a pair share access to the same sea or ocean; a variable that measures the average elevation of countries

---

[69] Our results are robust to using Germany or the United Kingdom as the frontier countries instead of the United States. Results with the Germany and UK baselines are available upon request. It is not surprising that the results using the United Kingdom as the baseline would be similar to those using the United States (the genetic and linguistic groups are the same, and only the religious plurality groups differ). Germans are genetically very close to the English, and like the United States the plurality religion is Protestant, so again the results change little when Germany is used as the baseline. In fact, any genetic group in Northwestern Europe is close to the English.

[70] Summary statistics for the European sample are available upon request. On these, brief observations are in order. First, as stressed by Giuliano, Spilimbergo, and Tonon (2006), genetic distance is indeed correlated with geodesic distance and transportation costs (geodesic distance and freight costs themselves bear a .968 correlation with each other). This remains true, though the correlations are weaker, when relative genetic distance is considered. Second, as in the world sample, relative genetic distance is only weakly correlated with relative linguistic and religious distances. Third, the correlation of absolute log income differences with absolute genetic distance (.328) is smaller that their correlation with relative genetic distance to the English (.409), as our model predicts.

[71] Again, introducing distance metrics and freight costs relative to the frontier (either the United Kingdom, the United States, or Germany) did not change our results concerning genetic distance. These regressions are available upon request.

TABLE IX

Results for the European Data Set (Two-Way Clustered Standard Errors)

| | (1) No controls, simple genetic distance | (2) No controls, relative genetic distance | (3) Add geography | (4) 1870 income data |
|---|---|---|---|---|
| $F_{\mathrm{ST}}$ genetic distance in Europe | 28.134 | | | |
| | (14.605)* | | | |
| $F_{\mathrm{ST}}$ genetic distance, relative to the English | | 45.222 | 39.333 | 39.842 |
| | | (22.193)** | (18.708)** | (11.052)** |
| Absolute difference in latitudes | | | −0.800 | 0.467 |
| | | | (0.723) | (1.397) |
| Absolute difference in longitudes | | | 0.233 | 1.022 |
| | | | (0.129)* | (1.171) |
| Geodesic distance (1,000s of km) | | | −0.344 | −0.150 |
| | | | (0.306) | (0.177) |
| =1 for contiguity | | | −0.136 | −0.204 |
| | | | (0.073)* | (0.063)** |
| =1 if either country is an island | | | −0.078 | 0.039 |
| | | | (0.087) | (0.086) |
| =1 if pair shares at least one sea or ocean | | | −0.159 | −0.063 |
| | | | (0.137) | (0.070) |
| Average elevation between countries | | | −0.028 | −0.049 |
| | | | (0.223) | (0.141) |
| Freight rate (surface transport) | | | 16.532 | −4.004 |
| | | | (14.387) | (5.938) |
| =1 if either country is landlocked | | | 0.074 | [a] |
| | | | (0.178) | |
| Constant | 0.378 | 0.382 | −2.079 | 1.130 |
| | (0.099)** | (0.084)** | (2.293) | (0.947) |
| # of observations | 325 | 325 | 325 | 171 |
| # of countries | 26 | 26 | 26 | 19 |
| Standardized beta (%) | 31.69 | 39.80 | 34.61 | 59.28 |
| $R^2$ | .10 | .16 | .21 | .39 |

*Notes*. Dependent variable: Difference in log per capita income across pairs (in 1995 for columns (1)-(3), in 1870 for column (4)). Two-way clustered standard errors in parentheses. [a]Dropped due to singularity.
*Significant at 10%.
**Significant at 5%.

that lies on the direct path between two countries (for instance, the average elevation between France and Austria is the average elevation of France, Germany, and Austria); and their measure of freight costs from the Import Export Wizard. In addition to these controls, we included additional measures of isolation: a dummy for contiguity, a dummy taking a value of 1 if at least one country in a pair is landlocked, and a similar dummy variable for islands. Together, the inclusion of these variables reduces the standardized effect of relative genetic distance from 41.08%

to 34.61%, but the effect remains larger than in the world sample, and statistically significant at the 5% level. Although it is clearly crucial to control for geographic factors here, these do not seem to play nearly as important a role in income differences as they apparently do in bilateral trade between European countries.

The last column of Table IX uses income differences in 1870 as the dependent variable. Although we lose seven countries for lack of income data in the Maddison source, the results on genetic distance relative to the English are quantitatively and statistically much stronger than those for the contemporary period, consistent with the findings reported in Section IV.G. Upon impact, a major new innovation such as the Industrial Revolution diffuses in proportion to how genetically distant countries are from the frontier.[72]

*Controlling for Linguistic and Religious Distance.* In a short comment on our work, Fearon (2006) examined the interrelationships between genetic and linguistic distance within Europe. Regressing income levels on genetic distance from the English, geodesic distance to the United Kingdom, and linguistic distance from the English language (using both his data based on linguistic trees and the lexicostatistical data), he found that genetic distance was generally robust to the inclusion of these variables.[73] We reexamine this issue using our bilateral methodology and with our full set of geographic controls. Table X presents the results.[74] The bottom line is that the magnitude of the genetic distance effect is not affected by the inclusion of linguistic and religious distance. We refer the reader to Section IV.F for an interpretation of these results.

---

[72] It is convenient (and surely not a coincidence) that the baseline population for calculating relative genetic distance, the English, is the plurality group both in the United States (the frontier in 1995) and in the birthplace of the Industrial Revolution (the frontier in 1870).

[73] Fearon (2006) presented one regression with 22 observations using the lexicostatistical data where the $t$-statistic on the coefficient on genetic distance to the English fell to 1.5. We have replicated this regression with all 23 European countries for which the lexicostatistical and linguistic trees data are available and found that genetic distance remained statistically significant at the 5% level (Iceland, Hungary, and Finland are missing from these regressions due to lack of data on one or the other linguistic distance measure). These results are available upon request.

[74] While our baseline sample features 26 countries, we lack observations on linguistic and religious distance for Iceland, and we lack lexicostatistical data for Hungary and Finland, where Indo-European languages are not spoken. As a result, columns (1) and (3) present baseline regressions, including geographic controls only, for comparison.

TABLE X

CONTROLLING FOR RELIGIOUS AND LINGUISTIC DISTANCE IN THE EUROPE DATA SET (TWO-WAY CLUSTERED STANDARD ERRORS)

| | (1) Baseline | (2) Linguistic and religious distance | (3) Baseline (smaller sample) | (4) % cognate, plurality |
|---|---|---|---|---|
| $F_{ST}$ genetic distance, relative to the English | 41.691<br>(18.875)** | 42.485<br>(19.310)** | 44.252<br>(20.209)** | 44.096<br>(20.288)** |
| Linguistic distance, plurality, relative to English | | −0.224<br>(0.109)** | | |
| Religious distance, plurality, relative to Protestants | | 0.107<br>(0.171) | | |
| 1 − % cognate, plurality, relative to English | | | | 0.045<br>(0.186) |
| Constant | −1.537<br>(2.025) | −1.445<br>(1.973) | −1.339<br>(2.161) | −1.314<br>(2.231) |
| # of observations | 300 | 300 | 276 | 276 |
| # of countries | 25 | 25 | 24 | 24 |
| Standardized beta (%) | 37.55 | 38.26 | 39.52 | 39.38 |
| $R^2$ | .21 | .22 | .25 | .25 |

*Notes*. Dependent variable: absolute value of log income differences, 1995. Two-way clustered standard errors in parentheses. All columns include the following controls (estimates not reported): absolute difference in latitudes, absolute difference in longitudes, geodesic distance, dummy for contiguity, dummy = 1 if either country is landlocked, dummy = 1 if pair shares at least one sea or ocean, average elevation between countries, freight rate (surface transport). Compared to Table IX, in columns (1) and (2) Iceland is dropped due to missing data on linguistic and religious distance from Fearon. In columns (4) and (5) Hungary and Finland are dropped because their languages are not Indo-European, and thus not part of the lexicostatistical data set.

*Significant at 10%.

**Significant at 5%.

## V. CONCLUSION

In this paper we make two contributions: (1) For the first time, we document a statistically and economically significant positive relationship between measures of genetic distance and cross-country income differences, even controlling for measures of geographical and climatic distances, transportation costs, and measures of historical, linguistic, and religious distance. (2) We provide an economic interpretation of these findings, in terms of barriers to the diffusion of development from the world technological frontier.

Our interpretation is based on two main ideas. The first idea is that, on average, genetic distance captures divergence in characteristics that are transmitted across generations within populations over the very long run. Genetic distance, measuring the time since two populations shared common ancestors, provides an ideal summary of differences in slowly changing genealogically transmitted characteristics, including habits and customs. The second idea is that such differences act as barriers to the diffusion of development from the world technological frontier.

The empirical evidence is consistent with this barriers interpretation. In line with our framework, the effect on economic distance associated with *relative* genetic distance from the world technological frontier is larger than the effect of absolute genetic distance. We also found that the effect has varied across time and space in ways that support our diffusion-from-the-frontier interpretation: the effect increased in the first part of the nineteenth century, peaked in 1870, and slightly decreased afterward, consistent with the view that relative genetic distance captures barriers to the diffusion of the Industrial Revolution. Some evidence, particularly the results for European countries, suggests that these differences may stem in substantial part from cultural (rather than purely biological) transmission of characteristics across generations.

Although our analysis provides a general macroeconomic framework to interpret our empirical findings, the study of the specific microeconomic mechanisms through which the effects operate is left for future research. An analysis of microeconomic data may shed light on the relations among genetic distance, vertical characteristics, imitation costs, and the spread of specific innovations.[75] Interestingly, we have found that linguistic and religious distances, two culturally transmitted characteristics, only slightly reduce the effect of genetic distance on income differences, therefore suggesting a role for other slow-changing biological and/or cultural traits—including differences in customs, norms, or values. These traits are inherently harder to measure, particularly within the long-term macroeconomic perspective that we have adopted. Another natural extension of our work would be to investigate whether and how genetic distance affects bilateral exchanges and interactions between different groups and societies, both peacefully (trade, foreign direct investment) and nonpeacefully (conflict and wars). Finally, it would be interesting to link our results to the vast literature on demographics and economic

---

[75] For instance, recent microeconomic analysis of international technological diffusion finds an important role for ethnic scientific communities, consistent with our interpretation (Kerr 2008).

growth and explore the connections between genetic distance, intergenerationally transmitted characteristics, and demographic patterns.[76]

A final consideration is about policy implications. A common concern with research documenting the importance of variables such as genetic distance or geography is pessimism about policy implications. What use is it to know that genetic distance explains income differences, if one cannot change genetic distance, at least in the short run? These concerns miss a bigger point: available policy variables may have a major impact not on genetic distance itself, but on the magnitude of the effect of genetic distance on income differences. This magnitude has changed over time, and can change further. If we are correct in interpreting our results as evidence for long-term barriers across different societies, significant reductions in income disparities could be obtained by encouraging policies that reduce such barriers, including efforts to translate and adapt technological and institutional innovations into different histories and traditions, and to foster cross-societal exchanges and openness. More work is needed—at the micro as well as macro level—to understand the specific mechanisms, market forces, and policies that could facilitate the diffusion of development across societies with distinct long-term histories.

## APPENDIX I: DEFINITION OF $F_{\mathrm{ST}}$

In this Appendix we illustrate the construction of $F_{\mathrm{ST}}$ for the simple case of two populations ($a$ and $b$) of equal size, and one gene that can take only two forms (allele 1 and allele 2). Let $p_a$ and $q_a = 1 - p_a$ be the gene frequencies of allele 1 and allele 2, respectively, in population $a$.[77] The probability that two randomly selected alleles at the given locus are *identical* within the population (homozygosity) is $p_a^2 + q_a^2$, and the probability that they are different (heterozygosity) is

$$h_a = 1 - \left(p_a^2 + q_a^2\right) = 2p_a q_a. \tag{13}$$

---

[76] For instance, Coale and Cotts Watkins (1986) documented the correlation between cultural similarity and the time paths of fertility across Europe (see also Richerson and Boyd [2004, Chapter 5]). Galor (2005) provides an in-depth discussion of the economic literature on demographics and growth.

[77] Note that because $p_a + q_a = 1$ we also have $(p_a + q_a)^2 = p_a^2 + q_a^2 + 2p_a q_a = 1$.

By the same token, heterozygosity in population $b$ is

$$h_b = 1 - \left(p_b^2 + q_b^2\right) = 2p_b q_b, \tag{14}$$

where $p_b$ and $q_b = 1 - p_b$ are the gene frequencies of allele 1 and allele 2, respectively, in population $b$. The average gene frequencies of allele 1 and 2 in the two populations are, respectively,

$$\overline{p} = \frac{p_a + p_b}{2} \tag{15}$$

and

$$\overline{q} = \frac{q_a + q_b}{2} = 1 - \overline{p}. \tag{16}$$

Heterozygosity in the *sum* of the two populations is

$$h = 1 - (\overline{p}^2 + \overline{q}^2) = 2\overline{pq}. \tag{17}$$

Average heterozygosity is measured by

$$h_m = \frac{h_a + h_b}{2}. \tag{18}$$

$F_{\mathrm{ST}}$ measures the variation in the gene frequencies of populations by comparing $h$ and $h_m$:

$$F_{\mathrm{ST}} = 1 - \frac{h_m}{h} = 1 - \frac{p_a q_a + p_b q_b}{2\overline{pq}} = \frac{1}{4}\frac{(p_a - p_b)^2}{\overline{p}(1 - \overline{p})}. \tag{19}$$

If the two populations have identical allele frequencies ($p_a = p_b$), $F_{\mathrm{ST}}$ is zero. On the other hand, if the two populations are completely different at the given locus ($p_a = 1$ and $p_b = 0$, or $p_a = 0$ and $p_b = 1$), $F_{\mathrm{ST}}$ takes the value 1. In general, the higher the variation in the allele frequencies across the two populations, the higher is their $F_{\mathrm{ST}}$ distance. The formula can be extended to account for $L$ alleles, $S$ populations, and different population sizes and to adjust for sampling bias. The details of these generalizations are provided in Cavalli-Sforza, Menozzi, and Piazza (1994, pp. 26–27).

## APPENDIX II: SPATIAL CORRELATION

This Appendix illustrates why spatial correlation may be present in our bilateral analysis. Consider three countries, 1, 2, and 3. Observations on the dependent variable, $|\log y_1 - \log y_2|$ and $|\log y_1 - \log y_3|$, will be correlated by virtue of the presence of country 1 in both observations. Conditioning on the right-hand-side variables (which are bilateral in nature) should reduce cross-sectional dependence in the errors $\varepsilon_{12}$ and $\varepsilon_{13}$, but we are unwilling to assume that observations on the dependent variable are independent conditional on the regressors. In other words, simple least-squares standard errors will lead to misleading inferences due to spatial correlation.

With $N$ countries, there are $N(N-1)/2$ distinct pairs. Denote the observation on absolute value income differences between country $i$ and country $j$ as $dy_{ij}$. Pairs are ordered so that country 1 appears in position $i$ and is matched with all countries from 2 to $N$ appearing in position $j$. Then country 2 is in position $i$ and is matched with 3 to $N$ appearing in position $j$, and so on. The last observation has country $N-1$ in position $i$ and country $N$ in position $j$. We denote the nonzero off-diagonal elements of the residual covariance matrix by $\sigma_m$, where $m$ is the country common to each pair.

A simple example when the number of countries is $N = 4$ is illustrative. In this case, under our maintained assumption that the error covariances among pairs containing a common country $m$ are equal to a common value $\sigma_m$, the covariance matrix of the vector of residuals $\varepsilon$ is of the form

$$\Omega = \operatorname{cov}\begin{pmatrix} \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{34} \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 & & & & & \\ \sigma_1 & \sigma_\varepsilon^2 & & & & \\ \sigma_1 & \sigma_1 & \sigma_\varepsilon^2 & & & \\ \sigma_2 & \sigma_3 & 0 & \sigma_\varepsilon^2 & & \\ \sigma_2 & 0 & \sigma_4 & \sigma_2 & \sigma_\varepsilon^2 & \\ 0 & \sigma_3 & \sigma_4 & \sigma_3 & \sigma_4 & \sigma_\varepsilon^2 \end{pmatrix}.$$

This clearly demonstrates the presence of cross-sectional (spatial) correlation. It is important to note, however, that our data are not linearly dependent; that is, there is additional information brought in by considering the bilateral approach. One major reason is that the dependent variable is the absolute difference in log income, not just the difference in log income. It is easy to show that taking absolute values greatly reduces spatial dependence in the dependent variable. Another major reason is that we are conditioning on right-hand-side variables (such as geodesic distance and genetic distance) that are truly bilateral in nature; that is, our empirical model is *not* the result of simply differencing a "level" specification across cross-sectional units.

TUFTS UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH
UNIVERSITY OF CALIFORNIA, LOS ANGELES AND NATIONAL BUREAU OF ECONOMIC RESEARCH

## REFERENCES

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

——, "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 117 (2002), 1231–1294.

Alcalá, Francisco, and Antonio Ciccone, "Trade and Productivity," *Quarterly Journal of Economics*, 119 (2004), 612–645.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," *Journal of Economic Growth*, 8 (2003), 55–194.

Ashraf, Quamrul, and Oded Galor, "Human Genetic Diversity and Comparative Economic Development," working paper, Brown University, 2008.

Barro, Robert J., and Xavier Sala-i-Martin, "Technological Diffusion, Convergence and Growth," *Journal of Economic Growth*, 2 (1997), 1–26.

Bisin, Alberto, and Thierry Verdier, "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics*, 105 (2000), 955–988.

——, "The Economics of Cultural Transmission and the Evolution of Preferences," *Journal of Economic Theory*, 97 (2001), 98–319.

Boyd, Robert, and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago, IL: University of Chicago Press, 1985).

Boyd, Robert, and Joan B. Silk, *How Humans Evolved* (New York, NY: Norton, 2003).

Brezis, Elise, Paul Krugman, and Daniel Tsiddon, "Leapfrogging in International Competition: A Theory of Cycles in National Technological Leadership," *American Economic Review*, 83 (1993), 1211–1219.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Robust Inference with Multi-Way Clustering," NBER Technical Working Paper No. T0327, 2006.

Case, Anne, "Spatial Patterns in Household Demand," *Econometrica*, 59 (1991), 953–965.

Cavalli-Sforza, Luigi L., and Francesco Cavalli-Sforza, *The Great Human Diasporas* (Reading, MA: Addison Wesley, 1995).

Cavalli-Sforza, Luigi L., and Marcus W. Feldman, *Cultural Transmission and Evolution* (Princeton, NJ: Princeton University Press, 1981).

Cavalli-Sforza, Luigi L., Paolo Menozzi, and Alberto Piazza, *The History and Geography of Human Genes* (Princeton, NJ: Princeton University Press, 1994).

Clark, Gregory, *A Farewell to Alms. A Brief Economic History of the World* (Princeton, NJ: Princeton University Press, 2007).

Clark, Gregory, and Susan Wolcott, "Why Nations Fail: Managerial Decisions and Performance in Indian Cotton Textiles, 1890–1938," *Journal of Economic History*, 59 (1999), 397–423.

Coale, Ansley J., and Susan Cotts Watkins, *The Decline of Fertility in Europe* (Princeton, NJ: Princeton University Press, 1986).

Dawkins, Richard, *The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution* (Boston, MA: Houghton Mifflin, 2004).

Desmet, Klaus, Michel Le Breton, Ignacio Ortuno Ortin, and Shlomo Weber, "Stability of Nations and Genetic Diversity," working paper, Universidad Carlos III, 2007.

Diamond, Jared, *The Third Chimpanzee. The Evolution and Future of the Human Animal* (New York, NY: Harper Collins, 1992).

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

Dyen, Isidore, Joseph B. Kruskal, and Paul Black, "An Indoeuropean Classification: A Lexicostatistical Experiment," *Transactions of the American Philosophical Society*, 82 (1992), 1–132.

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

Fearon, James, "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth,* 8 (2003), 195–222.

——, "Is Genetic Distance a Plausible Measure of Cultural Distance?," working paper, Stanford University, 2006.

Gallup, John L., Andrew D. Mellinger, and Jeffrey D. Sachs, "Geography and Economic Development," NBER Working Paper No. 6849, 1998.

Galor, Oded, "From Stagnation to Growth: Unified Growth Theory," in *Handbook of Economic Growth*, Philippe Aghion and Steven Durlauf, eds. (Amsterdam, the Netherlands: North-Holland, 2005).

Galor, Oded, and Omer Moav, "Natural Selection and the Origin of Economic Growth," *Quarterly Journal of Economics*, 117 (2002), 1133–1191.

Giuliano, Paola, Antonio Spilimbergo, and Giovanni Tonon, Genetic, Cultural and Geographical Distances, unpublished, International Monetary Fund, 2006.

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

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Cultural Biases in Economic Exchange," NBER Working Paper No. 11005, 2004.

——, "Does Culture Affect Economic Outcomes?," *Journal of Economic Perspectives*, 20 (2006), 23–48.

Hall, Robert, and Jones, Charles I., "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics*, 114 (1999), 83–116.

Heston, Alan, Robert Summers, and Bettina Aten, "Penn World Table Version 6.1," Center for International Comparisons at the University of Pennsylvania (CICUP), 2002.

Howitt, Peter, and David Mayer-Foulkes, "R&D, Implementation, and Stagnation: A Schumpeterian Theory of Convergence Clubs," *Journal of Money, Credit and Banking*, 37 (2005), 147–177.

Hummels, David, and V. Lugovskyy "Usable Data? Matched Partner Trade Statistics as a Measure of Transportation Cost," *Review of International Economics*, 14 (2006), 69–86.

Jobling, Mark A., Matthew E. Hurles, and Chris Tyler-Smith, *Human Evolutionary Genetics. Origins, Peoples & Diseases* (New York, NY: Garland Science, 2004).

Kerr, William R., "Ethnic Scientific Communities and International Technology Diffusion," *Review of Economics and Statistics*, 98 (2008), 518–537.

Kimura, Motoo, "Evolutionary Rate at the Molecular Level," *Nature*, 217 (1968), 624–626.

Kremer, Michael, "Population Growth and Technological Change: 1,000,000 B.C. to 1990," *Quarterly Journal of Economics*, 108 (1993), 681–716.

Limao, Nuno, and Anthony Venables, "Infrastructure, Geographical Disadvantage, Transport Costs and Trade," *World Bank Economic Review*, 15 (2001), 451–479.

Maddison, Angus, *The World Economy: Historical Statistics* (Paris, France: OECD Development Center, 2003).

Masters, William A, and Margaret S. McMillan, "Climate and Scale in Economic Growth," *Journal of Economic Growth*, 6 (2001), 167–186.

McNeill, John Robert, and William H. McNeill, *The Human Web: A Bird's Eye View of Human History* (New York, NY: Norton, 2003).

Mecham, R. Quinn, James Fearon, and David Laitin, Religious Classification and Data on Shares of Major World Religions, unpublished, Stanford University, 2006.

Olsson, Ola, and Douglas A. Hibbs Jr., "Biogeography and Long-Run Economic Development," *European Economic Review*, 49 (2005), 909–938.

Parente, Stephen L., and Edward C. Prescott, "Barriers to Technology Adoption and Development," *Journal of Political Economy*, 102 (1994), 298–321.

——, *Barriers to Riches* (Cambridge, MA: MIT Press, 2002).

Richerson, Peter J., and Robert Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution* (Chicago, IL: University of Chicago Press, 2004).

Rogers, Everett M., *Diffusion of Innovations* (New York, NY: The Free Press, 1962).

Sachs, Jeffrey, "Tropical Underdevelopment," NBER Working Paper No. 8119, 2001.

Shennan, Stephen, *Genes, Memes and Human History. Darwinian Archaeology and Cultural Evolution* (London: Thames and Hudson, 2002).

Tabellini, Guido, "Culture and Institutions: Economic Development in the Regions of Europe," IGIER Working Paper No. 292, 2005.

Weitzman, Martin, "On Diversity," *Quarterly Journal of Economics*, 107 (1992), 363–405.

Whitehead, Alfred North, *Science and the Modern World* (Cambridge, UK: Cambridge University Press, 1931).