Stability of Nations and Genetic Diversity*

Klaus Desmet† Michel Le Breton‡ Ignacio Ortuño-Ortín§ Shlomo Weber¶

July 2008

Abstract

This paper presents a model of nations where *culturally heterogeneous* agents vote on the optimal level of public spending. Larger nations benefit from increasing returns in the provision of public goods, but bear the costs of greater cultural heterogeneity. This tradeoff induces agents' preferences over different geographical configurations, thus determining the likelihood of secession or unification. We provide empirical support for choosing genetic distances as a proxy of cultural heterogeneity and by using data on genetic distances, we examine the stability of the current map of Europe. We then identify the regions prone to secession and the countries that are more likely to merge. Furthermore, we estimate the welfare gains from European Union membership.

JEL Classification Codes: H77, D70, F02, H40.
Keywords: nation formation, genetic diversity, cultural heterogeneity, secession, unification, European Union

---

*We thank Lola Collado, Andrés Romeu, Christian Schultz and Romain Wacziarg for helpful comments. Financial aid from the Spanish Ministerio de Educación y Ciencia (SEJ2005-05831), the Fundación BBVA 3-04X and the Fundación Ramón Areces is gratefully acknowledged.

†Universidad Carlos III, Getafe (Madrid), Spain, and CEPR.

‡Université de Toulouse I, GREMAQ and IDEI, Toulouse, France.

§Universidad Carlos III, Getafe (Madrid), Spain

¶Southern Methodist University, Dallas, USA, CORE, Catholic University of Louvain, Louvain-la-Neuve, Belgium, and CEPR.

1 Introduction

Recent decades have witnessed large-scale map redrawing. Some countries, such as the Soviet Union and Yugoslavia, have broken up, while others have moved towards much closer cooperation, as epitomized by the European Union, and to a lesser extent, ASEAN (Association of South East Asian Nations). The theoretical arguments suggest that the size of a nation is determined by the trade-off between scale economies that benefit larger nations and the costs of population heterogeneity that in general favor smaller countries (see, e.g., Alesina and Spolaore, 1997, 2003, and Bolton and Roland, 1997). The natural question is whether this theory has an empirical support. Although there is some indirect evidence supportive of this tradeoff,[1] a direct empirical exploration has so far been lacking. Our paper addresses this shortcoming.

In our theoretical model with multiple regions and countries, agents in every country vote on the optimal level of public spending while taking into account increasing returns in the provision of public goods. However, the utility derived from public goods is decreasing in the country's degree of cultural heterogeneity. Assuming that the tax rate in every country is chosen by majority voting, we compare regional welfare across different political arrangements. In particular, we can study whether regions or countries would like to unite or secede. The stability concept we use requires that the majority of agents in every region affected by such a move should support the rearrangement (see Alesina and Spolaore, 1997).

Since the main goal of the paper is the empirical investigation of countries' stability, the most crucial issue in linking our model to the data is the empirical measurement of cultural heterogeneity. We accept the view that a degree of mixture between two populations over the course of history is positively correlated with the similarity of their cultural values. Since populations that have experienced more mixing — or populations that have become separated more recently — are closer genetically, there should be a positive correlation between genetic and cultural distances. We therefore use genetic distances amongst populations as a proxy for cultural distances.[2] Note that

[1] Alesina, Spolaore and Wacziarg (2000) uncover a positive relation between the number of countries and the degree of trade openness, and Alesina and Wacziarg (1998) find that smaller countries are more open. These results are consistent with trade liberalization making it easier for ethnic and cultural minorities to secede.

[2] See Guiso, Sapienza and Zingales (2005) and Spolaore and Wacziarg (2006) for applications of genetic distances to economics.

we view genetic distances as a record of mixing, and *not* as an indicator of the link between genes and human behavior.

To provide an empirical support to our study of nations' stability we examine the current map of Europe.[3] Specifically, we address the following questions: Is the current map of Europe stable? What are the weak links in that map? Which regions are more likely to secede? Which countries stand a better chance to cooperate and possibly unite? Who gains and who loses from the formation of the European Union? By determining the parameter values that are consistent with our stability concept, we identify regions that are more likely to separate, and countries that are more likely to unite. By raising the perceived cost of cultural heterogeneity component in regions' utility functions, we find out that the Basque Country would be the most likely region to break away from the the existing country. Likewise, by decreasing the perceived cost of cultural heterogeneity, we search for most likely candidates for the unification and show that Denmark and Norway, two neighbors united from the Middle Ages until 1814, are on the top of the list. It interesting to point out that the other pairs that stand to gain most from unification — Austria-Switzerland, Belgium-Netherlands, and Norway-Sweden — fit the same pattern: relatively similar population size, similar levels of GDP per capita, and genetic similarity.[4] More generally, we find that unions between large and small countries are unlikely, as the larger country would not find it fiscally attractive. However, if we were to abstract from size differences, a union between, say, Germany and Switzerland, would become beneficial. As expected, unions between countries with very different levels of GDP per capita are quite unlikely, since the implied transfers would not be of interest to the richer one.

Another issue we explore is the gains from European Union membership. The idea is to view the European Union as a new country formed by the merger of previously independent nations. This allows us to examine who gains and loses from being a member of the union. Depending on the methodology used, Ireland or Portugal come out as the winners, whereas Germany always comes in last, being the only country that loses from being a part of the EU. Of course, these findings

[3] This setup could easily be applied to other regions of the world.

[4] In fact, two of these pairs were united for a part of their history. Belgium and the Netherlands were united from 1384 to 1581, and briefly again after the Treaty of Waterloo, from 1815 to 1830. Sweden and Norway were united in the 14th century and were under the same crown from 1814 to 1905.

are not only the result of cultural heterogeneity. A country's size and level of development also matter.[5] Larger countries already reap much of the benefits from increasing returns, and richer countries would lose from redistribution in a union with poorer bedfellows.

But in some cases culture clearly matters. Compare Belgium and Greece, two countries with a population of about 10 million each. Since GDP per capita is 60% higher in Belgium, one would expect Greece to gain more from EU membership than Belgium. That is indeed the case if cultural differences are ignored. However, once cultural heterogeneity is taken into account, this situation is reversed. This is because in our data set Greece is the country which is culturally most distant from the rest of the EU, whereas Belgium is the "genetic capital" of the Union.

Given the importance of cultural heterogeneity in our framework, we now turn to discussing some theoretical and empirical issues related to this notion. To model cultural heterogeneity, we rely on a matrix of cultural distances between nations. We refer to this measure as *metric heterogeneity*. Preferences are such that, all else equal, an agent prefers to be part of a nation which minimizes cultural distances. In other words, each agent ranks nations on the basis of how culturally distant they are. The notion of metric heterogeneity we employ is similar to the one described in the literature on cooperative games where players are characterized by their location in a network or in a geographical space. In such a framework the gains from cooperation increase when the distances among the players in the coalition decrease. Le Breton and Weber (1995) focus on the case where two-person coalitions may form and characterize the patterns for which there is a stable group structure. In contrast to their work, we do not allow for unlimited monetary lump sum transfers among players in the group.

Instead of relying on genetic distances as a proxy for cultural distances, an alternative would be to use data from social surveys on individuals' values. However, the answers to many questions in opinion polls are arguably biased by short term events, such as the political business cycle. Since we are interested in long-term decisions — secessions or unifications — information gathered from surveys or opinion polls may not be the most appropriate. Nevertheless, we do explore this type of information, and find a strong correlation between distances based on social surveys and genetic distances. We view this result not as an argument for an extensive use of opinion polls, but rather

[5] We base our analysis on the European Union with 15 members, before the last two enlargements.

as lending support to the view that genetic distances are a reasonable proxy for cultural distances.

In addition to genetic distances or social surveys, geographical or linguistic distances may capture the same type of information. Indeed, the relation between genes, languages and geography has been extensively studied in population genetics (see, e.g., Sokal, 1987, and Cavalli-Sforza et al., 1994). However, even after controlling for languages and geography, we find that populations that are similar in genes tend to give more similar answers to opinion polls.[6]

Needless to say, our main assumption — more population mixing implies smaller cultural differences — is open to debate. Some authors claim that mixing is not necessary for cultural diffusion to happen (see, e.g., Jobling et al., 2004). It might be the case that, say, Danes have not mixed much with Germans in the last 20 generations, so that there genetic distance is relatively large. However, cultural diffusion might have taken place through books, newspapers, the education system, religion, etc., making their preferences quite similar. The question of whether the transmission of culture takes place through migration flows and the mixing of populations (demic diffusion) has generated yet another debate in population genetics. Cavalli-Sforza et al. (1994) and Chikhi et al. (2002) have argued for the dominant role played by demic diffusion. Their view has been supported by Spolaore and Wacziarg (2006) in their study of the diffusion of innovation, whereas Haak et al. (2005) offer an opposite view regarding the diffusion of farming in Europe.[7]

The use of genetic distances requires our theory to embed those distances in a multi-dimensional space. Since the values of genetic distances are based on information of many different genes, they are represented in a multi-dimensional space. Thus, in contrast to much of the existing theoretical work, population heterogeneity in our model is multi-dimensional. Although there may be certain policy issues for which a one-dimensional space suffices,[8] in general this is too restrictive. For example, if agents who reside in the same county have to decide on the geographic location of a public facility, this problem is, by nature, two-dimensional. Also, agents with the same income may have different views regarding the desired level of redistribution within society. Thus, the search for an optimal public policy is naturally a multi-dimensional problem.

[6] A recent paper by Giuliano et al. (2006) argues that in the case of trade genetic distances cease to be significant once geographical distances are properly measured. In contrast, our focus is on cultural distances, not on trade.

[7] See also Ashraf and Galor (2007) on the effect of cultural diffusion on technological innovation.

[8] See, e.g., Alesina and Spolaore (1997).

Other relevant work, though not in the context of the trade-off between scale economies and cultural heterogeneity, includes the *landscape theory* of Axelrod and Bennett (1993), aimed at predicting the European alignment during the Second World War. They consider a two-bloc setting where each nation is characterized by its propensity to work with other nations. Given the partition of all nations into two blocs, the *frustration* of a nation is determined as the sum of its propensities towards nations outside the bloc it belongs, and *energy* is then the weighted sum of the frustrations of all countries. Using the 1936 data, Axelrod and Bennett show that a local energy minimum over two-bloc structures almost exactly corresponds to the wartime alignment in Europe. Also, Spolaore and Wacziarg (2005) estimate the effect of political borders on economic growth and run a number of counterfactual experiments to examine how the union of different countries would affect growth. However, they do not take into account cultural heterogeneity. A recent paper of Alesina, Easterly and Matuszeski (2006) explores the poor economic performance of "artificial" states, where borders do not match a division of nationalities.

The rest of the paper is organized as follows. Section 2 presents the theoretical framework. Section 3 studies the stability of Europe by exploring the likelihood of secessions and unions between country pairs. Section 4 analyzes the gains of European Union membership. Section 5 contains an exhaustive study of the full stability of Europe. Section 6 provides empirical support for using genetic distances as a proxy of cultural heterogeneity. Section 7 concludes.

2 The Model

2.1 General Framework

The world W is partitioned into countries, each consisting of one or several regions. Each individual in W resides in one of the regions. The set of regions is denoted by N. In the rest of the discussion, the set of regions is taken as given, whereas the partition of the world into countries can change. The population of country C is given by

$$p(C) = \sum_{I \in C} p(I),$$

where $p(I)$ is the population of region I. The summation extends over all regions I belonging to C.

There are two types of heterogeneity in this model, cultural and income. Within regions there is only income heterogeneity. In other words, there is intra-regional income heterogeneity, but no intra-regional cultural heterogeneity, so that individuals in a region may have different incomes but are culturally homogeneous. Within countries that consist of multiple regions both types of heterogeneity, cultural and income, will be present.

For any two regions $I, J \in N$, we call $d(I, J)$ the cultural distance between a resident of I and a resident of J. In the empirical part of our investigation we identify $d(I, J)$ with the genetic distance between region I and region J. Obviously, $d(I, J) = d(J, I)$ for all I and J. Given that cultural heterogeneity is only present across regions, $d(I, I) = 0$ and $d(I, J) > 0$ for all $I \neq J$. We normalize distances such that $0 \leq d(I, J) \leq 1$. We denote by D the matrix $D = (d(I, J))_{I,J \in N}$. The weighted cultural distance between a resident of region I, that belongs to country C, and all other residents of C is

$$H(I, C) = \sum_{J \in C} \frac{p(J)d(I, J)}{p(C)}. \tag{1}$$

The value of $H(I, C)$ represents the degree of *cultural heterogeneity* experienced by a resident of region $I \in C$.

The income distribution in region I is given by the density function $f_I(y)$ with support $[\underline{y}, \bar{y}]$ which is common for all regions. The total income in I is denoted by $Y(I)$:

$$Y(I) = \int_{\underline{y}}^{\bar{y}} y f_I(y) dy. \tag{2}$$

Similarly, $Y(C)$ denotes the total income in country C.

Agents' utility depends on private consumption, c, public consumption, g, and the degree of cultural heterogeneity they face. We adopt the following quasi-linear expression for the utility of an individual in region $I \in C$:

$$u(c, g, I, C) = c + V(g, H(I, C)), \tag{3}$$

where V is twice continuously differentiable, strictly concave and increasing in the amount of public good g. We assume that cultural heterogeneity reduces the utility an agent derives from the consumption of the public good g. Thus, V is decreasing in the second argument, the level of cultural heterogeneity faced by a resident of region I in country C.[9]

Two comments are in place. First, the functional form chosen here is of a more general nature than those in Alesina and Spolaore (1997), Alesina, Baqir and Easterly (1999) and Alesina, Baqir and Hoxby (2004), where the degree of heterogeneity is represented by the type or location of the public good. In fact, the theoretical results we will discuss hold for an even more general class of utility functions given by:

$$u(c, g, I, C) = \Psi(c + V(g, H(I, C))) + \Phi(H(I, C)),$$

where Φ is an arbitrary continuous function, and Ψ is increasing, continuous, and concave, in addition to satisfying some additional technical conditions. It can be shown that this class contains the isoelastic functions $\Psi(u) = u^\delta$, with $0 < \delta \leq 1$, and, in particular, $\Psi(u) = u$, as in our specification in (3).

Second, in our setting individuals prefer more culturally homogenous countries. Alesina, Baqir and Hoxby (2004) offer two reasons for this "homogeneity bias". One is that individuals who share a common background may have similar preferences over public goods. The other is that, even if individuals have similar preferences to those in other groups, they may still prefer to interact with members of their own groups. As Alesina, Baqir and Hoxby (2004) point out, if the function H stands for the disutility of interacting with others, then H may rise in the case of interaction with culturally distinct individuals.

Public goods are financed through a proportional tax rate $\tau, 0 \leq \tau \leq 1$, so that taxation is redistributive. For simplicity, we assume that the price of the public and the private good are both equal to 1. Furthermore, taxation does not involve deadweight losses, so that if country C selects the tax rate τ, the level of the public good will be $\tau Y(C)$. The indirect utility of an individual i with income y_i, residing in region I in country C, that adopts the tax rate τ, can be presented as

$$v(y_i, \tau, I, C) = y_i(1 - \tau) \ + \ V(\tau Y(C), H(I, C)). \tag{4}$$

The tax rate τ in every country C, denoted $\tau(C)$, is chosen by majority voting. Note that for every country C the preferences of every agent $i \in C$ over tax rates are single-peaked. Denote

[9] If agents reside in one-region countries, intra-regional cultural homogeneity implies that for every one-region country I the value of $H(I, I)$ is equal to zero.

by $\tau(y_i, I, C)$ the preferred tax rate for an individual i with income y_i who resides in region $I \in C$:

$$\tau(y_i, I, C) = \arg\max_{\tau \in [0,1]} v(y_i, \tau, I, C) \tag{5}$$

Note that, given single-peaked preferences, majority voting yields the tax rate preferred by the median agent.

It is crucial to point out that the preferences of a region's median income agent over different geographical arrangements "represent" those of the majority of its residents. This is referred to as the *decisiveness* of the median agent (Gans and Smart, 1996). For every region I denote by $y_m(I)$ the median income. And for every region I in C let $v_m(I, C)$ be the value of the indirect utility of the median income agent in I when the tax rate in C is given by $\tau(C)$:

$$v_m(I, C) = v(y_m(I), \tau(C), I, C).$$

We then have the following result, the proof of which is relegated to the Appendix:

Lemma - Median Decisiveness: *For every region I and two different countries C and C' with $I \in C \bigcap C'$ we have*

$$p(\{i \in I | v(y_i, \tau(C), I, C) > v(y_i, \tau(C'), I, C')\}) > \frac{1}{2} p(I)$$

if and only if

$$v_m(I, C) > v_m(I, C').$$

The Lemma states that if the median income agent of a region I prefers to be in country C, rather than in country C', then a majority of agents in that region also prefer C to C'. Note that agents correctly foresee the tax rate that will be chosen by majority vote in the corresponding geographical arrangement.

We would like to remark that the median decisiveness links our model to the theory of *hedonic games*, introduced by Drèze and Greenberg (1980), where the payoff of a player depends exclusively upon the group to which she belongs. In our framework the benefit of a region from being part of a certain country depends solely on regions in that country and is independent of the number and composition of other countries. Thus, our nation formation game is a hedonic game.[10]

Now we turn to the identification of stable partitions of countries. There are several stability concepts that have been applied in the literature (see, e.g., Alesina and Spolaore, 1997, Jéhiel and Scotchmer, 2001, Bogomolnaia et al., 2005). In this paper, we use a stability concept of *Limited Right of Map Redrawing*, called *B-stability* in Alesina and Spolaore (1997) and *equilibrium with admission by majority vote* in Jéhiel and Scotchmer (2001), which requires, subject to majority voting, the unanimous approval of any border redrawing by all affected regions.[11]

For every partition $\pi = \{C_1, \ldots, C_K\}$ and every region $I \in N$ denote by $C^I(\pi)$ the country in π that contains I. We then have the following definition:

Domination relation: Partition π' dominates partition π if for every region affected by the shift from π to π', the majority of its residents prefer the new arrangement π' over the old one π. That is, for every region I with $C^I(\pi) \neq C^I(\pi')$ we have

$$p(\{i \in I | v(y_i, \tau(C^I(\pi')), I, C^I(\pi')) > v(y_i, \tau(C^I(\pi)), I, C^I(\pi))\}) > \frac{1}{2} p(I).$$

This concept of domination allows us to precisely define stability under *Limited Right of Map Redrawing*:

Limited Right of Map Redrawing (LRMR): Let partition π be given. A partition $\pi' \neq \pi$ generates credible *map redrawing* if π' dominates π. A partition π is *LRMR-stable* or simply *stable* if it cannot generate credible map redrawing.

Now we introduce our definition of efficiency. Consistent with the median decisiveness, it is based on the indirect utilities of the median income agents in the regions:

Efficiency: Let Π be a set of world partitions. We call $\pi \in \Pi$ *median-efficient* if it maximizes

$$\sum_{I \in N} v_m(I, C^I(\pi))$$

[10] However, our game is not 'additively separable' which rules out the direct application of the results by Banerjee, Konishi and Sönmez (2001) and Bogomolnaia and Jackson (2002). Also, the contribution by Milchtaich and Winter (2002), where players compare groups on the basis of the distance between their own characteristics and the average characteristics of the group, share some common features with our work.

[11] If the majority requirement is replaced by an unanimous consent, this stability concept is reminiscent of *contractual* and *individual* stability studied by Greenberg (1977), Drèze and Greenberg (1980), Bogomolnaia and Jackson (2002).

over all world partitions $\pi \in \Pi$.

This concept of efficiency is useful when searching for stable partitions. Indeed, the following proposition states that not only is it always possible to find a LRMR-stable partition, but that every median-efficient partition is stable. The opposite is not always true though.

Proposition: *The set of LRMR-stable partitions and the set of median-efficient partitions are both nonempty. Moreover, every median-efficient partition is LRMR-stable.*

The proof of the proposition is presented in the Appendix.

2.2 Our Specification

Before bringing our theoretical model to the data, we make some additional assumptions on agents' utilities. We adopt the following quasi-linear functional form for the utility of an individual in region $I \in C$:

$$u(c, g, I, C) = c + \alpha(Z(I, C)\ g)^{\beta}, \tag{6}$$

where $\alpha > 0$ and $\beta > 0$ are exogenously given parameters, and $Z(I, C)$ is a 'discount factor', whose range is between 0 and 1.

Since cultural heterogeneity reduces the utility an agent derives from the consumption of the public good g, the value of $Z(I, C)$ is negatively correlated with the cultural heterogeneity faced by a resident of region I in country C. More specifically, we assume that for such an agent the discount factor is given by

$$Z(I, C) = 1 - H(I, C)^{\delta}, \tag{7}$$

where $\delta \in [0, 1]$.

The parameter δ is important in two respects. First, since $H(I, C)$ is between 0 and 1, the smaller is δ, the greater is the cost of heterogeneity. If δ is very small, the value of $Z(I, C)$ in a multi-regional country is close to zero. In other words, a small δ implies that in such a country any amount of public consumption becomes almost useless. Second, the smaller is δ, the more convex is the discount factor Z. For small values of δ, the discount factor exhibits a high degree of convexity, so that the relative effect of increasing heterogeneity on Z is larger at lower levels of heterogeneity.

If agents reside in one-region countries, the discount factor $Z(I, I)$ is equal to one, regardless of the value of δ. Thus, the utility of agents in one-region countries becomes:

$$u(c, g) = c + \alpha g^{\beta}.$$

The indirect utility of an individual i with income y_i, residing in region $I \in C$, where the tax rate is τ, is

$$v(y_i, \tau, I, C) = y_i(1 - \tau) \; + \; \alpha \; (Z(I, C) \; \tau \; Y(C))^{\beta}. \tag{8}$$

We can now explicitly derive $\tau(y_i, I, C)$, the preferred tax rate for an individual i with income y_i who resides in region $I \in C$. It is easy to see that the (interior) solution to (5) is

$$\tau(y_i, I, C) = \left(\frac{y_i}{\alpha \; \beta \; (Z(I, C) Y(C))^{\beta}} \right)^{\frac{1}{\beta - 1}} \tag{9}$$

Notice that, in general, for $I, J \in C$ we have $Z(I, C) \neq Z(J, C)$. In other words, the cost of cultural heterogeneity tends to be different for agents living in different regions of the same country. As a result, two individuals with the same income level, but residing in different regions of country C, typically have different preferred tax rates. This implies that the median agent in country C does not necessarily coincide with the agent with the median income in C. This feature has important consequences for the empirical part of the paper. Finding the preferred tax rate of a coalition of regions forming a country becomes more laborious than just finding the preferred tax rate of the median income agent. Of course, when a country is formed by only one region, this problem disappears, and the agent with the median income becomes the decisive one in determining the tax rate.

3 Stability of Europe

In this section we investigate whether we can find values of parameters that render the current map of Europe stable according to our *Limited Right of Map Redrawing* stability concept. Using information on cultural distances between European regions and countries, our goal is to find values of α, β and δ that yield a LRMR-stable partition of Europe.

This exercise is of interest for a number of reasons. First, as a way of validating our theoretical framework, it seems important that the set of parameter values consistent with stability is not empty. Second, our analysis allows us to determine which regions are more likely to separate, and which countries are more likely to form a union. For instance, by increasing the cost associated with cultural heterogeneity, we can check which region would be the first one to secede. We can thus pinpoint the 'weak' links in the current map of Europe.

3.1 Data

The most important data issue is to specify the matrix of cultural distances D. As already mentioned, we use genetic distances between populations.[12] The best-known reference is Cavalli-Sforza et al. (1994), who collected data from different sources to construct a matrix of genetic distances for a large number of populations across the world.[13] To carry out our exercise, it is important to have information, not just on countries, but also on regions. Indeed, to limit the range of δ from above and from below, it is not enough to make sure that no existing countries want to unite, we also must guarantee that no existing regions want to separate. The matrix of Cavalli-Sforza et al. (1994) is therefore appropriate, as it contains information on 22 European countries and 4 European regions (Basque Country, Sardinia, Scotland and Lapland).[14] Table A.1 in the Appendix reproduces the matrix. Although it leaves out a number of relevant regions (Flanders, Catalonia, Brittany, Northern Italy, Corsica, etc.), the fact of having at least some regions is conceptually enough to allow us to estimate δ.[15]

The other data we need are standard. Data on population and GDP per capita (measured in PPP) are for the year 2000, and come from Eurostat, the Penn World Tables and the International Monetary Fund. Data on income distribution come from the World Income Inequality Database v.2.0a, collected by the United Nations University. Since those data are not available for all years,

[12]See Hartl and Clark (1997) for an introduction to population genetics, and Jorde (1985) for a discussion on the use of the different types of genetic distances to measure human population distances.

[13]The distances in Cavalli-Sforza et al. (1994) are based on large sample sizes and use information about many different genes. Most of the frequencies used to obtain those distances come from allozymes, instead of from direct 'observation' of the DNA sequence, a technique which is now available. However, Cavalli-Sforza et al. (2003) argue that these new techniques and data do not change the basic results.

[14]Given the small population of Lapland, less than 100,000 and spread over three countries, we do not use this region in our subsequent analysis. We also drop Yugoslavia, as that country disintegrated in the 1990s.

[15]One possibility would be to incorporate more recent data from other sources, such as the ALFRED database, available at http://alfred.med.yale.edu/alfred/index.asp. However, merging the data would require a laborious and complex effort. Since our goal is to illustrate how data on genetic distances can be used to study issues of stability, we prefer to stick to the high quality data provided in Cavalli-Sforza et al. (1994).

we take the year which is closest to 2000. The income distributions of regions are taken to be the same as those of the countries they belong to.

For those countries for which we have information on regions, we need to distinguish in the data between the country, the region, and the country net of the region. Take the case of Spain. If the question is whether the Basque Country wants to separate, the two relevant decision makers are the Basque Country and the rest of the Spain. However, if the question is whether Spain wants to unite with Portugal, the two relevant decision makers are Spain (including the Basque Country) and Portugal.

3.2 Estimation Strategy

Our strategy is to first calibrate α and β using data on a set of European and OECD countries, so that we are left with only one degree of freedom, the parameter δ. To calibrate α and β, we assume away cultural heterogeneity within countries.[16] This amounts to assuming that each country is made up by one region. In that case, the tax rate adopted by country C is

$$\tau(C) = \left(\frac{y_{m(C)}}{\alpha \, \beta \, (Y(C))^{\beta}} \right)^{\frac{1}{\beta - 1}} \tag{10}$$

where $y_{m(C)}$ is the median income in C. As can be seen from (10), we need data on the tax rate, $\tau(C)$, median income, $y_{m(C)}$, and total income, $Y(C)$. For the tax rate, we take the ratio of government spending on public goods to total GDP. It is not entirely obvious how to measure spending on public goods. To get as close as possible to what is a public good, we want to focus on activities where congestion is limited. We use a number of alternative measures. All data come from the Government Finance Statistics (GFS) database, collected by the IMF. A first measure takes the sum of general public services, defense, public order and safety, environmental protection, and economic affairs. A second measure takes only general public services. And a third measure focuses exclusively on defense. As will be shown later, these alternative measures do not lead to qualitatively different results. We use data for all European and OECD countries in the GFS database. Depending on the measure used, we have information on 27 to 30 countries.

To calibrate α and β, we estimate (10) by applying nonlinear least squares. The results for

[16] As a robustness check, we re-estimate α and β for a subset of countries which are relatively homogeneous, and use those alternative estimates for our subsequent analysis. This does not change our results qualitatively.

each of the three measures of government spending are reported in Table 1. Standard errors are given in brackets.

	α	β
General public services, defense public order, environment economic affairs	-287 (529)	-0.0322 (0.0709)
General public services	25.80 (26.79)	0.0833 (0.0627)
Defense	-6.42 (3.27)	-0.1917 (0.1625)

Table 1: Estimation of α and β

Using these measures of α and β, we now compute the range of δ for which the current map of Europe is LRMR stable. In principle, checking for stability would require us to analyze all possible partitions of the 21 countries and 3 regions we focus on. However, the number of such partitions is too large (445,958,869,294,805,289). We therefore limit our analysis to all possible separations (Basque Country-Spain, Scotland-Britain, Sardinia-Italy) and all possible mergers between country pairs. In as far as large unions start off small, focusing on unions between country pairs is not unrealistic.[17]

Using this setup, for Europe to be LRMR stable, two conditions need to be satisfied:

1. There is no unanimity between a region and the country it is part of to separate, i.e., there is no majority in both the region and the country it belongs to in favor of secession.

2. There is no unanimity between any pair of countries to unite, i.e., there is no majority in each of the two countries to unite.

We start by analyzing the condition for no region to secede. Consider the three regions in our database (Basque Country, Sardinia, and Scotland) and the three countries they belong to (Spain, Italy, and Britain). For secession to occur, there needs to be a majority in both affected parts. For instance, if the Basque Country is to separate, a majority of Basques and a majority of

[17]We return to the issue of unions between more than two countries in Section 5.

the population in the rest of Spain should approve. Therefore, in this context 'Spain' is defined as 'Spain without the Basque Country', and likewise for Italy and Britain. In the case secession does not occur, the agent with the median income in region $I \in C$ enjoys utility level

$$v_m(I,C) = v(y_m(I), \tau(C), I, C),$$

If, instead, region I secedes, the utility of the agent with the median income becomes

$$v_m(I,I) = v(y_m(I), \tau(I), I, I),$$

Under median representation region I prefers to remain part of country C if $v_m(I,C) \geq v_m(I,I)$. Since the utility of forming part of country C depends on the parameter δ, we write the net gain of the union for the median income agent of region I as

$$g_{I,C}(\delta) \equiv v_m(I,C) - v_m(I,I) \tag{11}$$

We now need to consider the same condition for the other affected part, i.e., the median income agent of 'Spain without the Basque Country' or of 'Britain without Scotland'. The net gain of the union for the median income agent of the rest of the country C/I can be written as

$$g_{C/I,C}(\delta) \equiv v_m(C/I,C) - v_m(C/I,C/I) \tag{12}$$

According to our definition, to prevent a secession, it suffices that one of the remaining parts prefers to remain united. Thus, a first *necessary* condition for the current European partition to be stable is the existence of a nonempty set of the parameter δ for which *at least* one of the functions (11) and (12) is positive for each of the pairs Basque Country-Spain, Sardinia-Italy, and Scotland-Britain. The set of δ for which secession does not occur can be defined as

$$S^R \equiv \{\delta | \max\{g_{I,C}(\delta), g_{C/I,C}(\delta)\} \geq 0, \text{for all } I \in \ \{\text{Sardinia, Basque Country, Scotland}\}\}$$

The range of δ for which this condition holds for the relevant secessions in our data set is obtained numerically.

We now analyze the condition that ensures no country pairs unite. To determine the preferred tax rate in a possible union between, say, C and C', we need to identify the median voter. Because the 'discount factor' Z is not the same for all agents, this implies that the median voter need not coincide with the median income agent of the union. To solve this problem, we proceed in the following way. We compute the average income of an agent in each decile of the income distribution for both countries C and C'. This, together with data on population and income, allows us to determine for the union of C and C' the preferred tax rate of each one of these agents. In the case of the union between two countries, this gives us 20 tax rates. Given that preferences over tax rates are single peaked, we can find the optimal tax rate for the decisive agent. This is done by ordering the 20 tax rates mentioned above, and taking the one which corresponds to half of the population of the union.

The net gain obtained by the median income agent in country C from joining country C' can be written as

$$g_{C,C'}(\delta) \equiv v_m(C, C \cup C') - v_m(C, C)$$

A second *necessary* condition for LRMR stability is that there is no pair of countries C, C' such that it is in the interest of both to join. In other words, there is no pair C, C' such that $g_{C,C'}(\delta) > 0$ and $g_{C',C}(\delta) > 0$. The set of δ for which no two nations want to unite can be defined as

$$S^N \equiv \{\delta | \min\{g_{C,C'}(\delta), g_{C',C}(\delta)\} \leq 0, \text{for all } C, C']$$

Combining the necessary conditions for 'no secession' and 'no union', a *necessary* condition for LRMR stability is that the set

$$S \equiv S^N \cap S^R$$

is non empty. It is clear that S is an interval on the real line, and we write $S \equiv [\underline{\delta}, \overline{\delta}]$.

3.3 Secessions and Unions between Country Pairs

To numerically compute whether there exists a range of δ that renders the current map of Europe stable, we take the values of α and β estimated before. Taking government spending to be the sum of general public services, defense, public order and safety, environmental protection, and economic affairs, Table 1 gives us $\alpha = -282$ and $\beta = -0.0322$. Numerical computation then shows that $S = [0.0285, 0.1575]$. A first conclusion is therefore that the set S is nonempty. This result is robust to the alternative definitions of government spending in Table 1.

We can now look at which regions are more likely to secede, and which country pairs are more likely to unite. To understand how this can be done, note that if $\delta < 0.0285$, cultural distances are given so much weight, that we cannot prevent certain regions to break away. By progressively lowering δ, we can then rank regions, depending on the risk they pose to the union. Likewise, if $\delta > 0.1575$, the weight put on cultural distances is not enough to prevent some currently independent nations from uniting. By progressively increasing δ, we can rank country pairs, depending on how likely they are to unite.

Table 2 focuses on the likelihood of secessions. As can be seen, the Basque Country is the more likely one to break away, followed by Scotland and Sardinia. This ranking is unchanged under a number of robustness checks.[18]

1	Basque Country
2	Scotland
3	Sardinia

Table 2: Likelihood of secession

Table 3 focuses on the likelihood of unions between country pairs. The first column consists of the benchmark case. Austria and Belgium are the two countries most likely to unite: both are small, have similar populations, and similar levels of GDP per capita. According to the Cavalli-Sforza matrix, they are also genetically close. Remember that present-day Belgium became part of Austria with the Treaty of Utrecht (1713), following the Spanish War of Succession, and remained under Habsburg rule until the French invasion of 1794. The next pairs which stand to gain most from unification — Switzerland-Belgium, Denmark-Norway, Austria-Switzerland, and Belgium-Netherlands — fit the same pattern: small countries, comparable levels of GDP per capita, and genetically similar. Again, the presence of Belgium in many of these pairs is not surprising: this 'genetic capital' of Europe, located on the border between Latin and Germanic Europe since Roman times, often served as Europe's 'battlefield'. The ranking shows that unions between large and small countries are unlikely as the larger country would not derive much fiscal benefit from such unions. There is one exception though: Poland-Belgium. Since the larger country of the two, Poland, is also the poorer one, this union still has the potential of being mutually beneficial. Likewise, unions between two large nations are not common, as on their own they already benefit from substantial

[18] In particular, we used the two alternative definitions of government spending of Table 1. In addition, we also checked for α plus or minus its standard error, and β plus or minus its standard error.

increasing returns in the provision of public goods. The only two such unions in the top-10 occupy the last two positions: Germany-Britain and France-Germany.

	Benchmark	**Geographically contiguous**	**Same population**
1	Austria-Belgium	Denmark-Norway	Denmark-Netherlands
2	Switzerland-Belgium	Austria-Switzerland	Austria-Switzerland
3	Denmark-Norway	Belgium-Netherlands	Belgium-Netherlands
4	Austria-Switzerland	Norway-Sweden	Germany-Switzerland
5	Belgium-Netherlands	Germany-France	Germany-Belgium
6	Belgium-Poland	France-Britain	Belgium-Britain
7	Switzerland-Denmark	Czech Republic - Hungary	Switzerland-Belgium
8	Norway-Sweden	France-Italy	Switzerland-Netherlands
9	Germany-Britain	Denmark-Sweden	Germany-Netherlands
10	France-Germany	Netherlands-Britain	Austria-Belgium

Table 3: Likelihood of unions

The second column in Table 3 restricts possible unions to country pairs that are geographical neighbors.[19] In that case, Denmark and Norway are the two countries most likely to unite. They are followed by Austria-Switzerland, Belgium-Netherlands, and Norway-Sweden. Not surprisingly, those pairs tend to have a common historical path. Norway was a part of the Danish crown from the Middle Ages until 1814. Belgium and the Netherlands were united under Burgundy, Habsburg and Spain from 1384 to 1581, and again after the Treaty of Waterloo, from 1815 to 1830. Sweden and Norway were under the same crown from 1814 to 1905, not counting a brief common spell in the 14th century.

The third column in Table 3 runs a counterfactual by assuming that all countries have the same population of 26 million, corresponding to the average of the countries in our data set. When abstracting from different population sizes, the most likely union is between Denmark and the Netherlands. In fact, the genetic distance between the two is the smallest one in the Cavalli-Sforza matrix. Relations between both countries became strong during the Eighty Years War between the Netherlands and Spain in the 16th and 17th centuries, when a large number of Dutch migrated to Denmark, turning the Netherlands into one of the most important export markets for

[19] In the case of islands, such as Britain, or peninsulas, such as Denmark, we interpret this as countries which are geographically 'close'.

Denmark. When ignoring differences in population sizes, unions between, for instance, Germany and Switzerland, or Germany and Belgium, become increasingly likely. This suggests that the most important obstacle to a union between, say, Germany and Switzerland, is their different sizes. Other unions, such as between Belgium and Poland, now become less likely. Indeed, what made Poland attractive to Belgium in the benchmark case was its large size.

As a robustness check, Table 4 uses alternative definitions of government spending. The first column takes general public services to be the measure of public spending. Using the corresponding α and β from Table 1, we re-estimate which countries are most likely to unite. The second column follows the same procedure, using defense as the measure of public spending. As can be seen, in both cases, the results are similar to the benchmark case. In fact, the five most likely unions remain the same. Further robustness checks on α and β do not change the results. In particular, when we take β plus or minus its standard error, and re-optimize the value of α, the five most likely unions do not vary. The same result obtains when taking α plus or minus its standard error.

	General public services	**Defense**
1	Austria-Belgium	Austria-Belgium
2	Switzerland-Belgium	Switzerland-Belgium
3	Denmark-Norway	Denmark-Norway
4	Austria-Switzerland	Austria-Switzerland
5	Belgium-Netherlands	Belgium-Netherlands
6	Switzerland-Denmark	Switzerland-Denmark
7	Norway-Sweden	Poland-Belgium
8	Germany-Britain	Norway-Sweden
9	Poland-Belgium	Germany-Britain
10	France-Germany	France-Germany

Table 4: Likelihood of unions, using alternative definitions of public spending

4 The Gains of European Union Membership

In this section we use our model to estimate the gains of being a member of the EU-15. Our goal is two-fold. First, we want to see which countries gain most and which lose most from being part of the European Union. Second, we would like to understand how taking into account cultural distances affects the ranking of those gains.

The idea is to view the European Union as a new country formed by the merger of previously independent nations. We can then compare the utility of being inside or outside the EU. In terms of data, we focus on the 14 member states of the EU-15 for which we have information.[20] If country C is part of the European Union, the utility of its median income agent is $v_m(C, EU)$, where EU is the set of members of the European Union. Country C's relative gain from becoming part of the EU is:

$$g_{C,EU}(\delta) \equiv \frac{v_m(C, EU) - v_m(C, C)}{v_m(C, C)}$$

The relative gains of being part of the European Union depends on the value of δ. Assuming the current map of Europe is stable, our previous estimations indicate that δ belongs to the set $S = [0.0285, 0.1575]$. Since it is not obvious *which* value of δ to choose within that range, we assume that all the elements of S are equally likely. To compute the relative welfare gain of being a member of the EU, we therefore take the average of $g_{C,EU}(\delta)$ over all the parameters in S, namely

$$g_{C,EU} \equiv \int_S g_{C,EU}(\delta) dF \tag{13}$$

where F is the uniform distribution over the interval S. We take an approximation $\widehat{g}_{C,EU}$ by computing

$$\widehat{g}_{C,EU} \equiv \sum_{i=0}^{1000} g_{C,EU}(\underline{\delta} + \frac{\overline{\delta} - \underline{\delta}}{1000} i) \tag{14}$$

Table 5 reports the ranking of relative utility gains of the different member states of the EU-15.[21] According to our computations, Ireland is the country that gains most, followed by Denmark. Germany is the only country that loses from EU membership, although the gains in the larger countries — Italy, Britain, France and Spain — are relatively small.

Different variables — population size, GDP per capita, income distribution, and cultural heterogeneity — affect this ranking. Table 5 seems to suggest a strong correlation between population size and relative gains. However, population cannot be the entire explanation. Greece, Belgium and Portugal, for instance, all have a population size of around 10 million, but Greece gains less than Belgium, and Belgium gains less than Portugal. The difference between Belgium and Portugal can be attributed to GDP per capita. Richer countries are forced to redistribute

[20] Data on cultural distances are missing for Luxembourg.

[21] We limit ourselves to reporting the ranking, and not the relative utility gain for each country, as this measure is not meaningful.

more, and may therefore be less interested in uniting. However, this does not explain the difference between Belgium and Greece. This is where cultural distances come in: Belgium is the *least* distant from the average European country, whereas Greece is the *most* distant. This explains why Greece, in spite of being nearly 40% poorer than Belgium, gains less from membership in the EU.

	Country	Population	Cultural distance	GDP per capita
1	Ireland	3.8	0.095	126
2	Denmark	5.3	0.045	126
3	Finland	5.1	0.105	113
4	Portugal	10	0.051	80
5	Austria	8.1	0.043	126
6	Belgium	10.2	0.027	117
7	Sweden	8.87	0.067	119
8	Greece	10.6	0.142	73
9	Netherlands	15.9	0.041	120
10	Spain	40.3	0.056	92
11	France	59	0.032	114
12	Britain	58.6	0.034	112
13	Italy	56.9	0.042	113
14	Germany (-)	82	0.031	112

Table 5: Ranking of relative utility gains from being member of EU

To understand the role of cultural distances, we recompute the gains from being part of the EU, setting all distances between all countries to zero. The results are reported in Table 6. As expected, Greece now gains more than Belgium. When abstracting from cultural distances, Greece goes up 4 ranks and Belgium goes down 2 ranks. France also swaps places with Italy. Given that France is culturally closer to the European average, it gains less than Italy if we do not take into account culture.

Rather than focusing on relative utility gains, one can compute a ranking based on monetary gains. We do so by calculating the relative increase in per capita income, r, all agents in country C should receive to render its median agent indifferent between joining the EU (and not receiving the additional income $ry_m(C)$) and remaining outside the EU (and receiving $ry_m(C)$). The relative increase (decrease) in income is a measure of the relative monetary gains (losses) from being part of the EU. To determine r for each nation C we solve the following equation:

$$y_m(C)(1+r)(1-\tau'(C)) + \alpha(\tau'(C)Y(C)(1+r))^\beta \\ = y_m(C)(1-\tau(EU)) + \alpha(Z_{(C,EU)}\tau(EU)Y(EU))^\beta \tag{15}$$

where $\tau'(C)$ is the optimal tax rate for the median income agent of country C, given that everyone's income in C is multiplied by $(1+r)$.

Country	Change ranking
Ireland	0
Finland	1
Denmark	-1
Greece	4
Portugal	-1
Austria	-1
Sweden	0
Belgium	-2
Netherlands	0
Spain	0
Italy	2
Britain	0
France	-2
Germany (-)	0

Table 6: Relative utility gains of being member of EU (no cultural distances)

Table 7 reports the relative increase in income that leaves the decisive agent in each country indifferent between joining or not joining the EU-15. The ranking we obtain is similar to the one based on utility. Germany is the only country that loses (nearly 2% of income), whereas Portugal is the one that gains most (about 26% of income). There are some differences though. Greece now gains more than Belgium, albeit, in spite of being poorer, still less than Portugal. This is no longer the case when ignoring cultural differences. As can be seen in the last column of Table 7, Greece moves up one position, whereas Portugal moves down one. Thus, Greece becomes the country that gains most from EU membership, ahead of Portugal. When not abstracting from cultural differences, Spain, in spite of its larger size, benefits more than the Netherlands.

Country	Monetary gain (%)	Population	Cultural distance	GDP per capita	Ranking (no distance)
Portugal	25.93	10	0.051	80	-1
Greece	22.39	10.6	0.142	73	1
Ireland	18.96	3.8	0.095	126	0
Finland	16.79	5.1	0.105	113	0
Denmark	16.12	5.3	0.045	126	0
Belgium	14.43	10.2	0.027	117	-2
Austria	14.21	8.1	0.043	126	0
Sweden	12.89	8.87	0.067	119	2
Netherlands	10.82	15.9	0.041	120	-1
Spain	4.84	40.3	0.056	92	1
France	0.79	59	0.032	114	-2
Britain	0.61	58.6	0.034	112	0
Italy	0.52	56.9	0.042	113	2
Germany	-1.76	82	0.031	112	0

Table 7: Relative monetary gain from being member of EU

5 Full Stability

In Section 3 we argued that a complete study of LRMR stability for Europe would exceed our computing capacity. Indeed, for the 21 countries and 3 regions in our data set, this would amount to checking 445,958,869,294,805,289 possible partitions. Moreover, determining who is the agent with the median optimal tax rate in each partition is extremely laborious, because, due to cultural heterogeneity, the decisive agent need not coincide with the median income agent. This is one reason for why in Section 3 we limited our analysis to unions of two countries. The other reason is that in a dynamic framework, where larger unions between multiple countries start off as smaller unions between a few, focusing on country pairs is of interest *per se*.

In this section we revisit the problem of full stability. By introducing two restrictions, we are now able to check for all possible partitions. First, instead of looking at the entire Europe, we focus on the EU-15, and leave out the peripheral countries, such as Ireland, Finland and Sweden. In absence of data for Luxembourg, this leaves us with 11 countries, and ‘only’ 678,570 possible partitions. Second, we assume that in each country the level of the public good is chosen to maximize the total utility of its residents. It is easy to see that maximizing total utility in a nation is equivalent to maximizing the population-weighted average of the utility of the mean income residents of the different regions. In that case, the tax rate adopted in country C is the solution to

$$\overline{\tau}(C) = \arg \max_{\tau \in [0,1]} \sum_{I \in C} p(I) v(\overline{y}(I), \tau, I, C) \tag{16}$$

where $\overline{y}(I)$ is the mean income in region I. One can easily show that the solution to (16) is given by

$$\overline{\tau}(C) = \left(\frac{1}{\alpha \ \beta \ \sum_{I \in C} p(I) \ (Z_{(I,C)})^{\beta}} \right)^{\frac{1}{\beta - 1}} \frac{1}{Y(C)} \tag{17}$$

To compute the tax rate (17), we only need information on population, total GDP and cultural distances without identifying the median agent for each partition. As a result, calculating welfare for each of the 678,570 partitions becomes a computationally feasible task.

One can easily define LRMR stability when the decisive agent is the mean income agent, by changing 'the majority of its residents' in the definition of Domination Relation by the 'mean income agent'. In this case, the corresponding definition of efficiency should be mean-efficiency, rather than median-efficiency, and a similar result as the one in Proposition 1 holds.

We want to emphasize that we adopted this approach with the sole goal of simplifying the problem computationally. From a theoretical point of view, this simplification may come at a cost. However, from an empirical point of view it turns out that this 'mean agent' framework is a good 'proxy' of the previous approach. To reach this conclusion, we recalculated our derivations in Sections 3 and 4, using a 'mean agent' rather than a 'median agent' framework. Since none of the previous results changed, we believe that adopting this simplification does not come at the cost of losing realism. Our empirical results would likely be very similar if we were able to do the exercise using a median voter framework.

We compute total welfare for each of the $678,570$ partitions and select the partition that yields the maximum. The result depends, obviously, on the chosen value of the parameter δ. We find that, at an accuracy level of 0.00001, there exists a 'critical' value of $\delta^* = 0.04066$, such that for $\delta < \delta^*$ the current partition of Europe maximizes total welfare, and therefore is efficient and LRMR-stable, whereas for $\delta > \delta^*$ the union of all countries maximizes total welfare, so that the EU would be efficient and LRMR-stable. In other words, the only two efficient partitions of Europe is either full integration or full independence.

This result is subject to two caveats. First, the absence of intermediate configurations is not a general feature of the model. One can easily generate examples for subsets of the countries analyzed in this paper for which the efficient partition implies the union of some, but not all, countries. For example, in the case of Sweden, Denmark and Greece for values of $\delta \in [0.18, 0.21]$ the efficient partition consists of the union of only Denmark and Sweden. Second, in our model we do not impose any restrictions on how unions are formed. Even if a union between all countries is the efficient outcome, whether a full union is reached or not would depend on the dynamics of how unions are formed. The literature on whether preferential trade agreements are building blocks or stumbling blocks to global free trade may be of interest here.

6 Genetic and Cultural Distances

In this section we offer the arguments to support our choice of using genetic distances as a proxy for cultural distances among populations. The question we ask is *Are genetic distances correlated to cultural distances?* We propose the following strategy to answer this question: compare the matrix of genetic distances from Cavalli-Sforza et al. (1994) with the answers given in the World Values Survey (WVS) to questions on "cultural values". In particular, we take the 430 questions included in the sections on Perceptions of Life, Family and Religion and Moral from the four waves currently available online at http://www.worldvaluessurvey.org/.

We use these questions from the WVS to calculate cultural distances between 14 European nations. Each question has q different possible answers and we denote by $x_{i,j} = (\, x^1_{i,j}, x^2_{i,j}, \ldots, x^q_{i,j})$ the vector of relative answers to question i in nation j. For example, suppose that question i has three possible answers, a, b and c. The vector $x_{i,j} = (1/2, 0, 1/2)$ indicates that in nation j, half of the people answer a, and the other half c. We construct a *matrix of opinion poll distances* between the nations such that the (j, k) element of the matrix represents the average Manhattan distance between nation j and nation k and is given by

$$w_{jk} = \sum_{i=1}^{430} \sum_{s=1}^{q} \left| x^s_{i,j} - x^s_{i,k} \right| \tag{18}$$

We denote the resulting matrix by W, which is reported in Table A.2 in the Appendix. All our results are robust to the usage of the Euclidean distance instead of the Manhattan distance in (18).

6.1 Descriptive Statistics

We wish to verify whether matrix W is correlated with the matrix of genetic distances D, i.e, whether genetically close countries provide similar answers to the questions in the World Values Survey. Figure 1 exhibits a scatter plot to get a better visualization of a possible correlation. The y-axis represents WVS distances and the x-axis genetic distances. If genetic distance from nation i to nation j is "x" and the WVS distance is "y ", then in the plot there is a corresponding point with coordinates (x, y). Thus, the x-coordinate of a point in the plot comes from the coefficient in matrix D and the y-coordinate comes from the corresponding coefficient in matrix W. We can see that Figure 1 suggests a strong correlation between WVS and genetic distances.

6.2 A More Formal Test

Due to the triangle inequality property, the elements of a distance matrix are not independent, so that we cannot use standard methods of least square estimation to test for (linear) correlation between the matrices D and W. A method often used in Population Genetics is the *Mantel Test* which is a nonparametric randomization procedure.[22]

Mantel's test statistic is the correlation coefficient, r, of the distance matrices D and W. The significance of the correlation is evaluated via random permutation of the rows and corresponding columns of D and W. For each random permutation, the correlation r is computed. After a sufficient number of iterations, the distribution of values of r is generated and the critical value of the test at the chosen level of significance is found from this distribution. In our case, the correlation coefficient between matrices D and W is 0.64 and the hypothesis of non-positive correlation is strongly rejected based on a Mantel test with $100,000$ replications (p-value of 0.00014). This highly significant correlation provides a foundation for the use of the matrix of genetic distances as a proxy for the cultural heterogeneity among European countries.

If the defense for using matrix D is based on its correlation with the matrix of distances W, one might claim that it would be better to directly use W for our analysis. However, the matrix W is based on opinion polls, and although we focus on questions related to people's long term preferences, their answers may still be distorted by short term events. In that sense, we are

[22] See Mantel (1967), Sokal and Rohlf (1995), and Legendre and Legendre (1998). For the use of the Mantel test in economics, see Collado et al. (2005).

interested in analyzing the correlation between W and D, *not* because W is an unbiased measure of the true cultural distances, but because a lack of positive correlation would raise doubts about using D as a proxy for those unknown cultural distances.

An additional criticism might be that there are better proxies for the cultural distances than the genetic distances among populations. A natural alternative to our matrix D could be the matrix of geographical distances between countries. Thus, we compute a matrix G of geographical distances among our European countries.[23] Since we do not observe the true matrix of cultural distances there is no fully satisfactory way to assess which matrix, either D or G, is a better proxy. However, it is possible to test whether genetic distances are more than just a proxy for geographic distances. In other words, it might be the case that once we control for geography, the matrix of genetic distances G is no longer correlated with the matrix W.

In order to investigate this possibility we perform a multiple variable Mantel test to determine the significance of the correlation coefficient of the D and W matrices, controlling for G.[24] The correlation is now 0.32, significantly greater than zero (p-value of 0.02). Thus, after controlling for how geographically close populations are, we still find that populations that are similar in genes tend to be similar in their answers to the opinion polls.

The amount of mixing between populations might also be influenced by the languages spoken by them. One would expect that two populations with the same language have experienced more mixing than populations speaking quite unrelated languages. We therefore study whether the correlation between genetic distances and cultural distances still holds after controlling for both linguistic and geographic distances. To do so, we construct a matrix L of linguistic distances between all our populations.[25] We then perform a second multiple variable Mantel test to determine

---

[23]Geographic distances were calculated "as the crow flies", and the coordinates of each region were obtained from Simoni et al. (2000). This matrix and all the other matrices calculated in the paper, as well as the software programs used for the computation of correlation tests, are available from the authors upon request.

[24]To do so, we follow Smouse et al. (1986) who extend the Mantel bivariate test to the context of multiple control variables.

[25]Our measure of distance between languages is based on the proportion of cognates between Indo-European languages elaborated by Dyen, Kruskal and Black (1992). See Ginsburgh et al (2005) and Desmet et. al (2005) for an application of these distances to economics. The linguistic distances between populations are calculated using the information from the Ethnologue Project on the number of people speaking each language in each country. We set the distance between Finland and any other country to 1, the maximum possible distance. The matrix L of linguistic distances, and the details of its construction, are available form the authors upon request.

the significance of the correlation coefficient of the D and W matrices, controlling for G and L. The correlation is now 0.28, still significantly greater than zero (p-value of 0.04). To understand what this means, consider the following example. Say country i is geographically equidistant from j and k, and the same language is spoken in j and k. In that case country i will be closer to country j than to country k in the answers given to the WVS if the genetic distance between i and j is smaller than between i and k.[26]

The significant positive correlation between genetic distances and World Values Survey distances therefore holds up, even when controlling for geographic and linguistic distances. To the best of our knowledge, this is the first time a clear correlation between genetic distances and modern cultural distances has been reported in the literature. This result provides an argument in favor of using genetic distances as a proxy for cultural distances between populations.

7 Further Research

By using data on cultural distances between regions and nations, this paper has empirically explored the stability of Europe. There are at least three main areas for future research. First, integration and cooperation between regions and countries may take many different forms. Regions may have high degrees of autonomy, without fully seceding. Countries may closely cooperate, without fully uniting. By incorporating those possibilities into the theoretical framework, one could empirically study the degree of decentralization and cooperation. Second, certain recent events, such as the breakup of the Soviet Union or the enlargement of the EU, can be analyzed within the framework we propose. Third, the dynamics of nation formation warrants further attention. Large coalitions, such as the present day EU, started off being much smaller. Since there is likely to be path-dependence in coalition formation, understanding these dynamics is important.

8 Appendix

Proof of the Lemma: Consider a region I and two different countries C and C' such that $I \in C \bigcap C'$. First, suppose that the inequality $v(y_i, \tau(C), I, C) > v(y_i, \tau(C'), I, C')$ holds for more

[26]Performing an alternative multiple variable Mantel test to determine the significance of the correlation between W and G, controlling for D and L, gives a positive but less significant correlation, p-value=0.10.

than half of region I's population. By (4), this inequality can be rewritten as

$$y_i(\tau(C') - \tau(C)) > V(\tau Y(C'), H(I, C')) - V(\tau Y(C), H(I, C)). \tag{19}$$

The range of y_i that satisfies (19) is an interval, and since it contains more than half of region I's population, the interval must include the median agent $y_m(I)$, for whom (19) should hold as well.

Assume now that $v_m(I, C) > v_m(I, C')$. By (19), we have

$$y_m(I)(\tau(C') - \tau(C)) > V(\tau Y(C'), H(I, C')) - V(\tau Y(C), H(I, C)). \tag{20}$$

If $\tau(C') - \tau(C) \geq 0$ then (20) holds for all $y_i > y_m(I)$ and some $y_i < y_m(I)$. If $\tau(C') - \tau(C) < 0$ then (20) holds for all $y_i < y_m(I)$ and some $y_i > y_m(I)$. In both cases, more than half of I's residents have the same preferences over C and C' as the median agent $y_m(I)$. **Q.E.D.**

Proof of the Proposition: For every $\pi \in \Pi$ denote

$$R(\pi) = \sum_{I \in N} v_m(I, C^I(\pi)).$$

Then π is a median-efficient partition if and only if

$$R(\pi) = \max_{\pi' \in \Pi} R(\pi').$$

Since Π is a finite set, there exists a median-efficient partition π. Let us show that it is LRMR-stable. Indeed, if not, then there is a partition π' that dominates π. Then a median agent in every region affected by a shift from π to π', would be better off at π'. Since in regions that are not affected by a shift, there is no change in utility, we have $R(\pi') > R(\pi)$, a contradiction to the median-efficiency of π. **Q.E.D.**

Table A.1: Matrix of Genetic Distances (from Cavalli-Sforza et. al., 1994), max distances = 1000

	Bas	Sa	Au	Fr	Ge	Be	Dk	Ne	En	Ire	Nor	Sc	Sw	Gr	It	P	Sp	Fi
Basque	0																	
Sardinia	261	0																
Austria	195	294	0															
France	93	283	38	0														
Germany	169	331	19	27	0													
Belgium	107	256	16	32	15	0												
Denmark	184	348	27	43	16	21	0											
Netherlands	118	307	38	32	16	12	9	0										
England	119	340	55	24	22	15	21	17	0									
Ireland	145	393	115	93	84	75	68	76	30	0								
Norway	195	424	61	56	21	24	19	21	25	79	0							
Scotland	146	357	74	62	53	59	40	48	27	29	58	0						
Sweden	168	371	80	78	39	34	36	41	37	94	18	74	0					
Greece	231	190	86	131	144	103	191	199	204	289	235	253	230	0				
Italy	141	221	43	34	38	30	72	64	51	132	88	112	95	77	0			
Portugal	145	340	48	48	51	31	77	60	46	115	73	97	78	103	44	0		
Spain	104	295	69	39	69	42	80	76	47	113	97	100	99	162	61	48	0	
Finland	236	334	77	107	77	63	96	123	115	223	94	166	82	150	94	119	159	0

Table A.2: Cultural distances (World Values Survey), max distance=100.

	Au	Fr	Ge	Be	Dk	Ne	En	Ire	Sw	Gr	It	P	Sp	Fi
Austria	0													
France	28	0												
Germany	19	27	0											
Belgium	20	16	23	0										
Denmark	34	26	31	27	0									
Netherlands	30	25	27	21	26	0								
England	25	22	25	20	27	22	0							
Ireland	31	32	38	26	36	31	22	0						
Sweden	30	26	27	26	22	23	24	34	0					
Greece	27	32	32	29	41	38	28	32	37	0				
Italy	23	24	28	22	34	29	22	23	32	24	0			
Portugal	23	29	28	25	41	37	27	28	38	28	18	0		
Spain	24	22	26	19	32	26	22	24	32	30	19	21	0	
Finland	27	34	27	30	34	31	26	37	28	30	32	32	32	0

References

[1] Alesina, A., Baqir, R. and W. Easterly (1999), "Public Goods and Ethnic Divisions Size of Nations," *Quarterly Journal of Economics* 114, 1243-1284.

[2] Alesina, A., Baqir, R. and C. Hoxby (2004), "Political jurisdictions in Heterogenous Communities," *Journal of Political Economy* 112, 348-396.

[3] Alesina, A., Easterly, W. and J. Matuszeski (2006), "Artificial States," NBER Working Paper #12338.

[4] Alesina, A. and E. Spolaore (1997), "On the Number and Size of Nations," *Quarterly Journal of Economics* 112, 1027-56.

[5] Alesina, A. and E. Spolaore (2003), *The Size of Nations*, MIT Press, Cambridge, MA.

[6] Ashraf, Q. and O. Galor, (2007), "Cultural Assimilation, Cultural Diffusion and the Origin of the Wealth of Nations," mimeo.

[7] Axelrod, R. (1997), "Choosing Sides,"Chapter 4 in *The Complexity of Cooperation*, Princeton University Press, Princeton, NJ.

[8] Axelrod, R. and D.S. Bennett (1993), "A Landscape Theory of Aggregation," *British Journal of Political Science*, 23, 211-233.

[9] Banerjee, S., Konishi, H. and T. Sömnez (2001), "Core in a Simple Coalition Formation Game,"*Social Choice and Welfare* 18, 135-153.

[10] Bogomolnaia and M.O. Jackson (2002), "The Stability of Hedonic Coalition Structures," *Games and Economic Behavior* 38, 201-230.

[11] Bogomolnaia, A., Le Breton, M., Savvateev, A. and S. Weber (2007), "Stability under Unanimous Consent, Free Mobility and Core," *International Journal of Game Theory* 35(2), 185-204.

[12] Bolton, P. and G. Roland (1997), "The Breakup of Nations: A Political Economy Analysis," *Quarterly Journal of Economics* 112, 1057-1090.

[13] Campbell, J.Y. (1999) "Asset Prices, Consumption, and the Business Cycle" in *Handbook of Macroeconomics*, vol. 1c, J. B. Taylor and M. Woodford, eds., Elsevier, Amsterdam.

[14] Cavalli-Sforza, L.L. and W.F. Bodmer (1999), *The Genetics of Human Populations*, Dover Publications, Mineloa, NY.

[15] Cavalli-Sforza, L.L. and M. W. Feldman (2003), "The Application of Molecular Genetic Approaches to the Study of Human Evolution," *Nature Genetic Supplement* 33, 266-275.

[16] Cavalli-Sforza, L.L., Menozzi P., and A. Piazza (1994), *The History and Geography of Human Genes*, Princeton U.P., Princeton, NJ.

[17] Chikhi, L., Nichols, R.A., Barbujani, G. and M.A. Beaumont (2002), "Y Genetic Data Support the Neolithic Demic Diffusion Model," *Proceedings of National Academy of Sciences* 99 (17), 11008-11013.

[18] Collado, L. , Ortuño-Ortín, I. and A. Romeu (2005), "Vertical Transmission of Consumption Behavior and the Distribution of Surnames", mimeo, University de Alicante.

[19] Desmet, K., Ortuño-Ortín, I. and S. Weber (2005), "Peripheral Diversity and Redistribution," CEPR Discussion Paper # 5112.

[20] Drèze, J. and J. Greenberg (1980), "Hedonic Coalitions: Optimality and Stability,"*Econometrica* 48, 987-1003.

[21] Dyen I., Kruskal J.B., P. Black (1992), "An Indo-European Classification: a Lexicostatistical Experiment," *Transactions of the American Philosophic Society* 82, 1-132.

[22] Friend, I. and M. Blume, (1975), "The Demand for Risky Assets," *American Economic Review* 65, 900-922.

[23] Gans, J. and M. Smart (1996), "Majority Voting with Single-Crossing Preferences," *Journal of Public Economics* 59, 219-237.

[24] Ginsburgh, V., Ortuño-Ortín, I., and S. Weber, (2005), "Disenfranchisement in Linguistically Diverse Societies: The Case of the European Union," *Journal of the European Economic Association* 3, 946-965.

[25] Giuliano, P., Spilimbergo, A., and G. Tonon (2006) "Genetic, Cultural and Geographical Distances," CEPR Discussion Paper # 5807.

[26] Greenberg, J. (1977) "Existence of an Equilibrium with Arbitrary Tax Schemes for Financing Local Public Goods," *Journal of Economic Theory* 16, 137-150.

[27] Guiso, L., Sapienza, P. and L. Zingales (2005), "Cultural Biases in Economic Exchange", mimeo.

[28] Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Werner Alt, K. and J. Burger (2003), "Ancient DNA from the first European farmers in 7500-year-old Neolithic sites", *Science*, 310, 1016-1018.

[29] Hartl, D. L. and A. G. Clark (1997), *Principles of Population Genetics*, Third Edition, Sinauer, Subderland, MA.

[30] Jéhiel, P. and S. Scotchmer (2001), "Constitutional Rules of Exclusion in Jurisdiction Formation," *Review of Economic Studies* 68, 393-413.

[31] Jobling, M.A., Hurles, M.E. and C. Tyler-Smith (2004), *Human Evolutionary Genetics. Origins, People & Disease*, Garland Science, New York, NY.

[32] Jorde, L.B. (1985), "Human Genetic Distance Studies: Present Status and Future Prospects," *Annual Review of Anthropology* 14, 343-373.

[33] Le Breton, M. and S. Weber (1995), "Stability of Coalition Structures and the Principle of Optimal Partitioning," in *Social Choice, Welfare and Ethics*, Barnett,W., Moulin, H., Salles, M. and N. Schofield, eds., Cambridge University Press, Cambridge, MA.

[34] Legendre P. and L. Legendre (1998). *Numerical Ecology*, Elsevier, New York, NY.

[35] Mantel N. (1967), "The Detection of Disease Clustering and a Generalized Approach," *Cancer Research* 27, 209-220.

[36] Milchtaich, I. and E. Winter (2002), "Stability and Segregation in Group Formation," *Games and Economic Behavior* 38, 318-346.

[37] Rosser, Z.H. *et al.* (2000), "Y-Chromosomal Diversity in Europe is Clinical and Influenced Primarily by Geography, Rather than by Language," *American Journal of Human Genetics* 67, 1526-1543.

[38] Shiller, R. and S. Athanasoulis (1999), "World Income Components: Measuring and Exploiting International Risk Sharing Opportunities," *National Bureau of Economic Research*, Working Paper # 5095.

[39] Simoni, L., Calafell, F., Pettener, D., Bertrantpetit, J. and G. Barbujani (2000), "Geographic Patterns of mtDNA Diversity in Europe", *American Journal of Human Genetics* 66, 262-268.

[40] Smouse P.E., Long J.C. and R.R. Sokal (1986), "Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence," *Syst. Zool* 35, 627-632.

[41] Sokal, Robert R. (1987), "Genetic, Geographic and Linguistic Distances in Europe," *PNAC*, vol 18, 1722-1726.

[42] Sokal, R.R. and F.J. Rohlf (1995), *Biometry: The Principles and Practice of Statistics in Biological Research*, W. H. Freeman and Company, New York, NY.

[43] Spolaore, E. and R. Wacziarg (2005), "Borders and Growth," *Journal of Economic Growth* 10, 331-386.

[44] Spolaore, E. and R. Wacziarg (2006), "The Diffusion of Development," mimeo.

[45] UNU-WIDER World Income Inequality Database, Version 2.0a, June 2005. http://www.wider.unu.edu/wiid/wiid.htm.