The Stability and Breakup of Nations: A Quantitative Analysis*

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

July 2011

Abstract

This paper quantitatively analyzes the stability and breakup of nations. The tradeoff between increasing returns in the provision of public goods and the costs of greater cultural heterogeneity mediates agents' preferences over different geographical configurations, thus determining the likelihood of secessions and unions. After calibrating the model to Europe, we identify the regions prone to secession and the countries most likely to merge. We then estimate the implied monetary gains from EU membership. As a test of the theory, we show that the model can account for the breakup of Yugoslavia and the dynamics of its disintegration. We find that economic differences between the Yugoslav republics determined the order of disintegration, but cultural differences, though small, were key to the country's instability. The paper also provides empirical support for the use of genetic distances as a proxy for cultural heterogeneity.

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

*We thank Lola Collado, Andrés Romeu, Mattias Polborn, Gérard Roland, Tom Romer, Christian Schultz, Romain Wacziarg, two anonymous referees, and the participants of the PIER Political Economy Conference at University Pennsylvania helpful comments. Financial aid from the Spanish Ministry of Science (ECO2008-01300) and the Fundación BBVA 3-04X 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, New Economic School, Moscow, Russia, 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 closer cooperation, as epitomized by the European Union, and to a lesser extent, ASEAN. Theory suggests 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 favor smaller countries (see, e.g., Alesina and Spolaore, 1997, 2003, and Bolton and Roland, 1997). Although there is some reduced-form evidence supportive of this tradeoff,[1] an empirical and quantitative exploration of its plausibility within a structural model has so far been lacking. Our paper addresses this shortcoming by proposing a theoretical model, calibrating it to the European experience, and analyzing its predictions for the breakup of Yugoslavia and for other 'weak links' in the European map.

In our theoretical model with multiple regions and countries, agents in every country vote on the optimal level of public spending, 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 a region's welfare across different political arrangements. This allows us to study whether regions or countries would like to unite or secede. We use two alternative stability concepts: one requires that any rearrangement should have the support of the majority in each affected region, while another allows for unilateral secessions.

Since the main goal of the paper is to empirically explore the stability of countries, the most crucial issue in linking our model to the data is the empirical measurement of cultural heterogeneity. We accept the view that the degree of mixing 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] It is

[1] Alesina, Spolaore and Wacziarg (2000) uncover a positive relation between the number of countries and the degree of trade openness; Alesina and Wacziarg (1998) find that smaller countries are more open; Sorens (2005) finds that larger richer regions with a different language are more likely to support secessionist movements; and Sambanis (2006) provides a descriptive overview of this literature.

[2] For applications of genetic data in economics, see Spolaore and Wacziarg (2009), who use genetic distance between populations to study the diffusion of economic development, and Ashraf and Galor (2008), who study the relation between development and genetic diversity within populations.

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

To assess the empirical plausibility of the theory, we calibrate our model to the current European situation. In particular, we determine the parameter values that are consistent with the European map being stable. We then ask the following questions:[3] 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? To answer these questions, we start by increasing the cost associated with cultural heterogeneity, and check which region would be the first to secede. We find that the Basque Country and Scotland are the most likely regions to break away. Consistent with this finding, these are the only two regions in Western Europe where local governments have called for referendums on self-determination. Likewise, by decreasing the perceived cost of cultural heterogeneity, we determine which countries are most likely to unite. When focusing on neighboring countries, Austria and Switzerland top the list of 231 possible pairs, followed by Denmark-Norway, France-Britain, France-Italy, and Belgium-Netherlands. These results suggest that unification is more likely when countries have similar population sizes, similar levels of GDP per capita, and similar cultures. Note that two out of these five pairs (Denmark-Norway and Belgium-Netherlands) were actually united for parts of their history. As a first application of the theory we then explore the gains (and losses) from European Union membership. Focusing on the EU-15, we find that the monetary gains as a share of GDP go from 0% in the case of Germany to 13% in the case of Portugal. All else equal, richer countries gain less than poorer countries (Sweden gains less than Portugal), and cultural distances also reduce gains (Greece gains less than Portugal, in spite of being poorer).

As a more direct test of our theory we analyze whether our calibrated model is able to account for the breakup of Yugoslavia. We start by determining whether the theory predicts that any of the six Yugoslav republics wants to unilaterally break away from the rest of Yugoslavia. If at least one republic prefers to leave the union, we conclude that Yugoslavia is unstable. We then go one step further and study the dynamic process of disintegration. We assume that the republic that gains most from secession breaks away first. Once that republic has left, we recompute the incentives for secession for the remaining republics, given that they are now part of a diminished Yugoslavia. The next republic to go is then the one who gains most from leaving whatever is left

[3] Although we focus on Europe, this setup could easily be applied to other parts of the world.

over of Yugoslavia. This process continues until either the whole country has disintegrated or no republic experiences gains from secession.

We find that Yugoslavia was less stable than the rest of Europe. In particular, using parameter values for which, say, the Basque Country and Scotland, have no incentive to secede, Slovenia and Croatia did prefer to leave Yugoslavia. Overall, the model accounts well for the order of disintegration: Slovenia and Croatia secede first, followed possibly by Bosnia and Macedonia, whereas Montenegro prefers to stay in the union. These predictions are broadly consistent with historical events. After the dissolution of the all-Yugoslav Communist Party in 1990 Slovenia and Croatia were the only two republics pushing for more devolution, and they both declared independence on the same day, June 25, 1991. Macedonia and Bosnia followed close behind, and by 1992 only the union between Montenegro and Serbia was left. Although Montenegro eventually also declared independence from Serbia, this happened 15 years later, in 2006, after a small majority voted in favor of self-determination. Another relevant question in the context of Yugoslavia is what was the relative importance of culture and economics in the country's disintegration. Although cultural distances between the different Yugoslav republics were not particularly large, we find that economic differences alone were not big enough to pull the country apart. If Yugoslavia had been culturally homogenous, the model predicts the country would have stayed together. That being said, the order of disintegration is largely driven by economics. The first two republics to leave, Slovenia and Croatia, were, respectively, the richest republic and the largest republic, whereas the republic that never wants to leave according to the model, Montenegro, is very small, with a population of just over half a million.

The ability of the model to account for the breakup of Yugoslavia and to closely match the order of its disintegration lends credibility to our theory. It also provides a new tool for the analysis of the stability of nations. In the last couple of years there has been much talk about the viability of countries such as Iraq and Bolivia to survive as unified states. Although these questions are difficult and complex, and answering them requires in-depth knowledge that goes beyond the simplicity of our framework, our paper shows that having a structural model that can make quantitative predictions can shed light on these important policy questions.

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 heterogeneous 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.[4]

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 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 (see, e.g., Fearon, 2003). 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.[5]

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 30 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

[4] 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.

[5] A 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.

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.[6]

Since we use information of many different genes, the use of genetic distances requires our theory to embed those distances 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,[7] 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.

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. A major drawback of their work is the absence of economic forces. Other related work is Spolaore and Wacziarg (2005) who 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. Also of interest is a recent paper of Alesina, Easterly and Matuszeski (2010) that 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 provides empirical support for using genetic distances as a proxy of cultural heterogeneity. Section 4 calibrates the model to Europe, explores the likelihood of secessions and unions, and

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

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

analyzes the gains from EU membership. Section 5 tests the theory by analyzing whether the model can account for the breakup of Yugoslavia and the order of its disintegration. Section 6 concludes.

2 The Model

In this section we describe the structure of the model. The world is made up of agents, regions and countries. Agents reside in regions and are geographically immobile. Different agents are endowed with different levels of income. The set of regions is exogenously given. Regions are culturally homogeneous. Countries consist of one or more regions. Cultural differences across regions imply that multi-region countries are culturally heterogeneous. The partition of the set of regions into countries can change, and our goal is to analyze the stability of partitions. Agents derive utility from the consumption of a private good and a public good, which is financed through a country-wide proportional tax rate. The utility derived from the public good is decreasing in the degree of cultural heterogeneity. Tax rates are decided by majority vote at the country level. If a change in partition affects the country a region belongs to, the utility of an agent in that region changes, because the tax rate, the level of the public good, and the degree of cultural heterogeneity change. By imposing *intermediate preferences*, the median income agent in each region represents the majority of its residents, in spite of the multi-dimensionality problem. We consider two different stability concepts to assess the possibility of map redrawing. One requires the consent of all affected regions, whereas the other allows for unilateral secession. In what follows we present the model's setup, discuss the preferences over different partitions, define the stability concepts, and end by presenting the specific functional forms we will use in the empirical section.

2.1 Setup

The world is partitioned into countries, indexed by C, each consisting of one or several regions, indexed by I or J. The set of regions, denoted by N, is exogenously given, whereas the partition of regions into countries can change. Denote by Π the set of possible partitions of regions into countries, and by π an element of that set. Regions are populated by agents, who are geographically immobile. The population of country C is then

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

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

There are two types of heterogeneity, cultural and income. Within each region, there is income heterogeneity, but no cultural heterogeneity. That is, residents of a same region may have different incomes, but are culturally identical. In countries that consist of multiple regions both types of heterogeneity, cultural and income, are present. That is, residents of different regions, in addition to possibly having different incomes, may be culturally different.

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.[8] Obviously, $d(I, J) = d(J, I)$ for all I and J, and $d(I, I) = 0$ for all I. The *cultural heterogeneity* experienced by a resident of region $I \in C$, denoted by $H(I, C)$, is the weighted cultural distance between that resident and all other residents of C. This corresponds to the expected cultural distance between an agent from region I and a randomly drawn agent from country C. Hence,

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

Each agent i is endowed with a certain amount of income, y_i. The income distribution in region I is given by the cumulative density function $F_I(y)$ with total mass $p(I)$, whose support $[\underline{y}, \bar{y}]$ is common to all regions. We assume that $F_I(y)$ is continuous and strictly increasing. The total income in I is denoted by $Y(I)$:

$$Y(I) = \int_{\underline{y}}^{\bar{y}} y \; dF_I(y) \tag{2}$$

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

The utility of resident i of region $I \in C$, $u(x, g, H(I, C))$, depends on private consumption, x, the level of the public good in his country, g, and the degree of cultural heterogeneity he faces, $H(I, C)$:

$$u(x, g, H(I, C)) \tag{3}$$

The utility function u is twice continuously differentiable, strictly concave and increasing in x and g. For simplicity, the prices of the private good and the public good are both set to 1.

Cultural heterogeneity reduces the utility, so that u is decreasing in its last argument, $H(I, C)$. Alesina, Baqir and Hoxby (2004) offer two reasons for this *homogeneity bias*, reflecting

[8] In the empirical part of our investigation we identify $d(I, J)$ with the genetic distance between regions I and J.

the preference of agents for more culturally homogeneous communities. One is that individuals who share a common background may have similar preferences over public goods. Another is that, even if individuals have similar preferences to those in other groups, they may still prefer to interact with members of their own group. The specific functional form we will use in the empirical section will assume that cultural heterogeneity enters negatively by reducing the utility an agent derives from the consumption of the public good.

Public goods are specific to each country, and are financed through a proportional tax rate $\tau, 0 \leq \tau \leq 1$, chosen by majority voting. As is obvious, this implies redistribution.[9] If country C selects tax rate τ, the level of the public good will be $\tau Y(C)$. The indirect utility v of resident i of region $I \in C$, with income y_i and paying a tax rate τ, can be written as

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

Note that for every country C the preferences of every agent $i \in C$ over tax rates are single-peaked. Denote by $\tau(y_i, I, C)$ the preferred tax rate for an individual i with income y_i who resides in region $I \in C$. Hence,

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

2.2 Preferences over Partitions

If a region changes the country it belongs to, residents in that region will face a different degree of cultural heterogeneity, a different tax rate, and a different level of public good.[10] An agent residing in region I, faced with the choice of I belonging to C or C', prefers the first alternative if

$$v(y_i, \tau(C), I, C) > v(y_i, \tau(C'), I, C')$$

where $\tau(C)$ is the tax rate chosen by majority voting at the country-level.[11] This implies that an agent understands that the resulting tax rate depends on the country his region belongs to.

Since changing the country a region belongs to affects the tax rates, public goods, and cultural heterogeneity, it is not immediately obvious who will be a region's decisive agent. We assume that agents have *intermediate preferences* of Grandmont (1978). This assumption, satisfied

[9] Consistent with the empirical literature, the degree of redistribution will be decreasing in the level of cultural heterogeneity. We return to this point in Section 2.4.

[10] The level of the public good depends both on the tax rate and the total income of the country.

[11] Since preferences over τ are single-peaked, this $\tau(C)$ is well defined.

by a large class of functions,[12] solves the multi-dimensionality issue, as it implies that the preferences of a region's median income agent represent those of the majority of its residents. That is, if the median income agent of region I prefers to be part of country C, rather than of C', then a majority of agents in region I also prefer to be part of country C.[13] This is referred to as the *decisiveness* of the median agent (Gans and Smart, 1996).

Median decisiveness will facilitate our analysis of the stability of a partition, as it allows us to only focus on the preferences of the median income agents of the different regions. It is important to point out that in our framework the benefit of a region from being part of a certain country depends solely on the composition of that country and not on the number and composition of the other countries. This 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.[14]

2.3 Stability

We now turn our attention to the stability of partitions. Several stability concepts have been applied in the literature.[15] We consider two such concepts in our paper. The first one, called *Limited Right of Map Redrawing*, requires, subject to majority voting, the approval of any map redrawing by all affected regions.[16] The second concept, called *unilateral secession*, is a modification of the *core*, and it requires only the approval, by majority voting, of the seceding region.[17]

[12] See, e.g., Greenberg and Weber (1985) and Demange (1994). In particular, it contains the functional forms considered 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, it also contains the utility functions of the type $u(x, g, I, C) = x^{\delta} + \Phi(g, H(I, C))$, where Φ is an arbitrary continuous function, with $0 < \delta \leq 1$. This includes the case the quasi-linear case with δ equal to 1, considered in our empirical investigation.

[13] This result is proven in Proposition 1 of the working paper. See Desmet et al. (2009).

[14] 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.

[15] See, e.g., Alesina and Spolaore (1997), Jéhiel and Scotchmer (2001), Bogomolnaia et al. (2007).

[16] This stability concept is referred to as *B-stability* in Alesina and Spolaore (1997) and *equilibrium with admission by majority vote* in Jéhiel and Scotchmer (2001). If the majority requirement were to be replaced by unanimous consent, this stability concept would be reminiscent of *contractual individual stability*, studied by Greenberg (1977), Drèze and Greenberg (1980), Bogomolnaia and Jackson (2002).

[17] This resembles *C-stability* in Alesina and Spolaore (1997).

2.3.1 LRMR-Stability

The *Limited Right of Map Redrawing* stability concept, denoted LRMR-stability, says that a partition is stable if there is no other partition preferred by a majority of agents in *each* region affected by the change in partition. In other words, changing a partition would need the consent of all affected parties. The Canadian Clarity Act of 2000, which regulates the possible secession of a Canadian province, is an example. Under this act, secession requires an amendment to the constitution, thus involving negotiations with the governments of *all* provinces. We now formalize the LRMR-stability concept, and argue that the set of LRMR-stable partitions is nonempty.

Definition 1: LRMR-Stability. *A partition $\pi \in \Pi$ is LRMR-stable if there is no $\pi' \in \Pi$ for which the majority of residents of every region affected by the shift from π to π' prefers partition π' over π.*

Proposition 1. *The set of LRMR-stable partitions is nonempty.*

The proof is presented in the Appendix.

Recall that Π was defined as the set of *possible* partitions. When empirically analyzing the stability of a given partition, it may be convenient to restrict the set of *possible* partitions to be smaller than the set of *all* partitions. Indeed, the formation of a country may be limited by various geographical, political or historical considerations. For example, it is unusual for countries to consist of geographically disconnected regions. Previous violent conflicts or wars may also prevent the co-existence of certain regions within the same country. Some of these restrictions, which are exogenous to the model, will play a role in the empirical analysis.

2.3.2 ULS-Stability

LRMR-stability assumes that no map redrawing can take place without the consent of *all* affected regions. However, in some countries or federations there are constitutional provisions that allow for unilateral secession. For example, Article 39(1) of the Ethiopian constitution reads "every nation, nationality, and people in Ethiopia has the unconditional right to self-determination, including the right to secession". Of course, the absence of constitutional provisions that regulate exit does not mean unilateral secessions are never attempted, with or without success. The secession of the Netherlands from Spain in the 16th century, the split up of Norway and Sweden in 1905, the failed

secession of the Confederate States in 1861, and the break-up of Yugoslavia in the last decade of the 20th century, are just some examples.

To deal with this possibility, we introduce the *unilateral secession* stability concept, denoted ULS-stability, which says that a partition is stable only if no majority in any of the regions wants to break away. Let $C^I(\pi)$ denote the country in π that contains region I.

Definition 2: ULS-Stability. *A partition $\pi \in \Pi$ is ULS-stable if there is no region I for which the majority of its residents prefers I to be a separate country instead of I being a part of country $C^I(\pi)$.*

It is trivial to see that the set of ULS-stable partitions is non-empty. Indeed, a partition that consists of each region being a separate country is always ULS-stable. Therefore,

Proposition 2. *The set of ULS-stable partitions is nonempty.*

As in the case of LRMR-stability, in our empirical applications we may wish to restrict the set of possible partitions Π to take into account certain limitations that have not been explicitly modeled. For example, the core region of a country, in its capacity of hegemon, may perceive additional benefits from keeping the country together, thus preventing it from wanting to leave the union.

When studying the possible break-up of a country, we are not only interested in its stability, but also in understanding the relative incentives of different regions to secede, and the possible dynamics of disintegration. To address this issue, let us introduce a secession function, $s(I, C)$, which measures the propensity of I to secede from C. In the empirical section, $s(I, C)$ will be a function of the utility difference between separating and remaining within the union, $v_m(I, I) - v_m(I, C)$. Consider the following process. Take $\max_I s(I, C^I(\pi_0))$, where π_0 is the starting partition. If it is non-positive, no secession takes place, and the original partition is ULS-stable. If not, the region with the maximum value secedes, generating a new partition, π_1. We then take $\max_I s(I, C^I(\pi_1))$. If it is non-positive, no further secession takes place. If not, the region with the maximum value secedes, generating the partition π_2, and so on. Since the process is finite, there is a terminal partition, which is obviously ULS-stable.

2.4 Our Specification

To bring our theoretical model to the data, we adopt the following quasi-linear functional form for the utility of a resident of region $I \in C$:

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

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. This utility function satisfies the *intermediate preferences* assumption.

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

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

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.

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 $I \in C$:

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

As can be seen, the optimal tax rate is decreasing in the level of cultural heterogeneity, $Z(I, C)$.[18]

[18] This is consistent with the findings of La Porta et al. (1999) and Alesina et al. (2003) who provide evidence of a negative relation between ethnolinguistic diversity and the willingness to redistribute. These papers measure diversity without taking into account the distance between groups, i.e., how distinct the different groups are. When including ethnolinguistic distances between groups, the negative relation between diversity and redistribution is further strengthened (Desmet et al., 2009). For a survey of this literature, see Ginsburgh and Weber (2011).

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.

3 Genetic and Cultural Distances

In our quantitative analysis we will use genetic distances as a proxy for cultural distances between populations. In the introduction we argued that genetic distances are the best objective measure of the intensity of population mixing. Therefore, in as far as cultural diffusion happens through mixing, genetic distances should be a reliable measure of cultural distances. To further justify our choice of proxy, we study the correlation between genes and culture by comparing a matrix of genetic distances to a matrix of cultural distances. The genetic distances come from Cavalli-Sforza et al. (1994),[19] whereas the cultural distances are based on the World Values Survey (WVS).

[Figure 1 about here]

To compute cultural distances from the WVS, we take the 430 questions related to *Perception of Life, Family,* and *Religion and Moral* from the four waves currently available online at http://www.worldvaluessurvey.org/. Each question has q different possible answers and we denote by $a_{i,j} = (\ a^1_{i,j}, a^2_{i,j}, \ldots, a^q_{i,j})$ the vector of relative answers to question i in nation j. For example, suppose that question i has three possible answers, a^1, a^2 and a^3. The vector $a_{i,j} = (1/2, 0, 1/2)$ indicates that in nation j, half of the people answers a^1 and the other half a^3. We construct a

[19] See Hartl and Clark (1997) for an introduction to population genetics, and Jorde (1985) and Cavalli-Sforza et al. (1994) for a discussion on the use of the different types of genetic distances to measure human population distances. 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.

matrix of cultural distances between nations such that the (j,k) element of the matrix represents the average Manhattan distance between nation j and nation k:

$$w_{jk} = \frac{1}{430} \sum_{i=1}^{430} \sum_{s=1}^{q} \left| x_{i,j}^{s} - x_{i,k}^{s} \right| \tag{10}$$

All our results are robust to the usage of the Euclidean distance instead of the Manhattan distance in (10).

Both distance matrices are reported in the Appendix. Table B.1 gives the matrix of genetic distances, denoted by D, and Table B.2 gives the matrix of cultural distances, denoted by W. Although matrix D covers more countries, and also some regions within countries, the comparison between both matrices is done on the basis of 14 European countries.

As a first step to verify whether genetically close countries provide similar answers to the questions in the World Value Survey, Figure 1 shows a scatter plot representing genetic distances on the horizontal axis and WVS distances on the vertical axis. Each point in the scatter plot corresponds to a nation pair (i,j), where the x-coordinate is the genetic distance between i and j, and the y-coordinate is the WVS distance between i and j. Figure 1 suggests a strong positive relation between genetic and cultural distances.

We now compute the correlation between both matrices, and assess its statistical significance. Due to the triangle inequality property, the elements of a distance matrix are not independent, so that we cannot use standard methods to determine the statistical significance of the observed correlation between matrices D and W. To address this issue, we rely on the *Mantel test*, a nonparametric randomization procedure often used in Population Genetics.[20] The significance of the correlation is evaluated via the random permutation of rows and columns of one of the matrices. For each random permutation, the correlation r between the two matrices is re-computed. After a sufficient number of permutations, the distribution of values of r is generated, and the critical value of the test at the chosen level of significance is determined.

The correlation coefficient between matrices D and W is 0.56 and the hypothesis of non-positive correlation is strongly rejected based on a Mantel test with $100,000$ replications (p-value of 0.000). This highly significant correlation provides the foundation for our use of the matrix of genetic distances as a proxy for the cultural heterogeneity among European countries. This

[20] 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).

correlation is robust to the different subgroups of questions in the World Value Survey we consider. Computing the corresponding the correlations with *Perceptions of Life* (307 questions), *Family* (77 questions) and *Religion and Moral* (165 questions) yield 0.45, 0.67 and 0.68.

If the justification for using matrix D is based on its correlation with the matrix W, one might argue 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 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.

Of course there may be other proxies for cultural distances apart from genetic distances. Two alternatives come to mind: geographic distances and linguistic distances. To explore this possibility, we compute a matrix of geographic distances, G, and a matrix of linguistic distances, L, between European countries.[21] The correlation between geographic distances (G) and cultural distances (W) is 0.52 (p-value of 0.000) and the correlation between linguistic distances (L) and cultural distances (W) is 0.42 (p-value of 0.000). Not surprisingly, these correlations are also positive and statistically significant. In fact, as has been argued by Cavalli-Sforza et al. (1988), there is a strong link between genetic distances, geographic distances and linguistic distances. Subsequent studies, such as the one by Belle and Barbujani (2007), have tended to confirm this view.

To determine which distance measure is a more convincing proxy of cultural distances, we compute several partial correlations. We start by determining the significance of the partial correlation between genetic distances (D) and cultural distances (W), after controlling for geographic distances (G) and linguistic distances (L). This is done by using a partial Mantel test, which consists of constructing a matrix of residuals, D', of the regression of D on G and L, and a matrix of residuals, W', of the regression of W on G and L, and then comparing D' and W' using a standard Mantel test.[22] The correlation between D' and W' is 0.28, still significantly greater than zero (p-value of 0.04). To understand what this means, consider the following example. Say country i

[21] Geographic distances were calculated "as the crow flies", and the coordinates of each region were obtained from Simoni et al. (2000). Linguistic distances between countries are computed using data on the linguistic composition of each country from the Ethnologue project and data on the distance between Indo-European languages from Dyen, Kruskal and Black (1992). For the use of linguistic distances in economics, see Fearon (2003), Ginsburgh et al. (2005) and Desmet et al. (2009).

[22] See Smouse et al. (1996).

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. We now assess the significance of the partial correlation between geographic (G) and cultural distances (W), after controlling for genetic (D) and linguistic distances (L). We find a partial correlation of 0.20. Though positive, it is not statistically significant (p-value of 0.105). A similar exercise for the partial correlation between linguistic (L) and cultural distances (W), after controlling for genetic (D) and geographic distances (G), gives a partial correlation of 0.11 with a p-value of 0.185.

In other words, when controlling for the other two distances, only the partial correlation between genetic distances and cultural distances continues to be positive and statistically significant at the 5% level. Although linguistic and geographic distances are also reasonable proxies for cultural distances between populations, this result provides an argument in favor of using genetic distances.

4 Stability of Europe

In this section we use information on cultural distances between European regions and countries to estimate values of α, β and δ that yield an LRMR-stable partition of Europe. This exercise is of interest in itself, since as a way of validating our theoretical framework, it seems important that the set of parameter values consistent with stability is not empty. Once we have calibrated the parameter values to the stability of Europe, we explore the 'weak links' in the current map of Europe by analyzing which regions have a higher propensity to secede and which country pairs have a higher propensity to unite. This exercise aims to illustrate the basic forces at work in the theoretical model. As a first real application of the theory, we then analyze the gains (and losses) from European Union membership. A second application, on the breakup of Yugoslavia, will be discussed in Section 5.

4.1 Data

We rely on genetic distances to proxy for cultural distances. To find parameter values that render the European map stable, 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 must also guarantee that no existing regions want to separate. The matrix of Cavalli-Sforza et al. (1994), reproduced in Table B.1 in the Appendix, is

therefore appropriate, as it contains information on 22 European countries and 4 European regions (Basque Country, Sardinia, Scotland and Lapland).[23,24] Although it leaves out a number of relevant regions (Flanders, Catalonia, Brittany, Northern Italy, Corsica, etc.), it does include the only two regions in Europe (Scotland and the Basque Country) where in recent years local governments have called for referendums on self-determination.[25] Given that we have data on two of the regions with most popular support for greater devolution, and even outright independence, this should allow us to estimate a reasonable lower bound on δ.

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, 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.

4.2 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. Because we have no comprehen-

[23] 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.

[24] Genetic distances in Cavalli-Sforza et al. (2004) are between populations rather than between nations. However, in the case of Europe most of the ethnic groups defined by Cavalli-Sforza coincide with the corresponding nations. For example, what Cavalli-Sforza et al. (2004) call "French" coincides with the population of France, and includes different regional groups, such as Bretons or Flemish.

[25] In Scotland, the SNP has committed to holding an independence referendum, whereas in the Basque Country the regional government's *Plan Ibarretxe* that called for a referendum in 2008 got overturned in Spanish courts. In terms of popular support, according to an ICM poll conducted for The Sunday Telegraph in 2006, Scottish independence is not only backed by 52% of Scots, but by an even higher 59% of English voters (http://www.telegraph.co.uk/news/main.jhtml?xml=/news/2006/11/26/nunion26.xml). Note, however, that when polls give three options, including greater devolution, instead of outright independence, the support for independence tends to drop to about 20%-30%. In the case of the Basque Country, a poll by the Spanish *Centro de Investigaciones Sociologicas* in 2005 found 35.3% of Basques in favor of independence and 39.4% against.

sive data on genetic distances between regions within countries, our approach is to estimate α and β abstracting from within-country cultural heterogeneity. For this approach to be reasonable, we focus on a set of countries that are culturally relatively homogeneous. For want of data on genetic diversity, we use information on linguistic diversity to define such a set. Based on the ranking of linguistic diversity by Desmet et al. (2009), we only include countries in the bottom one third of that ranking, and take those countries to be culturally homogeneous for our estimation procedure.

The tax rate adopted by a culturally homogeneous country C is

$$\tau(C) = \left(\frac{y_{m(C)}}{\alpha \ \beta \ (Y(C))^{\beta}} \right)^{\frac{1}{\beta - 1}} \tag{11}$$

where $y_{m(C)}$ is the median income in C. To estimate α and β, we thus need data on the tax rate, $\tau(C)$, the median income, $y_{m(C)}$, and the total income, $Y(C)$. For the tax rate, we take the ratio of government spending on public goods to total GDP. As a measure of public goods we take 'general public services' from the Government Finance Statistics (GFS) database, collected by the IMF. We use data for those European and OECD countries in the GFS database that are in the bottom one third of countries in terms of linguistic diversity. This gives us information on a total of 15 countries.

We estimate (11) by applying nonlinear least squares. The results are reported in Table 1. In addition to focusing on the sample of 15 culturally homogeneous countries, for the purpose of subsequent robustness checks we consider a number of other alternatives to estimate α and β. First, we re-estimate the parameter values on the entire sample of European and OECD countries. Although this doubles the sample size, it comes at the cost of including countries that are culturally relatively heterogeneous. Second, rather than considering all countries to be culturally homogeneous, we re-estimate the parameter values assuming that all countries have a common (positive) degree of cultural heterogeneity. In particular, we set the cultural discount factor, $Z(I, C)$, equal to 0.3 for all countries. Recall that $Z(I, C)$ has a value between 0 and 1, with lower values indicating higher degrees of cultural heterogeneity. The resulting estimates of α and β for these alternative approaches are also reported in Table 1.

	α	β
Only linguistically homogeneous countries	22.05	0.095
All countries	25.80	0.0833
All countries ($Z(I, C) = 0.3$)	51.05	0.052

Table 1: Estimation of α and β

Now that we have estimates for α and β, we compute the range of δ that guarantees the stability of the current map of Europe. We restrict our analysis to the set of partitions which contain the current map, all possible secessions (Basque Country-Spain, Scotland-Britain, Sardinia-Italy), and all possible mergers between country pairs. In other words, we do not consider *all* partitions. For example, we do not allow for unions between more than two countries, or for regions switching allegiances from one country to another. In as far as larger unions start off small, and in as far as a region must first secede before joining another country, our restricted set of partitions seems reasonable. In any case, in Section 4.5 we will return to the issue of unions between more than two countries.

To determine the range of δ that ensures LRMR-stability, 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 the seceding region and in the rest of the country. For instance, if the Basque Country is to separate, a majority of Basques and a majority of the population in the rest of Spain should approve.[26] This implies that for the Basque Country and the rest of Spain to remain united, it is enough that one of the two parties prefers to maintain the union.

Formally, suppose region I is part of country 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 $\tau(C)$. The median income agent in region I prefers to stay within the union than to secede if

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

Recall that the higher is δ, the smaller is the cost of heterogeneity, and thus the stronger the preference to remain united. The net gain of the union for the median agent of region I can thus be written as an (increasing) function of δ,

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

A similar condition must hold for the rest of the country. By analogy, for the median income agent of the rest of the country C/I, the net gain of keeping country C united, rather than letting region

[26] If we had data on more Spanish regions, then LRMR-stability would require a majority in all Spanish regions. However, our database never has information on more than one region within each country, so that any possible secession needs the approval of the majority in the seceding region and of the majority in the rest of the country taken as a whole.

I secede, is

$$g_{C/I,C}(\delta) \equiv v_m(C/I, C) - v_m(C/I, C/I) \geq 0. \tag{13}$$

As before, this net gain is an increasing function of δ, since a higher value of δ implies attaching a lower cost to cultural heterogeneity.

In order to prevent secession under LRMR-stability, it suffices that *one* of the two parties, the potentially breakaway region or the rest of the country, prefers to remain united. Thus, a first *necessary* condition for the current European partition to be LRMR-stable is the existence of a nonempty set of the parameter δ for which *at least* one of the functions (12) and (13) 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 thus 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 is obtained numerically. In particular, we find that for any $\delta \geq 0.0284$, no region will secede from their respective country. The no-secession condition thus provides a lower bound on the value of δ. This is consistent with the fact that higher values of δ imply lower costs of cultural heterogeneity.

We now analyze the condition ensuring that no country pairs want to unite. To determine the preferred tax rate in a possible union between, say, C and C', we first need to identify the median voter. Because the cultural heterogeneity 'discount factor' Z is not the same for all agents, 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, 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 the European map to be LRMR-stable is that there is no country pair for which a majority in both countries wants to unite. Formally, 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 is then

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

Numerically, we find that for any $\delta \leq 0.147$, no two countries want to join. The no-union condition thus provides an upper bound on δ. If the value of δ surpasses this upper bound, the importance attached to cultural heterogeneity is no longer sufficient to keep certain country pairs from uniting.

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

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

is non empty. It is clear that S is an interval on the real line, and we write $S \equiv [\underline{\delta}, \overline{\delta}]$. From the numerical values mentioned before, we find that $S = [0.0284, 0.147]$. A first conclusion is therefore that the set of parameter values that render Europe stable is non-empty. Using the two alternative estimates of α and β in Table 1 to determine the interval S gives us nearly identical results. In what follows we use the lower and upper bounds of the interval S to identify the 'weak links' in the European map. We will then use these calibrated parameter values to analyze the gains from European Union membership and the breakup of Yugoslavia.

4.3 Secessions and Unions between Country Pairs

To illustrate the different forces at work in the model, we analyze which European regions have a higher propensity to secede, and which European country pairs have a higher propensity to unite. Recall that if $\delta < 0.0284$, the weight given to cultural distances in the utility function becomes large enough, so that we can no longer prevent secessions from occurring. Therefore, by progressively lowering δ, we can rank regions according to their propensity of secession. Likewise, if $\delta > 0.147$, the weight put on cultural distances becomes small enough, so that some currently independent nations would prefer to unite. Thus, by progressively increasing δ, we can rank country pairs according to their propensity to unite. The aim of this exercise is to gain a better understanding of the model.

For the case of secessions, the model predicts that the Basque Country has the highest propensity to break away, followed by Scotland and Sardinia. This ranking is unchanged under a

number of robustness checks.[27] These results are consistent with the observation that the Basque Country and Scotland are the only two regions in Western Europe that in recent years have called for referendums on self-determination.

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

Table 2: Propensity to unite

For the case of unions, Table 2 ranks the propensity of country pairs uniting. Out of the 231 possible unions, the first column reports the top-10 of most likely unions. Austria and Switzerland head the ranking: both are small, have similar population sizes, and nearly equal levels of GDP per capita. In addition, they are genetically very close. Many of the other pairs in the top-10, such as Denmark-Norway, Austria-Belgium, and Belgium-Netherlands, fit the same pattern. Unions between rich and poor countries are unlikely, since redistribution would be too high for the rich country's liking. Unions between large and small countries are absent from our list too: the economic gains for the large country would be too limited. However, not all unions are between small countries. Although large countries stand to gain less from unification than small countries, if they are sufficiently similar, they may still prefer to join. This is the case of France and Britain, which rank in fourth position. These two countries have identical population sizes (a difference of 1%), identical levels of GDP per capita (a difference of 2%), and they are in the 10th percentile of genetically least distant countries. This last observation is common to all country pairs in the top-10: all are genetically close, none drops below the 20th percentile of most similar country pairs. Not surprisingly, many of these country pairs were actually united at some point in history. 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

[27] In particular, we used the range of δ, based on the alternative estimates of α and β in Table 1, and also analyzed the results using two alternative definitions of government spending: 'defense' or 'general public services, defense, public order, environment and economic affairs'.

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 Nordic countries continue to closely cooperate up to the present day in a variety of ways, as does the Benelux. Although reunification may not be on the political agenda, recent polls show that more than two thirds of Dutch support a union with the Flemish part of Belgium, whereas about half of the Flemish want to reunite with the Netherlands.[28]

If we were to limit ourselves to geographically contiguous regions, some unions, such as Hungary-Czech Republic and Norway-Sweden would now make it into the top-10, whereas others, such as Belgium-Sweden would drop out (see last column in Table 2). Counterfactual exercises can be used to further enhance our understanding of the model. For example, if all countries had the same population (equal to the average of European countries), then some country pairs, such as Germany and Switzerland, would become likely candidates for unification. Or if all country pairs had the same genetic distances (equal to the average between European countries), some pairs, such as Belgium-Netherlands, would cease to make it to the top-10.

4.4 The Gains of European Union Membership

We now use our model to estimate the gains from being a member of the EU-15. Given that we are relying on our theoretical framework, we are interpreting the EU to be a full-blown political union with common taxes and public policies. In that sense, our analysis should not be viewed as estimating the benefits from EU membership in its current form, but rather as a counterfactual exercise. Our goal is two-fold. First, we want to see which countries gain most from being part of the European Union. Second, we want 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 focus on the monetary gains (equivalent variation) from being part of the EU-15. In particular, we compute the relative increase in per capita income, $r(C)$, that all agents in country C should receive to render its median agent indifferent between joining the EU (and not receiving the additional income $r(C)y_m(C)$) and remaining outside the EU (and receiving $r(C)y_m(C)$). The relative increase (decrease) in income is a measure of the relative monetary gains (losses) from being part of the EU. To determine $r(C)$ for each nation C we solve the following

[28] This is based on polls by television station RTL4 and by the newspapers *De Standaard* and *De Morgen* in 2007.

equation:

$$y_m(C)(1+r(C))(1-\tau'(C)) + \alpha(\tau'(C)Y(C)(1+r(C)))^\beta \tag{14}$$
$$= y_m(C)(1-\tau(EU)) + \alpha(Z_{(C,EU)}\tau(EU)Y(EU))^\beta$$

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 $(1+r(C))$. In terms of data, we focus on the 14 member states of the EU-15 for which we have information.[29]

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 $S = [0.0284, 0.147]$. 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 $r(C)$ over all the parameters in S, namely

$$r(C) \equiv \int_S r(C,\delta)dF \tag{15}$$

where F is the uniform distribution over the interval S. We take an approximation $\widehat{r}(C)$ by computing

$$\widehat{r}(C) \equiv \sum_{i=0}^{50} r(C, \underline{\delta} + \frac{\overline{\delta}-\underline{\delta}}{50}i) \tag{16}$$

Table 3 reports the ranking of relative monetary gains of the different member states of the EU-15. According to our computations, Portugal is the country that gains most, followed by Greece. At the other extreme is Germany, a country that is essentially indifferent about being a part of the EU-15, gaining a meager 0.3% from membership. Except for Spain, gains in the larger countries are relatively small, around 1.5% of GDP.

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

Country	Monetary gain (%)	Population	Cultural distance	GDP per capita	Ranking (no distance)
Portugal	13.1	10	0.051	80	-1
Greece	11.9	10.6	0.142	73	1
Ireland	8.9	3.8	0.095	126	0
Finland	8	5.1	0.105	113	0
Denmark	7.5	5.3	0.045	126	-1
Belgium	7	10.2	0.027	117	-3
Austria	6.8	8.1	0.043	126	-1
Sweden	6.4	8.87	0.067	119	1
Netherlands	5.7	15.9	0.041	120	-1
Spain	4.1	40.3	0.056	92	5
Britain	1.6	58.6	0.034	112	-1
France	1.5	59	0.032	114	-1
Italy	1.5	56.9	0.042	113	2
Germany	0.3	82	0.031	112	0

Table 3: Relative monetary gain from being member of EU

Different variables — population size, GDP per capita, income distribution, and cultural heterogeneity — affect this ranking. Table 3 seems to suggest a strong correlation between population size and relative gains. However, population is not the entire explanation. Greece, Belgium and Portugal, for instance, all have a population size of around 10 million, but Portugal gains more than Greece, and Greece gains more than Belgium. The difference between Belgium and Portugal can be attributed to GDP per capita. Richer countries are forced to redistribute more, and may therefore be less interested in uniting. In contrast, the difference between Greece and Portugal is due to cultural distances: Greece is genetically the *most* distant from the average European country, so that, in spite of being poorer than Portugal, it gains less from EU membership.

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 change in the ranking is given in the last column of Table 3. As expected, Greece now exchanges positions with Portugal. Another interesting case is Belgium. Being the country which is *least* distant from the EU average, it now loses three positions, and now gains less than Sweden. One of the larger countries, Spain, moves up 5 positions, and now gains more than many of the smaller countries, such as the Netherlands, Austria, Belgium or Sweden.

4.5 Full Stability

In the previous sections we limited the set of feasible partitions to the current partition, all unions between country pairs, all secessions, and the EU-15. In order to expand this set, one has to face the constraint of computing capacity. Indeed, the number of all possible partitions of the 21 countries and the 3 regions in our database amounts to 445,958,869,294,805,289. 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 4.3 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 address the problem of global stability and consider the set of *all* possible partitions. To cut through the computational problem, we introduce two restrictions. First, instead of looking at all of Europe, we focus on the EU-15, and leave out some of the peripheral countries, such as Ireland, Finland and Sweden. In the 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

$$\overline{\tau}(C) = \arg \max_{\tau \in [0,1]} \sum_{I \in C} p(I) v(\overline{y}(I), \tau, I, C) \tag{17}$$

where $\overline{y}(I)$ is the mean income in region I. One can easily show that the solution to (17) is given by

$$\overline{\tau}(C) = \left(\frac{1}{\alpha \ \beta \ \sum_{I \in C} p(I) \ (Z_{(I,C)})^{\beta}} \right)^{\frac{1}{\beta - 1}} \frac{1}{Y(C)}. \tag{18}$$

To compute the tax rate (18), we only need information on population, total GDP and cultural distances, without having to identify the median agent for each partition. As a result, calculating welfare for each of the 678,570 partitions becomes a computationally feasible task. It is straightforward to adjust the LRMR-stability concept to make the mean income agent the decisive one. We only need to replace 'the majority of its residents' in Definition 1 by the 'mean income agent'.

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 4.2 and 4.3, using a 'mean agent' rather than a 'median agent' framework. Since none of the previous results changed, adopting this simplification does not seem to come at the cost of losing realism.

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 is therefore LRMR-stable, whereas for $\delta > \delta^*$ the union of all countries maximizes total welfare, so that the EU would be LRMR-stable. In other words, the only two LRMR-stable partitions of Europe is either full integration or full independence.

This result is subject to two caveats. First, the absence of configurations other than full union or full disintegration 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, if we were to focus on Sweden, Denmark and Greece, then for values of $\delta \in [0.18, 0.21]$, the LRMR-stable 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 obviously 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.

5 Breakup of Yugoslavia

In this section we analyze the dynamics of the breakup of Yugoslavia. Within the framework of our model, we ask two questions. Was Yugoslavia unstable? And, if so, what does our model say about the dynamics of its disintegration? Our strategy in answering these questions is as follows. Using the calibrated parameter values for α, β and δ from Section 4, we start by analyzing whether any (or several) of the Yugoslav republics wants to unilaterally break away from the rest of Yugoslavia. If at least one republic prefers to leave the union, we conclude that Yugoslavia is unstable. We then

go one step further by studying the dynamic process of disintegration. We assume that the republic that gains most from secession breaks away first. Once that republic has left, we recompute the incentives for secession for the remaining republics, given that they are now part of a diminished Yugoslavia. The next republic to go is then the one who gains most from leaving whatever is left over of Yugoslavia. This process continues until either all republics have become separate nations or no republic experiences further gains from secession in which case the country stops disintegrating.

Of course our goal is to compare the outcome predicted by the model with what happened in reality. Without going into detail, it is therefore useful to briefly recall the historical timeline of the breakup. After the dissolution of the all-Yugoslav Communist Party in 1990 and multi-party elections in the different republics, Slovenia and Croatia pushed for increasing devolution and independence, while Serbia and Montenegro favored Yugoslav unity. Unable to reach their objectives within a unified Yugoslavia, on June 25, 1991, Slovenia and Croatia became the first republics to declare independence. While in Slovenia independence was obtained without any significant violence, in Croatia war broke out because of opposition from its Serbian minority. A couple of months later, in September 1991, Macedonia followed suit, and seceded in a peaceful manner, without any resistance on the part of the Yugoslav authorities. Half a year later, on April 5, 1992, Bosnia declared independence, after holding a referendum on self-determination. As a reaction, the Serbian minority in Bosnia seceded from the newly created state, and the Bosnian war ensued. De facto, Yugoslavia was now down to a union between Serbia and Montenegro. During the next decade, the two remaining republics gradually drifted apart, culminating in Montenegro becoming independent in May 2006, after 55.5% of its voters supported self-determination in a referendum.

5.1 Data

We compute genetic distances between the six Yugoslav republics — Serbia, Montenegro, Macedonia, Bosnia, Slovenia and Croatia — using allele frequencies of different genes. Given that for the Yugoslav republics Cavalli-Sforza et al. (1994) provide information on only a limited number of genes, we rely on the Allele Frequency Database (ALFRED) at the Yale Medical School to complement our data.[30] As in Cavalli-Sforza (1994), we use the F_{ST} measure to compute genetic

[30] Whenever possible, we use genetic data at the level of the republics, rather than at the level of populations, to compute genetic distances between the republics. In particular, the raw data of Cavalli-Sforza et al. (2004) provide information on the geographic location of the samples.

distances, thus making our distance matrix comparable to the one we used for European countries in previous sections. Table 4 gives the genetic distances between the six Yugoslav republics. With the exception of Montenegro, these distances are based on a reasonable number of genes. In any case, as we argue later, the exact value of these different distances does not matter as much as the average order of magnitude.[31] Abstracting from Montenegro, the average distance between two republics is 33, similar to the distance between, say, France and the Netherlands, but larger than the distances between, for example, Belgium and the Netherlands or Germany and Austria.

	Croatia	**Bosnia**	**Serbia**	**Slovenia**	**Macedonia**	**Montenegro**
Croatia	0	25.12	35.26	26.68	67.38	23.12
Bosnia	25.12	0	22.08	9.15	18.50	0.00
Serbia	35.26	22.08	0	15.36	82.92	8.50
Slovenia	26.68	9.15	15.36	0	33.12	14.44
Macedonia	67.39	18.50	82.92	33.12	0	0.21
Montenegro	23.13	0.00	8.50	14.44	0.21	0

Table 4: Genetic distances between Yugoslav republics

For GDP per capita and population we consider two different dates, one before the breakup, 1984, and one after the breakup, 2004. The 1984 data are the last ones published by the Statistical Yearbook of Yugoslavia *(Statisticki Godišnjak Jugoslavije)* and the 2004 data are some of the first published for all ex-Yugoslav republics by Eurostat. The output data for 1984 are based on the alternative socialist concept of *gross social product*, which does not include so-called nonproductive services, such as education, public administration and defense. Although we are only interested in *relative* output per capita, using gross social product may introduce some error if the weight of these nonproductive services differ across republics. To address this issue, it is desirable to complement the social product measure with a more standard output measure. In our benchmark exercise we will therefore take the average of the output per capita and population data of 1984 and 2004.[32] To make our data comparable to the ones of Europe, GDP per capita is expressed relative to the EU-25. As can be seen in Table 5, differences across republics were substantial, especially in the case of Slovenia, which had a GDP per capita 165 percent higher than the Yugoslav average.

[31] To be precise, leaving out Montenegro, the average number of genes used in computing the genetic distance between two republics is 12. Although this is substantially less than the average number of genes used by Cavalli-Sforza et al. (1994) to compute genetic distances between European countries, it is enough to give us an order of magnitude. In the case of Montenegro, the average number of genes used is only 3, so the order of magnitude of distances involving Montenegro may be off.

[32] Though not in the model, another reason for taking the average of 1984 and 2004 would be if agents were forward-looking and took into account how secession changes future output per capita.

Lastly, data on income distribution come, as in our exercise on Europe, from the World Income Inequality Database v.2.0a. For lack of data for individual republics, we take the year closest to 1990 for Yugoslavia, and use the same income distribution for all six republics.

	GDP per capita EU25=100	**Population ('000s)**
Croatia	47	4583
Bosnia	20	4316
Serbia	24	9650
Slovenia	88	1960
Macedonia	19	2111
Montenegro	23	642

Table 5: GDP per capita and population (averages 1984-2004)

5.2 The Dynamics of Breakup

We now use the calibrated parameter values of α, β and δ to see whether the theoretical model can account for the breakup of Yugoslavia. While we have assigned exact values to α and β, recall that there is a range $\delta \in [0.0284, 0.147]$ consistent with Europe being stable. High values within that range correspond to agents attaching little weight to cultural heterogeneity, whereas low values within that range imply agents care a lot about cultural differences. In light of what actually happened in Yugoslavia, we allow for unilateral secessions, and thus take ULS-stability to be the relevant stability concept. That is, it is enough for a majority in a given republic to be in favor of independence for the republic to break away. Furthermore, we assume that the set of all possible partitions excludes Serbia from unilaterally seceding. This reflects Serbia being the hegemon.[33]

We represent the incentives to secede by the monetary gains (equivalent variation) from leaving the union.[34] To do so, we determine the "income tax on secession", $t(I, C)$, that renders the median agent of the seceding republic, I, indifferent between breaking away or remaining part of the union of republics that have not yet seceded, C.[35] If, say, a 10% "secession tax" is needed to convince a republic to stay within the union, then the relative monetary gain from secession is

[33] This captures the idea that Serbia, by being the dominating center, derived enough of its income per capita from being the center, so that it would be unwilling to secede. Since our theoretical model does not allow for a region's income per capita to change with secession, we simply assume that Serbia never secedes unilaterally.

[34] The main results in this section go through if we represent the incentives to secede by welfare gains instead of by monetary gains.

[35] The income tax on secession (t) is the same as minus the income subsidy needed to keep regions within the union (what we called r in the case of the European Union).

approximately 10%. To determine $t(I,C)$ for each region I that is part of union C we solve for $t(I,C)$ in the following equation:

$$\begin{aligned} & y_m(I)(1-t(I,C))(1-\tau(I)) + \alpha(\tau(I)Y(I)(1-t(I,C)))^{\beta} \\ = \quad & y_m(I)(1-\tau(C)) + \alpha(Z_{(I,C)}\tau(C)Y(C))^{\beta} \end{aligned} \tag{19}$$

where the optimal tax rate of the median income agent of region I, $\tau(I)$, is calculated for the after "secession tax" income levels.

	Monetary gains from breakup *Low weight on cultural heterogeneity*		
	Round 1	**Round 2**	**Round 3**
Slovenia	0.02		
Croatia	-0.012	0.011	
Bosnia	-0.242	-0.196	-0.120
Macedonia	-0.326	-0.278	-0.201
Montenegro	-0.463	-0.423	-0.364

Table 6: Relative monetary gain from breakup ($\delta = 0.147$)

Table 6 reports the relative monetary gains of unilateral secession, as measured by $t(I,C)$, experienced by each republic. A positive number indicates that the median agent of a given region is better off seceding than remaining within the union. By setting δ to its highest value within the calibrated range $[0.0284, 0.147]$, we are assuming that little weight is attached to cultural heterogeneity. The first column (Round 1) shows that when all of Yugoslavia is united, only Slovenia has an incentive to secede. However, once Slovenia quits the union, the second column (Round 2) shows that Croatia follows suit, as it no longer gains from staying within the now smaller union, composed of Yugoslavia minus Slovenia. Once Slovenia and Croatia have left, no other republic has an incentive to secede, as reflected by the negative numbers in the third column (Round 3). However, if agents give more weight to cultural distances, the disintegration does not stop there. This can be seen in Table 7, which does the same exercise, but now sets δ to its lowest possible value within the calibrated range $[0.0284, 0.147]$. As expected, this increases the incentive to secede. The order of secession is now: Croatia, Slovenia, Bosnia and Macedonia. Compared to the first exercise reported in Table 8, the first two countries are the same, but because cultural differences now weigh more, Bosnia and Macedonia also break away. Once those four republics have seceded, the model predicts a ULS-stable union between Serbia and Montenegro.

	Monetary gains from breakup *High weight on cultural heterogeneity*				
	Round 1	**Round 2**	**Round 3**	**Round 4**	**Round 5**
Croatia	0.085				
Slovenia	0.063	0.079			
Bosnia	0.013	0.059	0.117		
Macedonia	-0.056	-0.009	0.049	0.10	
Montenegro	-0.244	-0.211	-0.164	-0.121	-0.096

Table 7: Relative monetary gain from breakup ($\delta = 0.0284$)

No matter which δ we use within the calibrated range $[0.0284, 0.147]$, Slovenia and Croatia always secede. Recall that after the dissolution of the all-Yugoslav Communist Party in 1990 these were the only two republics pushing for more devolution. When enough weight is attached to cultural heterogeneity, Bosnia and Macedonia also break away. The only republic that never wants to leave is Montenegro. These predictions are broadly consistent with historical events. The four republics that possibly secede according to the model — Slovenia, Croatia, Bosnia and Macedonia — all did so between 1991 and 1992. Although Montenegro eventually also declared independence from Serbia, this happened 15 years later, in 2006, after a relatively small majority voted in favor of self-determination. The model also does well in matching the order of disintegration, with the exception that it predicts Bosnia seceding before Macedonia, whereas in reality it happened the other way around. This discrepancy may be related to the existence of Serbian enclaves in Bosnia, which made its independence more difficult than that of Macedonia.[36]

To analyze the robustness of the model's predictions, we carry out a number of additional experiments. In the benchmark experiment of Tables 6 and 7 the measure of GDP per capita is the average of 1984 and 2004. Although we focused on this average to avoid possible issues with the *gross social product* measure of 1984, it may nonetheless be informative to see how our results change if we redo our exercise with the 1984 data. When doing so, we find that if we attach a high weight to cultural heterogeneity ($\delta = 0.0284$), the results are virtually unchanged, compared to the benchmark: all republics, with the exception of Montenegro, secede. The only difference is that Croatia and Bosnia now break away before Slovenia**.** As we will see throughout the rest of this section, the order of Slovenia and Croatia often gets reversed. This is not surprising, given that both republics declared independence on the very same day, June 25, 1991. Continuing with the

[36] In fact, in 2009 tensions have resurfaced in Bosnia, with Bosnian Serb leaders threatening to secede from the fragile Bosnian federation.

data of 1984, if we attach a low weight to cultural heterogeneity ($\delta = 0.147$), no republic wants to unilaterally secede, so that Yugoslavia remains united. The weaker incentives for the rich republics to secede reflect the smaller differences between poor and rich republics in the 1984 data, compared to the 2004 data. If we use the data for 2004, the results are the same as in the benchmark case of Tables 6 and 7: we get complete disintegration (with the exception of Montenegro) if cultural heterogeneity gets a high weight, whereas only Slovenia and Croatia break away if cultural heterogeneity gets a low weight.

The overall picture that emerges from these experiments is that the disintegration of Yugoslavia was a close call. Whether all of Yugoslavia was bound to break up, whether only Slovenia and Croatia were likely to leave, or even whether there was a possibility of the union staying together, essentially depends on how much weight is attached to cultural distances. This is a reasonable outcome: if under all possible parameter values the model were to predict the break-up of Yugoslavia, then one could wonder why it did not happen earlier, or how the country came into being in the first place.[37]

Although the breakup of Yugoslavia may have been a close call, our results clearly indicate that Yugoslavia was more unstable than the rest of Europe. Indeed, for the range of parameter values for which the model predicts the disintegration of Yugoslavia, the rest of Europe is stable. That is, compared to, say, the Basque Country and Scotland, two of Europe's regions where the demands for self-determination are greatest, Slovenia and Croatia had a greater incentive to secede. If any nation in Europe was unstable, it was Yugoslavia.[38] This finding holds up when we apply the same stability ULS-stability concept to both Yugoslavia and the rest of Europe. In other words, if we allow for unilateral secession, Slovenia and Croatia are more keen on breaking away than either the Basque Country or Scotland.

Another relevant question is what was the role of culture and what was the role of economics in Yugoslavia's breakup. After all, it has often been argued that cultural distances between the different Yugoslav republics were not particularly great. Our matrix of genetic distances between

[37] It is important to remember that Yugoslavia was not an accident of history. The country's foundation was laid by the Corfu Declaration in 1917 signed by Serbs, Croats and Slovenes. Its preamble stated that Serbs, Croats and Slovenes were "the same by blood, by language, by the feelings of their unity, by the continuity and integrity of the territory..." Consistent with what we said before, cultural differences were not that huge, and depending on the weight given to them, the model predicts Yugoslavia surviving as a union or breaking apart.

[38] This statement is limited to the countries and regions analyzed in Section 4, which do not include, amongst others, the Czech Republic and Slovakia.

the different republics largely subscribes to this view: the average distance of 33 is similar to the distance between Scotland and England or between Belgium and France. However, the fact that these cultural distances are relatively small does not imply that they do not matter. To see this, we run a counterfactual experiment by assuming that Yugoslavia was culturally completely homogeneous. In that case we find that all republics want to remain within the union. Cultural differences, though small, were essential for the country's disintegration. Economic differences, though large, were not enough to cause the breakup of Yugoslavia.

However, the order of disintegration as predicted by the model is driven by economics. The first two republics to leave, Slovenia and Croatia, were, respectively, the richest republic and the largest republic, whereas the republic that never wants to leave according to the model, Montenegro, is very small, with a population of just over half a million. Although some republics were culturally more distant than others, this does not matter in determining the order of breakup. If we re-do our experiments setting all distances between all republics equal to the average distance of 33, all results go through. In fact, if we set a slightly higher number of 50 or a slightly lower number of 25, the order of incentives for secession remains largely unchanged. Given that some of the distances in our matrix were based on a limited number of genes, and are thus less reliable, this is reassuring because it suggests that the results do not depend on the exact distances, but rather on the average order of magnitude.

As a final test for the robustness of our results, we ask whether a republic's decision to secede, when it does, is a dominant strategy. To give an example, consider 'round 3' in Table 9, which predicts that Bosnia leaves the union of Serbia, Macedonia, Bosnia and Montenegro. To see whether Bosnia's decision is a dominant strategy, we ask whether Bosnia would still want to secede from any possible partition of Serbia, Macedonia, Bosnia and Montenegro.[39] The answer is 'yes'. More generally, it turns out that only allowing for secession when it is a dominant strategy does not change any of the results.

6 Concluding Remarks

This paper has proposed a theory where larger countries benefit from scale economies in the provision of public goods but suffer from greater cultural heterogeneity. It has then empirically and

[39] As before, we assume that Serbia never secedes.

quantitatively explored the theory by calibrating the model to Europe. It has identified the 'weak links' in the current European map by determining which regions are most likely to secede and which countries are most likely to unite. As a first application, we have computed the predicted gains (and losses) from European Union membership. As a second application, we have analyzed whether the model can account for the breakup of Yugoslavia. The model not only shows that Yugoslavia was more unstable than the other European countries, it also generates an order of disintegration that closely matches the historical record of the country's breakup.

This paper is the first to quantitatively analyze the stability of nations by exploring the tradeoff between cultural heterogeneity and scale economies. Though the model is simple and stylized, the possibility of making quantitative predictions within a structural model opens up a set of new possibilities. To give just one example: in recent years there has been much interest in the stability of Iraq. While understanding this issue obviously requires in-depth knowledge of Iraq that goes much beyond the simplicity of this framework, being able to make quantitative predictions based on a calibrated model should be a useful input into this debate.

To make such quantitative models richer, we see at least three main areas that would benefit from 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, large coalitions, such as the present-day European Union, started off much smaller. Since there is likely to be path-dependence in coalition formation, better understanding these dynamics is important. Third, map redrawing may have effects that go beyond tax rates and public goods. For example, central (capital) regions may not want to see peripheral regions go, because much of their wealth stems from being the center.

A Proofs

Proof of Proposition 1: 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 $\tau(C)$. For every region $I \in N$ and partition $\pi \in \Pi$ denote by $C^I(\pi)$ the country in π that contains I. For every partition $\pi \in \Pi$ denote

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

We say that 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 π^* is $LRMR$-stable. If not, then there is a partition π' such that the median agent in every region affected by the map redrawing from π^* to π', would be strictly better off at π'. Since in regions that are not affected by such map redrawing there is no change in utility, we have $R(\pi') > R(\pi^*)$, a contradiction to the median-efficiency of π^*. **Q.E.D.**

B Additional Tables

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

	Bas	Sar	Aut	Cze	Fra	Ger	Pol	Rus	Swi	Bel	Den	Net	Eng	Ice	Ire	Nor	Sco	Swe	Gre	Ita	Por	Spa	Fin	Hun
Basque	0																							
Sardinia	261	0																						
Austria	195	294	0																					
Czech	159	327	36	0																				
France	93	283	38	72	0																			
Germany	169	331	19	52	27	0																		
Poland	146	282	72	64	66	47	0																	
Russia	140	266	64	75	59	60	30	0																
Switzerland	165	353	12	31	23	10	60	78	0															
Belgium	107	256	16	43	32	15	40	51	14	0														
Denmark	184	348	27	54	43	16	69	80	19	21	0													
Netherlands	118	307	38	66	32	16	54	57	16	12	9	0												
England	119	340	55	60	24	22	70	79	28	15	21	17	0											
Iceland	221	396	153	173	146	106	144	169	115	78	88	101	76	0										
Ireland	145	393	115	117	93	84	150	160	86	75	68	76	30	99	0									
Norway	195	424	61	76	56	21	58	90	33	24	19	21	25	74	79	0								
Scotland	146	357	74	104	62	53	121	129	59	59	40	48	27	111	29	58	0							
Sweden	168	371	80	90	78	39	82	110	55	34	36	41	37	106	94	18	74	0						
Greece	231	190	86	126	131	144	177	161	148	103	191	199	204	288	289	235	253	230	0					
Italy	141	221	43	77	34	38	64	75	44	30	72	64	51	143	132	88	112	95	77	0				
Portugal	145	340	48	46	48	51	65	98	53	31	77	60	46	149	115	73	97	78	103	44	0			
Spain	104	295	69	65	39	69	117	122	43	42	80	76	47	163	113	97	100	99	162	61	48	0		
Finland	236	334	77	175	107	77	139	153	112	63	96	123	115	157	223	94	166	82	150	94	119	159	0	
Hungary	153	279	40	69	70	46	25	30	57	52	78	71	70	172	152	77	124	99	88	61	63	118	115	0

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

	Aut	Fra	Ger	Bel	Den	Net	Eng	Ire	Swi	Gre	Ita	Por	Spa	Fin
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., Devleeschauwer, A., Easterly, W., Kurlat, S. and R. Wacziarg (2003), "Fractionalization", *Journal of Economic Growth*, 8, 155-194.

[4] Alesina, A., Easterly, W. and J. Matuszeski (2010), "Artificial States," *Journal of the European Economic Association*, forthcoming.

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

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

[7] Alesina, A., Spolaore, E. and R. Wacziarg (2000), "Economic Integration and Political Disintegration," *American Economic Review* 90, 1276-1296.

[8] Alesina, A. and R. Wacziarg (1998), "Openness, Country Size and Government," *Journal of Public Economics* 69, 305-321.

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

[10] Ashraf, Q. and O. Galor, (2008), "Human Genetic Diversity and Comparative Economic Development," mimeo.

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

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

[13] Belle, E.M.S. and G. Barbujani, (2007), "Worldwide Analysis of Multiple Microsatellites: Language Diversity Has a Detectable Inßuence on DNA Diversity", *American Journal of Physical Anthropology*, 133, 1137-1146.

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

[15] 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.

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

[17] 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.

[18] Cavalli-Sforza, L.L., Piazza, A, Menozzi, P. and J. Mountain (1988), "Reconstruction of Human Evolution: Bringing Together Genetic, Archaeological and Linguistic Data", *Proceedings of the National Academy of Sciences*, 85, 6002- 6006.

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

[20] 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, 11008-11013.

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

[22] Demange, G. (1994), "Intermediate preferences and stable coalition structures," *Journal of Mathematical Economics*, 23, 45-58.

[23] Desmet, K., Le Breton, M., Ortuño-Ortín, I. and S. Weber (2009), "The Stability and Breakup of Nations: A Quantitative Analysis," VIVES Discussion Paper No. 10, KU Leuven.

[24] Desmet, K., Ortuño-Ortín, I. and S. Weber (2009), "Linguistic Diversity and Redistribution," *Journal of the European Economic Association*, 7, 1291-1318.

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

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

[27] Fearon, J.D. (2003). Diversity by Country, *Journal of Economic Growth*, 8, 195-222.

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

[29] 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.

[30] Ginsburgh, V. and S. Weber (2011), *How Many Languages Do We Need? Economics of Linguistic Diversity*, Princeton University Press, forthcoming.

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

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

[33] Greenberg, J. and S. Weber (1986), "Strong Tiebout Equilibrium under Restricted Preferences Domain," *Journal of Economic Theory*, 38, 101-117.

[34] Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D., Werner Alt, K. and J. Burger (2005), "Ancient DNA from the first European farmers in 7500-year-old Neolithic sites", *Science*, 310, 1016-1018.

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

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

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

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

[39] La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and R. Vishny (1999), "The Quality of Government," *Journal of Law, Economics and Organization*, 15, 222-79.

[40] 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.

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

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

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

[44] Sambanis, N. (2006), "Globalization, Decentralization, and Secession: A Review of the Literature and Some Conjectures." in: G. Ranis, D. Cameron and A. Zinn, eds., *Globalization and Self-Determination: Is the Nation-State under Siege?*, New York: Routledge.

[45] 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.

[46] 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.

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

[48] 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.

[49] Sorens, J. (2005), "The Cross-Sectional Determinants of Secessionism in Advanced Democracies," *Comparative Political Studies*, 38, 304-326.

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

[51] Spolaore, E. and R. Wacziarg (2009), "The Diffusion of Development," *Quarterly Journal of Economics*, forthcoming.

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