DISCUSSION PAPER SERIES

No. 5630

THE DIFFUSION OF DEVELOPMENT

Enrico Spolaore and Romain Wacziarg

INSTITUTIONS AND ECONOMIC PERFORMANCE (FORMERLY TRANSITION ECONOMICS)

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP5630.asp**

ISSN 0265-8003

THE DIFFUSION OF DEVELOPMENT

Enrico Spolaore, Tufts University
Romain Wacziarg, Stanford University and CEPR

Discussion Paper No. 5630
April 2006

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INSTITUTIONS AND ECONOMIC PERFORMANCE (FORMERLY TRANSITION ECONOMICS)**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Enrico Spolaore and Romain Wacziarg

CEPR Discussion Paper No. 5630

April 2006

ABSTRACT

The Diffusion of Development*

This paper studies the barriers to the diffusion of development across countries over the very long-run. We find that genetic distance, a measure associated with the amount of time elapsed since two populations' last common ancestors, bears a statistically and economically significant correlation with pairwise income differences, even when controlling for various measures of geographical isolation, and other cultural, climatic and historical difference measures. These results hold not only for contemporary income differences, but also for income differences measured since 1500 and for income differences within Europe. We uncover similar patterns of coefficients for the proximate determinants of income differences, particularly for differences in human capital and institutions. The paper discusses the economic mechanisms that are consistent with these facts. We present a framework in which differences in human characteristics transmitted across generations - including culturally transmitted characteristics - can affect income differences by creating barriers to the diffusion of innovations, even when they have no direct effect on productivity. The empirical evidence over time and space is consistent with this 'barriers' interpretation.

JEL Classification: O11 and O57
Keywords: barriers, genetic distance and income differences

Enrico Spolaore
Department of Economics
Tufts University
Medford
MA 02155
USA
Tel: (1 617) 627 4068
Email: enrico.spolaore@tufts.edu

Romain Wacziarg
Graduate School of Business
Stanford University
518 Memorial Way
Stanford CA 94305-5015
USA
Tel: (1 650) 723 6069
Fax: (1 650) 725 7979
Email: wacziarg@gsb.stanford.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=128278

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=139873

* We thank Alberto Alesina, Robert Barro, Francesco Caselli, Steven Durlauf, Jim Fearon, Oded Galor, Yannis Ioannides, Pete Klenow, Andros Kourtellos, Ed Kutsoati, Peter Lorentzen, Paolo Mauro, Sharun Mukand, Gérard Roland, Antonio Spilimbergo, Chih Ming Tan, David Weil, Bruce Weinberg, Ivo Welch, and seminar participants at Boston College, Fundação Getulio Vargas, Harvard University, the IMF, INSEAD, London Business School, Northwestern University, Stanford University, Tufts University, UC Berkeley, UC San Diego, the University of British Columbia, the University of Connecticut and the University of Wisconsin- Madison for comments. We also thank Robert Barro and Jim Fearon for providing some data. The usual disclaimer applies.

Submitted 17 March 2006

1 Introduction

What explains the vast differences in income per capita that are observed across countries? Why are these differences so persistent over time? In this paper, we argue that barriers to the diffusion of development prevent poor countries from adopting economic practices, institutions and technologies that make countries rich. We argue that these barriers are not only geographic, but also human. We propose and test the hypothesis that cross-country differences in human characteristics that are transmitted with variations from parents to children create barriers to the diffusion of development. These *vertically transmitted characteristics* include cultural features, such as language and habits, among other characteristics of human populations.

In recent years a large empirical literature has explored the determinants of income levels using cross-country regressions, in which the level of development, measured by income per capita, is regressed on a set of explanatory variables.[1] In this paper, we depart from this usual methodology. We are not primarily concerned with the factors that make countries rich or poor. Instead, we are concerned with the barriers that prevent poor countries from adopting better income determinants, whatever they are. To do so, we use income differences between pairs of countries as our dependent variable, and various measures of distance as regressors. Our approach allows us to consider measures of distance between countries across several dimensions, and to investigate whether those distances play a role as barriers to the diffusion of development.[2]

For the first time, we document and discuss the relationship between genetic distance and differences in income per capita across countries. We use genetic distance between populations as a measure of their degree of similarity in vertically transmitted characteristics (VTCs). Measures

[1] Recent contributions to this literature include Hall and Jones (1999), Acemoglu, Johnson and Robinson (2001), Easterly and Levine (2003), Alcalá and Ciccone (2004), Rodrik, Subramanian and Trebbi (2004), Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) among others.

[2] There is a voluminous literature on cross-country income convergence, dating back to Baumol (1986). In the neoclassical literature, convergence occurs because the marginal return to capital is higher in countries farther from their steady-state, which depends, among other things, on the level of technology (the "A" parameter). In contrast, we seek to shed light on the factors that prevent or facilitate the diffusion of productivity-enhancing innovations across countries. In this respect, our paper is closer to the approach in Barro and Sala-i-Martin (1997), where technological diffusion drives convergence. Policy-induced constraints on the diffusion of technology are analyzed by Parente and Prescott (1994, 2002). Policy experimentation and imitation across neighbors are studied by Mukand and Rodrik (2005). Unlike these contributions, we consider more broadly the barriers to the diffusion of technological and institutional characteristics in the very long run.

of genetic distance between populations are based on aggregated differences in allele frequencies for various *loci* on a chromosome. In this paper we use measures of F_{ST} distance, also known as coancestor coefficients. F_{ST} distances, like most measures of genetic diversity, are based on indices of heterozygosity, the probability that two genes at a given locus, selected at random from the relevant populations, will be different (heterozygous). Since most genetic differences tend to accumulate at a regular pace over time, as in a kind of molecular clock, genetic distance is closely linked to the time since two populations' last common ancestors - that is, the time since two populations were in fact the same population. Hence, genetic distance can be used to determine paths of genealogical relatedness of different populations over time (phylogenetic trees).[3]

The main findings of this paper are fourfold. First, measures of genetic distance between populations bear a statistically and economically significant effect on differences in income per capita, even when controlling for various measures of geographical isolation, and other cultural, climatic and historical difference measures. Second, the effect of genetic distance holds not only for contemporary income differences, but also for income differences measured since 1500. While the effect is always large, positive, and significant, the magnitude of the effect has varied over time in an interesting way. The effect declined from 1500 to 1820, went up dramatically, peaked at the time of the Industrial Revolution, and steadily declined afterwards. Third, the effect of genetic distance on income differences is *larger* for countries that are geographically *closer*. Finally, the effect of genetic distance holds not only for contemporary and historical worldwide income differences, but also for income differences within Europe. The magnitude of the effect of genetic distance is *larger* within European countries than across countries from all continents.

In a nutshell, the correlation between genetic distance and income differences is extremely robust over time and space, but also presents important variations over those dimensions. These variations over time and space provide valuable clues about the economic interpretation of the effect.

What is the economic meaning of this effect? One possibility is that this correlation may just reflect the impact of variables affecting both genetic distance and income differences. If that were the case, controlling for those variables would eliminate the effect of genetic distance on

[3] Our main source for genetic distances between human populations is Cavalli-Sforza, Menozzi and Piazza (1994). The classical reference on evolutionary rates at the molecular level is Kimura (1968). A recent textbook reference on human population genetics is Jobling, Hurles and Tyler-Smith (2004). For a nontechnical discussion of these concepts see Dawkins (2004).

income difference. We control for a large number of reasonable suspects (geographical and climatic differences, measures of geographical isolation, etc.). In particular, we control for geography and region-specific differences that may impede the diffusion of development, as emphasized by Jared Diamond in his influential book *Guns, Germs and Steel* (1997).[4] We find that these geographical and regional variables often do have an effect on income differences, but that their inclusion does not eliminate the effect of genetic distance as an independent explanatory variable. Moreover, the effect of genetic distance on income differences holds within Europe, where geographic differences are much smaller.

Our empirical analysis opens the door to a causal interpretation of the relationship between genetic distance and income differences. What mechanisms can explain a causal link? It is important to stress that a link from genetic distance to income differences is *not* evidence that the mechanisms themselves are genetic. On average populations that are more genetically distant have had more time to diverge in a broad variety of characteristics transmitted intergenerationally. These include characteristics that are passed on genetically, through DNA, but also some that are passed on non-genetically, i.e. culturally.[5] As long as these cultural characteristics are transmitted to younger generations from genetically related individuals, they will be correlated with genetic distance. Language is an obvious example. While humans are genetically predisposed to learn *some* language, there is no gene for speaking Japanese or Italian. However, people who speak the same language tend to be closely related genetically because most children learn their language from their parents. Moreover, since languages (and other deep cultural characteristics) change gradually over time, people who speak more similar languages also tend to be closer to each other genealogically.[6]

Therefore, one should not view genetic distance as an exclusive measure of distance in DNA-transmitted characteristics. It is more appropriate to interpret genetic distance as a general metric

[4] See also Olsson and Hibbs (2005).

[5] By vertical cultural transmission we mean any transmission of characteristics from parents to children that does not take place through DNA, such as language. Evolutionary models of cultural transmission have been developed, by Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985). For a nontechnical discussion see Cavalli-Sforza and Cavalli-Sforza (1995, chapter 8). Economic models of cultural transmission from parents to children have been provided by Bisin and Verdier (2000, 2001). Galor and Moav (2003) present an innovative theory of long-term economic growth in which a key role is played by evolutionary changes in preference parameters that are genetically transmitted across generations. For an in-depth discussion of these issues, see also Galor (2005).

[6] See Cavalli-Sforza, Menozzi and Piazza, 1994, pp. 96-105.

for genealogical distance between populations, capturing overall average differences not only in genetically transmitted features but also in culturally transmitted characteristics. In this paper we will define vertically transmitted characteristics (VTCs) to be all characteristics passed on from parents to children, whether through DNA or culturally. If we take this broader perspective, we can interpret the effect of genetic distance on income differences as evidence of an important role for vertically-transmitted characteristics, reflecting divergent historical paths of different populations over the long run.[7]

Rather than addressing the "nature versus nurture" debate, which is beyond the scope of our analysis, we interpret our findings as evidence for the economic importance of long-term divergence in VTCs of different populations: the diffusion of development is impeded by barriers arising from differences in VTCs. That said, it is also true that we find clues pointing to cultural transmission rather than purely genetic transmission as a likely mechanism behind our results. For instance, as we already mentioned, we find large effects of genetic distance on income differences within Europe. That is, genetic distance explains income differences between populations that are geographically close, have shared very similar environments, and have had a very short time to diverge genetically (in many cases, less than a few thousand years). Since cultural change is much faster than genetic change, and most genetic change, especially in the short-run, is neutral (i.e. unrelated to natural selection), our findings are consistent with cultural transmission as a key mechanism explaining persistent income differences.[8]

While we do not wish to push the distinction between genetic and cultural transmission too far, we do stress a different distinction. That is the distinction between a *direct effect* and a *barrier effect*. VTCs have a direct effect if they enter directly into the production function - say, by improving total factor productivity. An example would be the transmission of a more productive work ethic from parents to children. By contrast, a barrier effect occurs if different characteristics between populations prevent or reduce the diffusion of productivity-enhancing innovations (more productive technology, institutions, etc.). For example, differences in language may have no direct

[7] Moreover, as we briefly discuss in Section 2.2, any sharp distinction between "genetic" and "cultural" characteristics may be misleading, since the economic impact of genetic and cultural characteristics is likely to depend on their combination and interaction.

[8] 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, 1997), Cavalli-Sforza and Cavalli-Sforza (1995), Olson (2002) and Richerson and Boyd (2004).

bearing on productivity but may act as obstacles to the introduction of innovations arising from populations with different languages. Another way of stating this idea is to say that differences in VTCs are obstacles to the *horizontal diffusion of development*.[9] Hence, in principle, genetic distance may explain income differences because of direct effects (some populations have more productive VTCs than others), barrier effects (different VTCs prevent the horizontal diffusion of innovations), or both. It is worth noting that either effect would be sufficient to account for the correlation we document.[10]

Generally, a precise decomposition of the two effects is conceptually and empirically difficult, as some VTCs may have both direct and barrier effects. However, our data provide clear indications that VTCs act at least in part as *barriers* to the diffusion of development. First, there exists a *negative interaction* between genetic distance and geographical distance. That is, we find that genetic distance has a bigger effect on income differences for country pairs that are geographically close. This result is consistent with a simple model in which geographical and genetic distance are both barriers to the diffusion of innovations. The intuition is straightforward: if genetic distance acts as a barrier, it matters more for countries that are nearby, and face lower geographical barriers to exchange with each other, while it is less important for countries that are far away, and would learn little from each other anyway due to geographic distance.[11] Second, there is a pattern in the effect of genetic distance on income differences over time, from 1500 to today: while the effect of genetic distance is always large, positive and significant, it varies over time in an interesting way. The effect declined from 1500 to 1820, spiked up and peaked in 1870, and steadily declined again afterwards. This is consistent with the interpretation of genetic distance as related to barriers to the diffusion of innovations. The effect of barriers should peak when a major innovation is introduced and initially adopted only by the populations that are closest to the innovator (such as the industrial revolution in the 19th century), but decline over time as the major innovation spreads to more distant populations.

Our paper is organized as follows. In Section 2, we present a simple analytical framework to help

[9] In the anthropology literature, vertical transmission takes place across (usually related) generations, while horizontal transmission takes place across (possibly unrelated) groups of people belonging to the same generation.

[10] In fact, Section 2.1 presents a simple model in which barriers associated with differences in neutral VTCs are *sufficient* to explain a positive effect of genetic distance on income differences.

[11] By contrast, a simple model where VTCs and geographical characteristics directly affect the production function would not typically imply a negative interaction term.

in the interpretation of our empirical work. Our simple model illustrates a) the link between genetic distance and distance in VTCs, and b) the link between differences in VTCs and the diffusion of innovations across populations. A key point is to show that random divergence in neutral VTCs is sufficient to generate income differences if those VTCs are barriers to the horizontal diffusion of innovations. An extension to non-neutral VTCs strengthens the link between genetic distance and income differences. This section also presents and discusses a general taxonomy of the different channels through which genetic distance may affect income differences. Section 3 discusses the data and the empirical methodology used in this paper. Since we regress pairwise differences in income on distance measures, we face a problem of spatial correlation, and address this estimation issue using a new econometric methodology. Section 4 presents our empirical results: consistent with our theoretical framework, we document that genetic distance is positively related to pairwise differences in income per capita and in its proximate determinants. Section 5 concludes.

2 Theoretical Framework

In the first part of this section (Section 2.1) we propose a simple analytical framework to study the diffusion of technological and institutional innovations across societies and its relationship with genetic distance In the second part (Section 2.2) we provide a general discussion of the channels through which genetic distance may affect income differences, and briefly discuss these channels in relation to the existing literature.

2.1 VTCs, Genealogical Distance and the Horizontal Transmission of Innovations

2.1.1 Setup

Our model starts from the following assumptions:

a) Innovations may be transmitted vertically (across generations within a given population) and horizontally (across different populations).

b) The horizontal diffusion of innovations is not instantaneous, but is a function of barriers to technological and institutional diffusion.

c) Barriers to technological and institutional diffusion across societies are a function of how far societies are from each other as a result of divergent historical paths.

Productive knowledge is summarized by a positive real number A_{it}. We assume a linear technology $Y_{it} = A_{it}L_{it}$, where L_{it} is the size of the population, which implies that income per capita is given by $y_{it} \equiv Y_{it}/L_{it} = A_{it}$.

For simplicity, we summarize all other relevant characteristics of a society (cultural habits and traditions, language, etc.) as a point on the real line. That is, we will say that at each time t a population i will have cultural characteristics q_{it}, where q_{it} is a real number.[12] These characteristics are transmitted across generations with variations.[13] Over time, characteristics change (vocabulary and grammar are modified, some cultural habits and norms are dropped while new ones are introduced, etc.). Hence, at time $t+1$ a population i will have different characteristics, given by:

$$q_{it+1} = q_{it} + \eta_{it+1} \tag{1}$$

where q_{it} are the characteristics inherited from the previous generations, while η_{it+1} denotes cultural change.

By the same token, the dynamics of productive knowledge includes vertical transmission across generations as well as changes (innovations), that is:

$$A_{it+1} = A_{it} + \Delta_{it+1} \tag{2}$$

where Δ_{it+1} denotes change in productivity due to technological and institutional innovations. Changes may take place because of original discovery by agents that belong to population i and/or because of successful imitation/adaptation of innovations that were discovered elsewhere. The diffusion of technological and institutional innovations can be viewed as a special case of cultural transmission.

We are interested in the long-run process of vertical and horizontal transmission of innovations across populations at different genealogical (i.e., genetic) distances from each other - that is, with

[12] Of course, this is a highly simplified and reductive way of capturing cultural differences. In general, culture is a highly elusive and multi-faceted concept. In a well-known survey over fifty years ago Kroeber and Kluckhohn (1952) listed 164 definitions of culture proposed by historians and social scientists. See also Boyd and Richerson (1985).

[13] A note on semantics is in order: while we call these characteristics "cultural" for illustrative purposes, q_i could be easily reinterpreted to include also genetically transmitted characteristics. The key points are that those characteristics a) must be passed with variation from one generation to the next, and b) must affect the probability of adopting innovations from populations with different characteristics.

different distances from their last common ancestor.[14] To capture these relationships in the simplest possible way, we will assume the following intergenerational structure. At time 0, there exists only one population, with cultural characteristics q_0 (normalized to zero) and productive knowledge A_0.[15] At time 1 the population splits in two distinct populations (population 1 and population 2). At time 2, population 1 splits in two populations (populations 1.1 and 1.2), and population 2 splits in two populations (populations 2.1 and 2.2). This structure provides us with the minimum number of splits we need to have variation in genealogical distances between populations at time 2. We can measure genetic distance between populations by the number of genealogical steps one must take to reach the closest common ancestor population. Let $d(i, j)$ denote the genetic distance between populations i and j. Populations 1.1 and 1.2 have to go back only one step to find their common ancestor (population 1), while populations 1.1 and 2.1 have to go back two steps to find their common ancestor (population 0), as illustrated in Figure 1. Therefore, we have:

$$d(1.1, 1.2) = d(2.1, 2.2) = 1 \tag{3}$$

and:

$$d(1.1, 2.1) = d(1.1, 2.2) = d(1.2, 2.1) = d(1.2, 2.2) = 2 \tag{4}$$

What is the relationship between genealogical distance, cultural change and technological change? In order to explore these issues, it is useful to consider the following benchmark assumptions:

A1. At each time t two populations i and j with $A_{it} = A_{jt}$ face an identical probability π_t of discovering an *original* innovation that would increase productive knowledge by Δ_t.

This assumption means that cultural characteristics q_{it} *per se* do not have a *direct* effect on the rate of technological progress: two populations with different cultural characteristics but identical levels of productive knowledge face identical probabilities of expanding the technological frontier.

[14] By genealogical distance we mean the number of generations that separate two agents (in our case, populations) from their last common ancestor. Two sisters are at a genealogical distance equal to one. Two first cousins are at a genealogical distance equal to two. Geneticists do not express genetic distances in terms of the number of generations back to a common ancestor, but use measures of genetic similarity across populations that are closely correlated with the number of genealogical steps. More details on the definition and construction of genetic distance measures are provided in Section 3.

[15] In this analysis we will abstract from differences in size across populations, and assume that all populations have identical size. For a discussion of the relationship between size and productivity, see Alesina, Spolaore and Wacziarg (2005).

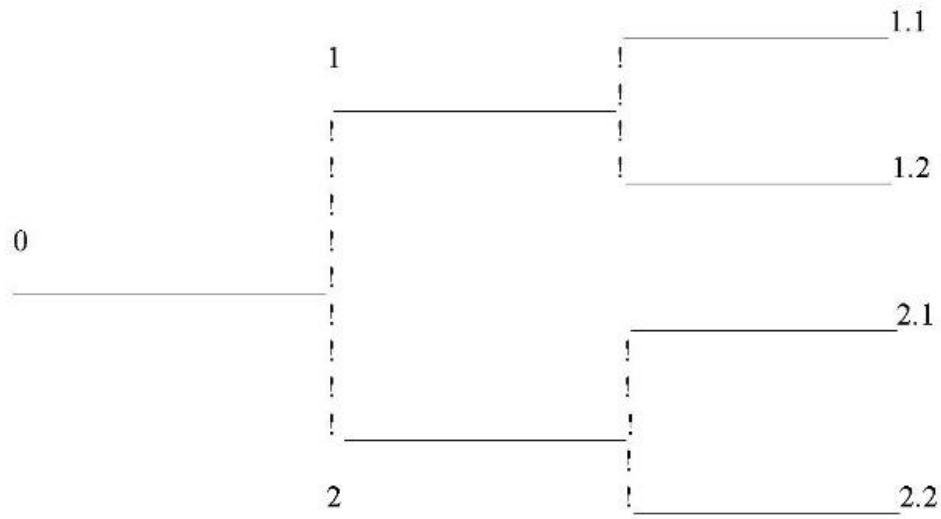

Figure 1: Population Tree.

In other words, cultural characteristics are assumed to be *neutral* with respect to the process of innovation.[16]

By contrast, it is reasonable to assume that the process of *imitating* somebody else's innovation is a function of the cultural distance between the innovator and the imitator. That is, we assume:

A2. If an innovation is introduced by some population i with cultural characteristics q_i, the extent to which a population j, with cultural characteristics q_j, can increase its own technological knowledge through the imitation and adaptation of population i's innovation will depend on the "cultural distance" between the two populations, that is, on $|q_j - q_i|$.

But how do different populations end up with differing cultural characteristics? For the purposes of this analysis, we will consider a simple model of cultural divergence ("mimetic drift"):[17]

A3. Cultural transmission follows a random walk, in which cultural characteristics are transmitted vertically across populations, while "cultural change" is white noise.

[16] We will relax this assumption below.

[17] If we reinterpret the variable q_i as including some genetic characteristics, we can reinterpret the random process as a case of "mimetic plus genetic drift". In this model we refer to q_i as a purely cultural variable to stress that VTCs include cultural features, not just genetic ones.

Clearly, this is a highly stylized approximation of more complex phenomena, but it does provide a simple way to capture the dynamics of changes in neutral cultural characteristics. Specifically, we will assume that for each population i cultural characteristics are given by:

$$q_i = q_{i'} + \eta_i \tag{5}$$

where $q_{i'}$ are the characteristics of the closest ancestor (population 0 for populations 1 and 2, population 1 for populations 1.1 and 1.2, population 2 for populations 2.1 and 2.2), and η_i is equal to $\eta > 0$ with probability 1/2 and $-\eta$ with probability 1/2.[18]

When cultural characteristics follow the above process, we can immediately show that on average cultural distance between two populations is increasing in their genealogical distance. Specifically, in our example, the expected cultural distance between populations at a genealogical distance $d(i,j) = 1$ is:

$$E\{|q_j - q_i| \mid d(i,j) = 1\} = \eta \tag{6}$$

while populations at a genealogical distance $d(i,j) = 2$ have twice the expected cultural distance:

$$E\{|q_j - q_i| \mid d(i,j) = 2\} = 2\eta \tag{7}$$

The above relationships imply that, on average, populations that are closer genealogically will also be closer culturally:

$$E\{|q_j - q_i| \mid d(i,j) = 2\} - E\{|q_j - q_i| \mid d(i,j) = 1\} = \eta > 0 \tag{8}$$

This is not a deterministic relationship: it is possible that two populations that are genealogically distant will end up with *more similar cultures* than two populations which are more closely related. But that outcome is less likely to be observed than the opposite. In summary, we have:

Result 1

On average, greater genealogical (i.e., genetic) distance is associated with greater cultural distance.

[18] We assume this process to keep the algebra as simple as possible, without loss of generality. The key assumption is that η_i must be white noise.

2.1.2 Innovations and Diffusion with Neutral VTCs

We are now ready to study the relationship between diffusion of innovations, cultural change,and genetic distance within our framework.

First of all, consider the case in which inter-population barriers to the horizontal diffusion of innovations are prohibitive. In other terms, consider the case in which there is no horizontal transmission of innovations, but just vertical transmission. To fix ideas, suppose that at time $t = 1$, each of the two existing populations (1 and 2) could independently increase its inherited productivity A_0 by $\Delta > 0$ with probability π. Assuming that no other innovation takes place at time 2, what are the expected differences in income across populations at time 2?

Populations with the same closest ancestor will inherit the same productive knowledge (either A_0 or $A_0 + \Delta$) and will not differ in income per capita. That is:

$$E\{|y_j - y_i| \mid d(i,j) = 1\} = 0 \tag{9}$$

On the other hand, populations 1 and 2 will transfer different technologies to their descendants if and only if one of the two population has successfully innovated at time 1 while the other population has not. This event takes place with probability $2\pi(1-\pi)$. Hence, expected income differences across populations with genealogical distance equal to 2 are given by:

$$E\{|y_j - y_i| \mid d(i,j) = 2\} = 2\pi(1-\pi)\Delta \tag{10}$$

Not surprisingly, when technological innovations diffuses only via vertical transmission, income differences are strongly correlated with genealogical distance:

$$E\{|y_j - y_i| \mid d(i,j) = 2\} - E\{|y_j - y_i| \mid d(i,j) = 1\} = 2\pi(1-\pi)\Delta > 0 \tag{11}$$

The relationship is stronger the higher is the variance of innovations across population (which, in our example, is measured by $\pi(1-\pi)$), and it is highest at $\pi = 1/2$.

By contrast, if there were *no barriers* to the horizontal transmission of innovations across populations, all societies would have the same income per capita independently of their genealogical distance.[19] In general, genealogical distance matters for income differences if and only if there are barriers to horizontal diffusion. Let us consider the case in which barriers are positive but

[19] For all populations at time 1 and 2, we would have $y = A_0 + \Delta$ with probability $1-(1-\pi)^2$ and $y = A_0$ with probability $(1-\pi)^2$.

not prohibitive. That is, if population i innovates and increases its productivity to $A+\Delta$ while population j does not innovate, population j will be able to increase its total factor productivity to:

$$A + \max\{(1 - \beta|q_j - q_i|)\Delta, 0\} \tag{12}$$

where β measures barriers to the diffusion of innovations across populations with different VTCs q_i and q_j.[20] Now, populations 1 and 2 will *not* end up with the same technology if and only if a) only one of the two populations finds the innovation (an event with probability $2\pi(1-\pi)$), and b) the two populations are culturally different - that is, one experienced a cultural change equal to η while the other experienced $-\eta$ (an event with probability 1/2). If both a) and b) hold (an event with probability $\pi(1-\pi)$), one of the two populations will have productivity equal to $A_0 + \Delta$ while the other will have productivity equal to $A_0 + (1 - 2\beta\eta)\Delta$. If no additional diffusion can take place at time 2 (that is, if horizontal transmission is possible only for contemporaneous innovations), we have:

$$E\{|y_j - y_i| \mid d(i,j) = 2\} - E\{|y_j - y_i| \mid d(i,j) = 1\} = 2\pi(1-\pi)\beta\eta\Delta > 0 \tag{13}$$

The above equation shows that *income differences are increasing in genealogical distance if and only if there are positive barriers to diffusion ($\beta \neq 0$) and populations diverge culturally over time ($\eta \neq 0$).*

In the above example we have assumed that horizontal diffusion of the innovation introduced at time 1 takes place only contemporaneously - that is, at time 1. The analysis can be extended to allow for further horizontal transmission at time 2.

Consider the case in which at time t two populations (say, 1.1 and 1.2) have "inherited" technology $A_0 + \Delta$ by vertical transmission while the other two populations (say, 2.1 and 2.2) have inherited $A_0 + (1 - 2\beta\eta)\Delta$. From population 2.1's perspective, the unadopted innovation from period 1 is:

$$[A_0 + \Delta] - [A_0 + (1 - 2\beta\eta)\Delta] = 2\beta\eta\Delta \tag{14}$$

If we consider this situation as equivalent to the case in which populations 1.1 and 1.2 come up with a *new* innovation of size $2\beta\eta\Delta$, we can model the adoption of that innovation by population 2.1 as:

$$\Delta_{2.1} = [1 - \beta|\min\{|q_{2.1} - q_{1.1}|, |q_{2.1} - q_{1.2}|\}]2\eta\beta\Delta \tag{15}$$

[20] In the rest of the analysis, for simplicty and without loss of generality, we assume that $\beta|q_j - q_i| < 1$.

where the expression $\min\{|q_{2.1} - q_{1.1}|, |q_{2.1} - q_{1.2}|\}$ captures the fact that population 2.1 will adopt the innovation from the population that is culturally closer.

In this case, the expected income gap between populations at different genealogical distance is given by:

$$E\{|y_j - y_i| \mid d(i,j) = 2\} - E\{|y_j - y_i| \mid d(i,j) = 1\} = \pi(1-\pi)\beta^2\eta^2\Delta > 0 \tag{16}$$

which, again, implies a positive correlation between differences in income per capita and genealogical distance, as long as $\beta \neq 0$ and $\eta \neq 0$.[21]

We can summarize the above analysis as:

Result 2

Income differences across populations are increasing in genealogical (i.e., genetic) distance if and only if there are positive barriers to the diffusion of innovations ($\beta \neq 0$) and populations diverge culturally over time ($\eta \neq 0$).

2.1.3 Non-Neutral VTCs

The above results have been obtained under the assumption that VTCs are neutral - that is, they bear no direct effect on the production function and on the process of innovation itself, but only on the process of horizontal diffusion of innovations. The assumption can be relaxed by allowing a direct effect of cultural characteristics on the probability of innovating.[22] Specifically, assume that population i's probability of finding an innovation is given by:

$$\pi_i = \pi + \phi q_i$$

This means that a higher q_i is associated with more innovations and a lower q_i with less innovations. The analysis above can be viewed as the special case $\phi = 0$. Under this more general assumption equation (13) becomes:

$$E\{|y_j - y_i||d(i,j) = 2\} - E\{|y_j - y_i| \mid d(i,j) = 1\} = 2[\pi(1-\pi) + \phi^2\eta^2]\beta\eta\Delta \tag{17}$$

[21] An analogous equation can be obtained for innovations that occur in period 2. If the four populations inherit identical technologies from period 2 and each population can find an innovation of size Δ in period 2 with probability π we have $E\{|y_j - y_i||d(i,j) = 2\} - E\{|y_j - y_i||d(i,j) = 1\} = 2\pi^2(1-\pi)^2\beta\eta\Delta > 0$

[22] In the Appendix we will consider extensions in which cultural characteristics may directly affect not only the probability of innovating, but also the level of productivity once the innovation has been adopted.

This equation shows that the larger the direct impact of cultural characteristics on the probability of innovating, the stronger the relationship between expected income differences and genealogical distance, provided there are barriers to diffusion ($\beta \neq 0$) and cultural heterogeneity ($\eta \neq 0$). In other words, a direct effect of cultural characteristics on the innovation process strengthens the relationship between genealogical (i.e., genetic) distance and income gaps, as long as there are barriers to the diffusion of innovations, consistent with Result 2 above.

2.2 VTCs and Income Differences: A General Taxonomy

In the analytical framework presented above, we have illustrated a simple mechanism of development diffusion implying a positive correlation between genealogical distance (i.e. distance from the last common ancestors) and income differences. The central feature of the framework is the link between genealogical distance and the vertical transmission of characteristics across generations. In our model, we showed how differences in neutral characteristics (that is, characteristics that do not have a direct effect on productivity and innovations) can explain income differences by acting as barriers to the diffusion of innovation across populations. We then extended the model to include possible direct effects of VTCs on productivity. Specifically, in our framework we considered a direct effect of different characteristics on the probability of adopting productivity-enhancing innovations. We have seen how direct effects increase the magnitude of the correlation between genetic distance and income differences, but are not necessary for the existence of a positive correlation: barrier effects due to neutral VTCs are sufficient to explain a positive correlation between genealogical distance and income differences.

In our framework we have referred to the transmission of characteristics as cultural - that is, *not* directly related to the transmission of DNA from parents to children. We have done that for two reasons. One reason is conceptual: to provide a model that clearly shows how a direct link from DNA-transmitted characteristics to economic outcomes is not necessary for our results, as long as the vertical transmission of cultural characteristics takes place among genetically-related individuals (typically, parents and children). The second reason is substantial. Our focus is on income differences across different populations of Homo Sapiens Sapiens, taking place over a relatively short period in terms of genetic evolution, and we expect that over that time frame divergence in cultural characteristics have played an important role.

However, in principle the insights from our framework can be generalized to include a broader

set of channels through which characteristics are vertically transmitted. In general, characteristics can be transmitted across generations through DNA (genetic transmission, or GT - e.g. eye color) or through pure cultural interactions (cultural transmission, or CT - e.g., a specific language). Moreover, VTCs, whether transferred through GT or CT, may affect income differences because of a direct (D) effects on productivity or because they constitute barriers (B) to the transmission of innovations across populations. Hence, in general one can identify four possible combinations of mechanisms through which VTCs may affect income differences: a GT direct effect, a GT barrier effect, a CT direct effect, and a CT barrier effect.[23] 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

Taxonomy of the effects of VTCs

For instance, VTCs affecting the trade-off between quality and quantity of children in the theoretical framework proposed by Galor and Moav (2002) to explain the Industrial Revolution would be examples of GT direct effects (Quadrant I). GT barrier effects (Quadrant II) could stem from visible genetically-transmitted characteristics (say, physical appearance) that do not affect productivity directly, but introduce barriers to the diffusion of innovations and technology by reducing exchanges and learning across populations that perceive each other as different.[24] Direct effects of cultural characteristics have been emphasized in a vast sociological literature that goes back at least to Max Weber.[25] 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 (2004). The link between cultural characteristics and barriers (Quadrant IV) is at

[23] It is important to notice that these conceptual types should not be viewed as completely separable, but rather as points on a logical continuum, which may involve a mix of them. For a recent general discussion of the interactions between biological and cultural transmission, see Richerson and Boyd (2004). Recent results in genetics that are consistent with complex gene-culture interactions are provided by Wang et al. (2006).

[24] This effect is related to recent work by Guiso, Sapienzan and Zingales (2004), who argue that differences in physical characteristics may affect the extent of trust across populations. Visible differences across ethnic groups also play an important role in the analysis of ethnic conflict by Caselli and Coleman (2002).

[25] More recent references can be found in the edited volume by Harrison and Huntington (2000).

the core of our basic model, while our extension to non-neutral cultural characteristics may be interpreted as an example from Quadrant III.

It is worth pointing out that the distinction between GT and CT may be useful to fix ideas, but is not a clear-cut dichotomy. In fact, this distinction, essentially that between nature and nurture, may be misleading from an economic perspective, 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 U.S.). Consider a clearly genetic characteristic of an individual, for instance having two X chromosomes. This purely genetic 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.[26] By the same token, one can think of the differential impact of a given cultural 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 interactions in which culture affects genes is the spread of the gene for lactose tolerance in populations that domesticated cows and goats. In the interpretation of our empirical analysis we will not dwell much on the distinction between genetic and cultural transmission of characteristics, but interpret genetic distance as an overall measure of differences in the whole set of VTCs.

On the other hand, we will have more to say empirically on whether our estimates capture a direct versus a barrier effect of VTCs. Our model so far has ignored the role of geographic barriers in the diffusion of development. In Appendix 1, we present a variety of reduced form models, extending our basic framework to include geographic barriers. We focus on the interaction term between geographic barriers and barriers linked to differences in VTCs. We show that a negative effect of this interaction term on income differences arises naturally from models in which VTCs act as barriers. The basic intuition is simple and general: under the barriers interpretation, if populations are far apart geographically, then it should matter less that they are also distant along

[26] This is a variation on an example by Alison Gopnik in her comment to the now-famous 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.

some other dimension for the flow if development-enhancing innovations. In contrast, models where VTCs bear direct effects on income do not predict such a negative interaction term, as it is the presence or absence of traits that determines directly the extent of the income gain, irrespective of geographic distance. We will use these theoretical insights in our empirical work to test whether differences in VTCs act at least partly as barriers to the diffusion of development.

3 Data and Empirical Methodology

3.1 Data

Genetic Distance Since the data on genetic distance that we use as a measure of distance in vertically-transmitted characteristics is not commonly used in the economics literature, it is worth spending some time describing it.[27] Genetic distance measures the genetic similarity of two populations. The basic unit of analysis is the *allele*, or the variant taken by a gene. By sampling populations for specific genes, geneticists have compiled data on allele frequencies, i.e. the proportion of the population with a gene of a specific variant.[28] Differences in allele frequencies are the basis for computing summary measures of distance based on aggregated differences in allele frequencies across various genes (or *loci* on a chromosome). Following Cavalli-Sforza et al. (1994), we will use measures of F_{ST} distance, also known as coancestor coefficients (Reynolds et al., 1983). F_{ST} distances, like most measures of genetic diversity, are based on indices of heterozygosity, the probability that two genes at a given locus, selected at random from the relevant populations, will be different. The construction of F_{ST} distances can be illustrated for the simple case of two populations (a and b) of equal size, one locus, and two alleles (1 and 2). Let p_a and q_a be the gene frequency of allele 1 and allele 2, respectively, in population a.[29] The probability that two randomly selected genes at a given locus are *identical* within the population ("homozygosity") is

[27] As far as we know, this is the first study of the relationship between genetic distance and differences in income per capita across countries. Guiso, Sapienza and Zingales (2004), in a parallel study, use genetic distance between European populations as an instrument for a measure of trust in order to explain bilateral trade flows. This is quite different from our application, as we are interested in explaining income differences, not trade flows. Their results are consistent with our interpretation of genetic distance as related to barriers. An economic application of measures of genetic distance across different species is provided by Weitzman (1992).

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

[29] Therefore we have $p_a + q_a = 1$ and $(p_a + q_a)^2 = p_a^2 + q_a^2 + 2p_a q_a = 1$.

$p_a^2 + q_a^2$, and the probability that they are different ("heterozygosity") is:

$$h_a = 1 - p_a^2 + q_a^2 = 2p_a q_a \tag{18}$$

By the same token, heterozygosity in population b is:

$$h_b = 1 - p_b^2 + q_b^2 = 2p_b q_b \tag{19}$$

where p_b and q_b be the gene frequency 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:

$$\overline{p} = \frac{p_a + p_b}{2} \tag{20}$$

and:

$$\overline{q} = \frac{q_a + q_b}{2} \tag{21}$$

Heterozygosity in the *sum* of the two populations is:

$$h = 1 - \overline{p}^2 + \overline{q}^2 = 2\overline{pq} \tag{22}$$

By contrast, *average* heterozygosity is measured by:

$$h_m = \frac{h_a + h_b}{2} \tag{23}$$

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

$$F_{ST} = 1 - \frac{h_m}{h} = 1 - \frac{p_a q_a + p_b q_b}{2\overline{pq}} \tag{24}$$

If the two populations have identical allele frequencies ($p_a = p_b$), F_{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_{ST} takes value 1. In general, the highest the variation in the allele frequencies across the two populations, the higher is their F_{ST} distance. The formula can be extended to account for L alleles, S populations, different population sizes, and to adjust for sampling bias. The details of these generalizations are provided in Cavalli-Sforza et al. (1994, pp. 26-27).[30]

These measures of genetic distance have been devised mainly to reconstruct phylogenies (or family trees) of human populations. F_{ST} (which is also known as the coancestor coefficient) can be interpreted as the distance to the most recent common ancestors of two populations. Thus, in

[30] For a general discussion of measures of genetic distances, see also Nei (1987).

effect, genetic distance is related to how long two populations have been isolated from each other.[31] If two populations split apart as the result of outmigration, their genes start to change as a result of genetic drift (randomness) and natural selection. When calculating genetic distances in order to study population history and phylogenesis, geneticists concentrate on neutral characteristics that are not affected by strong directional selection occurring only in some populations and environments (Cavalli-Sforza et al., 1994, p. 36).[32] In other words, the term "neutral markers" refers to genes affected only by drift, not natural selection.

If the populations are separated, this process of change will take them in different directions, raising the genetic distance between them. The longer the period for which the separation lasts, the greater will genetic distance become. More specifically, the rate of evolution is the amount of evolutionary change, measured as genetic distance between an ancestor and a descendant, divided by the time in which it occurred. If drift rates are constant, genetic distance can be used as a molecular clock - that is, the time elapsed since the separation of two populations can be measured by the genetic distance between them. Figure 2, from Cavalli-Sforza et al. (1994), illustrates the process through which different human populations have split apart over time. Heuristically, genetic distance between two populations is captured by the horizontal distance separating them from the next *common* node in the tree.

In this paper we will use F_{ST} distance as a measure of "genealogical distance" between populations. Consistent with our theoretical framework, we expect a larger F_{ST} distance to reflect a longer separation between populations, and hence, on average, a larger difference in VTCs. The data itself is from Cavalli-Sforza et al. (1994), p. 75-76: we focus on the set of 42 world populations for which they report all bilateral distances, based on 120 alleles.[33] These populations are aggregated from

[31] Isolation here refers to the bulk of the genetic heritage of a given population. Small amounts of interbreeding between members of different populations do not change the big picture.

[32] 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, a recent reference is Jobling et al. (2004). The fact that genetic distance is calculated with respect to neutral genetic markets implies that genetic distance can provide an especially useful measure for the channels of Quadrants III and IV in Figure 2 (Section 2.3).

[33] Cavalli-Sforza et al. (1994) also provide a different measure of genetic distance (Nei's distance). Nei's distance, like F_{ST}, measures differences in allele frequencies across a set of specific genes between two populations. F_{ST} and Nei's distance have slightly different theoretical properties, but the differences are unimportant in practice as they are very highly correlated, and the choice of measures does not impact our results (as we show below).

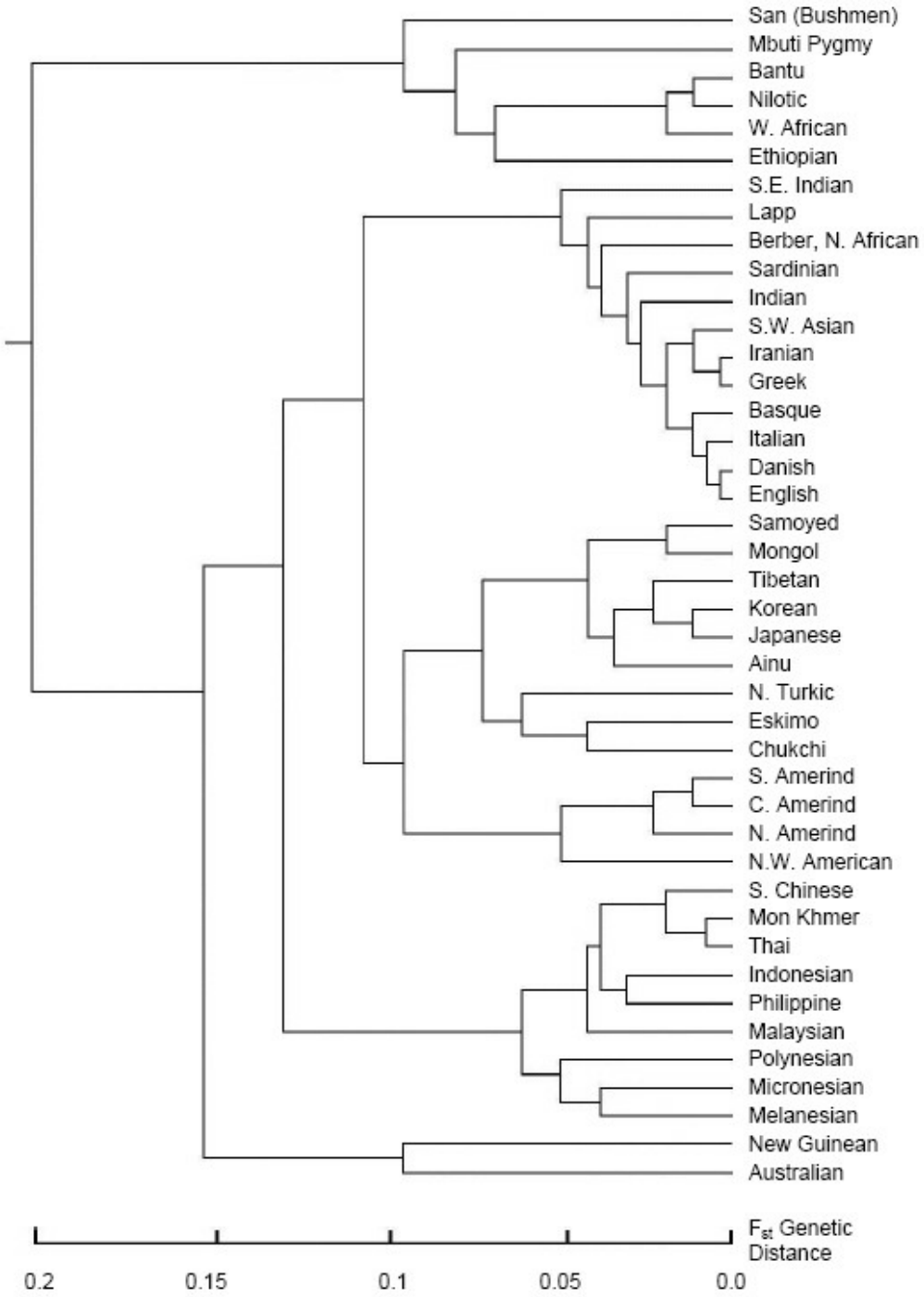

Figure 2: Genetic distance among 42 populations. Source: Cavalli-Sforza et al., 1994.

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.[34] 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, for which the genetic distance is 0.0021.[35] The mean genetic distance among the 861 available pairs is 0.1338.

Genetic distance data is available at the population level, not at the country level. It was thus necessary to match populations to countries. We did so using ethnic composition data from Alesina et al. (2003). In many cases, it was possible to match ethnic group labels with population labels from Cavalli-Sforza et al. (1994). The was supplemented with information from Encyclopedia Britannica when the mapping of populations to countries was not achievable from ethnic group data. Obviously, many countries feature several ethnic groups. We matched populations to the dominant ethnic group, i.e. the one with the largest share of the country's population.[36] Random error in the matching of populations to countries should lead us to understate the correlation between genetic distance and income differences, and we further discuss the issue of measurement error stemming from possible mismatches below.

The ethnic composition in Alesina et al. (2003) refers to the 1990s. This is potentially endogenous with respect to current 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 North America because of a favorable geographical environment.[37] In order to construct genetic distance between countries as of 1500

[34] In our empirical work, we make use of the more detailed data for Europe in order to extend our results.

[35] Among the more disaggregated data for Europe which 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 Lapp and the Sardinians. The mean genetic distance across European populations is 0.013. As can be seen, genetic distances are roughly ten times smaller on average across populations of Europe than in the World dataset. However, we still find that they significantly predict intra-Europe income differences.

[36] We have also computed measures of weighted genetic distance by using the data on each ethnic group within a country. The measure reflects the expected distance between two randomly selected individuals, one from each country in a pair. Weighted genetic distances are very highly correlated with unweighted ones, so as we show below for practical purposes it does not make a big difference which one we use.

[37] In fact, income differences are not very persistent at a long time horizon such as this - see Acemoglu et al. (2002). Our own data shows that pairwise log income differences in 1500 are uncorrelated with the 1995 series in the common

(in an effort to obtain a variable that is more exogenous than current genetic distance), we also mapped populations to countries using their ethnic composition as of 1500, i.e. prior to the major colonizations of modern times. Thus, for instance, while 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 main ethnic group is now of European descent (New Zealand, Australia, North America and some countries in Latin America). 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, possibly reducing measurement error.

Geographic Distance. In addition to genetic distance, we also used several measures of geographic distance. The first is a measure of the greater circle (geodesic) distance between the major cities of the countries in our sample. This comes from a new dataset compiled by researchers at CEPII.[38] This dataset features various measures of distance (between major cities, between capitals, weighted using several distances between several major cities, etc.), all of which bear pairwise correlations that exceed 99%. The dataset includes other useful controls such as whether pairs of countries share the same primary or official language, whether the countries are contiguous, whether they had a common colonizer, etc. We used some of these controls in our regressions.

The second measure of geographic distance that we use is latitudinal distance - i.e. simply the absolute value of the difference in latitude between the two countries in each pair: $G_{ij}^{LA} = |\text{latitude}_i - \text{latitude}_j|$. Latitude could be associated with climactic 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 latitude direction than along the longitude direction, because similar longitudes share the same climate, availability of domesticable animal species, soil conditions, etc. We should therefore expect countries at similar latitudes to also display similar levels of income. Finally, we control for a measure of longitudinal distance, $G_{ij}^{LO} = \left|\text{longitude}_i - \text{longitude}_j\right|$, to capture geographic isolation along this alternative axis.

sample (Table 1).

[38] The data is available free of charge at http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

Summary Statistics Table 1, Panel A displays correlations between our various measures of distance.[39] Perhaps surprisingly, these correlations are not as high as we might have expected. For instance, the correlation between geodesic distance and F_{ST} genetic distance is only 35.4% - though unsurprisingly it rises to 47.8% if genetic distance is measured based on populations as they were in 1500 (the colonization era acted to weaken the link between genetic distance and geographic distance by shuffling populations across the globe). The correlation between alternative measures of genetic distance, on the other hand, tends to be quite high: the F_{ST} measure and the Nei measure of genetic distance bear a correlation of 92.9%, so it should not matter very much which one we use. Finally, our various measures of genetic distance bear positive correlations between 14% and 22% with the absolute value of log income differences in 1995.[40] Together, these correlations suggest it may be possible to identify separately the effects of geographic and cultural barriers on the long-run diffusion of development.

3.2 Empirical Methodology

3.2.1 Specification

Since our empirical methodology is new, we describe it in some detail. We depart from the usual approach, which consists of regressing income levels or income growth on a set of determinants. Instead, we consider a specification in which the absolute *difference* in income between pairs of countries (or other dependent variables such as measures of institutions and human capital) is regressed on measures of distance between the countries in this pair. We computed income differences between all pairs of countries in our sample for which income data was available, i.e. $13,861$ pairs of countries. Thus our baseline specification is:

$$|\log y_i - \log y_j| = \beta_0 + \beta_1 G_{ij}^D + \beta_2 G_{ij}^G + \beta_3 G_{ij}^{LA} + \beta_4 G_{ij}^{LO} + \beta_5' X_{ij} + \varepsilon_{ij} \tag{25}$$

where G_{ij}^D is genetic distance, G_{ij}^G is geodesic distance, G_{ij}^{LA} is latitudinal distance, G_{ij}^{LO} is longitudinal distance, X_{ij} is a set of controls capturing other types of barriers and ε_{ij} is a disturbance

[39] Panel B presents means and standard deviations of the main variables, allowing us to assess the quantitative magnitudes of the effects estimated in the regressions that follow.

[40] The data on per capita income is purchasing power-parity adjusted data from the World Bank, for the year 1995. We also used data from the Penn World Tables version 6.1 (Summers, Heston 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.

term to be further discussed below.[41]

The reason our empirical specification must involve income differences rather than a single country's income level on the left hand side is that this makes the use of bilateral measures of barriers possible. There is no other way to quantify the impact of barriers, which are inherently of a bilateral nature, on income differences. Conceptually, therefore, we depart in a major way from existing methodologies: our regression is not directional, in the sense that the right-hand side variable takes on the same value for each country in the pair, i.e. our specification is not simply obtained by differencing levels regressions across pairs of countries.[42]

We should stress that equation (25) is a reduced form. That is, 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. Whether income differences are caused by these factors is the subject of a vast literature but is not primarily the subject of this paper. This paper is concerned with *barriers* to the diffusion of these more proximate causes of income differences: barriers work to explain differences in income presumably because they affect the adoption of technologies, norms for human and physical capital accumulation, the adoption of institutions conducive to differential economic performance, etc.. In Section 5 we will relate barriers to differences in human capital, institutions, investment rates, population growth rates and openness. Evaluating the role of genetic distance in affecting income differences through these various channels, however, does not form the core of our argument.

3.2.2 Estimation

We estimate equation (25) using a new methodology. In principle, if one is willing to assume that the measures of barriers are exogenous, equation (25) can be estimated using least squares. However, in this case usual methods of inference will be problematic. 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

[41] 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.

[42] This, obviously, would result in adding no new information relative to the levels regressions themselves. Our methodology is more akin to gravity regressions in the empirical trade literature than to levels or growth regressions in the literature on comparative development.

errors ε_{12} and ε_{13}, but we are unwilling to assume that observations on the dependent variable are independent conditional on the regressors.[43] In other words, simple least squares standard errors will lead to misleading inferences due to spatial correlation.

Before proceeding, we note the following observations and conventions: 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...N appearing in position j. Then country 2 is in position i and is matched with 3...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 non-zero off-diagonal elements of the residual covariance matrix by σ_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 σ_m, the covariance matrix of the vector of residuals ε is of the form:

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

In this context, controlling for a common-country fixed effect should account for the correlated part of the error term. For this we rely on well-known results cited in Case (1991), showing that fixed effects soak up spatial correlation, though in a context quite different from ours: we do not have longitudinal data, and the panel nature of our dataset comes from the fact that each country is

[43] Another feature that reduces the dependence across pairs is the fact that the dependent variable involves the *absolute value* of log income differences. Simple simulations show that under i.i.d. Normal income draws with moments equal to those observed in our sample (Table 1b), the correlation between absolute value differences in income for any two pair containing the same country will be about 0.22. Without taking absolute values, it is straightforward to notice that the correlation would be exactly 0.5.

paired with all the other countries in the dataset.[44] Following this observation, we model:

$$\varepsilon_{ij} = \sum_{k=1}^{N} \gamma_k \delta_k + \nu_{ij} \tag{26}$$

where $\delta_k = 1$ if $k = i$ or $k = j$, $\delta_k = 0$ otherwise, and ν_{ij} is a well-behaved disturbance term. We treat δ_k as fixed effects, i.e. we introduce in the regression a set of N dummy variables δ_k each taking on a value of one $N - 1$ times.[45] Given our estimator, the effect of the right-hand side variables is identified off the variation within each country, across the countries with which it is matched.

It is in principle possible to test for the presence of spatial correlation both in the model with common country fixed effect and without. Such tests, known as Moran's I-tests, require specifying a neighborhood matrix along which non-zero correlations are allowed. In our case, the neighborhood matrix is easy to conceptualize: its entries are 1 whenever there is a common country in a pair, zero otherwise. Unfortunately, the dimensionality of the neighborhood matrix the square of the number of observations. In the case of our worldwide dataset, there would be $192,127,321$ entries, making the problem computationally intractable. We have run I-tests for our smaller dataset of European countries, where the matrix has $87,616$ entries.[46] Without dummies, Moran's I test suggested the presence of spatial correlation. With common country dummies added to the specification, it did not. We find this to be encouraging.

To summarize, for each country we create a dummy variable equal to 1 if the country appears in a given pair. We then include the full set of N dummy variables in the regression. The inclusion of these fixed effects soaks up the spatial correlation in the error term resulting from the presence of each country multiple times in various country pairs. In addition, our standard errors

[44] Note that simply treating ε_i and ε_j as fixed effects, by including corresponding dummy variables in the regression, will not fully address our concern. This is because, with the exception of country 1 and country N, all countries will appear either in position i or in position j in different observations, inducing spatial correlation between these pairs. In the example above, for instance, country 2 appears in position i in observation 1, and in position j in observation 4, inducing spatial dependence between ε_{12} and ε_{23}. Simple country fixed effects would not soak up this dependence.

[45] The inclusion of fixed effects did not greatly alter the signs or magnitudes of the estimates of the slope coefficients on our variables of interest, compared to simple OLS estimation. In contrast, in line with our expectations, our common-country fixed effects technique resulted in standard errors that were quite different from (and generally much larger than) the (wrong) ones obtained with simple OLS.

[46] The specification being tested is the baseline specification of Column 6 in Table 6. Details of these tests are available upon request.

are heteroskedasticity-consistent (i.e. we correct standard errors to account for the fact that the diagonal elements of Ω might differ).[47]

4 Empirical Results

4.1 Baseline Results

The baseline estimates of equation (25) are presented in Table 2. Columns (1) through (5) feature our measures of distance entered one by one. The results are in line with expectations: greater distance, whether genetic, latitudinal, longitudinal or geodesic, is significantly associated with greater income differences. Univariate results using F_{ST} genetic distance (column 4) suggest that a one standard deviation increase in genetic distance is associated with a 0.198 increase in log income differences - 21.77% of this variable's standard deviation.[48] Columns (5) and (6) shows that it matters little whether we use Nei genetic distance rather than F_{ST} genetic distance. In fact, the impact of a standard deviation difference in Nei genetic distance is slightly larger than that of F_{ST} distance. We will focus on F_{ST} genetic distance for the remainder of this paper, since as we discussed it has a clear interpretation as genealogical distance.

Column (6) and (7) enters all three measures of distance together, for the Nei and F_{ST} distance measures respectively. Interesting results emerge. First, the magnitude of the coefficient on geodesic distance falls by over one third, suggesting that it was capturing at least in part the effect of genetic distance. Second, the coefficient on latitudinal distance also becomes smaller in magnitude. Hence, including genetic distance greatly reduces the effect of geographic barriers. Once controlling for

[47] There are obviously several ways to address the issue of spatial correlation in our context. An alternative we considered would be to do feasible GLS by explicitly estimating the elements of Ω, and introducing the estimated Ω as a weighing matrix in the second stage of the GLS procedure. While apparently straightforward, this is computationally demanding as the dimensionality of Ω is large - in our application we have over $13,000$ country pairs with available data on the variables of interest, and up to 167 covariance terms to estimate - so we leave this for future research. We also pursued several bootstrapping strategies based on selecting subsamples such that the problem of spatial correlation would not occur, generating results very similar to those we present here. Details and results from these bootstraps are available upon request. In contrast to these alternatives, our approach is computationally easy to implement.

[48] This percentage is the commonly used *standardized beta coefficient*, which we will use throughout as a convenient measure of the economic magnitudes of our estimates. The standardized beta coefficient is reported in the last line of each table.

genetic distance, the effect of a one standard deviation change in latitudinal distance on income differences is almost halved, with a standardized beta of 6.64%: the income differences observed unconditionally across latitudes (column 2) are not primarily geographical, but linked to differences in vertically transmitted characteristics captured by genetic distance. Finally, longitudinal distance bears a coefficient that is now negative and economically small.

Column (8) introduces a set of control variables that might proxy for different sets of barriers, linguistic, historical and geographic (the X_{ij} variables in equation (25)). Some of these variables, such as the one reflecting linguistic similarity and colonial history, are perhaps less exogenous with respect to income differences than the distance measures already considered, so results should be interpreted cautiously. Several lessons emerge. First, the signs of the coefficients are largely as expected. Contiguous countries tend to have more similar income levels, as do countries that were ever a single political entity. Countries that have had a common colonizer share more similar income levels, though the effect is small in magnitude. Unsurprisingly, income differences are greater when countries were ever in a colonizer-colonized relationship. On the other hand, having substantial fractions of the populations in each country speaking the same language bears a very small effect on income proximity, and the sign is the opposite of that expected (we will return to the issue of linguistic distance below). Second, these coefficients are estimated quite precisely. Third, and perhaps most importantly, the inclusion of these controls does not modify the estimates on our main variables of interest: the distance measures, in particular genetic distance, hardly change at all. Across specifications, variation in genetic distance accounts for about 20% of the variation in income differences.

Together, these results provide evidence consistent with the model of cultural barriers introduced in Section 2: genetic distance enters with a positive, statistically significant and economically meaningful coefficient in all specifications. Moreover, we find some weak empirical support for the idea that income differences are less pronounced along similar latitudes than across latitudes, once we control for genetic distance.

4.2 Extensions and Robustness

We now consider a number of extensions and robustness tests on the baseline specification in column (8) of Table 3. We enter additional controls, make use of alternative genetic distance data, and examine competing or complementary hypotheses on the diffusion of development.

The Interaction between Genetic and Geographic Distances. In column (1) of Table 3, we assess whether the effects of genetic distance might depend on geodesic distance, by adding the interaction between the two variables to the baseline specification. We find strong evidence of a negative interaction effect: genetic distance matters less for income differences when countries are far apart geographically, and the effect is statistically significant. As discussed in Section 2.2 and Appendix 1, we can interpret this result as evidence in favor of the barrier interpretation for genetic distance: the effect of genetic distance falls as physical distance increases, so these two types of barriers act at least in part as substitutes for each other. There would be no reason to observe this pattern of coefficients if genetic distance mattered *only* as the result of a genetic trait bearing a direct effect on income levels. The magnitude of the overall effect of genetic distance, evaluated at the mean of geodesic distance, is very close to the magnitudes estimated in the baseline specification (with a standardized beta coefficient of 17.96% versus 18.23%).

Weighted Genetic Distance. As we described above, in our baseline results we used data on genetic distance between the plurality populations of each country. Some countries, such as the United States or Australia, are made up of sub-populations that are distinct, and for which genetic distance data is available in the worldwide data on genetic distances between 42 populations (for example Australia is composed of Aboriginals and populations of English descent). To improve upon the measurement of genetic distance between countries, we made use of a measure of weighted genetic distance. The measure is computed as follows: Assume that country 1 contains populations $i = 1...I$ and country 2 contains populations $j = 1....J$, denote by s_{1i} the share of population i in country 1 (similarly for country 2) and d_{ij} the genetic distance between populations i and j. The weighted F_{ST} genetic distance between countries 1 and 2 is simply:

$$WF_{ST} = \sum_{i=1...I} \sum_{j=1...J} s_{1i} \times s_{2j} \times d_{ij} \tag{27}$$

The interpretation of this measure is straightforward: it represents the expected genetic distance between two randomly selected individuals, one from each country. Since there are few countries composed of more than one populations for which there is data available in the worldwide dataset of 42 populations, and since variation induced by the plurality groups tend to dominate by construction, the correlation between weighted genetic distance and the genetic distance measure based on plurality matching is very high - 93.4%. Unsurprisingly, therefore, in column (2) of Table 3 the magnitude of the effect of WF_{ST} is only slightly larger than in our baseline specification, possibly

reflecting a lesser incidence of measurement error using weighted distance. In what follows we continue to focus on the measure based on plurality match.

Possible Endogeneity of Genetic Distance. Next, we attempt to control for the possible endogeneity of genetic distance with respect to income differences. While differences in (neutral) allele frequencies between the populations of two countries obviously 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. Consider for instance 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 geographic factors were not properly accounted for in the regressions through included control variables, then genetic distance today could be positively related to income distance not because cultural distance precluded the diffusion of development, but because similar populations settled in regions prone to generating similar incomes.

To assess this possibility, column (3) of Table 3 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 problem identified above is likely to be most acute. The effect of genetic distance is now actually marginally *larger* than in column (8) of Table 2. The difference in latitudes becomes three times larger, an observation to which we shall return below. Next, we use our data on F_{ST} genetic distance as of 1500 as an instrument for current genetic distance in column (4) of Table 3. This variable reflects genetic distance between populations as they were before the great migrations of the modern era, and yet is highly correlated (65.8%) with current genetic distance, so it fulfills the conditions of a valid instrument. Again, the magnitude of the genetic distance effect is raised - in fact it is more than *doubled* in magnitude when 15th century genetic distance is used as an instrument for current genetic distance: a standard deviation change in genetic distance now accounts for 39.26% of a standard deviation in absolute log income differences. As is usual in this type of application, the larger estimated effect may come from a lower incidence of measurement error under IV - the matching of populations to countries is much more straightforward for the 1500 match. The results suggest that, if anything, our baseline results were understating the effect of genetic distance on income differences.

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 pre-industrial era, an effect that could have persisted to this day. Our estimates of the effect of latitudinal distance provided some evidence that this effect was still at play: in our regressions we have found evidence that differences in latitudes help explain some of the 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) 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.[49]

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").[50] In order to test Diamond's hypothesis, we added the Diamond gap to regressions explaining income differences in 1995 (column 5) and in 1500 (column 6). For the former regression, we restrict our sample to the Old World. 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 non-aboriginal populations who are there because of guns, germs and steel, i.e. thanks to their ancestors' Eurasian advantage. 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 come close to eliminating) the effect of genetic distance. In column (6), using 1500 income differences as a dependent variable, the Diamond gap is again significant and large in magnitude, despite the paucity of observations. This provides suggestive quantitative evidence in favor of Diamond's observation that the diffusion of development was faster in Eurasia. We also

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

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

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, so Diamond's hypothesis on the long-term diffusion of development is complementary to ours.

Climatic Factors. Our remaining robustness tests are concerned with attempts to better control for geographic factors. Since genetic distance is correlated with geographic distance, if we do not suitably control for geographic factors we might wrongly attribute to differences in VTCs part of the income difference that is really due to geography. We start by controlling explicitly for differences in climate between countries. Latitudinal difference may partly but not perfectly capture these climatic differences. We include as an additional control (column 7) a measure of climatic similarity based on 12 Koeppen-Geiger climate zones.[51] Our 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 control for the absolute difference in the percentage of land areas in tropical climates, also based on Koeppen-Geiger zones (column 8). In both cases, we find that the effect of genetic distance is raised slightly in magnitude. As expected, climatic differences bear a positive relationship with income differences, and, also as expected, the inclusion of the climatic difference variable reduces the magnitude of the effect of latitudinal differences.

Isolation controls. If our various measures of distance and geographic contiguity do not adequately capture the extent to which countries may be geographically isolated from each other, we may wrongly attribute to genetic distance the effect of geographic isolation. To address this possibility, we tried a variety of measures of geographic isolation in addition to those already included as controls in our baseline specification. In Table 4, column (1), we added a dummy variable taking on a value of 1 if either country in each pair is an island, and a dummy similarly defined for landlocked countries. Both, as expected, bear large and positive coefficients, consistent with the idea that pairs where at least one country is an island or landlocked will be more isolated from each other,

[51] The 12 Koeppen-Geiger climate zones are: tropical rainforest 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, is available at http://www.ciesin.columbia.edu/eidata/.

resulting on average in greater income differences. However, the inclusion of these variables does not affect the coefficient on our variable of interest, genetic distance.

Continent Effects. The largest genetic distances observed in our worldwide dataset occur between populations that live on different continents. One concern is that genetic distance may simply be picking up the effect of cross-continental barriers to the diffusion of development, i.e. continent effects. If this were the case, it would still leave open the question of how to interpret economically these continent effects, but to test explicitly for this possibility, we added to our baseline specification a dummy that takes on a value of 1 if the two countries in a pair are on the same continent (column 2 of Table 4) and a similarly defined set of dummies for six specific continents (column 3). The results are as expected - with the exception of Asia and Oceania, living on the same continent implies more similar income levels. The inclusion of continent dummies reduces but does not eliminate the effect of genetic distance, which remains statistically significant. We will provide further evidence on the within-continent effects of genetic distance using our European dataset.

Cultural distance. As we have emphasized, cultural factors such as language, religion, norms and values are all part of the set of VTCs captured by genetic distance. In principle, genetic distance will reflect differences in the whole set of VTCs. Are there specific cultural traits that genetic distance may capture, and that could be directly measurable? How much of the estimated effect of genetic distance is attributable to differences in specific measurable VTCs? Columns (4) through (6) of Table 4 include measures of linguistic distance and religious differences in an attempt to answer this question.[52]

We use data on linguistic trees from Fearon (2003). Linguistic trees graphically display the degree of relatedness of world languages. By counting the nodes separating language pairs in linguistic trees, it is possible to construct discrete measures of distance between languages: in Fearon's data up to 15 nodes may separate languages (our variables are re-scaled from 0 to 1, with 1 indicating high linguistic distance). Using data on the linguistic composition of countries, and matching languages to countries, we can then construct indices of linguistic distance between

[52] In our baseline specification we already included a measure of linguistic overlap - a dummy equal to 1 if at least 9% of the population in each country speaks the same language. This is a very crude measure of linguistic similarity, and here we seek to improve upon it.

countries. We did so, as for genetic distance, in two ways: first, we computed a measure of distance between plurality languages for each country pair. Second, we computed a measure of weighted linguistic distance, representing the expected linguistic distance between two randomly chosen individuals, one from each country in a pair (the formula is the same as that of equation 27). This should improve the measurement of linguistic distance for countries where different languages as spoken by different subpopulations. In columns (4) and (5) of Table 4, we include these two measures of linguistic distance in turn. The effects are statistically significant and of the expected sign, but they are small in magnitude: the standardized beta on linguistic distance is equal to 3.62% in column (4) and to 5.15% in column (5). Hence, our measures of linguistic distance can only account for a small fraction of the variation in income differences. Moreover, their inclusion does not affect the coefficient on genetic distance.[53]

Next, in column (6) we included an index of religious difference. The index is the average absolute value difference between the countries in a pair in the population shares of 11 major world religions, obtained from Barrett (2001).[54] Again, the variable takes on the expected positive sign. Its standardized beta is again small, at 7.66%. Moreover, its inclusion barely affects the coefficient on genetic distance compared to a regression obtained using the same sample (the estimated coefficient falls from 2.077 to 2.032).

These results have several interpretations. The most straightforward interpretation is that differences in VTCs captured by genetic distance are not primarily linguistic and religious in nature. This opens up the question of what they are, an important question for future research. Another interpretation is that the specific measures of linguistic and religious distance that we use are inadequate. Consider linguistic distance. We use a discrete measure based on counting common nodes from a linguistic tree. It is a well-known fact that linguistic trees and genetic trees are very similar, as demonstrated in Cavalli-Sforza et al. (1994, p. 98-105). The reason is straightforward: genes, like languages, are transmitted intergenerationally, and this insight is the basis for our interpretation of genetic distance as capturing the full set of VTCs, both genetic and cultural. Yet

[53] The coefficient on genetic distance is raised compared to column (8) of Table 2, but this is due to the slightly smaller sample of available country pairs when we include linguistic distance, not to the inclusion of the latter variable. The coefficient on genetic distance is actually slightly smaller when we control for linguistic distance when the samples are kept the same. This is what we would expect if genetic distance captured in part the effect of linguistic distance.

[54] These religions are Roman Catholics, Protestants, Orthodox Christians, Other Christians, Muslims, Hindus, Buddhists, Other East Asian Religions, Jews, Other Religions, Non-Religious Population.

the correlation between our measure of genetic distance and linguistic distance is only about 25%. This may be partly due to the fact that genetic distance is a *continuous* measure of distance, whereas the number of nodes is a discrete measure of linguistic distance: populations may share few common nodes but linguistic splits occurred recently, in which case one is overestimating distance, or they may share lots of common nodes but the last split occurred a long time ago, in which case one is underestimating distance. Better measures of linguistic distance based on lexicostatistical methods are not widely available outside of Indo-European groups (see Dyen, Kruskal and Black, 1992). Similarly, our measure of religious difference does not take into account the distance separating different religions but only the similarity in religious group shares. For instance, Protestants may be "closer" to Catholics than to Buddhists, a feature that is not be captured by our measure.

Whatever the interpretation, our results show that genetic distance is an overall measure of differences in VTCs that is robustly correlated with income differences, and we were not able to identify specific measures of cultural distance that could alter this conclusion.

4.3 Alternative Time Periods and Samples

Historical Income Data. Table 5 examines whether the pattern of coefficients uncovered for income differences in 1995 held for earlier periods in history. We used income per capita data since 1500 from Maddison (2003), and repeated our basic reduced form regression for 1500, 1700, 1820, 1870, 1913 and 1960.[55] For the 1500 and 1700 regressions, we use the early match for genetic distance, i.e. genetic distance between populations as they were in 1492, prior to the discovery of the Americas and the great migrations of modern times.[56] For the subsequent periods we use the current match. Table 5 shows that across periods, the coefficient on genetic distance is statistically significant and positive. Moreover, the magnitudes are much larger than for the current period: considering regressions obtained from a common sample of 26 countries (275 pairs) for which data is

[55] The data on income for 1960 is from the Penn World Tables version 6.1. For comparison, column (7) reproduces the results for 1995.

[56] 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, United Kingdom, United States. There were 275 pairs (26 countries) with available data for 1500 income, and 328 pairs (29 countries) for 1700. A noteworthy feature of this sample is that it contains no countries in Sub-Saharan Africa.

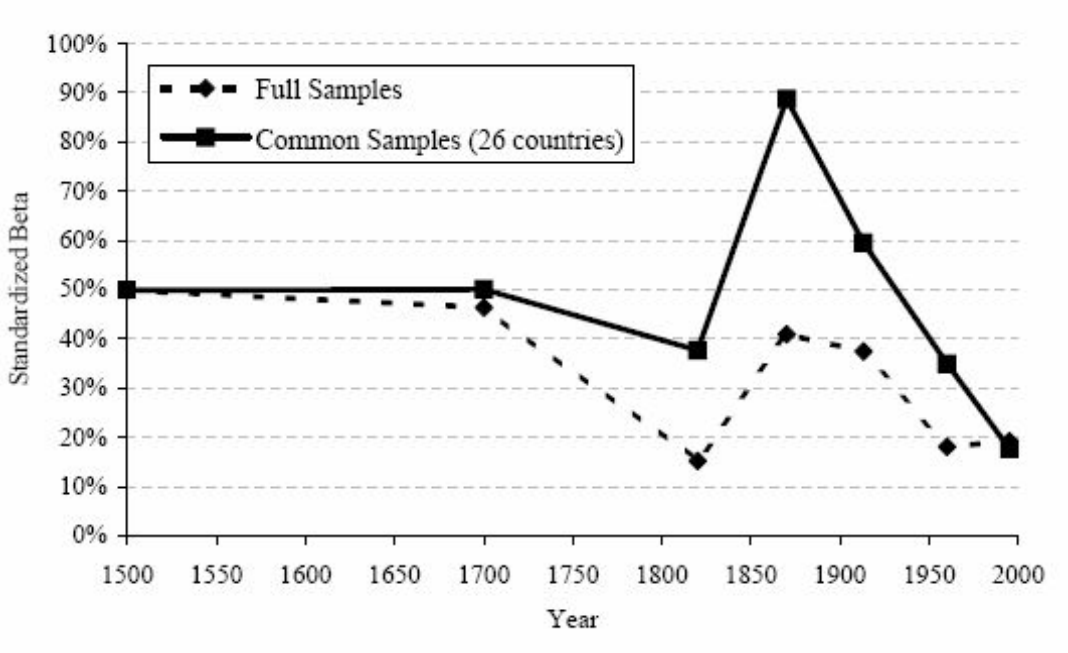

Figure 3: Time Path of the Effect of Genetic Distance, 1500-1995.

continuously available, standardized beta coefficients range from 34.94% (1960) to 88.75% (1870).[57] 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 (the correlation between income differences in 1500 and in 1995 is -0.051 for the 325 country pairs for which data is available). This fact is consistent with our interpretation of genetic distance as a barrier to the diffusion of innovations across populations.

Strong evidence for our model of barriers to diffusion can be obtained from plotting the time path of coefficients, as is done in Figure 3. The result is striking. The slope coefficient of genetic distance decreases gradually from 1500 to 1820, then spikes up in 1870 during the Industrial Revolution and declines thereafter. This is fully consistent with our model where a major innovation (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 declines.

Overall, the historical permanence of the effect of genetic distance on income disparities over the

[57] Similar orders of magnitudes are obtained from unrestricted samples though magnitudes were generally lower.

past five centuries and the specific time path of the effect in relation to the timing of the Industrial Revolution paint an empirical picture that is very consistent with the model of Section 2.

Genetic Distance across European Countries. Cavalli-Sforza et al. (1994) provide data on genetic distances within certain regions of the world, in particular Europe. These data are more disaggregated (i.e. cover more distinct populations) than the matrix of distances for 42 worldwide populations. For Europe, they present a distance matrix for 26 populations that can be matched readily to 25 European countries.[58] Analyzing these 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 European 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 within Europe have developed over the last couple of 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 in favor of the cultural interpretation rather than a genetic effect.

Table 6 presents the results. Much to our surprise, genetic distance is again positively and significantly associated with income differences. Moreover, while genetic distance across European countries are smaller than in the World sample, so are the income differences to be explained. Thus, we find that the effect of F_{ST} distance is large in magnitude. The baseline estimate in column (6) of Table 5, which includes several controls (we now exclude the colonial variables for obvious reasons) suggests that a one standard deviation change in the log of genetic distance accounts for 49.18% of the variation in log income differences. Thus, cultural barriers captured by genetic distance seem very strongly associated with income differences. Physical distance measures correspondingly bear small or insignificant coefficients, suggesting that geographic barriers are not a big hindrance to the diffusion of income across countries of Europe. The last column of Table 6 shows that genetic distance accounts for an even greater amount (79.59%) of the variation in income differences in

[58] These countries are Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Macedonia, Netherlands, Norway, Poland, Portugal, Russian Federation, Slovak Republic, Slovenia, Spain, Sweden, Switzerland and the United Kingdom.

1870, during the Industrial Revolution. Again, we take these differences in estimated magnitudes as consistent with the barrier interpretation.

4.4 Effects on the Proximate Determinants of Income Levels

Our approach to quantifying the barriers to the diffusion of development has been of a resolutely reduced-form nature. That is, we did not specify what economic factors make incomes similar or different, and have instead focused on the effects of geographic and cultural barriers on income differences directly. However, differences in income result from more proximate causes. Several prime candidates have been offered to explain differences in income per capita. These factors are summarized in the model of Section 2 by parameter A, and we now discuss them in greater detail.

In the tradition of the Solow model, steady-state income per capita is positively affected by *rates of factor accumulation* (in physical and human capital), and negatively affected by the depreciation of capital per worker, which is more rapid when *population growth* is faster. The level of total factor productivity, in growth accounting or income accounting exercises, has been found to account for much of the variation in growth and income levels.[59] What causes differences in the levels of total factor productivity, however, is largely unknown. On a general level, TFP is "technology", though the deeper determinants of the adoption of better technologies are generally left unspecified. A recent literature has stressed the importance of *institutions* as a determinant of productivity (the seminal contribution here is Acemoglu, Johnson and Robinson, 2001), and Glaeser et al. (2004) have recently reemphasized the importance of *human capital accumulation*, rather than institutions, as a central determinant of income levels. Finally, a large literature emphasized the role of *market size and openness* as a driver of growth and income levels. We consider these five proximate causes of income levels and examine the effects of our measures of distance on pairwise differences in the rate of physical capital accumulation, the rate of population growth, institutional quality, the stock of human capital and openness. These are meant to reflect the prime candidate explanations for

[59] For an excellent survey on this point, see Caselli (2005).

differences in income levels. Our empirical model can be summarized as:

$$\left.\begin{array}{c}\text{Genetic distance}\\ \text{Geodesic distance}\\ \text{Latitudinal distance}\\ \text{Contiguity, colonial past...}\end{array}\right\} \Rightarrow \left.\begin{array}{c}\text{Institutional quality differences}\\ \text{Human capital differences}\\ \text{Investment differences}\\ \text{Population growth differences}\\ \text{Openness differences}\end{array}\right\} \Rightarrow \text{Income Differences}$$

Our paper is primarily concerned with the first arrow in this diagram, and there are voluminous literatures and unresolved debates on the respective roles of the proximate determinants of income. Thus, we do not attempt to decompose income differences into differences into the underlying proximate cause, a task that is both a tall order and beyond the scope of this paper. We focus instead on investigating empirically the role of our distance measures as determinants of differences in the proximate determinants of income.

Institutions. Engerman and Sokoloff (1997) and Acemoglu, Johnson and Robinson (2001) suggest that the prime cause of economic development is the quality of a country's institutions as the outcome of a long-term historical process. If this is the case, countries that are distant in terms of institutional quality should also be distant in terms of income per capita. We use a commonly-used measure of institutional quality, the risk of expropriation variable (this variable, which ranges from 0 to 10, was used for instance in Acemoglu et al., 2001). Calculating the absolute value of the pairwise difference in the risk of expropriation for 1990, we regressed this variable on our various measures of geographic and cultural distance. Table 7, column (1) presents the results. The main measures of distance are significant when entered individually, but the log of geodesic distance becomes small in magnitude when genetic distance and latitudinal distance are entered alongside it.[60] This is interesting, as latitudinal distance had a small effect in the income differences regression. Our finding that latitudinal distance matters for institutional differences is consistent with the view that geographic and climatic factors have historically constituted hindrances to the diffusion of institutions conducive to higher incomes. Our estimates suggest that a one standard deviation change in genetic distance accounts for 17.99% of the variation in the difference in expropriation risk.

[60] Results for the specifications where distance measures are entered individually are available upon request.

Human Capital. In a recent paper taking issue with the literature on the primacy of institutions, Glaeser et al. (2004) suggest that variation in human capital is the fundamental cause of income differences across countries. To examine the role of geographic and cultural barriers in preventing countries from adopting high levels of human capital, we reran the specification of equation (25), replacing the left hand side variable with the absolute difference in the stock of human capital, measured by the average number of years of primary, secondary and tertiary schooling in the population aged 25 and above in 1990.[61] The results are presented in Table 7, column 2. The pattern of coefficients is similar to that obtained for institutions. A one standard deviation change in genetic distance accounts for 23.01% of a standard deviation in human capital differences.

Population Growth. In column (3) of Table 7, we examine the determinants of absolute differences in rates of population growth. In neoclassical growth models, rapid population growth reduces the steady-state level of income per worker. Thus, differences in population growth (in turn resulting mainly from differences in mortality and fertility) are thought to be associated with differences in income. Again, measures of cultural and geographic distance may affect how differently countries' populations grow.[62] We particularly expect geographic distance measures to be correlated with differences in population growth, as countries located closer to the equator tend to have higher rates of population growth. Indeed, Table 7, column (3) shows that latitudinal distance is positively related to population growth differences (where population growth is defined over the 1960-1990 period): the standardized beta for latitudinal distance is equal to 15%. Genetic distance again appears to be significantly related to the dependent variable, with a standardized beta of 12.88%. One interpretation of this finding is that differences in VTCs are associated with persistent differences in norms of behavior affecting population growth, particularly fertility which dominates the cross-country variation in population growth rates.[63]

Physical Capital Investment. The rate of investment in physical capital is also a determinant of steady-state income levels in the neoclassical model. How do the geographic and cultural barriers

[61] The human capital data is from Barro and Lee (2000). Again, this is a commonly used, if imperfect, measure of the stock of human capital.

[62] For a theoretical analysis of the relationships among VTCs, fertility and economic growth in the long run see Galor and Moav (2002).

[63] A regression relating differences in fertility rates to our distance measures reveals a very significant and large effect of genetic distance on fertility differences. These results are available upon request.

to the diffusion of development relate to differences in this proximate cause of income levels? Table 7, column (4) provides the corresponding estimates, using the Penn World Tables version 6.1 series on the investment share of GDP for 1990. The dependent variable is the absolute difference in these shares across country pairs. Distance measures bear positive signs when they are entered separately as well as jointly. Genetic distance is positively related to differences in investment rates, and its magnitude is economically significant (the standardized beta is equal to 12.52%). We can interpret this finding as indicating that differences in VTCs hinder the adoption of norms of investment behavior that are possibly conducive to superior economic outcomes.

Extent of the Market. Finally, there exists a very large literature on the causal links between international openness, the extent of the market, and economic development. Recent contributions that have emphasized the extent of the market as an important determinant of economic performance include Ades and Glaeser (1999), Alesina, Spolaore and Wacziarg (2000, 2005), Spolaore and Wacziarg (2005). We examined whether differences in market size are related to our various measures of distance. Our measure of market size is the conventionally used ratio of imports plus exports to GDP (in 1990), which proxies for access to world markets. Table 7, column (5) displays the results when using absolute differences in this openness variable as the dependent variable. Differences in openness seem unrelated to most of the variables included in our model, with the exception of the post-war common colonizer variable (this could capture an Africa effect) and the common language variable. Genetic distance appears unrelated to differences in trade openness.

To summarize, genetic distance is significantly positively associated with differences in four of the five proximate determinants of development that we considered, and these effects are generally large in magnitude. This confirms the results found using income differences directly, and provides suggestive evidence that genetic distance may act through several channels.

5 Conclusion

In this paper we have documented the following facts: First, differences in income per capita across countries are positively correlated with measures of genetic distance between populations. Second, genetic distance, an overall measure of differences in vertically transmitted characteristics across populations, bears an effect on income differences even when a large set of geographical and other variables are controlled for. Third, the patterns of relationships between income differences

and measures of genetic and geographical distances hold not only for current worldwide data but also for estimates of income per capita and genetic distance since 1500, as well as in a sample of European countries. Finally, similar patterns hold when the dependent variable is differences in human capital, institutional quality, population growth and investment rates.

These results strongly suggest that characteristics transmitted from parents to children over long historical spans play a key role in the process of development. In particular, the results are consistent with the view that the diffusion of technology, institutions and norms of behavior conducive to higher incomes, is affected by differences in vertically transmitted characteristics associated with genealogical relatedness: populations that are genetically far apart are more likely to differ in those characteristics, and thus less likely to adopt each other's innovations over time. The pattern of the effects of genetic distance in space and time, and the interaction with geographical distance, suggest that genetic distance is associated with important barriers to the diffusion of development Some evidence, particularly the results for European countries, also suggests that these differences may stem in substantial part from cultural (rather than purely genetic) transmission of characteristics across generations.

A final consideration is about policy implications. A common concern with research documenting the importance of variables like genetic distance or geography is pessimism about its 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 coefficient that measure the effect of genetic distance on income differences. That coefficient has been changing over time, and can change further. If we are correct in interpreting our results as evidence for long-term barriers across different populations and cultures, significant reductions in income disparities could be obtained by encouraging policies that reduce those barriers, including efforts to translate and adapt technological and institutional innovations into different cultures and traditions, and to foster cross-cultural exchanges. More work is needed - at the micro as well as macro level - in order to understand the specific mechanisms, market forces, and policies that could facilitate the diffusion of development across countries with distinct long-term histories and cultures.

References

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

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

Ades, Alberto and Edward Glaeser (1999), "Evidence on Growth, Increasing Returns and the Extent of the Market", *Quarterly Journal of Economics*, vol. 114, no. 3, August, pp. 1025-1045.

Alcalá, Francisco and Antonio Ciccone (2004), "Trade and Productivity", *Quarterly Journal of Economics*, vol. 119, no. 2, May.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003), "Fractionalization", *Journal of Economic Growth*, vol. 8, no. 2, June 2003, pp. 155-194.

Alesina, Alberto, Enrico Spolaore and Romain Wacziarg (2000), "Economic Integration and Political Disintegration", *American Economic Review*, vol. 90, no. 5, pp. 1276-1296, December.

Alesina, Alberto, Enrico Spolaore and Romain Wacziarg (2004), "Trade, Growth, and the Size of Countries", in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Amsterdam: North Holland.

Barrett, David, George Kurian and Todd Johnson (2001), *World Christian Encyclopedia (2nd edition)*, New York: Oxford University Press, 2 vols.

Barro, Robert J. and Jong-Wha Lee (2000), "International Data on Educational Attainment: Updates and Implications", *Harvard CID Working Paper* no. 42.

Barro, Robert J. and Xavier Sala-i-Martin (1997), "Technological Diffusion, Convergence and Growth", *Journal of Economic Growth*, vol. 2, no.1, pp. 1-26.

Baumol, William (1986), "Productivity Growth, Convergence and Welfare: What the Long Run Data Show", *American Economic Review*, vol. 76, pp. 1072-1085.

Bisin, Alberto and Thierry Verdier (2000), "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics*, vol. 105, no. 3, pp. 955-988.

Bisin, Alberto and Thierry Verdier (2001), "The Economics of Cultural Transmission and the Evolution of Preferences," *Journal of Economic Theory*, vol. 97, no. 1, pp. 298-319.

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

Case, Anne (1991), "Spatial Patterns in Household Demand", *Econometrica*, vol. 59, no. 4, July, pp. 953-965.

Caselli, Francesco (2005), "Accounting for Cross-Country Income Differences", in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Amsterdam: North Holland.

Caselli, Francesco and John Coleman (2002), "On the Theory of Ethnic Conflict", *working paper*, Harvard University.

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

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

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

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

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

Diamond, Jared (1997), *Guns, Germs, and Steel: The Fates of Human Societies.* New York: W. W. Norton & Company.

Dyen, Isidore, Joseph B. Kruskal and Paul Black (1992), "An Indoeuropean Classification: A Lexicostatistical Experiment", *Transactions of the American Philosophical Society*, vol. 82, no. 5.

Easterly, William and Ross Levine (2003), "Tropics, Germs, and Crops: the Role of Endowments in Economic Development", *Journal of Monetary Economics*, vol. 50, no. 1, January.

Engerman, Stanley L., and Kenneth L. Sokoloff (1997), "Factor Endowments, Institutions,and Differential Paths of Growth among New World Economies: A View from Economic Historians of the United States", in Stephen Haber, ed., *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800-1914*, Stanford: Stanford University Press.

Fearon, James (2003), "Ethnic and Cultural Diversity by Country", Journal of Economic Growth vol. 8, no. 2 June, pp. 195-222.

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

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

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

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

Hall, Robert and Jones, Charles I. (1999), "Why Do Some Countries Produce so Much More Output Per Worker than Others?", *Quarterly Journal of Economics*, vol. 114, no. 1, pp. 83-116, February.

Harrison, Lawrence E. and Samuel P. Huntington, eds. (2000), *Culture Matters: How Values Shape Human Progress*, New York: Basic Books.

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

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

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

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

Kroeber, A. and Kluckhohn, C. (1952), *Culture.* New York: Meridian Books.

Masters, William A and Margaret S. McMillan (2001), "Climate and Scale in Economic Growth," *Journal of Economic Growth*, vol. 6, no. 3, September.

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

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

Mukand, Sharun and Dani Rodrik (2005), "In Search of the Holy Grail: Policy Convergence, Experimentation, and Economic Performance", *American Economic Review*, vol. 95, no. 1, pp. 374-383, March.

Nei, Masatoshi (1987), Molecular Evolutionary Genetics, New York: Columbia University Press.

Olson, Steve (2002), *Mapping Human History. Discovering the Past Through Our Genes*, New York: Houghton Mifflin Company.

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

Parente, Stephen L. and Edward C. Prescott (1994), "Barriers to Technology Adoption and Development", *Journal of Political Economy*, vol. 102, no. 2, pp. 298-321.

Parente, Stephen L. and Edward C. Prescott (2002), *Barriers to Riches*, Cambridge: MIT Press.

Reynolds, J., B.S. Weir, and C.C. Cockerham (1983), "Estimation of the Coancestry Coefficient: Basis for a Short-term Genetic Distance", *Genetics*, vol. 105, pp. 767-779.

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

Rodrik, Dani, Arvind Subramanian and Francesco Trebbi (2004), "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development", *Journal of Economic Growth*, vol. 9, no.2, June.

Sachs, Jeffrey (2001), "Tropical Underdevelopment", *NBER Working Paper* #8119.

Sachs, Jeffrey (2005), *The End of Poverty: Economic Possibilities for Our Time*, New York: Penguin Books.

Spolaore, Enrico and Romain Wacziarg (2005), "Borders and Growth", *Journal of Economic Growth*, vol. 10, no 4, December.

Tabellini, Guido (2005), "Culture and Institutions: Economic Development in the Regions of Europe", *IGIER Working Paper* #292, June.

Wang, E.T., G. Kodama, P. Baldi, and R.K. Moyzis (2006), "Global Landscape of Recent Inferred Darwinian Selection for Homo Sapiens", *Procedeeings of the Natural Academy of Sciences*, January 3, pp.135-140.

Weitzman, Martin L. (1992), "On Diversity", *Quarterly Journal of Economics*, vol. 107, no. 2, pp. 363-405, May.

Appendix

In this Appendix we provide some simple models in which both geographical distance and distance in vertically transmitted characteristics (genetic distance) affect income differences by affecting barriers to the diffusion of innovations. We will show that a *negative* interaction between geographical distance and genetic distance arises naturally in those models. First we will present a simple reduced-form specification, in which we compare barrier effects and direct effects stemming from differences in geographical and VTCs. We will show that a negative interaction is consistent with barrier effects, while it is not implied by direct effects. We will then sketch two specific models of barriers, with more explicit microeconomic interpretations, and show that those models do in fact imply a negative interaction between geographical and genetic distance. The overall message of this Appendix is that a negative interaction between geographical distance and genetic distance is a robust implication of models in which both distances affect income differences through their effects on *barriers* to the diffusion of productivity-enhancing innovations.

Reduced-form Setup

Consider a reduced-form barrier model in which the probability that country i adopts country j's innovation is:

$$P(f, g) = P_1(f)P_2(g)$$

where f is genetic distance between i and j and g is geographical distance between i and j. Either distance reduces the probability of adoption - i.e., it constitutes a barrier to the diffusion of innovations: $dP_1/df < 0$ and $dP_2/dg < 0$. If we specify the probabilities as linear functions, we have:

$$P_1(f) = 1 - b_f f$$

$$P_2(g) = 1 - b_g g$$

and:

$$P(f, g) = 1 - b_f f - b_g g + b_f b_g f g$$

That is, the probability of adopting the same innovation is decreasing in f and g but *increasing* in the interaction $f \times g$. Hence, the expected income difference, which is inversely related to the probability of adopting the same innovation, will be increasing in f and g, but *decreasing* in the interaction $f \times g$, as in our regressions.

By contrast, suppose that vertically transmitted characteristics q's had a *direct* effect on total factor productivity, with $q_i > q_j$:

$$TFA_i = A_i = a_i(q_i)^\gamma$$

Consistently with our framework in Section 2, assume that the difference in vertically transmitted characteristics is a function of genetic distance. In particular, for analytical convenience and without any loss of generality, assume:

$$\ln q_i - \ln q_j = \Psi f$$

Suppose that society i comes up with an innovation that increases total factor productivity to some level $a_i > a_j$ If country j does not adopt the innovation, the difference in their levels of total factor productivity is given by:

$$\ln A_i - \ln A_j = \ln a_i - \ln a_j + \gamma(\ln q_i - \ln q_j) = \Phi + \gamma\Psi f$$

While if both countries adopt the innovation and achieve the same level of total factor productivity, we have:

$$\ln A_i - \ln A_j = \gamma(\ln q_i - \ln q_j) = \gamma\Psi f$$

By putting the two effects together (barriers effects to the adoption of the innovation, and direct effects from vertically transmitted characteristics), we have that the expected difference in total factor productivities between countries i and j is:

$$\ln A_i - \ln A_j = (1-P)\Phi(a_i - a_j) + \gamma\Psi f = [b_f f + b_g g - b_f b_g f g]\Phi + \gamma\Psi f = c_1 f + c_2 g + c_3 f g$$

where:

$$c_1 = b_f\Phi + \gamma\Psi > 0$$

$$c_2 = b_g\Phi > 0$$

$$c_3 = -b_f b_g \Phi < 0$$

This very simple reduced-form model implies a *positive* effect of genetic distance and geographical distance on income differences, and a *negative* interaction as long as both distances constitute *barriers* to the diffusion of innovations ($b_f > 0$ and $b_g > 0$). It is also worth stressing that a direct effect of genetic distance on productivity differences ($\gamma\Psi > 0$) increases the magnitude of c_1, but it is not necessary for $c_1 > 0$. In other terms, barrier effects alone are *sufficient* for $c_1 > 0$ and $c_2 > 0$, while they are *necessary and sufficient* for $c_3 < 0$.

Two Models of Barriers

In what follows we will briefly sketch two models with a more explicit microeconomic interpretation of the mechanisms through which barriers affect the adoption of innovations. Model 1 will focus on a physical interpretation of barriers: barriers may prevent societies from *observing* other societies's innovations Model 1 considers a cost interpretation of barriers: barriers increase the costs to adapt and imitate other societies' innovations. In either case, the model implies a negative interaction between geographical distance and genetic distance.

Model 1

Consider a model in which both geographical distance and genetic distance reduce the probability that a society would be able to observe another society's innovation.

Again, consider two countries (i and j), which are separated by N geographical steps and M cultural steps. An innovation discovered in country i must travel all N plus M steps in order to reach country j. At each cultural step there is a probability θ_f that the innovation will be "lost in translation," while at each geographical step there is a probability θ_g that the innovation will fail to make it through that geographical space, where $0 < \theta_f < 1$ and $0 < \theta_g < 1$.[64] Therefore the two countries' expected difference in productivity and income per capita will be a function of the total *probability that the innovation is lost*, i.e.:

$$P(N, M) = 1 - (1 - \theta_f)^M (1 - \theta_g)^N$$

It is immediately apparent that this probability is a positive function of the geographical distance (measured by N) and of the cultural distance (measured by M), and a negative function of their interaction:

$$\frac{\partial P(N, M)}{\partial M} = -\ln(1 - \theta_f)(1 - \theta_f)^M (1 - \theta_g)^N > 0$$

$$\frac{\partial P(N, M)}{\partial N} = -\ln(1 - \theta_g)(1 - \theta_f)^M (1 - \theta_g)^N > 0$$

$$\frac{\partial^2 P(N, M)}{\partial M \partial N} = -\ln(1 - \theta_f)\ln(1 - \theta_g)(1 - \theta_f)^M (1 - \theta_g)^N < 0$$

Hence, these results are consistent with our reduced-form model above: differences in expected productivity and income per capita are positively associated to measures of geographical and cultural

[64] Hence $\ln(1 - \theta_g) < 0$ and $\ln(1 - \theta_f) < 0$.

distance, and negatively associated with their interaction, when those measures represent barriers to the diffusion of productivity-enhancing innovations.

Model 2

A different mechanism that delivers analogous results is based on the assumption that geographical distance reduces the probability that the innovation is observed by the distant country, while cultural distance increases translation costs. Suppose that country i produces an innovation of size Δ, while country j does not. But now also assume that, when the innovation is observed in country j, translating it into the local productive system entails a cost C which is higher for higher cultural distance M. That is, $\frac{dC}{dM} > 0$.

The innovation will be adopted in country j if and only if $C(M) < \Delta$. Define as $\Phi\{C(M) - \Delta \geq 0\}$ the probability that the translation costs are too high, and the innovation is not adopted in country j. Since $\dfrac{dC}{dM} > 0$, we have $\dfrac{\partial \Phi}{\partial M} > 0$.

Again, the two countries' expected differences in incomes per capita will be a function of the probability that country j does not adopt country i's innovation. Then the probability that the two countries have different income per capita is given by the sum of a) the probability that geographical distance prevents country j from observing country i's innovation, and b) the probability that country j observes country i's innovation but fails to adopt it because the translation costs due to *cultural distance* are too high:

$$P(M, N) = [1 - (1 - \theta_g)^N] + (1 - \theta_g)^N \Phi\{C(M) - \Delta \geq 0\}$$

Again, we have that this probability $P(M, N)$ is increasing in N and M, while *their interaction is negative*:

$$\frac{\partial P(N, M)}{\partial N} = -[1 - \Phi\{C(M) - \Delta \geq 0\}] \ln(1 - \theta_g)(1 - \theta_g)^N > 0$$

$$\frac{\partial P(N, M)}{\partial M} = (1 - \theta_g)^N \frac{\partial \Phi}{\partial M} > 0$$

$$\frac{\partial^2 P(N, M)}{\partial M \partial N} = \ln(1 - \theta_g)(1 - \theta_g)^N \frac{\partial \Phi}{\partial M} < 0$$

Hence, this model is also consistent with the reduced-form above.

Table 1 – Summary Statistics for the Main Variables

Panel a. Simple Correlations among Distance Measures

	Geodesic distance	Diff. in absolute latitudes	Diff. in absolute latitudes	F_{ST} Gen. Dist.	F_{ST} Gen. Dist., 1500	Nei Gen. Dist.	Abs. log income diff. 1995	Abs. log income diff. 1500[a]
Difference in absolute latitudes	0.331	1						
Difference in absolute longitudes	0.843	0.060	1					
F_{ST} Genetic Distance	0.354	0.138	0.205	1				
F_{ST} Genetic Distance, 1500 match	0.478	0.166	0.305	0.658	1			
Nei Genetic Distance	0.318	0.154	0.171	0.929	0.606	1		
Abs. log income difference, 1995	0.015	0.104	-0.048	0.141	0.226	0.177	1	
Abs. log income difference, 1500[a]	0.159	0.155	0.069	-0.096	0.196	-0.086	-0.051	1
Abs. log income difference, 1700[b]	0.141	0.249	0.131	0.000	0.072	0.076	0.503	0.060

(number of observations: 13861 except [a]: 325 and [b]: 1431)

Panel b. Means and Standard Deviations

Variable	# Obs.	Mean	Std. Dev.	Min	Max
Worldwide Dataset					
Geodesic distance (1000s of km)	13861	7.939	4.499	0.010	19.904
Latitudinal distance	13861	0.283	0.205	0.000	1.060
Longitudinal distance	13861	0.731	0.584	0.000	3.500
F_{ST} Genetic distance	13861	0.121	0.083	0.000	0.338
F_{ST} Genetic distance, 1500	13861	0.126	0.076	0.000	0.356
Nei Genetic distance	13861	0.020	0.015	0.000	0.062
Abs. log income difference, 1995	13861	1.291	0.913	0.000	4.294
Abs. log income difference, 1500	325	0.327	0.237	0.000	1.012
Abs. log income difference, 1870	1431	0.647	0.488	0.000	2.110
European Dataset					
F_{ST} Genetic distance (Europe)	296	0.010	0.006	0.000	0.032
Abs. log income difference, 1995	296	0.547	0.431	0.006	1.652
Abs. log income difference, 1870	153	0.448	0.304	0.007	1.288

Table 2 - Baseline regressions
(common-country fixed effects, dependent variable: absolute value of log income differences, 1995)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Geodesic Distance	**Latitude Difference**	**Longitude Difference**	**FST Gen Dist**	**Nei Gen Dist**	**All Distances (Nei)**	**All distances (FST)**	**With controls**
Fst Genetic Distance				**2.388**			**2.104**	**2.000**
				(0.102)**			**(0.106)****	**(0.107)****
Nei Genetic Distance					**14.893**	**13.484**		
					(0.569)**	**(0.581)****		
Geodesic Distance	**0.026**					0.015	0.014	0.013
(1000s of km)	**(0.002)****					(0.003)**	(0.003)**	(0.003)**
Absolute difference		**0.577**				0.296	0.294	0.237
in latitudes		**(0.042)****				(0.047)**	(0.047)**	(0.048)**
Absolute difference			**0.081**			-0.078	-0.079	-0.086
in longitudes			**(0.014)****			(0.027)**	(0.027)**	(0.027)**
1 for contiguity								-0.388
								(0.059)**
1 if countries were or are								-0.258
the same country								(0.071)**
1 if common language								0.093
(9% threshold)								(0.025)**
1 for pairs ever in colonial								0.076
relationship								(0.077)
1 for common colonizer								-0.083
post 1945								(0.028)**
1 for pairs currently in								-1.591
colonial relationship								(0.359)**
# observations	13861	13861	13861	13861	13861	13861	13861	13861
(# countries)	(167)	(167)	(167)	(167)	(167)	(167)	(167)	(167)
Adjusted R-squared	0.75	0.75	0.75	0.76	0.76	0.76	0.76	0.76
Effect of 1 s.d. change in bold	12.66%	12.99%	5.17%	21.77%	24.05%	21.77%	19.18%	18.23%
regressor, % 1 s.d. income diff.								

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%.

Table 3 - Robustness Tests and Extensions, Part I
(common-country fixed effects, dependent variable: difference in log income per capita in 1995, except as stated, in column 5)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dist*GD Interaction	**Weighted GD**	**Without New World**	**IV with 1500 GD**	**Diamond Gap, w/o New World**	**Income 1500, Diamond Gap**	**Climatic difference control**	**Tropical difference control**
Weighted Fst Genetic Distance		**2.335 (0.120)****						
Fst Genetic Distance	**2.474 (0.212)****		**2.306 (0.173)****	**5.337 (0.207)****	**1.222 (0.199)****		**2.697 (0.129)****	**2.872 (0.129)****
Fst Genetic Distance, 1500 match						**1.703 (0.441)****		
Absolute difference in Latitudes	0.172 (0.053)**	0.369 (0.050)**	0.753 (0.112)**	0.244 (0.049)**	0.721 (0.111)**	0.300 (0.092)**	0.151 (0.068)**	0.352 (0.073)**
Absolute difference in Longitudes	-0.093 (0.027)**	0.003 (0.028)	-0.331 (0.095)**	0.046 (0.029)	0.013 (0.094)	0.009 (0.034)	-0.159 (0.050)**	-0.102 (0.052)*
Geodesic Distance (1000s of km)	0.023 (0.005)**	0.008 (0.003)**	0.015 (0.012)	-0.017 (0.004)**	-0.032 (0.012)**	-0.010 (0.008)	0.017 (0.006)**	0.010 (0.006)
Diamond Gap					0.390 (0.033)**	0.209 (0.029)**		
Measure of climatic difference of land areas, by 12 KG zones							0.037 (0.002)**	
Difference in % land area in KG tropical climates								0.147 (0.022)**
Distance * Fst Genetic Distance	-0.063 (0.024)**							
Observations	13861	11079	7626	13861	7626	325	10153	10153
(# countries)	(167)	(158)	(124)	(167)	(124)	(26)	(143)	(143)
Adjusted R-squared	0.76	0.78	0.75	-	0.76	0.81	0.78	0.77
Effect of 1 s.d. change in bold regressor, % 1 s.d. income diff.	17.96% [a]	20.52%	20.15%	48.66%	10.68%	39.26%	24.13%	25.70%

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%. All specifications except that of column 5 include dummies equal to 1 if countries are contiguous, if countries were or are the same country, if they share a common language (9% threshold), for pairs ever in a colonial relationship, for pairs with a common colonizer post 1945 and for pairs currently in a colonial relationship, as in column (8) of Table 2. The estimated coefficients for these controls (not reported) are available upon request.

[a]: effect evaluated at the mean of geodesic distance.

Table 4 - Robustness Tests and Extensions, Part II
(common-country fixed effects, dependent variable: difference in log income per capita in 1995)

	(1)	(2)	(3)	(4)	(5)	(6)
	Isolation controls	**Same continent control**	**Same continent controls**	**Linguistic distance, dominant**	**Linguistic distance, expected**	**Religious similarity**
Fst Genetic Distance	**2.028 (0.106)****	**1.391 (0.106)****	**0.810 (0.109)****	**2.738 (0.129)****	**2.692 (0.129)****	**2.032 (0.111)****
Absolute difference in Latitudes	0.248 (0.047)**	0.134 (0.047)**	-0.149 (0.054)**	0.332 (0.062)**	0.330 (0.062)**	0.379 (0.047)**
Absolute difference in Longitudes	-0.078 (0.027)**	-0.127 (0.027)**	-0.008 (0.027)	-0.148 (0.046)**	-0.156 (0.046)**	-0.040 (0.025)
Geodesic Distance (1000s of km)	0.010 (0.003)**	-0.007 (0.004)*	0.004 (0.004)	0.007 (0.006)	0.007 (0.006)	0.002 (0.003)
=1 if either country is an island	0.212 (0.042)**					
=1 if either country is landlocked	0.406 (0.046)**					
Same Continent Dummy		-0.412 (0.022)**				
Both in Asia			0.304 (0.037)**			
Both in Africa			-0.960 (0.039)**			
Both in Europe			-0.775 (0.048)**			
Both in North America			-1.773 (0.131)**			
Both in Latin America/Caribbean			-0.107 (0.044)**			
Both in Oceania			0.041 (0.163)			
Linguistic Distance Index, dominant languages				0.235 (0.077)**		
Linguistic Distance Index, Expected					0.469 (0.094)**	
Religious Difference, based on Barrett Data						1.443 (0.163)**
# Observations (# countries)	13861 (167)	13861 (167)	13861 (167)	10011 (142)	10004 (142)	12403 (158)
Adjusted R-squared	0.77	0.77	0.78	0.77	0.77	0.77
Effect of 1 s.d. change in bold regressor, % 1 s.d. income diff.	18.46%	12.65%	7.37%	24.91%	24.50%	18.49%

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%. All specifications include dummies equal to 1 if countries are contiguous, if countries were or are the same country, if they share a common language (9% threshold), for pairs ever in a colonial relationship, for pairs with a common colonizer post 1945 and for pairs currently in a colonial relationship, as in column (8) of Table 2. The estimated coefficients for these controls (not reported) are available upon request.

Table 5 - Regressions on Historical Income Data
(common-country fixed effects, dependent variable: absolute value of log income differences)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Income 1500	**Income 1700**	**Income 1820**	**Income 1870**	**Income 1913**	**Income 1960**	**Income 1995**
Fst Genetic Distance, 1500 match	**2.163** **(0.451)****	**3.033** **(0.460)****					
Fst Genetic Distance			**0.827** **(0.231)****	**3.248** **(0.349)****	**3.294** **(0.364)****	**1.469** **(0.105)****	**2.104** **(0.106)****
Absolute difference in latitudes	0.215 (0.105)**	0.547 (0.140)**	1.032 (0.068)**	0.833 (0.106)**	0.878 (0.118)**	0.538 (0.072)**	0.014 (0.003)**
Absolute difference in longitudes	-0.096 (0.051)*	-0.058 (0.058)	0.322 (0.039)**	0.327 (0.081)**	0.400 (0.093)**	-0.005 (0.042)	0.294 (0.047)**
Geodesic Distance (1000s of km)	0.014 (0.008)*	0.005 (0.011)	-0.035 (0.006)**	-0.055 (0.011)**	-0.059 (0.012)**	0.006 (0.005)	-0.079 (0.027)**
# Observations (# countries)	325 (26)	406 (29)	1176 (49)	1711 (59)	1891 (62)	5995 (110)	13861 (167)
Adjusted R-squared	0.80	0.85	0.84	0.79	0.77	0.79	0.76
Effect of 1 s.d. change in bold regressor, % 1 s.d. income diff.	49.86%	46.35%	15.20%	40.93%	37.44%	18.04%	19.18%
Common sample effect [a]	49.86%	50.04%	37.67%	88.75%	59.55%	34.94%	17.68%

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%.
[a]: Effect of 1 s.d. change in bold regressor, % 1 s.d. income diff., for the common sample of 325 observations (26 countries), computed to ensure comparability across columns.

Table 6 - Regressions for European Countries
(common country fixed effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Distance	**Latitude Difference**	**Longitude Difference**	**FST Gen Dist**	**All Distances**	**With controls**	**1870 Income**
Fst genetic distance in Europe				**35.110**	**32.108**	**33.480**	**49.511**
				(5.820)**	**(7.605)****	**(8.136)****	**(7.110)****
Geodesic Distance (1000s of km)	**0.212**				0.033	0.023	-0.135
	(0.042)**				(0.127)	(0.129)	(0.134)
Absolute difference in latitudes		**0.270**			-0.819	-1.042	0.513
		(0.365)			(0.868)	(0.875)	(1.029)
Absolute difference in longitudes			**1.651**		0.976	0.789	1.302
			(0.324)**		(0.683)	(0.678)	(0.810)
1 for contiguity						-0.042	0.069
						(0.085)	(0.084)
1 if countries were or are the same country						-0.033	-0.253
						(0.118)	(0.282)
1 if common language (9% threshold)						-0.323	-0.302
						(0.101)**	(0.122)**
# observations	296	296	296	296	296	296	171
(# countries)	(25)	(25)	(25)	(25)	(25)	(25)	(18)
Adjusted R-squared	0.77	0.74	0.77	0.78	0.79	0.80	0.80
Effect of 1 s.d. change in bold regressor, % 1 s.d. income diff.	38.53%	3.96%	89.55%	51.57%	47.16%	49.18%	79.59%

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%.

Table 7 - Regressions for the proximate determinants of income
(common-country fixed effects, dependent variable: absolute value difference in the variables appearing in the second row)

	(1)	(2)	(3)	(4)	(5)
	Institutions	**Human Capital**	**Population growth**	**Investment**	**Openness**
Fst Genetic Distance	**3.711** **(0.238)****	**6.259** **(0.356)****	**0.337** **(0.029)****	**10.243** **(0.780)****	**-0.044** **(0.037)**
Absolute difference in latitudes	1.813 (0.122)**	2.201 (0.236)**	0.392 (0.022)**	9.424 (0.484)**	-0.049 (0.023)**
Absolute difference in longitudes	-0.001 (0.089)	-0.059 (0.172)	0.067 (0.012)**	0.002 (0.352)	-0.026 (0.017)
Geodesic Distance (1000s of km)	-0.012 (0.011)	-0.020 (0.021)	-0.010 (0.002)**	-0.000 (0.042)	0.002 (0.002)
1 for contiguity	-0.211 (0.117)*	-0.367 (0.177)**	-0.036 (0.018)**	-0.268 (0.484)	-0.027 (0.022)
1 if countries were or are the same country	-0.058 (0.163)	-0.786 (0.251)**	-0.004 (0.020)	-0.910 (0.461)**	-0.044 (0.032)
1 if common language (9% threshold)	0.115 (0.047)**	0.251 (0.086)**	0.011 (0.007)	0.299 (0.202)	-0.030 (0.010)**
1 for pairs ever in colonial relationship	0.136 (0.114)	-0.234 (0.225)	0.054 (0.024)**	1.269 (0.612)**	0.037 (0.021)*
1 for common colonizer post 1945	-0.130 (0.062)**	-0.384 (0.102)**	-0.034 (0.009)**	-0.830 (0.243)**	-0.143 (0.015)**
1 for pairs currently in colonial relationship	-0.305 (0.192)	-0.297 (1.541)	-0.193 (0.261)	-6.059 (0.845)**	0.234 (0.038)**
# observations # countries	6328 113	5565 106	6105 111	8911 134	8911 134
Adjusted R-squared	0.83	0.78	0.79	0.80	0.87
Effect of 1 s.d. change in bold regressor, % 1 s.d. of dep. var.	17.99%	23.01%	12.88%	12.52%	-0.74%

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%