

IZA DP No. 2229

Genetic, Cultural and Geographical Distances

Paola Giuliano
Antonio Spilimbergo
Giovanni Tonon

July 2006

Genetic, Cultural and Geographical Distances

Paola Giuliano

*International Monetary Fund
and IZA Bonn*

Antonio Spilimbergo

*International Monetary Fund,
CEPR and WDI*

Giovanni Tonon

DFCI, Harvard University

Discussion Paper No. 2229

July 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Genetic, Cultural and Geographical Distances^{*}

This paper investigates how the measures of genetic distance between populations, which have been used in anthropology and historical linguistics, can be used in economics. What does the correlation between genetic distance and economic variables mean? Using the measure of genetic distance, a newly-collected database on transport costs, as well as more refined measures of geography within Europe, we show that i) geography explains both genetic distance and transportation costs between European countries, and ii) genetic distance does not explain economic outcomes once we control for geography. We conclude that genetic distance in economics capture transportation costs between countries and not cultural differences.

JEL Classification: Z10, F10

Keywords: transport costs, genetics, trade, geography

Corresponding author:

Paola Giuliano
International Monetary Fund
Research Department
700 19th Street NW
Washington, DC 20431
USA
E-mail: pgiuliano@imf.org

^{*} The views expressed in this paper are those of the authors and do not necessarily represent those of the IMF, its Board of Directors, or the views of the Dana Farber Cancer Institute. We thank Alberto Alesina, Andrea Bianchi, Francesco Caselli, Oya Celasun, David Hummels, Simon Johnson, Nuno Limao, Paolo Mauro, Enrico Spolaore, Arvind Subramanian, Antonio Terraciano, Alessandro Vanzetti, Romain Wacziarg, Luigi Zingales, and the participants in the spring 2006 NBER workshop on political economy for useful discussions, and Zhaogang Qiao and José Romero for excellent research assistantship.

I. INTRODUCTION

Cultural factors have a strong influence on economic and social phenomena. This proposition has a long intellectual history in Western thought going back to Greek and Roman writers. Social scientists have argued that cultural innovations have led to the development of capitalism (Weber, 1958) or that different historical and culture experiences in Italian regions have led to different development paths (Putnam, 1993). Despite this long tradition, until recently, there was little quantitative analysis of the effects of culture on economic outcomes, because culture is a very elusive concept and it is very challenging to proof a unequivocal causal relationship from culture to economic outcomes.²

Besides economics, other disciplines have grappled with the challenge of using quantitative measures to study culture. The pioneering work by Cavalli-Sforza (1994) has introduced the use of genetic analysis in social sciences, including archeology, paleo-anthropology, linguistics, history, and culture. After collecting an impressive database on genetic distances among various populations, Cavalli-Sforza has argued that there is a strong correlation between genetic patterns and ancient migrations, taking place especially in Neolithic times, which have ultimately determined language differences (Cavalli-Sforza, Menozzi, and Piazza, 1994).³ Cavalli-Sforza's successful and convincing use of genetics to deal with social phenomena such as languages has attracted the attentions of scholars from other social sciences, including economists, who were looking for exogenous and quantifiable measures of cultural differences.

Two intriguing papers propose using genetic distance as a proxy (Spolaore and Wacziarg, 2006) or as an instrument for culture (Guiso, Sapienza, and Zingales, 2005). Spolaore and Wacziarg interpret the Cavalli-Sforza's index as a measure of "vertically transmitted characteristics," reflecting different historical paths of populations over the long run, and show that bilateral differences in per-capita income levels are strongly correlated with genetic distance. Guiso, Sapienza, and Zingales (2005) argue that one specific cultural trait, the degree of bilateral trust between countries, is a very important determinant of international trade. Since trust is

² See Guiso et al. (2006) for a review of the recent literature.

³ Section 2 below defines the measure of genetic distance.

obviously endogenous, they propose genetic distance as an instrument for it. Being pre-determined, genetic distance is unlikely to be related to current economic activity (Guiso, Sapienza, and Zingales, 2005).

This paper questions the validity of the use of genetic distance in economics as a proxy or as an instrument for culture and proposes a different interpretation of the correlation between genetic distance and economic outcomes. Our starting point is that genetic distance and economic outcomes are both influenced by geographical variables. Contemporary genetic patterns in Europe were shaped by natural selection, migrations, and genetic drift, the latter ones largely determined by geographical impediments. After several millennia and despite advancements in transportation technology, the mountains, the rivers, and the seas, which shaped past migrations and genetic drift, continue to have an impact on modern transport costs and, ultimately, on trading flows between countries. As a result, the correlation between genetic distance and trade is largely spurious and disappears once geography is properly accounted for.

We make our point considering the trade among European countries. We chose Europe because there is a considerable overlap between genetically defined populations and politically defined countries.⁴ We chose trade because, among other economic outcomes, it is clearly connected to mutual trust, a cultural trait. In addition, trade allows us interesting robustness tests selecting only groups of goods for which the effect of trust or other cultural traits should be more relevant. Finally, gravity equations provide an established benchmark to test hypotheses in trade.

As a further robustness and to show that our results are not limited to trade, we also consider the role of genetic distance in explaining bilateral income differences among European countries as in Spolaore and Wacziarg (2006).

⁴Although the biological and the political concepts of “populations” are often used as synonymous, they are clearly different. For example within Italy, Sardinians are a population genetically very different from the rest of Italians (the genetic distance between Sardinians and Italians is 221 while the difference between Italians and Swedes is 95); similarly the Lapps are very different from the Finns. In our paper, we consider only the populations associated to political countries, this implies that from Cavalli-Sforza’s specification, we exclude Basque, Lapp, Sardinian and Scottish. Note that Spolaore and Wacziarg (2006) use data for 42 genetic populations to map into more than 120 countries in the world.

The present paper contributes to three lines of research. One is the study of the importance of culture on economic outcomes. The main challenge in the analysis of culture is to guarantee that apparently exogenous measures are not capturing omitted variables. The contribution of this paper is to show that genetic distance is in reality highly correlated with geographic variables and cannot be used as an instrument or a proxy for cultural variables.

Second, our paper contributes to the debate on the role of geography in economic development (see Rodrik, 2002, or Sachs, 2003, for a summary). Geography matters for development in a not obvious way, including by influencing the ethnic composition of a country. For instance, Acemoglu et al. (2001) show how geography had an impact on settlers' mortality and so on the pattern of colonization; Alesina et al. (2006) show that those countries, whose border shape does not reflect natural geographical barriers, experienced a lower level of economic development. Our paper provides a further example of the role that geography may play in an indirect, but not less powerful way, on economic development.

Finally, our paper contributes to the literature on the determinants of transportation costs. The paper provides an original contribution to this literature in two ways. Several authors have shown that the simple measures of (log)-distance is only a first approximation for true transport costs (Hummels, 1998; Limao and Venables, 2001); in the context of the gravity models, many studies have included geographical variables such as insularity or contiguity to complement the standard crude measure of distance. Building on this tradition, we have shown that major mountains, common seas and countries elevation also contribute to transport costs. Second, we use a new dataset on transportation costs. The currently used measures of transportation costs are indirect measures, plagued by measurement errors; our measure represents the actual transport costs, allowing us to study the importance of transportation costs on trade in a more reliable way.

The rest of the paper is organized as follows. Section II provides an overview of the available measures of genetic distance, highlighting how they are calculated, what they measure and their relationship with physical anthropology data, including anthropometric characters like stature or qualitative traits such as eye color or skin pigmentation. Section III shows that genetic distance may explain very well trade between European countries in a standard gravity equation with (log)-distance as a proxy of transport costs, but becomes insignificant once transport costs

or other variables capturing geographical features are introduced. Section IV discusses alternative uses of genetic distance in economics; section V concludes.

II. WHAT IS IN THE GENES?

Population genetics studies populations' genetic composition and their changes over time, focusing on genes that are present in at least two different forms (alleles) in the population. In its simplest form, the fundamental measurement in population genetics is the frequency at which alleles are found at any specific *gene locus* (allele frequency).⁵

Although not all alleles occur in all human populations, differences in alleles within local human populations are much greater than among different populations. Specifically, 93% of total human variability is found within local populations. The remaining 7% is found between populations (Rosenberg et al., 2002). As noted by Lewontin, "if everyone on earth becomes extinct except for the Kikuyu of East Africa, about 90% of all human variability would still be present in the reconstituted species" (Lewontin, 1984).

Subsets of the specific group of genes that varies between populations are used to reconstruct the evolutionary history of populations. Genetic variation among human populations derives mainly from gradations in allele frequencies of subset of genes rather than from distinctive alleles present in specific populations. It is only through the accumulation of

⁵ With the term *gene locus* (or, for simplicity, gene) we intend a sequence of DNA that encodes for a protein, and with the term *allele* we consider a particular form of a specific gene. Often alleles are distinguished for their effects on the phenotype (e.g., morphological, physiological or biochemical characteristics of an individual – or group of like individuals – that differ in this respect from other individuals), or, simply, for differences in pair sequence. Several different methods have been used to measure the genetic composition of a population. Some of these techniques are directly linked to DNA alterations. Classical analysis instead measured the result of DNA alterations, that is, protein variation. The most extensive and comprehensive studies on variants have been performed on protein polymorphism. The ongoing Human Genome Diversity Project (HGDP) and the International HapMap Project will soon provide a wealth of data and information linked directly to the DNA status, but the results available so far and the analysis performed on these data are not exhaustive. Preliminary analysis, however, supports the notion that the major tenets of the classical protein polymorphism analysis, as presented in its more comprehensive form by Cavalli-Sforza et al. in 1994, correlates closely with this new, more extensive scrutiny. For this reason, in the present study, we have relied on the data as provided in Cavalli-Sforza.

small allele-frequency differences across many loci that the genetic structure of a population, that is, the distinctive combination of allele frequencies, could emerge; see also below).

Several indices have been proposed to quantify the degree of genetic differentiation among two or more populations using series of gene frequencies. One such index is the F_{ST} distance, which measures the genetic variance between populations as a fraction of the total genetic variance. By construction, F_{ST} ranges between 0 and 1; the closer F_{ST} is to 1, the higher is the genetic distance between two populations. This index has shown a high degree of correlation with other measures of genetic distances and since the data provided by Cavalli-Sforza are expressed in F_{ST} this index will be used in this study.

Phenotypic characteristics (including anthropometric characters like stature or qualitative traits such as eye color or skin pigmentation) and the overall genetic structure of human populations are not related. For example, the pattern of overall genetic variation among populations differs substantially from traditional racial divisions (Figure 1). Morphologically similar peoples are not necessarily genetically similar overall.

These findings confirm that physical anthropology data are not reliable to reconstruct past migrations because external traits on which anthropometric studies are typically based on are particularly sensitive to natural selection. Only a very small fraction of the human genes is related to phenotypes that are under strong selection pressure (see for example Akey et al., 2002; Goldstein and Chikhi, 2002). In contrast, most of the genes that differ between populations and are used to compute genetic distance are selectively neutral, that is, they lack selective advantage (see “Neutral Theory of Evolution”, Kimura, 1968). As already clear to Darwin, neutral characters are best for reconstructing evolutionary history. If many genes used for the analysis show intercorrelated responses to the various environments in which human evolution has occurred, the measured genetic distance would be a reflection of the environments rather than of evolutionary history.

The absence of correlation between genetic distance and the color of the skin is particularly intriguing and would argue against a relationship between “cultural perception” and overall genetic features, as measured by Cavalli-Sforza et al., as well as by classical human population studies. Indeed, recent reports have suggested how skin pigmentation correlates with polymorphisms affecting single genes (Lamason et al., 2005; Soejima et al., 2006).

In contrast, genetic distance and geography are strongly correlated. Without using prior information about individual sampling locations, a clustering algorithm applied to multilocus genotypes from worldwide human populations produced genetic clusters largely coincident with major geographic regions (Rosenberg et al., 2005). For populations that are geographically close, genetic and geographic distances are often highly correlated, with genetic distance reaching an asymptote at about 1000-2600 miles on average (Figure 2).⁶ Moreover, small discontinuous jumps in genetic distance are present for most population pairs on opposite sides of geographic barriers (Rosenberg et al., 2005). This is also true for Europe, where sharp increases in genetic distance correspond to geographical impediments, including major mountains and seas (Barbujani and Sokal, 1990; see Figure 3).

In conclusion, the Cavalli_Sforza's measure of genetic distance, which has been used in economics, is very poorly correlated with external traits, which determine social perception of "races", including skin pigmentation and heights while it is correlated with geographical variables.

A. Genes, Culture, and Geography in Europe

The correlation between genetic distances and cultural variables is still controversial. While Cavalli-Sforza has convincingly argued that linguistic families are correlated with ancient migration and genetic patterns, the correlation with other cultural traits is at best tentative.

As we previously said, differences between populations arise largely through random genetic drift when they are separated by distance, geographical barriers or culture. Europe has been considered an excellent area to study the importance of the different factors because its archeology, linguistics and genetics are fairly well known.

Two recent studies, one for Northern European populations (Zerjal et al. 2000) and one for the entire Europe (Rosser et al., 2000) show that populations in Europe are related mainly on the basis of geography and not on the basis of linguistic affinity. Northern Europe shows

⁶ This issue is analyzed further in the empirical section, see in particular Table 1 for the geographic determinants of genetic distance.

linguistic and cultural diversity.⁷ At the same time, the Scandinavian Peninsula is separated from Finland and the Baltic countries by the Baltic Sea. Zerjal et al. (2000), using Y-chromosomal data, conclude that the major genetic difference in Northern Europe is geographical, distinguishing populations living in the Western and Eastern side of the Baltic. Language plays a less but still important part in the determination of genetic differences (they found that Latvians showed greater genetic similarity to the Lithuanians than to the Estonians). Using Y-chromosome data and extending the sample to 47 European countries⁸, Rosser et al. (2000) also find a strong and highly significant partial correlation between genetics and geography but a low and non-significant partial correlation between genetics and language.

From the evidence above, we conclude that, while the strong correlation between geography and genetic differences is uncontroversial, the relationship between genetic distance and some manifestation of culture is still argument of debate.

Even if the correlation between language families and genetic distance seems plausible, it is unclear how to use this correlation in economics. Economic ties or transmission of information between populations are facilitated by mutual comprehension but belonging to the same linguistic family is not necessarily a good measure of mutual comprehension. For instance, Indians from New Delhi are linguistically much closer to Icelanders than to Indians from Mumbai but this does not suggest any strong cultural commonality between Icelanders and Northern Indians. Moreover, to an Italian-speaker Hungarian, Hindi, or Armenian are equally incomprehensible despite the fact that the Italian is much closer historically to Armenian and Hindi than Hungarian. On the other hand, for an English-speaker French could be more intelligible than German despite the fact that German and French are both Germanic languages. In other words, belonging to the same historical linguistic group only in few cases helps communication.

⁷ In terms of language: Swedes and Norwegians belong to the Germanic subfamily of the Indo-Europeans, Latvian and Lithuanians belong to the Baltic subfamily of the Indo-Europeans and Finns, Saami and Estonians belong to the Uralic family. Cultural differences are defined by the authors as livelihood differences: the Saami were hunters and nomadic, whereas the other populations were mainly farmers.

⁸ The dataset is the biggest existing available on Y-chromosome diversity.

III. GENETIC DISTANCE, GEOGRAPHY, AND TRADE

This section analyzes the relationship between genetic distance and geography. Our first goal is to show how geography has shaped genetic differences within Europe. Our starting point is Figure 3 (Sokal et al., 1990), which shows the main genetic changes within Europe. Sokal et al. (1990) have identified 33 boundaries of sharp changes in gene frequencies across Europe and have shown that the zones of abrupt genetic change in European populations correspond mostly to geographical boundaries. Specifically, the authors have counted 22 physical, 4 mountainous, and 18 marine boundaries. “In the 22 cases in which there are both physical barriers and genetic boundaries, it is reasonable to postulate that the causal arrow is likely to go more from physical barriers to both genetic and linguistic differentiation, rather than in other directions” (Cavalli-Sforza, 1996, pag. 271).⁹ The importance of geography is also confirmed by classical genetic studies in humans and other organisms, also showing a strong association between geographic boundaries and genetic distance. Finally, note the ambiguous effect of sea. Ancient migrations often followed the sea coasts; sharing the same sea is a unifying factor. At the same time, crossing large seas was relatively complicated so islands are usually genetically isolated.

In order to investigate more systematically how geographical factors shape genetic distance, we run a regression with genetic distance as dependent variable and several geographical variables as control variables. The measure of genetic distance is derived from Cavalli-Sforza et al., p. 270 (with F_{ST} derived from the analysis of the allele frequencies of 88 genes). The choice of geographical variables, following Sokal et al. (1990), includes distance, number of mountains between countries, the presence of a common sea, and average terrain elevation between two countries (as defined below). In addition, all regressions have country fixed effects to control for country specific characteristics. The results presented in the first

⁹ Note also that some genetic boundaries cut countries in the middle (for instance in Germany, Italy, Finland, Island, Spain, and Greece.) This observation reinforces the point that sharp genetic differences exist within national population, i.e. national borders are not always genetic borders. Moreover some abrupt changes follow linguistic and not geographical lines, for instance between Germany and the Netherlands. In other cases, there are linguistic barriers but no big genetic discontinuity (see for instance Germany and France, or Germany and Poland).

three columns of Table 1 use different combinations of geographical variables and confirm that geographical measures and genetic distance among European countries are indeed correlated.¹⁰ The regressions reported in Table 1 and the literature reviewed above show that geography (including the distance between countries, the presence of major mountains chains, and common seas) plays a fundamental role in explaining genetic distance either by having determined past migration routes or by having separated populations, thereby contributing to the genetic drift.

Given the strong correlation between geography and genetic distance, we hypothesize that geography affects both genetic distance and, via transport costs, trade and that the correlation between trade and genetic distance is spurious. In the next section, we show that: i) the same geographic factors that contribute genetic distance are also important determinants of modern transportation costs; ii) in a standard gravity equation the impact of genetic distance on trade disappears once we introduce transport costs.

A. Data

The bilateral export data are obtained from the United Nations COMTRADE database revised by Feenstra et al. (2005), the time span for our analysis is 1975-2000. Our GDP data are obtained from the World Development Indicator of the World Bank; distance between capitals, common official language and contiguity dummies are obtained from a new dataset compiled at CEPII.¹¹

We use a newly constructed measure of transportation costs.¹² This measure (tc_{ij}) is taken from shipping company quotes collected by Import Export Wizard (IEW), a shipping company providing estimates of transportation costs around the world. IEW calculates the

¹⁰ Note that the effect of insularity is captured by country fixed effects; as a consequence, sharing the same sea is always a strongly unifying factor. In a regression which includes island dummies among the regressors, the sign on island dummies is positive and significant consistent with Sokal's findings.

¹¹ The data on bilateral exports can be found at <http://cid.econ.ucdavis.edu/>. The data on distance, language and contiguity is available at <http://www.cepii.fr/anglaisgraph/bdd/distances.htm>.

¹² For a review of the literature on transport costs see Anderson and van Wincoop (2004).

surface freight data based on a survey of inter-modal and marine tariffs from carriers around the world. The variable tc_{ij} is the cost in U.S. dollars of transporting a “1000 kg unspecified freight type load (including machinery, chemicals, etc.) with no special handling required, using the optimal combination of going through land and water to transport the goods.” The data refers to 2006. The advantage of this measure is that it represents the actual average transport costs and not an indirect measure or proxy, which are often plagued by measurement errors.¹³

We construct a set of measures of geographical barriers using information on sea, mountain chains, and the average elevation of countries. We define a variable (mountains) identifying the number of major mountain chains between countries. According to the World Atlas, major mountain chains in Europe are: the Alps, the Apennines, the Atlantic Highlands (which include the Kjolen in Norway and Sweden, and the Pennines in the UK), the Balkan Mountains, the Massif Central, the Meseta, the Pyrenees, the Urals, the Carpathian Mountains and the Caucasus. We define a dummy “common sea” equal to one if a pair of countries shares the same sea, which can be the Mediterranean, the Atlantic Ocean, or the Northern/Baltic Sea. Finally we construct a variable measuring the average elevation of countries which are between two trading partners. For instance, for the pair Germany-Italy this variable is equal to the average elevation of Germany, Austria, and Italy. This variable measures the difficulty of transportation/migration between countries. The sample statistics for the data are reported in Table A1.¹⁴

B. Determinants of Transportation Costs and Genetic Distance

In this section we analyze whether the measures of geographic distance we have constructed have an impact on transportation costs and genetic distance. The importance of geography in determining transportation costs is well established; however, the standard measure of geography - log distance between (the capitals of) two countries - is considered only a first

¹³ As a robustness check, we replicate our main results using a “matched partner” technique (see Limao and Venables, 2001); the results are discussed below.

¹⁴ An alternative would be the variance of the elevation of the countries. We prefer our measure because the cost of transportation is usually proportional to the average elevation.

rough approximation of transport costs (see Hummels, 1998, Limao and Venables, 2001, and Eaton and Kortum, 2002).¹⁵ A substantial literature in engineering has long observed that topographical characteristics, such as terrain variability, affect transportation costs (see Tsunokawa, 1983, and World Bank, 1987). For example, for the same horizontal distance, moving goods across variable terrain requires more energy and time. Similarly, Limao and Venables (2001) and the World Bank (1998) find that non primary export high performers are island countries and none is landlocked.

To understand the geographic determinants of transport costs, we run the following regression:

$$\log(tc_{ij}) = \alpha_0 + \alpha_1 \log(D_{ij}) + \alpha_2 \text{geography}_{ij} + \alpha_3 \text{country}_i + \alpha_4 \text{country}_j + \varepsilon_{ij}$$

where tc_{ij} are the transport costs, D_{ij} is the distance between country i and country j , *geography* is a vector with our proxies for geographical barriers (number of mountains, common sea, and average elevation of intermediate countries), country_i and country_j control for importer and trading countries fixed effects. The results of our regressions are reported in the last three columns of Table 1 for several combinations of geographical barriers. All variables have the expected sign. (Log)-distance, number of mountain chains, and average elevation between countries increase transportation costs, whereas the presence of a common sea reduces them. In conclusion, the same geographical determinants explain genetic distance and transport costs between countries.

C. Genes versus Geography in Explaining Trade

In this section, we show that, when we instrument transportation costs with those measures of geographical characteristics that influenced also migration patterns in the past (and

¹⁵ See Clark, Dollar, and Micco (2004) for a more nuanced measure of transport costs.

hence largely genetic distance nowadays), the impact of genetic distance on economic outcome (in this case trade) disappears. In order to do so, we estimate the following gravity equation:

$$\ln(X_{ijt}) = \beta_0 + \beta_1 \text{gendist}_{ij} + \beta_2 \ln(Y_{it}) + \beta_3 \ln(Y_{jt}) - \beta_4 \ln(D_{ij}) + \beta_5 C_{ij} + \beta_6 L_{ij} + \beta_7 E_{ijt} + \varepsilon_{ijt}$$

where X_{ijt} is the value of annual exports from country i to country j , gendist_{ij} is the genetic distance between country i and j as defined by Cavalli-Sforza et al. (1994), Y_{it} is the real GDP of country i , D_{ij} is the distance between i and j , C_{ij} is a dummy variable for geographic contiguity between country i and j , L_{ij} is a dummy variable for common language between i and j , E_{ijt} is a dummy equal to one if country i and country j both use the Euro at time t , and ε_{ijt} is the error term. We use a panel from 1975 to 2000, controlling for year and country of origin and country of destination fixed effects. We cluster the standard errors at the bilateral country pair level.

We begin by estimating a standard gravity equation. The first specification (Table 2, column 1) does not include genetic distance. Contiguity, distance and language have all the expected sign. The Euro variable is not significant. The importer's GDP has a negative and significant sign; since we include country of origin and country of destination fixed effects, the negative sign only captures the time variation of the series. As a next step, we introduce genetic distance among the regressors (specification 2). Genetic distance has a negative sign and is significant at the five percent level. From this regression, one would be inclined to infer that cultural distance has a negative impact on trade. Specification (2) could be misleading, since genetic distance could capture the effect of omitted variables (transport costs). We argue that genetic distance is related to trade because it is a proxy for geographic impediments, which increase transportation costs between countries.

To address this issue we introduce transportation costs. We first run an OLS regression; the significance of genetic distance disappears once we introduce transport costs. Column 3 reports the results. However, OLS regression may suffer from endogeneity bias because transport costs depend on the volume of trade (Hummels, 1999). For this reason, we instrument transport costs with a combination of our measures of geography. The IV regressions (Columns

4-5) ¹⁶ confirm that, when we introduce transportation costs, genetic distance is no longer significant, whereas the effect of transportation costs is negative and highly significant. Column 4 includes year and importer and exporter country fixed effects. Column 5 controls in addition for country specific linear trends for both importing and exporting countries.¹⁷ As before, transportation costs appear to be crucial in the determination of trade, while genetic distance is once again not significant.

For all our specifications we run the appropriate tests (see Stock and Yogo, 2002, and Moreira, 2004) to check that our instruments are weakly correlated with transportation costs. The F statistics for the joint hypothesis that the instruments' coefficients are zero in the first stage regression, has always a significance level lower than 5% for all specifications. In addition all the tables report the Hansen J statistics of over-identifying restrictions.

An alternative explanation?

So far, we have shown that the coefficient on genetic distance becomes insignificant once we control for transport costs. A possible explanation of this result could be that countries, which are separated by natural barriers, have developed different genetic make-ups as well as different cultures. In this case, transport costs, cultural differences, and genetic distances would be highly correlated with the consequence that our regressions would suffer from multi-collinearity and the t-statistics on genetic distance could be low without indicating necessarily that genetic distance is irrelevant in explaining trade flows. While the correlation between these variables is not so high to justify the presence of multi-collinearity, we take the issue seriously and propose the following tests based on separating the goods according to the importance of trust and cultural differences (Rauch classification) and according to easiness to transport (our new proposed classification based on bulkiness). The motivation is as follows.

¹⁶ We run several specifications. Table 2 reports our preferred specification, which uses average elevation and the presence of a common sea. We find very similar results using different combinations of our instruments; however the specification reported is the best in terms of the Hansen J statistics.

¹⁷ Anderson and van Wincoop (2003) argue that a term proxying for multilateral resistance should be included on the regressions. We include country dummies and country-specific trend.

If genetic distance is a proxy for cultural differences, countries with less genetically similar populations should trade relatively less in goods for which the role of trust and/or cultural similarities are more relevant. In order to test this hypothesis, we divide the goods into two categories: differentiated products and homogeneous goods following Rauch's classification (Rauch, 1999).¹⁸ Table 3 reports the results for regressions using the sample restricted to these two types of goods. Even after splitting the sample, the results still hold (genetic distance is not significant for differentiated goods and has the wrong sign for the homogeneous type.) Note also that the coefficients (instrumented) on transportation costs are larger and statistically significant for homogeneous goods in line with the idea that homogenous goods are, on average, heavier and more costly to move than other goods (Rauch, 1999).

While the Rauch classification is based on the existence of organized trade and referenced prices, we also propose a new classification based on the easiness to transport.¹⁹ Goods which are easy to transport because the values per weight is very high (for instance, diamonds, gold, electronics) should not be sensitive to transport costs. On the opposite, bulky goods which are difficult to move should be more sensitive to transport costs. On the other hand, both "easy to move" and "bulky" goods should be equally sensitive to genetic distance. We construct the index of bulkiness by looking at the freight to value ratio for the US imports from Mexico and Canada at 4 SITC digits.²⁰ We consider bulky all goods whose freight to value ratio is higher than the median, the other half of the goods is classified easy to transport. Table 4 reports the results of our basic specification for easy-to-transport and bulky goods; as before, transport costs are instrumented for using our geographical variables. As expected, the

¹⁸ Homogeneous goods are goods traded on organized exchanges, reference priced goods are goods not traded on organized exchanges but nevertheless possessing a reference price, differentiated goods are other commodities (Rauch, 1999). Over 60 percent of the value is in differentiated goods; about 20 percent belongs to reference priced goods, and the remaining to the organized trade.

¹⁹ We thank David Hummels for useful discussion on this particular index and for providing us the data to construct it.

²⁰ We use US data because they give detailed information on freight rate and values at 4 digit level. We choose import from Mexico and Canada because, being contiguous countries, all modes of transportation are used, including sea, land, and air. We make the assumption that the ranking of this ratio is the same in North America as in Europe.

coefficient on transport costs is large and highly significant for bulky goods while is not significant for easy-to-transport goods. Genetic distance is always insignificant.

In conclusion, splitting the samples between goods with organized markets or according to the easiness to transport leads to the same results: transport costs are significant for bulky goods and/or homogeneous goods while genetic distance is never significant where we expected it to be significant, i.e. in goods in which cultural differences may be relevant. We take this as a strong indication that, even though genetic distance, geographical barriers, and cultural differences are correlated, trade is explained mostly by transport costs.

Robustness checks

Given that we use a new database on transport costs, which was never used before in economics to our knowledge, it is legitimate to ask if our results are driven by the peculiarity of our new data. In order to confirm that our results hold also with standard, if very imprecise, measures of transport costs, we repeat the same regressions using indirect transport costs.

We construct indirect transport costs using a “matched partner” technique (see Limao and Venables, 2001). Exporting countries report trade flows exclusive of freight and insurance (*fob*) and importing countries report flows inclusive of freight and insurance (*cif*). The ratio between *cif* and *fob* can then be used to construct an indirect measure of transportation costs, i.e.

$itc_{ijt} = 1 - \frac{cif_{ijt}}{fob_{ijt}}$. The raw data come from the IMF’s Direction of Trade Statistics from 1975 to

2002. Following Hummels (forthcoming), we restrict our analysis to *ad valorem* transportation costs which lie in a range between 0 and 100 percent, considered a reasonable range of variation.

The results are reported in Table A3 and are very similar to the main specification. Genetic distance appears to be highly significant in the OLS regression, however its impact on trade is nil once transportation costs are introduced.

Differences in income and genetic distance

A very intriguing paper by Spolaore and Wazciarg (2006) shows that income and genetic differences between countries are systematically correlated. The authors interpret the results observing that genetic distance is a proxy for vertically transmitted cultural traits. According to the authors, the transmission of technological improvements, which are ultimately the cause of advancement of income per capita, would be easy between populations which share similar backgrounds.

We regress log-differences in per capita income for 1998 on genetic distance and we also find that genetic distance is a good predictor of bilateral income differences with a significance of 10 percent (Table 6, column 1), confirming Spolaore and Wazciarg's (2006) results.²¹ However, when we include geographical variables (average elevation between countries, and common sea), the coefficient on genetic distance loses significance.

Our results confirm Spolaore and Wazciarg's (2006) observation that barriers between countries are systematically correlated to differences in income with an important qualification: these results are due to geographical impediments between countries and there is no evidence that vertically transmitted characteristics as proxied by genetic distances play any role.

IV. POSSIBLE USES OF GENETIC DISTANCE IN ECONOMICS

In this section we discuss possible uses of genetic distance in economics. From the previous discussion it appears that the first obvious use of genetic distance is an instrument for transportation costs. Being exogenous and determined by geographical barriers, genetic distance could be simply used as an indirect measure for geographical impediments. In Table 5, we run a gravity equation using genetic distance as an instrument for transportation costs. As expected the coefficient on transportation costs is negative and significant.²² The first stage of this

²¹ Our regressions differ from Spolaore and Wazciarg (2006) in one important respect. We consider only Europe and not the entire world in order to avoid arbitrary decision of how to calculate genetic distance between countries and because we have much more information on geographic features for Europe. In line with Spolaore and Wazciarg (2006), we control for several measures of geography (including absolute differences in latitude and longitude), a dummy for contiguity, and the presence of a common language. In addition, we control for and country of origin and destination fixed effect.

²² As before, Column 1 controls for country of origin and country of destination fixed effects, whereas Column 2 adds country specific linear trends.

regression is reported in Table A4. As before, in the IV specification, not only the effect of transportation costs remains significant, but its magnitude increases substantially.

Besides being a proxy for geographical impediments, being a heritage of the past, a population's genetic make-up can also provide valuable insights about the history of a country. Inasmuch ancient history is believed to be an important determinant of current economic outcomes, genetics can add useful information. For instance, the genetic composition of a country, which was formerly a colony, can inform us on how much intermingling there was between colonizers and natives. Little intermingling will suggest that the colonizers' culture was simply imposed while more intermingling will indicate that population transfers also occurred.²³ Even more specifically, analysis of mitochondrial DNA and chromosome Y analysis can unveil the type of intermingling – if just male colonizers contributed or if whole nuclear families moved.

The genetic make-up of a population could also provide information on past environmental conditions, including the endemic diffusion of diseases such as malaria or the dietary habits, which shaped and acted upon specific phenotypes. Note that in this case, the measure of genetic distance proposed by Cavalli-Sforza, which is intentionally based on genes that likely do not affect phenotypes subject to natural selection, is not appropriate. Some of these applications are left for future research.

V. CONCLUSIONS

This paper analyzes the validity of genetic distance as an instrument or proxy for cultural distances. The use of genetic distance, which has been introduced in anthropology and historical linguistics by the pioneering work of Cavalli-Sforza, has been proposed also in economics by at

²³ We credit Simon Johnson for this specific idea. Note that the idea that genetic analysis is a good instrument to understand the pattern of cultural spreading was the original motivation of Cavalli-Sforza et al. (1994). During the Neolithic revolutions in Europe new artifacts or techniques (or cultures) appeared to spread westwards. However, it is unclear if this spreading of culture was brought about by a spread of population or not. Cavalli—Sforza has argued that genetic analysis is compatible with the hypothesis of migration. Similarly, the transfer of techniques from colonizers to colonies may have happened with or without population intermingling and the subsequent economic outcome and institutions may depend on the way techniques have spread.

least two papers (Guiso et al., 2004, and Spolaore et al., 2005). We think this application is inappropriate because genetic distance as it is not measuring the cultural variables which are of interest for economists.

Our starting point is the observation that the same geographical barriers which shaped genetic patterns play a role in contemporary transport costs. We confirm this intuition with simple regressions. We have taken the case of European trade for which data on genetic distance are uncontroversial and there is considerable (but not total) overlapping between genetic population and countries. If geographical barriers play a role for both genetic distance and modern day transport costs, the variable genetic distance in a trade equation simply captures an omitted variable. We show the point by presenting gravity equations with genetic distance and with or without transport costs. As expected, the coefficient on genetic distance becomes insignificant once transport costs are introduced.

We check the robustness of our results by repeating our regression restricting the sample to goods with an organized market (using Rauch classification) and for easiness to transport (using a new classification). The results are confirmed: bulky and homogeneous goods are particularly sensitive to transport costs while genetic distance does not play any role even in differentiated goods.

Finally, we consider the role of genetic distance in explaining income differences between countries and show that, once we properly control for geographic impediments, genetic distance is not a significant determinant of income differences. Our results are then not only limited to trade.

Our main conclusion is that genetic distance is a proxy for geographical differences and not for the cultural differences which are of interest to the economists. This is not to say that cultural variables are not relevant for economic outcomes, our results only show that genetic distance is not a good instrument or proxy for culture.

References

- Acemoglu, Daron, Simon Johnson and James A. Robinson, 2001, "The Colonial Origins of Comparative Development: and Empirical Investigation," *American Economic Review*, Vol. 91 No. 5 (December).
- Alesina, A., W. Easterly and J. Matuszeski (2006), "Artificial States", NBER WP 12328
- Akey, J. M., K. Zhang, L. Jin, and M.D. Shriver, 2002, "Interrogating a High-Density SNP Map for Signatures of Natural Selection," *Genome Research*, Vol. 12, pp. 1805-14.
- Anderson, J. and E. van Wincoop, 2003, "Gravity with Gravitas: a Solution to the Border Puzzle," *American Economic Review*, Vol. 93 No. 1, pp. 170-92.
- , 2004, "Trade Costs," *Journal of Economic Literature*, Vol. 42 No. 3, pp. 691-751.
- Cavalli-Sforza, L. L., and M. W. Feldman, 2003, "The application of Molecular Genetic Approaches to the Study of Human Evolution," *Nature Genetic Supplement*, Vol. 33 pp. 266-75.
- Cavalli-Sforza, L., P.Menozzi, and A. Piazza, 1994, *The History and Geography of Human Genes*, (Princeton, New Jersey: Princeton University Press).
- Eaton, J. and S. Kortum, 2002, "Technology, Geography and Trade," *Econometrica*, Vol. 70 No. 5, pp. 1741-79.
- Clark, X., D. Dollar, and A. Micco, 2004, "Port Efficiency, Maritime Transport Costs, and Bilateral Trade," *Journal of Development Economics*, pp. 417-50.
- Feenstra, R., R. Lipsey, H. Deng, A. Ma, H. Mo, 2005, "World Trade Flows: 1960-2000," NBER Working Paper 11040, (Cambridge, Massachusetts: National Bureau of Economic Research).
- Gallup, John Luke, and Jeffrey Sachs D., with Andrew D. Mellinger, 1998, "Geography and Economic Development," in *Annual World Bank Conference on Development Economics*, (Washington: The World Bank).
- Goldstein, D. B., and L. Chikhi, 2002, Human Migrations and Population Structure: What We Know and Why it Matters," *Annual Reviews Genomics Hum Genet*, Vol. 3 pp. 129-52.
- Guiso, L., Sapienza, P. and L. Zingales, 2005, "Cultural Biases in Economic Exchange," NBER Working Paper No. 11005, (Cambridge, Massachusetts: National Bureau of Economic Research).

- , 2006, “Does Culture Affect Economic Outcomes?” *Journal of Economic Perspectives*, Vol. 20, No. 2, pp. 23-48.
- Hofstede, G., 1980, *Culture’s Consequences: International Differences in Work-Related Values*, (Beverly Hills: Sage Publication).
- Hofstede, G., and R. McCrae, 2004, “Personality and Culture Revised: Linking Traits and Dimension of Culture,” *Sage Publications-Cross Cultural Research*
- Hummels, David, 1998, “Towards a Geography of Transport Costs,” Mimeo, (Chicago: University of Chicago, Department of Economics).
- Hummels, David, and V. Lugovskyy, 2006, “Usable Data? Matched Partner Trade Statistics as a Measure of Transportation Cost,” Forthcoming, *Review of International Economics*.
- Kimura, M., 1968, “Evolutionary Rate at the Molecular Level,” *Nature*, No. 217, pp. 624-26.
- Lamason, R. L., Mohideen, M. A., Mest, J. R., *et al.*, 2005, “A Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans Science,” SLC24A5 No. 310, pp. 1782-86.
- Lewontin, R.C., Rose L., and Kamin L., 1984, *Not in Our Genes: Biology, Ideology and Human Nature*, (New York: Pantheon Books).
- Limao, N. and Venables, A., 2001, “Infrastructure, Geographical Disadvantage, Transport Costs and Trade,” *World Bank Economic Review*, Vol. 15, pp. 451-79.
- McCrae, R. and J. Allik, 2002, *The Five-Factor Model of Personality Across Cultures*, (New York: Kluwer Academic/Plenum Publisher).
- Putnam, R., 1993, *Making Democracy Work*, (Princeton, New Jersey: Princeton University Press).
- Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., and M. W. Feldman, 2005, “Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure,” *PLoS Genet* No. 1, p. 70.
- Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi, 2002, “Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development,” NBER Working Paper 9305, (Cambridge, Massachusetts: National Bureau of Economic Research).
- Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L.A. Zhitovskyy, and M.W. Feldman, 2002, “Genetic Structure of Human Populations,” *Science* No. 298, pp. 2381-85.

- Sachs, Jeffrey, 2003, "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income," [NBER Working Paper 9490](#).
- Soejima, M., H. Tachida, T. Ishida, A. Sano, and Y. Koda, 2006, "Evidence for Recent Positive Selection at the Human AIM1 Locus in a European Population," *Mol Biol Evol*, No. 23, pp. 179-88.
- Spolaore, E. and R. Wacziarg, 2005, "The Diffusion of Development," mimeo
- Terraciano, A., et al., 2005, "National Character Does Not reflect Mean Personality Trait Levels in 49 Countries," *Science*, Vol. 310, (October), pp. 96-100.
- Weber, M., 1958, *The Protestant Ethic and the Spirit of Capitalism*, Translated by T. Parsons, (New York: Scribner's Library).
- Zerjal, T., et al., 2001, "Geographical, Linguistic and Cultural Influences on Genetic Diversity: Y-Chromosomal Distribution in Northern European Populations," *Molecular Biology and Evolution*, Vol. 18 No. 6, pp. 1077-87

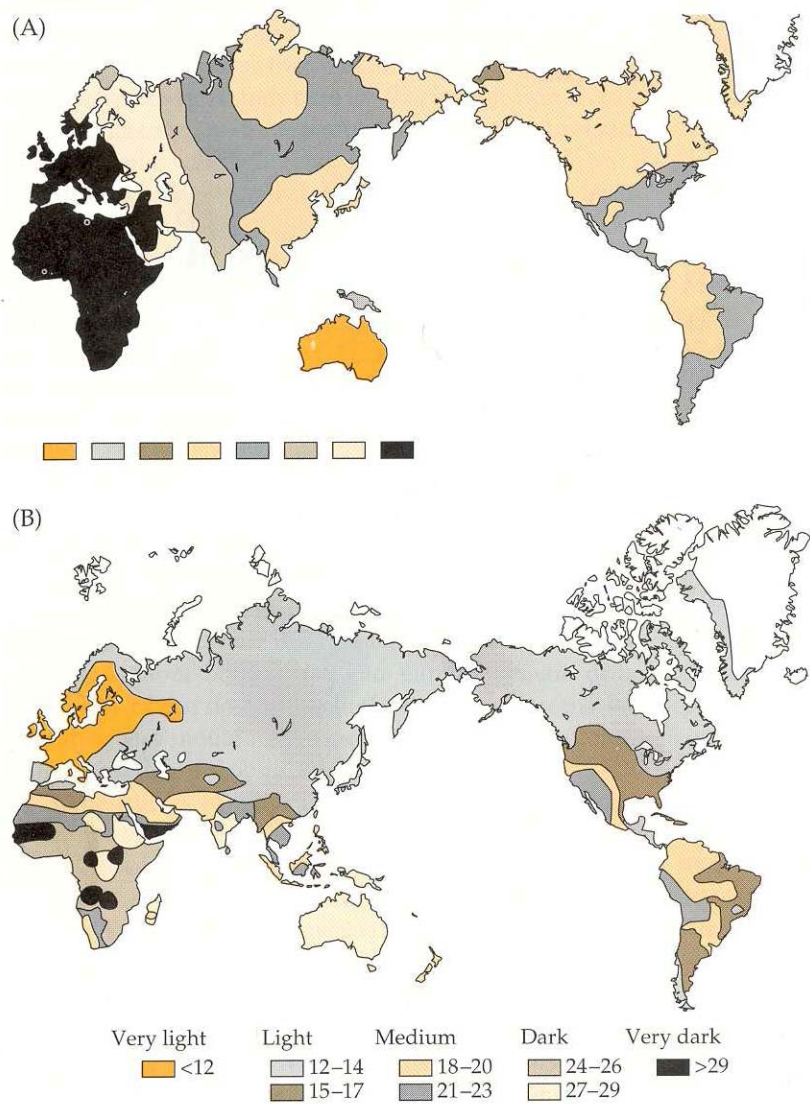
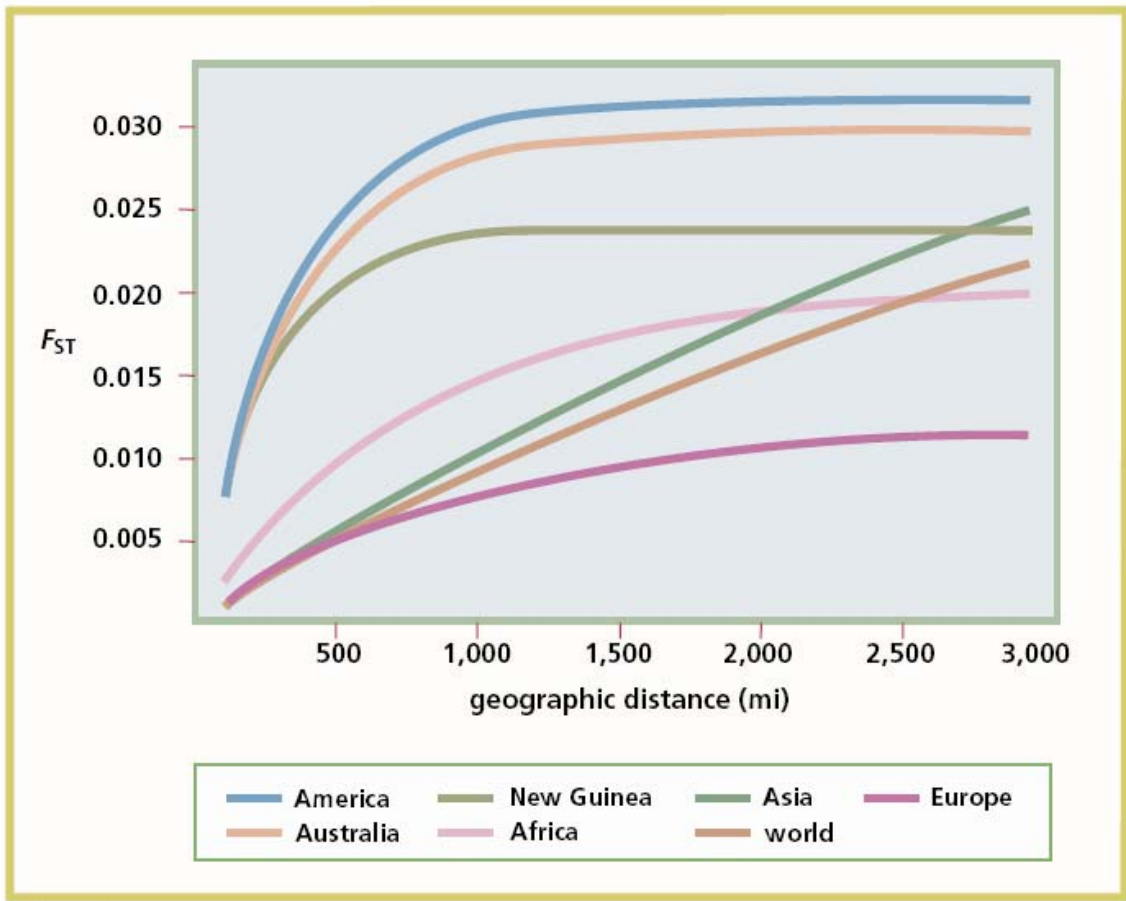


Figure 1. A. Division of the world's human population into eight classes of genetic similarity, based on overall difference and similarity at numerous enzyme and blood-group loci. The eight classes represented are arrayed in order of increasing difference. B. Geographic distribution of skin color, classified in eight grades of pigmentation intensity (adapted from Cavalli Sforza et al., 1994)



BOB CRIMI

Figure 2

Relationship between Genetic and Geographic Distance (Cavalli-Sforza and Feldman, 2003)

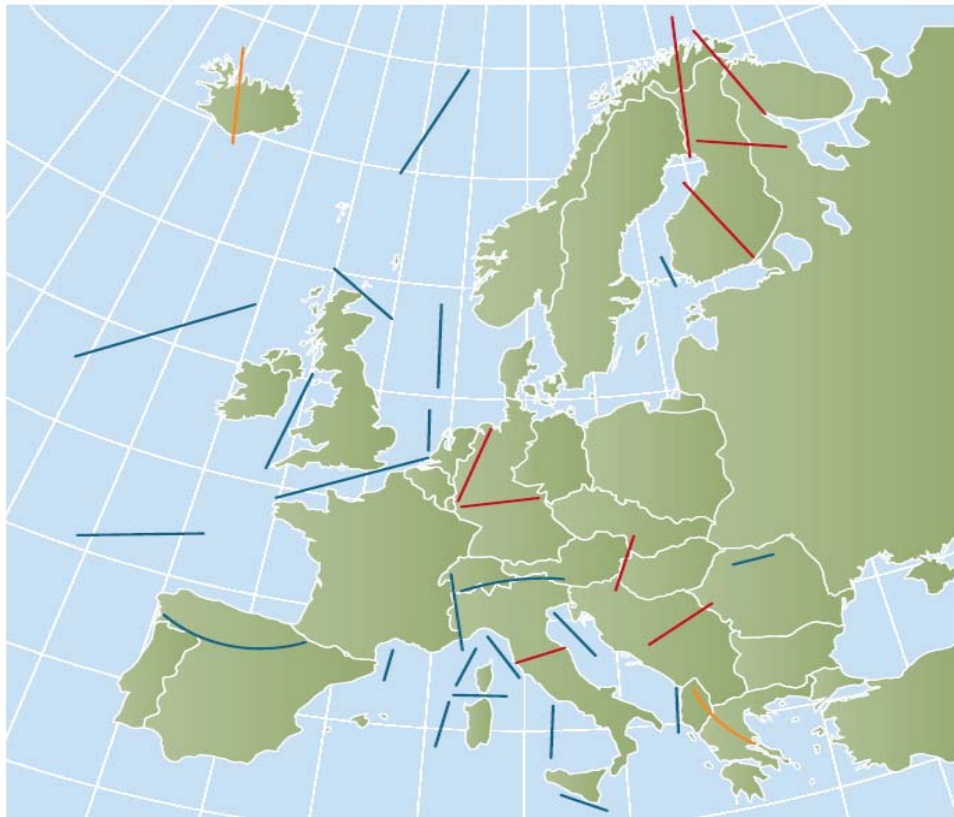


Figure 3

Zones of sharp genetic changes in Europe (Sokal, 1990)

Table 1.
Geographic Determinants of Genetic Distance and Transportation Costs

	Genetic Distance	Genetic Distance	Genetic Distance	$\ln(tc_{ij})$ Transp. cost	$\ln(tc_{ij})$ Transp. cost	$\ln(tc_{ij})$ Transp. cost
	(1)	(2)	(3)	(4)	(5)	(6)
Log (distance)	0.022*** (.0031)	.0173*** (.0037)	.0173*** (.0037)	.0871*** (.0036)	.0880*** (.0067)	.0856*** (.0039)
Number of mountain chains	.0010 (.0024)		-.0008 (.0025)	.0079*** (.0013)		.0069*** (.0014)
Average elevation between countries		.0414** (.0214)	.0441** (.0225)		.0381*** (.0133)	.0216 (.0139)
Common sea	-.0280*** (6.723)	-.0238*** (.0075)	-.0237*** (.0075)	-.0338*** (.0013)	-.0336*** (.0042)	-.0322*** (.0041)
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of dest. fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	470	470	470	1332	1332	1332
R-squared	0.95	0.95	0.95	0.93	0.93	0.93

* significant at 10%; ** significant at 5%, *** significant at 1%. Robust standard errors in parentheses

Table 2 Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Exports

	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>	<i>IV</i>	<i>IV</i>
	(1)	(2)	(3)	(4)	(5)
Transportation costs			-4.547*** (1.204)	-28.62** (11.47)	-27.92** (11.22)
Genetic Distance		-3.135** (1.537)	-1.047 (1.498)	10.02* (5.416)	10.08* (5.321)
Common Language	.7031*** (.1776)	.6708*** (.1652)	.7501*** (.1464)	1.169*** (.3118)	1.152*** (.3038)
Euro	.0921 (.0666)	.0800 (.0654)	.0812 (.0650)	.0880 (.0710)	.0254 (.0752)
Log (distance)	-.9623*** (.0881)	-.8531*** (.0941)	-.4499*** (.1306)	1.684* (1.005)	1.626* (.9789)
Contiguity	.1745 (.1153)	.2026* (.1157)	.2241** (.1112)	.3380* (.1845)	.3516** (.1785)
GDP exporter	.6959*** (.2307)	.7037*** (.2289)	.7248*** (.2281)	.8363*** (.2427)	.4468* (.2357)
GDP importer	-1.090** (.5181)	-1.080** (.5155)	-1.055** (.5164)	-.9225* (.5240)	.4109 (.3612)
Fixed year effects	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes
Country spec. linear trend	No	No	No	No	Yes
Observations	9408	9408	9408	9408	9408
Hansen J statistic				.6774	.4847
R-Squared	0.85	.85	.86	.99	.99

Significant at 10, %, ** significant at 5%, *** significant at 1%. Errors are clustered at the bilateral pair level. Instruments include average elevation and common sea.

Table 3. Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Exports

	Differentiated Goods				Homogeneous Goods			
	<i>OLS</i> (1)	<i>OLS</i> (2)	<i>IV</i> (3)	<i>IV</i> (4)	<i>OLS</i> (1)	<i>OLS</i> (2)	<i>IV</i> (3)	<i>IV</i> (4)
Transp. costs		-4.759*** (1.207)	-10.52 (6.841)	-9.562 (6.641)		-5.888*** (1.598)	-38.876** (16.31)	-39.70** (16.56)
Genetic Distance	-2.924* (1.583)	-.7781 (1.532)	1.816 (3.360)	1.804 (3.288)	-3.261 (2.117)	-.5897 (2.218)	14.379** (7.449)	14.86** (7.569)
Common Language	.6537*** (.1570)	.7361*** (.1411)	.8359*** (.1795)	.8111*** (.1747)	.7045*** (.2618)	.8085*** (.2378)	1.3916*** (.4288)	1.400*** (.4348)
Euro	.0746 (.0710)	.0757 (.0705)	.0771 (.0702)	.0445 (.0622)	.1230 (.0827)	.1246 (.0827)	.1335 (.0940)	-.0543 (.1122)
Log (distance)	-.7018*** (.0910)	-.2795** (.1290)	.2314 (.5907)	.1525 (.5713)	-1.241*** (.1292)	-.7279*** (.1755)	2.150 (1.405)	2.236 (1.423)
Contiguity	.3252*** (.1248)	.3468*** (.1217)	.3730*** (.1290)	.3869*** (.1239)	.2981** (.1565)	.3156** (.1518)	.4139* (.2487)	.4278* (.2509)
GDP exporter	.7783*** (.2655)	.8002*** (.2647)	.8268*** (.2617)	.3443 (.2405)		.4393 (.3011)	.5823* (.3315)	.4354 (.3612)
GDP importer	-.9610* (.5066)	-.9342* (.5073)	-.9017* (.5026)	.7244* (.3904)		.4994 (.3461)	.5749 (.3738)	-.5683 (.4195)
Fixed year effects	Yes	Yes	Yes	Yes		Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes		Yes	Yes	Yes
Country specific time effects	No	No	No	Yes		no	No	No
Observations	9363	9363	9363	9363	8981	8981	8981	8981
Hansen J Statistics			.3879	.1767			.1150	.1022
R-Squared	.87	.93	.99	.99	.99	.99	.99	.99

* significant at 10%; ** significant at 5%, *** significant at 1%. Standard errors are clustered at the bilateral pair level.

Table 4
Genetic Distance, Transportation Costs and Trade for “Easy to Transport” and “Bulky” Goods
Dependent Variable: Log Total Exports

	Easy to Transport Goods			Bulky Goods		
	<i>OLS</i> (1)	<i>OLS</i> (2)	<i>IV</i> (3)	<i>OLS</i> (1)	<i>OLS</i> (2)	<i>IV</i> (3)
Transportation costs		-6.081*** (1.385)	-7.133 (7.798)		-4.259*** (1.189)	-34.45*** (13.42)
Genetic Distance	-2.171 (1.610)	.4906 (1.598)	-.1243 (.6852)	-.8820*** (.0994)	-1.650 (1.578)	12.11* (6.198)
Common Language	.5272*** (.1699)	.6365*** (.1519)	.6554*** (.2010)	.7341*** (.1743)	.8098*** (.1575)	1.346*** (.3820)
Euro	.1138 (.0787)	.1149 (.0782)	.1151 (.0777)	.0897 (.0604)	.0903 (.0599)	.0949 (.0700)
Log (distance)	-.7606*** (.0957)	-.2181 (1.506)	-.1243 (.6852)	-.8820*** (.0994)	-.5022*** (.1324)	2.189* (1.186)
Contiguity	.1733 (.1347)	.1957 (.1322)	.1996 (.1347)	.2425** (.1189)	.2619** (.1151)	.3988* (.2130)
GDP exporter	.1284 (.333)	.1597 (.3301)	.1651 (.3315)	-.8695** (.4417)	-.8439** (.4425)	-.6622 (.4582)
GDP importer	.9140*** (.3118)	.9515*** (.3103)	.9580*** (.3107)	.7151*** (.2211)	.7394*** (.2198)	.9115*** (.2460)
Fixed year effects	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country spec. linear trend	No	No	No	No	No	No
Observations	9290	9290	9290	9498	9498	9498
Hansen J statistic			.59			.61
R-Squared	.88	.88	.99	.85	.85	.99

significant at 10, %, ** significant at 5%, *** significant at 1%. Errors are clustered at the bilateral pair level. Instruments include average elevation and common sea.

Table 5
Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Exports
Instrumenting Transportation Costs with Genetic Distance

	<i>IV</i> (1)	<i>IV</i> (2)
Transportation costs (direct transport cost)	-7.595** (3.265)	-6.737** (3.150)
Common Language	.7833*** (.1362)	.7622*** (.1380)
Euro	.0361 (.0664)	.0223 (.0659)
Log (distance)	-.1730 (.3263)	-.2397 (.1096)
Contiguity	.2437** (.1147)	.2593** (.1096)
GDP exporter	.8076*** (.2132)	.6507*** (.1989)
GDP importer	-.7772* (.4565)	.3753 (.3288)
Fixed year effects	Yes	Yes
Country dummies	Yes	Yes
Country spec. linear trend	No	Yes
Observations	10348	10348
R-Squared	.99	.99

* significant at 10%; ** significant at 5%, *** significant at 1%.
Standard errors are clustered at the bilateral pair level.

Table 6
Absolute Value of Log Income Differences, 1998

	<i>OLS</i> (1)	<i>OLS</i> (2)
Genetic Distance	1.462* (.8058)	1.196 (.7561)
Common Language	-.0734 (.0756)	-.0602 (.0792)
Log(distance)	-.1126 (.0910)	-.1218 (.0931)
Absolute Difference in Latitudes	.0112* (.0063)	.0112* (.0059)
Absolute difference in Longitudes	.0060 (.0039)	.0074* (.0040)
Contiguity		.0089 (.0644)
Average elevation between countries		-.0003 (.0002)
Common sea		-.1286** (.0603)
Country fixed effects	Yes	Yes
Observations	233	233
R-Squared	.95	.95

* significant at 10%; ** significant at 5%, *** significant at 1%.

Table A1
Summary Statistics

	Mean	Std. Dev.	Observations
Log Distance	7.134	.6365	1480
Contiguity	.0959	.2946	1480
Common Language	.0135	.1154	1480
Genetic Distance	99.22	115.24	470
Islands	.1277	.3338	1480
Landlocked	.1540	.3611	1480
Euro	.0235	.1517	26460
Number of mountains	.9202	.9420	1480
Number of shared rivers	.0770	.2955	1480
Average elevation b/w countries	159.08	122.04	1474
Common sea	.3594	.4800	1480
Log(GDP) Exporter	25.22	1.85	22280
Log(GDP) Importer	25.22	1.85	22280
Log(Total Exports)	11.29	2.74	25872
Log Transp. Costs (direct transport costs)	5.27	.1603	1332

Table A2
IV Regression, First Stage
Geographical Barriers as a Measure for Transportation Costs

	Log (Transp. Costs)
Average Elev. b/w countries	.0458** (.0212)
Common sea	-.0052 (.0046)
Common language	.0161* (.0092)
Euro	.0003 (.0012)
Genetic Distance	.4167*** (.0453)
Log(distance)	.0826*** (.0044)
Contiguity	.0059 (.0062)
GDP exporter	.0043 (.0029)
GDP importer	.0051 (.0062)
Observations	9408
R-squared	.92

* significant at 10%; ** significant at 5%, *** significant at 1%.
Standard errors clustered at the bilateral country pair level.

Table A3
Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Exports

	<i>IV</i> (1)	<i>IV</i> (2)	<i>IV</i> (3)
Transportation costs (indirect transport costs)	-1.492** (.6295)	-1.348** (.5977)	-2.884*** (.6578)
Genetic Distance			-2.082 (2.271)
Common Language	.6616** (.2918)	.6547** (.2706)	.8518** (.4168)
Euro	.0159 (.0831)	-.0716 (.0937)	-.0404 (.1092)
Log (distance)	-.4416* (.2585)	-.4726** (.2422)	
Contiguity	.2125 (.2708)	.2279 (.2479)	
GDP exporter	.7207*** (.2324)	.3486 (.2321)	.7257*** (.2596)
GDP importer	-1.151*** (.5295)	.1253 (.3785)	-1.129** (.5420)
Fixed year effects	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes
Country spec. linear trend	No	Yes	No
Observations	9095	9095	9095
Hansen J statistic	.2076	.3498	.1198
R-Squared	.99	.99	.98

significant at 10, %, ** significant at 5%, *** significant at 1%. Errors are clustered at the bilateral pair level. Instruments include numbers of mountains, average elevation and common sea. Column 3 also includes log(distance) and contiguity as instruments.

Table A4
IV Regression, First Stage
Genetic Distance as an Instrument for Transportation Costs

Genetic Distance	.4532*** (.0478)
Common Language	.0177* (.0096)
Euro	.0002 (.0013)
Log(Distance)	.0892*** (.0042)
Contiguity	.0044 (.0060)
GDP exporter	.0045 (.0028)
GDP importer	.0050* (.0030)
Observations	10348
R-squared	.92

* significant at 10%; ** significant at 5%, *** significant at 1%.
Standard errors are clustered at the bilateral pair level.