

Ethnic Polarization, Potential Conflict, and Civil Wars

Jose G. Montalvo Marta Reynal-Querol
Universitat Pompeu Fabra The World Bank
and IVIE*

March 2005

Abstract

This paper analyzes the relationship between ethnic fractionalization, polarization, and conflict. In recent years many authors have found empirical evidence that ethnic fractionalization has a negative effect on growth. One mechanism that can explain this nexus is the effect of ethnic heterogeneity on rent-seeking activities and the increase in potential conflict, which is negative for investment. However the empirical evidence supporting the effect of ethnic fractionalization on the incidence of civil conflicts is very weak. Although ethnic fractionalization may be important for growth, we argue that the channel is not through an increase in potential ethnic conflict. We discuss the appropriateness of indices of polarization to capture conflictive dimensions. We develop a new measure of ethnic heterogeneity that satisfies the basic properties associated with the concept of polarization. The empirical section shows that this index of ethnic polarization is a significant variable in the explanation of the incidence of civil wars. This result is robust to the presence of other indicators of ethnic heterogeneity, alternative sources of data for the construction of the indicators, and alternative data for civil wars.

JEL Classification numbers: D74, D72, Z12, D63

*Instituto Valenciano de Investigaciones Economicas. We are grateful for comments by Antonio Villar, Joan Esteban, Paul Collier, Tim Besley and two anonymous referees. We thank the participants of seminars at the World Bank, Institut de la Mediterranea, Toulouse, Brown University, the European Economic Association Meetings and the Winter Meetings of the Econometric Society. We would like to thank Sergio Kurlat, Bill Easterly and Anke Hoeffler for sharing their data with us. Financial support from the BBVA Foundation and the Spanish Secretary of Science and Technology (SEC2003-04429) is kindly acknowledged

1 Introduction

The increasing incidence of ethnic conflicts and its much publicized consequences have attracted the interest of many researchers in the social sciences. Many studies have addressed directly the issue of ethnic diversity and its effects on social conflicts and civil wars. Political scientists have stressed the importance of institutions in the attenuation or intensification of social conflict in ethnically divided societies. Recently economists have connected ethnic diversity with important economic phenomena like investment, growth or the quality of government (Easterly and Levine 1997, Alesina et al. 2003 and La Porta et al. 1999). The number of papers dealing with the effects of ethnic diversity on issues of economic interest is growing rapidly.

In this respect it is common in recent work to include as a regressor in empirical growth estimations an index of ethnic fractionalization. There are several reasons to include such an indicator. First, some authors have argued that ethnically diverse societies have a higher probability of ethnic conflicts, which may lead to a civil war. Otherwise the political instability caused by potential ethnic conflicts has a negative impact on investment and, indirectly, on growth. Second, ethnic diversity may generate a high level of corruption which, in turn, could also deter investment. Finally it has been argued that in heterogeneous societies the diffusion of technological innovations is more difficult, specially when there is ethnic conflict among groups in a country. Business as usual is not possible in a society with a high level of potential ethnic conflict, since this situation affects all levels of economic activities. Trade may be restricted to individuals of the same ethnic group; public infrastructures can have an ethnic bias; government expenditure may favor some ethnic groups, etc. The common element in all these mechanisms is the existence of an ethnic conflict which, through social and political channels, spreads to the economy.

However many empirical studies find no relationship between ethnic fractionalization¹, ethnic conflicts, and civil wars. There are at least three alternative explanations for this lack of explanatory power. First, it could be the case that the classification of ethnic groups in the Atlas Nadorov Mira (ANM), source of the traditional index of ethnolinguistic fractionalization (ELF), is not properly constructed. Some authors² have used other sources, different from the ANM, to construct datasets of ethnic groups for a large sample of countries. In general the correlation between the index of fractionalization obtained using these alternative data sources is very high (over 0.8). Second, Fearon

¹Measured by the index of ethnolinguistic fractionalization (ELF) using the data of the Atlas Nadorov Mira.

²Montalvo and Reynal-Querol (2000), Alesina et al (2003) or Fearon (2003).

(2003)³ has argued that it is important to measure the "ethnic distance" across groups in order to obtain indicators of cultural diversity. He measures these "distances" in terms of the proximity in a tree diagram of the families of languages of different countries. As in the case of alternative data sources, the correlation of the index of ethnic fractionalization, using these "distances", with the original ELF is very high, 0.82.

The third alternative is the one we pursue in this paper. Up to now the alternative data on ethnic diversity and distances of ethnic groups in a country have been aggregated using indices of fractionalization. However, it is not clear to what extent an index of diversity could capture potential ethnic conflict. In principle claiming a positive relationship between an index of fractionalization and conflicts implies that the more ethnic groups there are the higher is the probability of a conflict. Many authors would dispute such an argument. Horowitz (1985), which is the seminal reference on the issue of ethnic groups in conflict, argues that the relationship between ethnic diversity and civil wars is not monotonic: there is less violence in highly homogeneous and highly heterogeneous societies, and more conflicts in societies where a large ethnic minority faces an ethnic majority. If this is so then an index of polarization should capture better the likelihood of conflicts, or the intensity of potential conflict, than an index of fractionalization.

The objective of this paper is to analyze the empirical support for the link between ethnicity and conflict. We pursue this objective by reexamining the evidence on the causes of civil wars using alternative indices to measure ethnic diversity. In the empirical section we show that the index of ethnic polarization is a significant explanatory variable for the incidence of civil wars. This result is robust to the use of other proxies for ethnic heterogeneity, alternative sources of data, and the use of a cross section instead of panel data. Therefore it seems that the weak explanatory power of ethnic heterogeneity on the incidence of civil wars found by several recent studies is due to the use of an index of fractionalization instead of an index of polarization.

This paper is organized as follows. Section 2 describes the characteristics of the index of fractionalization and compares it with an index of polarization. Section 3 presents the empirical results obtained by applying the index of fractionalization and the index of polarization to data on ethnic diversity. It is shown that for very high levels of fractionalization the level of the index of polarization can be very low. In fact, for high levels of diversity the correlation between fractionalization and polarization is negative. In this section we also discuss the source of data on ethnic and religious heterogeneity. Section 4 analyzes the causes of civil wars and compares the empirical performance of

³See also Caselli and Coleman (2003).

the polarization index proposed in this paper vis-à-vis the fractionalization index and other indices of ethnic heterogeneity. Section 5 contains a set of robustness checks. Section 6 summarizes the conclusions.

2 Ethnic heterogeneity and potential conflict

Several authors have stressed the importance of ethnic heterogeneity in the explanation of growth, investment, and the efficiency of government or civil wars. Easterly and Levine (1997) find empirical evidence to support their claim that the very high level of ethnic diversity of countries in Africa explains an important part of their poor economic performance. However their theoretical arguments, as they recognize explicitly, are based on "polarized societies"⁴ not on highly fractionalized cases. The effect of ethnic polarization on growth follows a more indirect channel: the choice of poor public policies which, in the end, negatively influences long-run growth. In particular ethnic polarization transforms economic policy via a rent seeking mechanism. Additionally ethnic polarization generate problems in the design of structural policies related to infrastructure and education. La Porta et al. (1999) point out that ethnic diversity leads to corruption and low efficiency governments that expropriate the ethnic losers.

Several authors have interpreted the finding of a negative relationship between ethnic diversity and growth to be a consequence of the high probability of conflict associated with a highly fractionalized society. For this reason many papers use the index of ethnolinguistic fractionalization (ELF) as the indicator of ethnic heterogeneity. The raw data for this index come from the Atlas Narodov Mira (1964) compiled in the former Soviet Union in 1960. The index ELF was originally calculated by Taylor and Hudson (1972). In general any index of fractionalization can be written as

$$FRAC=1-\sum_{i=1}^N \pi_i^2 = \sum_{i=1}^N \pi_i(1 - \pi_i) \quad (1)$$

where π_i is the proportion of people that belong to the ethnic (religious) group i and N is the number of groups. The index of ethnic fractionalization has a simple interpretation as the probability that two randomly selected individuals from a given country will not belong to the same ethnic group.⁵

However many authors have found that, even though ethnic fractionalization seems to be a powerful explanatory variable for economic growth, it is not significant in the explanation of civil

⁴See pages 1205, 1232 or 1241.

⁵Mauro (1995) uses this index as an instrument in his analysis of the effect of corruption on investment.

wars and other kinds of conflicts. These results has led many authors to disregard ethnicity as a source of conflict and civil wars. Fearon and Laitin (2003) and Collier and Hoeffler (2002) find that neither ethnic fractionalization nor religious fractionalization have any statistically significant effect on the probability of civil wars.

We argue that one possible reason for the lack of explanatory power of ethnic heterogeneity on the probability of armed conflicts and civil wars is the measure for heterogeneity. In empirical applications researchers should consider a measure of ethnic polarization, the concept used in most of the theoretical arguments, instead of an index of ethnic fractionalization. We propose an index of ethnic polarization with the form

$$Q = 1 - \sum_{i=1}^N \left(\frac{1/2 - \pi_i}{1/2} \right)^2 \pi_i = 4 \sum_{i=1}^N \pi_i^2 (1 - \pi_i).$$

The original purpose of this index was to capture how far is the distribution of the ethnic groups from the $(1/2, 0, 0, \dots, 0, 1/2)$ distribution (bipolar), which represents the highest level of polarization⁶. This type of reasoning is frequently present in the literature on conflict⁷ and, in particular, on ethnic conflict. Esteban and Ray (1999) show, using a behavioral model and a quite general metric of preferences, that a two-point symmetric distribution of population maximizes conflict.

In addition Horowitz (1985) points out that ethnic conflicts will take place in countries where a large ethnic minority faces an ethnic majority. Therefore ethnic dominance, or the existence of a large ethnic group, although close to being a necessary condition for a high probability of ethnic conflict, is not sufficient. You also need that the minority is not divided into many different groups but is also large. The Q index captures the idea of a large majority versus a large minority as the worst possible situation since the index in this case is close to the maximum.

Collier and Hoeffler (1998) note "coordination cost would be at their lowest when the population is polarized between an ethnic group identified with the government and a second, similarly sized ethnic group, identified with the rebels." Collier (2001) also emphasizes that the relationship between ethnic diversity and the risk of violent conflicts is not monotonic. Highly heterogeneous societies have even a lower probability of civil wars than homogeneous societies. The highest risk is associated with the middle range of ethnic diversity.⁸ The Q index satisfies this condition.

Notice also that Fearon (2003) points out that the index of fractionalization, being not sensitive

⁶See also Reynal-Querol (2002).

⁷Montalvo and Reynal-Querol (2005) show how to obtain the Q index from a pure contest model.

⁸Horowitz (1985) also argues that there is less violence in highly homogeneous and highly heterogeneous countries.

to discontinuities, cannot capture important differences in ethnic structures. In particular the idea of majority rule is not well reflected by the index of fractionalization. By contrast the sensitivity of the Q is the highest when groups are close to 50%.

2.1 Fractionalization versus the Q index

How does fractionalization compare with the Q index? As mentioned above the index of fractionalization can be interpreted as the probability that two randomly selected individuals do not belong to the same group. Let's consider the case of two groups. In this situation the index of fractionalization can be written as

$$FRAC = 1 - \pi_1^2 - \pi_2^2 = \pi_1(1 - \pi_1) + \pi_2(1 - \pi_2) = 2\pi_1\pi_2$$

simply because $\pi_1 + \pi_2 = 1$.

Following the definition of the Q index we can write it, for the case of two groups, as

$$Q = 4(\pi_1(\pi_1(1 - \pi_1)) + \pi_2(\pi_2(1 - \pi_2))) = 4\pi_1\pi_2$$

which is equal to the index FRAC up to a scalar. When we move from two groups to three groups the relationship between FRAC and Q breaks down. For instance FRAC can be calculated for the case of three groups as

$$FRAC = \pi_1(1 - \pi_1) + \pi_2(1 - \pi_2) + \pi_3(1 - \pi_3)$$

In this case, and without considering the scale factor that bounds it between 0 and 1, the Q index is proportional to

$$Q \propto \pi_1(\pi_1(1 - \pi_1)) + \pi_2(\pi_2(1 - \pi_2)) + \pi_3(\pi_3(1 - \pi_3))$$

Comparing these two formulas we can see the basic difference between the interpretation of the fractionalization index and the meaning of the Q index. In FRAC each of the terms in the sum is the probability that two randomly selected individuals belong to different groups when one of them belongs to a particular group. For instance $\pi_i(1 - \pi_i)$ is the probability that two individuals belong to different groups when one of them belongs to group i . These probabilities have the same weight in each of the terms of the fractionalization index but they have weight equal to the relative size of group i in the case of the Q index. In the fractionalization index the size of each group has no effect

on the weight of the probabilities of two individuals belonging to different groups whereas in the Q index these probabilities are weighted by the relative size of each group.

Looking at both indices one may wonder how much large and small groups contribute to the value of the index with respect to their relative size. The different weighting scheme is crucial to answer this question. Let's define c_i as the proportional contribution of group i to the index of fractionalization, that is $c_i = \pi_i(1 - \pi_i)/(\sum \pi_i(1 - \pi_i))$. Define \tilde{c}_i as the proportional contribution of group i to the index of polarization, that is $\tilde{c}_i = \pi_i^2(1 - \pi_i)/(\sum \pi_i^2(1 - \pi_i))$. If all the groups have equal size the proportional contribution of each of the groups is equal to its relative size in both, fractionalization and polarization, that is $c_i = \tilde{c}_i = \pi_i$. Imagine now that we increase the size of one group by epsilon and decrease the size of another group by the same amount. Now the proportional contribution of the largest group in the index of fractionalization is smaller than its relative size, $c_i < \pi_i$, and the reverse happens for the smallest group. In the index of polarization the result is the opposite: the proportional contribution of the largest group in the index of polarization is larger than its relative size, $\tilde{c}_i > \pi_i$, and the reverse happens to the smallest group. Loosely speaking⁹ we can say that large (small) groups contribute to the index of polarization proportionally more (less) than their relative size. The opposite is true for the index of fractionalization: large (small) groups contribute to the index less (more) than their relative size.

3 From income inequality to ethnic fractionalization

The index of fractionalization has, at least, two theoretical justifications based on completely different contexts. In industrial organization the literature on the relationship between market structure and profitability has used the Herfindahl-Hirschman index to measure the level of market power in oligopolistic markets.¹⁰ The derivation of the index in this context starts with a noncooperative game where oligopolistic firms play Cournot strategies. Therefore the index can summarize the market power in games that work through the market.¹¹

The second theoretical foundation for the index of fractionalization comes from the theory of inequality measurement. One of the most popular measures of inequality is the Gini index, G , that has the general form

⁹Montalvo and Reynal-Querol (2002) for a formal proof of this claim.

¹⁰This index has been also used in antitrust cases.

¹¹However the index of fractionalization may not be appropriate when the structure of power works through political or military processes as they appear to follow rent-seeking or conflict models.

$$G = \sum_{i=1}^N \sum_{j=1}^N \pi_i \pi_j |y_i - y_j|$$

where y_i represent the income level of groups i and π_i is its proportion with respect to the total population. This formulation is specially suited to measure income and wealth inequality. However, if we want to measure ethnic diversity the "distance" between ethnic groups may be a very difficult concept to measure. In addition the dynamics of the "we" versus "you" distinction is more powerful than the antagonism generated by the "distance" between them. For these reason we may want to consider only if an individual belongs or does not belong to an ethnic group. If we substitute the Euclidean income distance $\delta(y_i, y_j) = |y_i - y_j|$, by a discrete metric (belong/do not belong)

$$\begin{aligned} \delta(y_i, y_j) &= 0 & \text{if } i = j \\ &= 1 & \text{if } i \neq j \end{aligned}$$

Therefore the discrete Gini (DG) index can be written as

$$DG = \sum_{i=1}^N \sum_{j \neq i}^N \pi_i \pi_j.$$

It is easy to show that the discrete Gini index (DG) calculated using a discrete metric is simply the index of fractionalization

$$DG = \sum_{i=1}^N \sum_{j \neq i}^N \pi_i \pi_j = \sum_{i=1}^N \pi_i \sum_{j \neq i}^N \pi_j = \sum_{i=1}^N \pi_i (1 - \pi_i) = (1 - \sum_{i=1}^N \pi_i^2) = FRAC.$$

3.1 From income polarization to discrete polarization and the Q index

We showed in the previous section that the index of fractionalization can be interpreted as a Gini index with a discrete metric (belong/do not belong to the group) instead of an Euclidean income distance. The Q index can be interpreted as the polarization measure of Esteban and Ray (1994) with a discrete metric. By imposing three reasonable axioms Esteban and Ray (1994) narrow down the class of allowable polarization measures to only one measure, P , with the following form

$$P = k \sum_{i=1}^N \sum_{j=1}^N \pi_i^{1+\alpha} \pi_j |y_i - y_j|$$

for some constants $k > 0$ and $\alpha \in (0, \alpha^*]$ where $\alpha^* \simeq 1.6$. Notice that when $\alpha = 0$ ¹² and $k = 1$ this polarization measure is precisely the Gini coefficient. Therefore the fact that the share of each group is raised to the $1 + \alpha$ power, which exceeds one, is what makes the polarization measure significantly different from inequality measures. The parameter α can be treated as the degree of "polarization sensitivity." If we substitute the Euclidean income distance $\delta(y_i, y_j) = |y_i - y_j|$, by a discrete metric (belong/do not belong), then we have what we call discrete polarization

$$DP(\alpha, k) = k \sum_{i=1}^N \sum_{j \neq i} \pi_i^{1+\alpha} \pi_j$$

The discrete nature (belong/do not belong) of the distance across groups has important implications for the properties of the index. In particular, and in contrast with the polarization index of Esteban and Ray (1994), there is only one level of polarization sensitivity ($\alpha = 1$) for which the discrete polarization measure satisfies the properties of polarization. In addition there is only one value of k ($k = 4$) such that the index DP ranges between 0 and 1. The Q index is precisely the index $DP(1, 4)$ ¹³.

The index of polarization of Esteban and Ray (1994) was initially thought as a measure of income or wealth polarization. As such it is difficult to implement empirically since its value depends critically on the number of groups, the value of k and the value of α ¹⁴. However in terms of income or wealth it is not clear which levels distinguish different groups with a common identity. Where does the middle class start? How "rich" is rich? This difficulty together with the uncertainty over the right parameter for α has reduced the empirical applicability of the polarization index. In the case of ethnic diversity the identity of the groups is less controversial. Additionally the discrete nature of the distance (belong/do not belong) fixes the values of α and k . This makes the Q index easily applicable to data on ethnic and religious diversity.

¹²Strictly speaking for $\alpha = 0$ this is not an index of polarization.

¹³For proofs of these claims and all the technical details on the relationship between fractionalization, polarization and the Q index see Montalvo and Reynal-Querol (2002).

¹⁴See Duclos et al. (2004) for a recent reconsideration of the empirical measurement of polarization with Euclidean distances.

4 The empirical relationship between fractionalization and polarization

In this section we compare the empirical content of measures of fractionalization and indicators of polarization. Keefer and Knack (2002) argue that their income based measures of polarization are similar to the Gini coefficient suggesting that in practice the divergence between income-based polarization and inequality is more theoretical than actual. However the difference between ethnic polarization and fractionalization is both theoretical and actual. Theoretically, as we showed in sections 2, discrete polarization and fractionalization represent quite different concepts. In this section we describe the alternative data sources for ethnic and religious heterogeneity and we show that the index of fractionalization and polarization are very different independent of the source of data used in their calculation.

4.1 Sources of data on ethnic heterogeneity

There are basically three sources of ethnolinguistic diversity across countries: the **World Christian Encyclopedia (WCE)**, the **Encyclopedia Britannica (EB)** and the **Atlas Narodov Mira (ANM)** (1964). For reasons that we have explained elsewhere¹⁵ we think the most accurate description of ethnic diversity is the one in the WCE, which contains details for each country on the most diverse classification level, which may coincide with an ethnolinguistic family or subfamilies, subpeoples, etc. We follow Vanhanen (1999) in taking into account only the most important ethnic divisions and not all the possible ethnic differences or groups. Vanhanen (1999) uses a measure of genetic distance to separate different degrees of ethnic cleavage. The proxy for genetic distance is "the period of time that two or more compared groups have been separated from each other, in the sense that intergroup marriage has been very rare. The longer the period of endogamous separation the more groups have had time to differentiate." This criterion is reasonable since we are using discrete distances and, therefore, we have to determine the identity of the relevant groups.

Another source of data on ethnic diversity is the **Encyclopedia Britannica (EB)**¹⁶ which uses the concept of geographical race. A third source of data on ethnolinguistic diversity is provided by the **Atlas Narodov Mira (ANM)** (1964), the result of a large project of the Department of Geodesy and Cartography of the State Geological Committee of the old USSR.

¹⁵For a detailed discussion of the differences between these data sources see Montalvo and Reynal-Querol (2000).

¹⁶This is the basic source of data on ethnic heterogeneity of Alesina et al. (2003).

There are also several possible sources of data on religious diversity. Barret’s (1982) **World Christian Encyclopedia (WCE)** provides information on the size of religious groups for a large cross-section of countries. The WCE has several well-known shortcomings when dealing with data on religion.¹⁷ **L’Etat des Religions Dans le Monde (ET)**, which is based on a combination of national data sources and the WCE, provides information on the proportions of followers of Animist and Syncretic cults, which we believe is important for the calculation of indices of religious heterogeneity. For this reason we use the ET as our primary source for the religious data.¹⁸ Alesina et al. (2003) use the data on religious diversity compiled by the **Encyclopedia Britannica (EB)**.¹⁹

4.2 Are empirical polarization and fractionalization very different?

Once we have described the different sources of data available to measure ethnic and religious heterogeneity we need to show the empirical relationship between both indices. Figure 1 presents the relationship between ethnolinguistic polarization and fractionalization using our data sources. It shows that for low levels of fractionalization the correlation between ethnic fractionalization²⁰ and polarization is positive and high. In particular, from our previous discussion in section 2.1 we know that when there are only two ethnic groups ethnic polarization is two times ethnic fractionalization. That is the reason why the slope of the line is 1/2 for ethnic polarization up to 0.4²¹. However for the medium range the correlation is zero and for high levels of fractionalization the correlation with polarization is negative.

Figure 2 presents the scatterplot of religious fractionalization versus religious polarization. It shows a similar pattern: for low levels of religious fractionalization the correlation with polarization is positive. However for intermediate and high levels of religious fractionalization the correlation is zero. Therefore the correlation is low when there is a high degree of heterogeneity, which is the interesting case.

Figures 3 and 4 confirm that the previous results do not depend on the source of data used in the construction of the indices. Figure 3 shows the relationship between the index of ethnic

¹⁷See L’Etat des Religions dans le Monde (1987) pages 7-9.

¹⁸Our secondary source is **The Statesman’s Yearbook (ST)** which is only based on national sources.

¹⁹The correlation of the indices constructed with the different sources of religious diversity is very high, as it was in the case of ethnic heterogeneity.

²⁰The index of ethnic fractionalization calculated with our data has a correlation of 0.86 with the index obtained using the Atlas Nadorov Mira (ELF). The correlation with the index of Alesina et al. (2003) is 0.83.

²¹Nevertheless we should notice that only in 3,6% of the countries the number of groups is equal to two.

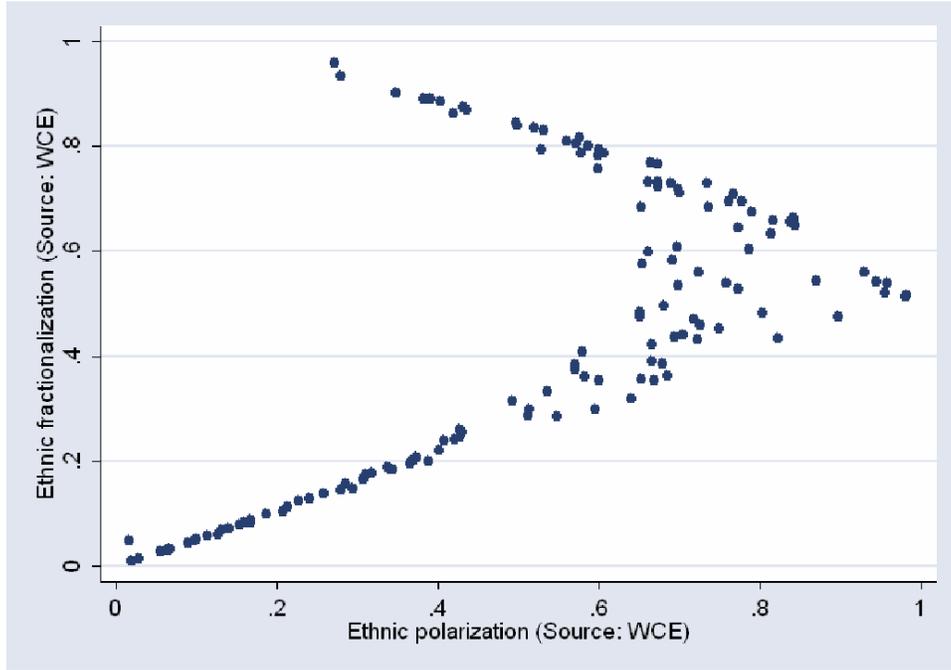


Figure 1: Ethnic fractionalization versus polarization. Source: WCE.

fractionalization and ethnic polarization constructed using the data of Alesina et al. (2003). The shape in figure 3 is very similar to the one in figure 1. Figure 4 shows ethnic fractionalization and polarization calculated using the data from the Atlas Nadorov Mira, the third basic source of data on ethnic diversity. The graph is very similar to figures 1 and 3.

A previous version of this paper²² shows that nine out of the ten most ethnically polarized countries have suffered a civil war during the sample period (1960-95). In the case of ethnic fractionalization only four out of the ten most fractionalized countries have suffered a civil war. It is interesting to describe the situation of a countries that have a high degree of polarization but a low degree of fractionalization (close or below the average). Guatemala is a good example of this situation. The ethnic composition of the population is 55% Ladino (Mestizo), 42% Maya (Amerindian) and 3% other small groups. This implies a very high degree of polarization (0.96), and a low level of fractionalization (0.52).

During the same sample period civil wars occurred in 7 out of 10 countries with the highest level of religious polarization. However only three out of the ten countries with the highest level

²²Montalvo and Reynal-Querol (2004).

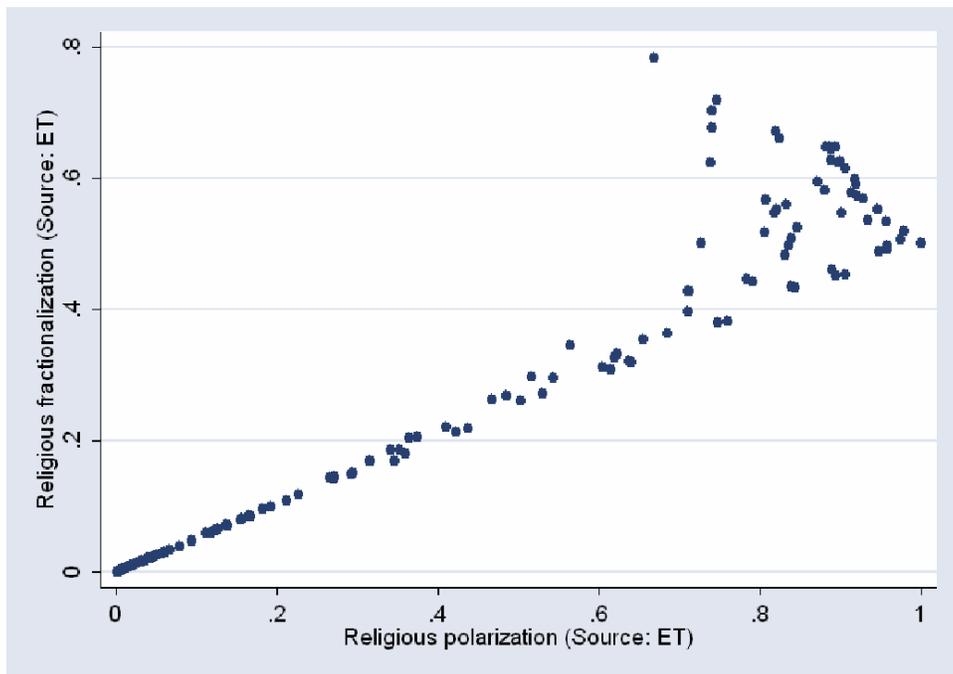


Figure 2: Religious fractionalization versus polarization. Source: ET.

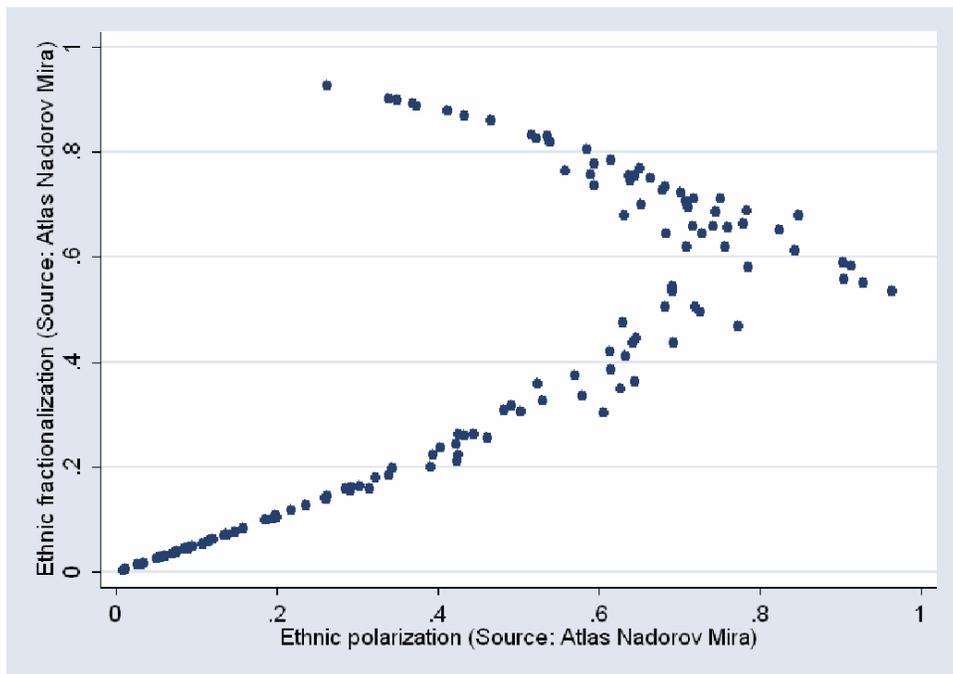


Figure 3: Ethnic fractionalization versus polarization. Source: Atlas Nadorov Mira.

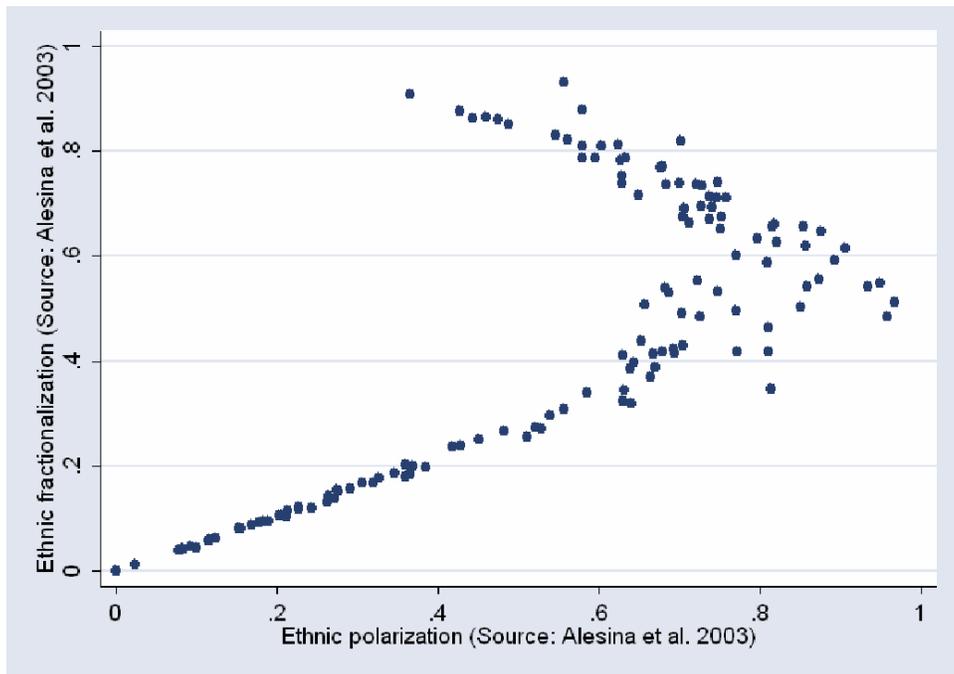


Figure 4: Ethnic fractionalization versus polarization. Source: Alesina et al. (2003).

of religious fractionalization suffered a civil war. For instance in Nigeria there is a high level of religious polarization between Christians (49%) and Muslims (45%) similar to the case of Bosnia (50% Christians and 40% Muslims). In both cases the degree of religious fractionalization is low.

5 Regression results

Several authors have stressed the importance of ethnic heterogeneity in many economic phenomena (growth, investment, etc.). One basic element that explains the relationship between heterogeneity and development is the existence of potential ethnic conflict that, through social and political channels spreads to the economy. There is no doubt that civil wars are traumatic events that damage economic development. We argued earlier that the index of polarization is a good indicator to capture the extent of social conflicts. But then, is it polarization or fractionalization that matters in the explanation of conflicts in heterogeneous societies?

In this section we present the estimation of a logit model for the incidence of civil wars as a function of polarization and fractionalization measures of ethnic and religious heterogeneity. The sample includes 138 countries during 1960-99. We divide the sample into five-year periods. The endogenous variable is the incidence of a civil war. We use the Peace Research Institute of Oslo (PRIO) dataset for civil wars. Our basic endogenous variable corresponds to the definition of intermediate and high intensity civil wars of PRIO, which we call PRIOCW. PRIO defines an intermediate and high intensity armed conflict²³ as a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 yearly battle-related deaths and a minimum of 1,000 during the course of the civil war. We focus only on civil wars, categories 3 and 4 of conflict of PRIO, which cover civil conflicts with and without interference from other countries.

The explanatory variables follow the basic specifications of Fearon and Laitin (2003), Doyle and Sambanis (2000) and Collier and Hoeffler (2002). Fearon and Laitin (2003) argue that income per capita is a proxy for "state's overall financial, administrative, police and military capabilities." Once a government is weak rebels can expect a higher probability of success. In addition a low level of income per capita reduces the opportunity cost of engaging in a civil war. Recently Miguel et al. (2004) have argued that the measurement of the impact of GDP growth on civil wars is complicated since there are endogeneity issues. Their set-up is very different from ours. They use annual data

²³See the Appendix I for more details on this definition.

and GDP growth. In this situation the potential endogeneity problem of GDP growth with respect to conflict is very high. For this reason Miguel et al. (2004) use rainfall as an instrument for GDP growth. We use periods of five years for civil wars and the GDP per capita at the beginning of each period. This set up reduces also the potential endogeneity problem.

The size of the population is another usual suspect in the explanation of civil wars. First, the usual definitions of civil war set always a threshold in the number of deaths, which suggests that one should control by population as a scale factor. Second, Collier and Hoeffler (2002) consider that the size of the population is an additional proxy for the benefits of a rebellion since it measures potential labor income taxation. Finally Fearon and Laitin (2003) indicate that a large population implies difficulties in controlling what goes on at the local level and increases the number of potential rebels that can be recruited by the insurgents.

Mountains are another dimension of opportunity since this terrain could provide a safe haven for rebels. Long distances from the center of the state's power also favors the incidence of civil wars, specially if there is a natural frontier between them, like a sea or other countries. Collier and Hoeffler (2002) point out that the existence of natural resources provide an opportunity for rebellion since these resources can be used to finance the war and increases the payoff if victory is achieved. Finally most of the literature considers the effect of democracy.

Therefore the explanatory variables for the core specification of the incidence of civil wars include the log of real GDP per capita in the initial year (LGDPC), the log of the population at the beginning of the period (LPOP), primary exports (PRMEXP), mountains (MOUNTAINS), noncontiguous states (NONCONT), and the level of democracy (DEMOCRACY). Using this core specification we check the empirical performance of indices of fractionalization and polarization as well as other measures of ethnic and religious heterogeneity.

5.1 Ethnic heterogeneity and the incidence of civil wars

Table 1 reports the results obtained using alternatively measures of fractionalization and polarization²⁴. The first column shows that the index of ethnolinguistic fractionalization (ETHFRAC) has no statistically significant effect on the incidence of civil wars. This result is consistent with Fearon and Laitin (2003) and Collier and Hoeffler (1998). However if we substitute the index of ethnic fractionalization by the Q index of ethnic polarization, ETHPOL, we find (column 2) a positive and statistically significant effect on the incidence of civil wars. The initial GDP per capita has a

²⁴All the tables show the z statistic tests calculated using the standard errors adjusted for clustering.

negative effect²⁵ in the incidence of civil wars while the log of population has a positive effect.²⁶ We find no significant effect of mountains, noncontiguous states or primary exports on the incidence of civil wars. Finally the level of democracy has a positive but not statistically significant coefficient. Column 3 checks the relative strength of the index of ethnic polarization versus fractionalization and shows that the coefficient on ethnic fractionalization is not significantly different from zero while the one on polarization is positive and significant.

[Insert Table 1 about here]

The effect of ethnic polarization is not only statistically significant but also economically important. Using the results in column 3, if the level of polarization increases from the average (0.51) to the level of Guinea (0.84) then the probability of conflict almost doubles. An increase in one standard deviation (0.24) of the average polarization increases the probability of conflict by 67%.

Another potential dimension of social heterogeneity that can generate conflictive situations is religion. Column 4 shows that religious fractionalization (RELFrac) is not statistically significant. Neither is the coefficient of religious polarization (RELPol) in column 5. Column 6 shows the basic logit regressions using both religious fractionalization and religious polarization. The coefficient of the index of religious fractionalization (RELFrac) is marginally insignificant while the index of religious polarization (RELPol) is statistically significant. When both indicators are included in the same specification, religious polarization has the expected positive sign but fractionalization has a negative impact on the probability of civil wars. This means that, conditional on a given degree of polarization, more religious diversity decreases the probability of a civil war. We argued before that a high number of different groups increases the coordination problems and, therefore, given a level of polarization, the probability of civil wars may be smaller. For instance Korea and Sri Lanka have the same level of religious polarization (0.72). However Sri Lanka, which suffered a civil war, has a degree of religious fractionalization of 0.49 while Korea, with a much higher level (0.79), did not experience a civil war.

In column 7 we include together the index of ethnic polarization and religious polarization. Only the estimated coefficient of the first one is statistically significant. If we add also as explanatory variables the degree of ethnic fractionalization and religious fractionalization (column 8) only the

²⁵Depending on the particular specification this effect could be statistically significant or not. In the next section we show that the coefficient of the initial GDP per capita is very significant and robust when we use other datasets on civil wars different from PRIOCW.

²⁶The same results are reported by Doyle and Sambanis (2000), Fearon and Laitin (2003) and Collier and Hoeffler (1998, 2002).

coefficient of ethnic polarization is significantly different from 0. It seems clear that ethnic polarization has a robust and powerful explanatory power on civil wars in the presence of other indices of fractionalization and polarization while the statistical relevance of religious polarization depends on the particular specification²⁷. Therefore in the rest of the paper we check the robustness of the results of table 1 using only ethnic polarization.

5.2 Robustness to alternative measures of heterogeneity

Table 2 reports the performance of the Q index in the presence of other indicators of ethnolinguistic heterogeneity. Column 1 displays, to simplify the comparisons, the results of table 1 for the core specification. Besides the indices of fractionalization and polarization the literature has proposed some other indicators of potential ethnic conflict. Collier (2001) notices that ethnic diversity could be not only an impediment for coordination but also an incitement to victimization. Dominance, or one ethnic group in a majority, can produce victimization and, therefore, increase the risk of a civil war. Therefore the effect of ethnic diversity will be conditional on being measured as dominance or fractionalization. In principle fractionalization should make coordination more difficult and, therefore, civil wars will be less probable since it will be difficult to maintain cohesion among rebels. Collier (2001) argues that the problem with the results in Easterly and Levine (1997) is that they are unable to distinguish between fractionalization and dominance. The empirical results reported by Collier (2001) seems to indicate that a good operational definition of dominance implies a group that represents between 45% and 90% of the population²⁸. However Collier and Hoeffler (2002) find that dominance, defined as mentioned above, has only a weak positive effect on the incidence of civil wars. In column 2 of table 2 we show that ethnic dominance (ETHDOM) does not have any significant effect in our core specification. When ethnic dominance is included with the Q index, column 3, its coefficient is not significant while ethnic polarization continues being a significant explanatory variable on the probability of civil wars. Caselli and Coleman (2002) propose another indicator which is the product of the largest ethnic group (ETHLRG) by primary exports (PRIMEXP). In column 4 we can see that this variable has a coefficient that is not significantly different from 0.

²⁷For a more detailed account of the performance of religious polarization in the context of many different specifications see Montalvo and Reynal-Querol (2000).

²⁸Collier (2001) justifies his choice by arguing that "the level of significance and the size of the coefficient of dominance reach a maximum when dominance is defined on the range of 45%-90% of the population". Since we want to check the robustness of our index Q to alternative measures we have chosen the "statistically most powerful" empirical definition for dominance.

In column 5 we show that the index of polarization is significant even when the product of the largest ethnic group by primary exports is included as an explanatory variable. Finally we could also include the size of the largest minority (LARGMINOR) as another way to proxy polarization. Column 6 shows that the coefficient on this new variable is not statistically significant while ethnic polarization continues to be significant even in the presence of this new variable (column 7).

[Insert table 2 about here]

6 Some additional test of robustness

The previous section has shown that the relevance of ethnic polarization in the explanation of civil wars is robust to the presence of other indicators of ethnic heterogeneity like fractionalization, dominance or the product of the size of the largest group by the proportion of primary exports. In this section we explore the robustness of previous results. In particular, we discuss: (a) different definitions of civil wars; (b) the inclusion of regional dummies or the elimination of particular regions; (c) the use of different data sources to construct the indices; (d) cross-section regressions covering the whole period.

6.1 The operational definition of civil war

In this section we check the robustness of the results to the use of an alternative definition of civil war. Up to this point we have worked with the definition proposed by PRIO for intermediate and high intensity armed conflicts²⁹, which we name PRIOCW. PRIO offers also series to construct armed conflicts that generate more than 25 deaths per year, PRIO25, and very intense armed conflicts (more than 1,000 deaths yearly), PRIO1000. Another source of data is Doyle and Sambanis (2000) (DSCW), who define civil war as an armed conflict with the following characteristics: "(a) it caused more than 1,000 deaths; (b) it challenged the sovereignty of an internationally recognized state; (c) it occurred within the recognized boundary of that state; (d) it involves the state as a principal combatant; (e) it included rebels with the ability to mount organized armed opposition to the state; and (f) the parties were concerned with the prospects of living together in the same political unit after the end of the war."³⁰

Finally Fearon and Laitin (2003) use a different operational definition of civil war (FLCW). For

²⁹Those causing more than 25 yearly deaths and a minimum of 1,000 deaths over the course of the war.

³⁰This definition is practically identical to Singer and Small (1994) in their Correlates of Wars project (COW).

these authors a violent conflict should meet the following criteria to be coded as a civil war: (1) it should involve the "fighting between agents of (or claimants to) a state and organized, non-state groups who sought either to take control of a government, take power in a region, or use violence to change government policies, (2) the conflict killed or has killed at least 1,000 over its course, with a yearly average of at least 100 deaths, (3) at least 100 were killed on both sides (including civilians attacked by rebels)."

[Insert table 3 about here]

Table 3 shows the proportion of armed conflicts over total observations using different definitions of armed conflict and different periodicity. The closest definitions are the PRIOCW and Doyle and Sambanis (DSCW). For annual data the proportion of armed conflicts ranges from 5.9% (PRIO1000) to 15.2% (PRIO25). For five years periods the proportions are between 10.1% and 22.2%. Finally if we consider the whole period the proportions range from 29.2% up to 53.6%.

[Insert table 4 about here]

Table 4 shows the results of the basic specification using the different definitions of armed conflicts. Column 1 shows that ethnic polarization is statistically significant when we use as dependent variable the definition of civil wars of Doyle and Sambanis (2000). In fact we can see that the size of the coefficient on ethnic polarization is very similar to the one obtained using the intermediate and high definition of armed conflict of PRIO (PRIOCW). We already argued that in practice the data of Doyle and Sambanis (2000) and the PRIOCW are very similar. Column 2 shows that ethnic polarization is marginally statistically significant if we use the definition of civil war of Fearon and Laitin (2003). Columns 3 and 4 show that the statistical significance of the coefficient on ethnic polarization is robust to the use of the other two definitions of PRIO. In fact it is interesting to notice that the coefficient that measures the effect of ethnic polarization on the probability of civil wars increases monotonically with the intensity of the conflict (2.05 including minor conflicts; 2.28 for intermediate and high intensity conflicts, and 2.33 for the most violent conflicts). Another interesting fact in columns 1 to 4 of table 4 is the robustness of the coefficient of initial GDP per capita. It seems that the relative weakness of the coefficient of this variable in tables 1 and 2 is due to the definition of civil war used (intermediate and high intensity types following PRIO).

Finally we should notice that using the data of Doyle and Sambanis (2000) and Fearon and Laitin (2003) the importance of initial level of democracy is much larger than using the dataset of PRIO. Since using the PRIO dataset democracy is very far from being statistically significant and it reduces the sample size we also consider the effect of excluding this variable from the specification. Column

5 shows that the results of table 1 are robust to the exclusion of the DEMOCRACY variable, but the sample size increases significantly due to the large number of missing data in that variable. Columns 6 to 9 show that the statistical significance of ethnic polarization in the explanation of civil wars is robust to the use of alternative datasets for the endogenous variable even if we do not consider the DEMOCRACY variable in the specification.

6.2 Robustness to regional effects

Are the results robust to including dummy variables for the different regions of the world? Are they robust to the elimination of regions that are considered specially conflictive? We investigate this questions in table 5. Columns 1 and 2 show that ethnic polarization is statistically significant in the presence of regional dummies³¹, with and without the inclusion of ethnic fractionalization, which is not significant. The elimination from the sample of the countries in Sub-Saharan Africa, column 3, does not affect the statistical significance of ethnic polarization. If we eliminate those African countries and include in the regression the index of ethnic fractionalization, column 4, then the coefficient on ethnic polarization is not significant. However, as we argued before, since ethnic fractionalization is not statistically significant it seems clear that its presence increases the standard error of the ethnic polarization estimated coefficient. Columns 5 and 6 show the robustness of ethnic polarization to eliminating from the sample the Latin American countries. Finally, columns 7 and 8 confirm that the effect of ethnic polarization on civil wars is robust to the elimination from the sample of the Asian countries.

[Insert table 5 about here]

6.3 The effect of alternative data sources for ethnic heterogeneity

One may wonder if part of the results in the previous sections are driven by the data used in the construction of the indices of polarization and fractionalization. We pointed out that there are three basic sources of data on ethnic heterogeneity: the World Christian Encyclopedia (base of our data), the Encyclopedia Britannica (source of the indices of Alesina et al. 2003) and the Atlas Nadorov Mira (ANM) (source of the well-known ELF). We argued before that the correlation between our indicators and the ones calculated using other sources of data is quite high. The Q index of polarization calculated using the row data of Alesina et al. (2003)³² has a positive

³¹The dummies are for Sub-Saharan Africa, Latin America and Asia.

³²We thank Sergio Kurlat and Bill Easterly for sharing with us the row data of Alesina et al. (2003).

(1.93) and statistically significant effect ($z=2.32$) on the incidence of civil wars (PRIOCW), opposite to what happens with the coefficient of the index of fractionalization calculated using the same source (estimated coefficient=1.27 and $z=1.67$). When we run the regression with the Q index of polarization calculated using the row data of the Atlas Nadorov Mira, we find out that it has a positive effect (estimated coefficient=2.35 and $z=3.33$) on the probability of civil wars, while the index of fractionalization calculated with the same dataset is not statistically significant (estimated coefficient=1.20 and $z=1.41$).

The results using other definitions of civil wars are equally supportive of the robustness of the results. For instance for intense civil wars (PRIO1000 definition) the coefficient on ethnic polarization calculated using the data of Alesina et al. (2003) is 1.95 ($z=2.22$). If ethnic polarization is calculated using the ANM then its estimated coefficient on the incidence of intense civil wars is 1.98 ($z=2.63$). In both cases ethnic fractionalization is not statistically significant.

6.4 Cross-section regressions

In the empirical section we have been working with a panel of countries divided in five-year periods. However it seems reasonable to perform a final robustness check running the logit regressions in a cross section. The dependent variable takes now value 1 if a country has suffered a civil war during the whole sample period (1960-1999) and zero otherwise. GDP per capita, population, democracy and primary exports are measured at the beginning of the period (1960). Table 6 shows that the index of ethnolinguistic polarization is significantly different from zero with (column 1) or without including the regional dummy variables (column 2)³³. The result is robust to the use of different dataset for civil wars like Doyle and Sambanis (2000), columns 3 and 4, or Fearon and Laitin (2003), columns 5 and 6.

[Insert table 6 about here]

7 Conclusions

Several recent papers have documented the negative effect of ethnic fractionalization on economic development. Some authors have argued that a high degree of ethnic fractionalization increases potential conflict, which has negative effects on investment and increases rent seeking activities.

³³If instead of ethnic polarization we include ethnic fractionalization the estimated coefficient is 1.50 with a z-statistic of 1.57.

However, many of the theoretical arguments supporting the effect of ethnic heterogeneity on potential conflict were developed in the context of polarized societies. In addition researchers use frequently the index of fractionalization to capture the concept of polarization. We argue that the measure of ethnic heterogeneity appropriate to capture potential conflict should be a polarization measure. In fact Horowitz (1985), in his seminal book on ethnic groups in conflict, points out that the most severe conflicts arise in societies where a large ethnic minority faces an ethnic majority. The index of ethnic fractionalization is not able to capture this idea appropriately.

We define an index of polarization based on a discrete metric that we call discrete polarization. It turns out that our index is related with the original index of income polarization of Esteban and Ray (1994). We describe a particular case of discrete polarization, the Q index, that satisfies the basic properties associated with the concept of polarization. Keefer and Knack (2002) argue that their income-based measures of polarization are very similar to the Gini coefficient suggesting that in practice the divergence between income-based polarization and inequality is more theoretical than actual. In this paper we have shown that the difference between ethnic polarization and fractionalization is both theoretical and actual.

In the empirical section we show that the index of ethnic fractionalization does not have a significant effect on the likelihood of conflicts. Therefore it is unlikely that ethnic fractionalization affects economic development through an increase in the probability of conflicts. This finding, however, does not mean that ethnic diversity has no role in the explanation of civil wars. In fact ethnic polarization is a significant explanatory variable for the incidence of civil wars if we use the Q index of polarization. This result is robust to the use of other proxies for ethnic heterogeneity, alternative sources of data, regional dummies and the use of a single cross section of data. Therefore it seems that the weak explanatory power of ethnic heterogeneity on the incidence of civil wars found by several recent studies is due to the use of an index of fractionalization instead of an index of polarization. In addition Montalvo and Reynal-Querol (2005) confirm that ethnolinguistic fractionalization has a direct negative effect on growth, probably due to its impact on the transmission of ideas. However they also find that an increase in ethnic polarization has an indirect negative effect on growth because it increases the incidence of civil wars and public consumption and reduces the rate of investment.

Appendix I: Definition of the variables

PRIOCW: Intermediate and war definition of armed conflict from PRIO. This is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths yearly and a minimum of 1,000 deaths over the course of the civil war. We only consider types 3 and 4 (internal armed conflicts).

PRIO1000: PRIO definition including armed conflicts that generate more than 1,000 deaths yearly (war definition following PRIO classification). We only consider types 3 and 4 (internal armed conflicts).

PRIO25: PRIO definition including armed conflicts that generate more than 25 deaths yearly (minor armed conflicts plus intermediate plus war following PRIO classification). We only consider types 3 and 4 (internal armed conflicts).

DSCW: Civil wars using the dataset of Doyle and Sambanis (2000). Their definition considers a conflict as a civil war if

- a) it caused more than 1,000 deaths,
- b) it challenged the sovereignty of an internationally recognized state,
- c) it occurred within the recognized boundary of that state,
- d) it involves the state as a principal combatant
- e) it included rebels with the ability to mount organized armed opposition to the state
- f) the parties were concerned with the prospects of living together in the same political unit after the end of the war.

This definition is nearly identical to the definition of Singer and Small (1994).

FLCW: The definition of civil war of Fearon and Laitin (2003) is a conflict that

- a) involves fighting between agents of (or claimants to) a state and organized, nonstate groups who sought either to take control of a government, to take power in a region, or to use violence to change government policies,
- b) the conflict killed at least 1000 over its course, with a yearly average of at least 100,
- c) at least 100 were killed on both sides (including civilians attacked by rebels). The last condition is intended to rule out massacres where there is no organization or effective opposition.

LGDP: Log of real GDP per capita of the initial period (1985 international prices) from the Penn World Tables 5.6. Updated with the data of the Global Development Network Growth Database (World Bank).

LNPOP: Log of the population at the beginning of the period from the Penn World Tables.5.6. Updated with the data of the Global Development Network Growth Database (World Bank).

PRIMEXP: Proportion of primary commodity exports of GDP. Primary commodity exports. Source: Collier and Hoeffler (2001).

MOUNTAINS: Percent Mountainous Terrain: This variable is based on work by geographer A.J Gerard for the World Bank's "Economics of Civil war, Crime, and Violence" project.

NONCONT: Noncontiguous state: Countries with territory holding at least 10,000 people and separated from the land area containing the capital city either by land or by 100 kilometers of water were coded as "noncontiguous." Source: Fearon and Laitin (2003)

DEMOCRACY: Democracy score: general openness of the political institutions (0=low; 10=high). Source: Polity IV dataset. We transform the score in a dummy variable that takes value 1 if the score is higher or equal to 4. This variable is very correlated with the variable Freedom of the Freedom House.

ETHFRAC: index of ethnolinguistic fractionalization calculated using the data of the World Christian Encyclopedia.

ETHPOL: index of ethnolinguistic polarization calculated using the data of the World Christian Encyclopedia.

ETHDOM: index of ethnic dominance. It takes value 1 if one ethnolinguistic group represents between 45% and 90% of the population. Source: WCE.

ETHLRG: proportion of the largest ethnic group. Source: WCE.

RELFRAC: index of religious fractionalization. Source: L'Etat des Religions dans le Monde and The Statesmen Yearbook.

RELPOL: index of religious polarization. Source: L'Etat des Religions dans le Monde and The Statesmen Yearbook.

References

- [1] Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and R. Wacziarg (2003) "Fractionalization." *Journal of Economic Growth*, 8, 155-194.
- [2] Atlas Narodov Mira (Atlas of the People of the World). Moscow: Glavnoe Upravlenie Geodezii i Kartografii, 1964. Bruck, S.I., and V.S. Apenchenko (eds.).
- [3] Barret, D. ed. (1982) *World Christian Encyclopedia*. Oxford University Press.
- [4] Caselli and Coleman (2002) "On the theory of ethnic Conflict." mimeo.
- [5] Collier, Paul (2001) "Implications of Ethnic Diversity." *Economic Policy* (April) .
- [6] Collier, P., and A. Hoeffler (2002) "Greed and Grievances in Civil Wars." Center for the Study of African Economies, WP2002-01.
- [7] _____ and _____ (1998) "On Economic Causes of Civil Wars." *Oxford Economic Papers* 50: 563-73.
- [8] Doyle, M. W., and N. Sambanis (2000) "International Peacebuilding: A Theoretical and Quantitative Analysis." *American Political Science Review* 94:4 (December).
- [9] Duclos, J., Esteban, J. and D. Ray (2004), "Polarization: concept, measurement and estimation." forthcoming in *Econometrica*.
- [10] Easterly, W., and Levine (1997) "Africa's growth tragedy: Policies and Ethnic divisions." *Quarterly Journal of Economics*.
- [11] *Encyclopedia Britannica* (2000). Chicago: Encyclopedia Britannica.
- [12] Esteban, J., and Ray (1994) "On the measurement of polarization." *Econometrica* 62(4): 819-851.
- [13] _____ and _____ (1999) "Conflict and Distribution." *Journal of Economic Theory* 87: 379-415.
- [14] Fearon, J. (2003), "Ethnic and cultural diversity by country," *Journal of Economic Growth*, 8, 195-222.

- [15] Fearon, J. and Laitin, D. (2003). "Ethnicity, Insurgency, and Civil War," *American Political Science Review* 97 (February)
- [16] Grossman, H. (2001), "The Creation of Effective Property Rights," *American Economic Review*, 91, 2, 347-352.
- [17] Horowitz, D. (1985). *Ethnic Groups in Conflict*. Berkeley: University of California Press.
- [18] Keefer, P. and S. Knack (2002) "Polarization, politics and property rights: links between inequality and growth." *Public Choice* 11: 127-154.
- [19] La Porta, R., Lopez de Silanes, F., Shleifer and R. Vishny (1999) "The Quality of Government." *Journal of Law, Economics and Organization* 15(1): 222-279.
- [20] *L'État des Religions*. Éditions La Découverte et Éditions du Cerf , Paris, 1987.
- [21] Mauro, P. (1995), "Corruption and Growth," *Quarterly Journal of Economics*, CX, 681-712.
- [22] Miguel, E., Satyanath, S. and E. Sergenti (2004), "Economic Shocks and Civil Conflicts: an Instrumental Variables Approach," forthcoming in the *Journal of Political Economy*.
- [23] Montalvo, J. G. and Reynal-Querol, M. (2000), "The Effect of Ethnic and Religious Conflict on Growth." <http://www.wcfia.harvard.edu/programs/prpes>.
- [24] _____ and _____ (2002), "Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth" UPF Working Paper 660.
- [25] _____ and _____ (2005), "Ethnic Diversity and Economic Development," *Journal of Development Economics*, 76, 293-323.
- [26] Reynal-Querol, M. (2002), "Ethnicity, Political Systems and Civil War," *Journal of Conflict Resolution*, 46 (1), 29-55.
- [27] Singer, J.D. and M. Small (1994) "Correlates of War Project: International and Civil War Data, 1816-1992." (ICPSR 9905). Ann Arbor, Mich.
- [28] Taylor, C., and M.C. Hudson (1972) *The World Handbook of Political and Social Indicators*, 2nd ed. (New Haven, CT: Yale University Press.
- [29] Vanhaven, T. (1999). "Domestic Ethnic Conflict and Ethnic Nepotism: A Comparative Analysis" *Journal Of Peace Research*, vol.36, no.1, pp. 55-73.

TABLE 1
Logit regressions for the incidence of civil wars.
Basic indicators of ethnolinguistic/religious heterogeneity.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-5.82 (2.06)	-6.26 (1.93)	-6.29 (2.01)	-5.27 (1.66)	-6.03 (1.85)	-6.89 (2.26)	-6.77 (1.94)	-7.47 (2.32)
LGDP	-0.28 (1.27)	-0.44 (1.99)	-0.42 (1.79)	-0.40 (1.44)	-0.32 (1.11)	-0.33 (1.13)	-0.37 (1.32)	-0.37 (1.33)
LPOP	0.34 (2.18)	0.41 (2.40)	0.40 (2.21)	0.40 (2.47)	0.39 (2.39)	0.43 (3.01)	0.40 (2.31)	0.43 (2.72)
PRMEXP	-0.90 (0.52)	-1.01 (0.54)	-1.07 (0.57)	-0.36 (0.21)	-0.56 (0.32)	-0.35 (0.21)	-1.21 (0.64)	-0.89 (0.48)
MOUNTAINS	0.00 (0.49)	0.00 (0.25)	-0.00 (0.19)	0.00 (0.36)	0.00 (0.41)	0.00 (0.29)	-0.00 (0.15)	-0.00 (0.16)
NONCONT	0.08 (0.13)	0.29 (0.49)	0.28 (0.48)	0.04 (0.07)	0.08 (0.13)	0.31 (0.49)	0.32 (0.52)	0.47 (0.79)
DEMOCRACY	0.07 (0.21)	0.03 (0.09)	0.03 (0.09)	0.10 (0.29)	0.10 (0.28)	0.01 (0.05)	0.03 (0.08)	-0.03 (0.09)
ETHFRAC	1.19 (1.89)		0.17 (0.19)					0.04 (0.05)
ETHPOL		2.37 (2.97)	2.28 (2.23)				2.27 (2.84)	2.09 (2.03)
RELFAC				0.37 (0.36)		-4.97 (1.65)		-4.45 (1.39)
RELPOL					0.73 (1.00)	3.90 (1.97)	0.44 (0.65)	3.29 (1.59)
Pseudo R ²	0.10	0.12	0.12	0.09	0.10	0.11	0.12	0.13
N	846	846	846	846	846	846	846	846

The sample includes 138 countries for the period 1960-99. The dependent variable is the incidence of civil wars following the definition of PRIO that includes intermediate and high intensity armed conflicts (PRIOCW). The method of estimation is logit. The absolute z-statistics in parenthesis are calculated using standard errors adjusted for clustering. Explanatory variables: LGDP: log of real GDP per capita in the initial year; LPOP: the log of the population at the beginning of the period; PRMEXP: primary exports (Collier and Hoeffler); MOUNTAINS: mountains; NONCONT: noncontiguous states; DEMOCRACY: degree of democracy (Polity IV); ETHFRAC: ethnic fractionalization (Source: WCE); ETHPOL: ethnic polarization (Source: WCE); RELFRAC: Religious fractionalization (Source: ET); RELPOL: religious polarization (Source: ET).

TABLE 2
 Logit regressions for the incidence of civil wars.
 Robustness to alternative indicators of ethnolinguistic heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C	-6.29 (2.01)	-4.82 (1.59)	-6.37 (2.03)	-5.07 (1.74)	-6.22 (1.93)	-5.10 (1.70)	-6.41 (1.96)
LGDP	-0.42 (1.79)	-0.49 (2.35)	-0.42 (1.94)	-0.40 (1.85)	-0.43 (1.95)	-0.49 (2.15)	-0.41 (1.76)
LPOP	0.40 (2.21)	0.40 (2.46)	0.41 (2.43)	0.40 (2.40)	0.40 (2.29)	0.42 (2.64)	0.39 (2.32)
PRIMEXP	-1.07 (0.57)	-0.17 (0.10)	-1.11 (0.60)	1.19 (0.50)	-0.52 (0.18)	-0.20 (0.11)	-1.25 (0.69)
MOUNTAINS	-0.00 (0.19)	0.00 (0.03)	-0.00 (0.21)	0.00 (0.38)	-0.00 (0.22)	0.00 (0.11)	-0.00 (0.26)
NONCONT	0.28 (0.48)	0.22 (0.37)	0.26 (0.46)	0.03 (0.06)	0.28 (0.49)	0.18 (0.30)	0.28 (0.46)
DEMOCRACY	0.03 (0.09)	0.06 (0.18)	0.04 (0.11)	0.09 (0.25)	0.03 (0.10)	0.07 (0.22)	0.03 (0.08)
ETHPOL	2.28 (2.23)		2.54 (2.79)		2.35 (2.82)		2.91 (2.62)
ETHFRAC	0.17 (0.19)						
ETHDOM		0.44 (1.16)	-0.14 (0.34)				
ETHLRG*PRIMEXP				-2.92 (0.78)	-0.98 (0.21)		
LARMINOR						2.22 (1.32)	-1.36 (0.61)
N	846	846	846	846	846	846	846
Pseudo R ²	0.12	0.09	0.12	0.09	0.13	0.10	0.12

The sample includes 138 countries for the period 1960-99. The dependent variable is the incidence of civil wars following the definition of PRIO that includes intermediate and high intensity armed conflicts (PRIOCW). The method of estimation is logit. The absolute z-statistics in parenthesis are calculated using standard errors adjusted for clustering. Explanatory variables: LGDP: log of real GDP per capita in the initial year; LPOP: the log of the population at the beginning of the period; PRIMEXP: primary exports (Collier and Hoeffler); MOUNTAINS: mountains; NONCONT: noncontiguous states; DEMOCRACY: degree of democracy (Polity IV); ETHFRAC: ethnic fractionalization (Source: WCE); ETHPOL: ethnic polarization (Source: WCE); ETHDOM: ethnic dominance (Source: WCE); ETHLRG*PRIMEXP: largest ethnic group by primary exports; LARMINOR: size of the largest minority (Source: WCE).

TABLE 3
 Proportion of observations with civil wars (1960-99).
 Alternative sources of data and time periods

	Annual	Five years periods	All period (1960-99)
PRI01000	5.9%	10.1%	29.2%
PRI0CW	10.8%	14.4%	33.1%
PRI025	15.2%	22.2%	53.6%
Doyle-Sambanis (DSCW)	11.4%	15.3%	35.5%
Fearon-Laitin (FLCW)	13.0%	16.6%	34.8%

TABLE 4
Logit regressions for the incidence of civil wars.
Comparing alternative data on civil wars

Dependent variable:	DSCW	FLCW	PRIO1000	PRIO25	PRIOCW	DSCW	FLCW	PRIO1000	PRIO25
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	-1.84 (0.76)	-2.47 (0.89)	-4.32 (1.54)	-4.16 (1.61)	-7.29 (2.59)	-3.26 (1.46)	-4.25 (1.67)	-5.22 (2.02)	-5.69 (2.44)
LGDP	-0.86 (3.89)	-1.19 (4.65)	-0.62 (2.78)	-0.62 (3.28)	-0.45 (1.93)	-0.82 (3.63)	-1.08 (4.52)	-0.66 (3.22)	-0.60 (3.14)
LPOP	0.35 (2.36)	0.51 (2.68)	0.30 (1.94)	0.39 (2.35)	0.47 (2.77)	0.44 (3.04)	0.58 (3.22)	0.38 (2.56)	0.48 (3.05)
PRIMEXP	-0.91 (0.54)	-0.55 (0.37)	-0.01 (0.01)	0.24 (0.20)	-0.87 (0.49)	-0.68 (0.45)	-0.37 (0.27)	0.04 (0.03)	0.27 (0.26)
MOUNTAINS	-0.00 (0.80)	0.00 (1.04)	0.00 (0.54)	0.00 (0.22)	-0.00 (0.20)	-0.00 (0.63)	0.00 (1.03)	0.00 (0.50)	0.00 (0.18)
NONCONT	0.25 (0.45)	0.90 (1.59)	0.30 (0.50)	0.69 (1.59)	0.16 (0.29)	0.18 (0.35)	0.82 (1.64)	0.13 (0.24)	0.54 (1.37)
DEMOCRACY	0.43 (1.25)	0.53 (1.65)	0.03 (0.09)	0.18 (0.68)					
ETHFRAC	-0.52 (0.65)	0.01 (0.01)	0.57 (0.62)	-0.06 (0.09)	0.18 (0.20)	-0.73 (0.92)	-0.14 (0.15)	0.57 (0.63)	-0.17 (0.23)
ETHPOL	2.31 (2.76)	1.95 (1.97)	2.33 (2.16)	2.05 (2.41)	2.31 (2.23)	2.32 (2.74)	2.11 (2.02)	2.35 (2.12)	2.13 (2.48)
Pseudo R ²	0.13	0.25	0.13	0.14	0.15	0.16	0.26	0.16	0.17
N	846	846	846	846	990	990	990	990	990

The sample includes 138 countries for the period 1960-99. The method of estimation is logit. The absolute z-statistics in parenthesis are calculated using standard errors adjusted for clustering. The endogenous variables are: PRIOCW: intermediate and high intensity armed conflict (PRIO); DSCW: Doyle and Sambanis (2000) definition of civil war; FLCW: Fearon and Laitin (2003) definition of civil war; PRIO1000: armed conflict generating more than 1,000 deaths yearly (PRIO); PRIO25: armed conflict generating more than 25 deaths yearly (PRIO). Explanatory variables: LGDPC: log of real GDP per capita in the initial year; LPOP: the log of the population at the beginning of the period; PRIMEXP: primary exports (Collier and Hoeffler); MOUNTAINS: mountains; NONCONT: noncontiguous states; DEMOCRACY: degree of democracy (Polity IV); ETHFRAC: ethnic fractionalization (Source: WCE); ETHPOL: ethnic polarization (Source: WCE);

Table 5
Robustness of the results to the inclusion of regional dummies
and the elimination of countries in specific regions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
C	-6.17 (1.74)	-6.07 (1.68)	-4.59 (1.19)	-4.23 (1.06)	-7.47 (1.93)	-7.60 (2.01)	-5.50 (1.69)	-5.39 (1.66)
LGDPC	-0.43 (1.84)	-0.41 (1.71)	-0.45 (1.78)	-0.43 (1.57)	-0.37 (1.59)	-0.34 (1.43)	-0.40 (1.72)	-0.43 (1.78)
LPOP	0.40 (2.34)	0.38 (2.08)	0.33 (1.75)	0.29 (1.37)	0.46 (2.35)	0.45 (2.24)	0.33 (2.01)	0.35 (1.93)
PRMEXP	-1.08 (0.56)	-1.15 (0.59)	-0.94 (0.45)	-1.14 (0.55)	-0.55 (0.30)	-0.60 (0.33)	-0.92 (0.43)	-0.81 (0.37)
MOUNTAINS	-0.00 (0.15)	-0.00 (0.11)	-0.01 (0.64)	-0.01 (0.65)	-0.00 (0.57)	-0.00 (0.50)	-0.00 (0.02)	-0.00 (0.09)
NONCONT	0.11 (0.17)	0.09 (0.14)	0.12 (0.20)	0.05 (0.09)	-0.06 (0.09)	-0.07 (0.11)	0.12 (0.15)	0.14 (0.17)
DEMOCRACY	0.09 (0.25)	0.09 (0.26)	0.01 (0.02)	-0.00 (0.02)	0.06 (0.16)	0.06 (0.16)	0.16 (0.43)	0.16 (0.44)
ETHPOL	2.48 (3.10)	2.35 (2.23)	2.40 (2.60)	1.98 (1.32)	2.20 (2.80)	2.12 (2.15)	2.46 (2.92)	2.61 (2.43)
ETHFRAC		0.26 (0.29)		0.63 (0.42)		0.19 (0.21)		-0.30 (0.32)
Reg. Dummies	Yes	Yes	No	No	No	No	No	No
Eliminated region	None	None	SAfrica	SAfrica	Laam	Laam	Asiae	Asiae
N	846	846	580	580	678	678	781	781
Pseudo R ²	0.13	0.13	0.12	0.12	0.12	0.12	0.11	0.11

The sample includes 138 countries for the period 1960-99. The dependent variable is the incidence of civil wars following the definition of PRIO that includes intermediate and high intensity armed conflicts (PRIOCW). The method of estimation is logit. The absolute z-statistics in parenthesis are calculated using standard errors adjusted for clustering. Explanatory variables: LGDPC: log of real GDP per capita in the initial year; LPOP: the log of the population at the beginning of the period; PRMEXP: primary exports (Collier and Hoeffler); MOUNTAINS: mountains; NONCONT: noncontiguous states; DEMOCRACY: degree of democracy (Polity IV); ETHFRAC: ethnic fractionalization (Source: WCE); ETHPOL: ethnic polarization (Source: WCE). Regional dummies: SAFRICA: Sub-Saharan Africa; LAAM: Latin America; ASIAE: Asia.

TABLE 6
Robustness regressions
Cross section logit regressions for the incidence of civil wars.

Endogenous Variable	PRIOCW (1)	PRIOCW (2)	DSCW (3)	DSCW (4)	FLCW (5)	FLCW (6)
C	-1.19 (0.37)	-1.04 (0.26)	2.23 (0.68)	5.84 (1.31)	4.91 (1.41)	8.37 (1.76)
LGDP	-0.63 (2.03)	-0.63 (1.61)	-1.01 (2.95)	-1.40 (2.94)	-1.23 (3.34)	-1.64 (3.20)
LPOP	0.35 (1.61)	0.34 (1.51)	0.37 (1.57)	0.34 (1.33)	0.25 (1.06)	0.24 (0.94)
PRIMEXP	1.19 (0.55)	1.29 (0.57)	-0.34 (0.15)	0.23 (0.10)	-0.20 (0.08)	0.90 (0.36)
MOUNTAINS	-0.00 (0.58)	-0.01 (0.45)	0.00 (0.42)	-0.01 (0.83)	0.01 (0.70)	0.01 (0.46)
NONCONT	0.02 (0.03)	0.05 (0.06)	-0.53 (0.59)	-0.56 (0.57)	-0.17 (0.19)	0.14 (0.14)
DEMOCRACY	0.32 (0.55)	0.35 (0.59)	-0.02 (0.04)	-0.09 (0.14)	-0.19 (0.30)	-0.03 (0.06)
ETHPOL	3.35 (2.46)	3.42 (2.48)	3.26 (2.37)	3.53 (2.44)	2.95 (2.15)	3.27 (2.26)
Reg. dummies	No	Yes	No	Yes	No	Yes
N	90	90	90	90	90	90
Pseudo R ²	0.17	0.17	0.26	0.27	0.29	0.31

The sample includes 138 countries for the period 1960-99. The method of estimation is logit. The endogenous variables are: PRIOCW: intermediate and high intensity definition of armed conflict of PRIO; DSCW: Doyle and Sambanis (2000) definition of civil war; FLCW: Fearon and Laitin (2003) definition of civil war. Explanatory variables: LGDP: log of real GDP per capita in the initial year; LPOP: the log of the population at the beginning of the period; PRIMEXP: primary exports (Collier and Hoeffler); MOUNTAINS: mountains; NONCONT: noncontiguous states; DEMOCRACY: degree of democracy (Polity IV); ETHPOL: ethnic polarization (Source: WCE). Regional dummies: SAFRICA: Sub-Saharan Africa; LAAM: Latin America; ASIAE: Asia.

Appendix 2. Ethnic polarization and fractionalization
(Source: WCE)

Country	ETHPOL	ETHFRAG
Afghanistan	0.786	0.603
Algeria	0.514	0.299
Angola	0.572	0.805
Argentina	0.579	0.408
Australia	0.492	0.315
Austria	0.240	0.128
Bahamas, The	0.705	0.441
Bahrain	0.569	0.383
Bangladesh	0.132	0.068
Barbados	0.366	0.199
Belgium	0.871	0.544
Benin	0.436	0.868
Bolivia	0.767	0.708
Botswana	0.650	0.485
Brazil	0.773	0.644
Burundi	0.512	0.286
Cameroon	0.576	0.817
Canada	0.672	0.767
Cape Verde	0.822	0.435
Central African Republic	0.578	0.787
Chad	0.665	0.768
Chile	0.723	0.432
China	0.661	0.599
Colombia	0.789	0.675
Comoros	0.127	0.061
Congo, Dem. Rep.	0.586	0.799
Congo, Rep.	0.674	0.721
Costa Rica	0.420	0.241
Cote d'Ivoire	0.432	0.874
Cyprus	0.652	0.357
Denmark	0.097	0.049
Dominica	0.370	0.202
Dominican Republic	0.725	0.460
Ecuador	0.837	0.657
Egypt, Arab Rep.	0.427	0.247
El Salvador	0.279	0.145
Ethiopia	0.778	0.695
Fiji	0.930	0.559
Finland	0.294	0.148
France	0.294	0.147
Gabon	0.519	0.834
Gambia, The	0.689	0.728
Germany	0.227	0.123
Ghana	0.661	0.731
Greece	0.186	0.099
Grenada	0.945	0.542
Guatemala	0.955	0.520
Guinea	0.843	0.649

Guinea-Bissau	0.532	0.829
Guyana	0.813	0.634
Haiti	0.207	0.104
Honduras	0.430	0.254
Hong Kong, China	0.066	0.034
Hungary	0.308	0.167
Iceland	0.055	0.028
India	0.348	0.901
Indonesia	0.529	0.793
Iran, Islamic Rep.	0.598	0.756
Iraq	0.665	0.390
Ireland	0.141	0.072
Israel	0.548	0.286
Italy	0.154	0.080
Jamaica	0.600	0.354
Japan	0.067	0.034
Jordan	0.982	0.515
Kenya	0.381	0.890
Korea, Rep.	0.028	0.014
Kuwait	0.980	0.513
Lesotho	0.343	0.185
Liberia	0.390	0.890
Luxembourg	0.596	0.298
Madagascar	0.017	0.050
Malawi	0.736	0.684
Malaysia	0.762	0.695
Mali	0.420	0.862
Malta	0.167	0.083
Mauritania	0.536	0.334
Mauritius	0.803	0.482
Mexico	0.654	0.576
Morocco	0.897	0.475
Mozambique	0.499	0.838
Myanmar	0.650	0.474
Nepal	0.652	0.682
Netherlands	0.214	0.113
New Zealand	0.366	0.196
Nicaragua	0.681	0.496
Niger	0.698	0.718
Nigeria	0.404	0.885
Norway	0.090	0.045
Oman	0.408	0.239
Pakistan	0.698	0.608
Panama	0.586	0.048
Papua New Guinea	0.669	0.354
Paraguay	0.310	0.174
Peru	0.817	0.658
Philippines	0.497	0.843
Poland	0.099	0.051
Portugal	0.020	0.010
Rwanda	0.401	0.221
Samoa	0.388	0.199

Saudi Arabia	0.114	0.059
Senegal	0.560	0.809
Seychelles	0.160	0.084
Sierra Leone	0.600	0.793
Singapore	0.666	0.421
Solomon Islands	0.258	0.139
Somalia	0.679	0.385
South Africa	0.718	0.469
Spain	0.693	0.436
Sri Lanka	0.749	0.452
St. Lucia	0.958	0.540
St. Vincent and the Grenadines	0.773	0.527
Sudan	0.699	0.711
Suriname	0.734	0.729
Swaziland	0.318	0.178
Sweden	0.337	0.189
Switzerland	0.724	0.560
Syrian Arab Republic	0.373	0.207
Taiwan, China	0.685	0.363
Tanzania	0.271	0.959
Thailand	0.582	0.361
Togo	0.673	0.732
Tonga	0.066	0.034
Trinidad and Tobago	0.842	0.662
Tunisia	0.167	0.087
Turkey	0.342	0.185
Uganda	0.279	0.932
United Arab Emirates	0.640	0.320
United Kingdom	0.571	0.373
United States	0.691	0.583
Uruguay	0.426	0.260
Vanuatu	0.285	0.155
Venezuela	0.758	0.539
Yemen, Rep.	0.063	0.032
Yugoslavia, FR (Serb./Mont.)	0.599	0.782
Zambia	0.606	0.787
Zimbabwe	0.698	0.534