Università
Ca'Foscari
Venezia

Scuola Dottorale di Ateneo
Graduate School
Dottorato di Ricerca in Economia
Ciclo XXIV
Anno di Discussione 2013

Ethnic Distribution, Effective Power and Ethnic Conflict

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: SECS-P/03

Tesi di dottorato di MATIJA KOVACIC , Matricola 955559

Coordinatore del Dottorato
Prof. Michele Bernasconi

Tutore del dottorando
Prof. Claudio Zoli

Co-tutore del dottorando
Prof. Sergio Currarini

The undersigned MATIJA KOVACIC, in his quality of doctoral candidate for a Ph.D. degree in Economics granted by the Università Ca' Foscari Venezia attests that the research exposed in this dissertation is original and that it has not been and it will not be used to pursue or attain any other academic degree of any level at any other academic institution, be it foreign or Italian.

Abstract

This thesis is investigating both in theoretical and empirical terms the relationship between the features of ethnic distribution and the probability of conflict. The thesis is composed of four chapters. The first chapter is the introduction to the thesis in which I briefly summarize the main problems related to the existing literature on ethnic diversity and conflict and I present an overview of the measures of ethnic diversity commonly used in the literature.

In the second chapter I propose a theoretical model that specifies the potential of conflict in a society as a function of the population distribution across ethnic groups. I axiomatically derive a general parametric class of indices of conflict potential that combines the groups' effective power and the between-groups interaction. The effective power of a group is a function of a group's relative size but it also depends on the relative sizes of all the other groups in the population. The interaction component, on the other hand, is given by the probability of interaction between the members of one group with those of other groups. I show that for certain parameter values the index reduces to the existing indices of ethnic diversity, while in general the indices combine in a non-linear way three different aspects of ethnic diversity, namely the fractionalization, the polarization and the ethnic dominance. The results of the model share some common features with the literature on conflicts in contests and the literature on voting power indices. In particular, the power component of the extreme element of the class of indices is intuitively related to the definition of voting power in a simple majority game. In that particular case, the value of the effective power is given by the relative Penrose-Banzhaf index of voting power calculated over the shares of populations associated to each ethnic group.

In the third chapter I investigate empirically the role of ethnic diversity in the explanation of the ethnic conflict outbreak. The empirical performance of the indices of conflict potential developed in the second chapter is tested against the existing distributional indices of ethnic diversity within the context of the commonly used logistic model that focuses on the onset of ethnic conflicts in a time range from 1946 to 2005. Together with the set of the explanatory variables for structural conditions and country characteristics, I take advantage of the recent "Ethnic Power Relations" data set which includes additional information on the political exclusion and competition along ethnic lines and it offers the best coding for ethnic wars. The results obtained suggest that the indices of conflict potential outperform the existing indices of ethnic diversity in the explanation of ethnic conflict onset. This evidence is robust to the inclusion of a larger set of regressors, time and regional controls as well as to several other estimation techniques.

The fourth chapter explores empirically the determinants of conflict duration with a particular attention to the potential role of ethnic diversity together with ethnic politics and competition dynamics. The first part of the chapter presents an overview of the existing literature on conflict duration, the main data cources and the related econometric issues. The second part of the chapter consists in a non-parametric and a parametric survival analysis of the duration of ethnic conflict where we address in detail the issues of non-proportionality of the hazard function, the unmeasured heterogeneity and the presence of "repeated events". The results suggest that there is a statistically significant and robust association between ethnic distribution and conflict duration, together with other commonly used explanatory variables in the literature on conflict duration.

Acknowledgements

This dissertation represents the end of a four years long journey, but also the beginning of another, longer and hopefully even more successful period of my life.

I am indepted to many people for making these four years an unforgettable experience. First and foremost, an immense thank goes to my parents and to their infinite support and help during all these years. Without your support and my decisiveness that you have always strongly encouraged, all this would have never been possible.

I am deeply grateful to my advisor Claudio Zoli. I have been very privileged to get to know and to collaborate with you over the last two years. I learned a lot from you about accademic life and research, how to tackle new problems and how to develop a critical way of thinking. Working with you has been a great pleasure to me. You have always been patient and encouraging and your hard work, enthusiasm and especially your approach to reserach and to students set an example.

Furthermore, an immense thank goes to Agar Brugiavini, to her infinite disponibility, politeness and comprehension. Your patience and your ability to overcome any obstacle that I met during all these years is unique. Thank you for encouraging me in my research.

I also thank my co-advisor Sergio Currarini for insightful comments and help. I am also grateful to Giacomo Pasini who helped me much with the empirical issues. Thank you also to Noemi Pace, Enrica Croda, Giuliano Petrovich, Luciano Pezzolo and to all those who trusted in me and who provided me any kind of support and help.

My long and sometimes endless days in San Giobbe would have been even longer if *They* were not there, my great colleagues and friends: Breno, Aks and Caterina. We were always helping each other in good and bad times. Thank you for your support and collaboration. A big thank goes to Breno for his confidence and his Brasilian jokes that sometimes only he was able to understand fully. A special thank goes to Aks, for his patience and indisputable willingness to help. Thank you also to Caterina for useful suggestions and time dedicated both to accademic and non accademic issues. I could never conclude this list without mentioning Laura, Dragana, Anna, Giulia, Luis and Ludmila. It was great to spend so much time with you.

An immense thank goes to my unique friends in Croatia. You have always made everything to make me feel less distant from you and from my adorable country. Only you can know how much I appretiate this. Don't bother about the order of these acknowledgements, you know very well what is your contribution. *Ne postoje rijeci koje bi mogle izreci koliko sam vam zahvalan, dragi moji.*

Venice, 30.01.2013

Matija Kovacic

To my loving parents.

Contents

List of Tables

List of Figures

Chapter 1

Introduction to the Thesis: Ethnic Diversity and Conflict

1.1 Ethnic Diversity, Effective Power and Conflict

Internal conflicts represent one of the main impediments for economic development. The negative consequences of conflicts are not only related to the destruction of infrastructure or human lives but they also undermine democratic political institutions, generate an uncertain environment for future investments and growth, and exacerbate the conditions that favor further insurgency. According to some scholars, the negative economic consequences of internal conflicts are so important that can be considered as one of the explanatory factors in the growing income gap between the world's richest and poorest countries.

Over the past decade many economists and other social scientists have worked to better understand the causes and the economic and social legacies of internal warfare. The literature on conflict can be classified into three main categories, each of which models conflict in a different way and emphasizes the role of different factors in the explanation of one of the three main aspects of conflicts, namely their onset, incidence or duration. Some scholars emphasize the role of structural conditions and country characteristics, others emphasize the role of the state as not an ethnically neutral institution but as an active agent of political exclusion along ethnic lines that may generate tensions between excluded groups and the state. Several other scholars put more emphasis on the role of ethnic or religious diversity and argue that societies that are more ethnically heterogeneous are also more likely to experience internal conflict.

Although ethnic demands and grievances play a prominent role in most recent conflicts, there is still no broad consensus of whether or not ethnic diversity is an important correlate to conflict. The literature on ethnic diversity and conflict relies on a variey of indicators capturing different features of ethnic distribution. From a distributional point of view, there are three main manifestations of ethnic diversity, namely the fractionalization, the polarization and the ethnic dominance. The indicators used to capture these phenomena rank different population distributions according to their conflict potential. For instance, the fractionalization index assumes an increasing relationship between ethnic fragmentation within a country and its conflict vulnerability. Ethnic polarization, on the other hand assumes that highly fragmented and highly homogeneous societies are less likely to experience conflict with respect to those societies composed by few prominent ethnic groups. Finally, ethnic dominance emphasizes the role of the numerical predominance of one ethnic group over other(s). The empirical evidence on the association between ethnic diversity and conflict, however, is very heterogeneous. That is, there is still no broad consensus on which distributional aspect of ethnic diversity is associated with conflict. Using the fractionalization index, several scholars find a negative correlation between ethnic fragmentation and conflict, others claim that ethnic fragmentation and conflict are positively correlated, while others again argue that there is no relationship at all between the two. Applying the polarization index, on the other hand, some scholars find a positive and statistically robust evidence between the incidence of conflict and the degree of ethnic polarization, while others apply the same concept to the explanation of conflict onset and find no statistically significant relationship between ethnic polarization and the probability of conflict outbreak. Several other scholars show that once we take into consideration the degree of political exclusion and competition along ethnic lines, distributional aspects of ethnicity (fractionalization and polarization) have no effect on the likelihood of ethnic and civil conflicts in general. Finally, some empirical studies find a positive and robust evidence of ethnic dominance and dismiss the role of ethnic heterogeneity.

The variation in results can be due to several factors, some of them already addressed in the literature. The way in which we define whether a country is experiencing a conflict in a given year and the marker that we use to code ethnic groups may have a significant impact on results. Moreover, there are three possible aspects of conflict, namely the onset (whether a conflict starts in a given period), the incidence (whether a conflict occurs in a given period in a given country), and duration (whether a conflict continues, having already occured in the previous period). Although these are derivatives of the same phenomenon, they should be understood and treated as distinct concepts. In this thesis, however, I propose another alternative explanation for the variation in results. Since each of the three different aspects of ethnic diversity is important under certain model specifications or with the alternative data sets used in the empirical analysis, it may be that all of them are actually important for conflict but their relative importance depends on the features of the population distribution across ethnic groups. Since all distributional indices of ethnic diversity combine the effects of groups' power and across - group interaction, under some population distributions the interaction component may dominate the power component and viceversa. In other words, for some population distributions, the relevant aspect of ethnicity could be the fractionalization, for others this could be the dominance or a combination between the two. Since the relative strenght of power and interaction may depend on the particular features of ethnic constellations within a society, a natural way to approach this kind of problem is to construct an index that weights different aspects of ethnicity according to the characteristics of the underlying population distribution across groups.

The groups' effective power measures the ability of a group to translate the alienation into effective voicing, protest or any other type of collective action. The index of ethnic polarization assumes that the effective power of a group depends only on its relative population size while the index of ethnic fractionalization assumes that effective power is constant across groups and it is normalized to one. The interaction component, on the other hand, is given by the probability that two individuals randomly selected from the population belong to different ethnic groups. The potential alienation that comes from the interaction becomes effective once it is materialized in the form of a protest or rebellion. How efficient groups are to do this, depends on their relative size, namely bigger is the group higher is its effective power. This interpretation of power and interaction is embedded in the basic model of social antagonism proposed by Joan Maria Esteban and Debraj Ray in 1994.

Here I argue that the effective power of a group is *not* necessarily proportional to its relative population size. Although one of the main components of the existing indices of ethnic diversity is the interaction, in those models there is no interaction at all between groups. Groups are assumed to be unitary actors. This means that here is no interaction between them. There is no doubt that this is not a realistic assumption. Some could argue that introducing another strategic component into the structure of the existing models would make them intractable and too complicated. This, however, is not true. Relaxing the unitary actor assumption could expand the range of rational explanations for conflict, and one of these is the presence of strategic behavior between groups. A more realistic interpretation of the groups power is to model it as the *ability* of a group to influence the outcome of any kind of interaction that can occur between different ethnic groups in a society. This intuition share some common features with the logic behind the concept of *voting power*. The term voting power refers to an index that captures the power of a voter to influence the outcome of a voting process. Higher power means higher number of voting configurations in which an agent can change the outcome of voting by changing his or her vote from "yes" to "no" and *viceversa*. In his famous critique of the practice of assigning voting weights proportional to the numbers of citizens in different legislative bodies as a means of implementing the "one man, one vote" requirement, Banzhaf (1965) proves that "voting power is not (necessarily) proportional to the number of votes a legislator may cast", and that "the number of votes is not even a rough measure of the voting power of the individual legislator". Voting power, hence, in contrast to the number of votes an actor possesses, is the *ability* of an actor to influence the outcome of voting in a collectivity. The ability of being *pivotal* depends on the number of possibilities or coalitions in which a voter is in the critical position of being able to change the voting outcome by changing his vote.

Even in the absence of strict ethnic voting, the logic behind the voting power indices fits quite well in the context of ethnic groups relations. Strategic *coalitions* between groups can be observed in almost all conflicts, no matter how strong is the perception of the antagonism between them. Groups may simply collaborate with the "adversary" if they find it profitable. A coalition may shift during the time, *i.e.* two groups that previously were on the opposite sides may decide to join together if the conditions of the environment or the relative strenght have changed during the time. So, the logic of strategic behaviour is not related only to ethnic voting or similar simple political "games" but also to many other everyday situations that occur in ethnically divided societies, whether they are in conflict or not.

In order to address all these issues, I propose a general parametric class of indices of conflict potential that combines the groups' effective power and the between-groups interaction. Differenty from the existing literature, the effective power of a group is defined as a function of a group's relative size but it also depends on the relative sizes of all the other groups in the population. The interaction component, on the other hand, is given by the probability of interaction between the members of one group with those of other groups. Each group is *not* treated as a unitary actor but it is assumed that groups can either act individually or form alliances in order to exploit increasing returns to coalition formation. However, all the coalitions are assumed to be equally probable, that is there is no particular endogenous mechanism of coalition formation embedded in the model. I show that under certain parameter values the indices of conflict potential reduce to the existing indices of ethnic diversity commonly used in the literature, while in general the indices combine in a non-linear way three different aspects of ethnic diversity, namely the fractionalization, the polarization and the ethnic dominance. The indices share some common features with the literature on voting power indices and the literature on conflicts in contests. Indeed, the groups' effective power in the extreme case of the index is given by their relative Penrose-Banzhaf index of voting power in a simple majority game. With the effective power defined in this way, the index attributes to the power and the interaction component a different weight according to the characteristics of the underlying population distribution. For some distributions, the interaction dominates power while for the others the power component dominates the interaction component.

The difference between the indices of conflict potential and the existing indices of ethnic diversity is not only theoretical but also actual. The empirical performance of the indices of conflict potential is tested against the existing distributional indices of ethnic diversity within the context of the commonly used logistic model that focuses on the onset of ethnic conflicts in a time range from 1946 to 2005. The index based on the relative Penrose-Banzhaf voting power index outperforms the other indices of ethnic diversity. This empirical evidence is robust to the inclusion of an additional set of regressors and to alternative model specifications. The results show that the potential of conflict is given as the product of the effects of power and interaction which in turn depends on the features of the population distribution across groups. Contrary to many other scholars, the empirical results provided in this thesis show that the aspects of ethnic distribution *does* matter in the explanation of ethnic conflict but only when properly combined in a single measure of ethnic diversity.

This thesis is structured as follows. In the remaining part of the introduction, I present an overview of the measures of ethnic diversity commonly used in the literature on ethnic diversity and conflict. Chapter 2 introduces a theoretical model that specifies the potential of conflict in a society as a function of the population distribution across ethnic groups together with the axiomatic derivation of a general parametric class of indices of conflict potential that combines the groups' effective power and the between-groups interaction. Chapter 3 offers a detaled empirical anlysis of the correlates of ethnic conflict onset and tests the empirical performance of the indices of conflict potential against the existing distributional indices of ethnic diversity. Finally, Chapter 4 explores empirically the determinants of ethnic conflict duration with a particular attention to the potential role of ethnic diversity together with ethnic politics and competition dynamics.

1.2 Distributional Indices of Ethnic Diversity

The quantitatively oriented empirical research on ethnic diversity and conflict has focused on the question of whether and what particular constellations of ethnic diversity are more conflict-prone. From a distributional point of view we can distinguish between three basic manifestations of ethnic diversity: ethnic fractionalization (or fragmentation), ethnic polarization and ethnic dominance. Each of these concepts is based on an a priori assessment on what feature of population distribution across ethnic groups is particularly conflict prone. The indices that we use to capture these phenomena rank ethnic distributions according to their conflict potential. These measures are all uni-dimensional in the sense that they have only one type of input: population proportion (or the relative size) of each group and do not incorporate any other information on inter-group disparity, such as political representation asymmetries or economic, wealth or social disparity which may be an important generator of the alienation beetween groups[1]. Moreover, groups are treated as unitary actors and there is no interaction between them.

The index of ethnic fractionalization is the discrete version of the well known Gini's inequality index while the index of discrete ethnic polarization is the discrete version of the Esteban and Ray's (1994) income polarization index. Although they have a similar structure, the two indices differ in the treatment of group sizes. The former does not take into account the effect of group size on the sense of identity, while the latter

[1]There are, however, several indices of ethnic diversity that incorporate information on ethnic diversity with other information on between-group dissimilarity. One of these is the so-called Bossert's Generalized Fractionalization Index. This index matches each individual within the society to each other individual and assign them a similarity rank which ranges between 0 and 1. Even if it was initially calculated at individual's level, the index can be modified to take into account groups of different sizes characterized by perfect withing group homogeneity. Cederman and Girardin (2007), on the other hand, propose a specific ethnic diversity index that combines distributional aspects of ethnic diversity with political information. More precisely, their index takes explicitly into account which group controls the state and the value of the index (and hence the likelihood of conflict), is higher smaller is the group that controls the state relative to the size of other groups excluded from power.

assumes positive returns to group size and "transforms" the group size into effective voicing power through a coefficient which is commonly known as the "polarization sensitivity" parameter. The positive spillovers from group size to effective voicing imply that the polarization rises when within-group dispersion is reduced, which runs directly *against* the ordering over distributions generated by second order stochastic dominance. The notion of polarization is, hence, a distinct concept from that of inequality in as much as we require inequality measures be consistent with Lorenz curve ordering (Esteban and Ray, 2010). Ethnic fractionalization is monotonically increasing in the number of equally sized groups while discrete ethnic polarization attains its maximum at symmetric bimodal distribution, *i.e.* in the presence of two groups with equal population size.

Collier and Hoeffler (2001, 2004) and Schneider and Wiesehomeier (2008) differentiate societies that are highly fractionalized from those with a dominant group using a dummy variable which is 1 if the largest group represents 45-90% and 60-90% of the overall population respectively[2]. They argue that dominance abruptly becomes a problem once the group exceeds 50% of the population (or electorate). To the best of our knowledge, there is no distributional index of ethnic diversity that is able to take explicitly into account the presence of ethnic dominance. The index of discrete ethnic polarization relies on the intuition that the most conflict prone situation is the one in which there is a small majority that faces a substantial minority. This implies that the index should rank a distribution in which one ethnic group possesses scarsely more than 50% of the entire population higher (according to conflict potential) than a distribution in which the biggest group has "only" 49% of the population. The

[2]Collier (2000) test different size ranges for the largest ethnic group and show that the level of significance and the size of the coefficient reach a maximum when dominance is defined on the range 45-90% of the population.

index of discrete ethnic polarization accomplishes this task in the case of two groups, but it is unable to do so for distributions with more than two groups [3].

In what follows, we briefly present the measures of ethnic fractionalization and discrete ethnic polarization commonly used in the quantitatively oriented empirical research on ethnic diversity and conflict.

a) ***Gini's Heterogeneity Index*** $\mathbf{G^H}$ ***or Fractionalization Index F***

In an article of 1912, Corrado Gini works out several formulae aiming at establishing suitable indices of variability for qualitative phenomena (or categorical variables), almost analogous to the ones well known for quantitative variables. One such index is directly derived from the variability measure most studied and applied by Gini, namely the *Gini's mean difference of order one* (MD). Given observations x_i of a quantitative variable X, and respective relative frequencies π_i, the mean difference is defined as the arithmetic mean of all possibile differences:

$$MD(\Pi) = \sum_i \sum_{j \neq i} \pi_i \pi_j |x_i - x_j| \tag{1.1}$$

If we put $|x_i - x_j| \equiv D_{ij}$ and interpret it as the distance between the categories i and j, the index assumes the following form:

$$MD(\Pi) = \sum_i \sum_{j \neq i} \pi_i \pi_j D_{ij} \tag{1.2}$$

[3] Suppose, for instance, that we want to compare the following distributions according to their conflict potential: $\Pi_1 = (0.51, 0.29.0.2)$ and $\Pi_2 = (0.49, 0.3, 0.21)$. The RQ index of discrete ethnic polarization assigns to the first distribution a value of 0.87664 and to the second 0.88116. According to the RQ index, hence, both distributions are almost equally conflict prone.

In order to obtain the heterogeneity index G^H, Gini assumes the uniform distance 1 between *any* two different categories and $D_{ii} = 0$ for all i:

$$G^H(\Pi) = \sum_i \sum_{j \neq i} \pi_i \pi_j 1(i \neq j) = \sum_i \pi_i (1 - \pi_i) = 1 - \sum_i \pi_i^2. \tag{1.3}$$

This measure has been widely used in the literature on ethnic diversity and conflict to measure the degree of (ethnic or linguistic) fragmentation of a society. We refer to it as *fractionalization* or *fragmentation* index *FRAC*. It is defined as the probability that two randomly selected individuals belong to different ethnic categories (groups). Perhaps the most famous is the index of "ethno - linguistic fractionalization", also called ELF, constructed by Taylor and Hudson (1972) using the data of the *Atlas Narodov Mira*.

There are other two similar indices of variation in group membership, namely the Simpson index of interaction (I) and the Herfindahl-Hirschman concentration index (*HHI*). The Simpson index was introduced in 1949 by Edward H. Simpson to measure the degree of concentration when individuals are classified into types. The same index was rediscovered by Orris C. Herfindahl in 1950. The square root of the index had already been introduced in 1945 by the economist Albert O. Hirschman. As a result, the same measure is usually known as the Simpson index in sociology and ecology, and as the Herfindahl index or the Herfindahl-Hirschman index (HHI) in economics[4]:

$$I = HHI = \sum_i \pi_i^2 \tag{1.4}$$

[4]Although originally proposed by Gini, the index of ethnic or linguistic fractionalization is often reffered to as the Hirschman - Herfindahl index.

Although originally proposed by Gini, the fractionalization index is often reffered to as the Hirschman - Herfindahl index. The difference between *FRAC* and *I* or *HHI* is that the latter represents the probability that two randomly selected individual belong to the *same* category (group). The fractionalization index is obtained as $1-I$ or $1-HHI$:

$$FRAC = 1 - HHI = 1 - \sum_i \pi_i^2 = \sum_i \pi_i(1-\pi_i)$$

The index of ethnic fractionalization, hence, is the sum of the function $f(\pi) = \pi(1-\pi)$ evaluated at the different π_i. It is a strictly quasiconcave function of the population share vector (Figure 1). It follows that F satisfies the following properties:

1. *Any transfer of population from a group to a smaller one increases FRAC.*

2. *For a given number of groups, FRAC is maximized at the uniform distribution, i.e. when $\pi_i = 1/n$ for all i. Given n different groups, its maximum is $(n-1)/n$. Since $lim_{n\to\infty}(n-1)/n = 1$, we have the following property:*

3. *Over the set of uniform distributions, FRAC increases with the number of groups.*

4. *The split of any group with population share π into two new groups with population shares $\pi^{'}$ and $\pi^{''}$, $\pi^{'} + \pi^{''} = \pi$, increases FRAC.*

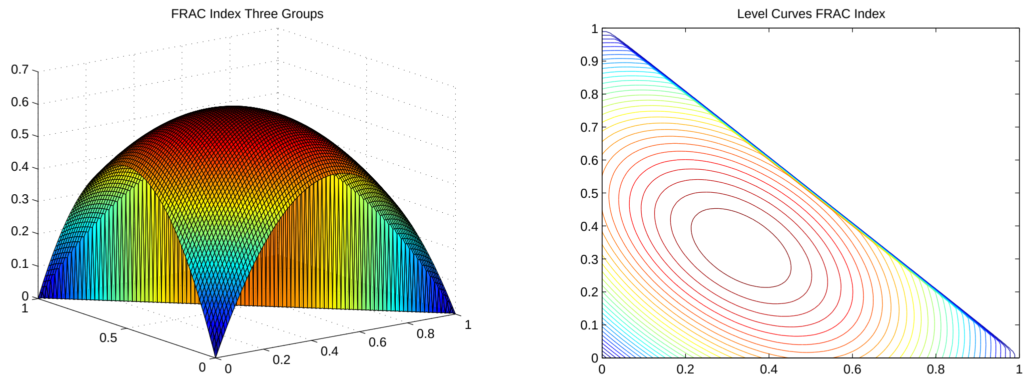


Figure 1.1: *FRAC* Index for the case of Three Groups

The fractionalization index assumes a linear relationship between increasing fragmentation and increasing conflict vulnerability, *i.e.* societies that are more fragmented from ethnical point of view, face a higher risk of (ethnic) conflict.

b) **Polarization Indices**

The concept of polarization is of recent use in Economics, but it has a long tradition in Political Science. However, its precise meaning remains somewhat ambigous. As Esteban and Schneider (2011) note: "Polarization is one of those ideas of which most social scientists believe to have an implicit understanding. As it is often the case, the overhasty acceptance of the seemingly obvious has, however, contributed to a considerable carelessness in the application of this concept". For instance, several authors have argued theoretically in terms of polarization but used as an empirical proxy the index of fractionalization. These, however, are two different and on occasion conflicting concepts. In a very broad sense, the well known income (or wealth) polarization measures aim to capture the distance between clustered groups (income classes) in a distribution. There are two main families of (income) polarization measures. One family of measures describes the clustering over distributions with arbitrary number of income classes while the other measures the clustering over distribution with only two income classes (with median income as the divide). Within the first family we have Esteban and Ray's (1991), Duclos, Esteban and Ray's (2004) [DER] polarization indices for continuous distributions and Esteban and Ray's (1994) [ER] polarization index for discrete distributions. Measures of bipolarization, on the other hand, begin with Wolfson (1994, 1997), based on Foster and Wolfson (1992, 2009), and by Wand and Tsui (2000).

Esteban and Ray (1994) conceptualize polarization as the sum of inter - personal antagonisms which "results from the interplay of the sense of group identification (group size) and the sense of alienation with respect to other groups (intergroup distance)" (Esteban and Ray, 2008). The ER index can be written as:

$$ER(\Pi) = K \sum_{i} \sum_{i \neq j} \pi_i^{1+\alpha_{ER}} \pi_j D_{ij} \tag{1.5}$$

where D_{ij} is the distance betwen any two groups i and j, $i \neq j$, $\alpha_{ER} \in (0, 1.6]$ and $K > 0$ is a contant. The parameter α_{ER} captures the importance of group identification or group power to translate the alienation into effective voicing and is regarded as a "polarization sensitivity parameter". Indeed, when α_{ER} is fixed to 0, and $K = 1$, the ER is precisely the Gini coefficient. As α_{ER} increases from zero, the divergence from inequality increases. When we assume $D_{ij} = D_i = 1$ for all $j \neq i$, the ER measure becomes:

$$ER(\Pi)_{\alpha_{ER}} = K \sum_{i} \pi_i^{1+\alpha_{ER}} (1 - \pi_i) \tag{1.6}$$

If we define groups using ethnic instead of income attributes, and assume $\alpha_{ER} = 0$ and $K = 1$, this measure is precisely the Gini's heterogeneity of fractionalization index. On the other hand, for $\alpha = 1$ and $K = 4$[5], the above measure is precisely the index of discrete ethnic polarization proposed by Montalvo and Reynal-Querol (2002, 2005), [RQ]:

$$RQ(\Pi) = 4\sum_{i} \pi_i^2(1 - \pi_i). \tag{1.7}$$

Montalvo and Reynal-Querol derive their results by imposing some "reasonable" axioms to the general class of polarization measures proposed by ER. They argue that the dichotomous nature of the distance across groups has important implications for the properties of the index and show that there is only one level of polarization sensitivity parameter ($\alpha = 1$) for which the discrete polarization measure satisfies a version of the properties of polarization. There is however one problematic aspect of the *RQ* index: eventhough constructed using discrete distances between individuals (groups), the index relies on some explicit results from the *ER* model, which is based on continous rather than discrete distances. Chakravarty and Maharaj (2011) develop some characterizations of the *RQ* index using alternative sets of independent axioms.

Similarly to the *FRAC* index which is defined as the probability that two randomly selected individuals belong to different ethnic categories (groups), the *RQ* index of discrete polarization is a positive multiple of the probability that out of *three* randomly selected individuals in the population, *two* will belong to the same group (Chakravarty and Maharaj, 2012). This index can also be interpreted as the

[5]For $K = 4$, the index is bounded between 0 and 1.

measure of a normalized distance of a particular distribution of ethnic groups from a bimodal distribution[6].

The RQ index of polarization defined in (1.7) is the sum of the function $f(\pi) = \pi^2(1-\pi)$ evaluated at the different π_i. While $FRAC$ is quasiconcave for all $\pi \in (0,1)$, the RQ index is convex for $\pi < 1/3$ and concave for $\pi > 1/3$. Figure 2 shows the shape of the RQ index of polarization for the case of three groups. The RQ index satisfies the following properties:

1. *A transfer of population from a group to a smaller one increases RQ if the size of both groups is greater than 1/3. When the size of both groups is lower than 1/3, RQ decreases.*

2. *For any given number of groups, RQ reaches its maximum when the population is concentrated on two equally sized groups only. The RQ index reaches its global maximum when there are two groups of equal size.*

3. *Over the set of uniform distributions, RQ decreases with the number of groups (starting from $n = 2$).*

4. *The split of any group with population share π into two new groups with population shares π' and π'', $\pi' + \pi'' = \pi$, increases RQ if and only if $\pi > 2/3$.*

It is worth noting here that the RQ index of polarization is *not* sensible to population transfers in the case of three groups with the relative size of one group fixed to 1/3. In that particular case, the RQ index of polarization equals 8/9 for any $\Pi = (\pi_1, 1-\pi_1-1/3, 1/3)$, $\pi_1 \in (0, 2/3)$. The RQ index, hence, ranks equally a distribution where one group has an absolute majority and another where all the groups have the same size, as long as the relative size of one group is equal to 1/3.

[6]This is actually the original form of the index (Reynal - Querol, 1998): $Q = 1 - \sum_i (\frac{1/2-\pi_i}{1/2})^2 \pi_i$.

5. *For $n = 3$ and $\pi_i = 1/3$ for some i, the RQ index is insensible to population transfers between the remaining two groups.*

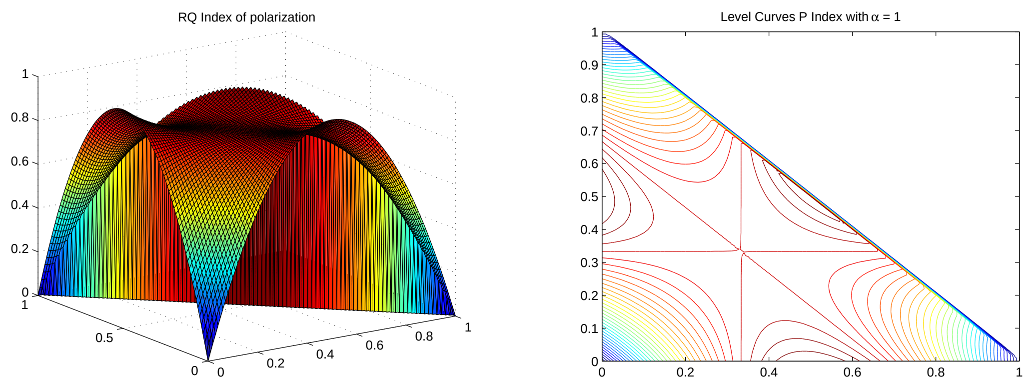


Figure 1.2: *RQ* Index for the case of Three Groups

At this point we can conclude that fractionalization and polarization (measured by *FRAC* and *RQ*) are two different and, on occasion, conflicting concepts. For instance, property 1 of *RQ* is not satisfied by *FRAC* where *any* equalization of group sizes increases the index while *RQ* can increase or decrese depending on the size of the groups involved. The second property is satisfied by both indices at least when there are only two groups. When the number of groups is greater than 2, the *FRAC* index is maximal when the population is uniformly distributed, while the *RQ* index continues to be maximal when the population is concentrated on two equally sized groups (Esteban and Ray, 2008). The third property is completely opposite for *FRAC* and *RQ*. By the fourth property any split increases *FRAC*, while *RQ* may either decrease or increase depending on the size of the splitted group. Finally, with three groups, the *FRAC* index is never insensible to population transfers between groups.

[?]

Chapter 2

Ethnic Diversity and Effective Power: The P Index of Conflict Potential

"The most severe ethnic conflict will arise [...] where a substantial ethnic minority faces an ethnic majority that can, given ethnic voting, win for sure in any national election" D. L. Horowitz, 1985.

2.1 The P Index of Conflict Potential

The empirical evidence on the association between ethnic diversity and conflict is ironically heterogeneous. Collier and Hoeffler (2004) show that the net effect of increased ethnic diversity is the sum of its effect on ethnic fragmentation and its effect on ethnic dominance. Ethnic fractionalization alone is not significant but it becomes significant in combination with ethnic dominance. Sambanis (2001, 2004) and Hegre and Sambanis (2006) find a positive and statistically robust association between ethnic fractionalization and conflict and argue that as a country becomes ethnically more fragmented, the probability of civil war increases. Collier (2001) shows that the interaction between ethnic and religious fractionalization (which they term as "social fractionalization") is negatively correlated with the likelihood of conflict and they conclude that the risk of civil conflict is lower in societies which are fractionalized by ethnicity and religion. Several other scholars have argued that the relationship between ethnic diversity and conflict is not monotonic and suggest, in line with Horrowitz (1985), that highly homogeneous and highly heterogeneous societies are less conflictual with respect to societies divided into a few prominent ethnic groups. Following this logic, Montalvo and Reynal-Querol (2002, 2005) argue that measures of fractionalization have failed to find a robust association between ethnic diversity and conflict because of the implicit assumption on the linearity between increasing fragmentation and conflict vulnerability. Applying their index of discrete ethnic polarization they find a positive and statistically significant association between ethnic polarization and the incidence (or occurence) of conflict[1]. Fearon and Laitin (2003), Fearon, Kasara and Laitin (2007) and Collier and Hoeffler (2004) however, consider onset rather than incidence and find no statistically significant correlation between ethnic polarization and conflict. Schneider and Wiesehomeier (2010) confirm that the relationship between ethnic polarization and conflict is ambigous and depends on the use of civil war incidence or civil war onset as an outcome variable. Finally, Cederman and Girardin (2007) and Cederman, Min and Wimmer (2009, 2010) show that once we account for the political dynamics of ethnic exclusion and competition, ethnic diversity in and of itself has no effect on the likelihood of civil conflict. While there is a wide consensus that low levels of development, large population and political exclusion and competition have a significant and robust impact on the likelihood

[1]A similar results can be obtained by considering the square of ethnic fractionalization - if it has a positive sign it means that conflict is most probable at the intermediate levels of fractionalization (Collier and Hoeffler, 1998).

of conflict, there is no clear evidence on the role of distributional aspects of ethnic diversity, *ceteris paribus*.

Although polarization and fractionalization capture different aspects of ethnicity, they still seem unable to establish a clear link between ethnic diversity and conflict. There are several possible explanation for the variation in results. First, there is no uniform criterion for determining whether a country is experiencing a conflict in a given year as is discussed at lenght by Sambanis (2004): some authors consider "real wars", *i.e.* those internal conflicts that count more than 1000 battle deaths in a single year (Fearon and Laitin, 2003; Collier and Hoeffler, 1998, 2004), others consider civil "conflicts", or those that count at least 25 battle deaths per annum (Montalvo and Reynal-Querol, 2005). Moreover, there are only few serious attempts to code *ethnic* conflicts which are different from civil conflicts (Sambanis, 2004; Cederman, Min and Wimmer, 2009). Second, there are at least three different aspects of conflict, namely *incidence*, *onset* and *duration*. Third, there are three main sources for data on ethnic distribution (World Christian Encyclopedia, WCE; Encyclopedia Britannica, EB; Atlas Narodov Mira, ANM) that use different criteria to define ethnic (and religious) groups. The results of a study may clearly depend on the data source used.

Without discussing the differences between data sources and conflict codings in detail, we want to propose another possible explanation for the variation in results. The above empirical evidence suggests that each of the three aspects of ethnic diversity is important under certain model specifications or with alternative data sets used in the empirical analysis. In other words, we do not have a strong evidence in favor of one aspect of ethnic diversity against another (with the exception of Montalvo and Reynal-Querol (2005) whose results, however, are not robust to alternative definition of the dependent variable, and the Collier and Hoeffler's dominance dummy). Here we argue that the relevance of each distributional aspect of ethnicity may depend on the features of the population distribution across ethnic groups. Since all diversity indices combine the effects of *interaction* and *power*, under some population distributions the interaction component may dominate the power component and *viceversa*. In other words, for some population distributions, the relevant aspect of ethnicity could be the fractionalization, for others this could be the dominance or a combination between the two.

We propose a parameterized index of ethnic diversity which we refer to as the *P Index of Conflict Potential*. Our starting point is the Esteban and Ray's (1994) model of conflict potential. According to the authors, the potential of conflict of a society is given by the sum of all inter-personal antagonisms. Antagonism or alienation derives from the distance between individuals (or groups) and it becomes "effective" once it is translated into effective voicing or protest. The efficiency of a group to translate the potential alienation into voicing depends on the cohesiveness within the group which in turn is determined by that group's relative size only: larger the group, higher is its "voicing" potential. Our approach departs from ER by two specific features. First, we assume that the effective power of a group depends not only on that group's relative size but also on the relative sizes of all the other groups in the population. This simply means that we allow the groups' effective power to depend on the features of the entire distribution of population across groups. The effective power of a group with a fixed population size may vary across distributions with the same number of groups in response to the variation in the relative sizes of the other groups. Second, we do not treat each group as a unitary actor but we assume that groups can either act individually or form alliances with other groups in order to exploit increasing returns to coalition formation (whether there are at all). We show that for certain values of the key parameter, the P index reduces to the existing indices of ethnic diversity (F and RQ), while in the limit it assumes a particular form that assigns different weights to the overall effects of power and across-group interaction on conflict potential according to the features of the underlying population distribution across groups. The index is able to capture the presence of an extreme form of ethnic dominance, which is intuitively related to the Penrose - Banzhaf (1946, 1965) and Shapley - Shubik (1954) definitions of voting power in a simple majority game. Indeed, the effective power in that case is given by the relative Penrose - Banzhaf index of voting power in a simple majority game. In what follows we present the way in which the P index is constructed starting from the general specification of the Esteban and Ray's model of social antagonism. We then axiomatically characterize the effective power function and we analyse the shape of the P index for different parameter values and different ethnic distributions. Finally, we explore whether the P index is substantially different from the RQ discrete polarization index and consider an alternative specification for the effective power function that relies on relative rather than absolute majority.

2.1.1 Effective Power

Let π_i be the relative population size of group i, where $i = 1, 2, ..., n$, and $\Pi = (\pi_1, \pi_2, ..., \pi_n)$ denotes the vector of groups' population shares. Esteban and Ray (1994) (henceforth ER) conceptualize conflict potential as the sum of all effective antagonisms between individuals or groups in the society. The antagonism or alienation felt by one individual towards other(s) is a function of the distance between them. Since, by assumption, individuals within each group are all alike (perfect intra-group homogeneity), the strength of alienation at the group's level is obtained as the sum of all the individual alienations. The alienation becomes effective once it is translated into some form of organized action, such as political mobilization, protest or rebellion. The power of a group to translate the overall alienation into effective voicing depends on the degree of cohesiveness within the group, which in turn depends on the group's relative size.

Formally, the power of a group is given by an increasing function $f(\pi_i)$ while the alienation between any two groups i and j, $j \neq i$, is proportional to the distance between them, D_{ij}. Putting together these two elements we obtain the *effective antagonism* felt by i towards j which is given by a continuous function $T(f(\pi_i), D_{ij})$. Finally, the level of conflict potential is postulated to be the sum of all the effective antagonisms:

$$ER(\Pi) = \sum_i \sum_{j \neq i} \pi_i \pi_j \; T(f(\pi_i), D_{ij}), \tag{2.1}$$

The function T is assumed to be strictly increasing in D whenever $(f, D) \gg 0$, and $T(f, 0) = 0$. Hence, bigger is the group, higher is its power to translate the alienation into protest, rebellion or any other organized action. The effective power of a group, however, depends *only* on its own relative size and not on the relative size of all the other groups. This means that, no matter how many groups there are or how the population is distributed among them, the effective power of each group is uniquely determined by its relative size.

Here, we go one step back and assume that the group's effective power depends (in addition to that group's relative size) also on the relative sizes of all the other groups. As in *ER*, we define a function Φ, that combines the group's effective power with the alienation felt towards other groups (defined on the distance between them). As in the case of *ER*, the potential of conflict in a society derives from the interaction between effective power and alienation:

$$P(\Pi) = \sum_i \sum_{j \neq i} \pi_i \pi_j \; \Phi(\pi_i, \Pi, \hat{D_{ij}}) \tag{2.2}$$

where $\hat{D_{ij}}$ is the distance between any two groups i and j from the population. In line with what suggested in Montalvo and Reynal-Querol (2005), we define $\hat{D}_{ij}$ using a discrete metric:

$$\hat{D_{ij}} = \begin{cases} 0 & \text{if} \quad i = j, \\ 1 & \text{if} \quad i \neq j. \end{cases}$$

There are several reasons that favor the use of a discrete metric to construct the "distances" across groups. First, there are no available and generally accepted measures of distance between ethnic or religious groups. [2] The measure of the distance between two ethnic categories is much more controversial than the identification of the list of ethnic groups. Second, the measurement of distances across groups may generate a larger measurement error that the "belong - does not belong to" criterion. Third, if the distance across groups is measured using the strength of the feeling of identity or political relevance then there is an important endogeneity problem (Montalvo and Reynal-Querol, 2005). Fourth, the use of a discrete metric to construct the distances simplifies significantly the structure of our model, especially because it makes the assumption on a symmetric probability distribution over coalitions even more plausible. The latter point will be clear in the next paragraph where we axiomatically derive the effective power function.

[2] Fearon (2003) suggests to measure "cultural fractionalization" based on the calculation of resemblance factors, while Vanhanen (1999) introduces an index of ethnic heterogeneity based on the concept of genetic distance which is measured as the period of time that two ethnic groups had not had close contact (the number of intergroup merriages is used to aproximate the strenght of contact). Esteban, Mayoral and Ray (2010) compute fractionalization and polarization using the data from Fearon (2003) and show that language distances embodied in polarization measure turn out to be very important correlates of ethnic conflict.

We assume that $\Phi(\pi_i, \Pi, 0) = 0$ and $\phi(\pi_i, \Pi) := \Phi(\pi_i, \Pi, 1)$ with ϕ not necessarily continous in π_i. Since $\sum \pi_i = 1$, the P index defined in (2.9) can be written as:

$$P(\Pi) = K \sum_i \phi(\pi_i, \Pi)\ \pi_i(1 - \pi_i). \tag{2.3}$$

The function $\phi(\pi_i, \Pi)$, $i = 1, 2, ..., n$ and $n \geq 2$, will be reffered to as the *effective power* associated to group i. Differently from ER, the effective power of a group i is a function of both π_i and Π_{-i}.

2.1.2 Axiomatic Derivation of the Effective Power Function

Let N be the set of all groups, *i.e.* $N = \{1, 2, ..., n\}$. Let Π denote the vector of relative population sizes for the n groups in the population. The set of all Π is the n dimensional simplex Δ^n. The effective power $\phi^n(\pi_i, \Pi)$ of any group $i \in N$ with relative population size π_i, given Π, is defined as:[3]

$$\phi^n(\pi_i, \Pi) : [0, 1] \times \Delta^n \to \Re_+.$$

We add a superscript n to distinguish between ethnic distributions characterized by different number of groups. The effective power function satisfies the following properties:

Axiom 1: Normalization (N)

[3]For ease of expoisiton here we consider $\Pi \in \Delta^n$ while for a given π_i only a subset of them is consistent with having one element equal to π_i.

For all $\pi_i \in [0,1]$, $\Pi \in \Delta^n$, and $n \geq 2$, then:

$$\sum_i \phi^n(\pi_i, \Pi) = 1, \quad i = 1, ..., n$$

with $\phi^n(0, \Pi) = 0$, and $\phi^n(1, \Pi) = 1$.

Normalization implies that the effective power of each ethnic group is contained in the interval $[0,1]$, *i.e.*, $0 \leq \phi^n(\pi_i, \Pi) \leq 1$. Note that we allow ϕ^n to be 1 (absolute power) or 0 (absence of power). This property will attribute to our measure a cardinal meaning.

Axiom 2: Monotonicity (M)

For all $\pi_i \in (0,1]$, $\Pi \in \Delta^n$, and $n \geq 2$, then:

$$\phi^n(\pi_i, \Pi) \geq \phi^n(\pi_j, \Pi) \;\; if \;\; \pi_i \geq \pi_j, \;\; \forall i, j; \; i \neq j.$$

The monotonicity axiom implies that, given any two groups with respective population shares π_i and π_j such that $\pi_i \geq \pi_j$, the effective power of the bigger group *cannot* be lower than the effective power of the smaller group. We allow, hence, the effective power of the bigger group to be equal to the effective power of the smaller one. The next property is implied by Axiom 2, and is obtained when $\pi_i = \pi_j$.

Axiom 2^a: Symmetry

$$\phi^n(\pi_i, \Pi) = \phi^n(\pi_j, \Pi) \;\; if \;\; \pi_i = \pi_j, \;\; \forall i, j; \;\; i \neq j.$$

The Symmetry property states that, if two groups are of equal size, then their effective power has to be the same. The reverse, however, is *not* necessarily true. In other words, the equality of effective power between any two groups is *not* an indicator of the equality of their relative sizes. Symmetry in combination with Normalization immediately implies that if all groups were to have identical relative size each one of them would have an effective power equal to $1/n$. This result will provide a reference for all the indices that we will obtain from the axiomatization. In fact a common feature of all the indices is that they will all exibith the same value for distributions where all the groups are of equal relative size. Morever this value will be proportional to the fractionalization index divided by n.

At this point we introduce two important assumptions:

i) Groups can either act individually or pool their strengths together through a coalition formation.

ii) If any two groups i and j form a coalition, the remaining groups belong to the "opponent" block. So we consider only the *bipartitions* of the population.

What is the *rationale* behind these two assumptions? Suppose that there are n ethnic groups in conflict with only one strategic endowment: human resources. With this simple "technology of conflict" (Hirshleifer, 1989), a small group that is interested in winning the context may find it profitable to join the forces with some other group(s) in order to contrast the adversary, even at the cost of the future division of power within the winning block. Consequently, a group that is big enough to ensure the victory alone will act as a unitary actor. Hence, one block or coalition is formed in order to contrast or challange the other block. Moreover, Skaperdas (1998), Tan and Wang (1997, 2010) and Esteban and Sakovics (2003) show that in a three groups contest, parties will have an incentive to form an alliance against the third if the formation of the alliance generates synergies which enhance the winning probability of the alliance. Skaperdas (1998) argue that this tendency is not only theoretical but also frequent in many real life situations and provide an example of the on and off alliance of the Bosnian and Croat forces against the more (strategically) well endowed Serb forces in Bosnia during the recent past.

Here we do not model any endogenous mechanism of coalition formation nor we are interested in which coalition is more probable than another. The probability distribution over coalitions hence is assumed to be symmetric. Symmetry is a desirable characteristic if we do not have information about the stable attitudes and differences among groups. Since we use a discrete metric to define distances, the assumption of symmetry seems even more plausible. Generally, the probability distribution over coalitions could also be asymmetric. For instance, if we were to attach to each individual with a clear ethnic or religious marker the level of income or wealth it possesses, we could define the probability of any coalition in terms of the similarity between the groups income or wealth attributes (the level of income or wealth is considered as a good proxy for political preferences). However, we do not have a good and complete data on income distribution among individuals that belong to different ethnic categories. As we will show later, even under these symplifying assumptions the distribution of the effective power between groups or blocks of groups will depend on the characteristics of the population distribution across them. This important feature of the effective power function will make the P index substantially different (both theoretically and empirically) from the existing distributional indices of ethnic diversity based on the assumption of groups as unitary actors.

In order to define the effective power for any arbitrary number of groups we start from the most simple case, *i.e.* from the case of a distribution with only two groups. Although the relevance of a coalition formation in that case is trivial, the results that we obtain for the case with two groups can be used to generalize the results for any arbitrary number of groups. This is because in the two groups case, there is only one possible bipartition, *i.e.* a block composed by group 1 versus another block composed by group 2 and *viceversa*.

Consider a population divided into two different ethnic groups with population shares π and $1-\pi$, $\pi > 0$. The relative effective power between groups is assumed to be a function of the groups relative size. Denoting with $\phi^2(\pi)$ and $\phi^2(1-\pi)$ the effective power of the groups, we define the *relative effective power* between groups as:

$$\frac{\phi^2(\pi)}{\phi^2(1-\pi)} = r(\rho) \quad where \quad \rho = \frac{\pi}{1-\pi}. \tag{2.4}$$

The index $\rho = \frac{\pi}{1-\pi}$ denotes the relative population size between the two groups.[4] The relative effective power between groups, hence, is defined as a function of the groups *relative population size*. We also know from the Symmetry Axiom that whenever $\pi = 1/2$ and, hence, $\rho = 1$ the groups will equally share the power, *i.e.* $r(1) = 1$. Moreover, given a functional form for ϕ^n, $r(\rho)$ derives directly by recalling other, $\pi = \rho/(1+\rho)$ and thus $r(\rho) := \frac{\phi^2(\rho/(1+\rho))}{\phi^2(1/(1+\rho))}$.

[4]Since $\pi \in (0,1)$, the coefficient ρ is the population shares *odds ratio*.

The relationship between the two has to satisfy the following property:

Axiom 3: Two Groups Relative Power Homogeneity - 2GRPH

Given Π and $\Pi^{'}$, let $\pi, \pi^{'} < 1/2 \Leftrightarrow \rho, \rho^{'} < 1$. Then, if $r(\rho), r(\rho^{'}) \neq 0$:

$$\frac{r(\lambda\rho)}{r(\rho)} = \frac{r(\lambda\rho^{'})}{r(\rho^{'})}; \;\; \forall\ \rho, \rho^{'} < 1,\ \lambda > 0\ \ s.t.\ \ \lambda\rho, \lambda\rho^{'} < 1.$$

We denote with **2GRPH*** stronger version of the axiom that requires that the condition holds for all $\rho, \rho^{'}, \lambda\rho, \lambda\rho^{'}$ not necessarily with values lower than 1, with $\rho, \rho^{'} \neq 0$.

In order to interpret the 2GRPH axiom, suppose that we start from a population distribution Π in which, for instance, the size of the smaller group is 40% of that of the bigger group (this means that $\rho = 0.4$). Now immagine that one part of the population from the second group migrates in a neighboring country such that the relative population size becomes 0.8, *i.e.* the size of the smaller group is now 80% of that of the bigger group. This is equivalent to say that the size of the smaller groups with respect to the size of the bigger one has doubled (i.e., $\lambda = 2$). Such a variation in the relative population size, may affect the *relative effective power* between the two groups. Now immagine a similar situation in which the size of one group with respect to the size of the other groups moves from 30% to 60%. As in the previous case, the relative population size ρ has doubled. The 2GRPH requires that the variation in the *relative effective power* is the same in both cases. In other words, no matter from where we start with respect to the relative size ρ, the variation in the relative effective power is always the same as long as the change in ρ is the same across the two distributions.

At this point we can state the first result:

Lemma 1

Given $n = 2$, the Effective Power of a group with population share π satisfies *Normalization*, *Monotonicity* and *2GRPH*, if and only if $\phi^2(\pi) = \phi^2_{\alpha,\beta}(\pi)$ for $\alpha \in \Re_+ \cup \infty$ and $\beta \in [0, 1]$ where

$$\phi^2_{\alpha,\beta}(\pi) := \begin{cases} \frac{\pi^\alpha}{\pi^\alpha+\beta(1-\pi)^\alpha} & \text{if} \quad \pi > 1/2, \\ 1/2 & \text{if} \quad \pi = 1/2, \\ \frac{\beta\pi^\alpha}{\beta\pi^\alpha+(1-\pi)^\alpha} & \text{if} \quad \pi < 1/2. \end{cases}$$

Proof of Lemma 1.

Sufficiency part.

Note that the obtained specification for $\phi^2_{\alpha,\beta}(\pi)$ satisfies the axioms considered.

Necessity part.

Consider axiom 2GRPH, requiring that $\frac{r(\lambda\rho)}{r(\rho)} = \frac{r(\lambda\rho^{'})}{r(\rho^{'})}$ for all $\rho, \rho^{'} < 1,\ \lambda > 0, \lambda\rho, \lambda\rho^{'} < 1$ with $r(\rho), r(\rho^{'}) \neq 0$.

We first consider the implications arising from this axiom, together with all the other axioms, where $r(\rho), r(\rho^{'}) \neq 0$, then we will move to the case where there exist ρ s.t. $r(\rho) = 0$.

Recall first that if Monotonicity holds then, $\phi^2(\pi) \leq \phi^2(\pi')$ if $\pi < \pi'$, while if Normalization holds $\phi^2(1-\pi) = 1-\phi^2(\pi)$ and therefore if $\pi < \pi'$ then $\frac{\phi^2(\pi)}{\phi^2(1-\pi)} \leq \frac{\phi^2(\pi')}{\phi^2(1-\pi')}$, thus by construction $r(\rho) \leq r(\rho')$ if $\rho < \rho'$, i.e., $r(\rho)$ is not decreasing. Note moreover that if 2GRPH holds then if $r(\rho) = r(\rho')$ for some $\rho < \rho'$ in some interval of $(0, 1)$ then given that we can set $\rho' = \lambda\rho$ the condition $\frac{r(\lambda\rho)}{r(\rho)}$ becomes $\frac{r(\rho')}{r(\rho)} = 1$ that holds in the interval and therefore, as λ varies, also for all other $\rho' \neq \rho$. As a result either $r(\rho)$ is constant and different from 0 for all $\rho < 1$ or it is strictly increasing, that is $r(\rho) < r(\rho')$ if $\rho < \rho'$. Here we focus on the latter case.

If $r(\rho), r(\rho^{'}) \neq 0$ then 2GRPH holds, let $\rho_0 := \lambda\rho \in (0, 1)$ that is $\lambda = \rho_0/\rho$ it follows that

$$\frac{r(\rho_0)}{r(\rho)} = \frac{r(\rho^{'} \cdot \rho_0/\rho)}{r(\rho^{'})} = g(\rho_0/\rho) \tag{A.1}$$

for some function $g(.)$. Note that if we set $\lambda < 1$, (we will discuss the implication of $\lambda > 1$ afterwards), then $\rho_0 < \lambda, \rho$ and $\rho_0/\rho < 1$, it then follows that $r(\rho_0) = g(\rho_0/\rho) \cdot r(\rho)$ for all $\rho_0, \rho < 1$ and $\rho_0/\rho < 1$. The functional equation therefore holds also if we swap ρ_0/ρ with ρ, and we obtain $r(\rho_0) = g(\rho) \cdot r(\rho_0/\rho)$ for all $\rho_0, \rho < 1$ and $\rho_0/\rho < 1$. As a result it would hold that:

$$r(\rho_0) = g(\rho_0/\rho) \cdot r(\rho) = g(\rho) \cdot r(\rho_0/\rho)$$

for all $\rho_0, \rho < 1$ and $\rho_0/\rho < 1$. Note that we have assumed that $r(\rho) > 0$ for all ρ, and therefore also $r(\rho_0), r(\rho_0/\rho) > 0$, which implies that $g(\rho) > 0$. we can then rewrite

$$\frac{g(\rho_0/\rho)}{g(\rho)} = \frac{r(\rho_0/\rho)}{r(\rho)} > 0$$

for all $\rho_0, \rho < 1$ and $\rho_0/\rho < 1$. Which is equivalent to set $g(\rho) = K \cdot r(\rho)$ for some $K > 0$. By substituting into (A.1) we obtain

$$r(\rho_0) = r(\rho) \cdot K \cdot r(\rho_0/\rho)$$

if we consider the function $\sigma(\rho) := K \cdot r(\rho)$ we have

$$\sigma(\rho_0) = \sigma(\rho) \cdot \sigma(\rho_0/\rho)$$

for all $\rho_0, \rho < 1$ and $\rho_0/\rho < 1$. The following is the (multiplicative) Cauchy functional equation specified for a domain where $\rho \in (0,1)$ and for $\sigma(\rho)$ strictly increasing. Note that the problem can be set equivalently to the one where the domain is on the strictly positive real line $\Re_{++}$ by simply setting $\sigma(\rho) := s(x)$ where $\rho = x/(1+x)$. The general solution for the restricted domain is in Eichhorn (1978) [see Theorem 1.9.13 and Remark 1.9.23]. It follows that

$$\sigma(\rho) = \rho^\alpha \text{ for all } \alpha > 0.$$

Moreover the case analyzed. earlier where $r(\rho)$ is constant can be summarized by the solution where $\alpha = 0$.

In fact, by substituting for $\sigma(\rho) := K \cdot r(\rho)$ with $K > 0$ one obtains that

$$r(\rho) = \beta \cdot \rho^\alpha \text{ for all } \alpha \geq 0,\ \beta > 0 \tag{A.2}$$

for all $\rho \in (0,1)$. Note that this solution, implies that 2GRPH holds also for all $\lambda > 1$.

Before analyzing the implications for the solution arising from other axioms we go back to the case where there exist ρ s.t. $r(\rho) = 0$. In this case 2GRPH does not hold.

However, we have just derived that for some ρ the function $r(\rho)$ is not equal to 0 then the solution (A.2) should hold. It follows that either $r(\rho) = 0$ for all $\rho \in (0, 1)$ or (A.2) hods. The former case can be embedded into (A.2) by setting $\beta = 0$.

We now move to consider the implications of the remaining axioms. Recall that by Normalization $\phi^2(1-\pi) = 1 - \phi^2(\pi)$, then by definition

$$\frac{\phi^2(\pi)}{\phi^2(1-\pi)} = \frac{\phi^2(\pi)}{1-\phi^2(\pi)} = \beta \cdot \frac{\pi^\alpha}{(1-\pi)^\alpha} \text{ forall } \alpha \geq 0,\ \beta \geq 0$$

where $\pi < 1/2$, that is $\phi^2(\pi) = \beta \cdot \frac{\pi^\alpha}{(1-\pi)^\alpha} [1 - \phi^2(\pi)]$ giving

$$\phi^2(\pi) = \frac{\beta \cdot \pi^\alpha}{\beta \cdot \pi^\alpha + (1-\pi)^\alpha}$$

for all $\alpha \geq 0$, $\beta \geq 0$, where $\pi < 1/2$. Note that by monotonicity $\phi^2(\pi) \leq \phi^2(1/2) = 1/2$, where the latter equality is obtained by Symmetry and Normalization. It follows that $\phi^2(1/2) = \frac{\beta}{\beta+1} \leq 1/2$ requires that $\beta \leq 1$. Which gives the desired result. The values for $\phi^2(\pi)$ for $\pi > 1/2$ are obtained by setting $\phi^2(\pi) = 1 - \phi^2(1-\pi)$ where $1 - \pi < 1/2$. ■

This functional form for the effective power is similar to the *ratio form contest success function* commonly used in the rent-seeking literature (Tullock (1967, 1980) (with $\alpha = 1$), Skaperdas (1996, 1998) and Nitzan (1991)). The only difference is the parameter β in the above formula. When $\beta = 1$ we obtain the ratio form contest success function in Tullock (1967, 1980) and Skaperdas (1996, 1998). However, the axiomatization of the effective power function differs from those in the literature. That is, the results obtained here *cannot* be obtained from the properties underlying a standard contest success function. For instance, Skaperdas (1996, 1998) assumes homogeneity of the relevant variables that are unbounded, which is not the case in our problem where π is bounded.

What is the form of the effective power function for different values of the parameter β?

Case 1: $\boldsymbol{\beta = 0}$

When $\beta = 0$ the parameter α plays no role and the effective power of a smaller group, *i.e.* a group with $\pi < 1/2$ is equal to 0 while the effective power of a bigger group is 1 (Figure 2.3). When groups have the same size, they equally share the power.

Case 2: $\boldsymbol{\beta \in (0, 1)}$

When $\beta \in (0, 1)$ the effective power of a group with the population share $\pi < 1/2$ is:

$$\phi^2_{\alpha,\beta}(\pi) = \frac{\beta\pi^\alpha}{\beta\pi^\alpha + (1-\pi)^\alpha} \quad for \quad \alpha \geq 0, \quad and \; \beta \in (0, 1).$$

Suppose, for instance $\beta = 1/2$. The effective power depends crucially on the value of the coefficient α. When $\alpha = 0$, the effective power of a group with $\pi < 1/2$ is $1/3$ (and the effective power of a group with $\pi > 1/2$ is $2/3$). When $\alpha = 1$, the effective power function is convex for all π. When $\alpha \to \infty$, the situation is identical to that with $\beta = 0$, namely the effective power of a group with $\pi > 1/2$ is 1 while the effective power of a group with $\pi < 1/2$ is equal to 0. Finally, by Lemma 1, when $\pi = 1 - \pi = 1/2$ the effective power of groups is equal to $1/2$ (Figure 2.3).

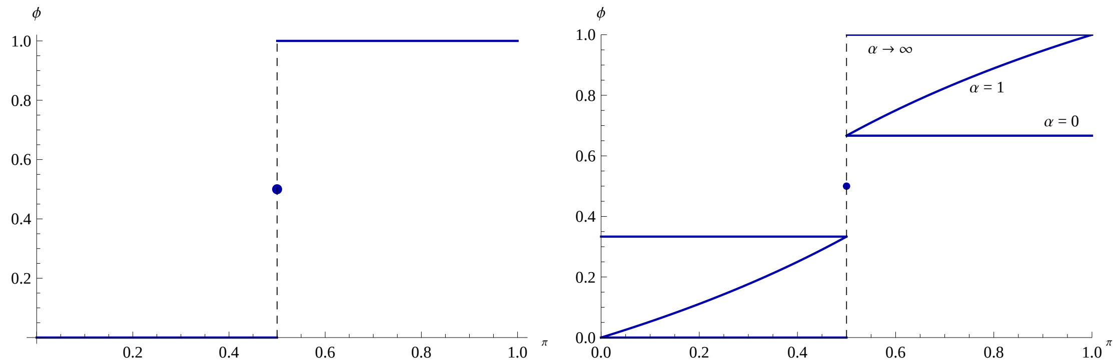


Figure 2.1: Effective Power with $\beta = 0$ and $\beta = 1/2$.

An interesting case occurs when $\alpha \to \infty$ and $\beta \in [0, 1]$. In that particular case the effective power function is given by the following expression:

$$\phi^2_{\infty,\beta} = \begin{cases} \frac{\beta}{1+\beta} & \text{if} \quad \pi > 1/2, \\ 1/2 & \text{if} \quad \pi = 1/2, \\ \frac{1}{1+\beta} & \text{if} \quad \pi < 1/2. \end{cases}$$

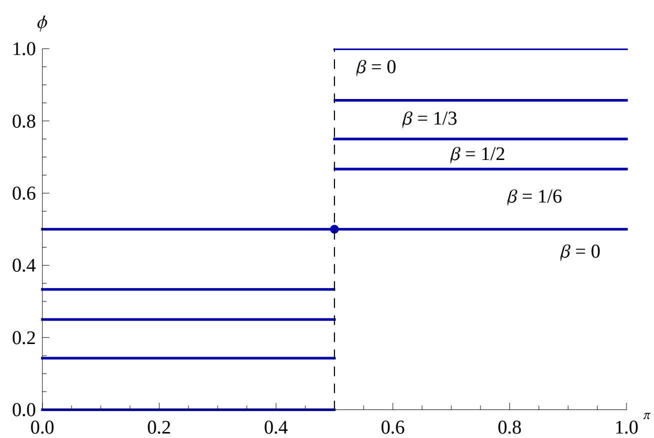


Figure 2.2: Effective Power with $\alpha \to \infty$ and $\beta \in [0, 1]$.

Case 3: $\boldsymbol{\beta = 1}$

When $\beta = 1$, the effective power function is given by the expression:

$$\phi^2_{\alpha,1} = \frac{\pi^\alpha}{\pi^\alpha + (1-\pi)^\alpha} \quad for \quad \alpha \geq 0, \quad and$$

$$\phi^2_{\infty,1} = \begin{cases} 1 & \text{if} \quad \pi > 1/2, \\ 1/2 & \text{if} \quad \pi = 1/2, \\ 0 & \text{if} \quad \pi < 1/2. \end{cases}$$

This functional form is analogous to the *ratio form contest success function* commonly used in the contest model, the workhorse of the formal conflict literature (Tullock, 1967, 1980; Skaperdas, 1996, 1998; Nitzan, 1991). A contest model assumes two contended parties, a rebel group and a government that face the problem of allocating resources between production and appropriation. While production is modeled in the standard manner, the appropriation is modeled using the above contest success function where inputs translate into the probability of one side winning the contest and consuming the opponent production in addition to their own. Contest models predict that the odds of winning increase with the relative effectiveness of the so-called "conflict technology" which can include any factor that influences effectiveness. In our case the conflict technology uses only one type of input, namely the groups' relative abundance of human resources or conflict labor (Esteban and Ray, 2008) and the probability of winning is interpreted as the power of a group to win a contest. The positive spillovers from a group size preponderance to the probability of winning (effective power) are captured by the parameter α (Figure 2.4).

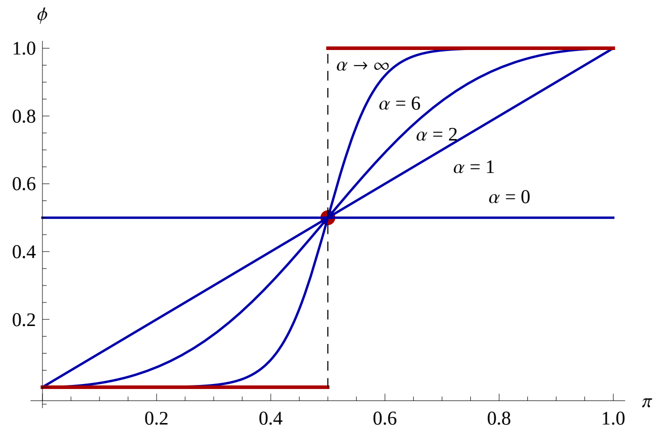


Figure 2.3: Effective Power with $\beta = 1$.

When $\alpha = 0$ the odds of winning are constant across groups, for $\alpha = 1$ the odds of winning are given by the groups relative population size while for $\alpha \to \infty$ the majoritarian group wins for sure, *i.e.* it holds the absolute power.

It is worth noting here that the case with $\alpha \to \infty$ and $\beta = 1$ corresponds to the case where $\beta = 0$. On the other side, with $\alpha = 0$ and $0 < \beta < 1$ the total amount of power assigned to the majoritarian group varies between $1/2$ and 1. In other words, as β increases from 0 to 1 the gap between the effective power of the majoritarian and the minoritarian group shrinks. For any $0 < \beta < 1$ the amount of power possessed by the minority is positive and as β approaches 1 it tends to equalize the power of the majoritarian group. The value of the coefficient β hence determines how much of the total effective power (which is normalized to 1) "belong" to the majoritarian group and *viceversa*.

Before moving to the general characterization for the case of more than two groups we would like to point out that some of the results obtained in Lemma 1 are not robust to different specifications of the axioms. In particular the following remarks hold.

Remark If one considers the modified axiom **2GRPH***, or requires that the solution to $\phi^2(\pi)$ is continuous for $\pi \in (0,1)$ then $\beta = 1$.

The first point is obtained by applying the general solution of the Cauchy functional equation behind the result in Lemma 1 in order to hold also for $\rho \geq 1$. If this is the case then the solution of the problem should be continuous for $\rho = 1$, which requires that the solution in Lemma 1 is continuous for $\pi = 1/2$, thereby leading to the case of $\beta = 1$.

Remark If one specifies **2GRPH** s.t. $\frac{r(\lambda\rho)}{r(\rho)} = \frac{r(\lambda\rho^{'})}{r(\rho^{'})} = r(\lambda)\ \forall\ \rho, \rho^{'} < 1,\ \lambda > 0$ s.t. $\lambda\rho, \lambda\rho^{'} < 1$ [denote it as **2GRPH****] then $\beta = 1$ or $\beta = 0$.

The above restrictions are obtained by requiring that the value of $r(\rho) = \beta\rho^{\alpha}$ as obtained in Lemma 1 satisfies the condition in the modified 2GRPH axiom. This implies that $\frac{\beta(\lambda\rho)^{\alpha}}{\beta\rho^{\alpha}} = \beta\lambda^{\alpha}$ which is satisfied only if $\beta = 1$. The case $\beta = 0$ also holds because the solution $r(\rho) = 0$ for all $\rho > 0$ is not affected by the specification of the 2GRPH axiom even in its modified form 2GRPH**.

Axiomatic Derivation of the Effective Power Function: n > 2

Suppose now $n > 2$ and $\beta = 1$. Given the set of all ethnic groups N, a coalition is defined as any subset of N. There are three particular types of coalitions, namely: the *grand* coalition or the coalition that contains all the groups; the *individual* coalition or the coalition that contains only one group (in this particular case a group is a unitary actor that contrasts the opponent block composed by all the other groups); and, finally, the *empty* coalition or the coalition that contains no group.

Since we assume that groups can either act individually and form alliances or blocks with other groups, any measure of their *effective* power should take this possibility into account. This means that, in addition to the case where one group "fights" alone against the rest (unitary actor), a measure of effective power has to take into account all the potential contributions to all the coalitions that a particular group can (theoretically) belong to.

Denote with C_i the set of all coalitions c, such that $i \in c$. In this set we include both the grand coalition and the $i's$ individual coalition. The value of any coalition $c \in C_i$ can be defined in terms of its "power". The power of any coalition $c \in C_i$ can be obtained by Lemma 1:

$$\phi^2(c) = \phi^2(\sum_{j \in c} \pi_j). \tag{2.5}$$

It follows that $\phi^n(\emptyset) = 0$ and $\phi^n(1) = 1$. In other words, the value of an empty coalition is 0 and the value of the grand coalition is 1.

We next define the *marginal contribution* of group i to the worth of any coalition $c \in C_i$ as:

$$m_i(c) := \phi^2(\sum_{j \in c} \pi_j) - \phi^2(\sum_{j \in c} \pi_j - \pi_i). \tag{2.6}$$

The sum of marginal contributions of group i over all coalitions in C_i is:

$$M_i = \sum_{c \in C_i} m_i(c). \tag{2.7}$$

The effective power of any group i will be a function of M_i but it will also depend on M_{-i}. However, as stated in the next axiom, what counts for the relative effective power between any two groups i and j is the ratio between their marginal contributions:

Axiom 4: Relative Effective Power (REP)

For any $i, j \in N$, $i \neq j$ and $n \geq 2$; $\exists\ g : \Re_+ \rightarrow \Re_+$, such that $\phi^n(\pi_j, \Pi) > 0$:

$$\frac{\phi^n(\pi_i, \Pi)}{\phi^n(\pi_j, \Pi)} = \frac{g(M_i)}{g(M_j)}.$$

This axiom states that the relative effective power between any two groups $i, j \in N$ depends on the their relative sum of marginal contributions to all the coalitions that they can theoretically belong to. However, in order to compare the effective power of any couple of groups from the population, it is sufficient to compare their marginal contributions. That is, no matter how many groups there are in the population or how the marginal contributions are distributed among them, the relative effective power between any two groups in comparison will be determined *exclusively* by their own M.

The relationship between marginal contributions and effective power is given by the following axiom:

Axiom 5: n Groups Relative Power Homogeneity (nGRPH)

Given two ethnic distributions, Π and Π' with the same number of groups, $n \geq 2$, if $\phi^n(\pi_j, \Pi) > 0$:

$$\frac{M_i}{M_j} = \frac{M_i^{'}}{M_j^{'}} \Rightarrow \frac{\phi^n(\pi_i, \Pi)}{\phi^n(\pi_j, \Pi)} = \frac{\phi^n(\pi_i^{'}, \Pi^{'})}{\phi^n(\pi_j^{'}, \Pi^{'})}.$$

Axiom 4 implies that the effective power of any two groups with same M has to be same. Moreover, if we compare two ethnic distributions with the same number of groups, if the ratio between the marginal contributions between any two groups from both distributions is the same, then their relative effective power has to be the same too.

At this point we can state the following theorem:

Theorem 1

The Effective Power of group i satisfies Axioms 1 - 5 if and only if:

$$\phi_{\alpha,\beta}^n(\pi_i, \Pi) = \frac{M_i^{\alpha,\beta}}{\sum_j M_j^{\alpha,\beta}}, \quad \forall i, j \in N; \quad i \neq j; \quad \alpha \in \Re_+ \cup \infty. \tag{2.8}$$

where $M_i^{\alpha,\beta}$ is obtained making use of $\phi_{\alpha,\beta}^2$.

Group i's effective power, hence, is defined as the relative sum of marginal contributions.

Proof of Theorem 1.

Sufficiency part.

Note that the obtained specification for $\phi_{\alpha,\beta}^n$ satisfies the axioms considered.

Necessity part.

Consider axiom REP. We first check the restrictions that make it consistent with the specification of ϕ^2 obtained in Lemma 1 applying 2GRPH.

For $n = 2$, the axiom REP requires that $\frac{\phi^2(\pi)}{\phi^2(1-\pi)} = \frac{g(M_i)}{g(M_j)}$ where M_i is associated to the group of share π and M_j to the other group. Note that by construction $M_i = 1-\phi^2(1-\pi)+\phi^2(\pi)$, and $M_j = 1-\phi^2(\pi)+\phi^2(1-\pi)$. Recalling that by Normalization $\phi^2(\pi)+\phi^2(1-\pi) = 1$, one obtains that $M_i = 2\phi^2(\pi)$, and $M_j = 2\phi^2(1-\pi)$. Thus REP requires that

$$\frac{\phi^2(\pi)}{\phi^2(1-\pi)} = \frac{g(2\phi^2(\pi))}{g(2\phi^2(1-\pi))}$$

for all $\pi \in (0, 1)$.

By letting $f(x) := g(2x)$ and recalling that $\phi^2(1-\pi) = 1-\phi^2(\pi)$ one obtains, when $\phi^2(\pi) > 0$,

$$\frac{f(\phi^2)}{\phi^2} = \frac{f(1-\phi^2)}{1-\phi^2}$$

for all $\phi^2 \in (0, 1]$, where ϕ^2 for short denotes $\phi^2(\pi)$. Recall that $\phi^2 = 1/2$ if $\pi = 1/2$.

The above functional equation is then consistent with setting $\frac{f(\phi^2)}{\phi^2} = h(\phi^2)$ if $\phi^2 < 1/2$, and $\frac{f(\phi^2)}{\phi^2} = h(1-\phi^2)$ for $\phi^2 > 1/2$, with $h(1/2) = 2f(1/2)$ for some function $h : (0, 1] \rightarrow \Re_+$.

It then follows that $g(2\phi^2) = f(\phi^2)$ for all values of the domain of $g(.)$ in $(0, 2]$ with

$$
\begin{aligned}
g(2\phi^2) &= h(\phi^2)\cdot\phi^2 \\
&= h(1-\phi^2)\cdot\phi^2 \text{ for } \phi^2 > 1/2.
\end{aligned}
$$

More generally $g(.)$ may depend on $\Pi_{-i,-j}$ and thus it can be written as related to a function $h(.)$ that depends on $\Pi_{-i,-j}$ if $n > 2$.

Thus for $M \in (0,2]$ one obtains that $g(M) = h_\Pi(M/2)\cdot M/2$ for $M \leq 1$, and $g(M) = h_\Pi(2 - M/2)\cdot M/2$ for $M > 1$.

Thus for the case where $M \in (0,2]$ sat with $M_j > 1$ and $M_i < 1$ then $\frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)} = \frac{g(M_i)}{g(M_j)} = \frac{h_\Pi(M_i/2)\cdot M_i}{h_\Pi(2-M_j/2)\cdot M_j}$.

By applying nGRPH one obtains that

$$
\frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)} = H(M_i/M_j)
$$

where $H(M_i/M_j)$ does not depends on Π.

By combining with the previous restrictions one obtains that this is the case only if $h_\Pi(.) = c > 0$.

That is if $M_j, M_i \in (0,2]$ then $\frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)} = \frac{M_i}{M_j}$. However, this is the case whenever π_i, π_j are sufficiently small.

The derived proportionality of $\frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)}$ hods for a given ratio $\frac{M_i}{M_j}$, but for appropriate choices of π_i and π_j when $n \geq 3$ one can guarantee that $M_j, M_i \in (0,2]$ and that $\frac{M_i}{M_j}$ can reach any positive value.

Thus we obtain that

$$\frac{\phi^n(\pi_i, \Pi)}{\phi^n(\pi_j, \Pi)} = \frac{M_i}{M_j} \tag{A.3}$$

for all M_j, M_i and all Π (that are consistent with π_i, π_j).

The specification of the two axioms NGRPH and REP lead to different restrictions on the final functional form thereby showing their independence.

The desired result is then obtained by imposing the Normalization axiom. In fact condition (A.3) implies that in the more general case $\phi^n(\pi_i, \Pi) = M_i \cdot w(\mathbf{M})$ where $\mathbf{M}$ denotes the distribution of all aggregated marginal contributions of each group, and $w(.)$ is a generic function, identical for all groups.

If this is the case then, by Normalization, $\sum_i \phi^n(\pi_i, \Pi) = \sum_i M_i \cdot w(\mathbf{M}) = w(\mathbf{M}) \cdot \sum_i M_i = 1$ thus, $w(\mathbf{M}) = 1/\sum_i M_i$, thereby leading to

$$\phi^n(\pi_i, \Pi) = \frac{M_i}{\sum_i M_i} \tag{A.4}$$

where the M_i components are obtained making use of the ϕ^2 in Lemma 1.

To conclude we are left to consider the case where $\phi^n = 0$ for some group j. In order to obtain this result is should be that $M_j = 0$. If this is not the case then there exists groups j whose ϕ^n is 0 irrespective of the value of M_j. Note however that M_j is not decreasing w.r.t. π_i, and thus by Monotonicity we should have that ϕ^n is 0 also for all groups i whose size is below π_j or what M_i is lower than M_j. But according to REP what is relevant is the ratio $\frac{M_i}{M_j}$ so, taking two groups one of which has $M_i > 0$ but $\phi_i^n = 0$ with $\frac{M_i}{M_j} \neq 0$ and $\phi_j^n > 0$ one can set the distribution such that $\frac{M_i}{M_j}$ is appropriately set at a desired positive value and therefore for all pairs $\frac{M_i'}{M_j'} < \frac{M_i}{M_j} \Rightarrow \frac{\phi^n(\pi_i',\Pi')}{\phi^n(\pi_j',\Pi')} \leq \frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)} = 0$, thereby leading to a situation where all groups execpt the largest one in all possible distributions have $\phi^n = 0$. Which is not consistent with the Normalization axiom. It follows that $\phi_i^n = 0$ only if $M_i = 0$. Making the result consistent with (A.4). ■

Given (2.15) the effective power of a group is a function of the relative size of all the groups in the population. Consider the following graphical example. Suppose $n = 3$ and $\pi_3 = 0.2$. Figure 1.4 shows $\phi^3(\pi_3, \Pi)$ as a function of π_1 and α (the relative population size of group 2 is simply $1 - \pi_1 - 0.2$) when the parameter β is fixed to 1. We can see that for $\alpha = 0$, $\phi^3(\pi_3, \Pi) = 1/3$ while for $\alpha = 1$, $\phi^3(\pi_3, \Pi) = \pi_1 = 0.2$. The situation changes for $\alpha \neq 0$ and $\alpha \neq 1$. The effective power asociated to group 3 varies with π_1 and, as $\alpha \to \infty$, $\phi^3(\pi_3, \Pi)$ equals 0.2 when $\pi_1 = 1/2$ and is 0 when $\pi_1 > 1/2$. Finally, for $\pi_1 \in (1/5, 1/2)$, $\phi^3(\pi_3, \Pi) = 1/3$.

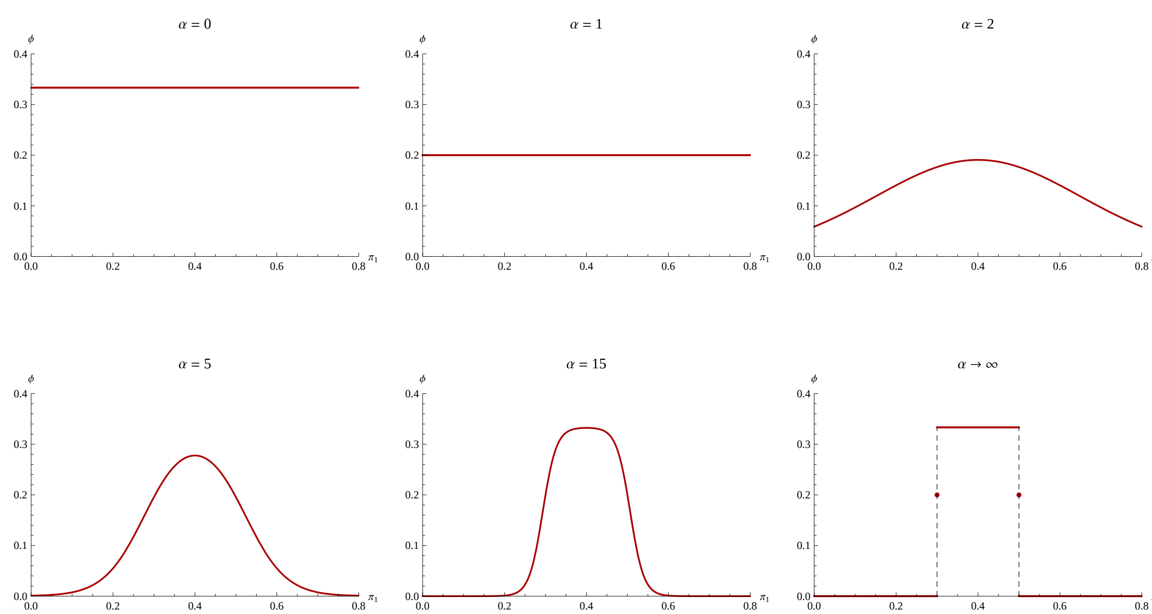


Figure 2.4: $\phi^3_\alpha(\pi_3, \Pi)$ as a function of α and π_1

For $n > 2$ and $\alpha > 1$, the effective power of any group i depends on both π_i and Π. As a consequence, the relative power of a group with a fixed population share π_i may vary significantly across different population distributions. That is, the relative power of a group with fixed population share may vary in response to the variation of the relative size of other groups.

Making use of the considerations in previous remarks we will restrict our attention in the analysis only to cases where $\beta = 0$ or $\beta = 1$.

2.1.3 Properties of the Effective Power Function

In this subsection we analyse graphically the properties of the effective power function for the case of three and four groups and we generalize the conclusions for the case of any arbitrary number of groups. In both cases the parameter β is fixed to 1.

The Case of Three Groups

In order to analyse the properties of the effective power function for the case of three groups, we classify the population distribution into three different categories:

- **A** : No group has absolute majority:
  - **A1** : Uniform Distribution
  - **A2** : Non-Uniform Distribution
- **B** : Relative size of one group is 0.5.
- **C** : One group has absolute majority.

A) *No group has absolute majority*

Let's consider first the case of a uniform distribution, that is when for each group $\pi_i = 1/n$ (A1). When the population is uniformly distributed across groups, the effective power of each group is equal to $1/3$ for any $\alpha \geq 0$. This result is also intuitive: when all groups have the same relative size, they contribute the same amount of "power" to all coalitions which they belong to. This result is driven by the Symmetry property of phi^n. The following figure represents ϕ^3_α for $\alpha \in [0, \infty)$. For expositional purposes we represent the cumulative effective power starting from group 1.

The blue line represents $\phi^3_\alpha(\pi_1, \Pi)$, the red line represents $\phi^3_\alpha(\pi_1, \Pi) + \phi^3_\alpha(\pi_2, \Pi)$ and the yellow line represents $\phi^3_\alpha(\pi_1, \Pi) + \phi^3_\alpha(\pi_2, \Pi) + \phi^3_\alpha(\pi_3, \Pi)$.

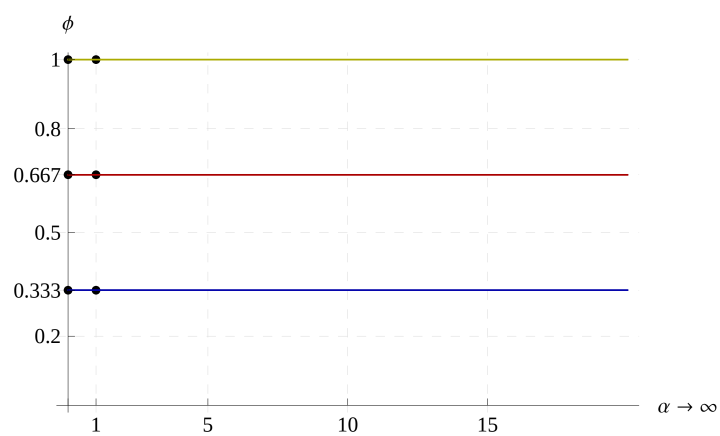


Figure 2.5: Cumulative $\phi^3_\alpha(\pi_i, \Pi)$ as a function of α; $\pi_i = 1/3$ for all i.

The situation changes when we move from $A1$ to $A2$. Figure 2.7 shows a cumulative ϕ^3_α when the population is distributed according to $\Pi = (0.45, 0.35, 0.2)$. When $\alpha = 0$ the effective power of each group is $1/3$. On the other hand, for $\alpha = 1$, the groups' effective power equals their respective population shares. As α approaches infinity, the effective power of each group converges to $1/3$. So, in the limit, the distribution of power across groups is identical to the case of a uniform distribution of population across groups (A1).

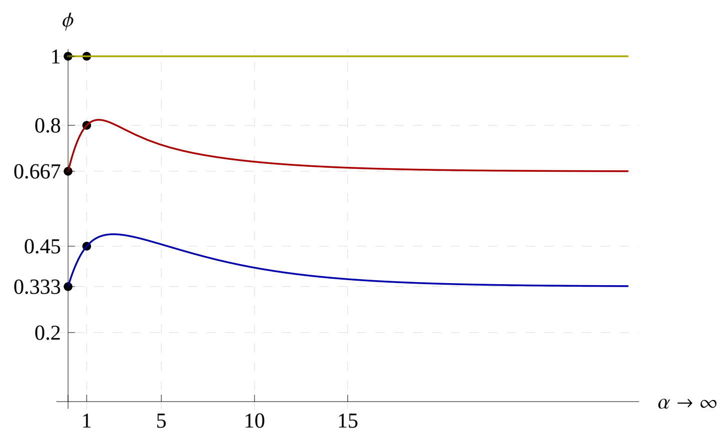


Figure 2.6: Cumulative $\phi^3_\alpha(\pi_i, \Pi)$ as a function of α; $\pi_i < 1/2$ for all i.

b) *The relative size of one group is 0.5*

Suppose now that the share of population belonging to one group, say group 1, is exactly a half. Consider the following distribution: $\Pi = (0.5, 0.3, 0.2)$.

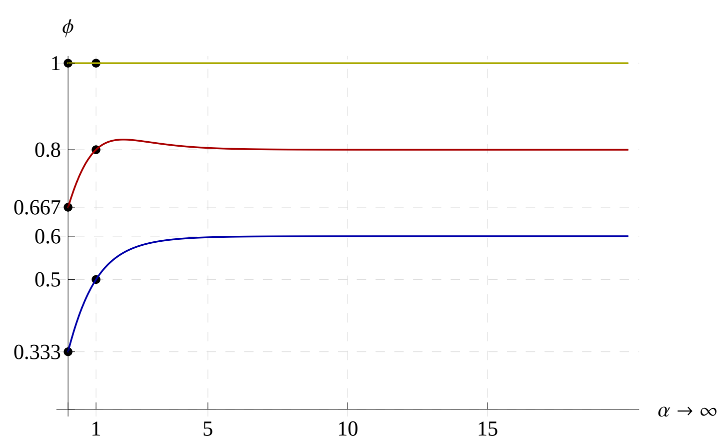


Figure 2.7: Cumulative $\phi^3_\alpha(\pi_i, \Pi)$ as a function of α; $\pi_1 = 1/2$.

For $\alpha = 0$ and $\alpha = 1$, the groups' effective power equals $1/3$ and π_i respectively. When $\alpha \to \infty$ the effective power of the biggest group (group 1) converges in limit to $3/5$, while the remaining two converge to $1/5$. The effective power of a group with $\pi = 1/2$ is three times higher than the effective power of the remaining two groups. Since effective power is given as the relative sum of marginal contributions, a group with population share of $1/2$ contributes three times more relative to the rest. This result is also intuitive: *any* coalition interested in reaching the absolute majority, *must* involve a group with $\pi = 1/2$.

c) *One group has absolute majority*

If the relative size of one group (say group 1) is greater than $1/2$, then the effective power of the numerically predominant group converges to 1 as α approaches infinity and the effective power of the remaining two groups vanishes. This result does not depend on how far is the relative size of the predominant group from $1/2$. In a situation with a hypothetical ethnic voting, a group with $\pi > 1/2$ can ensure a victory alone without relying on the "help" of other groups. So, its effective power is absolute and the "opposition" does not have any legal way to reach a majority.

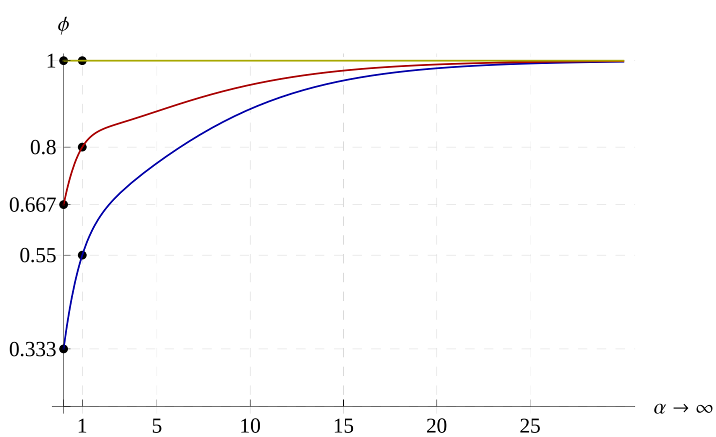


Figure 2.8: Cumulative $\phi^3_\alpha(\pi_i, \Pi)$ as a function of α; $\pi_1 > 1/2$.

In order to see how the effective power of a group with a fixed relative size may vary when as a response to the variation in relatives relative sizes of other groups, note that among all distributions that we considered up to now, the relative size of group 3 is constant and equal to 0.2. In the case in which the relative size of all groups is lower than 1/2, the effective power of group 3 converges in limit (as $\alpha \to \infty$) to 1/3; in a distribution with $\pi_1 = 0.5$, it is proportional to its population share while in the case in which group 1 is numerically predominand, the effective power of group 3 is equal to 0. We can conclude that as $\alpha \to \infty$, the effective power of group 3 varies as a function of the relative sizes of groups 1 and 2.

The Case of four groups

As in the case of three groups, we classify the population distributions into three different categories: no group has absolute majority with distinction between uniform and non-uniform distribution, one group has exactly a half of the population and one group has absolute majority.

When all the groups have the same relative size, their effective power is 1/n for any $\alpha \geq 0$. Consider now a non-uniform distribution in which no group has absolute majority. Let's start with $\Pi = (0.34, 0.31, 0.2, 0.15)$:

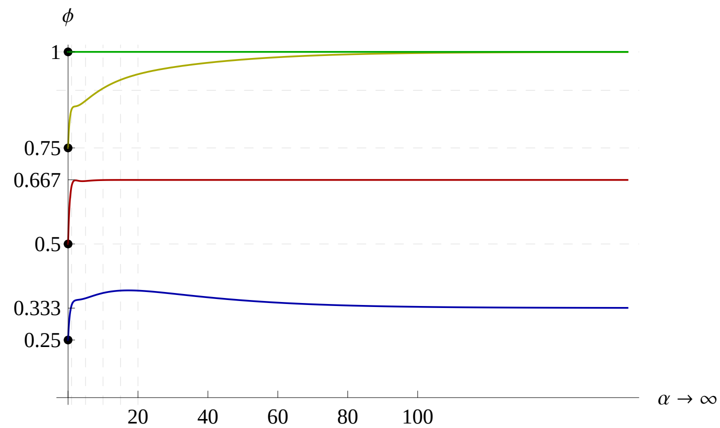


Figure 2.9: Cumulative $\phi^3_\alpha(\pi_i, \Pi)$ as a function of α; $\Pi = (0.34, 0.31, 0.2, 0.15)$.

We can see that as α approaches infinity the effective power of the smallest group converges to 0 while the power of the remaining three groups converges to $1/3$. This means that in the limit the smallest group in the population is powerless, although its relative population size is not negligible. The smallest group is not relevant in any coalition because, given the distribution of population across the remaining three groups, the relative population size of group 4 is not sufficiently high to bring any other group above the threshold level of 50%. However, if π_1 increases and either π_2 or π_3 decresaes, then group 4 may gain some power if the increase in group 1's relative population size is big enough to reach the 50% of the population when combined with group 4. In the next figure we illustrate a situation where this is the case even if π_4 decreases to 0.1.

Consider the following distribution: $\Pi = (0.45, 0.25, 0.2, 0.1)$:

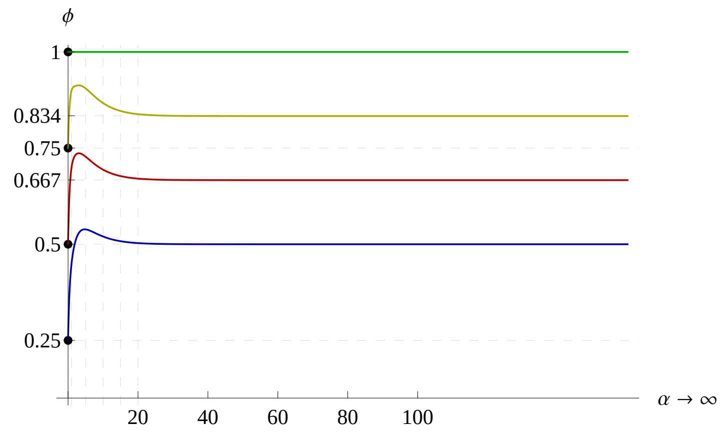


Figure 2.10: Cumulative $\phi^3_\alpha(\pi_i, \Pi)$ as a function of α; $\Pi = (0.45, 0.25, 0.2, 0.1)$.

As $\alpha \to \infty$, the effective power of the biggest group converges to 0.5, while the power of the remaining three groups converge to 1/6. It is worth noting here that a increases, the effective power may change non-monotonically.

b) *The relative size of one group is 0.5.*

Let's consider the following distribution: $\Pi = (0.5, 0.25, 0.15, 0.1)$:

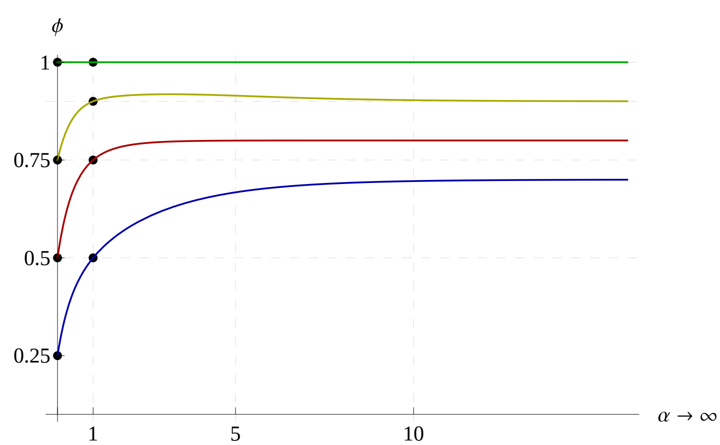


Figure 2.11: Cumulative $\phi^4_\alpha(\pi_i, \Pi)$ as a function of α; $\pi_1 = 1/2$.

The effective power of a group with $\pi = 1/2$ converges to 0.7 while the effective power of the remaining three groups converges to 0.1.

c) *One group has absolute majority.*

Suppose now that one group (say group 1) has absolute majority:

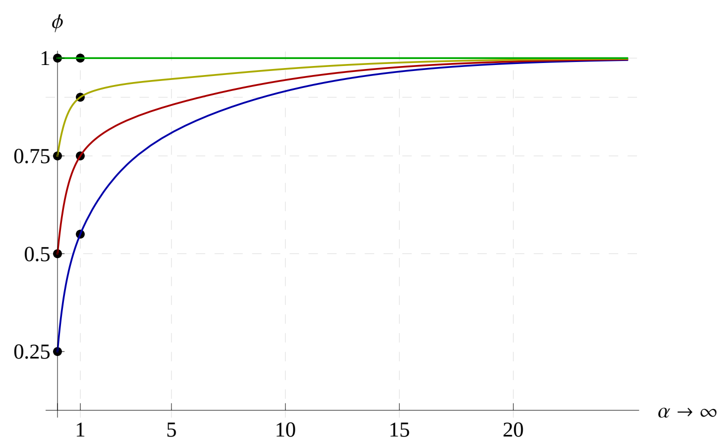


Figure 2.12: Cumulative $\phi^3_\alpha(\pi_i, \Pi)$ as a function of α; $\pi_1 > 1/2$.

As in the case of three groups, the power of the majoritarian group converges to 1 while the power of the remaining groups vanishes.

2.1.4 Effective Power as Decisiveness

The results from the above graphical inspection suggest that the effective power is *not* necessarily proportional to the groups relative size. This result is in line with the literature on *voting power*. The term voting power refers to an index that captures the power of a voter to influence the outcome of a voting process. Higher power means higher number of voting configurations in which an agent can change the outcome of voting by changing his or her vote from "yes" to "no" and *viceversa*. In his famous critique of the practice of assigning voting weights proportional to the numbers of citizens in different legislative bodies as a means of implementing the "one man, one vote" requirement, Banzhaf (1965) proves that "voting power is not (necessarily) proportional to the number of votes a legislator may cast", and that "the number of votes is not even a rough measure of the voting power of the individual legislator". Voting power, hence, in contrast to the number of votes an actor possesses, is the ability of an actor to influence the outcome of voting in a collectivity.

Our results for the case of three and four groups are in line with Banzhaf (1965). As $\alpha \rightarrow \infty$, the effective power of groups converges to their respective *relative Penrose-Banzhaf index of voting power* in a simple majority game:

Proposition 1

When $\alpha \rightarrow \infty$ and $\beta = 1$, the groups' Effective Power is given by their respective relative Penrose - Banzhaf Index of Voting power in a simple majority game, μ_i:

$$\phi^n_\infty(\pi_i, \Pi) = \mu_i, \quad \forall i, j \in N; \quad i \neq j. \tag{2.9}$$

(Proof in Appendix)

A simple majority game is a pair (N, v), where $N = (1, 2, ..., n)$ is the set of players and $v : 2^N \rightarrow \{0, 1\}$ is the charcteristic function which satisfies $v(\emptyset) = 0$, $v(N) = 1$ and $v(S) \leq v(T)$ whenever $S \subseteq T$. A coalition is winning if $v(S) = 1$, and coalitions with $v(S) = 0$ are called losing (Felsenthal and Machover, 1998). In a simple majority game a coalition is winning if the sum of all votes that players within the coalition may cast is higher than $1/2$ of the total votes in the population. The relative Penrose-Banzhaf Index is obtained by summing up marginal contributions of each player or group to the coalitions that s/he can theoretically belong to and dividing it with the sum of the marginal contributions of all players or groups. Since in a simple majority game a coalition can take only two values, namely 0 or 1, the marginal contribution of a player is 1 if it is able to switch the coalition from losing to winning and *viceversa*.

Here we do not deal explicitly with the distribution of votes across ethnic groups nor we are interested in the features of political system that characterizes a certain country. We have derived our results from two very simple and reasnable assumptions: groups form blocks in order to contrast the opponent, and the value of each block or coalition is given by a simple ratio form contest function widely applied in conflict and rent seeking literature.

Since ethnic voting is a prominent issue in almost all ethnically heterogeneous societies, the Penrose - Banzhaf index of voting power can be a good proxy for groups effective power. This is in sharp contrast with the logic underlying the existing indices of ethnic heterogeneity. Eventhough the interaction component is one of the main building blocks of discrete polarization, the RQ index implicitly assumes that there is no real interaction between groups, *i.e.* groups are assumed to be unitary actors.

Relaxing the unitary actor assumption could also expand the range of rational explanations for conflict, and one of these is the presence of strategic behavior between groups. Even in the absence of ethnic voting, strategic coalitions between groups can be observed in almost all ethnic conflicts, no matter how strong is the perception of the antagonism between them. The latter claim is in line with the so-called "opportunity-based" approach to conflict. In other words, a conflict is an industry where groups may collaborate with the "adversary" if they find it profitable. A coalition may shift during the time, *i.e.* two groups that previously were on the opposite sides may decide to join together if the conditions of the environment or the relative strenght have changed during the time. So, the logic of strategic behaviour is not related only to ethnic voting or similar political "games" but also to many other everyday situations that occur in ethnically divided societies, whether they are in conflict or not. This is particularly relevant from an economic point of view because of the well-known negative consequences of instability on economic life in general.

2.2 Properties of the P index of Conflict Potential

With the effective power function specified in (1.14), the P index of conflict potential is defined by the expression:

$$P_\alpha^n(\Pi) = K \sum_i \frac{M_i^\alpha}{\sum_j M_j^\alpha} \pi_i(1-\pi_i); \quad \alpha \in \Re_+ \cup \infty.$$

For $K = 4$ the index ranges between 0 and 1. A constant K is not only a scaling factor but it attributes to the P index a cardinal meaning.

For values of $\alpha \neq 0, \alpha \neq \infty$, the perceived power functions ϕ are continuous for $\pi \in [0,1]$. This property provides a consistency feature of $P_\alpha^n(\Pi)$ for comparisons between distributions with different number of groups n.

Consistency in groups elimination (CGE): Suppose that $\Pi = (\pi_1, \pi_2, ..., \pi_{n-1}, \pi_n)$ and $\tilde{\Pi} = (\tilde{\pi}_1, \tilde{\pi}_2, ..., \tilde{\pi}_{n-1})$ with $(\tilde{\Pi}, 0) = (\tilde{\pi}_1, \tilde{\pi}_2, ..., \tilde{\pi}_{n-1}, 0)$. Consider a sequence in Δ^n such that Π converges to $(\tilde{\Pi}, 0)$. The CGE property requires that

$$\lim_{\Pi \to (\tilde{\Pi},0)} P_\alpha^n(\Pi) = P_\alpha^{n-1}(\tilde{\Pi}). \tag{2.10}$$

In other words as one group disappears and distribution Π gets similar to $\tilde{\Pi}$ with the exception of the disappeared group, then the indices calculated for the two distributions should coincide. As we show this is the case only if $\alpha \neq 0$ and $\alpha \neq \infty$.

Proposition 2. $P^n_\alpha(\Pi)$ satisfies CGE if and only if $\alpha > 0$.

Proof of Proposition 2.

The result holds because of the continuity of $\phi^2_\alpha(\pi)$ for $\pi \in [0,1]$. Recall that $\phi^2_\alpha(0) = 0$ and $\phi^2_\alpha(1) = 1$. The continuity of $\phi^2_\alpha(\pi)$ for $\pi \in [0,1]$ is not satisfied for $\alpha \neq 0$ and $\alpha \neq \infty$.

Sufficiency part. By the construction of M^α_i, if $\phi^2_\alpha(\pi)$ is continuous, also M^α_i is continuous w.r.t. $\pi_i \in [0,1]$.

It then follows that for the group n s.t. $\pi_n \to 0$ we have that $\lim_{\pi_n \to 0} M^\alpha_i = 0$. In general, considering the remaining $n-1$ groups, by denoting with $\tilde{M}^\alpha_i$ the sum of marginal contributions for group i under $\tilde{\Pi}$ and M^α_i the one under Π we have that $\lim_{\Pi \to (\tilde{\Pi},0)} M^\alpha_i = \tilde{M}^\alpha_i$ for all $i = 1,2,3,...,n-1$. Given the continuous functional form for $P^n_\alpha(\Pi)$ w.r.t. M^α_i and the disstribution of π_i's then the CGE condition holds.

Necessity part.

We show that CGE is not always satisfied if $\alpha = 0$ or $\alpha = \infty$.

Consider the first case. The discontinuity of $\phi^2_0(\pi)$ takes place at $\pi = 0$ and $\pi = 1$. In particular $\phi^2_0(\pi) = 1/2$ for $\pi \in (0,1)$. The distribution of the marginal contributions of group n is given only by $\phi^2_0(\pi_n) - \phi^2_0(0) = 1/2$ and $\phi^2_0(\pi_n) - \phi^2_0(1-\pi_n) = 1/2$ thus for any $\pi_n > 0$ we have $M^0_n = 1$. This is not only the case for group n but also for all the other groups.

Given that $M_n^0 = 1$ we have that this is the case also for $\pi_n \to 0$, however what is required is that $M_n^0 \to 0$ if $\pi_n \to 0$. In fact $M_i^0 / \sum_j M_j^0 = 1/n$ for all groups in Π, but for all those in $\tilde{\Pi}$ we have $\tilde{M}_i^0 / \sum_j \tilde{M}_j^0 = 1/(n-1)$ thereby violating CGE.

Consider now the case $\alpha = \infty$. The discontinuity of $\phi_\infty^2(\pi)$ takes place at $\pi = 1/2$. The distribution of the marginal contributions of group n is affected when $\Pi \to (\tilde{\Pi}, 0)$ where both in Π and $\tilde{\Pi}$ there exists a set of grups (possibly the same in both distributions) whose aggregate share coincides with $1/2$. To illustrate the case consider for instance $n = 3$. Take $\Pi = (1/2, 1/2 - \pi_3, \pi_3)$ and $\tilde{\Pi} = (1/2, 1/2)$. Considering Π we get $M = (3, 1, 1)$ and thus the associated relative powers are $(0.6, 0.2, 0.2)$ this is the case even if we let $\pi_3 \to 0$. Thus CGE is not satisfied given that teh limit of $M_3^\infty / \sum_j M_j^0$ is different from 0. In fact in $\tilde{\Pi}$ we get $\tilde{M} = (1, 1)$ and thus the associated relative powers are $(0.5, 0.5)$ thereby violating CGE. ■

In the previous section we have seen how the effective power depends on the parameter α and on Π. Here, we are going to analyse the properties of the P index for different values of the coefficient α and for different population distributions. We show that for the case of two groups the parameter α plays no role and the P index reduces to the RQ index of discrete polarization which is twice the fractionalization index. When the population is split into more than two ethnic groups, the shape of the index crucially depends on the choice of the parameter α. For instance, for $\alpha = 0$ and $\alpha = 1$, the P index reduces, respectively, to the fractionalization index scaled by $1/n$ and to the RQ index of discrete polarization. As α increases, the P index departs from the RQ index and in the limit as $\alpha \to \infty$ it assumes a particular form that captures the presence of an extreme form of ethnic dominance.

2.2.1 The role of the coefficient α

The choice of the value for the coefficient α will yield a particular index of conflict potential. In what folows we consider the P index for $\alpha = 0$, $\alpha = 1$ and $\alpha \to \infty$.

- **Case 1**: $\alpha = 0$

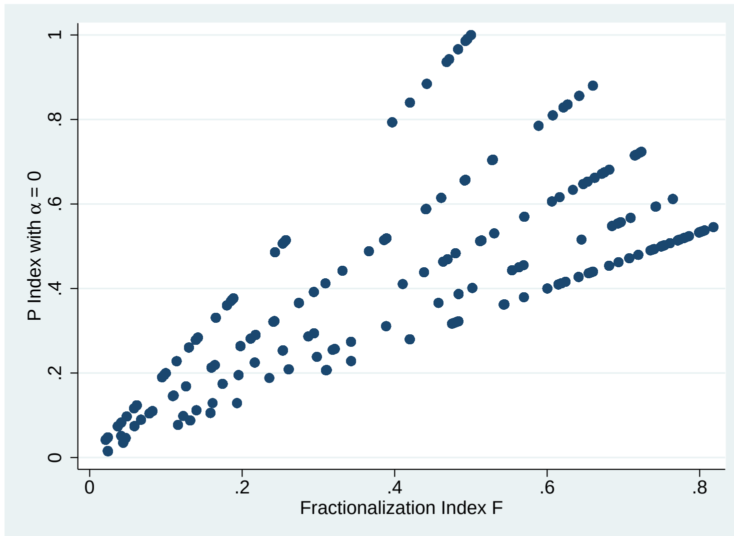
When $\alpha = 0$, the effective power of each group is constant and equal to $1/n$. This is because in (1.14) the sum of marginal contributions for each group is elevated at α, and since there are n different ethnic groups, the effective power of each of them is constant and inversely related to the number of groups in the population. With $\phi_0^n = 1/n$ for all i, the P Index of Conflict Potential becomes:

$$P_0^n(\Pi) = 4\frac{1}{n}\sum_i \pi_i(1-\pi_i) = 4\frac{1}{n} \times FRAC \tag{2.11}$$

It should be noted, however, that this is *not* exactly the fractionalization index because it is scaled by $1/n$. The fractionalization index attributes to each group a constant power which is normalized to 1. The effective power, hence, is independent of the groups' relative size and of the number of groups in the population. Consequently, the fractionalization index is shaped only by the interaction component and is defined as the probability that two individuals randomly selected from a population belong to different ethnic groups.

Here, we have a slightly different situation because the effective power assigned to each group is monotonically *decreasing* in n. It is still true that for a given n, the P_0 index and the fractionalization index provide the same ranking order. Figure 1.14 shows the scatterplot of P_0 and $FRAC$ for 141 countries from Cederman, Min and

Wimmer's (2009) *EPR* data set.



Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 2.13: *P index with* $\alpha = 0$ *versus* $FRAC$.

The relationship between P_0 and $FRAC$ assumes a particular "arrow" form, where each arrow corresponds to a certain n. The two indices are the same for $n = 4$, the steepest arrow corresponds to $n = 2$ while the flattest one corresponds to $n = 6$ (which is also the maximal number of groups considered). An interesting case occurs when all the groups have the same size. In that particular case, the P index with $\alpha = 0$ and the fractionalization index move in opposite directions. When the relative size of each group is $1/n$, the P index becomes:

$$P_0^n(\Pi) = 4\frac{1}{n}\frac{n-1}{n}$$

Despite its very simple structure, the P_0^n index has some interesting properties. In terms of the possibile relation with conflict potential it is indeed quite difficult to relate an increased probability of across group interaction to the increased conflict vulnerability. As n increases the probability of interaction increases but this does not necessarily leads to conflict because groups become smaller which reduces their chances to mobilize efficiently. Hence, there are two forces at play that should be taken into account: increased interaction versus reduced power. The index of fractionalization alone does not take this important aspect of conflict potential into account. The P_0^n index, on the other side, results much more informative: as n increases the contribution of interaction increases but it is rescaled by the power component, which decreses at a higher rate with respect to the interaction component shaping the index downwards. The results is that, with n equally sized ethnic groups, the maximum of conflict potential is reached in the case of a symmetric bimodal distribution. As n increases the value of the index converges to 0, despite the interaction component tends to infinity. In a conflict context with a continous ethnic fragmentation, interaction is weaker than power in generating conflict potential. So, the index combines a static and a dynamic component: it fully appraises the role of interaction but at the same time aligns with the logic undelying the measures of ethnic polarization, *i.e.* as n increases it puts more weight on power than on the increased interaction between groups.

Figure 1.15 shows the P index with $\alpha = 0$ (blue curve) and the fractionalization index (red curve) as a function of n where all groups have the same size[5].

[5]It should be stressed that P_0^n with $\pi_i = 1/n$ for all i behaves as the RQ index of discrete polarization.

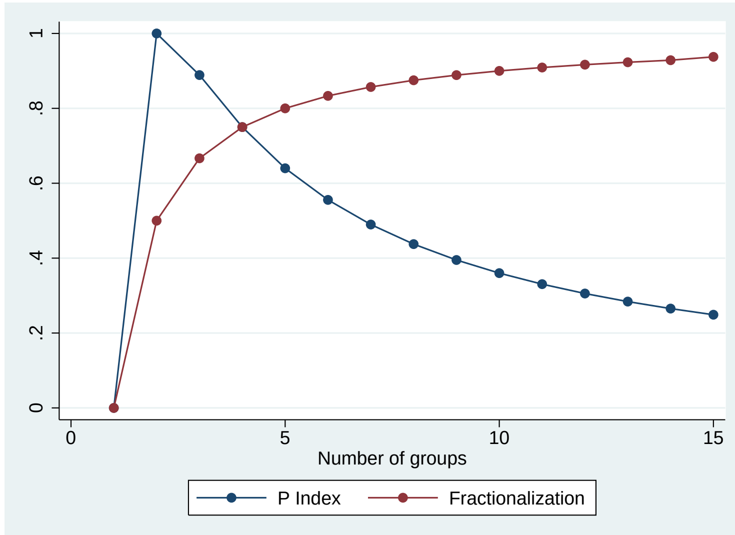


Figure 2.14: P Index with $\alpha = 0$ versus $FRAC$ index for n groups of equal size.

- **Case 2**: $\alpha = 1$

When $\alpha = 1$, the effective power of each group equals its relative population size. With $\phi_1^n = \pi_i$ for all i, the P index reduces to the RQ index of discrete polarization:

$$P_1^n(\Pi) = 4 \sum_{i=1}^{n} \pi_i^2 (1 - \pi_i) = RQ$$

The bigger is a group, the higher is its effective power to translate alienation into effective voicing. By effective voicing we mean any form of mobilization along ethnic lines or any other organized activity.

For $\alpha = 0$ and $\alpha = 1$, a group i's effective power depends only on n and π_i. In both cases, hence, Π_{-i} plays no role. The features of Π_{-i} become crucial for $\alpha \to \infty$.

- **Case 3**: $\alpha \to \infty$

As $\alpha \to \infty$, the effective power converges to the relative Penrose-Banzhaf Index of voting power in a simple majority game. Effective power of a group i is a function of *both* π_i and Π_{-i}. If we denote by π^* the relative size of the biggest group in the population and with γ_i the relative Penrose-Banzhaf Index of voting power associated to group i, the $P^n_\infty(\Pi)$ index can be written as:

$$P^n_\infty(\Pi) = \begin{cases} 4\pi^*(1-\pi^*) & \text{if} \quad \pi^* > 1/2, \\ (1-\theta_n)[4\pi^*(1-\pi^*)] + \theta_n P^n_0(\Pi) & \text{if} \quad \pi^* = 1/2, \\ 4\sum_i \gamma_i \pi_i (1-\pi_i) & \text{if} \quad \pi^* < 1/2. \end{cases}$$

where

$$\theta_n = \frac{n}{2^{n-1} + n - 2}$$

When the size of one group exceeds 1/2 the potential of conflict is determined only by that group's relative size. This is because the "opposition" is powerless. The P index with $\alpha \to \infty$ and $\pi_i > 1/2$ for some i is just the interaction component associated to a dominant group. As π of a dominant group approaches 1/2 the value of the index converges in limit to 1 (but it never reaches it). Similarly, when the size of adominant group increases, the overall interaction decreases. When no group has absolute majority the contribution of each group to the overall conflict potential is given by the product between their relative Penrose-Banzhaf index of voting power and their interaction component. Finally, with one group representing exactly a half of the population, the index is given as a convex combination between $P^n_\infty(\Pi)$ and $P^n_0(\Pi)$.

2.2.2 P Index for $n = 2$ and $n = 3$

In the case of two groups the parameter α plays no role and the P index reduces to:

$$P^2(\pi) = 4\pi(1-\pi).$$

With $n = 2$, P and $FRAC$ have identical shape, and in fact $P = 1/2 \times FRAC$. The only difference between them is their normalization. Both indices attain a maximum at a bimodal symmetric distribution, *i.e.* $\Pi = (\pi, 1-\pi) = (0.5, 0.5)$. The P index with $n = 2$ is identical to the RQ discrete polarization index (so the two curves overlap). In general, with only two groups all the indices provide the same ranking order and the only difference between them is their normalization.

Since for $n = 2$ all the indices are the same, the simplest way to look at the implications of different choices of the parameter α is to consider the case with three groups. With $n = 3$ all the indices can be expressed as a function of the relative size of two groups (since $\sum_i \pi = 1$). For expositional purposes, we decide to fix the size of one group (here group 3) to $1/3$ because i) we want to compare alternative population distributions with the uniform distribution, and ii) the RQ index of discrete polarization is *insensitive* to population transfers between groups when the relative size of one of them is fixed to $1/3$. Figure 1.17 shows P_0^3, P_1^3, P_5^3, P_{20}^3 expressed in terms of π_1.

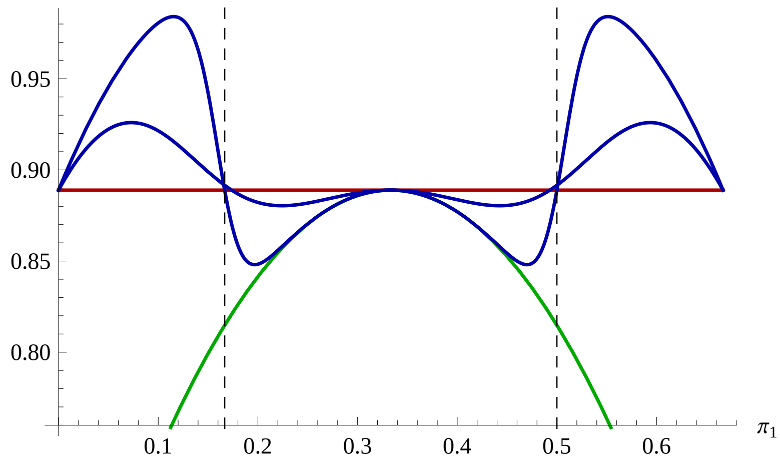


Figure 2.15: P Index for $\alpha = 0$, $\alpha = 1$, $\alpha = 5$ and $\alpha = 20$.

When all the groups have the same size, the P index is independent of α, *i.e.* it yields the same value for any $\alpha \in [0, 1]$. The P index with $\alpha = 1$ (actually the RQ index) is invariant with π_1, the shape of P_0^3 is identical to the shape of the fractionalization index while for $\alpha > 1$ the index becomes non monotonic in π_1.

As α approaches infinity, the shape of the P index becomes particularly interesting. With $n = 3$ and $\alpha \to \infty$, the P index of conflict potential is defined by the expression:

$$P_\infty^n(\Pi) = \begin{cases} 4\pi_1(1-\pi_1) & \text{if} \quad \pi_1 > 1/2, \\ \frac{2}{5} + \frac{3}{5}P_0^3(\Pi) & \text{if} \quad \pi_1 = 1/2, \\ P_0^3(\Pi) & \text{if} \quad \pi_1 < 1/2. \end{cases}$$

As we have already mentioned, with $n = 3$ and the size of one group fixed to $1/3$, the RQ index is constant and equal to $8/9$ for all Π. Montalvo and Reynal-Querol (2002) argue that ethnically polarized societies have a higher probability of being unstable and that such an instability has a negative impact on investment and, indirectly, on growth. According to this interpretation of the RQ measure, the fact that it does not vary with Π when the size of one group is fixed to $1/3$ implies that a country in which one group has an absolute (numerical) predominance should *not* be, ceteris paribus, more instable than a country in which the population is equally distributed across groups. The empirical evidence, however, suggests the opposite: one ethnic group being dominant almost doubles the risk of instability (Collier and Hoeffler, 2004). Interestingly, Montalvo and Reynal-Querol (2005) agree with Collier and Hoeffler in the sense that among all possible ethnic constellation, the one that sees a large ethnic minority facing a small majority is the worst and that their index captures well this idea. This, however, is true only for the case of three groups. Suppose, for instance that we compare two different ethnic distributions according to their conflict potential. In the first distribution there is one group that is scarsely predominant in terms of its relative size while in the second no group has absolute dominance, *i.e.* $\Pi_1 = (0.51, 0.29, 0.2)$ and $\Pi_2 = (0.49, 0.2, 0.21)$. According to the RQ index of discrete polarization, the potential of conflict associated with these two distributions is almost the same: $RQ(\Pi_1) = 0.876$ and $RQ(\Pi_2) = 0.881$ eventhough the first distribution is categorized as the worst one (small majority versus no majority).

In general, the RQ index is not sensible enough to take into account the conflict potential that derives from certain ethnic constellations. This is because the index considers power and interaction as two separated phenomena. Even before Collier and Hoeffler's empirical evidence on the importance of ethnic dominance, Horrowitz (1985) claimed that the most severe ethnic conflict will arise where a substantial ethnic minority faces an ethnic majority that can, given ethnic voting, win for sure in any national election. According to Horrowitz, hence, it is a (political) competition along ethnic lines that may, under certain conditions breed conflict. One of these conditions is the capability of one group to implement its own preffered outcome without appealing to arms or violence. Even in the absence of a predominant group, two or more groups can join forces in order to pursue some common (political or economic) interest. Traditional indices are not able to take any of these potentially important aspects of inter-group relationship into account (except for the case of two groups where the RQ index works well).

The P index with the Penrose - Banzhaf relative power, on the other hand, results much more sensible. Although it maintains a pure distributional nature and does not incorporate explicitly any additional information about the characteristics of the political system, voting rule or groups preferences, it fits quite well the Horrowitz's story. Figure 1.18 shows P_0^3 (green curve), P_1^3 (red line) and P_∞^3 *(blue curve)* expressed in terms of π_1.

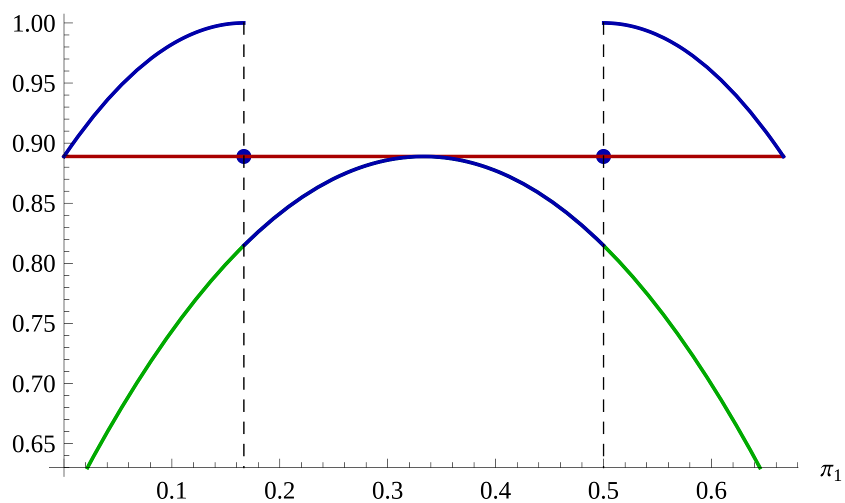


Figure 2.16: P Index for $\alpha = 0$, $\alpha = 1$ and $\alpha \to \infty$ ($n = 3$).

Starting from a uniform distribution, the P_∞^3 index follows the shape of the fractionalization index. As the size of π_1 increases, the society becomes less fragmented and the index decreases. When the relative size of group reaches $1/2$, the index "jumps" to $8/9$. Once π_1 exceeds $1/2$, the index reaches almost 1 and then decreases. The P_∞ index reaches almost one when the relative size of one group becomes scarsely higher than 1/2 because in that particular case a group in question can, given ethnic voting, win for sure in any national election. Collier and Hoeffler (2004) reasoned in terms of minority exploatation, and claim that when the size of the predominant group is scaresly higher than 1/2, the potential to exploit minority is highest and, hence its frustration in maximal. Since the minority in this case does not have access to conventional channels of achieving political change, use of arms or some other kind of combat technology is regarded a viable alternative strategy. However, as the size of the predominant group decreases, the potential of conflict also decreases. It is worth noting here that the P_∞ combines dominance (and, hence power) and interaction. Indeed, we can see that as long as we avoid the dominance, the conflict potential is entirely determined by the interation component - the shape of P_∞ follows the shape of the fractionalization index. The ability to separate these two components is particularly evident in the case of three groups. However, for any arbitrary number of groups, the P_∞ index behaves the same around and after the threshold value of .5.

Figures 1.19 and 1.20 show the P index for different values of the parameter α when the size of one group is fixed to 1/2 and 0.4 respectively.

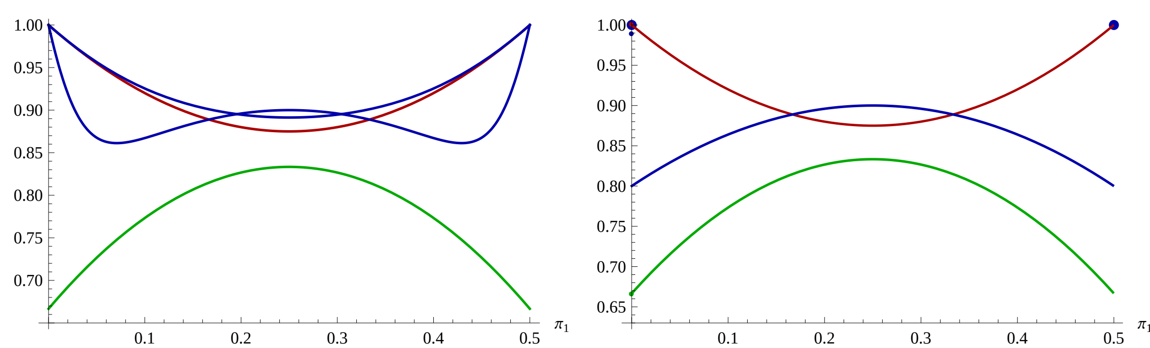


Figure 2.17: *P index for different α. Size of one group 1/2. (n=3)*

As before, when no group has absolute majority and $\alpha \to \infty$, the P index has the same shape as the fractionalization index. It attains its maximum at $\pi_1 = 0.25$ where the RQ is minimal. As the size of one group increases, the P index decreases because the interaction component is reduced and the power remains *constant* for all groups. The RQ index moves in opposite direction: as the relative size of one group increases, a distribution becomes closer to the bimodal one and the polarization increases. Similar patterns can be observed in Figure 1.18 where the size of one group is fixed to 0.4.

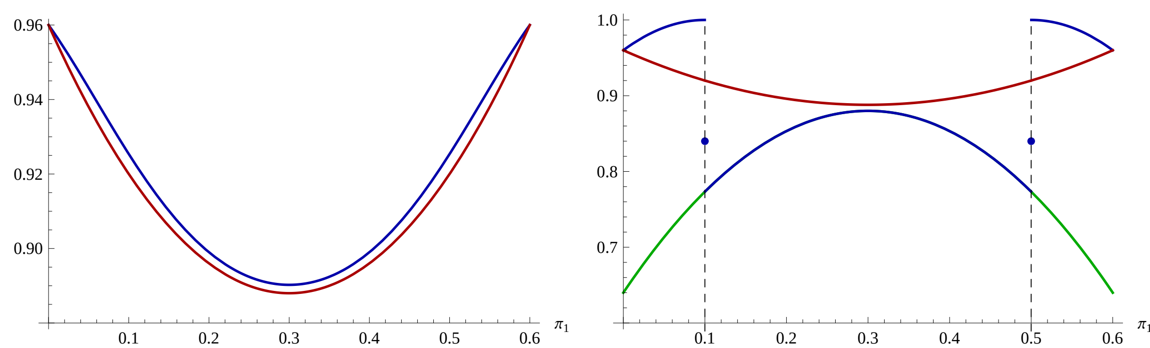


Figure 2.18: *P index for different α. Size of one group 0.4. (n=3)*

Finally, Figure 2.21 shows the shape of the P index in the unit simplex for $\alpha \to \infty$. The X and Y axis measure the size of two of the three groups. The Z axis represents the value of the index. If we compare this figure with Figure 1.1, we can see that as long as no group has one half of the population or more, the P_∞ index follows the shape of the fractionalization index. Once the relative size of one group exceeds $1/2$, the index reaches almost one and then decreases as the size of the majoritarian group increases. The area within the triangle corresponds to the upper part of the surface of the unit simplex for P^3_∞.

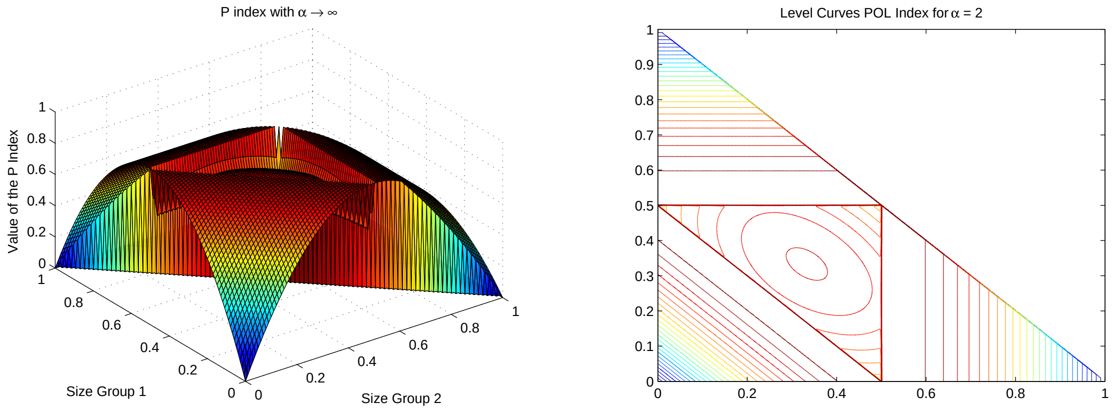


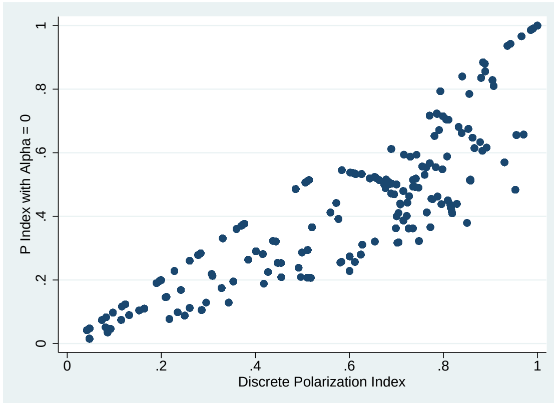
Figure 2.19: *P Index in the unit simplex for* $\alpha \rightarrow \infty$.

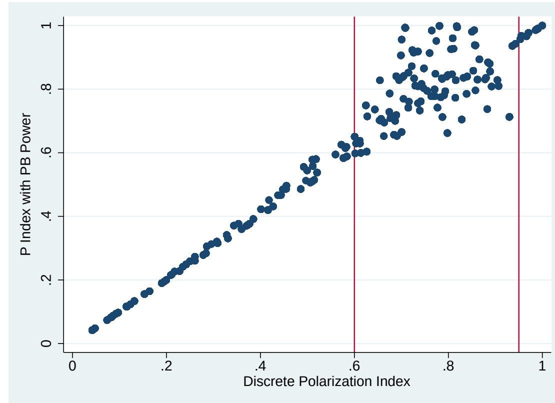
2.3 Is the *P* Index substantially different from the *RQ* Index?

Esteban and Ray (1994) define the conditions imposed by polarization using the interaction between changes in the euclidean distance of groups (defined in terms of income) and their relative size. Montalvo and Reynal - Querol (2005) redefine those conditions only in terms of groups size, since their index of polarization uses discrete distances between groups. They work with three groups, since this is the minimum number of groups that make the index of polarization different from the index of fractionalization. An index of discrete polarization with three groups should have basically three properties. The first property is that if we join the two smallest groups polarization should not decrease. The second property is that if we start from a distribution with two groups where one group is bigger than the other and we split the former into any number of sub-groups such that the biggest element is equal to the relative size of the group not affected by this partitioning, polarization should decrease. The third property is that any new distribution formed by shifting probability mass from one group equally to the other two groups must increase polarization.

The question that we want to address in this section is the following: Is the P index of conflict different enough from the RQ index of discrete ethnic polarization to be considered a radically different measure from the family of discrete polarization indices proposed by Montalvo and Reynal-Querol (2002, 2005)? In order to answer this question we procede in two steps. First, we analyse graphically the relationship between the P index and the RQ discrete polarization index using the data on ethnic distribution for 141 countries from the "Ethnic Power Relations" data set (Cederman, Min and Wimmer, 2009). Moreover, we compare the cumulative distributions of ethnic conflict onset and ethnic conflict incidence for both indices. Second, we check whether the P index satisfies the basic properties (axioms) of discrete polarization.

In order to compare the P index with the RQ index of discrete polarization, we consider the two extreme cases of our index, namely P_{∞}^{n} and P_{0}^{n}. Figure 2.22 shows the the relationship between the P index with $\alpha \to \infty$ and $\alpha = 0$ versus RQ discrete polarization index:





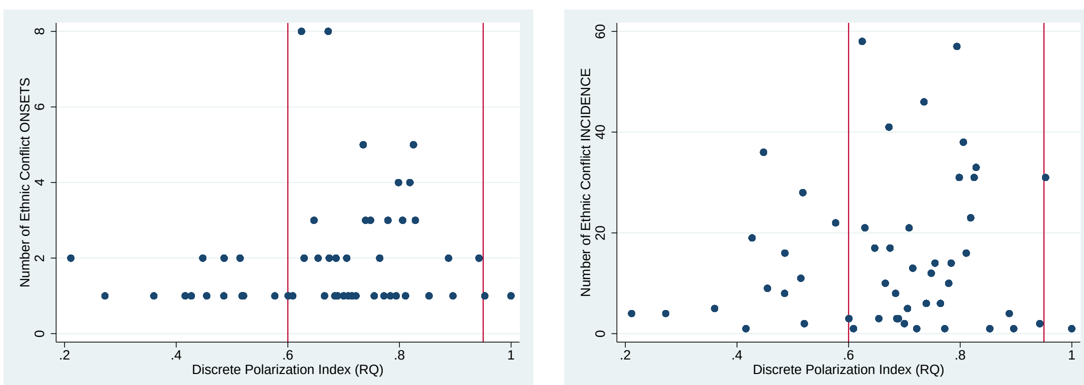
Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 2.20: *P index with* $\alpha = 0$ *and* $\alpha \to \infty$ *versus RQ discrete polarization index.*

The correlation between P_0^n and RQ is positive and relatively high (the coefficient of corrrelation is 0.88). The relationship between P_∞^n and RQ, on the other hand, is almost linear for $RQ < 0.5$ while for high values of RQ the coefficient of correlation between the two is low (0.38) indicating that there is some relationship between them but it is a weak one (this may be due to the fact that for $n = 2$ the two indices are the same). If we further restrict the range of RQ between 0.7 and 0.95 the correlation is almost 0 (0.05).

Since the RQ and the P_∞ index differ significantly only for high values of RQ, the question is whether this is sufficient to consider P_∞ as a different measure from the RQ index of polarization. This is not only a theoretical, but especially empirical issue since we want to analyse the relationship between conflict and different distributional aspects of ethnicity. Moreover, while P and RQ appear to be highly correlated against each other (for the entire interval [0,1]), quantitative enquiries aimed to test these indices against a particular conflict outcome *may* produce noticeable differences. For instance, Montalvo and Reynal-Querol (2002) directly test their measure of discrete polarization against the fractionalization index, using the same data (correlation between the two was around 0.85), and find that polarization is a significant correlate of ethnic conflict where fractionalization is not.

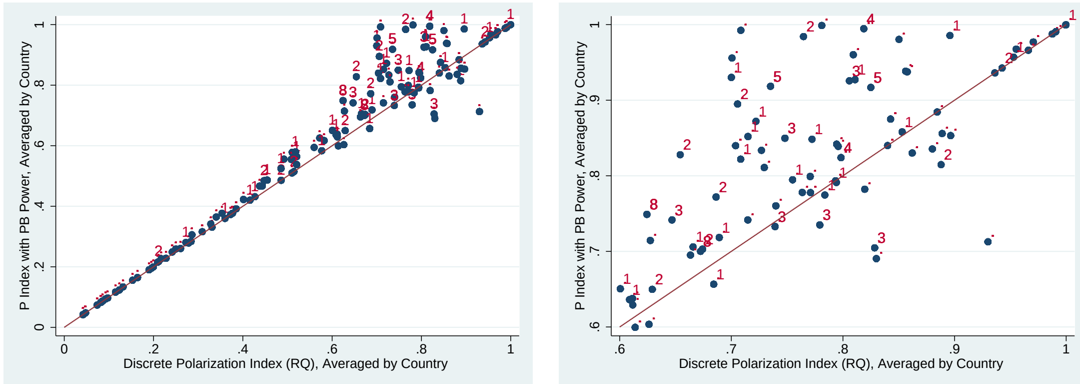
The following figure represents the frequency of ethnic conflict onset (whether an ethnic war starts in a given period) and incidence (whether ethnic conflict occurs at a given date in a given country) in relation to the RQ index (country level data). We can see that both for onset and incidence, the major concentration of ethnic conflict events lies in the range of RQ between 0.6 and 0.95, where the two indices differ most (Figure 2.24).



Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 2.21: Ethnic Conflict Onset and Incidence versus RQ Index

Since the correlation between the two indices for this range of the RQ index is very low, we can say that the P_∞ index treats differently the subsample of countries that are most conflict prone. Figure 2.23 shows P_∞ versus RQ with the labels for the quantity of ethnic conflicts [EW] onsets between 1946 and 2005 (red numbers).



Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 2.22: P_∞ *versus RQ with Ethnic War [EW] Label.*

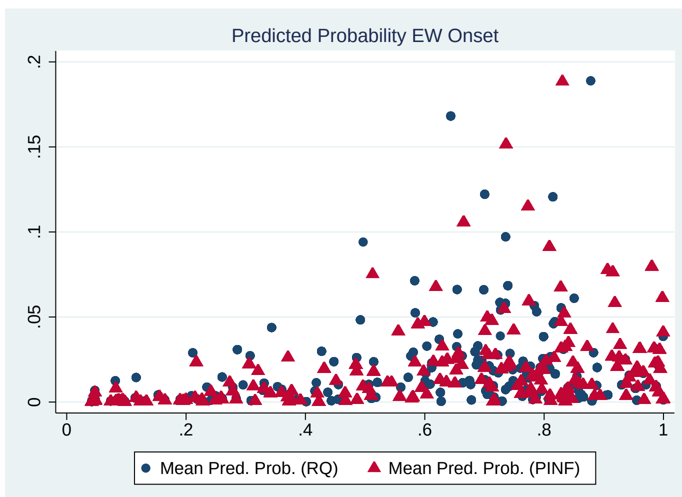
Almost all the countries with P_∞ larger than RQ have actually experience one or more ethnic conflicts during the period between 1946 and 2005. For instance, the P_∞ index for Chad and Iran is almost 1 (0.9981 and 0.9946, respectively) while the RQ index equals 0.81, as for the case of Canada and Algeria. While Canada and Algeria actually did not experience any conflict episode between 1946 and 2005, the number of ethnic conflicts in Chad and Iran was respectively 5 and 4. The very few examples of countries with the P_∞ index lower than the RQ index do not seem particularly conflict prone with 3 exceptions, namely Bosnia and Herzegovina, Sudan and Afganistan. The sensitivity of the two indices hence differs for the most conflict prone sub-sample of countries. This could make the two indices also differing in predictive power in terms of the probability of ethnic conflict outbreak.

Table 2.1 summarizes the incidence of ethnic conflict onsets for different ranges of the P_∞ index. Although P_∞ and RQ are significanlty different when the RQ index is larger than 0.6, in the following table we consider the entire range of RQ.

Table 2.1: Summary Statistics - Number of E. Conflicts P_∞ versus RQ

Index	♯ Conflicts(All)	♯ Countries	♯ Conflicts(Low)	♯ Countries
$P_\infty > RQ$	81	39	33	33
$P_\infty < RQ$	14	7	5	5
$P_\infty = RQ$	8	5	5	5
Total	103	51	43	43

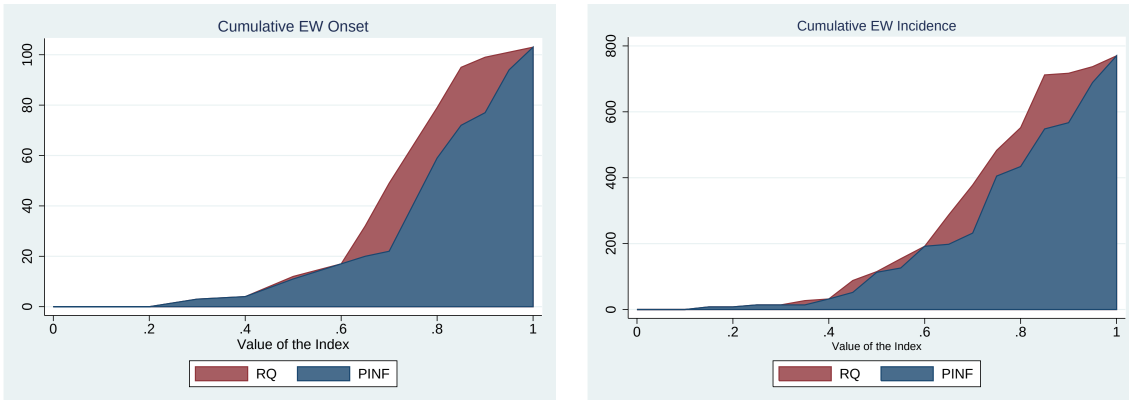
Figure 2.26 shows the predicted probability of ethnic conflict onset as a function of P_∞ (blue dots) and RQ (red triangles).



Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 2.23: Predicted Probability Ethnic Conflict Onset for RQ and P_∞.

Finally, Figure 2.25 shows the cumulative distribution function of the number of ethnic conflict onset and incidence [EW] with RQ (red area) and P_∞ (blue area). Both for ethnic conflict onset and ethnic conflict incidence, the RQ index of discrete ethnic polarization second - order stochastically dominates the P_∞^n index[6].



Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 2.24: Cumulative EW Onset and Incidence, RQ (red) versus P_∞ (blue).

In addition to the above graphical inspections, we finally check whether the P_∞ index satisfies the basic theoretical properties of the RQ index. The RQ index has basically two properties. The first property states that in a context with three ethnic groups with different population sizes[7], if we join the two smallest groups, the index should increase. Since the potential of conflict is given as a product of the effects of interaction and power, the net effect on conflict potential of merging the two smaller groups into one is positive because the aggregate effective power of the new block "multiplies" lower interaction. This property, however, holds *only* if there are iniatially three groups with different population sizes. As we have mentioned in the introductory section, the RQ index of discrete ethnic polarization is *not* sensitive to the transfers of probability mass from one group to another when the relatize size of one group is equal to $1/3$. Hence, if we start from a uniform distribution, *i.e.*

[6]However, the P_∞ index *cannot* be obtained as a tranformation of the RQ discrete polarization index.

[7]The authors work with three groups, since this is the minimum number of groups that make their measure different form the index of fractionalization.

$\Pi = (1/3, 1/3, 1/3)$, this property is only weakly satisfied because in that particular case the merge of any two groups does *not* alter the level of conflict potential in a society.

What about the P_∞ index of conflict potential? Does it satisfy this property? Although the P_∞ index does *not* assume that the effective power is necessarily increasing in the groups' size, it actually satisfies this property for any initial distribution of population Π, but only if the property is stated in a weak form. Suppose that we start from a distribution where the biggest group covers more than 50% of the population. In that particular case, the level of the P_∞ index is given by the interaction component associated to the majoritarian group irrespective of the relative size of the remaining two groups. If both the minoritarian groups merge into one, the level of P_∞ remains the same. On the other hand, when all the groups have less than 50% of the population, the merge of any two groups *must* increase the level of conflict potential because the newly formed block becomes majoritarian and its effective power is absolute. When all the groups have the same relative size, *i.e.* $\Pi = (1/3, 1/3, 1/3)$, the merge of any two groups into one does not alter the level of the index because the number of groups boils down to 2 and we know that in that case the P_∞ index is equal to the RQ index of discrete polarization.

The second property states that any new distribution formed by shifting probability mass from one group equally to the other two groups must increase RQ. Suppose that we start from a distribution with one big and two equally small groups. This property requires that if we shift probability mass from the biggest group equally to the other two groups, the value of the RQ index must increase. Obviously, this will be the case also for the P_∞ index but *only* when we the size of the biggest group is greater or equal than $1/2$ and greater than $1/3$.

In the next chapter we show that the difference between P_∞ and RQ is not only theoretical but also actual. The P_∞ index empirically performs much better than the RQ index.

2.4 Relative rather than Absolute Majority?

We formulate here a general version of the 2GRPH axiom. We essentialy require that the condition behind the original axiom holds for any number of groups, across different distributions and for any value of the relative population size between two groups. In order to formalize the property we denote by $\rho_{ij} = \frac{\pi_i}{\pi_j}$ the relative size of group i compared to group j. While $r^n(\rho_{ij}, \Pi) := \frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)}$ denotes the relative power of the two groups in the distribution of n groups. Let Δ denote the set of all Δ^n for $n \geq 2$.

Axiom 6: N Groups Strong Relative Power Homogeneity - NGSRPH.

Given $\Pi, \Pi^{'} \in \Delta$, if $r^n(\rho_{ij}), r^{n'}(\rho'_{ij}) \neq 0$:

$$\frac{r^n(\lambda\rho_{ij},\Pi)}{r^n(\rho_{ij},\Pi)} = \frac{r^{n'}(\lambda\rho'_{ij},\Pi')}{r^{n'}(\rho'_{ij},\Pi')} \; \forall \rho_{ij}, \rho_{ij}^{'} > 0, \lambda > 0, \Pi, \Pi^{'} \in \Delta.$$

Next result clarifies the implication of NGSRPH in combination with Normalization and Monotonicity.

Theorem 2.

The Effective Power $\phi^n(\pi_i, \Pi)$ of group i satisfies Axioms 1, 2, and 6 if and only if it coincides with:

$$\hat{\phi}^n_\alpha(\pi_i, \Pi) := \frac{\pi_i^\alpha}{\sum_j \pi_j^\alpha}, \text{ for } \alpha \geq 0. \tag{2.12}$$

Proof of Theorem 2.

Sufficiency part.

Note that the obtained specification for $\hat{\phi}^n_\gamma$ satisfies the axioms considered.

Necessity part.

Note that the combinations of the axioms lead to the result in Lemma 1 with $\beta = 1$ in the case of two groups ($n = 2$). In this case $r(\rho_{ij}) = (\rho_{ij})^\alpha$ for $\alpha \geq 0$ where Π is dropped from the general notation given that we consider only two groups and therefore ρ_{ij} provides all the relvant information for identifying Π. It then follows that $\frac{r^n(\lambda\rho_{ij},\Pi)}{r^n(\rho_{ij},\Pi)} = \frac{(\lambda\rho_{ij})^\alpha}{(\rho_{ij})^\alpha} = \lambda^\alpha$ for all $\rho_{ij} > 0, \lambda > 0$, for some $\alpha \geq 0$.

We can then restate the condition underlying NGSRPH as

$$\frac{r^n(\lambda\rho_{ij}, \Pi)}{r^n(\rho_{ij}, \Pi)} = \lambda^\alpha, \ \forall \rho_{ij} > 0, \lambda > 0, \Pi \in \Delta.$$

Then $r^n(\lambda\rho_{ij}, \Pi) = \lambda^\alpha \cdot r^n(\rho_{ij}, \Pi)$, by permuting ρ_{ij} with λ, given that both can be positive, we obtain $r^n(\lambda\rho_{ij}, \Pi) = {\rho_{ij}}^\alpha \cdot r^n(\lambda, \Pi)$, leading to the following funcitonal equation:

$$\lambda^\alpha \cdot r^n(\rho, \Pi) = \rho^\alpha \cdot r^n(\lambda, \Pi)$$

for all $\rho > 0, \lambda > 0, \Pi \in \Delta$. It follows that

$$\frac{r^n(\rho,\Pi)}{\rho^\alpha} = \frac{r^n(\lambda,\Pi)}{\lambda^\alpha}$$

thus $r^n(\rho,\Pi) = \Psi(\Pi,n)\cdot\rho^\alpha$ for all $\rho > 0, \Pi \in \Delta$ and $n \geq 2$.

Note houwever that Symmetry requires that $r^n(1,\Pi) = 1$. Thus $\Psi(\Pi,n) = 1$.

Recall that $r^n(\rho_{ij},\Pi) := \frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)}$ where $\rho_{ij} = \frac{\pi_i}{\pi_j}$, it then follows that

$$\frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)} = \left(\frac{\pi_i}{\pi_j}\right)^\alpha = \frac{(\pi_i)^\alpha}{(\pi_j)^\alpha}$$

for all $\pi_i, \pi_j, \Pi \in \Delta$ and $n \geq 2$. This implies that

$$\phi^n(\pi_i,\Pi) = \Upsilon(\Pi,n)\cdot(\pi_i)^\alpha.$$

Recalling the Normalization axiom we obtain $\sum_i \phi^n(\pi_i,\Pi) = \sum_i \Upsilon(\Pi,n)\cdot(\pi_i)^\alpha = 1$, that is $\Upsilon(\Pi,n) = 1/\sum_i (\pi_i)^\alpha$. The final result is therefore

$$\phi^n(\pi_i,\Pi) = \frac{(\pi_i)^\alpha}{\sum_i (\pi_i)^\alpha} \text{ for } \alpha \geq 0$$

as required.■

The derived class of Effective Power functions can also lead to an extreme case where $\alpha \to \infty$, in analogy with what obtained for ϕ^n_∞ in the results in Lemma 1 and Theorem 1.

Consider $\hat{\phi}^n_\alpha(\pi_i, \Pi) := \frac{\pi_i^\alpha}{\sum_j \pi_j^\alpha}$ then denote with $\hat{\phi}^n_\infty(\pi_i, \Pi) := \lim_{\alpha \to \infty} \hat{\phi}^n_\alpha(\pi_i, \Pi)$. It follows that denoting by $\pi^* = \max\{\pi_i\}$, we can rewrite $\hat{\phi}^n_\alpha(\pi_i, \Pi) := \frac{\left(\frac{\pi_i}{\pi^*}\right)^\alpha}{\sum_j \left(\frac{\pi_j}{\pi^*}\right)^\alpha}$, and obtain

$$\hat{\phi}^n_\infty(\pi_i, \Pi) = \begin{cases} 1/n^* & \text{if } \pi_i = \pi^* \\ 0 & \text{otherwise} \end{cases}$$

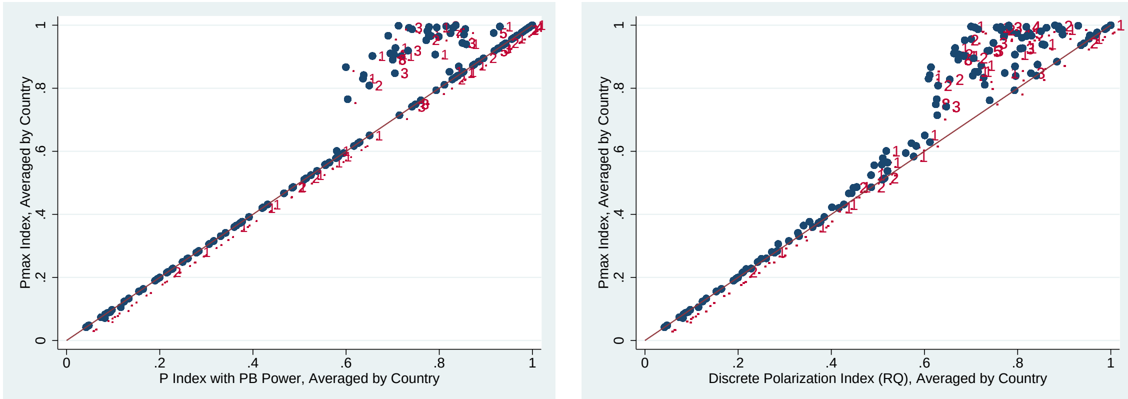
where n^* denotes the number of largest groups s.t. $\pi_i = \pi^*$.

When we combine the interaction component with the power component of the extreme case where $\alpha \to \infty$, we obtain the following index of conflict potential:

$$\hat{P}_\infty(\Pi) = 4 \sum_i^n \hat{\phi}^n_\infty(\pi_i, \Pi)\pi_i(1 - \pi_i) = 4\pi^*(1 - \pi^*). \tag{2.13}$$

In other words, the $\hat{P}_\infty$ index is given by the interaction component(s) associated to group(s) with the relative majority scaled by the number of such groups in the population. We will refer to this particular index as the P^{max} index of conflict potential. The index attains its maximum at $\pi^* = 1/2$.

The P^{max} index depends only on the size of the group with the relative majority and it is not affected by the relative sizes of all the other groups in the population. This makes the P^{max} index very different from the P_∞ index. Figure 2.25 shows the scatterplot of P^{max} versus P_∞ and RQ. The P^{max} index is equal to P_∞ and RQ in the case of two groups and when all the groups have the same population size. For all the other cases, P^{max} is larger than RQ and larger or smaller than P_∞.



Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 2.25: P^{max} *versus* P_{∞} *and RQ with Ethnic War [EW] Label.*

In the next chapter we show that P^{max} has a very weak empirical performance in the explanation of ethnic conflict onset with respect to the P_{∞} index. The weakness of the empirical performance of P^{max} is due to the fact that it is unable to account for the relative importance of different aspects of ethnic distribution in the generation of conflict potential as it was the case with the P_{∞} index which combines in a non-linear way the dominance, the polarization and the fractionalization into one single measure.

2.5 Concluding Remarks

The prominent theoretical and empirical literature on conflict does non provide a clear evidence on the role of ethnic diversity in conflict. Numerous studies have shown that the relationship between diversity and conflict is positive, negative, non monotonic or even not significantly different from zero. This variation in results may be due to the fact that there is no uniform criterion for determining a conflict episode or because there is a significant variation in data sources for ethnic diversity. In this chapter we offered another plausible explanation: the relevance of each distributional aspect of ethnic diversity may depend on the characteristics of the underlying population distribution across groups. The initial idea was to construct an index that takes this into account. This objective seems to be reached.

We proposed a new distributional index of ethnic diversity based on the general specification of the Esteban and Ray's model of social antagonism and on two simple, but crucial assumptions on groups power and across-group interaction. Although we were not interested in modelling the mechanism of coalition formation between groups nor in any other kind of groups' preferences, the results that we obtained are very informative: conflict potential is given as a weighted sum of the effects of across-group interaction and their relative effective power. Under some population distributions, the power component dominates the interaction component and generates the effects similar to the presence of an extreme form of ethnic dominance where the size of one group is scarsely higher than one half of the population. When the interaction components dominates the power component, the relevant concept of ethnicity is the fractionalization while for the intermediate case, what matters is the combination between the two. It is not important how big a group is but rather how *decisive* it can be in a hypothetical competiton between all the groups in the population. We show that a group can be powerless even when its size is not negligeable, which is in line with the literature on voting power in a simple majority games. Coalition formation and the power as decisiveness are crucial for the implications of our model. We aditionally confirm this by considering another alternative form for the effective power that relies on the relative rather than absolute majority and that does not allow any form of coalition formation between groups. The P index constructed in this way is unable to offer an insight to the interaction between different dimensions of conflict.

Our index is radically different from the existing indices of ethnic diversity eventhough it preserves a uni-dimensional nature. In this chapter we showed in detail that the P index with the Penrose-Banzhaf scores is a different measure form the RQ index of discrete ethnic polarization, at least from a theoretical point of view. In the next chapter we apply our indices to the empirical analyses of the correlates of ethnic conflict onset and we show that this difference is not only theoretical but also actual. The P index with Penrose-Banzhaf scores outperforms the existing indices of ethnic diversity and it is the only distributional index that is significantly correlated to the likelihood of ethnic conflict onset. This evidence is robust to the inclusion of an additional set of regressors, time and regional controls as well as to the alternative estimation methods.

Chapter 3

Ethnic Diversity and Ethnic Conflict: An Empirical Investigation

"Politics is more important than economics in causing ethnic civil war [...] and economic variables may be more important determinants of non-ethnic war onset"
N. Sambanis, 2001.

3.1 Introduction

Internal conflicts represent one of the main impediments for economic development. The negative consequences of conflicts are not only related to the destruction of infrastructure or human lives but also to the so-called "conflict trap" (Hegre et al., 2011): they undermine democratic political institutions, generate an uncertain environment for future investments and growth, and exacerbate the conditions that favor further insurgency. Moreover, internal conflicts are often contagious in the sense that massive refugees flow, diseases, lawlessness and the illegal trade in drugs, arms and minerals generate spillover effects into the countries neighboring conflict zones. The negative economic consequences of internal conflicts and their contagious effects are so important that can be considered as one of the explanatory factors in the growing income gap between the world's richest and poorest countries or regions (Collier et al, 2003). In order to prevent future conflicts, the determinants and causal mechanisms linking different social and structural factors to conflict must be identified.

Over the past decade many economists and other social scientists have worked to better understand the causes and the economic and social legacies of internal warfare. According to Cederman, Min and Wimmer (2009), the existing literature on conflict (mostly quantitative) may be classified into three main categories: Opportunity - based Approach, Ethnic and Religious Diversity - based Approach, and Minority - Mobilization - based Approach. According to the first strand of literature, the main correlates to civil conflict are structural conditions and country characteristics. This approach is based on a simple "cost-benefit" logic: it assumes that individuals (or groups) will participate in a rebellion if the expected payoff outweights the associated cost of such an engagement. In other words, a conflict is treated as any other economic activity in which individuals compare risks and benefits and then rationally choose whether to undertake the action (in this case rebellion) or not. Within the opportunity based aproach, the Fearon and Laitin's (2003) "insurgency model" and Collier and Hoeffler's (2004) approach can be considered as foundational. According to Fearon and Laitin, the main factors that increase the likelihood of conflict outbreak are poverty, large population, weak institutions and the share of the country's mountainous terrain. What matters for conflict is, hence, whether rebels can hide from the governmetal forces and retreat and whether economic opportunities are so poor that the life of a rebel becomes attractive[1]. In other words, a low opportunity

[1] There is a slight variation in the interpretation of the per capita income among scholars. Fearon and Laitin (2003) emphasize its correlation with state capacity, while Collier and Hoeffler (2004) link it to the opportunity costs facing potential rebels.

cost for rebellion increases the likelihood of confict. Other measures of grievances, such as vertical economic inequality (measured by the Gini Index), lack of democracy and state discrimination against minority groups are not as strong as structural conditions and country characteristics. Collier and Hoeffler (2004) emphasize the role of financial availability for rebellion activities, such as primary commodity exports or financial resources from diasporas. Regarding the cost of rebellion, the authors find that secondary education enrollment, growth rate and per capita income all have substantial and statistically significant effects that reduce conflict risk. Similarly, several other authors have stressed the positive impact of natural resources endowment for the opportunity for rebellion (Lujala, 2010; Lujala, Rod and Thieme, 2007).

With the exception of Collier and Hoeffler (2004), the opportunity based approach dismisses the role of ethnic and religious diversity in the explanation of internal conflict by arguing that ethnic grievances are too widespread to explain the rare event of internal conflict. The diversity based approach, on the other hand, claims that ethnicity *does* matter. According to this strand of literature, societies that are more ethnically heterogeneous are also more likely to experience internal conflict. The impact of ethnic and/or religious diversity is captured by several demographic indices of heterogeneity that rank different population distributions according to their conflict potential. All these measures are based on an a priori assessment of what aspect of population distribution across ethnic or religious groups is particularly conflict prone. There are three main manifestations of diversity: fractionalization, polarization and dominance. According to fractionalization, societies that are more fragmented face higher risk of conflict. The likelihood of conflict is, hence, monotonically increasing in the number of groups in the population (Sambanis, 2004; Hegre and Sambanis, 2006). Collier (2001) and Collier and Hoeffler (2004) proxy "social fractionalization" in a combined measure of ethnic and religious fractionalization. They calculate social fractionalization index as the product of the ethno-linguistic fractionalization and the religious fractionalization index plus the ethno-linguistic or religious fractionalization index. They show on the contrary to what expected by other works, that countries which are more fragmented from both the ethnic and the religious point of view face lower risk of internal warfare. One possible explanation for this result, according to the authors, is the fact that in highly fragmented societies the coordination between groups becomes more difficult and since groups are relatively small, the degree of intra-group cohesion is lower. Moreover, applying their dominance dummy (that takes the value of one if one single ethno-linguistic group makes up from 45 to 90% of the total population and zero otherwise) together with social fractionalization, they show that the risk of conflict is doubled with one group being numerically predominant. Similar evidence is offered by Schneider and Wiesehomeier (2008) (although their approach is substantially different from that of Collier and Hoeffler, 2004). Several other scholars assume that the relationship between ethnic and religious heterogeneity and conflict is not monotonic. Montalvo and Reynal-Querol (2005) apply the index of discrete ethnic polarization instead of fractionalization and find a significant and robust association between ethnic polarization and the incidence of civil conflict. This evidence is in line with the Horrowich's (1985) claim that highly homogeneous and highly heterogeneous societies are less conflictual with respect to those characterized by a few prominent ethnic groups.

A third stream of research studies the conditions under which minorities will mobilize against the state and under which conditions such mobilization will turn violent (Cederman, Min and Wimmer, 2009). This approach emphasizes the role of the state as not an ethnically neutral institution but an active agent of political exclusion along ethnic lines that generates tensions between competing and/or excluded and disriminated groups and the state. Cederman, Min and Wimmer (2009) show that once we account for the political dynamics of ethnic exclusion, discrimination and political competition, distributional aspects of ethnicity (fractionalization and polarization) have no effect on the likelihood of ethnic and civil conflict in general. This, however, does not mean that ethnicity does not play any role. Rather, ethnicity matters because the nation-state itself relies on ethno-national principles of political legitimacy: the state is ruled in the name of an ethnically defined people and rulers should care for their "own people" (Cederman, Min and Wimmer, 2009). This lack of universal inclusion is more likely in poor states as well as in states with weak civil institutions (which has some ressemblance with Fearon and Laitin's Insurgency model). In those states there are no alternative channels to pursuit national or ethnic interests, but it relies on dominance and forced exlusion of other "rivals" from desicion making process. Moreover, depending on the configuration of political power, similar political institutions can produce different consequences, while similar consequences can result from different constellations of power.

Our approach borrows from both the Diversity and the Minority - Mobilization Approach.[2] We fully agree with Cederman, Min and Wimmer (2009) that ethnic exclusion and political competition are important factors that breed conflict. However, we reject their claim that distributional aspects of ethnic diversity have no effect on the likelihood of conflict. We argue that the combination between certain ethno-political configurations and ethnic constellations may substantially increase the potential of conflict in a society. Distributional features of ethnicity may still be important, even after controlling for the quality of institutions, country characteristics and structural conditions as well as for political exclusion and discrimination along ethnic lines. The risk of conflict may still depend upon the characteristics of population distribution across groups, even in the presence of the "best" economic and political environment. Moreover, there may exist some empirical regularites between ethnic composition and ethnic exclusion. As previously mentioned, in poor states with weak social institu-

[2]We however control for several structural conditions and country characteristics variables from the "Opportunity-based Approach".

tions, the only way to control the state is to supress the opposition. So, there could be some relationship between the degree of political exclusion (defined in terms of the size of the excluded population) and the size of the ruling group.

The aim of this chapter is to investigate all these issues and in particular to provide an answer to our starting question: Are distributional aspects of ethnicity important for the explanation of conflict outbreak or is ethnic struggle driven exclusively by other factors suggested by the "Opportunity" and the "Minority mobilization" school of thought? We consider only "ethnic conflicts", *i.e.* conflicts "over ethno-national self-determination, the ethnic balance of power in government, ethno-regional autonomy, ethnic and racial discrimination, and language and other cultural rights" (Cederman, Min and Wimmer, 2009, p.326). We concentrate only on ethnic conflicts for three reasons. *First*, ethnic wars made up for the largest number of civil wars since World War II. Moreover, since the end of the Cold War, the share of ethno-nationalist wars reached 75%. Ethnic demands and grievances, hence, played a prominent role in most recent conflicts. *Second*, political mobilization mostly occurs along ethnic lines. In fact, political leaders appeal to the ideal self-rule and fair representation to mobilize their followers against the threat of ethnic dominance by others (Wimmer, Cederman and Minn , 2009). Similarly, Sambanis (2001) identifies significant differences between the determinants of ethnic and non - ethnic civil wars. Since ethnic and non - ethnic wars differ in term of both motivations and final objectives, it is very likely that they have different causes. Sambanis (2001) shows that ethnic wars are predominantly caused by political grievance and they are unlikely to occur in politically free, *i.e.* democratic societies, while non-ethnic conflicts are mostly driven by economic motives. *Third*, the data set that we use to construct our indices identifies all *politically relevant* ethnic categories around the world and records changes in politically relevant categories over time. The latter point makes our decision to focus on ethnic conflicts only even more plausible, and fits quite well with the structure of our measure of ethnic diversity.

We procede in two steps. First, we compare the empirical performance of the P index of conflict potential with the existing indices of ethnic diversity as well as with several dominance dummy variables in the explanation of ethnic conflict *onset*. In the previous chapter we have showed that there is substantial difference between the P index with Penrose-Banzhaf relative scores (P_∞) and the RQ index of discrete ethnic polarization (P_1) and we have offered a first insight into the empirical relationship between the two. Here we show that the difference between the RQ and the P_∞ index is not only theoretical but also actual. The P index with Penrose - Banzhaf relative scores performs better than all the existing indices of ethnic diversity in the explanation of ethnic conflict onset. Moreover, we show that the goodness of fit for the family of P indices (P_α) is an increasing function of the coefficient α. Second, we show that the distributional aspects of ethnic diversity *cannot* be discarded even after controlling for ethnic politics variables suggested by Cederman, Min and Wimmer (2009) and the dominance dummy variables from Collier and Hoeffler (2004) and Schneider and Wiesehomeier (2008). We confirm empirically our theoretical claim that the P_∞ index is a result of the combination between interaction and dominance (power). Although ethnic dominance and ethnic fractionalization are jointly significant, they do not "survive" the inclusion of the P_∞ index. For all model specifications we use the standard pooled logistic regression with robust standard errors clustered by country. Since ethnic conflict is a rare event, we also perform a rare event logistic estimation. In order to control for the independence of observations both across time and across countries we perform a random effects logistic estimation. We also test our results with the "low intensity ethnic conflicts" as a dependent variable and the main results do not change. Finally, we compare the empirical performance of the P_∞ index against an alternative indicato, the P^{max} index. Our results provide a *strong* *empirical support* for the P_∞ index of conflict potential as a robust correlate to ethnic conflict onset.

This chapter is structured as follows. In the next section we present our main data sources and the related econometric issues. We then present the empirical relationship between the P index of conflict potential for different values of the coefficient α and the fractionalization index. Section 4 presents our main results and section 5 concludes.

3.2 Data Sources and Econometric Issues

3.2.1 Data on Ethnic Diversity

Despite the large number of studies dealing with the implications of ethnic diversity on conflict and development, there are only few sources for data on ethnic groups that are commonly used. In literature there are mainly three sources of data on ethnic groups: the World Christian Encyclopedia, the Encyclopedia Britannica and the Altas Narodov Mira (1964). All of them use different definitions of ethnicity and different approaches. The Encyclopedia Britannica, for example, uses a concept of geographic race while the World Christian Encyclopedia takes into account linguistic and religious differences between groups. The Altas Narodov Mira was the main source for the index of ethnolinguistic fractionalization commonly used in studies on ethnic conflicts. Some authors have made the effort to develop an own data collection on ethnic diversity. Vanhanen (1999), Fearon (2003) and Montalvo and Reynal-Querol (2005), for example, state that only the most important ethnic divisions of a country should be taken into account for reasons of simplicity and clarity.

For the purposes of our analysis, probably the most appropriate data set is the "Ethnic Power Relations" [EPR henceforth] data set recently provided by Cederman, Min and Wimmer (2009)[3]. The *EPR* data set identifies all *politically relevant* ethnic categories around the world and measures access to executive-level state power for members of these ethnic categories in all the years from 1946-2005. The data sets includes 155 sovereign states with a population of at least one million and a surface area of at least 500 square kilometers as of 2005. In collecting the data, the authors relied on the expert input of nearly one hundred students of ethnic politics to assess formal and informal degrees of political participation and exclusion along ethnic lines. The coding of ethnic groups did *not* limit the possibilities to any existing ethnic group list, such as Minority at Risk Data Set [MAR henceforth] developed by Gurr and his collegaous (Gurr et al. 1993; Gurr 2000)[4], *Atlas Nardov Mira* (1964) or Fearon (2003). The coders were asked not to exclude groups based only on their relative population size, since even small groups can be significant at the national or regional level. In addition to the ethnic coding, the degree of access to central level state power for representatives of each group for each time period was determined. The access to power is reported in absolute terms, *i.e.* the data set does not provide information on under or over representation of groups relative to their population size. This data set improves significantly on previous efforts to code ethnic groups' access to power, such as Cederman and Girardin (2007), who rely on static measures of inclusion and limit their sample to Eurasia and North Africa.

The Definition of Ethnicity

In order to define "ethnicity" the authors follow the Weberian Tradition and define ethnicity as "a subjectively experienced sense of commonality based on a belief in com-

[3]http://dvn.iq.harvard.edu/dvn/dv/epr

[4]http://www.cidcm.umd.edu/mar/data.asp

mon ancestry and shared culture" (Cederman, Min and Wimmer, 2009). There are many different markers that may be used to indicate common ancestry: common language, adherence to the same religion or faith, similar phenotypical characteristics.[5] The authors' definition of ethnicity includes ethnolinguistic, racial and ethnoreligious groups but not tribes and clans "that conceive of ancestry in genealogical terms" (Cederman, Min and Wimmer, 2009). Moreover, ethnic groups may be hierarchically nested and contain several levels of differentiation, not all of which are politically relevant (Wimmer, 2008).

An ethnic category is politically relevant if at least one significant political actor claims to represent the interests of that group in the national political arena, or if members of an ethnic group are sistematically and intentionally discriminated against in the domain of public politics[6]. By "significant" political actor they mean a political organization, which is not necessarily a party, that is *active* in the national political arena. One group is discrimated against if there is an intentional political exclusion of the entire ethnic community from decision making, either at the national or at the regional level. The authors hence do not rely for instance on a possible educational disadvantages or discrimination in the labor or credit markets. An ethnic category hence becomes politically relevant as soon as there is a *minimal* degree of political mobilization or intentional political discrimination along ethnic lines. The coding of politically relevant ethnic groups is close to the MAR project but it also provides

[5]A phenotype refers to observable characteristics or traits, such as the individuals physiological features and behavior. In general, a phenotype results from the expression of an individual's genes as well as the influence of environmental factors and the interactions between the two.

[6]The access to central state power of those who claim to represent an ethnic group is classyfied into three general classes: absolute power, power sharing regimes, and exclusion from central power. In the case of the absolute power, the political elites do not significantly share the power with other political representatives and there are two sub-types of absolute power: monopoly (other representatives completely excluded) and dominance (there is some limited inclusion of members of other ethnic groups). Power sharing regimes are those that divide the power between the representatives of different ethnic groups with the distinction between *senior* and *junior* partners in a coalition (which in turn depends on the relative importance of positions held by a certain ethnic representative). For more details see the *Online Appendix* in Cederman, Min and Wimmer, 2009.

coding for majoritarian and dominant groups and it does not restrict attention only to politically excluded minorities (as it is the case in MAR profect).

Since politically relevant ethnic categories may change over time, the authors divided the time period and provided separate codings for each subperiod. The list of ethno-politically relevant categories, in some cases, was changing from one sub-period to another, either because certain categories ceased to be or became relevant for the first time, or because higher or lower level of ethnic differentiation became salient (Wimmer, Cederman and Min (2009): 326).

The Restricted *EPR* Data Set

We restrict the sample to 141 countries for two reasons. First, in some countries like Tanzania or Democratic Republic of Congo, the data on ethnic composition were too disaggregated or incomplete. Second, given the complex structure of the P index of conflict potential, we were forced to consider countries with no more than 6 ethnic groups[7]. However, we were particularly careful in deciding which countries to include into the analysis. We first ranked all ethnic groups in descending order according to their relative population size, and then choose the first six biggest ethnic categories. Countries in which the number of ethnic groups was more than 6 and the sum of the population sizes of groups ranked below the sixth biggest ethnic group was exceeding 10% of the population, were excluded from the analysis. Moreover, the relative population size of each potentially excludable ethnic category could not be substantial (not more than 5%). In such a way the marginal impact of excluded groups on the value of our indices of ethnic diversity is minimized. Moreover, the number of countries for which we were forced to "eliminate" some ethnic groups is

[7]The average and the median number of groups in Cederman, Min and Wimmer is 4.5 and 3 respectively.

low. As a result we consider 6544 out of 7155 observations from the original data set, which represents a reduction of only 8.5%.

3.2.2 Data on Ethnic Conflict

The difference between conflict data sets concerns the question of fatality thresholds for inclusion. There are two main approaches to fatality thresholds. First, the "Correlates of War" [COW] project sets the threshold for a civil war at 1000 battle deaths for the entire civil war. Using this criterion, the COW data set identifies 214 conflict episodes between 1816 and 1997[8]. Second, the Uppsala/PRIO "Armed Conflict Dataset" [ACD][9] (Gledish et al., 2002) defines armed conflict as any armed and organized confrontation between government troops and rebel organizations, or between army fractions, that reaches an annual battle death threshold of al least 25 people. Massacres and genocides are excluded because the victims are neither organized nor armed. Communal riots and pogroms are also excluded because the government is not directly involved.

Cederman, Min and Wimmer (2009), however, extend the ACD data set by coding each conflict for whether rebel organizations pursued ethnonationalist aims and recruited along ethnic lines. The authors identify as "ethnic" [...] "the aims of achieving ethnonational self-determination, a more favorable ethnic balance of power in government, ethnoregional autonomy, the end of ethnic and racial discrimination, language and other cultural rights" (Cederman, Min, Wimmer (2009): 326). All other wars are defined as *non-ethnic*. Examples of non-ethnic conflict include the various military coups in Argentina and the civil wars in China, Greece and Algeria. For a conflict to be ethnic, hence, armed organizations have to both explicitly pursue ethnonationalist

[8] A new version of the data set is now available for the time span that goes from 1816 up to 2007.
[9] http://www.prio.no/Data/

aims, motivations, and interests and recruit fighters and forge alliances on the basis of ethnic affiliations. The data set is based on a pooled time series that contains country-year observations coded as one if an ethnic war started within that observation and as a zero for all other cases. The authors identify 215 armed conflicts fought between 1946 and 2005, 110 of which were ethnic conflicts.

For the purposes of our analysis, the most suitable data set for conflict coding is the Uppsala/PRIO "Armed Conflict Dataset" (ACD) dataset with Cederman, Min, and Wimmer's (2009) coding for ethnic conflict. The reason why we decided to use this data set is two-fold. The first motivation is related to the fact that internal conflicts are *not* a homogeneous phenomenon. Conflicts differ both in objectives and motivations and are likely to have different causes (Sambanis, 2001). Since our main aim is to understand the potential role of *ethnic* factor in conflicts, it does not make any sense to aggregate all types of internal conflicts into one category. Civil conflicts in Greece, Algeria or recent war in Syria certainly were not motivated by ethnic grievances. Considering all types of conflicts as the same phenomenon is simply equivalent to mixing two different dependent variables. The second advantage of using the ACD data set concerns the question of fatality thresholds for inclusion. The COW's high 1000 battle deaths threshold excludes several important and long-lasting armed confrontations between government and rebels, like the well-known Northern Ireland dispute or the 58 years long conflict in Myanmar. Moreover, Gates and Strand (2004) argue that the COV data set is biased against small countries because the probability of reaching a threshold level of battle deaths in small countries is lower than it is in a more populous states. The low threshold in the ACD data set helps to mitigate (although it does not solve it completely) this problem. Since we restrict the original data set to 141 out of 155 countries we reduce the number of ethnic conflicts episodes from 110 to 103 which represents a reduction of only 6.3% of the total number of ethnic conflicts in *EPR*.

3.2.3 Explanatory Variables

The list of explanatory variables that we consider in our regression models is the one commonly used in conflict research (Fearon and Laitin, 2003; Collier and Hoeffler, 2004; Montalvo and Reynal-Querol, 2002, 2005; Sambanis, 2001; Hegre and Sambanis, 2006; Cederman, Min and Wimmer, 2009):

GDP per Capita

We control for the development of the country by including the first lag of GDP per capita. According to the literature, we expect a low GDP per capita to increase the risk of ethnic conflict. As indicated by Fearon and Laitin (2003), young men face lower opportunity costs by joining a rebellion and the state has lower capacity to distribute the resources equally among different ethnic groups. The GDP per Capita data comes from Cederman, Min and Wimmer (2009) and originates from Penn World Table 6.2. The data are in constant 2000 US Dollars. These data were also used to calculate the annual growth rate over the period 1946-2005.

Population Size

The GDP per capita and the size of the population are the two most robust variables in the civil war literature. In order to account for the size of the country, we include the natural logarithm of the first lag of population. We expect larger countries to be associated with higher probability of ethnic war. It is because in large countries is much more difficult to control social movements at the local level and there is a higher "human resource" to potentially recruit to an insurgency for a given level of income. Moreover, the absolute population size becomes even more relevant within the context of distributional indices that we relate to conflict potential. For instance, countries with the same number of ethnic groups but very different population sizes may have different propensities to conflict, *ceteris paribus*.

Oil Production per Capita

There are two different (but quite similar) visions of the importance of oil production in the model of conflict. First, as suggested by Ross (2003), when rebels can obstruct the extraction of oil (or other natural resources), the likelihood of secessionist movements increases. Second, Buhaug (2006) argue that oil matters in conflict over an existing state because oil resources are usually controlled by the central government and this increases the incentives to capture a state (Cederman, Min and Wimmer, 2009). Most studies use either a dummy variable for oil exporter or calculate the share of oil exports to GDP. However, this overly simplifying or overly aggregating way of accounting for the importance of oil production increases significantly the risk of collinearity with other variables in the model. For instance, the share of oil exports to GDP is not independent of the strength of other economic sectors in a country. For this reason, we decide to include the data for oil production per capita (in barells) based on Wimmer and Min (2006) from Cederman, Min and Wimmer's (2009) data set. What we expect is that oil exporting countries should have a higher probability of ethnic conflict.

Mountainous Terrain, Noncontiguous Territory and New State

Fearon and Latin (2003) suggest that mountainous countries are, ceteris paribus, more conflict prone, because rebels can hide and retreat. We adopt the mountainous terrain data from their data set. The data are taken from the A.J.Gerrard's (1990) project on mountains environment. As emphasized by Fearon and Laitin (2003), this coding does not take into consideration other types of rough terrain that can be favorable to rebellion, such as junglas or swamps, nor it takes into account the distribution of food in relation to mountains. Following Fearon and Laitin (2003) we expect a higher percentage of mountainous terrain to be associated with a higher probability of ethnic conflict onset. On the other hand, countries with the territory holding a least 10 000 people and separated from the land area containing the capital city either by land or by 100 kilometers of water are coded as "Noncontiguous". The data source is Fearon and Laitin (2003). Fearon and Laitin (2003) argue that the civil war should be more likely when rebels are separated from the territory of the center by water or land distance. Examples of a noncontiguous states include Angola (separated from Portugal), East Pakistan (today Bangladesh) separated from West Pakistan, but also India, Colombia, Malaysia and Philippines. Finally, we construct a dummy variable coded as 1 for the first two years of independence. Fearon and Laitin (2003) show that the odds of civil war onset are 5.4 times higher in the first two years of a state's independent existence. We expect that a newly independent states are more likely to experience ethnic conflicts as well.

Democracy and Autocracy

According to Hegre et al. (2001), among others, democratic societies are able more than other political regimes to solve internal disputes. Autocracies, on the other hand, can suppress rebellions by using force or threatening mass violence. Moreover, the achievement of self-determination, a more favorable ethnic balance of power in government, ethnoregional autonomy, the end of ethnic and racial discrimination, language and other cultural rights, results more suitable for demoracies than for other political regimes. Ethnic wars should therefore be less likely in strongly democratic and strogly autocratic regimes. Reynal - Querol (2002) and Sambanis (2001) confirm this hypothesis. We hence expect the likelihood of ethnic conflict to be higher in countries which are neither democracies nor autocracies. We take the data from Cederman, Min and Wimmer (2009). They use the Polity IV data set (PIV)[10] that examines qualities of democratic and autocratic authority in governing institutions, rather than discreet and mutually exclusive forms of governance. The governing authority spans from fully institutionalized autocracies through mixed, or incoherent, authority regimes (termed "anocracies") to fully institutionalized democracies. The PIV captures this regime authority spectrum on a 21-point scale ranging from -10 (hereditary monarchy) to +10 (consolidated democracy). The Polity scores can also be converted to regime categories: "autocracies" (-10 to -6), "anocracies" (-5 to +5), and "democracies" (+6 to +10). The Polity scheme consists of six component measures that record key qualities of executive recruitment, constraints on executive authority, and political competition. It also records changes in the institutionalized qualities of governing authority. The Polity data include information only on the institutions of the central government and on political groups acting, or reacting, within the scope of that authority. It does not include consideration of groups and territories that are actively removed from that authority (i.e., separatists or "fragments") or segments of the population that are not yet effectively politicized in relation to central state politics.

Instability

By instability we mean the "previous regime change". The regime change is defined as any change in the Polity Score of 3 points or more over the prior three years. The data are taken from the EPR data set and are based on PIV. In line with the existing literature, we expect instability to be positively associated with the onset of ethnic

[10] http://www.systemicpeace.org/polity/polity4.htm

conflict.

Share of the Excluded Population

Cederman, Min and Wimmer (2009) argue that a high degree of exclusion from central power increases the likelihood of rebellion because it decreases a state's political legitimacy. This makes it easier for political leaders to mobilize a following among their ethnic constituencies and challange the government. In order to account for the degree of exclusion along ethnic lines, we include the natural logarithm of the share of the population excluded from central government.[11] When one group is excluded from central government, it can however have some influence at the subnational level (regional autonomy) ot it can be completely powerless or discriminated against. A group is powerless when its elite representatives hold no political power at the national or regional levels without being explicitly discriminated against. A group is discriminated if group members are subject to "active, intentional, and targeted discrimination with the intent of excluding them from both regional and national power" (Cederman, Min and Wimmer (2009)). Examples include African Americans until the civil rights movement and Guatemalean Indians until the end of the civil war[12]. We expect that armed rebellions are more likely when the state excludes large sections of the population from central government on the basis of their ethnic origin. Moreover, the risk of conflict is expected to be particularly high under certain conbinations of ethnic constellations and ethnic exclusion. For instance, when a small majority excludes a substantial minority from central power may significanlty

[11]Cederman, Min and Wimmer (2009) assume that increases in the share of the excluded population have a greater effect on the likelihood of conflict at lower levels of exclusion than at higher levels, and they therefore use a logged transformation of this variable. They hypothesize that "the initial break with the ethnonational principles of legitimacy of modern nation-states carries more political risk than does the shift to an even more exclusionary ethnocracy" (Cederman, Min and Wimmer, 2009, p.327).

[12]See "Online Appendix", Ethnic Politics and Armed Conflict. A configuational analysis of a new global dataset, Cederman, Min and Wimmer, 2009.

increase the likelihood of rebellion.

Number of Power Sharing Partners

Cederman, Min and Wimmer (2009) also assume that infighting is more likely when many partners share government power, that is, in states characterized by a segmented center. Greater the number of political partners, the more likely the alliances will shift, increasing the fear of losing out in the ongoing strugle over the distribution of government spoils. In states with only one ethnically defined elite in power, such ethnic infighting is impossible. Following the authors we include the number of power sharing groups represented by ethnic elites. We expects that the likelihood of conflict increases when large number of ethnic groups shares government power and engages in competitive rivalry, *ceteris paribus*. We term this variable as the *degree of center segmentation*.

Past Imperial History

This variable is aimed at measuring the cohesion of the state. Namely, the cohesion of the state decreases the longer the pre-independance history of indirect rule in an empire and the larger the size of the population. Min and Wimmer (2006) calculate the percentage of years spent under imperial rule between 1816 and independence. The authors count all the years during which a territory was a colonial or imperial dependency (including of the Soviet union and other communist empires) or the heartland of an empire (Turkey under the Ottomans or Austria under the Habsburgs, but not the "mother country" of an empire with seaborne colonies, like Portugal). Moreover, the time spent under imperial rule may significanly impact the quality of institutions, both formal and informal, such as the perception of corruption at the central level or in general the role of the government. Past Imperial History should be positively correlated to the onset of ethnic conflict.

Ethnic Diversity Indices

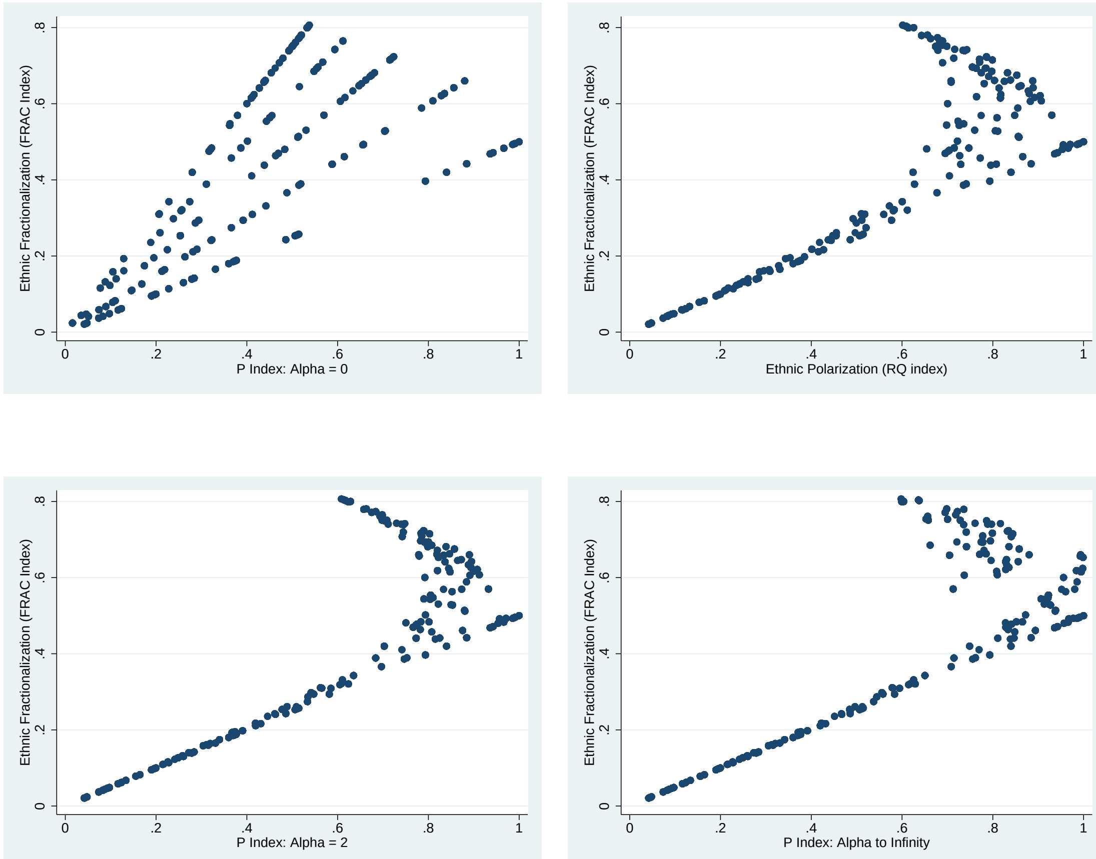
In order to take into account the distributional aspects of ethnicity, we calculate the index of ethnic fractionalization, the index of discrete ethnic polarization as well as the P index of conflict potential. Regarding the P index, we consider three different values for the coefficient α, namely $\alpha = 1$, $\alpha = 2$ and $\alpha \to \infty$. [13] In addition to the above mentioned distributional indices we construct the Collier and Hoeffler's (2004) and Schneider and Wiesehomeier's (2008) ethnic dominance dummy variables. In order to compare the fit of our models with different dummy variables for dominance as well as the relative strenght of our indices with respect to the dominance dummies, we also consider several other dummy variables for dominance using different threshold levels of the groups relative population size such as $50 - 55\%$, $50 - 65\%$ and $50 - 75\%$. Finally, we consider an alternative specification for the P index introduced in the previous chapter, namely the P_{max} index that is calculated using the relative population size of the biggest group in the population (π^*), *i.e.* $P_{max} = 4\pi^*(1 - \pi^*)$.

3.3 The Empirical Relationship Between *P*, *RQ* and *FRAC*

We start with the empirical relationship between our parametric index and the existing indices of ethnic diversity. Figure 3.1 shows the scatterplot of the P index with

[13] We have also calculated the P index for $\alpha = 5$. However, we use the results for the P index with $\alpha = 5$ only to compare the goodness of fit of the corresponding model with other models that include the P index with $\alpha = 1$, $\alpha = 2$ and $\alpha \to \infty$.

$\alpha = 0$, $\alpha = 1$ (actually the RQ index), $\alpha = 2$ and $\alpha \to \infty$ versus fractionalization. The values of the indices are calculated using the data for 141 countries from the *Ethnic Power Relations* data set (Cederman, Min and Wimmer, 2009). Points on the graph are country-year observations. Countries in which politically relevant ethnic categories did not change over time are attributed only one point (actually, all the country-year observations overlap since the value of the indices is constant over time). On the other hand, countries which experienced changes in politically relevant categories over time are attributed more than one point. However, since the variations in the ethno-political constellations over time were never large, the multiple records for such countries are very close to each other and in some cases it is difficult to distinguish them. We have also calculated the mean value of the indices over the period 1946-2005 and we applied them in some graphical representations later in this chapter. The choice of whether to consider means or sigle country-year observations is purely arbitrary and it is motivated only by a greater clarity in graphical representations. Our main empirical results are *not* affected at all by this particular choice.



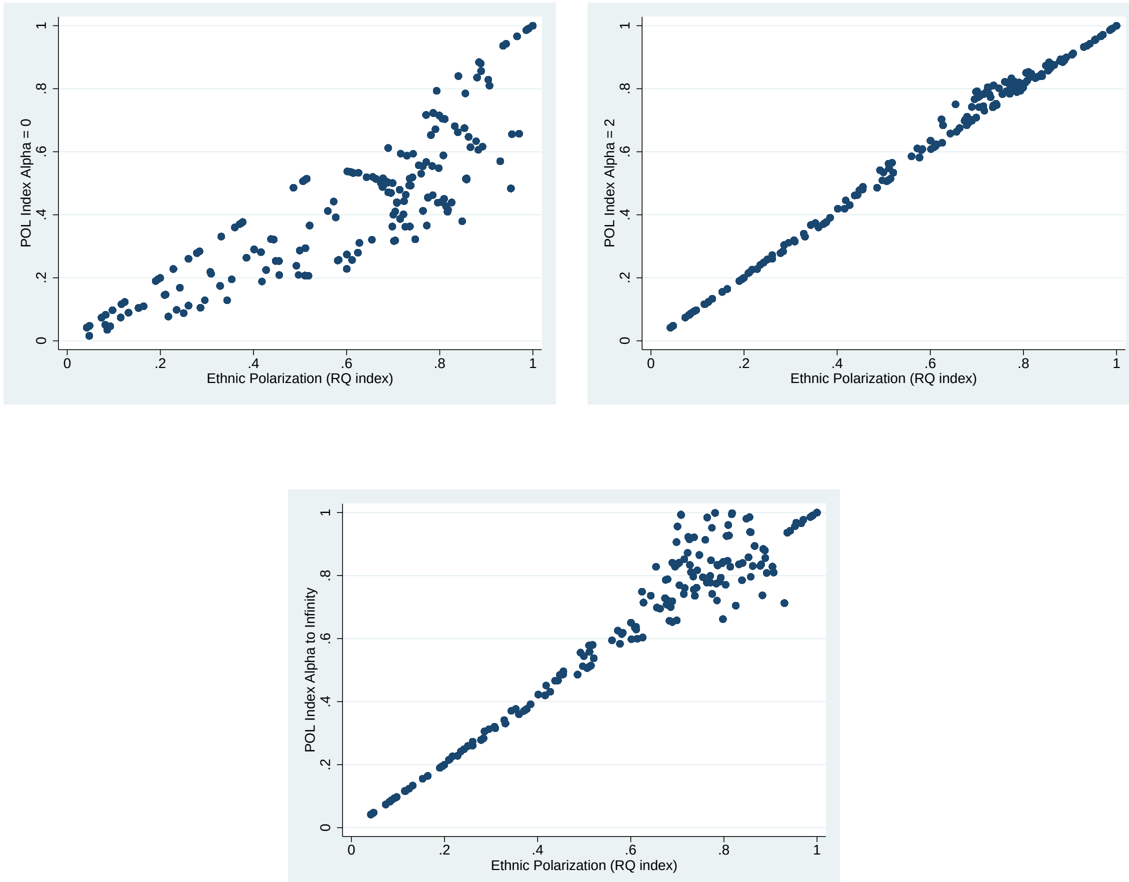
Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 3.1: P and RQ versus $FRAC$

The first figure shows the scatterplot of P_0^n versus fractionalization. The particular "arrow" form of the plots highlights the fact that P_0^n provides the same ranking order as $FRAC$ for a *given* number of ethnic groups. As in the case of the fractionalization index with a given n, P_0^n attains its maximum when all the groups have the same size. The main difference between them is that P_0^n decreases with n while $FRAC$ is monotonically increasing in n. The relationship between RQ and $FRAC$, on the other hand, shows that for low levels of fractionalization, the correlation between ethnic fractionalization and ethnic polarization is positive and high, for the medium range the correlation is very low (almost 0) and for high levels of fractionalization the correlation is negative. The two indices are the same for $n = 4$, the steepest arrow corresponds to $n = 6$ and the flattest one corresponds to $n = 2$.

The relationship between discrete ethnic polarization and ethnic fractionalization, hence, traces a broad quadratic form, with greater variance at the "peak" of the inverted U-curve. Finally, the relationship between P_{∞}^{n} and fractionalization is similar to the RQ polarization index. However, we can observe two separated patterns: almost a linear relationship with the slope of the line of $1/2$ for P_{∞}^{n} up to 0.8 with a slight concentration of points at the highest values of P_{∞}^{n} and a separated cloud of points in the range of P_{∞}^{n} between 0.6 and 0.9.

Figure 3.2 presents the relationship between the RQ index of discrete polarization and the P index for $\alpha = 0$, $\alpha = 2$ and $\alpha \to \infty$.



Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 3.2: RQ versus P with $\alpha = 0$, $\alpha = 2$ and $\alpha \to \infty$.

The relationship between P_2^n and RQ is almost linear. The P_2^n index is slightly higher than the polarization index for the range of RQ between 0.6 and 0.9. The situation changes as α increases. For $\alpha \to \infty$, the P index differs significantly for $RQ \in [0.65, 0.95]$ where the correlation between the two is low (correlation coefficient 0.3). In the previous chapter we argued that this range of RQ corresponds exactly to the intermediate range of $FRAC$ where the majority of ethnic conflict occurs. In the next section we show that the difference between the two measures is not only theoretical but also actual. The relationship between RQ and P with $\alpha = 0$ on the other hand is less clear. The two indices provide the same ranking order only when

$n = 2$. When the number of groups is greater than two, the P_0^n index is always lower than the RQ index.

3.4 Models and Findings

This section evaluates the impact of the P index on conflict behavior. We do so within the context of Fearon and Laitin's (2003) and Cederman, Min and Wimmer's (2009) logistic model that focuses on the onset of ethnic wars in a time range from 1946 to 2005. Ethnic conflict onset is a binary dependent variable coded as 1 in the first year of an armed conflict and 0 otherwise. Ethnic conflict onset is modeled as follows:

$$Pr(y_{it}|EW_{i,t-1}, X_{it}) = \Lambda(\beta_0 + \beta_1 EW_{i,t-1} + \beta_2 y_{i,t-1} + X'_{it}\beta_k + \epsilon_{it}) = \Lambda(z),$$

where $\Lambda(z) = exp(z)/(1 + exp(z))$, $i = 1, ..., 141$ and $t = 1946, ..., 2005$. The variable $EW_{i,t-1}$ is a lagged dummy variable for ethnic conflict (whether there was an active conflict in the previous period), $y_{i,t-1}$ is lagged per capita income, and X'_{it} is a vector of K explanatory variables and controls. The probability of ethnic conflict onset is hence the conditional probability of being in a conflict at time t given that there was peace in $t - 1$.

In order to analyse how the probability of ethnic conflict onset varies across countries we use a (pooled) logistic regression. However, for all model specification we also perform the probit estimation. Since the models have the same number of parameters (although being non-nested) we choose the model with a higher log-likelihood, which is always the logit model. In all model specifications we correct for error correlation over time for a given country by calculating a cluster - robust standard errors.[14] In such a way we obtain a cluster - robust standard error estimates. The number of clusters - countries in our data set is high enough to obtain an accurate inference. We also compare the results with a random effects panel estimator and test for the independence of observations.[15] Since ethnic war is a rare event and since the standard logistic regression can sharply underestimate the probability of such events, we also perform a rare event logit estimation (King and Zeng, 2000).

We control for possible time trends by including the *number of peace years* since the outbreak of the previous war, a cubic spline function on peace years as well as the regional time trends. For the sake of space we do not show time controls variables in the regression results tables, except for the regional time trends. Regional dummies[16] are included in all model specifications. We do not show them for the sake of space.

Table 3.1 provides summary statistics for the core variables in our dataset.

---

[14] In such a way we allow for the correlation of observations within each country (which induces correlation in error terms) but we avoid different countries to have correlated errors.

[15] The random effects logistic estimation allows us to control for the combined effect of all omitted country - specific covariates that cause some countries to be more conflict prone than others.

[16] We include 5 regional dummy variables: Asia, Sub-Saharan Africa, Latin America, East Europe and North Africa and Middle East.

Table 3.1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	Median
Country Characteristics					
Gdp/L (Natural Log)	6.053	7.385	0.028	110.315	3.296
Growth (Annual)	0.108	0.678	-13.775	23.351	.056
Population (Natural Log)	9.147	1.342	5.581	13.902	9.057
Oil/L (In Barrels)	2.169	13.512	0	272.403	0.001
Political Regime and Institutions					
Democracy (D)	0.358	0.479	0	1	0
Anocracy (D)	0.232	0.422	0	1	0
Political Instability (D)	0.126	0.332	0	1	0
New State (D)	0.025	0.157	0	1	0
Ethnic Politics					
% Excl. Population (Nat. Log)	1.941	1.546	0	4.585	2.079
E. Groups in Power	1.746	1.879	0	14	1
Imperial Past (% Time)	0.476	0.313	0	1	0.497
Territory					
Mountains (%, Natural Log)	2.187	1.413	0	4.421	2.442
NC State (D)	0.149	0.356	0	1	0
Ethnic Distribution Indices					
P Index with $\alpha = 0$ $[P_0]$	0.414	0.246	0.015	0.999	0.409
P Index with $\alpha = 1$ $[RQ]$	0.562	0.283	0.042	0.999	0.627
P Index with $\alpha = 2$ $[P_2]$	0.579	0.29	0.042	0.999	0.675
P Index with $\alpha \to \infty$ $[P_\infty]$	0.591	0.302	0.042	0.999	0.657
P^{max}	0.638	0.334	0.042	0.999	0.805
Fractionalization	0.396	0.247	0.021	0.818	0.441
Ethnic Dominance Dummies					
CH (45-90%)	0.548	0.497	0	1	
SW (60-90%)	0.375	0.484	0	1	
DD (50-55%)	0.609	0.487	0	1	
DD (50-60%)	0.631	0.482	0	1	
DD (50-65%)	0.661	0.473	0	1	

3.4.1 Explaining Ethnic War Onset

Table 3.2 presents the basic specification of our model of ethnic conflict onset. [17] Five variables are always significant with the same sign across all model specifications. The level of Gdp per capita is negatively correlated with the probability of conflict outbreak while the size of the country and political regimes that are nor autocracies nor democracies are positively associated with the probability of ethnic conflict outbreak. This is in line with Doyle and Sambanis (2000), Fearon and Laitin (2003), Collier and Hoeffler (2001, 2004), Montalvo and Reynal-Querol (2002, 2005) and Cederman, Min and Wimmer (2009), among others. The results also confirm the Fearon and Laitin's (2003) claim that countries with high shares of mountainous terrain are more conflict prone because rebels can hide and retreat. The previous regime change, oil production per capita and non-contiguous state variables do not result significant for any model specification. Among ethnic diversity variables, only the RQ index of discrete ethnic polarization and the P index are significantly different from zero. However, the P index with Penrose-Banzhaf relative scores outperforms the RQ index of discrete ethnic polarization. Column 5 and column 7 check the relative strength of RQ and P_∞ versus fractionalization. Column 7 shows that the coefficient on ethnic fractionalization is not significantly different from zero, while the coefficient on P_∞ is positive and highly significant. More interestingly, since the goodness of fit of the model that includes both P_∞ and $FRAC$ is practically the same as the one in Coulmn 3, we can conclude that fractionalization does not add much information to the model. This does not mean that ethnic fractionalization is never important but it simply means that the P_∞ index is able to *extract* almost all the information relative to the impact of the interaction between groups.

[17] The models were implemented in Stata 12.1. *Mathematica* was used for the computation of P, RQ, $FRAC$ and P^{max} indices.

Table 3.2: *Logit Regression:* **Basic Model** - *Ethnic War* **Onset** - *EPR(2009) Data Set: 141 Countries*

EW Onset	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Gdp/L	**-0.128*****	**-0.129*****	**-0.132*****	**-0.121*****	**-0.125*****	**-0.128*****	**-0.132*****	**-0.135*****
	0.044	0.044	0.044	0.043	0.044	0.045	0.045	0.045
Population	**0.389*****	**0.389*****	**0.389*****	**0.363*****	**0.375*****	**0.383*****	**0.390*****	**0.405*****
	0.084	0.085	0.087	0.081	0.084	0.086	0.089	0.090
Democracy	0.223	0.239	0.258	0.171	0.206	0.231	0.259	0.298
	0.322	0.321	0.320	0.327	0.327	0.328	0.327	0.323
Anocracy	**0.569****	**0.571****	**0.560****	**0.541****	**0.552****	**0.562****	**0.561****	**0.580****
	0.234	0.234	0.237	0.225	0.227	0.228	0.232	0.244
Oil/L	0.013	0.013	0.012	0.011	0.012	0.013	0.013	0.012
	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.007
Mountains	**0.186****	**0.186****	**0.192****	**0.202****	**0.189****	**0.187****	**0.192****	**0.225****
	0.087	0.088	0.089	0.088	0.089	0.088	0.089	0.091
Instability	0.139	0.139	0.156	0.165	0.150	0.144	0.155	0.160
	0.255	0.255	0.252	0.258	0.260	0.259	0.253	0.247
NC State	0.333	0.312	0.276	0.339	0.327	0.311	0.276	0.225
	0.504	0.507	0.526	0.512	0.511	0.509	0.526	0.536
New State	**2.167*****	**2.165*****	**2.194*****	**2.211*****	**2.180*****	**2.171*****	**2.194*****	**2.254*****
	0.713	0.706	0.700	0.719	0.714	0.706	0.699	0.692
RQ	**1.335****				1.049			-3.161
	0.649				0.887			1.887
P($\alpha = 2$)		**1.560****				1.439		
		0.658				0.883		
P($\alpha \to \infty$)			**1.750*****				**1.766****	**4.218*****
			0.632				0.794	1.554
FRAC				1.119	0.462	0.209	-0.034	
				0.577	0.836	0.862	0.856	
Constant	**-10.778*****	**-10.930*****	**-11.087*****	**-10.283*****	**-10.651*****	**-10.871*****	**-11.095*****	**-11.116*****
	1.432	1.431	1.411	1.422	1.415	1.425	1.414	1.387
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6294	6294	6294	6294	6294	6294	6294	6294
N. Countries	141	141	141	141	141	141	141	141
Pseudo R^2	0.1363	0.1385	**0.1421**	0.1349	0.1367	0.1386	**0.1422**	**0.1463**
Wald Chi2	172.757***	167.142***	166.569***	205.979***	190.052***	175.048***	166.556***	168.655***
Bic	1041.887	1039.718	1036.111	1043.342	1050.287	1048.394	1044.856	1040.772
Aic	900.193	898.024	894.417	901.647	901.845	899.952	896.415	892.330

Notes: *The sample includes 141 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high ethnic conflict. The method of estimation is logit. The absolute z-statistics are calculated using standard errors adjusted for clustering on countries.* ** $p < 0.05$, *** $p < 0.01$.

When we include P_∞ together with the RQ index of ethnic polarization, the coefficient on P_∞ is positive and highly significant while the coefficient on discrete ethnic polarization is not statistically diferent from zero and it has a negative sign.

Since RQ is actually the P index with $\alpha = 1$, the goodness of fit measured by the Pseudo R^2 is increasing in α. Indeed, the highest value of the Pseudo R^2 corresponds to $\alpha \to \infty$ (Figure 3.3). Similarly, the Akaike (AIC) and the Bayesian Information Criterion (BIC) criterion suggest that the model with the best fit is the one that includes the P_∞ index. We have also calculated the Somers' D statistic which provides an estimate of the rank correlation of the observed binary response variable (ethnic war onset) and the predicted probabilities. Since it can be used as an alternative indicator of model fit, we compared its value for the P index with different α. The results are in line with the previous conclusions based on the Pseudo R^2 and on the other informational criteria.

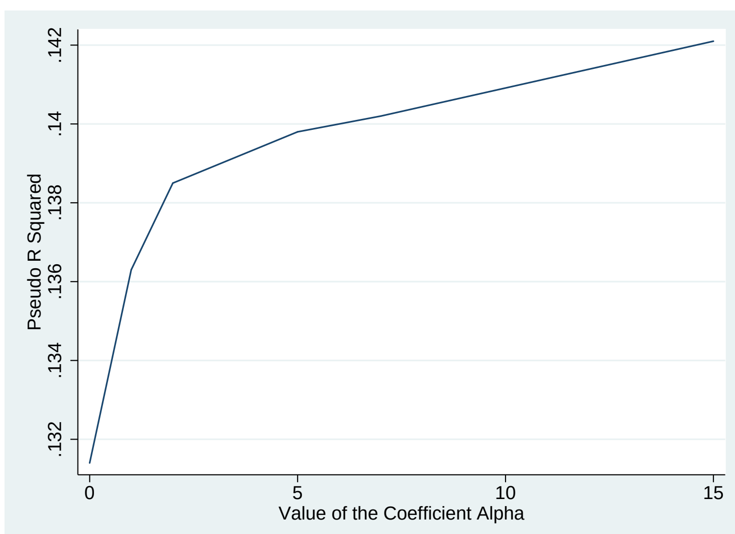


Figure 3.3: Goodness of Fit as a function of the coefficient α.

Conflict Potential as a Combination of Interaction and Dominance

The P_∞ index of conflict potential combines two important aspects of ethnic diversity, namely the interaction and the dominance. Column 1 in Table 3.3 shows that the Collier and Hoeffler's dominance dummy (defined as 1 if the relative size of the biggest group in the population is between 45 and 90%) is significantly different from zero with correct sign. Column 2 shows that the RQ index of discrete ethnic polarization does not "survive" in the baseline regression once we control for dominance. This evidence is in line with Collier and Hoeffler (2004). Only the P index with Penrose-Banzhaf relative scores (P_∞) and the fractionalization index (FRAC) remain significant in combination with the dominance dummy. Note that the model that includes both $FRAC$ and dominance is very similar in terms of the goodness of fit to the one that includes only the P_∞ index (Column 3, Table 2). Similar results are obtained with the Schneider and Wiesehomeier's (2008) ethnic dominance dummy (defined as 1 if the relative size of the biggest group in the population is between 60% and 90%) (Column 5 - 8). Comparing the Akaike (AIC) and the Bayesian Information Criterion (BIC) of the models that include both the dominance dummies and the fractionalization index, we see that they have almost the same value. Moreover, since smaller values of BIC and AIC imply a better model fit, we can conclude that both models perform better with respect to the other models. As before, the Somers' D estimate of the rank correlation of the observed binary response variable and the predicted probabilities is roughly the same (0.65) for both models.

In Table 3.4 we test the empirical performance of the P_∞ index against several alternative definitions of ethnic dominance. More precisely, we construct three additional dummy variables for ethnic dominance coded as 1 if the relative size of the biggest group in the population lies between 50 - 55%, 50 - 60% and 50 - 65%. As an exception we highlight the coefficients that are significant at the 0.1 level in green.

The results from Columns 1-3 suggest that the presence of an ethnic group with the population share between 50 - 55% *alone* does not increase significantly the probability of ethnic conflict. The index of ethnic fractionalization in combination with ethnic dominance, on the other hand, is significant at the 0.1 level. The P_∞ index of conflict potential is highly significant (Column 2). The dominance dummy in combination with the P_∞ index is not significantly different from zero and its coefficient is reduced.

A similar pattern can be observed with the 50 - 60% dominance dummy variable. The index of ethnic fractionalization is significant at the 0.1 level while the dominance dummy variable fails to reach statistical significance at the 0.1 level. The P_∞ index of conflict potential, on the other hand, is significant at the 0.01 level with the correct sign. Finally, in Colums 7-9 we introduce the 50 - 65% dominance dummy. The index of ethnic fractionalization and the ethnic dominance dummy are both significant at the 0.1 and 0.05 level respectively. However, once we include P_∞ in the model, the dominance dummy becomes insignificant.

Table 3.3: *Logit Model - Ethnic War* **Onset** *- EPR (2009) Data Set: Collier and Hoeffler's (2004) (CH) and Schneider and Wiesehomeier's (2008) (SW) dominance dummies included.*

EW Onset	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Gdp/L	**-0.150*****	**-0.145*** **	**-0.142*** **	**-0.145*** **	**-0.143*** **	**-0.144*** **	**-0.144*** **	**-0.142*** **
	0.049	0.048	0.048	0.047	0.051	0.051	0.050	0.051
Population	**0.453*** **	**0.429*** **	**0.416*** **	**0.402*** **	**0.433*** **	**0.412*** **	**0.409*** **	**0.364*** **
	0.096	0.096	0.096	0.092	0.091	0.094	0.096	0.087
Democracy	0.267	0.267	0.284	0.258	0.229	0.252	0.282	0.196
	0.336	0.332	0.329	0.328	0.332	0.326	0.326	0.318
Anocracy	**0.635*** **	**0.605** **	**0.588** **	**0.566** **	**0.642*** **	**0.615** **	**0.605** **	**0.572** **
	0.243	0.240	0.240	0.237	0.242	0.239	0.241	0.232
Oil/L	**0.016** **	**0.015** **	0.014	0.013	**0.018** **	**0.017** **	**0.016** **	**0.016** **
	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Mountains	**0.170** **	**0.164** **	**0.178** **	0.162	**0.196** **	**0.182** **	**0.193** **	**0.196** **
	0.080	0.082	0.086	0.083	0.080	0.081	0.083	0.083
Instability	0.087	0.106	0.132	0.127	0.096	0.110	0.129	0.136
	0.248	0.249	0.249	0.249	0.254	0.255	0.253	0.254
NC State	0.366	0.334	0.281	0.298	0.367	0.319	0.258	0.271
	0.501	0.505	0.521	0.534	0.485	0.496	0.518	0.519
New State	**2.160*** **	**2.150*** **	**2.178*** **	**2.175*** **	**2.180*** **	**2.161*** **	**2.190*** **	**2.203*** **
	0.714	0.706	0.696	0.698	0.729	0.712	0.700	0.712
CH (45-90%)	0.614**	0.487	0.332	0.811***				
	0.275	0.278	0.291	0.306				
SW (60-90%)					0.445	0.478	0.405	0.872**
					0.257	0.265	0.261	0.358
RQ		0.943				1.430**		
		0.640				0.684		
P($\alpha \to \infty$)			1.428**				1.717***	
			0.623				0.645	
FRAC				1.742**				2.219***
				0.686				0.844
Constant	**-10.995*** **	**-11.134*** **	**-11.283*** **	**-11.124*** **	**-10.779*** **	**-11.226*** **	**-11.448*** **	**-10.988*** **
	1.472	1.457	1.435	1.398	1.487	1.463	1.453	1.434
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6294	6294	6294	6294	6294	6294	6294	6294
N. Countries	141	141	141	141	141	141	141	141
Pseudo R^2	0.1377	0.1399	0.1437	0.1449	0.1349	0.1404	0.1451	0.1448
Wald Chi2	129.707***	137.717***	146.344***	144.573***	140.809***	150.392***	152.067***	166.023***
Bic	1040.536	1047.064	1043.366	1042.108	1043.327	1046.558	1041.931	1042.264
Aic	898.842	898.622	894.924	893.666	901.633	898.117	893.489	893.822

Notes: *The sample includes 141 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high ethnic conflict. The method of estimation is logit. The absolute z-statistics are calculated using standard errors adjusted for clustering on countries.* ** $p < 0.05$, *** $p < 0.01$.

Table 3.4: *Logit Model - Ethnic War* **Onset** *- EPR (2009) Data Set: Alternative Specifications for Dominance Dummies.*

EW Onset	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Gdp/L	**-0.146*****	**-0.143*****	**-0.139*****	**-0.147*****	**-0.139*****	**-0.140*****	**-0.146*****	**-0.143*****	**-0.140*****
	0.043	0.043	0.042	0.044	0.042	0.042	0.044	0.043	0.043
Population	**0.448*****	**0.445*****	**0.401*****	**0.451*****	**0.440*****	**0.405*****	**0.463*****	**0.446*****	**0.419*****
	0.086	0.090	0.086	0.087	0.090	0.087	0.092	0.091	0.093
Democracy	0.010	0.038	-0.048	0.019	0.056	-0.041	0.111	0.059	0.053
	0.358	0.358	0.363	0.356	0.353	0.361	0.366	0.358	0.367
Anocracy	**0.468****	**0.422**	**0.395**	**0.454**	**0.429**	**0.381**	**0.484****	**0.425**	**0.415**
	0.237	0.238	0.225	0.243	0.244	0.231	0.243	0.238	0.232
Oil/L	0.011	0.012	0.008	**0.014**	**0.013**	0.011	**0.015**	**0.013**	0.012
	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Mountains	**0.175**	**0.166**	**0.167**	**0.186****	**0.175****	**0.179****	**0.213****	**0.177**	**0.206****
	0.089	0.091	0.096	0.084	0.086	0.090	0.088	0.092	0.093
Instability	0.121	0.145	0.156	0.117	0.139	0.152	0.139	0.145	0.168
	0.250	0.249	0.253	0.251	0.251	0.256	0.254	0.251	0.257
NC State	0.407	0.279	0.372	0.363	0.238	0.326	0.237	0.241	0.215
	0.482	0.538	0.516	0.473	0.531	0.501	0.520	0.524	0.547
New State	**2.259*****	**2.235*****	**2.292*****	**2.238*****	**2.229*****	**2.269*****	**2.253*****	**2.228*****	**2.282*****
	0.729	0.715	0.724	0.736	0.719	0.732	0.745	0.719	0.740
DD (50-55%)	0.743	0.312	0.733						
	0.553	0.572	0.521						
DD (50-60%)				0.341	-0.252	0.267			
				0.431	0.467	0.430			
DD (50-65%)							**0.611****	0.080	**0.561****
							0.275	0.329	0.271
P$(\alpha \to \infty)$		**1.659*****			**2.013*****			**1.693****	
		0.598			0.665			0.728	
FRAC			**1.265**			**1.221**			**1.160**
			0.654			0.649			0.675
Constant	-10.464***	-11.248***	-10.297***	-10.457***	-11.439***	-10.277***	-10.737***	-11.297***	-10.565***
	1.492	1.514	1.439	1.499	1.555	1.449	1.528	1.533	1.477
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6294	6294	6294	6294	6294	6294	6294	6294	6294
N. Countries	141	141	141	141	141	141	141	141	141
Pseudo R^2	0.1447	0.1533	0.1490	0.1424	0.1532	0.1464	0.1468	0.1528	0.1503
Wald Chi2	194.872***	221.043***	273.621***	187.398***	232.854***	265.535***	200.143***	227.139***	269.267***
Bic	1077.353	1077.541	1081.810	1079.597	1077.624	1084.364	1075.178	1078.034	1080.484
Aic	901.922	895.363	899.632	904.166	895.446	902.186	899.746	895.856	898.306

Notes: *The sample includes 141 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high ethnic conflict. The method of estimation is logit. The absolute z-statistics are calculated using standard errors adjusted for clustering on countries.* ** $p < 0.05$, *** $p < 0.01$.

Ethnic Conflict Onset, Regional Time trends and Ethnic Politics

Ethnic war is a very unpredictable phenomenon. It has varied more across countries than across time. Figure 3.4 shows the global onset of ethnic conflict between 1946 and 2005.

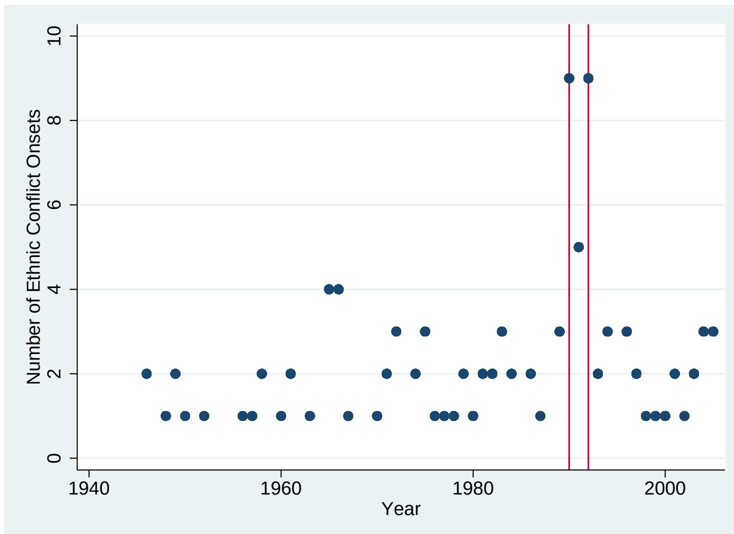


Figure 3.4: Ethnic Conflict Onset over Time

The frequency of ethnic conflict episodes is more or less stable over time. There is, however, a significant increase in the number of ethnic conflicts at the beginning of the 1990s. Is this peak related to one particular geographical region[18] or it is evenly distributed among all the regions in the world? This is an interesting issue because there may be some variation in within-region ethnic conflict onset due to factors that are region-specific over time. In order to answer this question we construct a *regional time trend dummy variables* that takes the following form:

[18] As we have previously mentioned, we divide the world in 5 geographical regions: Asia, Latin America, Sub-Saharan Africa, Eastern Europe and North-Africa and Middle East.

$$RegTrend = RegDummy + RegDummy * Time.$$

If there is no region-specific time effect, the product term of the regional dummy variable and time will be zero.

The second important issue that we address here is the political dimension of ethnic conflicts. As we have previously mentioned, the difference between ethnic (or identity) conflicts and other types of internal (or non-identity) conflicts lies in their motivations and objectives. In order to take into account for one possible source of motivations, we follow Cederman, Min and Wimmer (2009) and include in the regression three *ethnic politics variables*: the share of the population excluded from central government, the number of power sharing partners and the percentage of years spent under imperial rule between 1816 and independence.

Table 3.5 shows the results of our estimation when we include into regression a set of ethnic politics variables together with the regional time trends. In line with Cederman, Min and Wimmer (2009), the degree of ethnic exclusion is statistically significant for all model specification. Since the exclusion and discrimination of one or more ethnic groups "decreases a state's legitimacy and makes it easier for political leaders to mobilize a following among their ethnic constituencies and challange the government" (Cederman, Min and Wimmer, 2009), a high degree of ethnic exclusion increases the likelihood of ethnic conflict. Center segmentation is also significant with the correct sign in all model specifications. The greater the number of political partners, the more likely coalitions will shift, increasing the fear of loosing the share of the government cake and increasing the likelihood of conflict outbreak.

Table 3.5: *Logit Model: Ethnic War* **Onset** *- EPR (2009) Data Set:* **Extensive Model***: Share of the population excluded from central government, Center Segmentation and Imperial Past (EPR data set) and Regional Time Trends.*

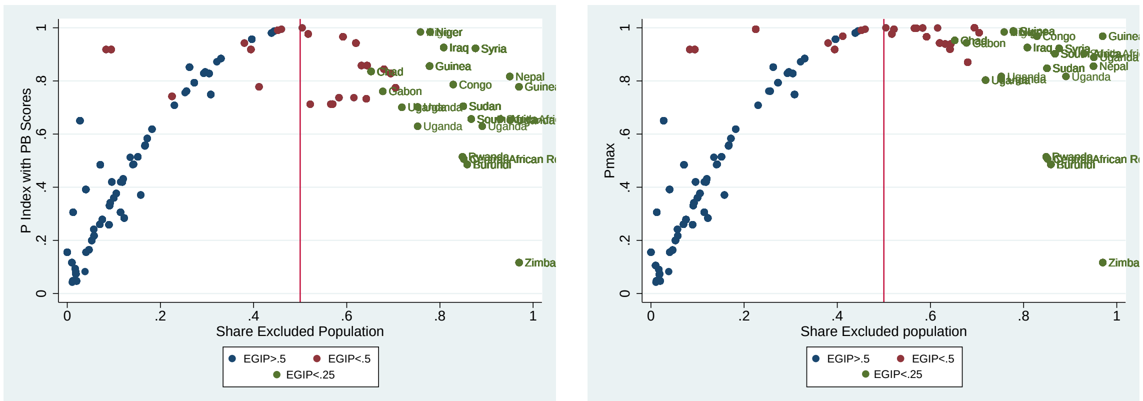
EW Onset	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Gdp/L	**-0.142*****	**-0.142*** **	**-0.142*** **	**-0.142*** **	**-0.141*** **	**-0.140*** **	**-0.141*** **
	0.045	0.045	0.045	0.045	0.045	0.045	0.046
Population	**0.359*** **	**0.363*** **	**0.379*** **	**0.354*** **	**0.361*** **	**0.369*** **	**0.394*** **
	0.084	0.085	0.088	0.084	0.085	0.087	0.094
Excl. Pop.	**0.341*** **	**0.326*** **	**0.312*** **	**0.362*** **	**0.355*** **	**0.354*** **	**0.364*** **
	0.121	0.117	0.111	0.123	0.122	0.120	0.118
Center Segm.	**0.118*** **	**0.116*** **	**0.113*** **	**0.114** **	**0.138*** **	**0.158*** **	**0.177*** **
	0.044	0.044	0.042	0.051	0.049	0.048	0.050
Imperial Past	0.535	0.529	0.581	0.546	0.517	0.492	0.557
	0.574	0.567	0.560	0.583	0.570	0.570	0.579
Democracy	-0.050	-0.049	-0.042	-0.037	-0.065	-0.075	-0.077
	0.444	0.441	0.436	0.444	0.450	0.450	0.448
Anocracy	0.347	0.345	0.331	0.350	0.361	0.375	0.377
	0.233	0.233	0.239	0.229	0.230	0.232	0.240
Oil/L	**0.017** **	**0.017** **	**0.016** **	**0.017** **	**0.018** **	**0.019** **	**0.020** **
	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Mountains	0.153	0.154	0.162	0.156	0.152	0.152	**0.162** **
	0.088	0.090	0.091	0.088	0.085	0.083	0.081
Instability	0.224	0.225	0.246	0.223	0.228	0.235	0.272
	0.269	0.269	0.266	0.269	0.270	0.268	0.266
NC State	0.095	0.074	0.009	0.123	0.084	0.047	-0.045
	0.471	0.484	0.516	0.453	0.472	0.486	0.512
New State	**2.331*** **	**2.324*** **	**2.323*** **	**2.354*** **	**2.325*** **	**2.312*** **	**2.315*** **
	0.718	0.715	0.713	0.714	0.717	0.712	0.707
EEurope*Y	**0.100*** **	**0.100*** **	**0.102*** **	**0.098*** **	**0.099*** **	**0.100*** **	**0.102*** **
	0.019	0.019	0.020	0.019	0.019	0.020	0.020
RQ	0.763				1.089		
	0.713				0.903		
P$(\alpha = 2)$		1.060				1.685	
		0.708				0.914	
P$(\alpha \to \infty)$			**1.465** **				**2.253*** **
			0.687				0.869
FRAC				0.322	-0.546	-1.084	-1.661
				0.677	0.865	0.921	1.016
Constant	**-10.430*** **	**-10.633*** **	**-11.073*** **	**-10.159*** **	**-10.495*** **	**-10.794*** **	**-11.341*** **
	1.383	1.408	1.471	1.383	1.381	1.417	1.510
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6284	6284	6284	6284	6284	6284	6284
N. Countries	141	141	141	141	141	141	141
Pseudo R^2	0.1615	0.1629	**0.1664**	0.1603	0.1618	0.1639	0.1690
Wald Chi2	494.738***	475.630***	448.151***	453.790***	644.007***	628.446***	584.584***
Bic	1086.559	1085.211	1081.721	1087.701	1095.036	1092.940	1087.896
Aic	890.932	889.584	886.094	892.074	892.663	890.567	885.523

The level of Gdp per capita is negatively correlated with the outbreak of ethnic conflict while the size of the population has a strong and positive effect. Ethnic conflict, hence, is more frequent in large and economically poorer countries. Instability and mountainous terrain play a key role in the Fearon and Laitin's (2003) insurgency model but receive rather limited support here. The mountainous variable is significant only when the P_∞ index is included into regression together with the fractionalization index while instability is insignificant for all model specifications. Oil production per capita is associated with resource competition theories and receives a full support here. Although democracy and anocracy have the correct sign they do not reach a significance at the 0.05 level. If we compare this evidence with the baseline model (Table 3.2) we can conclude that the impact of political institutions on the likelihood of ethnic conflict outbreak goes through political exclusion, discrimination and competition at the center. Indeed, the anocracy dummy variable that was highly significant in the baseline model here is not significantly different from zero and the magnitude of its coefficient is reduced. Moreover, the results from table 3.5 suggest that ethnic conflict is more frequent during the first two years of independence. The regional time trends are all insignificant except the one for the East-European countries. The two outliers detected in Figure 3.4 refer, hence, to the East - European countries (Balkans and the former Soviet Union) which experienced several ethnic conflicts at the beginning of the 1990s after the fall of communism.

Finally, the only measure of ethnic diversity that "survives" the inclusion of ethnic politics variables is the P_∞ index. The P index with the Penrose-Banzhaf relative scores is positive and significant at the 0.05 level. Cederman, Min and Wimmer (2009) show that ethnic diversity as such (conceived in their work as fractionalization or polarization) has no robust effect on the likelihood of ethnic conflict outbreak once they account for the political dynamics of ethnic exclusion and competition. We agree with them regarding the importance of ethnic politics but we disagree with the claim that distributional aspects of ethnicity do not matter. Our results confirm that there are some features of ethnic distribution that are particularly conflict prone, even after controlling for a series of different economic, political, structural and geographical characteristic. The robustness of this evidence confirms the intuition behind the P_∞ index: there is no unique and universally "dangerous" ethnic configuration, rather they are all important if combined in a proper way. It is worth noting here that the inclusion of ethnic politics variables resets the statistical significance of anocracy and mountains in the baseline model while the oil production per capita becomes significant at the 0.05 level in all model specifications.

Is there any relationship between political exclusion and ethnic distribution?

Poor countries with weak institutions and large fraction of ethnic population excluded from central power are, hence, particularly conflict prone. These countries are often characterized by a non democratic and repressive political systems that exploit ethnic groups that are excluded from power. The opportunity to exploit a minority is highest when one ethnic group constitutes a small majority. In that particular case the P_∞ index is maximal. Figure 3.5 shows the scatterplot of the P_∞ and the P^{max} index versus the share of the population excluded from central decision making for non - democratic countries with only *one* ethnic group in power. Those countries account for 73% of all ethnic conflicts in our dataset (75 versus 103). The relationship between ethnic exclusion and P_∞ seems to follow a certain empirical regularity. Blue dots represent countries in which the size of the group in power exceeds 50% of population, red dots represent countries in which the group in power has less than 50%, while the green dots belong to the countries where a small minority controls the state (<25% of the population). The first graph shows the scatter plot of P_∞ versus relative population share of group(s) excluded from central power and the second plots the P^{max} index versus excluded population.



Source: *Ethnic Power Relations (EPR) Data Set, Cederman, Min and Wimmer (2009).*

Figure 3.5: P_∞ and P^{max} versus Ethnic Exclusion (Share): Non democratic states with low Segmentation at the center.

These two graphical representations suggest an interesting stylized fact: countries with very low degrees of political exclusion and discrimination have also very low P_∞ index of conflict potential. Remember that P_∞ is low when the size of the dominant group is very high. As P_∞ increases, the share of the population excluded from central power monotonically increases (the correlation coefficient between the two is equal to 0.9). At the highest values of P_∞ (which corresponds to all those countries in which the size of the biggest group is slightly larger than 50%), the share of the excluded population is also around 50%. This means that countries with a scarsely dominant ethnic group in power, are inclined to exclude the rest of the population from decision making at the center or even at the regional level. The evidence suggested by this simple graphical inspection may reflect an interesting *causal mechanism* linking different features of the population distribution across groups to the degree of their political exclusion. There may be, hence, some endogenous relationship between ethnic distribution and political exclusion. The intuition behind this finding, however, is not completely new. Collier and Hoeffler (2004) suggest that countries with a small majority and a substantial minority are particularly conflict prone because the potential to exploit a minority is highest. When the share of the exluded population is above 50% the situation is less clear. However, we should note that almost all countries with high level of political exclusion along ethnic lines are those in which the size of the ruling elite (group) is not larger than 25%. So, in those countries a small minority controls the state.

The second figure represents the share of the exluded population with respect to the P^{max} index. Remember that P^{max} has only one input, namely the relative population size of the biggest group in the population. Given this particular structure of the index, we can easily relate the share of the population excluded from central power to the relative size of the biggest group in the population. For instance, we can check that countries that exclude 30% of the population are those in which the size of the ethnic group in power is exactly 70% ($P^{max} = 0.84$). Similarly, those countries that exclude 10% of the population are those governed by an ethnic group covering 90% of the population ($P^{max} = 0.36$). The situation regarding the countries labeled with green dots is now much clearer than before: almost all green dotted countries are those in which the ruling ethnic group constitutes less than 25% of the population but where the biggest group (which is discriminated against) has between 50 and 70% or between 30 and 50% of the population. Those countries continuously suffer from ethnic conflict.[19]

[19]Cederman and Girardin (2007) claim that the risk of conflict is highest when a small minority controls the state. They propose a particular index N^* of ethnonationalist exclusiveness that maps ethnic configurations onto political violence which is based on a dyadic relationships between the ethnic group in power and all the other peripheral groups. They apply the index empirically and find a positive and statistically significant relationship between N^* and the onset of civil conflicts in Euroasian and North - African countries.

Apart from an interesting insight of a possible relationship between the distributional aspects of ethnic diversity and the propensity of countries to exclude part of their population from power, this seemingly obvious empirical regularity may highlight another, that is well-known issue in the quantitative literature on conflict, namely the problem of endogeneity. Is the impact of ethnic diversity on the propensity for conflict direct or it goes through other, indirect channels? If ethnic diversity did not have any impact the other country charactersitics, a positive association between ethnic diversity and conflict would be attributed to some form of the primordial hate between individuals from different ethnic groups. It is, however, very hard to accept this interpretation of ethnicity even for the most convinced primordialist scholars. Although in this work we continuously emphasize the role of ethnic diversity *itself* in generating the potential of conflict, we are fully aware that ethnic diversity alone cannot be solely responsible for the emergence and for the persistence of ethnic conflicts. By this we mean that different features of ethnic diversity may (probably) affect the internal political and social fabric of a country, *i.e.* its formal and informal institutional arrangements. Although it is beyond the scope of this thesis, we believe that this is an extremely fasciating topic that should be examined further.

Rare Event Logistic Estimation

One of the inherent problems in conflict data is the relative rareness of events (Gates 2001; King and Zeng 2001). Rare events indicate "binary dependent variables characterized as by dozens to thousands of times fewer ones (events such as wars or coups) than zeroes (nonevents)" (King and Zeng 2001, p. 693). The basic problem is that the maximum likelihood estimation of the traditional logistic model suffers from small sample bias. Moreover, the degree of bias strongly depends on the relative number of "success" cases (when the binary dependent variable takes the value 1) with respect to "no event" cases. This means that "the country-specific indicator variables corresponding to the all-zero countries perfectly predict the zeroes in the outcome variable" (Gates 2001; King and Zeng 2001; King 2001). King and Zeng (2001) proposed an alternative estimation method to reduce the bias (this method is very similar to the Penalized Likelihood or Firth Method). Without entering into a detailed discussion of the method proposed by King and Zeng (2001) we can summarize the logic underlying this particular estimation technique as a method that assumes a lower mean square error and increases the probability of an avent, in this case the probability of the rare events. The authors argue that the effects of these methods will be largest when the number of observations is small, *i.e.* under a few thousand, and the events are rare, *i.e.* under 5% or so (King and Zeng, 2001). It should be noted, however, that the problem is *not* the rarity of events, but rather a small number of cases of success compared to the number of fails. That is, the problem is different when one has, for instance, 1000 observations and 20 success events (objects of interest) and another situation with 10 000 and 200 success. In both cases the proportion of the rare event is the same, namely 2%. The rare events logistic estimation results particularly useful in the former case rather than in the latter one. Since our data set counts something more than 6 000 observations and the number of rare events is 103 (or 1.64%), we expect that the results from the rare event logistic estimation do *not* differ significantly from thise obtained using the standard logistic estimation.

Table 3.6 shows the results of the "Rare Events Logit Estimation". The sign and the level of significance of our covariates are similar to those from Table 3.5. The magnitude of the coefficient of the P_∞ index is slightly reduced as it is the coefficient of the ethnic exclusion and center segmentation. The magnitude of the effect of oil production per capita is bigger than in the previous models. However, we conclude that accounting for rareness akes no substantial difference to our results. Collier and

Hoeffler (2004), Sambanis (2001) and Cederman, Min and Wimmer (2009) report similar findings.

In addition to the clustering on country, we perform an additional robustness check for the non-independence of observation over countries and over time. Table 3.7 shows the results of the random effects logistic estimation method. We do not find any substantial differences in results. The test of the correlation coefficient ρ is never significant which means that country - year observations *are* independent. The sign and level of significance of other covariates to ethnic conflict are similar to those obtained with the standard logistic and the rare event logit estimation method. The only difference is the *oil production per capita* variable which is not significantly different from zero for the random effects logistic estimation.

Finally, we test the robustness of our results against the P^{max} index. The P^{max} index was introduced in the last section of the previous chapter. The structure of the P^{max} index is much simpler than the P_∞ index and it relies on the relative rather than absolute majority. However, the cost of this simplification is its inability to capture the interaction between different aspects of ethnic distribution. Table 3.8 shows the estimation results when we include the P^{max} in the regression: it is significantly different from zero only in Model 4 when included together with the fractionalization index.

Table 3.6: *Logit Model: Ethnic War* **Onset** *- EPR (2009) Data Set:* **Rubustness Check A: Rare Events Logit.**

EW Onset	Model1	Model2	Model3	Model4	Model5	Model6	Model7
Gdp/L	**-0.146*** **	**-0.147*** **	**-0.147*** **	**-0.146*** **	**-0.145*** **	**-0.144*** **	**-0.145*** **
	0.045	0.045	0.045	0.045	0.045	0.045	0.046
Population	**0.351*** **	**0.355*** **	**0.370*** **	**0.345*** **	**0.351*** **	**0.359*** **	**0.382*** **
	0.084	0.085	0.088	0.083	0.085	0.086	0.093
Excl. Pop.	**0.325*** **	**0.309*** **	**0.296*** **	**0.347*** **	**0.339*** **	**0.338*** **	**0.347*** **
	0.120	0.116	0.110	0.122	0.121	0.120	0.117
Center Segm.	**0.111** **	**0.109** **	**0.106** **	**0.108** **	**0.132*** **	**0.150*** **	**0.169*** **
	0.044	0.043	0.042	0.051	0.049	0.048	0.049
Imperial Past	0.529	0.524	0.573	0.534	0.506	0.481	0.541
	0.571	0.565	0.558	0.580	0.568	0.568	0.577
Democracy	0.012	0.013	0.019	0.026	-0.002	-0.013	-0.017
	0.442	0.439	0.434	0.442	0.448	0.447	0.445
Anocracy	0.363	0.361	0.347	0.367	0.378	0.391	0.390
	0.232	0.232	0.238	0.228	0.228	0.231	0.239
Oil/L	**0.055*** **	**0.055*** **	**0.055*** **	**0.055*** **	**0.055*** **	**0.055*** **	**0.055*** **
	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Mountains	0.158	0.159	0.166	0.161	0.156	0.157	**0.165** **
	0.088	0.089	0.090	0.088	0.084	0.083	0.080
Instability	0.233	0.235	0.256	0.233	0.236	0.244	0.281
	0.268	0.267	0.265	0.268	0.269	0.267	0.265
NC State	0.074	0.055	-0.005	0.100	0.064	0.031	-0.053
	0.468	0.482	0.513	0.451	0.470	0.484	0.510
New State	**2.287*** **	**2.279*** **	**2.278*** **	**2.309*** **	**2.278*** **	**2.266*** **	**2.269*** **
	0.714	0.712	0.710	0.711	0.714	0.708	0.704
EEurope*Y	**0.085*** **	**0.086*** **	**0.088*** **	**0.084*** **	**0.085*** **	**0.085*** **	**0.087*** **
	0.019	0.019	0.020	0.019	0.019	0.019	0.020
RQ	0.729				1.057		
	0.710				0.899		
P($\alpha = 2$)		1.009				1.622	
		0.705				0.910	
P($\alpha \to \infty$)			**1.407** **				**2.172** **
			0.684				0.865
FRAC				0.293	-0.554	-1.068	-1.623
				0.674	0.861	0.917	1.011
Constant	**-9.800*** **	**-9.984*** **	**-10.402*** **	**-9.554*** **	**-9.846*** **	**-10.124*** **	**-10.648*** **
	1.376	1.402	1.464	1.377	1.374	1.410	1.502

Notes: *The sample includes 141 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high ethnic conflict. The method of estimation is Rare Event Logit. The absolute z-statistics are calculated using standard errors adjusted for clustering on countries. Estimates for peace-years, three natural cubic splines and regional dummy variables not reported.* ** $p < 0.05$, ***$p < 0.01$.

Table 3.7: *Logit Model: Ethnic War* **Onset** *- EPR (2009) Data Set: 141 Countries :* **Rubustness Check B: Random Effects Logit.**

EW Onset	Model1	Model2	Model3	Model4	Model5	Model6	Model7
Gdp/L	**-0.140****	**-0.140****	**-0.142****	**-0.140****	**-0.140****	**-0.140****	**-0.141****
	0.059	0.058	0.058	0.059	0.059	0.058	0.059
Population	**0.363*****	**0.366*****	**0.379*****	**0.359*****	**0.365*****	**0.371*****	**0.394*****
	0.113	0.112	0.108	0.113	0.112	0.111	0.108
Excl. Pop.	**0.340*****	**0.325*****	**0.312*****	**0.359*****	**0.353*****	**0.352*****	**0.364*****
	0.109	0.107	0.103	0.114	0.112	0.111	0.108
Center Segm.	**0.122****	**0.119****	**0.114****	0.115	**0.141****	**0.160****	**0.177*****
	0.056	0.055	0.051	0.067	0.070	0.069	0.064
Imperial Past	0.570	0.558	0.587	0.588	0.553	0.516	0.557
	0.557	0.553	0.531	0.561	0.557	0.550	0.528
Democracy	-0.024	-0.026	-0.036	-0.009	-0.039	-0.059	-0.077
	0.365	0.363	0.358	0.366	0.367	0.365	0.357
Anocracy	0.384	0.377	0.341	0.388	0.393	0.395	0.377
	0.279	0.279	0.281	0.281	0.279	0.278	0.268
Oil/L	0.017	0.017	0.016	0.017	0.018	0.019	0.020
	0.024	0.024	0.023	0.024	0.023	0.023	0.022
Mountains	0.165	0.165	0.166	0.168	0.163	0.160	0.162
	0.111	0.111	0.110	0.111	0.110	0.109	0.103
Instability	0.247	0.246	0.252	0.248	0.249	0.249	0.272
	0.284	0.283	0.283	0.284	0.284	0.283	0.279
NC State	0.160	0.135	0.029	0.184	0.144	0.091	-0.045
	0.456	0.459	0.464	0.450	0.455	0.456	0.424
New State	**2.363*****	**2.351*****	**2.330*****	**2.388*****	**2.355*****	**2.332*****	**2.315*****
	0.652	0.651	0.649	0.651	0.652	0.650	0.644
EEurop*Y	**0.102****	**0.102****	**0.103****	**0.100****	**0.101****	**0.101****	**0.102****
	0.042	0.043	0.042	0.042	0.042	0.042	0.042
RQ	0.796				1.104		
	0.702				0.959		
P($\alpha = 2$)		1.079				1.688	
		0.686				0.929	
P($\alpha \to \infty$)			**1.466****				**2.253*****
			0.614				0.775
FRAC				0.379	-0.520	-1.066	-1.661
				0.797	1.110	1.112	1.038
Constant	**-10.736*****	**-10.884*****	**-11.133*****	**-10.482*****	**-10.772*****	**-10.963*****	**-11.342*****
	1.728	1.733	1.717	1.696	1.729	1.731	1.625
Constant*	-2.238	-2.413	-3.720	-2.178	-2.325	-2.799	-11.858
	1.970	2.339	9.102	1.848	2.151	3.466	14.997
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6284	6284	6284	6284	6284	6284	6284
N. Countries	141	141	141	141	141	141	
Prob $\geq$ *chibar2*	0.296	0.328	0.457	0.283	0.314	0.385	0.497
Wald Chi2	99.959***	101.151***	109.270***	100.665***	100.499***	103.824***	115.707***
Bic	1095.018	1093.760	1090.455	1096.116	1103.546	1101.601	1096.642
Aic	892.645	891.387	888.082	893.743	894.428	892.482	887.523

Notes: *The dependent variable is the onset of ethnic conflict. The method of estimation is Random Effects Logit. The absolute z-statistics are calculated using standard errors adjusted for clustering on countries.* ** $p < 0.05$, ***$p < 0.01$.

Table 3.8: *Logit Model: Ethnic War Onset - EPR (2009) Data Set: 141 Countries:* **Rubustness Check C**: $\mathbf{P^{MAX}}$.

EW ONSET	Model1	Model2	Model3	Model4
Gdp/L	**-0.143*** **	**-0.144*** **	**-0.141*** **	**-0.142*** **
	0.045	0.045	0.046	0.045
Population	**0.367*** **	**0.376*** **	**0.391*** **	**0.387*** **
	0.085	0.088	0.095	0.089
Excl. Pop.	**0.317*** **	**0.310*** **	**0.368*** **	**0.355*** **
	0.115	0.114	0.118	0.120
Center Segm.	**0.099** **	0.076	**0.171*** **	**0.163*** **
	0.045	0.046	0.058	0.049
Imperial Past	0.552	0.575	0.632	0.513
	0.560	0.557	0.622	0.564
Democracy	-0.030	0.011	-0.082	-0.060
	0.440	0.442	0.442	0.453
Anocracy	0.333	0.326	0.366	0.378
	0.235	0.239	0.239	0.239
Oil/L	**0.016** **	0.015	**0.019** **	**0.019** **
	0.008	0.008	0.008	0.008
Mountains	0.158	0.168	**0.163** **	0.159
	0.091	0.096	0.080	0.083
Instability	0.224	0.224	0.283	0.242
	0.268	0.265	0.270	0.265
NC State	0.059	0.030	0.008	-0.038
	0.488	0.492	0.507	0.508
New State	**2.335*** **	**2.365*** **	**2.325*** **	**2.317*** **
	0.716	0.720	0.708	0.710
$\mathbf{P^{max}}$	0.932	2.465	-2.987	**2.455** **
	0.608	1.573	2.266	1.120
RQ		-1.935		
		1.986		
$\mathbf{P(\alpha \to \infty)}$			4.225	
			2.265	
FRAC				-2.467
				1.343
Constant	**-10.579*** **	**-10.537*** **	**-11.336*** **	**-11.009*** **
	1.408	1.403	1.533	1.433
Time Controls	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes
Reg. Time Trend	Yes	Yes	Yes	Yes
N. Observations	6284	6284	6284	6284
N. Countries	141	141	141	141
Pseudo R^2	0.1598	0.1612	0.1653	0.1619
Wald Chi2	421.175***	441.841***	559.315***	552.876***
Bic	1085.254	1092.958	1087.069	1091.035
Aic	889.627	890.585	884.696	888.662

Notes: *The sample includes 141 countries for the period 1946-2005. The absolute z-statistics are calculated using standard errors adjusted for clustering on countries.* ** $p < 0.05$, *** $p < 0.01$

3.4.2 Explaining Low-Intensity Ethnic Conflict Onset

Finally, we check if there is any difference between the correlates to ethnic conflicts that are clasified as *low intensity*, *i.e.* conflicts that do not reach the threshold of 1000 battle death a year, and the general category of ethnic conflicts (low, intermiade and high intensity conflicts together). The internal conflict in Myanmar, the longest ongoing war in the world that began after the country's independence in 1948 from the United Kingdom or the conflict in Northern Ireland are good examples of low intensity ethnic conflicts. The conflict in Burma, however, is characterized by three separated episodes of increased intensity, namely 1948-1952, 1968-1980, and 1983-1995.

Table 3.9 shows the results of our estimation. We can see that for all model specifications (except for Model 4) center segmentation is significant at the 0.05 level with correct sign. In Column 7 also the degree of ethnic exclusion and discrimination is statistically significant. As for the previous analysis, among all ethnic diversity indices only the P_∞ index is significantly different from zero. Moreover, its magnitude is almost doubled with respect to the RQ index and tripled with respect to the fractionalization index. Explanatory variables that were robustly significant when ethnic conflicts were clasified into one category do not result significant when we restrict our attention to low intensity conflicts only.

Table 3.9: *Logit Model:* **Low Intensity** *Ethnic War Onset - EPR (2009) Data Set: 141 Countries*

EW Onset	Model1	Model2	Model3	Model4	Model5	Model6	Model7
Gdp/L	**-0.107****	**-0.107****	**-0.106****	**-0.106****	**-0.108****	**-0.108****	**-0.111****
	0.046	0.047	0.050	0.045	0.048	0.049	0.056
Population	**0.383*****	**0.388*****	**0.412*****	**0.368*****	**0.393*****	**0.407*****	**0.466*****
	0.121	0.121	0.119	0.121	0.123	0.124	0.131
Excl. Pop.	0.190	0.181	0.162	0.214	0.211	0.213	**0.231****
	0.125	0.123	0.114	0.130	0.123	0.122	0.116
Center Segm.	**0.142****	**0.143****	**0.143****	0.125	**0.177****	**0.196****	**0.250*****
	0.063	0.063	0.067	0.073	0.080	0.082	0.084
Imperial Past	1.063	1.068	1.240	1.084	1.104	1.142	1.509
	1.231	1.235	1.202	1.251	1.278	1.304	1.371
Democracy	0.221	0.220	0.217	0.256	0.204	0.195	0.167
	0.568	0.568	0.579	0.565	0.577	0.580	0.606
Anocracy	0.491	0.483	0.445	0.505	0.518	0.523	0.521
	0.414	0.412	0.412	0.419	0.428	0.429	0.435
Oil/L	0.015	0.015	0.014	0.015	0.017	0.019	0.022
	0.014	0.014	0.013	0.015	0.015	0.015	0.014
Mountains	-0.001	0.002	0.018	-0.008	-0.000	0.004	0.012
	0.147	0.149	0.154	0.147	0.143	0.143	0.144
Instability	-0.020	-0.023	-0.006	-0.029	-0.009	-0.008	0.033
	0.465	0.462	0.454	0.464	0.464	0.463	0.455
NC State	0.600	0.583	0.520	0.615	0.582	0.551	0.459
	0.641	0.666	0.747	0.588	0.637	0.664	0.756
New State	1.371	1.362	1.317	1.415	1.368	1.357	1.297
	1.155	1.153	1.152	1.157	1.158	1.158	1.165
RQ	1.518				1.989		
	0.880				1.298		
$\mathbf{P}(\alpha = 2)$		1.748				2.462	
		0.903				1.303	
$\mathbf{P}(\alpha \to \infty)$			**2.747****				**3.963*****
			1.112				1.235
FRAC				0.801	-0.954	-1.474	-3.068
				1.250	1.891	1.950	2.027
Constant	**-11.767*****	**-11.998*****	**-13.022*****	**-11.151*****	**-11.925*****	**-12.279*****	**-13.626*****
	1.835	1.857	1.921	1.840	1.818	1.840	1.877
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6284	6284	6284	6284	6284	6284	6284
N. Countries	141	141	141	141	141	141	141
Pseudo R^2	0.1485	0.1504	**0.1623**	0.1442	0.1491	0.1519	**0.1686**
Wald Chi2	294.414***	286.570***	263.088***	287.108***	296.033***	288.035***	268.397***
Bic	666.019	665.087	659.339	668.063	674.450	673.117	665.003
Aic	470.392	469.460	463.712	472.436	472.077	470.744	462.630

Notes: *The sample includes 141 countries for the period 1946-2005. The dependent variable is the onset of the low intensity ethnic conflict. The method of estimation is logit. The absolute z-statistics are calculated using standard errors adjusted for clustering on countries. Estimates for peace-years, three natural cubic splines and regional dummy variables not reported.* ** $p < 0.05$, ***$p < 0.01$.

3.4.3 Marginal Effects

In the previous sections we have showed that the difference between P and RQ is not only theoretical but also actual. The P_∞ index results much more accurate in predicting the outbreak of ethnic conflict. Since RQ and P_∞ are used as a proxies for conflict potential, countries that are most conflict prone (in terms of the frequency of conflict events) should be assigned higher values of these indices. Figure 3.5 shows the number of ethnic conflict [EW] outbreaks for both RQ (red dots) and for the P_∞ index (blue squares) when both of them are larger than 0.85.

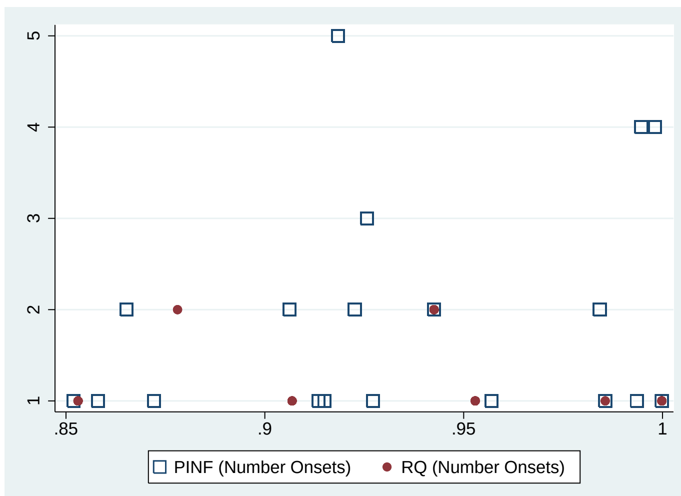


Figure 3.6: Number of EW Onsets for RQ, $P_\infty > 0.85$

While 36 ethnic conflicts are associated with the highest values of P_∞, the RQ index "captures" only 9 conflict episodes. In other words, the number of conflict events is 4 times higher at the highest levels of P_∞ than at the highest levels of the RQ index of discrete ethnic polarization. The RQ index, hence, results much less "selective" with respect to the P_∞ index of conflict potential.

The effect of ethnic heterogeneity measured by P_∞ index is also economically important. Using the results in column 1 and column 3 from Table 2.3, an increase in one standard deviation (0.3015) of the average P_∞ (0.591) increases the probability of ethnic conflict by 52% while an increase in one unit of the standard deviation (0.2825) of the average RQ (0.5618) increases the probability of ethnic conflict by 23%. If P_∞ increases from the average to the maximum value, the probability of conflict increases by 81.31% compared to 39.42% for the RQ index. This means that, if the level of P_∞ increases from the level of Belarus (0.591) to the level of Niger (0.984), the probability of ethnic conflict increases by 77.19% while in the case of the RQ index, this increase is substantially lower, namely 16.83%. Since the number of ethnic conflict episodes in Niger is twice the number of onsets in Belarus, the P_∞ index may result a more accurate measure of conflict potential than the RQ index.

The P_∞ and the RQ index mostly differ for the range of RQ between 0.6 and 0.95. For instance, when $RQ = 0.8$, the P_∞ index can range from 0.62 (Sierra Leone - 0 Ethnic Conflict)[20]) to 1 (Iran - 4 Ethnic Conflicts). With all the variables at their means, the probability of conflict predicted by P_∞ varies by 73.6%. Although Iran and Sierra Leone have similar ethnic distributions and, hence the RQ index yields almost the same level of polarization, the ethnic distribution of Iran is characterized by ethnic dominance since the biggest group in the population possesses 53% of the entire population versus 48% in the case of Sierra Leone. The difference of only 5 percentage points can actually make a difference, and the P_∞ index is sensitive enough to capture it.

Figure 2.6 plots the predicted annual probability of ethnic conflict as a function of the P_∞ index. The first subfigure shows the predicted probability of ethnic war for

[20]The conflict that begun in 1991 in Sierra Leone was not motivated by ethnic cleavages. It started when the Revolutionary United Front (RUF), with support from the special forces of Charles Taylor's National Patriotic Front of Liberia (NPFL), intervened in Sierra Leone in an attempt to overthrow the Joseph Momoh government.

RQ and P_∞ when all other variables are at their means. The remaining two figures display the marginal effect of P_∞ on the probability of ethnic conflict outbreak for different levels of economic development and different population size. If $P_\infty \to 0$, ethnic conflict is improbable for the richest quintile of countries. As P_∞ increases, the probability of conflict steadily increases for all quintiles of income. However, the probability that an ethnic conflict will escalate is more then 4 times higher for the poorest subsample of countries than for the richest one. Similarly, at $P_\infty = 1$ the conflict propensity of the smallest state is 4 times lower with respect to that of the biggest state.

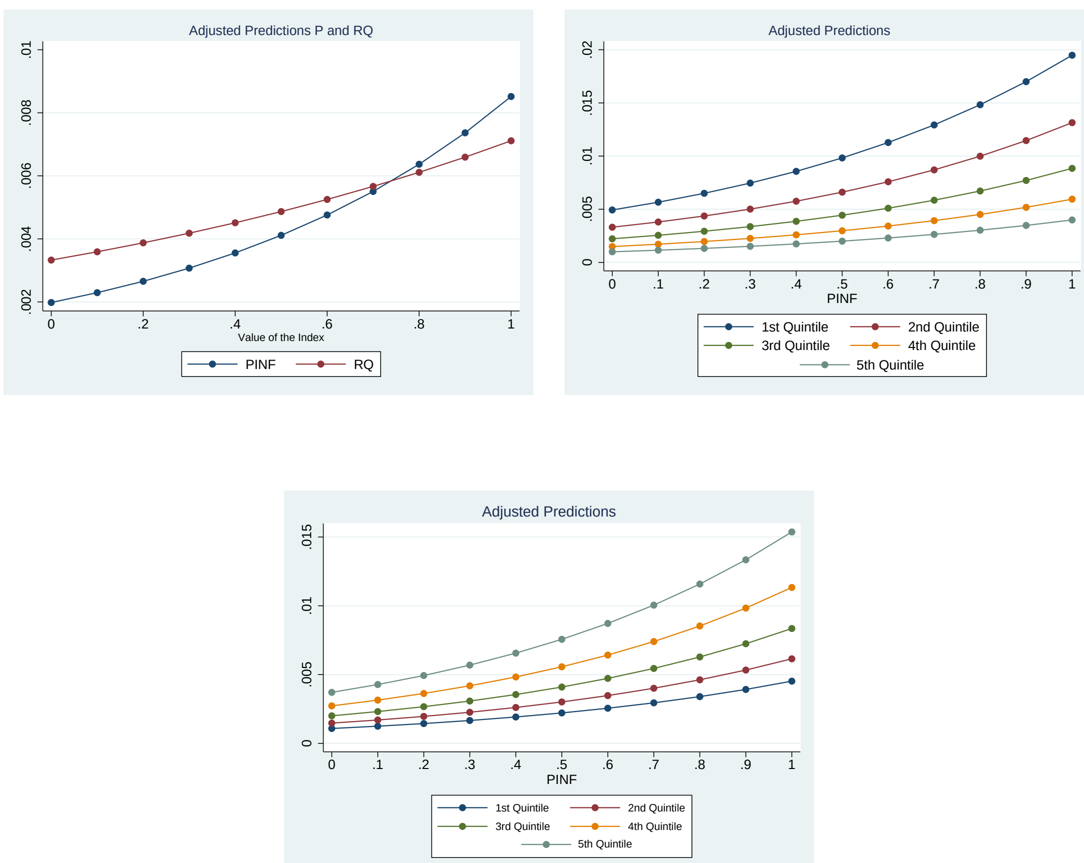


Figure 3.7: Marginal Effects P_∞ Index

The average impact of P_∞ on the likelihood of ethnic conflict outbreak is inscreasing in the share of the excluded population, in the number of power sharing partners and in past imperial history. Figure 2.7 shows the average impact of P_∞ for different levels of political exclusion, center segmentation and percentage of years spent under imperial rule.

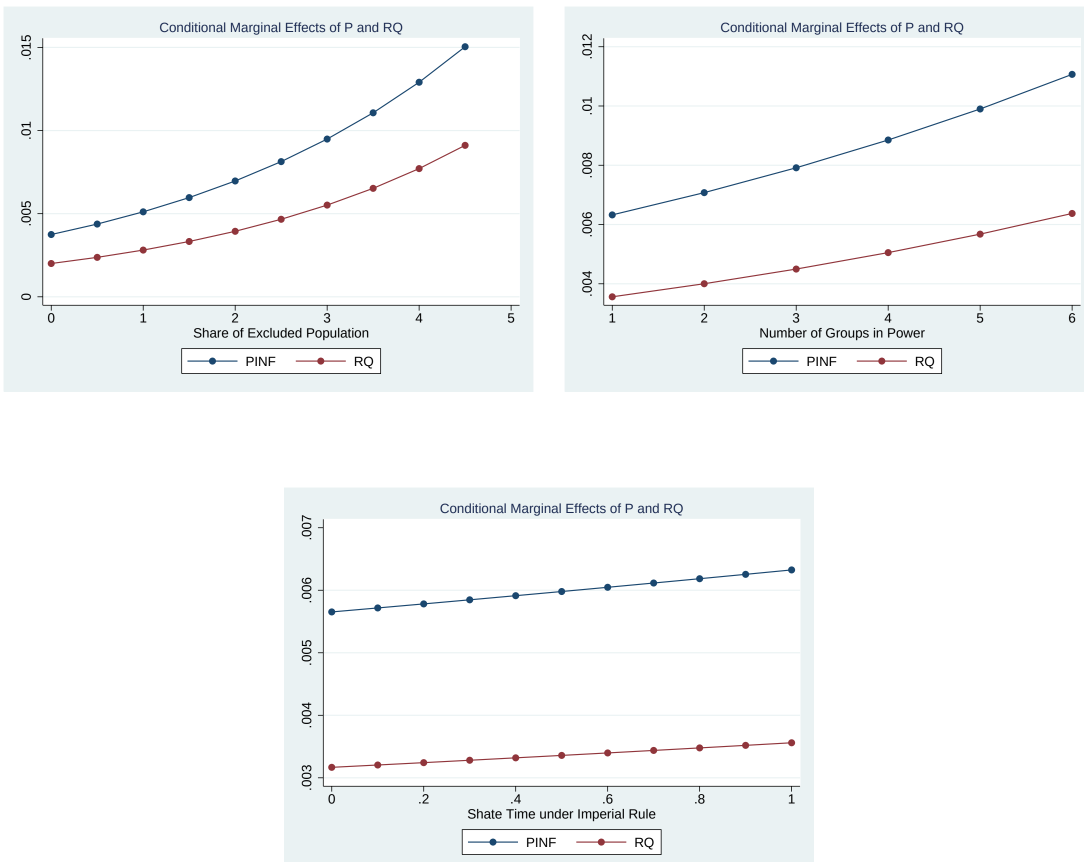


Figure 3.8: Marginal Effects P_∞ Index

When the size of the biggest group is scarsely higher than 0.5 (that is where the P_∞ index attains its maximum for any $n \geq 2$), the likelihood of ethnic conflict is more than 3 times higher when the rest of population is excluded from decision making or discriminated against with respect to the case in which no group is excluded from the central decision making process. For the range of P_∞ between

0.9 and 1 and complete political exclusion or discrimination of the minority, the number of ethnic conflicts is 4.5 higher with respect to the case in which the minority is included in central decision making. A key reason for this is that members of ethnic group(s) that are intentionally and categorically excluded from political, economic or any other simbolic benefit that derives from participation, feel entitled to fight for political change. However, high degrees of exclusion restrict their "legal" possibilieties to achieve this goal, so the use of arms is sometimes regarded a viable alternative strategy. Similarly, with one ethnic group being scarsely predominant, and other groups included in central decision making, the risk of conflict is increasing in the number of groups. Passing from 2 to 6, the risk of conflict increases 1.5 times (Figure 2.7). We can conclude that countries characterized by small ethnic majority and high degrees of exclusion along ethnic lines are particularly conflict prone. Similarly, countries with high levels of ethnic inclusion but with high competition at the center face higher probability of conflict with respect to countries with only few power sharing partners. The risk of conflict in this case is due to increased competition at the center and more groups there are, more the coalitions will shift, generating instability and conflict potential. In both cases, however, the probability of conflict decreases in the size of the dominant ethnic group.

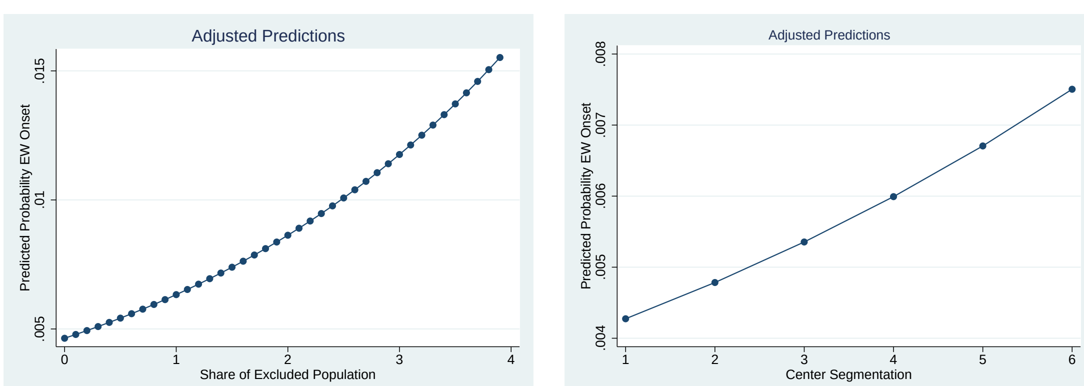


Figure 3.9: Predicted Probability of EW Onset.

3.5 Concluding Remarks

We opened this chapter by asking a simple but at the same time very delicated question: Are distributional aspects of ethnic diversity important for the explanation of ethnic conflict outbreak or is ethnic struggle driven exclusively by other economic, social and political factors? Any attempt to answer this question requires caution because the relationship between ethnic diversity and conflict is far from being simple. However, the results that we obtain in this chapter are quite convincing: ethnic diversity can be considered as an important correlate to ethnic conflict onset.

The P index of conflict potential with the relative Penrose-Banzhaf scores results to be robustly correlated with the onset of ethnic conflicts. Moreover, the index performs better than any other distributional index of ethnic diversity (dominance dummy variables included) in the explanation of the ethnic conflict outbreak. The superiority of our index over the existing indices of ethnic diversity lies in its particular structure, *i.e.* in the way in which it treats the across groups interaction and the groups power. The index combines in a non linear way three different concepts of ethnicity, namely the fractionalization, the polarization and the ethnic dominance by implicitly assigning different weight to the interaction and the power component according to the features of the underlying population distribution across groups. This particular feature of the index proves to be crucial for the explanation of the diversity-conflict nexus.

Our results suggest that the way in which the population is distributed across ethnic groups is important for the explanation of ethnic conflict. This evidence is robust to the inclusion of an additional set of regressors containing several ethnic politics variables, and to different methods of estimation. Even after controlling for structural conditions, country characteristics and time and regional specific effects, the only measure of ethnic diversity that still remains significant is the P_∞ index. The robustness of our results suggest hence that ethnic diversity matters only when the two main components of the conflict generating mechanism (interaction and power) are properly combined into one single measure. Our empirical results confirm this intuition since the P_∞ index of conflict potential resets the significance of these two components considered separately but they turn to be highly significant once combined in a proper way within the structure of our index.

The effect of ethnic heterogeneity measured by P_∞ index is also economically important. An increase in one unit of the standard deviation of the average P_∞ increases the probability of ethnic conflict by 52%. If P_∞ increases from the average to the maximum value, the probability of conflict increases by 81.31%. At the highest levels of P_∞, the probability of an ethnic conflict onset is more then 4 times higher for the poorest subsample of countries than for the richest one. When the relative size of the biggest group in the population is scarsely higher than 0.5 (that is where the P_∞ index attains its maximum for any $n \geq 2$), the likelihood of ethnic conflict is more than 3 times higher when the rest of population is excluded from decision making or discriminated against with respect to the case in which no group is excluded from the central decision making process.

The results obtained in this chapter, however, should be taken with a certain caution. The effect of the P_∞ index of conflict potential *cannot* be interpreted as a causal impact of ethnic diversity on the potential of conflict. Eventhough our results suggest that ethnic distribution is an important factor in the determination of conflict potential, it tells nothing about the direction of its impact. It may be, for instance, that certain ethnic distributions "imply" certain political constellations. This is an interesting issue that requires further examination.

Chapter 4

Ethnic Diversity and the Duration of Ethnic Conflict

4.1 Introduction

Peace and conflict studies have analyzed different aspects of civil conflicts, namely their onset, incidence and duration or continuation. Although these are derivatives of the same social phenomenon, they should be understood and treated as distinct concepts (Montalvo and Reynal-Querol 2010). While the determinants of all three aspects of conflict might overlap, they are not necessarily the same. The factors that motivate masses to mobilize and fight or that influence their efficiency to bring out a rebellion may be different from those that influence the intensity or duration of fighting. For instance, it is quite difficult to claim that high levels of poverty should generally increase the duration or intensity of fighting. As Costalli and Moro (2011) claim, the more reasons to fight do not necessarily lead to more zealous and more aggressive fighting. Similarly, it is not clear why autocracies should experience longer conflicts given that they may suppress with force rebel movements more efficiently with respect to democratic and/or partially inclusive societies.

Most of the literature examines correlates of civil war onset, meaning that the dependent variable takes the value 1 for country years with a civil war onset, and zero for other country years. Some drop country years with ongoing civil war (Collier and Hoeffler 2004), others include them and introduce a variable indicating whether war was ongoing in the previous year as a control (Fearon and Laitin 2003).[1] Other researchers look at the correlates of civil war incidence, meaning that the dependent variable is coded as 1 for every country year with at least one civil war occurring, and zero otherwise (Montalvo and Reynal-Querol 2005). A problem with the latter approach is that the estimated coefficients are quite complicated averages of the "effect" of a covariate on both the onset and the duration of civil conflicts. Since onset and duration are two different aspects of the same phenomenon, they also require different estimation techniques. A more natural alternative hence is to study determinants of conflict onset and conflict duration separately using survival models for the latter aspect of conflicts (Montalvo and Reynal - Querol, 2010; Collier, Hoeffler and Soderbom 2004; Fearon 2004, DeRouen and Sobek, 2004, Gates and Strand, 2004).

The exercise aimed to investigate a potential association between structural conditions, resource availability, country characteristics and features of ethnic distribution and the duration of ethnic conflict is not an easy task, just as it was not in the case of ethnic conflict onset. Our results from the previous chapter strongly suggest that countries with low levels of economic development, large populations and high degrees of political exclusion and competition along ethnic lines face higher risk of conflict. Moreover, we show that ethnic diversity is an important correlate to ethnic conflict.

[1] The latter approach avoids dropping onsets that occur when another civil war is already in progress.

We apply our index of conflict potential that assigns different "weights" to power and interaction in the determination of conflict potential according to the characteristics of the underlying population distribution across ethnic groups and find that, once the power and the interaction are properly combined into one single measure, ethnic diversity becomes a significant and robust correlate to ethnic conflict onset. The likelihood of ethnic conflict hence is given as a non-linear combination between the effects of the three main aspects of ethnic diversity, namely the dominance, the polarization and the fractionalization.

In this chapter we go one step further and explore empirically another important aspect of ethnic conflicts, namely their duration or persistence. As we have previously mentioned, the duration of an event as a dependent variable is different from the mere occurance of the event. The probability of ethnic conflict onset is the conditional probability of experiencing a conflict in a certain point in time given that there was no conflict in the previous period. The hazard or duration of an ethnic conflict measures the probability of experiencing a conflict between two points in time given that there was a conflict in the previous period. Understanding what makes ethnic conflicts longer, and which are the propagation mechanisms that make conflicts persistent and economically extremely costly is important in order to define the policies that need to be implemented to shorten them. As in the case of ethnic conflict onset, there is a potential problem of endogeneity which makes the analysis on the causal mechanisms very hard, so we limit ourselves to investigating the potential association between several explanatory variables and the duration of ethnic conflict. We are particularly interested in analyzing whether the features of ethnic diversity together with ethnic politics and competition dynamics are a possible dimensions of the propagation mechanisms of ethnic conflicts.

This chapter is structured as follows. In Section 2 we briefly summarize the main findings from the existing literature on conflict duration. Section 3 presents the key concepts of survival analysis and the main empirical issues together with a non-parametric and a parametric analysis of the duration of ethnic conflicts. Section 4 concludes.

4.2 Related Literature

The empirical studies on civil conflict onset and incidence greatly outnumber those of duration. If we consider ethnic rather than civil conflicts in general, the number of studies reduces further. The study of civil conflict duration begun in 2004 when the most prominent articles on the subject were published in a special issue of the *Journal of Peace Research*. As in the case of civil conflict onset, two empirical studies can be considered as foundational, namely the Fearon's (2004) and the Collier, Hoeffler and Soderbom's (2004) investigation of the correlates to civil conflict duration. Although these two approaches deal with the same concept (duration or persistence of civil conflicts), they differ significantly in the definition of a war and in the set of covariates used to explanain the forces that make conflicts persistent and extremely costly.

Fearon (2004) classifies the wars into five different categories[2]: Coups and popular revolutions, Post 1991 civil conflicts in Balkans and in the former Soviet Union, Decolonization wars, the so-called "Sun of Soil" wars and finally the Contraband financing. The author defines a coup-related civil war as a conflict between groups interested in taking control of a state and that are led by individuals which were part of the central government or armed forces. The coup in Argentina in 1955 or in Iraq in 1959, among others, fail into this category. A popular revolution on the other

[2]These categories are *not* mutually exclusive since some conflicts in his data set have more than one of the five attributes. This point was emphasized also by the author (Fearon, 2004, p. 277)

hand is defined as a conflict that involves mass demonstrations in the capital city in favor of deposing the regime in power (Cuba in 1958 or Iran in 1978). The second category includes recent conflicts in the South-Eastern Europe (followed by the fall of communism) and in the former Soviet block. Wars against the formal colonial empires, such as that in French Algeria or the Mau Mau rebellion in Kenya are classified as Decolonization wars. The "Sun of Soil" wars are defined as conflicts over natural and land resources between a "peripheral" ethnic minority and state supported migrants of a dominant ethnic group. Finally, the Contraband financing conflicts include all the conflicts in which the rebel group has access to funds from contrabant such as opium, diamonds or coca. Relying on the *Armed Conflict Datased* [ACD], the author considers conflicts with at least 1000 battle deaths over its course (which is a somewhat different definition of conflict with respect to the commonly accepted threshold of battle deaths on a yearly basis) and with a yearly average of at least 100 battle deaths.[3] Applying a multivariate Weibull estimation of the duration on these categories and the common determinants such as the level of economic development, population, democracy, ethnic war dummy and ethnic fractionalization, the author finds that civil wars emerging from coups or revolutions, those in South-Eastern Europe and the former Soviet Union as well as anti-colonial wars tend to be relatively short. By contrast, "sons of the soil" wars and the contraband financing are on average longer. After controlling for different types of civil conflicts, all the other explanatory variables have no statistically significant effect on duration.

Gates and Strand (2004) address in detail the problems related to the definition of civil conflicts and to a number of measurement problems that may lead to selection bias. In the previous chapter we have argued that the criteria that one uses to determine whether a country is experiencing a conflict in a particular time period

[3]In order to exclude genocides and massacres where there is no organized rebel involvement, the author requires at least 100 battle deaths on the side of the government (including civilians attacked by rebels).

may significantly affect the results. The analysis of conflict duration requires even more caution regarding the definition of the dependent variable. The first issue is concerned about which unit of time is used to determine the conflict duration. Most duration analysis measure time in years and very few of them measure duration in months (Fearon (2004) included). Gates and Strand (2004) argue that this is particularly problematic for very short conflicts eventhough some of them counted more than 1000 battle-deaths. When the time is measured in years, the distribution of wars is truncated at one year irrespective of their effective duration in months. Hence, there is no distinction between conflicts that erupt and finish in one month and those that last for 10 months or for the entire year. The authors provide a new dataset that addresses this issue by measuring conflicts in days rather than years or months. Moreover, they consider a threshold of 25 battle deaths in order to determine whether there is a conflict in a particular time period. Together with the set of explanatory variables suggested by Fearon (2004), the authors include several other variables such as the presence of international involvement in the conflict, a dummy for parallel conflicts, the relative change in the level of the per capita income (growth rate) and political instability variables. They find a robust effect for the intensity of the conflict, political instability, anocratic governments and the presence of parallel conflicts. However, their analysis does not consider any control for the ethnic diversity of the countries.

Collier, Hoeffler and Soderbom (2004) use the threshold of 1000 battle deaths to define a war, montly data and a piecewise exponential duration model. They show that the duration of conflicts is positively related to low per capita income, income inequality, population size and to a moderate degree of ethnic diversity measured by the ethnic fractionalization index. The relationship between conflict duration and ethnic diversity is non-monotonic since they find a positive and statistically significant effect of the square term of ethnic fractionalization on the hazard of peace. Factors that shorten conflicts are a decline in the prices of the primary commodities that the country exports and external military intervention on the side of the rebels. The latter effect was not statistically significant in Gates and Strand (2004). They also find that peace was less probable in the 1980s and 1990s then it had been previously.

Finally, Montalvo and Reynal-Querol (2010) investigate the explanatory power of ethnic polarization on the duration of civil conflicts. They follow Gates and Strand (2004) and consider the lower threshold for civil war (25 battle deaths) and the Weibull specification for the hazard function. The factors that lenghten civil conflicts are the size of the population and the degree of ethnic polarization while the oil exporting countries seem to fight shorter wars, which is opposite to the findings in Fearon (2004).

The empirical approach that we adopt in this chapter is somewhere between Montalvo and Reynal-Querol (2010) and Gates and Strand (2004). Although we are completely aware of the fact that choosing years rather than months or days is not the best option for measuring the duration time, we decide to follow Montalvo and Reynal-Querol (2010) and Fearon (2004) and consider years as the relevant unit of time. The data on conflict onset and duration are taken from the ACD that distinguishes between real "wars" with annual battle deaths exceeding 1000 and minor wars or "conflicts" with a minimum of 25 annual battle deaths. We consider that a conflict has finished if there is at least one year of peace before a new ethnic conflict breaks out. Differenty from the above mentioned authors*Aware of the different nature of ethnic and non ethnic conflicts, Fearon (2004) distinguishes between non ethnic, mixed or ambiguous and ethnic conflicts. See Fearon (2004), p. 18.*, we consider ethnic rather than civil conflicts in general. The precise coding for ethnic conflicts is taken from the "Ethnic Power Relations" data set [EPR] provided by Cederman, Min and Wimmer (2009). Our main aim is to assess the degree of importance of the features of ethnic distribution on the duration of ethnic conflicts. We do so by applying our index of conflict potential introduced in the second chapter. We follow the existing literature on conflict duration and model the hazard function using the Weibull specification and address in detail the issue of non proportionality. We also run a semi-parametric Cox estimation and the accelerated failure version of the Weibull model. Finally, we perform two additional robustness checks: i) in addition to the clustering on countries, in order to control for the non-independence of country-year observations we also calculate the robust standard errors clustered by conflict episodes, ii) we control for the presence of parallel conflicts by introducing a dummy variable coded as 1 if there is more than one conflict ongoing in the same country. The results are remarkably robust.

4.3 Empirical Investigation of Ethnic Conflict Duration

4.3.1 Survival Analysis: Basic Concepts

Survival analysis is aimed to model *time-to-event* or *duration data*.[4] The main outcome of interest hence is the time interval between the beginning of an event and the time of its termination. The key concepts of survival analysis are the *hazard function* and the corresponding *survival function*. The hazard function shows how the *hazard rate* evolves with time. The rate of hazard is defined as the instantaneous rate at which a randomly selected unit of observation known to be active up to a certain point in time will "fail" at the end of that time period. The hazard rate hence gives us the proportion of units of observations that "fail" per unit time. The cumulative or

[4]Time to event or duration data are also known as *transitory* or *survival time* data. In this work we refer to the time-to-evet data as the duration data.

integrated hazard at any point in time equals the area under the instantaneous hazard curve up until that particular time period. It shows hence the cumulative probability that the event (failure) has occured up to any point in time. In our context where the unit of observation is a conflict episode, the hazard rate is given as the conditional probability of conflict ending at any point in time and the cumulative hazard shows the cumulative probability that a conflict has ended up to any point in time.

The survival function, on the other hand, shows the cumulative proportion of the units of observation who have "survived" as a function of time. Obviously, the hazard function and the survival function are "two sides of the same coin": an increasing hazard function implies a decreasing survival function. This is because the rate of "failure" is just one minus the rate of survival. In our context, if the probability of conflict ending increases with time, the conflict survival rate will decrease with time. That is, the hazard of *peace* increases while the probability of *conflict continuation* decreases. It is very important to bear this in mind especially in light of the interpretation of the coefficients in the empirical part of the analysis.

4.3.2 Duration of Ethnic Conflict - A Non Parametric Survival

Our first task is to describe the data graphically using a survival curve. A survival curve permits us to analyse the temporal pattern in the data and to identify an appropriate distributional form for the data. The most commonly used non-parametric estimation procedure is the *Kaplan-Meier* method [KM] (also known as the product limit estimator), alternative to the *life tables* and the Nelson Aalen method. It is called a non-parametric method because it does *not* include any particular covariate into analysis, *i.e.* it analyses the hazard and the survival rates without conditioning them on any country characteristics. The Kaplan-Meier estimator is based on estimating conditional probabilities at each time point when an event occurs and taking the product limit of those probabilities to estimate the survival rate at each point in time. The KM estimator of survival at time t is given by the following expression (Stevenson, 2009):

$$\hat{S}(t) = \prod_{j=1}^{k} \frac{n_j - d_j}{n_j}, 0 \leq t \leq T^* \tag{4.1}$$

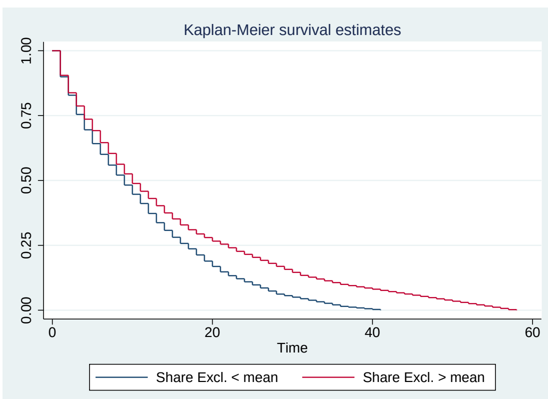
where n_j denote the number of ethnic conflict episodes (which we refer to as the units at risk) just before time t_j, *i.e.* n_j is the number of active conflicts at the beginning of time t_j, and d_j denote the number of ended conflicts in time t_j. Time T^* is the last year of the most durature conflict in the sample ($T^* = 58$). Each point of the KM function, hence, corresponds to the cumulative proportion of conflicts that have "survived" up to a corresponding point in time.

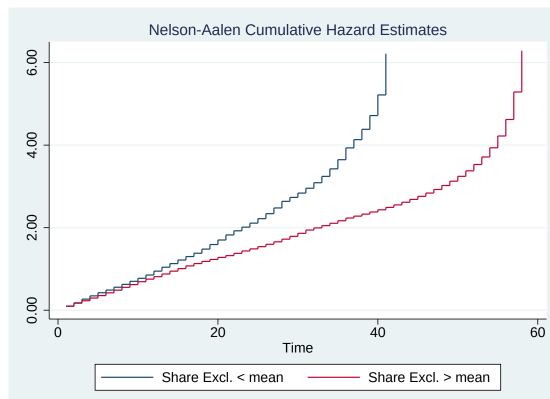
Figure 4.1 shows the KM survival function and the cumulative (integrated) hazard function[5] for several variables of interest. The list of variables includes: the degree of political exclusion and competition along ethnic lines, a dummy variable whether a country is characterized by inclusive or non-inclusive political system (democracy versus autocracy), the size of the population and a dummy variable that takes the value 1 for all those countries with the P_∞ index above the sample median (0.65) and 0 otherwise. The x-axis represents time-to-event in years (time to peace) and y-axis represents the probability of conflict survival at any time t, $0 \leq t \leq T^*$.

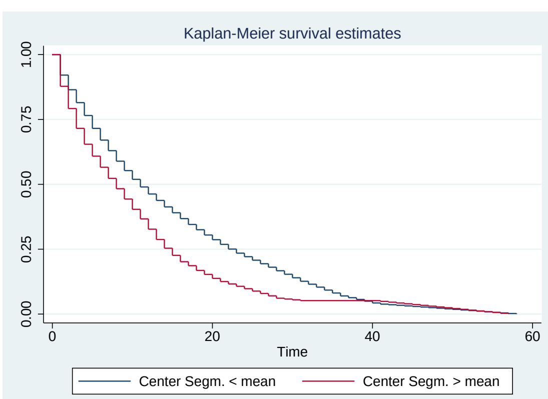
As we can see, the KM curve is a step function that drops at the time of each event. Thus, the KM curve represents an estimate of survival as a function of time. The first sub-figure shows that the survival function for the countries with the share of

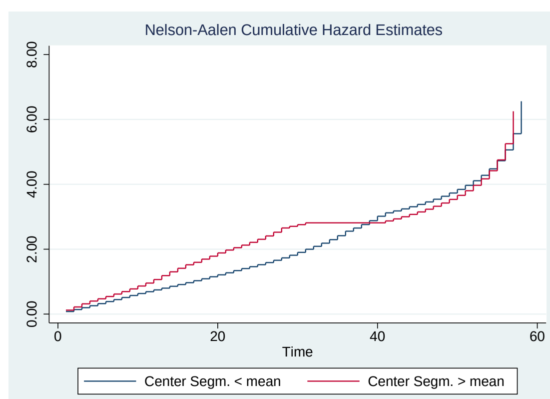
[5]We could have estimated also the smoothed hazard that uses a kernel - based smoothing of the hazard contributions. The smoothed value at a given time is based on a weighted average of the values in the neighbourhood of that point.

the excluded population above the sample mean dominates everywhere the function for countries with lower ethnic exclusion. This means that the duration of ethnic conflicts is positively related to the degree of political exclusion and discrimination. The results of the KM estimation also suggest that countries with the limited or absent political inclusion fight shorter ethnic wars. Countries with high levels of the P_∞ index have lower cumulative hazard curve which implies that the probability of reaching peace at any point in time is lower with respect to the countries with lower P_∞. Since the survival and the cumulative hazard curves cross for the population and the center segmentation variables, their impact on ethnic conflict duration is less clear.









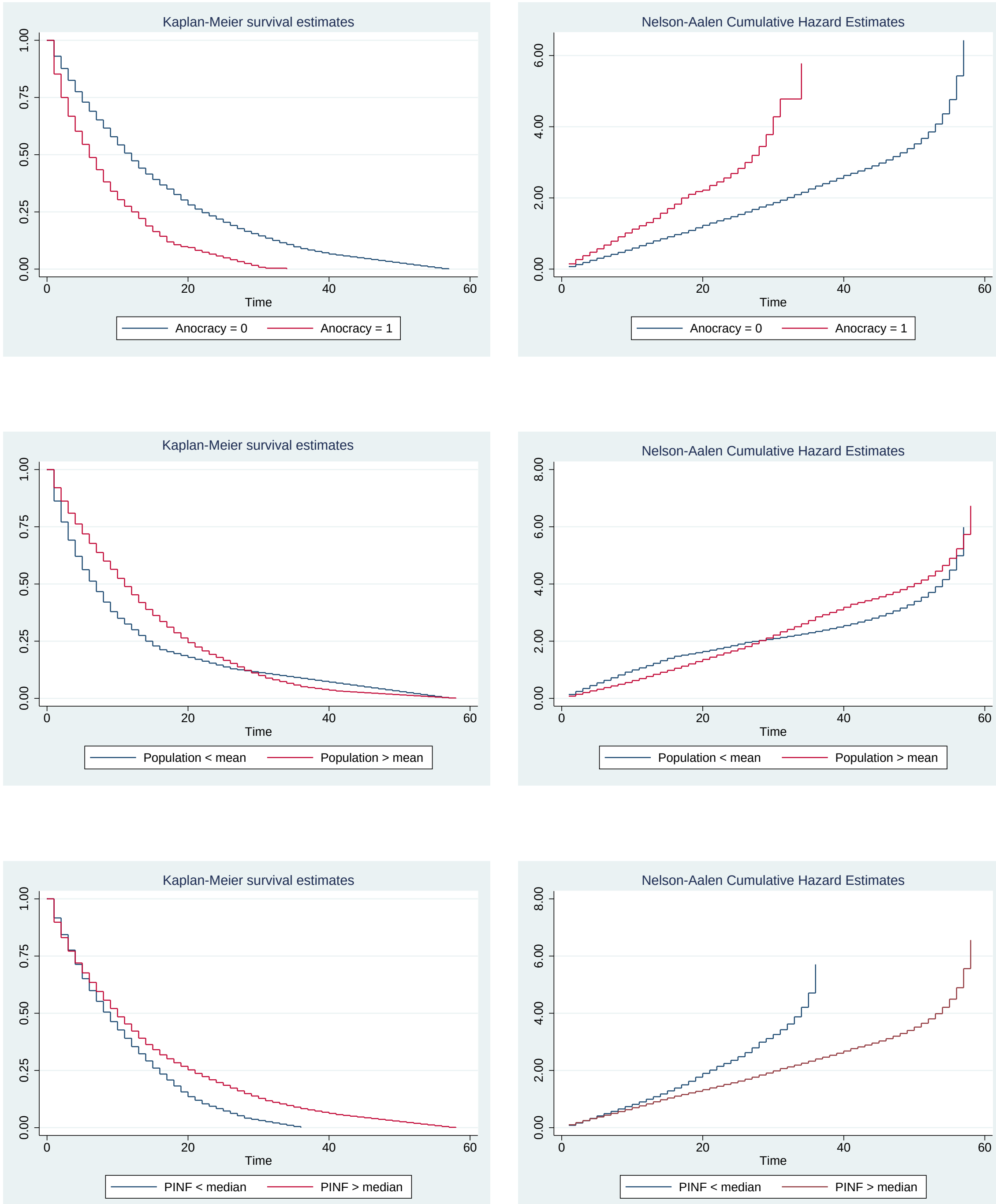


Figure 4.1: Kaplan - Meier Survival function estimation and cumulative hazard function for different explanatory variables.

In order to assess whether the observed sub-group differences between survival curves and hazard functions are statistically significant, we performed the Breslow's and the Logrank test for equality of survivor functions. The Breslow test is more powerful when there is little or no censoring and when the hazard functions are not parallel. Both tests confirm that the observed sub-group differences in the survival functions are statistically significant.

4.3.3 Duration of Ethnic Conflict - A Parametric Survival

The hazard and the corresponding survival rates in Figure 4.1 were modelled as a function of survival time. That is, we did not include any other covariate into anaysis. Now we add an additional dimension to the specification, *i.e.* we allow the hazard and the survival rates to vary between countries depending on their characteristics. There are several different alternatives to deal with the analyses of duration in a parametric and a semi - parametric setting. Each of these alternatives differs in terms of the parametric distribution used to model the hazard function. The most commonly used parametric distributions in the literature on conflict duration are the exponential and the Weibull distributions. The exponential distribution is described only by the mean and it assumes that the instantanous hazard does not vary with time, *i.e.* the hazard function is a horizontal straight line. The Weibull distribution, on the other hand, is less restrictive than the exponential distribution. It is described not only by the mean but also by the shape parameter. Depending on the sign of the shape parameter, the hazard function can either increase, decrease or remain constant with time. In the latter case the Weibull hazard function is equivalent to the exponential hazard function.

Collier, Hoeffler and Soderbom (2004) use a piecewise exponential model to overcome the restrictions implied by the simple exponential model characterized by a contant hazard function. However, most of the literature uses a Weibull model when running a parametric estimation (Fearon, 2004, Gates and Strand, 2004 and DeRouen and Sobek, 2004) since this specification is quite flexible in terms of the implied hazard functions (Montalvo and Reynal-Querol, 2010). Here we adopt the latter approach.

Let $h(t)$ denote the hazard function, λ the scale parameter and p the shape of the hazard function. Assume for the moment that the hazard rate depends only on time t. If the hazard rate is distributed according to the Weibull specification, it takes the following form (Stevenson, 2009):

$$h(t) = \lambda \cdot p \cdot t^{p-1} \tag{4.2}$$

If the shape parameter p is less than 1 instantaneous hazard monotonically decreases with time, if p equals 1 instantaneous hazard is constant over time (and in this case it is equivalent to the exponential hazard function) and if p is greater than 1 instantanous hazard increases with time. In order to see whether the data are consistent with the Weibull specification for the hazard function, we plot a log cumulative hazard as a function of log time. If the choice of the Weibull distribution is correct we should obtain an approximately straight line. Figure 4.2 shows the corresponding plot:

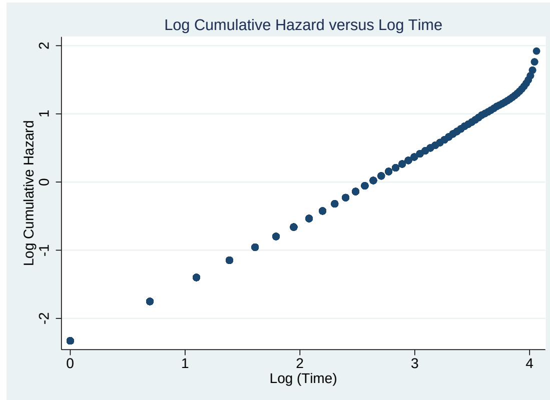


Figure 4.2: Log cumulative hazard as a function of Log time.

The cumulative hazard increases steadily with time at a constant rate up to $t \approx 30$. Figure 4.3 plots the cumulative hazard ratio for countries with low and high P_∞ as well as for those with larger and smaller populations.

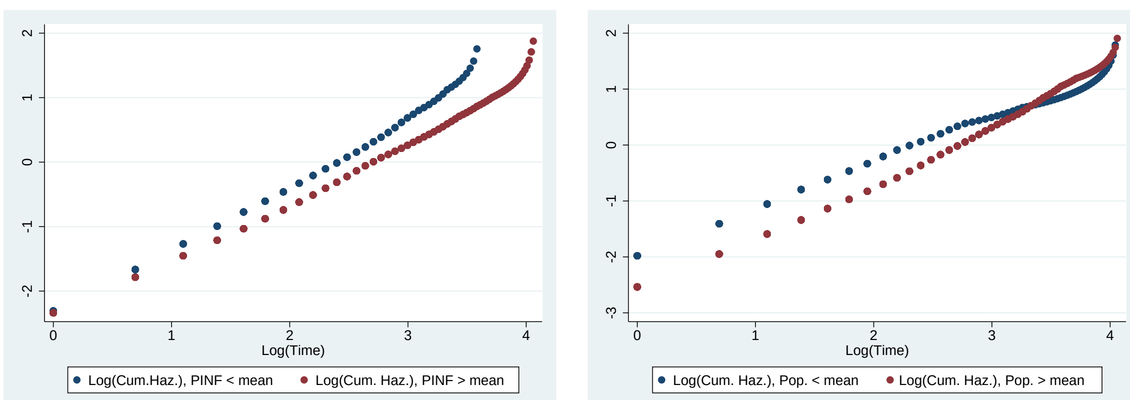


Figure 4.3: Cumulative Hazard Ratio for High and Low P_∞ and Population.

The above graphical representations suggest that the hazard of peace increases with time for both variables. However, the rate of increase in the hazard ratio is *not* proportional for different sub-groups of the covariates. For instance, the hazard function for both low and high P_∞ increases with time but at different rates: countries with lower P_∞ not only have lower survival rates (and, hence higher hazards of peace) in any time period, but they also increase at a higher rate. This means that the impact of P_∞ depends *both* on its levels and on time. The hazard function for different population sizes is less clear. The two hazard functions cross at high t but they seem proportional for short and medium-lenght ethnic conflicts.

Whether the impact of one variable is time dependent is a very important issue, especially because the assumption of proportionality is embedded in the Weibull specification for the hazard function. In order to make this point clearer, lets consider the Weibull specification for the hazard function with only one covariate, x_i that can take only two values, x_{i1} and x_{i2} - x can be a dummy variable that takes the value 1 for $P_\infty > median$ (x_{i2}) and zero otherwise (x_{i1}). The scale parameter λ is now a function of a covariate x_i, that is $\lambda = exp(\beta_0 + \beta_1 x_i)$. The hazard function with a covariate x_i becomes:

$$h(t|x_i) = p \cdot exp(\beta_0 + \beta_1 x_i) \cdot t^{p-1}. \tag{4.3}$$

The ratio of the hazard function of two different values of the covariate, x_{i1} and x_{i1} is:

$$\frac{h(t|x_{i1})}{h(t|x_{i2})} = \frac{p \cdot exp(\beta_0 + \beta_1 x_{i1}) \cdot t^{p-1}}{p \cdot exp(\beta_0 + \beta_1 x_{i2}) \cdot t^{p-1}} = exp(\beta_1(x_{i1} - x_{i1})). \tag{4.4}$$

This ratio is a *constant* proportion that depends only on the covariate and not on time (coefficient β_1 is the same for both x_{i1} and x_{i2}). Thus it is called a proportional hazards model. The assumption of proportionality is often violated. We first estimate our models without controlling for the possibile time dependence of our covariates and in the next section we perform several tests for the validity of the proportionality assumption.

Table 4.1 shows the estimation of the duration of ethnic conflicts under the set of expanatory variables commonly used in the literature[6] and a Weibull specification for the hazard function. The table reports the coefficients associated to each explanatory variable and not the *hazard ratios*. We can obtain the corresponding hazard ratio estimates by simply taking the exponential of the coefficient, namely $Hz = e^{coeff.}$. In order to account for the non-independence of observations from the same state we specify the robust standard errors clustered by country. This is a very important issue especially because many countries experience several ethnic conflicts that follow one another and it is very unprobable that these events are independent. One ethnic conflict may breed another, but it is also true that many conflicts appear to cease only because the number of battle deaths does not reach the threshold level in some period in time, but as soon as the number of victims turns to be higher than the threshold level, the *same* conflict appears as a new conflict episode. Obviously, these two apparently different conflicts cannot be treated as independent of each other.

[6] All these variables were explained in the previous chapter.

Table 4.1: *Weibull Hazard Model - Ethnic War* **Duration**

EW Duration	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Gdp/L	**-0.133***	**-0.112***	**-0.111***	**-0.115***
	0.045	0.042	0.042	0.044
Population	**-0.130**	**-0.214***	**-0.219***	**-0.235***
	0.120	0.120	0.127	0.124
Anocracy	0.346	0.119	0.103	0.096
	0.350	0.344	0.360	0.373
Democracy	-0.093	-0.371	-0.389	-0.331
	0.421	0.369	0.385	0.420
Oil/L	-0.088	-0.031	-0.031	-0.026
	0.060	0.051	0.051	0.052
Mountains	**-0.154***	-0.059	-0.061	-0.061
	0.086	0.097	0.099	0.100
Instability	**0.647***	**0.617***	**0.627***	**0.638***
	0.158	0.153	0.146	0.149
NC State	**0.939***	**0.833***	**0.841**	**0.925***
	0.311	0.317	0.328	0.350
P($\alpha \to \infty$)	**-1.193**	**-1.343**	**-1.362**	**-1.460**
	0.571	0.570	0.576	0.594
Excluded Pop.		**-0.311***	**-0.299**	**-0.265**
		0.115	0.128	0.119
Center Segm.			0.010	0.017
			0.037	0.038
Growth				**-0.207***
				0.118
Peace Duration				**0.099***
				0.009
Constant	-0.123	1.419	1.419	1.132
	1.225	1.250	1.254	1.170
Constant	**0.329***	**0.371***	**0.372***	**0.451***
	0.061	0.062	0.063	0.058
Reg. Dummies	Yes	Yes	Yes	Yes
N. Observations N	742	742	742	741
N. Countries	51	51	51	51
p	1.39	1.4498	1.4501	1.5697
Ln p	0.3293	0.3709	0.3716	0.45
Wald Chi2	106.740***	120.936***	130.604***	445.327***
Bic	1966.106	1922.954	1929.286	1806.758
Aic	1901.575	1853.814	1855.536	1723.814

Notes: *The estimation method is a hazard model with Weibull specification. The absolute z-statistics are calculated using a robust and cluster-adjusted estimator. The Hazard Ratios can be obtained as:* $HR = exp(coeff)$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The estimates suggest that the hazard rate is increasing over time ($p > 1$) but at decreasing rate ($p < 2$). Negative coefficients indicate lower hazard rates, *ceteris* *paribus*. In other words, lower hazard rates imply lower conditional rates of conflict ending and hence longer survival times. For instance, the results from Model 1 suggest that countries with high levels of the P_∞ index experience longer conflicts with respect to the countries with lower P_∞. This result is in line with the evidence obtained from the KM non parametric survival estimation. In Models 2 and 3 we add the share of the excluded population and the number of ethnic groups in power (our proxy for center segmentation). The size of the population now results significant at the 0.1 level and it is positively correlated to the ethnic conflict duration. Countries with large populations have a higher probability of experiencing persitent conflicts. The impact of ethnic exclusion is positive and statistically significant in all model specifications. Interestingly, the above results suggest that countries with a non-contigous territory and frequent political instability experience shorter ethnic conflicts. This result is driven by conflicts in colonial, and hence non-contigous territories which were relatively brief.[7] Gates and Strand (2004) find a positive and statistically significant relationship between political instability and civil conflict duration but they use a very different definition of political instability. Instead of using a dummy variable coded as 1 when a country experience a change in the Polity Score of 3 points or more over the prior three years (as we do here), the authors define political instability as the time since a regime changes.

The only "counterintuitive" result is the effect of the per capita income which is positively associated with the duration of ethnic conflicts in all model specifications. Similar evidence is obtained for the annual growth rate (Model 4). However, plotting duration against income reveals two outliers: the 28 years long conflict in "British" Northern Ireland and the 57 years long Israeli-Palestinian dispute (Figure 4.4). Both

[7]Fearon (2004) shows that the mean duration of non-contigous or decolonization conflicts is about 1.6 shorter than the median case. Although the author consider all civil wars and not only ethnic conflicts, some of the conflict episodes were certainly attributed an ethnic dimension and were mostly motivated by division of power among different ethnic groups.

countries are characterized by a relatively high levels of income per capita (calculated as the average over the period 1946-2005) and a very durature conflicts.

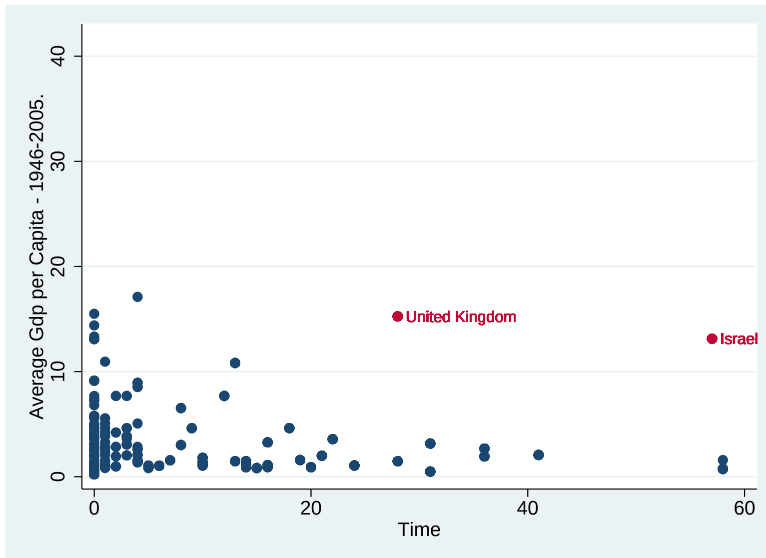


Figure 4.4: Average Gdp per Capita versus Ethnic Conflict Duration.

After dropping these two observations from the analysis, all the variables have the expected sign (Table 4.2).[8] More precisely, the level of economic development and the annual Gdp per capita growth rate are now negatively associated with ethnic conflict duration, though their estimates are statistically insignificant (Model 6). Although with the expected sign, ethnic exclusion is not significantly different from zero, while the degree of center segmentation is statistically significant at the 0.1 level. Finally, the longer is the period of peace before the outbreak of a conflict, shorter is its duration. This evidence may be due to the absense of significant conflict history that breeds further insurgency.

[8]The first 4 columns replicate the estimation results from Table 4.1

Table 4.2: *Proportional Hazard Model - Ethnic War* **Duration**: *Models 1-4 with all countries; Models 5-6 Israel and UK Excluded.*

EW Duration	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Gdp/L	**-0.133*****	**-0.112*****	**-0.111*****	**-0.115*****	0.064	0.062
	0.045	0.042	0.042	0.044	0.060	0.060
Population	-0.130	**-0.214***	**-0.219***	**-0.235***	**-0.297****	**-0.313*****
	0.120	0.120	0.127	0.124	0.116	0.114
Democracy	-0.093	-0.371	-0.389	-0.331	-0.366	-0.324
	0.421	0.369	0.385	0.420	0.462	0.497
Anocracy	0.346	0.119	0.103	0.096	0.057	0.037
	0.350	0.344	0.360	0.373	0.302	0.317
Oil/L	-0.088	-0.031	-0.031	-0.026	-0.034	-0.028
	0.060	0.051	0.051	0.052	0.054	0.054
Mountains	**-0.154***	-0.059	-0.061	-0.061	-0.092	-0.076
	0.086	0.097	0.099	0.100	0.100	0.101
Instability	**0.647*****	**0.617*****	**0.627*****	**0.638*****	**0.663*****	**0.659*****
	0.158	0.153	0.146	0.149	0.141	0.152
NC State	**0.939*****	**0.833*****	**0.841****	**0.925*****	**0.484***	**0.543***
	0.311	0.317	0.328	0.350	0.262	0.286
P($\alpha \to \infty$)	**-1.193****	**-1.343****	**-1.362****	**-1.460****	**-0.943****	**-1.059****
	0.571	0.570	0.576	0.594	0.452	0.484
Excluded Pop.		**-0.311*****	**-0.299****	**-0.265****	-0.168	-0.152
		0.115	0.128	0.119	0.130	0.119
Center Segm.			0.010	0.017	0.059	**0.065***
			0.037	0.038	0.037	0.037
Growth				**-0.207***		-0.263
				0.118		0.166
Number Peace Y.				**0.099*****		**0.092*****
				0.009		0.008
Constant	**0.329*****	**0.371*****	**0.372*****	**0.451*****	**0.350*****	**0.432*****
	0.061	0.062	0.063	0.058	0.049	0.044
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	742	742	742	741	659	658
N. Countries	51	51	51	51	49	49
p	1.39	1.4498	1.4501	1.5697	1.4194	1.5401
Ln p	0.3293	0.3709	0.3716	0.45	0.3502	0.4318
Wald Chi2	106.740***	120.936***	130.604***	445.327***	199.921***	616.231***
Bic	1966.106	1922.954	1929.286	1806.758	1748.230	1639.407
Aic	1901.575	1853.814	1855.536	1723.814	1676.378	1558.602

Notes: *The estimation method is a hazard model with Weibull specification. The absolute z-statistics are calculated using a robust and cluster-adjusted estimator. The Hazard Ratios can be obtained as:* $HR = exp(coeff)$.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Testing the Proportional Hazards Assumption

The Weibull model introduced so far does not impose any particular form for the hazard function, *i.e.* depending on the shape parameter p, it can be either increasing or discreasing or even constant in which case it is equivalent to the exponential hazard function. Although the Weibull model satisfies the general properties of proportional hazard models, it does not guarantee it though. Proportional hazard models assume that the effect of the covariates may increase or decrease the hazard by a proportionate amount at *all* points in time (the ratio of the hazard functions for two different strata of the same covariate is a *constant* proportion that depends *only* on the covariate and *not* on time). This means that the corresponding hazard functions should be parallel. For instance, the hazard curve for the countries with the P_∞ index above the median should be parallel to the hazard function for the countries with P_∞ below the median. The proportionality assumption hence requires that the rate of change of the hazard with time is constant among the two sub-groups of countries. Obviously, the *level* of the hazard for each sub-group of countries may differ across time, but its rate of change should be constant.

In the previous section we have analysed the shape of the KM survival functions for different levels of our main covariates. The KM survival functions and the cumulative hazard functions were never parallel indicating that the proportionality assumption is violated. Figure 4.5 shows the estimated hazard function and the corresponding survival curves for countries with P_∞ above the median (0.65) versus countries with P_∞ below the median.[9] The first two figures refer to the entire set of countries while the remaining two exclude United Kingdom and Israel.

[9]In order to obtain the estimates we run the regression with a dummy variable that takes the value 1 whether P_∞ is above the median and 0 otherwise. This variable is significant at the 0.05 level as it was the coefficient of the P_∞ index in its continuous version.

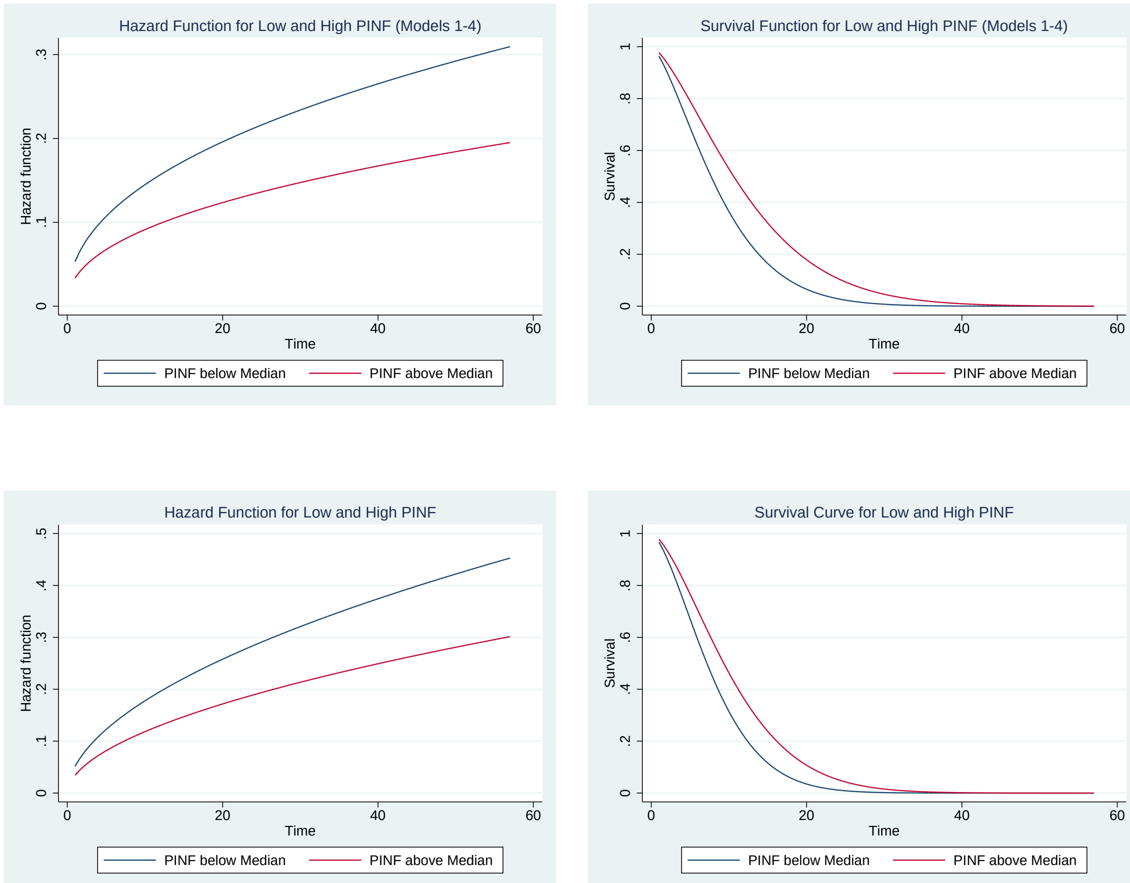


Figure 4.5: Hazard Function and the corresponding Survival Curve for Low and High P_∞. Covariates at means.

The hazard function for countries with high P_∞ is lower than the hazard function for the low P_∞. Similarly, the conflict survival times in countries with high levels of P_∞ are longer than in the rest of the countries. The two survival curves, however, *diverge* with time. The effect of P_∞ is hence *time dependent*. In order to control for the non-proportionality we have mainly two possibilities. We can test for changes in parameter values for coefficients estimated on a sub-sample of the data defined by time or we can interact the covariates (or some of them) with time and test the coefficients.

Here we choose the latter approach. We include interactions of the covariates with the logaritm of time in the baseline regression. We choose to use the logaritm of time because this is the most common function of time used in models with time dependent covariates but any other function of time could be used. If a time dependent variable is significant this indicates a violation of the proportionality assumption for that specific predictor. Table 4.3 shows the results when the interaction terms for P_∞, population size, share of the excluded population and center segmentation are included into regression. We perform the estimation for both samples of countries, *i.e* with and without United Kingdom and Israel. The first column in each part of the table reports our baseline regression estimates.

The results are in line with the patterns observed in Figures 4.1 and 4.6. Three out of four interaction terms are significant which indicates a violation of the proportionality assumption for those specific covariates.[10] As t increases the hazard function for both the low and the high P_∞ countries increases but at *different* rates. The coefficient of the interaction between P_∞ and time suggests that countries characterized by low levels of the P_∞ index will not only reach the end of a conflict before those with the high levels of P_∞ but they will do this faster. In other words, the probability of conflict termination for the low P_∞ countries increases in t at a higher rate with respect to the high P_∞ countries. The same logic applies to the ethnic exclusion and population variables.

[10]The F test that the four interaction variables have zero coefficients is rejected at the 0.001 level.

Table 4.3: *Weibull Hazard Model - Ethnic War* **Duration**: *Models 1-2 with all countries; Models 3-4 Israel and UK Excluded.*

EW Duration	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Interaction Terms	**No**	**Yes**	**No**	**Yes**
Gdp/L	**-0.115*****	-0.055	0.062	0.001
	0.044	0.034	0.060	0.052
Growth	**-0.207***	-0.119	-0.263	-0.110
	0.118	0.187	0.166	0.283
Peace Duration	**0.099*****	-0.013	**0.092*****	-0.016
	0.009	0.021	0.008	0.021
Population	**-0.235***	**1.640*****	**-0.313*****	**1.598*****
	0.124	0.238	0.114	0.276
Excluded Pop.	**-0.265****	**1.432****	-0.152	**1.301***
	0.119	0.621	0.119	0.731
Center Segm.	0.017	0.159	**0.065***	0.129
	0.038	0.205	0.037	0.248
Anocracy	0.096	**-0.360****	0.037	**-0.403****
	0.373	0.181	0.317	0.184
Democracy	-0.331	**-0.510****	-0.324	**-0.453***
	0.420	0.242	0.497	0.268
Oil/L	-0.026	**0.125****	-0.028	**0.109***
	0.052	0.051	0.054	0.065
Mountains	-0.061	**0.145***	-0.076	0.125
	0.100	0.075	0.101	0.077
Instability	**0.638*****	-0.045	**0.659*****	-0.010
	0.149	0.193	0.152	0.183
NC State	**0.925*****	-0.123	**0.543***	-0.229
	0.350	0.337	0.286	0.351
$P(\alpha \to \infty)$	**-1.460****	**4.800****	**-1.059****	**4.981****
	0.594	2.134	0.484	2.312
Variables Interacted with Time				
$P(\alpha \to \infty)$*Time		**-3.027*****		**-2.995****
		1.100		1.223
Population*Time		**-0.838*****		**-0.840*****
		0.102		0.135
Excl.Pop*Time		**-0.865*****		**-0.765****
		0.282		0.348
CenterSegm*Time		-0.095		-0.068
		0.120		0.143
Constant	1.132	**-26.125*****	0.585	**-25.607*****
	1.170	3.433	1.105	3.885
Constant	**0.451*****	**2.649*****	**0.432*****	**2.624*****
	0.058	0.096	0.044	0.105
N. Observations	741	741	658	658
N. Countries	51	51	49	49
Wald Chi2	445.327***	**460.550*****	616.231***	**452.044*****
Bic	1806.758	**-1435.061**	1639.407	**-1221.004**
Aic	1723.814	**-1536.437**	1558.602	**-1319.767**

We also perform a semi-parametric Cox estimation and check for the validity of proportionality assumption. The coefficients and hazard ratio estimates are similar to those obtained with the Weibull specification. Table 4.4 shows the results.

Table 4.4: *Cox Model - Ethnic War* **Duration**: *Models 1-4 with all countries; Models 5-6 Israel and UK Excluded.*

EW Duration	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Gdp/L	**-0.148*****	**-0.128*****	**-0.128****	**-0.127****	0.056	0.046
	0.052	0.050	0.050	0.051	0.064	0.061
Population	-0.099	-0.174	-0.179	-0.186	**-0.276****	**-0.277*****
	0.114	0.114	0.123	0.115	0.111	0.106
Anocracy	0.376	0.178	0.163	0.159	0.090	0.077
	0.338	0.337	0.355	0.352	0.305	0.305
Democracy	0.018	-0.228	-0.245	-0.185	-0.271	-0.216
	0.414	0.381	0.401	0.418	0.489	0.498
Oil/L	-0.084	-0.032	-0.032	-0.028	-0.033	-0.028
	0.060	0.053	0.053	0.051	0.056	0.055
Mountains	**-0.168***	-0.080	-0.081	-0.075	-0.100	-0.080
	0.088	0.101	0.103	0.100	0.104	0.101
Instability	**0.569*****	**0.541*****	**0.550*****	**0.547*****	**0.601*****	**0.580*****
	0.149	0.146	0.139	0.140	0.135	0.142
NC State	**1.001*****	**0.918****	**0.925****	**0.957*****	**0.511***	**0.544***
	0.351	0.357	0.365	0.370	0.288	0.299
P($\alpha \to \infty$)	**-1.104****	**-1.201****	**-1.220****	**-1.235****	**-0.839***	**-0.880****
	0.531	0.539	0.545	0.547	0.432	0.448
Excl. Pop.		**-0.275*****	**-0.264****	**-0.230****	-0.139	-0.123
		0.106	0.116	0.109	0.116	0.107
Center Segm.			0.009	0.012	**0.065***	**0.063***
			0.034	0.035	0.035	0.035
Growth				-0.171		-0.229
				0.117		0.145
Peace Duration				**0.080*****		**0.073*****
				0.007		0.006
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	742	742	742	741	659	658
N. Countries	51	51	51	51	49	49
Wald Chi2	107.862***	108.359***	113.300***	399.614***	173.524***	546.577***
Bic	8174.115	8140.724	8147.100	8059.830	7141.575	7064.285
Aic	8118.802	8080.802	8082.569	7986.102	7078.705	6992.458

Notes: *The estimation method is a Cox model. The absolute z-statistics are calculated using a robust and cluster-adjusted estimator. The Hazard Ratios can be obtained as:* $HR = exp(coeff)$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.5 shows the results of testing the proportionality assumption embedded in the estimation of the Cox model.

Table 4.5: Summary Statistics - Duration

Variable	ρ	p	ρ	p
	"All"		"No"	
Gdp/L	-0.10162	**0.0000*** **	-0.03622	**0.0015*** **
Growth	0.04192	**0.0183** **	0.01623	0.3249
Peace Duration	0.01928	0.3761	-0.02758	0.2449
Population	-0.03851	**0.0001*** **	-0.01654	0.2146
Excluded Population	-0.06627	**0.0000*** **	-0.05755	**0.0001*** **
Center Segmentation	0.00472	0.7456	0.03065	0.0778
Anocracy	0.00452	0.6363	0.04521	**0.0002**
Democracy	0.08028	**0.0000*** **	0.04181	**0.0000*** **
Oil/L	0.00792	0.4732	0.04940	0.0004
Mountains	-0.04381	**0.0004*** **	-0.02697	0.0636
Instability	-0.07456	**0.0000*** **	-0.03886	**0.0092**
NC State	0.10420	**0.0000*** **	0.06428	**0.0000*** **
P($\alpha \to \infty$)	-0.05684	**0.0000*** **	-0.03795	**0.0226** **
Global Test		**0.0000*** **		**0.0000*** **

The parameter ρ represents the coefficient of correlation between the scaled Shoenfeld residuals and time. Figure 4.8 shows the relationship between the scaled Shoenfeld residuals and time for P_∞ and ethnic exclusion. In a well behaved proportional hazard model the residuals are scattered around zero and the line fitted to the residuals has 0 slope. The main idea behind this test is that if the assumption of the proportional hazard holds for a particular covariate then the residuals for that covariate will *not* be related with time.

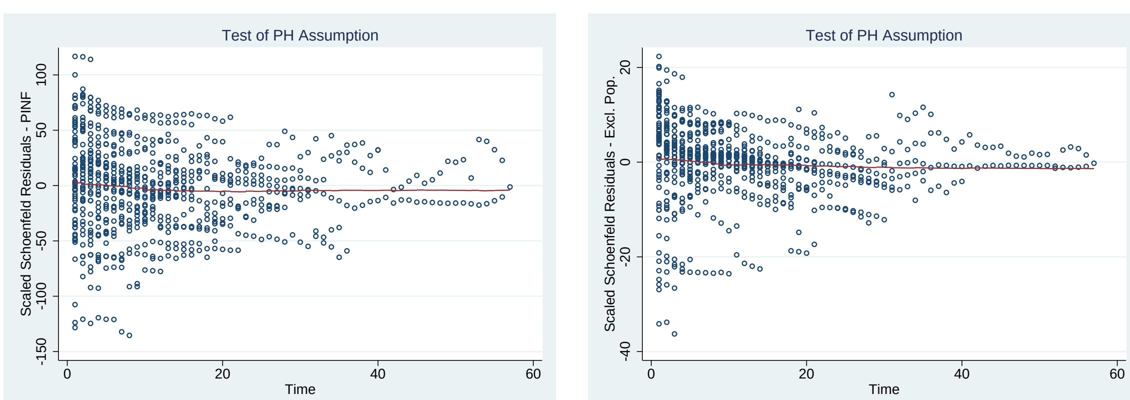


Figure 4.6: Residuals versus Time.

As we can see, the slope of the line fitted to the resuduals is negative and statistically significant for both the P_∞ index and the ethnic exclusion variable. We do not report the relationship betwen the residuals and time for other time varying covariates for the sake of space. However, the patterns are similar to those reported here although the line fitted to the residuals for some covariates has a positive slope.

Robustness Ckecks

In all the previous model specifications the observations were clustered by country. In such a way we accounted for the non-independence of observations from the same country. Here we perform an additional robustness check and we cluster the observations on conflicts. The results from Table 4.6 indicate that in terms of statistical significance the estimators are remarkably robust. The standard errors are slighly reduced with respect to the previous models, but the effect is not substantial.

In addition to the problems of unobserved heterogeneity and the presence of repeated events, there is another important issue discussed in Gates and Strand (2004), namely the presence of parallel conflicts which may significantly contribute to the prolungation of a conflict. In order to control for the possible effects of parallel conflict episodes, we construct a dummy variable that is coded as 1 if there is more than one ethnic conflict ongoing in the same country. The results are reported in Table 4.7. The parallel conflicts dummy variable is significant at the 0.001 level in all model specification except when we control for the non-proportionality. Compared to the baseline model, the magnitude of the coefficient of P_∞ is slightly higher while the magnitude of the population variable is reduced. Center segmentation is now significant at the 0.01 level and its magnitude is increased. Finally, the informational criteria (BIC and AIC) suggest that the latter model specification performs better.

Table 4.6: *Weibull Model - Ethnic War* **Duration**: *Models 1-4 with all countries; Models 5-7 Israel and UK Excluded.* **Cluster on Conflict**

EW Duration	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Gdp/L	**-0.133*** **	**-0.112*** **	**-0.111*** **	**-0.115*** **	0.064	0.062	0.001
	0.043	0.041	0.042	0.044	0.057	0.055	0.053
Population	-0.130	**-0.214** **	**-0.219***	**-0.235***	**-0.297*** **	**-0.313*** **	**1.598*** **
	0.103	0.107	0.127	0.124	0.107	0.107	0.252
Anocracy	0.346	0.119	0.103	0.096	0.057	0.037	**-0.403** **
	0.342	0.338	0.360	0.373	0.299	0.316	0.192
Democracy	-0.093	-0.371	-0.389	-0.331	-0.366	-0.324	**-0.453***
	0.401	0.354	0.385	0.420	0.446	0.484	0.247
Oil/L	-0.088	-0.031	-0.031	-0.026	-0.034	-0.028	0.109*
	0.059	0.051	0.051	0.052	0.053	0.053	0.063
Mountains	**-0.154***	-0.059	-0.061	-0.061	-0.092	-0.076	**0.125***
	0.080	0.093	0.099	0.100	0.095	0.095	0.071
Instability	**0.647*** **	**0.617*** **	**0.627*** **	**0.638*** **	**0.663*** **	**0.659*** **	-0.010
	0.160	0.156	0.146	0.149	0.146	0.156	0.168
NC State	**0.939*** **	**0.833*** **	**0.841** **	**0.925*** **	**0.484** **	**0.543** **	-0.229
	0.298	0.301	0.328	0.350	0.239	0.261	0.373
P($\alpha \to \infty$)	**-1.193** **	**-1.343** **	**-1.362** **	**-1.460** **	**-0.943** **	**-1.059** **	**4.981** **
	0.574	0.571	0.576	0.594	0.440	0.469	2.118
Exc. Pop.		**-0.311*** **	**-0.299** **	**-0.265** **	-0.168	-0.152	**1.301***
		0.114	0.128	0.119	0.127	0.116	0.698
Center Segm.			0.010	0.017	**0.059***	**0.065***	0.129
			0.037	0.038	0.035	0.035	0.222
Growth				**-0.207***		-0.263	-0.110
				0.118		0.181	0.315
Peace Duration				**0.099*** **		**0.092*** **	-0.016
				0.009		0.008	0.020
P($\alpha \to \infty$)*Time							**-2.995*** **
							1.155
Population*Time							**-0.840*** **
							0.124
Exc.Pop*Time							**-0.765** **
							0.334
CenterSegm*Time							-0.068
							0.130
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	742	742	742	741	659	658	658
N. Countries	51	51	51	51	49	49	49
Wald Chi2	113.186***	117.106***	130.604***	445.327***	165.325***	444.248***	369.902***
Bic	1966.106	1922.954	1929.286	1806.758	1748.230	1639.407	-1221.004
Aic	1901.575	1853.814	1855.536	1723.814	1676.378	1558.602	-1319.767

Notes: *The estimation method is a hazard model with Weibull specification. The absolute z-statistics are calculated using a robust and cluster-adjusted estimator. The Hazard Ratios can be obtained as:* $HR = exp(coeff)$.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.7: *Weibull Model - Ethnic War* **Duration**: *Models 1-2 with all countries; Models 3-4 Israel and UK Excluded.* **Parallel Conflicts**

EW Duration	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Gdp/L	**-0.139*** **	0.005	-0.055	0.001
	0.041	0.064	0.034	0.051
Population	-0.160	**-0.231** **	**1.640*** **	**1.599*** **
	0.109	0.111	0.238	0.278
Excl. Pop.	-0.190	-0.096	**1.432** **	**1.296***
	0.118	0.118	0.643	0.760
Center Segm.	0.045	**0.087*** **	0.159	0.129
	0.029	0.029	0.206	0.251
Anocracy	0.002	-0.035	**-0.359** **	**-0.404** **
	0.294	0.273	0.181	0.187
Democracy	-0.684	-0.631	**-0.509** **	**-0.460***
	0.432	0.531	0.236	0.262
Oil/L	**-0.088***	-0.074	**0.125** **	**0.109***
	0.051	0.052	0.051	0.064
Mountains	0.087	0.067	**0.144***	0.129
	0.089	0.093	0.083	0.090
Instability	**0.772*** **	**0.799*** **	-0.045	-0.006
	0.141	0.151	0.192	0.178
NC State	**1.083*** **	**0.739** **	-0.123	-0.225
	0.336	0.310	0.330	0.339
Parallel Conflicts	**-1.186*** **	**-1.059*** **	0.003	-0.021
	0.272	0.248	0.191	0.198
P($\alpha \to \infty$)	**-1.344** **	**-1.085** **	**4.798** **	**4.996** **
	0.576	0.546	2.153	2.363
Growth	-0.143	-0.197	-0.119	-0.110
	0.130	0.170	0.187	0.282
Peace Duration	**0.096*** **	**0.090*** **	-0.013	-0.016
	0.009	0.008	0.021	0.021
P($\alpha \to \infty$)*Time			**-3.026*** **	**-3.005** **
			1.115	1.263
Population*Time			**-0.839*** **	**-0.840*** **
			0.102	0.135
Exc.Pop*Time			**-0.865*** **	**-0.762** **
			0.293	0.362
CenterSegm*Time			-0.095	-0.067
			0.121	0.145
Reg. Dummies	Yes	Yes	Yes	Yes
N. Observations	741	658	741	658
N. Countries	51	49	51	49
Wald Chi2	311.638***	488.985***	486.184***	485.503***
Bic	1715.547	1568.887	-1428.454	-1214.551
Aic	1627.995	1483.593	-1534.438	-1317.803

Notes: *The estimation method is a hazard model with Weibull specification. The absolute z-statistics are calculated using a robust and cluster-adjusted estimator. The Hazard Ratios can be obtained as:* $HR = exp(coeff)$.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.5 shows the median and the mean duration of ethnic conflicts in our sample. We refer to the whole sample of countries with *"All"* while *"No"* indicates the restricted sample, *i.e.* without United Kingdom and Israel.

Table 4.8: Summary Statistics - Duration

Variable	**Sample**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Mean Duration	(All)	12.868	6.532	0.354	40.445
Median Duration	(All)	11.343	5.758	0.312	35.652
Mean Duration	(No)	11.019	5.611	0.423	25.048
Median Duration	(No)	9.65	4.914	0.371	21.937

The mean duration of ethnic conflicts between 1946 and 2005 was 12.9 years (if we exclude the two "outliers", the mean duration reduces to 11 years). More interesting for the purposis of our analysis is to look at the mean and the median duration of ethnic conflicts in countries with high P_∞ versus countries with low P_∞ (as before, to separate these two groups we use the median level of P_∞). Table 4.6 shows the summary statistics for ethnic conflict duration both for low and for high P_∞ as well as for the entire sample (labeled with "All") and without two "outliers" (labeled with "No").

Table 4.9: Summary statistics - Duration - High and Low P_∞

Variable	P_∞	**Sample**	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Mean Duration	High	(All)	14.017	6.696	0.354	40.445	541
Mean Duration	Low	(All)	9.762	4.869	0.592	30.937	200
Median Duration	High	(All)	12.356	5.903	0.312	35.652	541
Median Duration	Low	(All)	8.605	4.292	0.522	27.271	200
Mean Duration	High	(No)	12.053	5.703	0.423	25.048	541
Mean Duration	Low	(No)	8.220	4.244	0.464	20.294	200
Median Duration	High	(No)	10.556	4.995	0.371	21.937	541
Median Duration	Low	(No)	7.199	3.717	0.406	17.773	200

Countries with P_∞ above the median experience (on average) 4.32 more years of conflict (3.83 years more without UK and Israel).

Marginal Effects

Table 4.5 shows the marginal effects of our explanatory variables on the duration of ethnic conflict. The calculation of marginal effects of the explanatory variables is base on the models which do not include the interaction terms (Column 1 and Column 3 in Table 4.2).

Table 4.10: Marginal Effects

	"All"	**"All"**	**"No"**	**"No"**
Variable	**Mean D.**	**Median D.**	**Mean D.**	**Median D.**
Gdp/L	**0.703***	**0.542***	-0.349	-0.391
	0.259	0.189	0.339	0.438
Growth	**1.264***	0.975	1.489	1.668
	0.707	0.603	0.926	1.050
Peace Duration	**-0.604***	**-0.466***	**-0.519***	**-0.581***
	0.067	0.149	0.056	0.181
Population	**1.435***	1.106	**1.771***	**1.984***
	0.781	0.794	0.666	1.056
Excluded Pop.	**1.621****	**1.250***	0.863	0.967
	0.727	0.726	0.672	0.845
Center Segm.	-0.104	-0.080	**-0.368***	-0.413
	0.232	0.179	0.212	0.266
Anocracy (d)	-0.580	-0.438	-0.208	-0.231
	2.236	1.724	1.786	1.991
Democracy (d)	2.107	1.738	1.946	2.289
	2.784	2.199	3.150	3.726
Oil/L	0.161	0.124	0.161	0.181
	0.312	0.214	0.305	0.305
Mountains	0.372	0.287	0.432	0.484
	0.609	0.434	0.567	0.576
Instability (d)	**-3.469***	**-2.472***	**-3.336***	**-3.403***
	0.743	0.805	0.698	1.156
NC State (d)	**-4.969***	**-3.295***	**-2.826****	**-2.903****
	1.694	1.222	1.422	1.386
$P(\alpha \to \infty)$	**8.923****	**6.881****	**5.997****	**6.719***
	3.622	3.147	2.724	3.456

The marginal impact of the P_∞ index is positive and statistically significant. If P_∞ was to increase from 0 to 1, the mean duration of conflict would increase by 8.9 or 6 years. An increase in P_∞ of one unit of standard deviation (.3015) from its mean (0.591) value would increase the duration by 2.68 ("All") or 1.8 ("No"). Conflicts over discountigous territory tend to be aproximately 5 or 3 years shorter than the rest.

One additional year of peace before the outbreak of conflict shortens its duration by 0.6 or 0.5 years. Finally, the marginal effect of the population size is large and highly significant: one standard deviation increase in the population size from the mean increases the duration of conflict by 1.93 or 2.38 years. Finally, an increase in the share of the exlcuded population from its mean (6.7%) by one unit of standard deviation (4.7%) increases the mean duration of conflicts by 2.51 years.

Figure 4.6 shows graphically the predicted mean duration of ethnic conflicts for different quintiles of the population and for several possibile duration of the peace time preceding conflict as function of the P_∞ index of conflict potential.

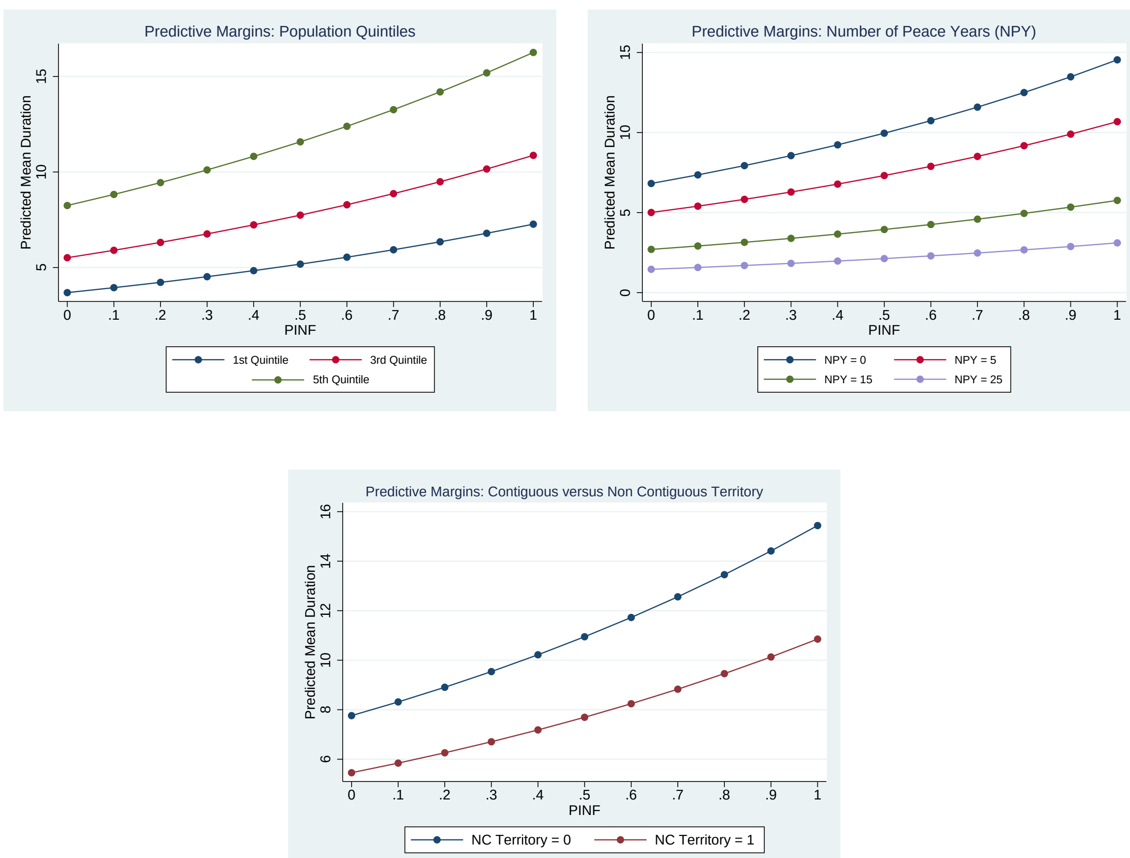


Figure 4.7: Predicted Mean Duration of Conflict.

4.4 Concluding Remarks

In this chapter we argue that onset and duration are two distinct concepts that should be understood and treated differently. While the determinants of both aspects of conflict might overlap, they are not necessarily the same. Our results from the previous chapter strongly suggest that countries with low levels of economic development, large populations and high degrees of political exclusion and competition along ethnic lines face higher risk of conflict. Moreover, we show that ethnic diversity is an important correlate to ethnic conflict. We apply our index of conflict potential that combines different aspects of ethnic distribution in a non-linear way and find that, once the power and the interaction are properly combined into one single measure, ethnic diversity becomes a significant and robust correlate to ethnic conflict onset.

In this chapter we went one step further and explored empirically the correlates to ethnic conflict duration or persistence. We were particularly interested in analyzing whether the features of ethnic diversity together with ethnic politics and competition dynamics are a possible dimensions of the propagation mechanisms of ethnic conflicts. However, as in the case of ethnic conflict onset, the problem of potential endogeneity did not allow us to address the issues of causality, so we limited our attention to the association between several explanatory variables and the duration of ethnic conflict.

The novelty of our approach is twofold. First, to the best of our knowledge there is only one empirical study that considers ethnic conflicts and not civil conflicts in general. Second, we address in detail the issues of non-proportionality of the hazard function and analyse the role of ethnic politics and ethnic competition in the determination of ethnic conflict persistence. Our results strongly support the role of the P_∞ index in making ethnic conflicts more persistent and economically extremely costly. We show that ethnic conflicts in countries with high levels of P_∞ are on average

4 years longer than conflicts in countries with low levels of P_∞. Moreover, if the P_∞ index was to increase from 0 to 1, the mean duration of ethnic conflict would increase by 6 years. The effect of P_∞, however, varies both in levels and with time. Countries with lower P_∞ have lower probabilites of conflict continuation and the probability of conflict ending increases at a higher rate with respect to the rest of the countries.

We also find that autocracies fight shorter wars with respect to more inclusive political systems. The effects of the Gdp per capita and the annual growth rate are not significantly different from zero, while large populations have singificant positive impact on conflict duration. The number of peace years and the non-contiguity of the countries territory make conflicts shorter while the share of the excluded population increases significantly the mean duration of ethnic conflicts. Instability results negatively related to conflict duration. However, together with the negative impact of the non-contiguous territory, the latter evidence can be explained with the fact that countries with non-contiguous territory and frequent government turnovers are also those characterized by decolonization conflicts which, according to the existing literature, were on average significantly shorter than the rest of ethnic conflicts.

We address the problems of unmeasured heterogeneity and the presence of "repeated events" by clustering the observations by country and by conflict. The problem of contemporaneous events occuring in a single country is controled by the inclusion of a dummy variable for the presence of paralle conflicts. The latter variable is highly significant but it does not alter the significance of our main covariates of interest. Moreover, we check the robustness of our results to alternative estimation techniques and the results do not change.

Before concluding, we want to emphasize that we are completely aware of the presence of many open questions that should be addressed in the analysis of conflict duration. Eventhough we believe some of the problems have been successfully addressed in this chapter, many things still remain to be improved. One of these concerns the problem of the precise dating of ethnic or civil conflicts in general. Here we measure time in years, but this is surely not the optimal choice. For all this reasons, the analysis provided in this chapter can be considered as preliminary, but it can represent a good starting point for the future investigation of ethnic conflict duration and the policy interventions needed to shorten long and persistent conflicts.

Bibliography

1. Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat and R. Wacziarg (2003) "Fractionalization",Journal of Economic Growth 8: 155-194.

2. Atlas Narodov Mira, Atlas of the People of the World. 1964. Moscow: Glavnoe Upravlenie Geodezii i Kartografii .Bruck, S.I., and V.S. Apenchenko (eds.).

3. Banzhaf, J. F. (1965): Weighted Voting Doesn't Work: A Mathematical Analysis, Rutgers Law Review, 19, 317-343

4. Cameron A. and Pravin K. Trivedi, Microeconometrics Using Stata, Revised Edition, 2010, Stata Press books

5. Cederman, L.E., A. Wimmer and B. Min. 2009. "Ethnic Politics and Armed Conflict: A configurational analysis of a new global data set." American sociological Review 74: 316-337.

6. Cederman, L.E., A. Wimmer and B. Min. 2009. "Ethnic Politics and Armed Conflict: A configurational analysis of a new global data set - ONLINE APPENDIX" American sociological Review 74 (2).

7. Cederman, Lars-Erik and Luc Girardin. 2007. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies." American Political Science Review 101:173-85.

8. Chakravarty, S. R. and B. Maharay. 2012. Ethnic Polarization ordering and indices. Journal of Economic Interaction and Coordination 7: 99-123.

9. Chakravarty, S. R. and B. Maharay. 2011. Measuring Ethnic Polarization. Social Choice and Welfare 37: 431-452.

10. Collier, Paul and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." Oxford Economic Papers 56:563-95.

11. Collier, Paul, Anke Hoeffler, and Dominic Rohner. 2006. "Beyond Greed and Grievance: Feasibility and Civil War." Center for the Study of African Economics, Working Paper 10.

12. Collier, P., A. Hoeffler and M. Soderbom. 2004. On the duration of civil war. Journal of Peace Research 41(3): 253-73

13. DeRouen, K. and D. Sobek. 2004. The dynamics of civil war duration and outcome. Journal of Peace Research 41 (3): 303-320.

14. Duclos, J., Esteban, J. and D. Ray. 2004. Polarization: concept, measurement and estimation. Econometrica 74: 1737-72.

15. Easterly, W., and R. Levine. 1997. Africa's growth tragedy: Policies and Ethnic divisions. Quarterly Journal of Economics CXII(4): 1203-1250.

16. Eichhorn, W. (1978): Functional Equations in Economics. London, Addison-Wesley.

17. Elbadawi, Nicholas and Nicholas Sambanis. 2000. "Why Are There So Many Civil Wars in Africa? Understanding and Preventing Violent Conflict." Journal of African Economics 9:244-69.

18. Esteban, J. M and D. Ray. 1994. On the measurement of Polarization. Econometrica, 62 (4)

19. Esteban, J. M and D. Ray. 2008. Polarization, Fractionalization and Conflict. Journal of Peace Research, 45: 163-182.

20. Esteban, J. M and D. Ray. 1999. Conflict and Distribution. Journal of Economic Theory, 87: 379-415.

21. Esteban, J. M and G. Schneider. 2008. Polarization and Conflict: Theoretical and Empirical Issues. Journal of Peace Research, 45 (2): 131-141.

22. Felsenthal, D. and M. Machover (1998): The Measurement of Voting Power, Edward Elgar, Cheltenham.

23. Fearon J. and D. Laitin. 2003. Ethnicity, Insurgency, and Civil War. American Political Science Review 97(1): 75-90.

24. Fearon, J. 2004. Why do some wars last so much longer than others?. Journal of Peace Research 41: 275-301.

25. Fearon, James D., Kimuli Kasara, and David D. Laitin. 2007. "Ethnic Minority Rule and Civil War Onset." American Political Science Review 101:187-93.

26. Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." American Political Science Review 97:1-16.

27. Foster, J. and M.C. Wolfson (2009), "Polarization and the Decline of the Middle Class: Canada and the US", Journal of Economic Inequality forthcoming.

28. Gates, S. and H. Strand. 2004. Modelling the duration of civil wars: measurement and estimation issues. Mimeo, Centre for the Study of Civil Wars, PRIO.

29. Gini, C. (1912). Variabilita' e mutabilita'. Studi economico giuridici della Facolta' di Giurisprudenza dell'universita' di Cagliari, a. III, parte II.

30. Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand. 2002. Armed Conflict 1946-2001: A New Dataset. Journal of Peace Research 39 (5): 615-37.

31. Hegre, H. and N. Sambanis. (2006). Sensitivity Analysis of Empirical Results on Civil War Onset. Journal of Conflict Resolution 50: 508-535.

32. Hegre, Havard, Tanja Ellingsen, Scott Gates, and Nils Petter Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816-1992." The American Political Science Review 95:33-48.

33. Hegre, Havard and Nicholas Sambanis. 2006. "Sensitivity Analysis of Empirical Results on Civil War Onset." Journal of Conflict Resolution 50:508-35.

34. Hirshleifer, J. (1989). Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. Public Choice 63: 101-112.

35. Hirshleifer, J. (1991). The paradox of power. Economics and Politics 3: 177-200.

36. Hirshleifer, J. (1995). Anarchy and its breakdwon. Journal of Political Economy 103: 26-52

37. Horowitz, Donald. 1985. Ethnic Groups in Conflict. Berkeley, CA: University of California Press.

38. Montalvo, J. G. and M. Reynal-Querol. 2002. Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth. UPF Working Paper 660.

39. Montalvo, J. G. and M. Reynal-Querol. 2005. Ethnic polarization, potential conflict and civil wars. American Economic Review 95 (3): 796-816.

40. Montalvo, J. G. and M. Reynal-Querol. 2005. Ethnic Diversity and Economic Development. Journal of Development Economics 76: 293-323.

41. Montalvo, J. G. and M. Reynal-Querol. 2002. Ethnicity, Political Systems, and Civil Wars. Journal of Conflict Resolution 46 (1): 29-54.

42. Montalvo, J. G. and M. Reynal-Querol. 2010. Ethnic polarization and the duration of civil wars. Economics of Governance 11: 123-143.

43. Posner, Daniel. 2004. "Measuring Ethnic Fractionalization in Africa." American Journal of Political Science 48:849-63.

44. Sambanis, Nicholas. 2001. "Do Ethnic and Nonethnic Civil Wars Have the Same Causes?" Journal of Conflict Resolution 45:259-82.

45. Sambanis, Nicholas. 2004. "What is Civil War? Conceptual and Empirical Complexities of an Operational Definition." Journal of Conflict Resolution 48:814-58.

46. Sambanis, Nicholas. 2009. "What is an Ethnic War? Organization and Interests in Insurgencies." Yale University, Department of Political Science, New Haven, CT. Unpublished manuscript.

47. Shapley, L.S. (1953), A Value for n-Person Games, in H.W. Kuhn and A.W. Tucker (eds.), Contributions to the Theory of Games II (Annals of Mathematics Studies), Princeton: Princeton University Press, pp. 307-317.

48. Shapley, L.S. and M.J. Shubik (1954): A Method for Evaluating the Distribution of Power in a Committee System, American Political Science Review, 48, 787-792.

49. Skaperdas, S. 1998. On the formation of alliances in conflict and contests. Public Choice 96: 25-42.

50. Stevenson, M. (2009). An Introduction to survival analysis. EpiCentre, IVABS, Massey University, Unpublished manuscript.

51. Tullock, G. 1980. Efficient rent seeking. In J.M. Buchanan, R.D. Tollison and G. Tullock (Eds.), Toward a theory of the rent-seeking society. College Station: Texas $A\&M$ University Press.

52. Vanhanen, Tatu. 1999. "Domestic Ethnic Conflict and Ethnic Nepotism: A Comparative Analysis." Journal of Peace Research 36:55-73.

53. Wang, Y.Q. and K.Y. Tsui (2000), "Polarization Orderings and New Classes of Polarization Indices," Journal of Public Economic Theory 2, 349-363

54. Wimmer, Andreas. 2002. Nationalist Exclusion and Ethnic Conflicts: Shadows of Modernity. Cambridge, UK: Cambridge University Press.

55. Wimmer, Andreas. 2008. "The Making and Unmaking Of Ethnic Boundaries: A Multi-Level Process Theory." American Journal of Sociology 113:970-1022.

56. Wimmer, Andreas and Brian Min. 2006. "From Empire to Nation-State: Explaining Wars in the Modern World, 1816-2001." American Sociological Review 71:867-97.

57. Wolfson, M. C. 1994. When inequalities diverge. American Economic Review, Papers and Proceedings 84 (2): 353-8.

58. Wolfson, M.C. (1997), "Divergent Inequalities: theory and empirical results", Review of Income and Wealth 43 401-421

Estratto per riassunto della tesi di dottorato

Studente: Matija Kovacic
Matricola: 955559
Dottorato: Economia
Ciclo: 24

Titlo della tesi: Ethnic Distribution, Effective Power and Ethnic Conflict // Distribuzione Etnica, Potere Effettivo e Conflitto Etnico

Abstract: This thesis is investigating both in theoretical and empirical terms the relationship between the features of ethnic distribution and the probability of conflict. The thesis is composed of four chapters. The first chapter is the introduction to the thesis in which I briefly summarize the main problems related to the existing literature on ethnic diversity and conflict and I present an overview of the measures of ethnic diversity commonly used in the literature.

In the second chapter I propose a theoretical model that specifies the potential of conflict in a society as a function of the population distribution across ethnic groups. I axiomatically derive a general parametric class of indices of conflict potential that combines the groups' effective power and the between-groups interaction. The effective power of a group is a function of a group's relative size but it also depends on the relative sizes of all the other groups in the population. The interaction component, on the other hand, is given by the probability of interaction between the members of one group with those of other groups. I show that for certain parameter values the index reduces to the existing indices of ethnic diversity, while in general the indices combine in a non-linear way three different aspects of ethnic diversity, namely the fractionalization, the polarization and the ethnic dominance. The results of the model share some common features with the literature on conflicts in contests and the literature on voting power indices. In particular, the power component of the extreme element of the class of indices is intuitively related to the definition of voting power in a simple majority game. In that particular case, the value of the effective power is given by the relative Penrose-Banzhaf index of voting power calculated over the shares of populations associated to each ethnic group.

In the third chapter I investigate empirically the role of ethnic diversity in the explanation of the ethnic conflict outbreak. The empirical performance of the indices of conflict potential developed in the second chapter is tested against the existing distributional indices of ethnic diversity within the context of the commonly used logistic model that focuses on the onset of ethnic conflicts in a time range from 1946 to 2005. Together with the set of the explanatory variables for structural conditions and country characteristics, I take advantage of the recent "Ethnic Power Relations" data set which includes additional information on the political exclusion and competition

along ethnic lines and it offers the best coding for ethnic wars. The results obtained suggest that the indices of conflict potential outperform the existing indices of ethnic diversity in the explanation of ethnic conflict onset. This evidence is robust to the inclusion of a larger set of regressors, time and regional controls as well as to several other estimation techniques.

The fourth chapter explores empirically the determinants of conflict duration with a particular attention to the potential role of ethnic diversity together with ethnic politics and competition dynamics. The first part of the chapter presents an overview of the existing literature on conflict duration, the main data cources and the related econometric issues. The second part of the chapter consists in a non-parametric and a parametric survival analysis of the duration of ethnic conflict where we address in detail the issues of non-proportionality of the hazard function, the unmeasured heterogeneity and the presence of "repeated events". The results suggest that there is a statistically significant and robust association between ethnic distribution and conflict duration, together with other commonly used explanatory variables in the literature on conflict duration.

Estratto: Questa tesi di dottorato consiste in un'analisi teorica ed empirica della relazione tra diverse caratteristiche della distribuzione etnica e conflitto. La tesi e' divisa in 4 capitoli. Il primo capitolo e' l'introduzione alla tesi in cui si riassumono i principali problemi riscontrati nella letteratura esistente insieme ad un breve ripasso degli indicatori della diversita' etnica.

Nel secondo capitolo si propone un modello teorico che specifica il potenziale di conflitto in funzione alle caratteristiche della distribuzione etnica. Si deriva in modo assiomatico una famiglia di indicatori che combinano il potere effettivo dei gruppi con la probabilita' di interazione tra diversi gruppi etnici. Il potere effettivo di un gruppo non dipende solo dalla sua grandezza ma anche dalla grandezza di tutti gli altri gruppi nella societa'. Si dimostra che per certi valori dei parametri, gli indicatori proposti assumono forma delle misure della diversita' etnica esistenti nella letteratura, mentre in generale l'indicatore combina in modo non-lineare i tre principali aspetti della diversita' etnica: frammentazione, polarizzazione e dominanza. La componente di potere dell'elemento estremo della famiglia di indicatori e' intuitivamente legata alla definizione del potere di voto di Banzhaf e Shapley-Shubik. In quel caso particolare la componente di potere e' esattamente uguale all'indice relativo del potere di voto di Banzhaf.

Il terzo capitolo consiste in una dettagliata analisi empirica sul ruolo della diversita' etnica nel conflitto. Si verifica la performance empirica degli indicatori sviluppati nel secondo capitolo rispetto alle misure esistenti nella letteratura. Tutto questo va fatto nel contesto di un modello logistico dove come variabile dipendente si ha una variabile binaria uguale ad uno nell'anno in cui scoppia il conflitto e 0 nel resto degli anni. Si dimostra che la differenza tra gli indicatori del potenziale di conflitto e altre

misure esistenti non e' solo teorica ma e' anche empirica. Gli indicatori del potenziale di conflitto hanno un potere esplicativo singificativamente superiore rispetto ad altre misure della diversita' etnica. Questa evidenza e' robusta.

Nel quarto capitolo si studiano le determinati della durata dei conflitti etnici usando sia le tecniche non parametriche che quelle parametriche. I risultati confermano l'importanza degli aspetti distributivi dell'etnicita' per la persistenza dei conflitti insieme a diverse variabili politiche, tra i quali il grado di esclusione e di discriminazione etnica nel contesto decisionale (politico).