

Linguistic fractionalization and conflict in the Basque Country

Ana Carmen Díaz and Javier Gardeazabal
University of the Basque Country

March 16, 2011

Abstract

This paper investigates the relationship between linguistic fractionalization and conflict. We conduct what we think is an interesting case study for the Basque Country. The Basque Country has experienced a long period of nationalist conflict during which terrorist organization Euskadi Ta Askatasuna (ETA) has produced more than eight hundred assassinations. During the forty years of Franco's dictatorship, most of which were before ETA committed its first assassination, the use of the Basque language was banned. Therefore, linguistic fractionalization in the Basque Country could be a source conflict. We report evidence showing that, on the contrary, linguistic fractionalization actually reduces the level of terrorism. We also find that political fractionalization increases the level of terrorism. These results are robust to the inclusion of economic, demographic, educational and well-being indicators. Further robustness checks using different measures of linguistic and political fractionalization leave the results unchanged.

We would like to thank Esteban Klor for his helpful comments and Manuel Bagues and Asier Miondo for providing us with electoral and geographical data. Financial support from the 7th Framework Programme of the European Commission through EUSECON project is gratefully acknowledged. Mailing address: Departamento de Fundamentos del Análisis Económico II, Universidad del País Vasco, Lehendakari Aguirre 83, 48015 Bilbao, Spain. E-mail: javier.gardeazabal@ehu.es

1 Introduction

In his “The clash of Civilizations?” Huntington (1993) advanced the hypothesis that the fundamental sources of conflict will not be economic or ideological but factors such as language, tradition, culture and religion. Differences in these factors, the argument goes, are more fundamental than economic or ideological differences. The rich can become poor and vice versa, he argues, but Russians cannot become Estonians.

This paper focuses on one of these factors, linguistic diversity, and its relationship with conflict. The relationship between linguistic diversity and conflict is important because conflict is thought to be detrimental for economic growth and, therefore, understanding the sources of conflict would help economic development. Ethno-linguistic diversity has been found to be detrimental for the quality of institutions and economic performance, e.g. Easterly and Levine (1997) and Alesina et al (2003), what can be interpreted as lending support to theories of interest group polarization and conflict.

Establishing a causal link between linguistic fragmentation and conflict is intricate for several reasons. A first difficulty is the fact that it is difficult to distinguish between linguistic and ethnic groups: ethnic groups are often defined by their language. Therefore, when researchers try to assess the effect of ethnic and linguistic fragmentation on conflict, there is no clear cut distinction between the contribution of ethnic and linguistic diversity to conflict. This is probably why many researchers focus on ethno-linguistic fragmentation, instead of considering both separately. A notable exception is Wimmer and Cederman (2009) who found that, once ethnic fractionalization is controlled for, linguistic fractionalization is not a significant determinant of armed conflict. To the extent that some ethnic groups are defined by their language, ethnic fractionalization might capture some linguistic fragmentation and, hence, may take some explanatory power away from linguistic fragmentation.

A second difficulty is the nature of conflict itself. Conflict can manifest in several forms, from strikes, demonstrations and riots to guerrilla warfare, terrorism and civil war. Conflict can be thought of as a latent variable of which researchers observe proxy measurements such as a binary indicator of participation in an armed conflict or the number of terrorist assassinations. Alternative conflict measurements result in different causal effects. Laitin (2000) reports evidence suggesting that language grievances are not associated with group violence while Abadie (2006) finds that linguistic fractionalization is a determinant of terrorism risk.

The third difficulty is measurement. Linguistic diversity manifests when different

social groups speak different languages, i.e. society is fragmented in different linguistic groups. Linguistic fragmentation in a country or region can be measured in different ways. The leading measure used in the literature is an index of fractionalization defined as the probability that two randomly chosen individuals belong to different linguistic groups. An alternative way of measuring linguistic fragmentation is the concept of polarization, defined as the distance between the frequency distribution of linguistic groups to the bipolar distribution, which yields the highest level of polarization. A country with many linguistic groups would have a high level of fractionalization and a low level of polarization. On the contrary, a country with two linguistic groups, each with half of the population, would have the highest polarization and a lower level of fractionalization. Montalvo and Reynal-Querol (2005) provide evidence suggesting that ethno-linguistic polarization increases the likelihood of internal armed conflict while ethno-linguistic fractionalization does not.

In this paper we conduct what we think is an interesting case study on the relationship between linguistic fractionalization and conflict using data from the Basque Country.¹ There are reasons to believe that conflict in the Basque Country could have some linguistic roots. Up until the middle of the nineteenth century Basques had a peaceful relationship with the Spanish Crown. After the Carlists wars, when the Spanish government revoked part of the Basque autonomy, Basques begun demanding higher levels of self-governance, but it was not until 1968 when terrorist organization *Euskadi Ta Askatasuna* (ETA) produced its first victim. After the Spanish civil war, during Franco's dictatorship (1936-1975), the use of the Basque language was either banned or repressed depending on the particular circumstances and timing. Therefore, it is reasonable to think that after a long period of repression, the linguistic issue in the Basque Country could be a source of conflict. In addition, Basque is the last remaining pre-Indo-European language in Western Europe and hence, very different from Spanish. This linguistic distance between the two main languages spoken in the Basque Country could potentially induce conflict if, as Desmet, Ortuño-Ortín and Wacziarg (2009) claim, fractionalization measures based on deep cleavages originating thousands of years ago are better predictors of conflict than more superficial ones originating only hundreds of years ago. Other bilingual regions of Spain, Catalonia and Galicia, constitute *prima facie* cases of this theory. Catalonia and Galicia, also with nationalistic demands, have not experience a terrorist conflict, and their languages, Catalan and

¹Hereinafter, we use the term Basque Country to refer to the Spanish *Comunidad Autónoma Vasca*, an autonomous region comprising the Spanish provinces of *Alava*, *Guipúzcoa* and *Vizcaya*.

Galician, share with Spanish a much closer ancestor: Latin.

Another interesting feature of our case study is the ethno-linguistic cut of the population under study. Ethno-linguistic diversity in the Basque Country has a strong linguistic basis. Although it has long been argued that Basque people were a genetic isolate and their genomes have been considered the most direct descendants of those who dwelt in Europe before the spread of agriculture, e.g. Dupanloup et al. (2005), recent research based on genome-wide analysis shows that Basques are not genetically different from other populations in Spain, e.g. Laayouni, Calafell and Bertranpetit (2010). Therefore, the difference between the two ethno-linguistic groups in the Basque Country is fundamentally linguistic.

Our identification strategy relies on the geographical and temporal variation in linguistic fragmentation. The Basque Country exhibits a fairly high geographical heterogeneity of Basque language use, providing a source of variability in linguistic fractionalization. Although linguistic fractionalization is bounded to change very little within a short period, as a result of its long-banned status, the population of Basque speakers has experienced a relatively high increase during the sample period analyzed, providing an additional source of variability difficult to observe in any other region or country.²

Our findings indicate that linguistic fractionalization in the Basque Country reduces the level of terrorism activity. Our interpretation of this result is that higher levels of linguistic fractionalization opens a channel through which Basque speakers can manifest their nationalism by using their language regularly. Linguistic fractionalization serves as a escape valve for conflict, reducing the probability of engaging in terrorist activities. When fractionalization is low, however, nationalism cannot be manifested through the use of the Basque language, either because most people speak Basque or only a few. Our empirical evidence also shows that political fractionalization is associated with higher levels of terrorist activity. Therefore, contrary to Huntington's hypothesis, our findings point to political rather than ethno-linguistic roots of conflict.

2 The data set

The data set is a panel of all 251 municipalities of the Basque Country over the 1981-2008 period with annual frequency. Our outcome variable is terrorism activity measured

²Campos and Kuzeyev (2007, forthcoming) find that 26 transition economies experimented a relatively important change in ethnic fractionalization over a two decades time period. However, linguistic fractionalization remained fairly constant over that period.

by the annual number of assassinations in each municipality, these data are from *Asociación de Víctimas del Terrorismo*. Of course, the number of terrorist assassinations is only a proxy measurement of the latent conflict. There are reasons to believe that the number of terrorist assassinations in a municipality may actually be a good measurement of the latent level of conflict in that municipality. It is well known that ETA targets its victims within security forces, politicians and entrepreneurs through its network of militants who exert physical and psychological violence on its victims, see Martín-Peña et al. (2010). This violence would range from coercion and extortion at the lower end to kidnappings and murder at the higher end. ETA’s network informs actual perpetrators of the location and living practices of targeted victims. So the observation of an assassination can be taken as an indication of the existence of a terrorist network within the area. Under these circumstances, the geographical dissemination of assassinations indicates conflict intensity.

Political fractionalization is measured using vote shares in Spanish general elections computed from the electoral data from the Spanish *Ministerio del Interior*. All the other municipal data comes from the EUSTAT. The frequency of observation differs depending on the variable. We have constructed an annual data set by interpolating several variables. Detailed discussion of data sources and transformations is included in an appendix. All fractionalization measures are computed as one minus the Herfindahl-Hirschman index. Polarization indexes are computed as in Reynal-Querol (2002). Linguistic fractionalization and polarization are measured in three different ways based on mother’s tongue, language spoken at home and Basque knowledge. These different measures allow us to carry out robustness checks that can only seldom be performed. Table 1 reports summary statistics on all variables.

As commented previously, this data set exhibits a high variability of linguistic fractionalization across municipalities. Both linguistic fractionalization and polarization are bounded to lay in the unit interval. Table 1 shows that linguistic fractionalization ranges from 0 to 0.6655 according to *linguistic fractionalization 1* (basque knowledge) and this variability is fairly uniform across measures of linguistic fractionalization. Variability is even higher for the measures of linguistic polarization: it ranges from 0 to 0.9598 according to *linguistic polarization 1*. Therefore, we will make use of this source of variability to identify the potencial effect of linguistic fractionalization/polarization on terrorism assassinations.

3 Econometric methodology

The outcome variable is the annual number of assassinations in each municipality, a count variable. Therefore, we use count regression estimation methodology. Table 1 shows that the mean number of annual assassinations is much lower than its variance. Therefore, the Poisson model, which assumes mean and variance are equal, cannot be used. Instead, we fit panel data negative binomial regressions. Let y_{it} denote the number of assassinations in municipality i in year t and X_{it} a $1 \times K$ vector of covariates. According to the negative binomial distribution, the probability that we observe $y_{it} = 0, 1, 2, 3, \dots$ assassinations in a given municipality and year is assumed to be

$$P(y_{it} | X_{it}, \tau_i) = \frac{\Gamma(y_{it} + \lambda_{it})}{\Gamma(y_{it} + 1)\Gamma(\lambda_{it})} \left(\frac{\tau_i}{1 + \tau_i}\right)^{y_{it}} \left(\frac{1}{1 + \tau_i}\right)^{\lambda_{it}}$$

where $\ln \lambda_{it} = X_{it}\beta$, β is a vector of parameters. The conditional mean and variance of the number of assassinations are $E(y_{it} | X_{it}, \tau_i) = \tau_i \lambda_{it} = \tau_i \exp(X_{it}\beta)$ and $V(y_{it} | X_{it}, \tau_i) = (1 + \tau_i)\tau_i \lambda_{it} = (1 + \tau_i)\tau_i \exp(X_{it}\beta)$, so the dispersion, the ratio of the variance to the mean is $(1 + \tau_i)$. The dispersion parameter, τ_i , can vary across municipalities but it is constant over time. Most of the evidence reported below assumes the dispersion parameter is fixed, although for robustness checks we also report evidence under the assumption that the dispersion parameters are random.

4 Empirical results

Columns (1) and (2) in Table 2 show the negative binomial regression of the number of terrorist assassinations on linguistic and political fractionalization respectively. Linguistic fractionalization measured by Basque knowledge has a negative and significant effect on the number of terrorist assassinations while political fractionalization has a positive and significant effect on the number of terrorist assassinations. The sign and significance of these fractionalization indexes remains the same when both are included together in column (3). Columns (4) to (10) report count regressions when educational and service availability variables are included as controls. Only the number of service centers per capita is significant in these regressions, with the coefficients on linguistic and political fractionalization retaining their sign, magnitude and significance. Table 3 shows in columns (1) to (10) other count regressions where several demographic and economic variables are included as control variables. While none of these additional con-

trols are significant, the same results obtain, linguistic fractionalization lowers terrorist conflict while political fractionalization increases it.

Next we perform a series of robustness checks using alternative measures of linguistic and political fractionalization. In particular, our data set includes linguistic classifications based on mother's tongue and language used at home. Political fractionalization can be computed using a nationalistic versus non-nationalistic classification and right versus left wing classification. Table (4) reports in column (1) to (9) the results of running negative binomial regressions for the nine possible combinations of linguistic and political fractionalization measures. Linguistic fractionalization remains significant and similar in magnitude when linguistic fractionalization is computed based on either the mother tongue or language spoken at home classifications. The coefficient estimate on political fractionalization retains its sign but not its significance when political fractionalization is measured using a nationalistic versus non-nationalistic classification or a right versus left wing classification.

The finding that linguistic fractionalization reduces terrorist conflict could be due to the specific measure of linguistic diversity used. García-Montalvo and Reynal-Querol (2005) showed that it is not ethno-linguistic fractionalization but ethno-linguistic polarization what helps explaining a country participation in armed conflict. A country like Chad, with 131 living languages must necessarily have a high fractionalization, but a low level of polarization which would be maximal in a community with two languages with half of the population each. As the Basque Country fits more into the case of a polarized linguistic community, linguistic polarization could be a source of conflict in the Basque Country. As a further set of robustness tests, Table 5 reports the same regressions as Table 4 where linguistic and political fractionalization are substituted by Reynal-Querol's (2002) measure of polarization. The coefficients on linguistic polarization are somewhat smaller (in absolute value) than in the previous table, but the same overall result obtains: linguistic polarization reduces terrorist conflict while political polarization increases it although not significantly.

Conflict can be thought of as a continuous latent variable of which we only observe a proxy, the number of terrorist assassinations. Under this interpretation, an observation of a municipality in the Basque Country with no assassinations in a given year does not mean there was no conflict in that municipality. It is reasonable to believe that observations with no assassinations are simply a realization of a stochastic process that could have yielded a different outcome. Under the latent variable view, conflict is smoother than the actual number of assassinations. If we could observe the continuous

latent variable, inference would be easier, as the dependent variable would not be a count variable. We have computed a smooth kernel estimate of the number of assassinations in each municipality. The kernel estimate of the number of assassinations is a weighted average of the number of assassinations in the neighbor municipalities within a given radius. A detailed explanation of the kernel estimate is given in the appendix. We have regressed this smoother measure of conflict on the same sets of right hand side variables and reported the results in Table 6. Since the dependent variable in this case is a non-negative continuous variable, we use least squares to fit the linear regression with the (log) of the smooth measure of conflict as dependent variable.

A caveat worth exploring is the possibility of endogeneity of political fractionalization arising from reverse causality from terrorism to political fractionalization. In the right hand side of the estimating equation the index of political fractionalization was computed using vote shares from the previous election, so it can be considered as a lagged right hand side variable. However, vote shares, and therefore political fractionalization, is known to exhibit a high level of persistence, e.g. Dolado et al. Table 7 reports the results of fitting instrumental variable regressions.

The results presented in this paper point towards causality from linguistic fractionalization to terrorist conflict. However, there is an interpretation of the results which would invalidate this causality link. If terrorist organization ETA targets its victims in locations where linguistic fractionalization is low, causality would run backwards and our previous findings would be invalidated.

5 Conclusions

This paper provides evidence on the relationship between linguistic fragmentation and conflict, using data from the Basque Country. During the forty years of Franco's dictatorship the Basque language was banned, therefore the terrorist conflict in the Basque Country could have some linguistic roots. Our findings indicate that, on the contrary, linguistic fractionalization reduces the likelihood of terrorist assassinations. We find political fractionalization to be a good predictor of terrorist assassinations. These findings are robust with respect to the measure of linguistic and political fragmentation used. In particular, the results are qualitatively the same when we use measures of linguistic fractionalization based on different classification criteria of linguistic groups: mother's tongue, language use at home and language knowledge. The results also obtain when

political fractionalization is measured using party vote shares, right-left classification or nationalistic versus non-nationalistic party affiliation. The results are also robust to the fragmentation measure used: fractionalization or polarization.

Our reading of these results is that linguistic fractionalization may help reducing the level of conflict because it provides a way for people to show their nationalism without having to engage in terrorist activities. A low linguistic fractionalization, however, does not allow people to show their nationalism, either because most people or only a few speak Basque.

Our findings can be understood as proof of a causal link from linguistic fractionalization and terrorist activity. However, they can also be consistent with terrorists targeting municipalities with low levels of linguistic fractionalization.

Appendix: variable definition

Unless otherwise noted, data were retrieved from the Basque Statistical Institute EUSTAT. These are census data.

- **Terrorist assassinations.** Annual number of terrorist assassinations for each municipality. These data are from *Asociación de Víctimas del Terrorismo*.
- **Basque knowledge.** There are three knowledge of Basque levels. *Euskaldunes* those people whose mother tongue is Basque and who understand and speak it well. *Quasi-euskaldunes* those whose mother tongue is Spanish or another language except Basque and who understand and speak Basque well and whose everyday spoken language at home is Basque. *Erdaldunes* are those whose mother tongue is Spanish or another language except Basque and who do not speak Basque at all and neither understand nor read it well. Raw data measured in number of people in each category for each municipality transformed into proportions of population in each category.
- **Mother tongue.** The first language learned in infancy up until the age of 3, distinguishing between children who were exposed only to Basque in their early infancy, those who only learned Spanish, those who were exposed to both Basque and Spanish almost simultaneously and finally, those children whose mother tongue was neither Basque nor Spanish. Raw data measured in number of people

in each category for each municipality transformed into proportions of population in each category.

- **Language spoken at home.** The language most frequently spoken within the household. The same categories as used for Mother Tongue are admitted: Basque, Spanish, both or other. These are census data. Raw data measured in number of people in each category for each municipality transformed into proportions of population in each category.
- **Votes.** Number of valid votes for each party: Partido Popular (PP), Partido Socialista Obrero Español (PSOE), Izquierda Unida (IU), Partido Nacionalista Vasco (PNV), Herri Batasuna (HB), Eusko Alkartasuna (EA) and Aralar, These data are from the Spanish Ministry of Interior.
- **Infant education.** Number of children from birth to the age of six attending infant education measured as a fraction of population.
- **School attendance.** Number of students attending private schools, public schools or ikastola. Ikastola is a type of primary and/or secondary school in which students are taught either entirely or predominantly in the Basque measured as a fraction of population.
- **Private Schools attendance.** Number of students attending private schools measured as a fraction of population.
- **Public Schools attendance.** Number of students attending public schools measured as a fraction of population.
- **Ikastola attendance.** Number of students attending ikastolas measured as a fraction of population.
- **Social services centers.** Number of centers from which social services are rendered in per capita terms.
- **Public outpatients centers.** Number of public outpatients' centers in per capita terms.
- **Population.** Inhabitants of the municipality.

- **Population density.** The number of inhabitants of the municipality divided by the surface area measured in squared kilometers.
- **Emigration.** Number of people who migrated from a municipality of the Basque Country to a municipality outside the Basque Country. Measured as a fraction of population.
- **People born in other provinces.** Measured as a fraction of population.
- **People born in a foreign country.** Measured as a fraction of population.
- **People born in the Basque Country.** Measured as a fraction of population.
- **Unemployment rate.** Proportion of active population not working on the reference date.
- **Land used in agriculture.** Total used arable land and permanent grazing pastures measured in hectares.
- **Financial institutions:** Number of banks and saving banks as a fraction of population.

Variable transformations

- Interpolation. In order to have a balanced panel data set, several series were interpolated. The procedure used for interpolation is as follows. Suppose we observe variable x_t and x_{t+s} with $s \geq 2$, $x_{t+j} = x_t + j \frac{x_{t+s} - x_t}{s}$ for $j = 1, \dots, s$. All variables interpolated are stocks. The following table indicates the variables interpolated, the dates available and interpolated

Variable	Dates available	Sample period after interpolation
Terrorist assassinations	1981-2008	[1981-2008]
Euskaldunes	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Quasieuskaldunes	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Erdaldunes	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Linguistic fractionalization 1	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Linguistic polarization1	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Mother tongue Basque	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]

Mother tongue Spanish	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Mother tongue both	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Linguistic fractionalization 2	1981, 1986, 1991, 1996, 2001, 2007	[1981-2006]
Linguistic polarization2	1981, 1986, 1991, 1996, 2001, 2008	[1981-2006]
Language spoken at home Basque	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Language spoken at home Spanish	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Language spoken at home both	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Linguistic fractionalization 3	1981, 1986, 1991, 1996, 2001, 2007	[1981-2006]
Linguistic polarization3	1981, 1986, 1991, 1996, 2001, 2008	[1981-2006]
Vote share psoe	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Vote share pp	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Vote share iu	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Vote share pnv	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Vote share hb	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Vote share ea	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Vote share ucd	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Vote share aralar	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
All parties	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Political fractionalization1	1989, 1993, 1996, 2000, 2004, 2009	[1989-2008]
Political polarization1	1989, 1993, 1996, 2000, 2004, 2010	[1989-2008]
Nationalist	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Non Nationalist	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Political fractionalization 2	1989, 1993, 1996, 2000, 2004, 2009	[1989-2008]
Political polarization2	1989, 1993, 1996, 2000, 2004, 2010	[1989-2008]
Left wing	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Right wing	1989, 1993, 1996, 2000, 2004, 2008	[1989-2008]
Political fractionalization 3	1989, 1993, 1996, 2000, 2004, 2009	[1989-2008]
Political polarization3	1989, 1993, 1996, 2000, 2004, 2010	[1989-2008]
Infant education	1983-2006	[1983-2006]
School attendance	1983-2005, 2010	[1983-2010]
Private school attendance	1983-2005, 2010	[1983-2010]
Public school attendance	1983-2005, 2010	[1983-2010]
Ikastola attendance	1983-2005, 2010	[1983-2010]
Social Service centers	1984-2006	[1984-2006]

Public outpatients centers	1988-2006	[1988-2006]
Population in thousands	1981, 1986, 1991, 1996, 2001, 2006	[1981-2006]
Population density	1981, 1985, 1986, 1991, 1993, 1996, 2000, 2004, 2006	[1981-2006]
Emigration	1988- 2007	[1988-2007]
People born in other provinces	1981, 1986, 1991, 1996, 2001	[1981-2001]
People born in the Basque Country	1981, 1986, 1991, 1996, 2001	[1981-2001]
People born in a foreign country	1981, 1986, 1991, 1996, 2001	[1981-2001]
Unemployment rate	1986, 1991, 1996, 2001	[1986-2001]
Land used in agriculture	1989, 2000	[1989-2000]
Financial institutions	1984,1985,1987,1989,1993, 1996, 2001, 2004, 2007	[1984-2007]

- The index of fractionalization computed as $FRAC = 1 - \sum_{i=1}^N \pi_i^2$ where π_i is the proportion of people who belong to the linguistic (political) group i , and N is the number of groups. The index of fractionalization has a simple interpretation as the probability that two randomly selected individuals will not belong to the same linguistic (political) group. We measure the linguistic fractionalization in three different ways: (i) *linguistic fractionalization 1* uses the Basque knowledge categories, (ii) *linguistic fractionalization 2* uses the mother tongue categories and (iii) *linguistic fractionalization 3* uses the language spoken at home categories.
- We measure political fractionalization in three different ways: (i) *political fractionalization 1* uses the vote shares of all parties, (ii) *political fractionalization 2* classifies parties in two groups, nationalist parties (Partido Nacionalista Vasco, Herri Batasuna, Eusko Alkartasuna and Aralar) and non-nationalist parties (Partido Popular, Partido Socialista and Izquierda Unida), (iii) *political fractionalization 3* classifies parties in two groups, left-wing (Partido Socialista, Izquierda Unida, Herri Batasuna, Eusko Alkartasuna and Aralar) and right wing (Partido Popular, Partido Nacionalista Vasco).
- The index of linguistic (political) polarization originally proposed by Reynal-Querol (2002) can be written is $RQ = 1 - \sum_{i=1}^N \left(\frac{0.5-\pi_i}{0.5}\right)^2 \pi_i$.
- The kernel estimate is constructed as follows. The raw data x_{it} are the number of assassinations in municipality i in period t . The spatially weighted index is defined as

$$x_{it}^* = \frac{3}{4} \sum_{j=1}^J \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right) I\left(\frac{d_{ij}}{h} \leq 1\right) x_{jt} \quad (1)$$

where J is the number of municipalities, d_{ij} is the distance between municipality i and j , h is the bandwidth parameter and $I(a) = 1$ if a is true and zero otherwise. The weighting function is the Epanechnikov kernel, $\int_0^1 \frac{3}{4} (1 - u^2) du = 1$. We tried different values of the bandwidth parameter h , which in our case is the radius of the circle defining neighbor municipalities. We used $h = 5, 10, 20$ Km.

References

- [1] Abadie, Alberto, 2006, "Poverty, Political Freedom, and the Roots of Terrorism," *American Economic Review Papers and Proceedings* 96, 50-56.
- [2] Alesina, Alberto, Arnaud Devleeschauwer, William Esasterly, Sergio Kurlat and Romain Wacziarg, 2003, "Fractionalization," *Journal of Economic Growth* 8, 155-194.
- [3] Campos, Nauro and Vitaliy Kuzev, 2007, "On the Dynamics of Ethnic Fractionalization," *Journal of Political Science* 51(3): 620-639.
- [4] Campos, Nauro, Ahmad Saleh and Vitaliy Kuzev, forthcoming, "Dynamic Ethnic Fractionalization and Economic Growth," *Journal of International Trade and Economic Development*
- [5] Isabelle Dupanloup, Giorgio Bertorelle, Lounès Chikhi and Guido Barbujani, 2004, "Estimating the Impact of Prehistoric Admixture on the Genome of Europeans" *Molecular Biology and Evolution* 21(7), 1361-1372.
- [6] Klaus Desmet, Ignacio Ortuño-Ortín and Romain Wacziarg, 2009. "The Political Economy of Ethnolinguistic Cleavages," NBER Working Papers 15360.
- [7] Easterly, William and Ross Levine, 1997, "Africa's Growth Tragedy: Policies and Ethnic divisions," *Quarterly Journal of Economics* 112, 1203-1250.
- [8] Huntington, Samuel P. 1993, "The Clash of Civilizations?" *Foreign Affairs* 72, 22-49.
- [9] Laayouni, Hafid, Calafell, Francesc and Bertranpetit, Jaume, 2010, "A genome-wide survey does not show the genetic distinctiveness of Basques," *Human Genetics* 127, 455-458.
- [10] Laitin, David D., 2000, "Language conflict and violence: the straw that strengthens the camel's back." *European Journal of Sociology*, 41, 97-137.
- [11] Montalvo, José G. and Marta Reynal-Querol, 2005, "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review* 95. 796-816.
- [12] Reynal-Querol, Marta, 2002, "Ethnicity, Political Systems, and Civil Wars," *Journal of Conflict Resolution* 46, 29-54.

- [13] Wimmer, Andreas, and Lars-Erik Cederman, 2009, "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set." *American Sociological Review* 74, 316-337.

Table 1: Summary statistics

Variables	Unit	Obs	Mean	Std. Dev.
Terrorist assassinations	Number of assassinations	7028	0.0418327	0.3110098
Euskaldunes	Persons	6475	2365.41	5557.005
Quasieuskaldunes	Persons	6475	1620.5	6751.366
Erdaldunes	Persons	6475	4268.152	19240.78
Linguistic fractionalization 1		6475	0.3594574	0.1959888
Linguistic polarization1		6475	0.5757381	0.2705667
Mother tongue Basque	Persons	5230	1691.765	3377.224
Mother tongue Spanish	Persons	5230	6232.514	26951.41
Mother tongue both	Persons	5230	338.2415	1220.496
Linguistic fractionalization 2		5230	0.2667106	0.1989455
Linguistic polarization2		5230	0.4737445	0.338223
Language spoken at home Basque	Persons	3985	1136.305	2048.858
Language spoken at home Spanish	Persons	3985	6496.123	27757.26
Language spoken at home both	Persons	3985	700.659	1887.587
Linguistic fractionalization 3		3985	0.3300926	0.2292531
Linguistic polarization3		3985	0.5205538	0.3278208
Votes psoe	Number of votes	7028	869.505	4173.238
Votes pp	Number of votes	7028	627.3042	3786.11
Votes iu	Number of votes	7028	217.74	1106.804
Votes pnv	Number of votes	7028	930.3216	3882.086
Votes hb	Number of votes	4016	344.2727	1366.588
Votes ea	Number of votes	7028	266.5041	971.4303
Votes ucd	Number of votes	6024	22.58068	232.885
Votes aralar	Number of votes	7028	24.231	170.2927
All parties	Number of votes	4312	5145.307	17845.01
Political fractionalization1		5020	0.7086126	0.1659941
Political polarization1		5020	0.7674948	0.1210017
Nationalist	Number of votes	7028	1417.784	5349.482
Non Nationalist	Number of votes	7028	1733.904	8974.95
Political fractionalization 2		4312	0.3987867	0.1393915
Political polarization2		4312	0.7639438	0.2438412
Left wing	Number of votes	7028	1574.707	6675.334

Right wing	Number of votes	7028	1576.981	7623.653	
Political fractionalization 3		4312	0.4479816	0.1045106	
Political polarization3		4312	0.8535458	0.1606412	
Infant education	% of population	4441	34.95502	38.53956	1.63
School attendance	% of population	6001	0.1308047	0.2672023	
Private school attendance	% of population	6001	0.0752189	0.1310761	
Public school attendance	% of population	6001	0.0443521	0.2036315	
Ikastola attendance	% of population	6001	0.0112336	0.0620744	
Social Service centers	% of population	4732	2.439932	2.65795	
Public outpatients centers	% of population	5479	0.8081776	1.135112	
Population in thousands	Thousands of people	6501	8447.399	30562.88	
Population density	Number of people by Km2	6318	523.9716	1647.476	
Emigration	% of population	4731	5.978708	6.357038	
People born in other provinces	% of population	5250	0.1537737	0.1114023	
People born in the BasqueCountry	% of population	5250	0.8374786	0.1124026	0.47
People born in a foreign country	% of population	5250	0.0087477	0.0082476	
Unemployment rate	% active population	3984	0.150115	0.060267	
Land used in agriculture	Hectareas	2658	43.17298	18.11427	7.22
Financial institutions	% of population	5123	1.053617	1.422785	

Table 2: Terrorist assassinations, linguistic and political fractionalization. Education and services controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Linguistic fractionalization 1 (basque knowledge)	-6.2360** (1.4198)		-4.7253** (1.6720)	-6.0675** (2.2970)	-4.8661** (1.6881)	-4.7407** (1.6639)	-4.3645* (1.7306)	-4.8062** (1.6762)	-4.8278** (1.4644)	-5.7952** (2.1390)
Political fractionalization 1 (all parties)		3.8862** (0.7942)	2.6743** (0.7896)	2.7431** (0.9553)	2.5888** (0.7989)	2.5968** (0.7970)	2.7263** (0.7944)	2.6597** (0.7892)	2.1123** (0.8143)	2.3795** (0.8138)
Infant education				-0.0076 (0.0104)						
School attendance					0.7322 (0.8040)					
Public school attendance						0.8355 (0.8027)				
Private school attendance							-2.7112 (4.6290)			
<i>Ikastola</i> attendance								4.2245 (6.1115)		
Service centers									-0.2232* (0.0975)	
Public outpatients' centers										-0.6633 (0.4635)
Constant	2.4649* (0.9733)	-3.2560** (0.7628)	-0.4508 (1.4065)	0.5588 (1.9924)	-0.4147 (1.4218)	-0.4740 (1.4012)	-0.5886 (1.3784)	-0.4274 (1.4062)	0.3574 (1.2586)	0.7494 (1.9887)
Number of observations	6475	5020	4483	3289	4483	4483	4483	4483	4483	4483
Number of municipalities	250	251	250	204	250	250	250	250	250	250

Standard errors in parentheses: * significant at 5%; ** significant at 1%.

Table 3: Terrorist assassinations, linguistic and political fractionalization: economic and demographic controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Linguistic fractionalization 1 (basque knowledge)	-4.6502** (1.7168)	-4.2655* (1.7362)	-4.3678* (1.6992)	-4.6485* (2.2759)	-3.0939* (1.5511)	-4.8372* (2.3343)	-2.5543 (1.7752)	-4.1627* (1.7376)	-5.2597** (1.8645)
Political fractionalization 1 (all parties)	2.6819** (0.7907)	2.7382** (0.8003)	2.6919** (0.7968)	1.4409 (0.7545)	1.5681* (0.7727)	1.4696 (0.7500)	1.7770* (0.7446)	1.8571 (1.0696)	2.4950** (0.8301)
Population in thousands	-0.0038 (0.0208)								
Population density		-0.0008 (0.0010)							
Emigration			-0.0214 (0.0337)						
People born in other provinces				4.3385 (4.5432)					
People born in a foreign country					9.6426 (28.7115)				
People born in the Basque Country						-4.6975 (4.6188)			
Unemployment rate							-5.2487 (3.7289)		
Land used in agriculture								0.0277 (0.0194)	
Financial institutions									-0.1918 (0.3304)
Constant	-0.4670 (1.4081)	-0.5656 (1.3825)	-0.5824 (1.3324)	0.1863 (1.0938)	0.2792 (1.2506)	4.8254 (4.5551)	1.6352 (2.3623)	-0.4038 (2.1192)	0.4953 (1.9858)
Number of observations	4483	4443	4470	3237	3237	3237	3237	2658	4195
Number of municipalities	250	250	250	249	249	249	249	248	249

Table 4: Robustness exercise: different measurements of linguistic and political fractionalization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Linguistic fractionalization 1 (basque knowledge)	-4.8278** (1.4644)	-6.4629** (1.5979)	-6.2555** (1.5860)						
Linguistic fractionalization 2 (mother tongue)				-7.2614** (1.5990)	-7.6775** (1.6304)	-7.6868** (1.6055)			
Linguistic fractionalization 3 (language spoken at home)							-5.6261** (1.3309)	-5.9608** (1.3704)	-6.0093** (1.3803)
Political fractionalization 1 (all parties)	2.1123** (0.8143)			1.9382* (0.8427)			1.2024 (0.9714)		
Political fractionalization 2 (nationalistic/non-nationalistic)		0.7162 (0.9537)			0.1131 (1.0258)			-0.3160 (1.0962)	
Political fractionalization 3 (right/left)			0.2946 (1.2344)			0.4445 (1.2800)			0.1745 (1.3854)
Service centers	-0.2232* (0.0975)	-0.3662** (0.1324)	-0.3558** (0.1364)	-0.1874* (0.0954)	-0.3545** (0.1295)	-0.3665** (0.1351)	-0.1950 (0.1019)	-0.2341 (0.1337)	-0.2493 (0.1368)
<i>Ikastola</i> attendance				5.7319 (6.6448)	6.5827 (6.4520)	6.8916 (6.5349)			
Public schools attendance		5.6174 (3.5387)	5.4360 (3.5359)						
Public outpatients' centers		-1.0009* (0.4759)	-1.0039* (0.4743)		-0.8967* (0.4445)	-0.8886* (0.4465)		-0.6442 (0.4277)	-0.6239 (0.4305)
Emigration							-0.0822 (0.0501)		
Financial institutions							-0.7343 (0.3868)	-0.8965* (0.4514)	-0.9232* (0.4487)
Constant	0.3574 (1.2586)	2.1125 (1.1976)	2.2424 (1.2710)	0.1284 (1.0505)	1.8370* (0.8655)	1.6736 (0.9907)	1.5723 (1.2973)	2.6749* (1.2846)	2.4465 (1.3847)

Table 5: Robustness exercise, Polarization measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Linguistic polarization 1 (basque knowledge)	-3.5157** (1.2350)	-5.1019** (1.2108)	-4.7345** (1.2175)						
Linguistic polarization 2 (mother tongue)				-4.0937** (0.9625)	-4.2893** (0.9423)	-4.2942** (0.9347)			
Linguistic polarization 3 (language spoken at home)							-3.9899** (0.9228)	-3.9584** (0.9094)	-3.9442** (0.9342)
Political polarization 1 (all parties)	2.3583 (1.2761)			2.4536 (1.2874)			0.5752 (1.5245)		
Political polarization 2 (nationalistic/non-nationalistic)		0.5816 (0.6878)			-0.2692 (0.7207)			-0.6246 (0.7530)	
Political polarization 3 (right/left)			-0.0630 (1.2033)			-0.5131 (1.2492)			-1.0966 (1.4154)
Service centers	-0.2479* (0.1009)	-0.3528** (0.1298)	-0.3272* (0.1324)	-0.2061* (0.0958)	-0.3336** (0.1252)	-0.3292** (0.1264)	-0.2126* (0.1026)	-0.2192 (0.1278)	-0.2237 (0.1262)
<i>Ikastola</i> attendance				7.2356 (6.6660)	6.4401 (6.4091)	5.8632 (6.5664)			
Public schools attendance		6.3281 (3.5681)	6.1667 (3.5769)						
Public outpatients' centers		-0.9944* (0.4764)	-1.0000* (0.4715)		-0.8859* (0.4368)	-0.8957* (0.4383)		-0.6560 (0.4232)	-0.6743 (0.4303)
Emigration							-0.0858 (0.0506)		
Financial institutions							-0.7714 (0.3943)	-0.8502 (0.4500)	-0.9062* (0.4383)
Constant	0.7440 (1.7509)	2.5254* (1.2568)	2.8539* (1.3976)	-0.2243 (1.3954)	2.0509* (0.8894)	2.3009 (1.2527)	2.3770 (1.6992)	3.0843* (1.2864)	3.6162* (1.6610)