Working Paper Series

Ethnic Distribution, Effective Power and Conflict

Matija Kovacic and Claudio Zoli

ECINEQ WP 2013 – 294

ECINEQ 2013 – 294
April 2013

www.ecineq.org

Ethnic Distribution, Effective Power and Conflict

Matija Kovacic†
University Ca' Foscari Venice

Claudio Zoli
University of Verona

Abstract

Ethnic heterogeneity can potentially be related to the occurrence of conflicts with long-lasting economic effects. Two main measures of ethnic heterogeneity are employed in the econometric literature on ethnic diversity and conflict: the Gini heterogeneity or fractionalization index and the discrete polarization index. However, still no broad consensus is reached on which distributional aspect of ethnic diversity is associated with the outbreak of conflict. In this paper we argue that the relative importance of each pattern of ethnic diversity depends on the trade-off between the groups' power and its interaction with other groups. Following the Esteban and Ray [On the measurement of polarization, Econometrica, 62(4), 1994] approach to social antagonism, we axiomatically derive a parametric class of indices of conflict potential that combines the groups' effective power and the between-groups interaction. We use a discrete metric to define the distances between groups and we do not treat each group as a unitary actor. Moreover, we assume that the effective power of a group depends not only on its own relative size but also on the relative size of all the other groups in the population. We show that for certain parameter values the obtained indices reduce to the existing indices of ethnic diversity, while in general the derived indices combine in a non-linear way three different aspects of ethnic diversity, namely the fractionalization, the polarization and the ethnic dominance. The power component of the extreme element of the class of indices is given by the relative Penrose-Banzhaf index of voting power. The results from our empirical exercise show that the derived extreme index outperforms the existing indices of ethnic diversity in the explanation of ethnic conflict onset.

Keywords: Ethnic distribution, Conflict, Polarization, Fractionalization, Power indices, Dominance.

JEL Classification: O11, Z13, O57, D63, D72, D74.

† **Contact details: Matija Kovacic:** Department of Economics, University Ca' Foscari, Canareggio 873, 30121 Venice; Tel: +39 340 3349885; E-Mail: matija.kovacic@unive.it
Claudio Zoli:.Department of Economics, University of Verona, Via dell'Artigliere, 19, 37129 Verona; Tel: +39 045 8028479; E-Mail: claudio.zoli@univr.it

"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. "Ethnic groups in conflict".

1 Introduction

Ethnic heterogeneity can potentially be related to the occurrence of conflicts with long-lasting economic effects. Although ethnic grievances play a prominent role in most recent conflicts, still no broad consensus is reached on whether or not ethnic heterogeneity is an important correlate to conflict. In order to quantify the degree of ethnic heterogeneity, the literature relies on a variety of indicators capturing different features of population distribution across ethnic groups. From a distributional point of view we can distinguish between three basic manifestations of ethnic diversity: ethnic fractionalization, ethnic polarization and ethnic dominance. The index of ethnic fractionalization is directly derived from the Gini's inequality index by assuming the uniform distance between any two different ethnic groups (Alesina et al, 2003)[1] while the index of discrete ethnic polarization is the discrete version of the Esteban and Ray's (1994) income polarization index (Montalvo and Reynal-Querol, 2002, 2005). 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. Ethnic dominance, on the other hand, emphasizes the role of the size predominance of one ethnic group over others and is captured by an indicator that takes the value of one if the relative size of one single group exceeds some threshold and zero otherwise.

The empirical evidence on the association between ethnic diversity and conflict is very heterogeneous. Applying the fractionalization index, Sambanis (2001, 2004) and Hegre and Sambanis (2006) find a positive and statistically robust association between ethnic fractionalization and ethnic conflict and argue that as a country becomes ethnically more fragmented, the probability of ethnic civil conflict increases. Collier (2001) and Collier and Hoeffler (2004) show that the interaction between ethno-linguistic and religious fractionalization (which they term as "social fractionalization") is negatively correlated with the likelihood of conflict because ethnic diversity makes rebellion harder since rebel cohesion becomes more costly. The "benign effects of social fractionalization" (Collier and Hoeffler, 2004, p.588), however, disappear in the presence of ethnic dominance, *i.e.*, with one ethnic group covering between 45% and 90% of the population. On the other side, Fearon and Laitin (2003) and Fearon, Kasara and Laitin (2007) find no significant effect of ethnic and religious fractionalization on the likelihood of civil conflict outbreak. Similarly, Cederman and Girardin (2007) and

[1] Although originally proposed by Corrado Gini in 1912 as an index of variability for categorical variables, the index of fractionalization is often compared with the Hirschman - Herfindahl concentration index.

Cederman, Min and Wimmer (2009, 2010) show that once we account for the political exclusion and competition along ethnic lines, ethnic diversity "... in and of itself has no effect on the likelihood of civil conflict ..." (Cederman, Min and Wimmer, 2009, p.319). 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) apply their index of discrete ethnic polarization and find a positive and statistically significant association between ethnic polarization and the incidence (or the occurrence) of conflict.[2] Schneider and Wiesehomeier (2010), on the other hand, find that the relationship between ethnic polarization and conflict is ambiguous and depends on whether it is considered civil war incidence or civil war onset as an outcome variable while Collier and Hoeffler (2004) find no statistically significant relationship between ethnic polarization and the risk of conflict outbreak.

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 explanations 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 length by Sambanis (2004) and Gates and Strand (2004): some authors consider "real wars", *i.e.*, those internal conflicts that count more than 1000 battle deaths in a single year (Sambanis, 2001; Fearon and Laitin, 2003; Collier and Hoeffler, 1998, 2004), others consider civil "conflicts", or those that count at least 25 battle deaths in a single year (Montalvo and Reynal-Querol, 2005). Moreover, there are only few serious attempts to code *ethnic* conflicts which are different from civil conflicts (Sambanis, 2001, 2004; Cederman, Min and Wimmer, 2009). Second, there are at least three different aspects of conflict that can be considered, namely *incidence*, *onset* and *duration*. Finally, there are three main sources for data on ethnic distribution (World Christian Encyclopedia, WCE; Encyclopedia Britannica, EB; and Atlas Narodov Mira, ANM) that use different markers to identify ethnic (and religious) groups.

Without discussing the differences between data sources and conflict codings in detail, here we propose another possible explanation for the variation in these results. The above mentioned 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 (but they are robust to the alternative definition of conflict in terms of a specific "battle-death" threshold), and the Collier and Hoeffler's (2004) dominance dummy analyses. Here we argue that the relevance of each distributional aspect of ethnicity may depend on the features of the population distribution across ethnic groups. All diversity indices combine

[2] A similar result can be obtained by considering the square of ethnic fractionalization - if it has a positive sign it means that the probability of conflict is highest at the intermediate levels of fractionalization, *i.e.*, where the polarization is high (Collier and Hoeffler, 1998).

the effects of *between groups interaction* and *groups power*, however for some population distributions the interaction component may dominate the power component and *viceversa*. That is the relative importance of these two components and, hence, the relative importance of different distributional aspects of ethnic diversity may depend on the characteristics of the population distribution across ethnic groups. This means that for some population distributions the relevant aspect of ethnicity may be the fractionalization, for others this may be the dominance or a combination between the two.

In this paper 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) [ER henceforth] model of social antagonism. According to ER, 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: the bigger is the group, the 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 parameter values, the P index reduces to the existing indices of ethnic diversity (fractionalization - $FRAC$ and discrete ethnic polarization - RQ), while in the limit it assumes a particular form that assigns different weights to the overall effects of power and between groups 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 component of the index in that case is given by the relative Penrose - Banzhaf index of voting power in a simple majority game.

The empirical performance of the derived 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 index of voting power outperforms the other indices of ethnic diversity. This empirical evidence is robust to the inclusion of an additional set of regressors, alternative model specifications, and to several estimation techniques. The results show that the potential of conflict is given as a product of the effects of power and interaction which relative importance depends on the features of the population distribution across groups. Contrary to many other scholars, the empirical results provided in this paper show that the aspects of ethnic distribution *do* matter for the explanation of ethnic conflict but only if properly combined in a single measure of ethnic diversity that takes into account the relative importance of interaction and power.

The paper is organized as follows. In Section 2 we present the way in which the P index is constructed starting from the general specification of the ER's model of social antagonism and we axiomatically characterize the effective power function. Section 3 analyses the shape of the P index for different parameter values and different ethnic distributions while in Section 4 we explore whether the derived indices are substantially different from the existing indices of ethnic diversity using the data on ethnic distribution for a large set of countries. Section 5 presents our main empirical results and Section 6 concludes.

2 The P Index of Conflict Potential

Consider a population partitioned into n non-overlapping groups. 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. 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.

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:

$$P(\Pi) = \sum_i \sum_{j \neq i} \pi_i \pi_j \ \Phi(\pi_i, \Pi, \hat{D_{ij}}) \tag{1}$$

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 (2002, 2005), we define $\hat{D}_{ij}$ using a discrete metric:

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

The use of a discrete metric to define the distances between groups is favored by several reasons.[3] First, there are no generally accepted measures of distance between ethnic or religious groups: measuring the distance between two ethnic categories is much more difficult than the identification of ethnic groups. Second, any attempt to measure the distances across groups may generate a larger measurement error than the "belong - does not belong to" criterion. Third, as suggested by Montalvo and Reynal-Querol (2008), 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. 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.

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

$$P(\Pi) = K \sum_i \phi(\pi_i, \Pi)\ \pi_i(1 - \pi_i). \tag{2}$$

The function $\phi(\pi_i, \Pi)$ for $i = 1, 2, ..., n$ where $n \geq 2$, will be referred 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 Axiomatic Derivation of the Effective Power Function

Let N be the set of all groups, *i.e.*, $N = \{1, 2, ..., n\}$. The set of all vectors Π of relative population sizes for the n groups is in 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:[4]

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

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

Axiom 1 Normalization (N) *For all $\pi_i \in [0, 1]$, $\Pi \in \Delta^n$, and $n \geq 2$, then*

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

[3] This classification is borrowed from Montalvo and Reynal-Querol, 2008, pg. 1836.

[4] For ease of exposition 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.

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

Normalization implies that the effective power of each ethnic group is bounded in the interval $[0, 1]$. Note that we allow ϕ^n to be 1 (absolute power) or 0 (absence of power). The Normalization 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*

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

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 2A Symmetry (S) *For all* $\Pi \in \Delta^n$, *and* $n \geq 2$, *then*

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

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 exhibit the same value for distributions where all the groups are of equal relative size. Moreover, this value will be proportional to the fractionalization index divided by n.

At this point we introduce two crucial 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 challenge the other block. Moreover, Skaperdas (1998), Tan and Wang (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 that 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 ..." (Skaperdas, 1998, p.27).

Here we do not model any endogenous mechanism of coalition formation nor we are interested in which coalition is more likely to form 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 differences among groups. Since we use a discrete metric to define distances, the assumption of symmetry seems even more plausible. In general, 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).[5] 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 simplifying 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 first consider the simpler 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 $r(\cdot)$ of the groups relative size. Denoting with $\phi^2(\pi)$ and $\phi^2(1-\pi)$ the effective power of the

[5]For instance, Van der Waal, Achterberg, and Houtman (2007) study 15 different countries (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, UK, Ireland, Italy, the Netherlands, Norway, Switzerland, and the US) for the period from 1956 to 1990, and find that income is a significant predictor of ideological preferences.

groups, we define the *relative effective power* between groups as:

$$\frac{\phi^2(\pi)}{\phi^2(1-\pi)} = r(\rho) \quad \text{where} \quad \rho = \frac{\pi}{1-\pi}. \tag{3}$$

The index $\rho = \frac{\pi}{1-\pi}$ denotes the relative population size between the two groups.[6] 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 $r(\rho) := \frac{\phi^2(\rho/(1+\rho))}{\phi^2(1/(1+\rho))}$.

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

Axiom 3 Two Groups Relative Power Homogeneity (2GRPH) *Given* Π *and* Π', *let* $\pi, \pi' < 1/2 \Leftrightarrow \rho, \rho' < 1$. *Then, if* $r(\rho), r(\rho') \neq 0$*:*

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

We denote with ***2GRPH**** *a 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 imagine 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 group 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 imagine a similar situation in which the size of one group with respect to the size of the other group 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 2.1 *Let* $n = 2$, *the Effective Power of a group with population share* π *satisfies* N, M *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*

---

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

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

Proof in Appendix.

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. When groups have the same size, they equally share the power. When $\beta \in (0,1)$ the effective power of a group with the population share $\pi < 1/2$ is:

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

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 2.1, when $\pi = 1 - \pi = 1/2$ the effective power of groups is equal to $1/2$.

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

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

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

$$\phi^2_{\alpha,1} = \frac{\pi^\alpha}{\pi^\alpha + (1-\pi)^\alpha} \quad for \quad \alpha \geq 0, \quad \text{\textit{and}} \tag{4}$$

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

This functional form for the effective power is similar to the *ratio form contest success function* commonly used in the rent-seeking literature (Tullock, 1980 (with $\alpha = 1$), Skaperdas, 1996, 1998 and Nitzan, 1991). 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.[7] 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.[8] 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. With $n = 2$ and $\beta = 1$, the parameter α can be interpreted as the *elasticity* of the relative effective power with respect to the relative population size. 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 2.1 are not robust to different specifications of the axioms. In particular the following remarks hold.

Remark 1 *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 2.1 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 2.1 is continuous for $\pi = 1/2$, thereby leading to the case of $\beta = 1$.

Remark 2 *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* $\beta = 0$.

[7]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.

[8]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.

The above restrictions are obtained by requiring that the value of $r(\rho) = \beta\rho^\alpha$ as obtained in Lemma 2.1 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: $\mathbf{n > 2}$

Consider now $n > 2$. 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 or 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, $c \subseteq 2^N$ 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 2.1:

$$\phi^2(c) = \phi^2(\sum_{j \in c} \pi_j). \tag{5}$$

It follows that $\phi^n(0) = 0$ and $\phi^n(N) = 1$. In other words, the power of an empty coalition is 0 and the power 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 (Shapley, 1953):

$$m_i(c) := \phi^2(\sum_{j \in c} \pi_j) - \phi^2(\sum_{j \in c} \pi_j - \pi_i). \tag{6}$$

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

$$M_i = \sum_{c \in C_i} m_i(c). \tag{7}$$

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 for* $\phi^n(\pi_j, \Pi) > 0$ *we have*

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

This axiom states that the relative effective power between any two groups $i, j \in N$ depends on 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 pair of groups, 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* $\phi^n(\pi_j, \Pi) > 0$ *then*

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

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, and 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 2.2 *The Effective Power of group i satisfies Axioms 1 - 5 if and only if:*

$$\phi^n_{\alpha,\beta}(\pi_i, \Pi) = \frac{M_i^{\alpha,\beta}}{\sum_j M_j^{\alpha,\beta}}, \quad \forall i, j \in N; \quad i \neq j; \quad \alpha \in \Re_+ \cup \infty. \tag{8}$$

where $M_i^{\alpha,\beta}$ *is obtained making use of* $\phi^2_{\alpha,\beta}$. *Group i's effective power, hence, is defined as the relative sum of marginal contributions.*

Proof in Appendix.

Given (8), the effective power of a group can be 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 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 $\beta = 1$. 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 associated 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 $\pi_1 \in (3/10, 1/2)$, $\phi^3(\pi_3, \Pi) = 1/3$.

For $n > 2$ and $\alpha \neq \{0, 1\}$, 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.

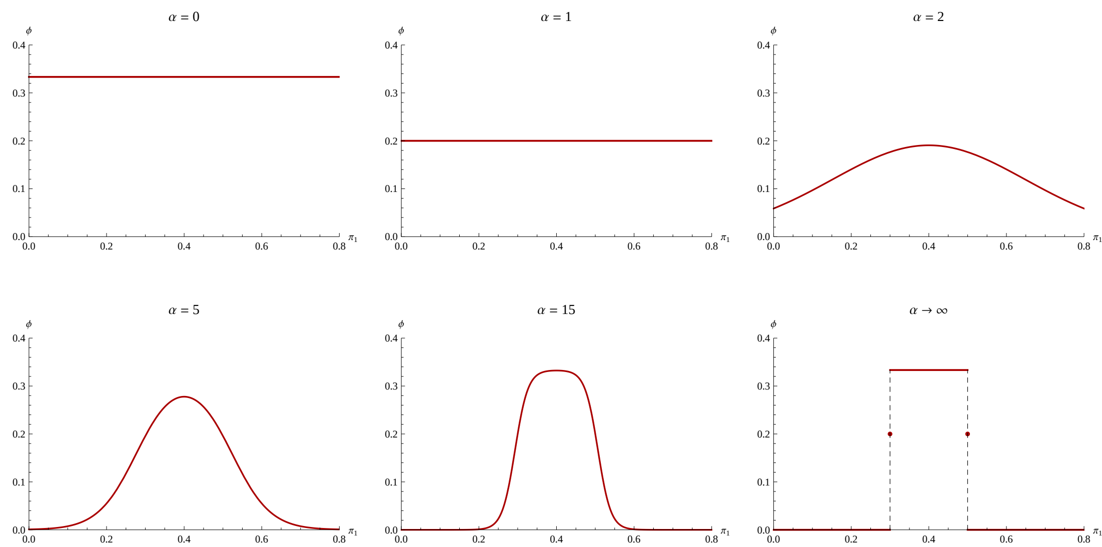


Figure 1: $\phi^3_\alpha(\pi_3, \Pi)$ as a function of α and π_1.

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

2.1.1 Effective Power as Decisiveness

The results from the previous section suggest that the effective power is *not* necessarily proportional to the groups relative size. This result shares some common features 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*. Voting power, hence, is the ability of an actor to influence the outcome of voting in a collectivity (Banzhaf, 1956). In his critique of the practice of assigning voting weights proportional to the numbers of citizens in different legislative bodies ("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 are in line with Banzhaf (1965). As $\alpha \to \infty$, the effective power of groups converges to their respective *relative Penrose-Banzhaf index of voting power* in a simple majority game:

Remark 3 *When $\alpha \to \infty$ and $\beta = 1$, the group i's Effective Power is given by its respective relative Penrose - Banzhaf Index of Voting power in a simple majority game.*

A simple majority game is a pair (N, υ), where $N = (1, 2, ..., n)$ is the set of players and $\upsilon : 2^N \to \{0, 1\}$ is the characteristic function which satisfies $\upsilon(\emptyset) = 0$, $\upsilon(N) = 1$ and $\upsilon(S) \leq \upsilon(T)$ whenever $S \subseteq T$. A coalition is winning if $\upsilon(S) = 1$, and coalitions with $\upsilon(S) = 0$ 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 it 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. However, 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. Even though 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.

Even in the absence of ethnic voting, coalitions between groups can be observed in many 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 strength have changed during the time. So, the logic of strategic behavior 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.

3 Properties of the P index of Conflict Potential

With the effective power function specified in (8), 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.

The shape of the index is determined by the interplay of two forces: *groups power* and *between groups interaction*. The relative importance of these two components for the determination of conflict potential depends on the features of the population distribution across groups, and crucially on the parameter α.

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): *Let* $\Pi := (\pi_1, \pi_2, ..., \pi_{n-1}, \pi_n)$ *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* $(\tilde{\Pi}, 0)$*, then*

$$\lim_{\Pi \to (\tilde{\Pi},0)} P_\alpha^n(\Pi) = P_\alpha^{n-1}(\tilde{\Pi}). \tag{9}$$

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 3.1 $P_\alpha^n(\Pi)$ *satisfies CGE if and only if* $\alpha > 0$.

Proof in Appendix.

In what follows we analyze 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.

3.1 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 follows we consider the P index for $\alpha = 0$, $\alpha = 1$ and $\alpha \to \infty$.

Case 1. When $\alpha = 0$, the effective power of each group is constant and equal to $1/n$. The effective power of each group is hence 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:

$$P_0^n(\Pi) = 4\frac{1}{n}\sum_i \pi_i(1-\pi_i) = 4\frac{1}{n} \cdot FRAC \tag{10}$$

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 in aggregate 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^n index and the fractionalization index provide the same ranking order. 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_0^n index becomes:

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

while $FRAC = \frac{n-1}{n}$.

Despite its very simple structure, the P_0^n index exhibits some interesting properties. In terms of the possible 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 lead 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 decreases 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 1. In a conflict context with a continuous 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 underlying the measures of ethnic polarization, *i.e.*, as n increases it puts more weight on power than on the increased interaction between groups.

Case 2. 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 proportionally 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 $\alpha = 1$, hence, 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 all the other values of α and in particular for $\alpha \to \infty$.

Case 3. 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 $P_\infty^n(\Pi)$ index can be written as:

$$P_\infty^n(\Pi) = \begin{cases} 4\pi^*(1 - \pi^*) & \text{if} \quad \pi^* > 1/2, \\ (1 - \theta_n)[4\pi^*(1 - \pi^*)] + \theta_n P_0^n(\Pi) & \text{if} \quad \pi^* = 1/2, \\ 4 \sum_i \gamma_i \pi_i (1 - \pi_i) & \text{if} \quad \pi^* < 1/2. \end{cases}$$

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 the 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 a dominant 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 covering exactly one half of the population, the index is given as a convex combination between $P_\infty^n(\Pi)$ when $\pi^* > 1/2$ and $P_0^n(\Pi)$.

3.2 P Index for Two and Three Groups

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

$$P_\alpha^2(\pi, 1-\pi) = 4\pi(1-\pi).$$

With $n = 2$, P_α^2 and $FRAC$ have identical shape, and in fact $P = (1/2) \cdot FRAC$. The only difference between them is their normalization. Both indices attain a maximum at symmetric bimodal distribution, *i.e.*, $\Pi = (\pi, 1-\pi) = (0.5, 0.5)$. In general, with only two groups all the indices provide the same ranking order. Regarding the interaction and the power component, in the case of two groups only the interaction matters in the determination of conflict potential.

Since for $n = 2$ all the indices are the same, the simplest way to analyze the implications of different choices of 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.

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:

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

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 investments and economic 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 unstable than a country in which the population is equally distributed across groups. The empirical evidence, however, suggests the opposite: with one ethnic group being dominant the risk of instability is almost doubled (Collier and Hoeffler, 2004). Interestingly, Montalvo and Reynal-Querol (2005) agree with Collier and Hoeffler (2004) in the sense that among all possible ethnic configurations, 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 two 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 scarcely predominant in terms of its relative population size while in the second no group has absolute majority, *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 $RQ(\Pi_2) = 0.881$ even though the first distribution is categorized as the worst one (small majority versus no majority).

In general, the RQ index is not sensitive enough to take into account the conflict potential that can potentially derive from certain ethnic constellations. This is because the index considers power and interaction as two separated phenomena. Even before Collier and Hoeffler's (2004) empirical evidence on the importance of ethnic dominance, Horrowitz (1985) has pointed out 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 preferred 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 accurate. 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 2 shows P_0^3 (green curve), P_1^3 (red line) and P_∞^3 *(blue curve)* expressed in terms of π_1. 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 1 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_∞^3 index reaches almost one when the relative size of one group becomes scarcely higher than $1/2$ because in that particular case a group in question has the absolute power and the "opposition" is powerless.

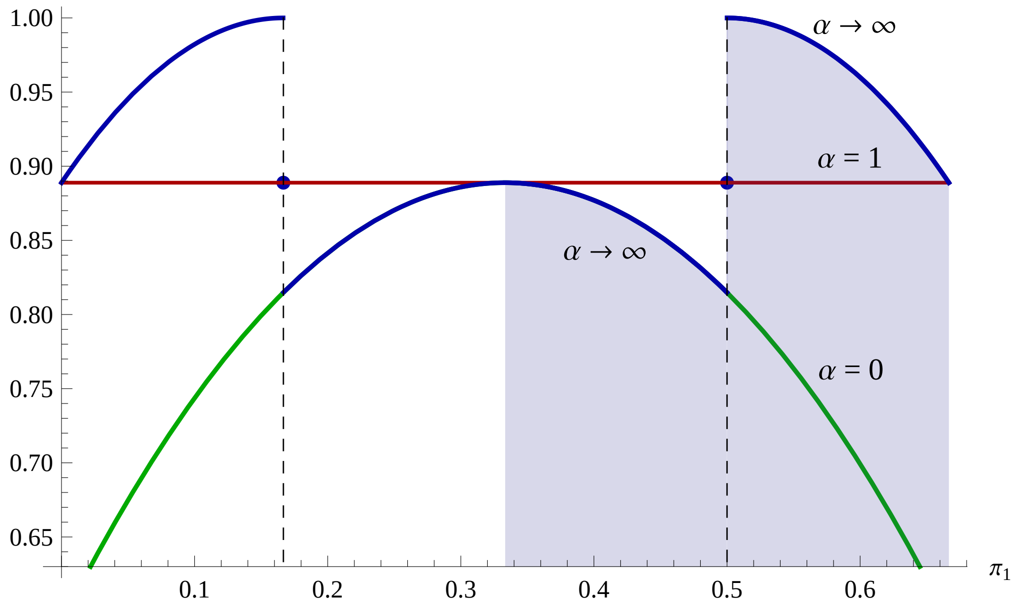


Figure 2: P Index for $\alpha = 0$, $\alpha = 1$ and $\alpha \to \infty$ $(n = 3)$.

Collier and Hoeffler (2004) reason in terms of minority exploitation, and claim that when the size of the predominant group is scarcely higher than $1/2$, the potential to exploit the minority is highest and, hence its "frustration" is maximal. Since the minority in this case does not have access to conventional channels for 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 increases, the potential of conflict decreases. It is worth noting here how the P_∞^3 combines dominance (and, hence power) and interaction. As long as we avoid the dominance, the conflict potential is entirely determined by the interaction component - the shape of P_∞^3 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_∞^3 index behaves the same around and after the threshold value of $1/2$, where the presence of dominance shifts the fractionalization (interaction) curve up without altering its shape.

4 How much does α matter? A first insight into data

In the previous section we have shown how the choice of the parameter α determines the shape of the P index. The difference between the P_1^n index (which is the RQ index of discrete ethnic polarization) and the extreme element of the class of indices (P_∞^n) lies in the capacity of the latter to combine different aspects of ethnic diversity into one single measure and to be sensitive to the presence of an extreme form of ethnic dominance. In this section we analyze graphically the relationship between the P index for $\alpha = 0$ and $\alpha \to \infty$ and the RQ index using the data on ethnic distribution for 138 countries from the "Ethnic Power Relations" data set (Cederman, Min and Wimmer, 2009). In the next section we test the empirical performance of the P index for different values of the coefficient α against the fractionalization index and the ethnic dominance dummy variables within the context of the commonly used logistic model that focuses on the onset of ethnic conflicts.

Figure 3 shows the relationship between P_{∞}^{n} and P_{0}^{n} versus RQ discrete polarization index. The correlation between P_{0}^{n} and RQ is positive and relatively high (the coefficient of correlation between the two is 0.82). 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 very weak (0.1).[9]

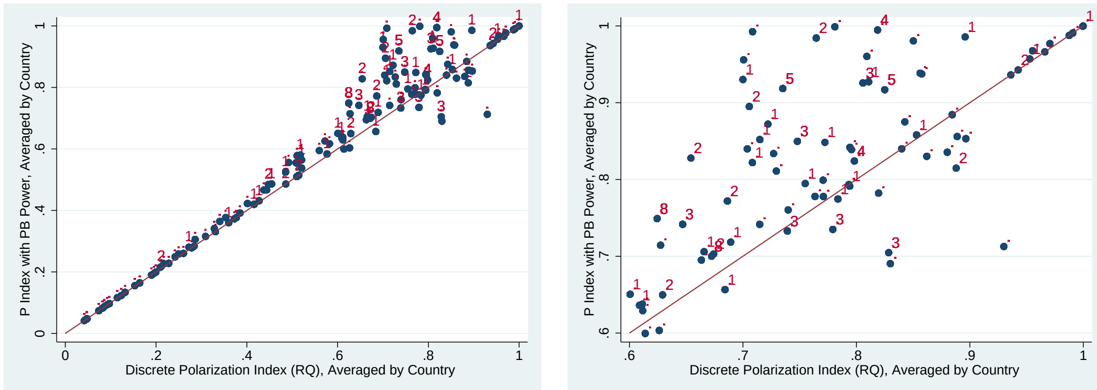
Source: *Ethnic Power Relations (EPR) Data set, Cederman, Min and Wimmer (2009).*

Figure 3: P index with $\alpha = 0$ and $\alpha \to \infty$ versus RQ discrete polarization index.

Since the RQ and the P_{∞}^{n} index differ significantly only for high values of RQ, the question is whether this is sufficient to consider P_{∞}^{n} as a different measure from the RQ index of polarization. This is not only a theoretical, but it is particularly an empirical issue since we want to analyze the relationship between conflict and different distributional aspects of ethnicity. Moreover, while P_{∞}^{n} and RQ appear to be highly correlated against each other (for the entire interval [0,1]), empirical exercises 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 conflict where fractionalization is not.

[9] In general, the difference between P_{∞}^{n} and RQ ranges from -0.217 to 0.286.

Figure 4 shows P^n_∞ versus RQ with the labels for the quantity of ethnic conflict [EW] onsets between 1946 and 2005 (red numbers).



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

Figure 4: P^n_∞ versus RQ with Ethnic War [EW] Label.

Almost all the countries with P^n_∞ larger than RQ have experienced one or more ethnic conflicts during the period between 1946 and 2005. For instance, the P^n_∞ index for Chad and Iran is almost 1 (0.9981 and 0.9946, respectively) while the RQ index equals 0.81, as approximately 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^n_∞ index lower than the RQ index do not seem particularly conflict prone with 3 exceptions, namely Bosnia and Herzegovina, Sudan and Afghanistan. The sensitivity of the two indices hence differs for the most conflict prone sub-sample of countries. This, however, does not necessarily imply that their predictive power (in terms of the probability of ethnic conflict) differs significantly. In the next section we test the empirical performance of the derived indices and we show that the predictive power of the P^n_∞ index is significantly higher than the predictive power of the other indices of ethnic diversity in the explanation of ethnic conflict outbreak.

5 Empirical Performance of the P Index of Conflict Potential

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 conflicts in a time range from 1946 to 2005. Ethnic conflict onset is a binary variable that takes the value of 1 in the first year of an armed conflict and 0 otherwise. In all model specifications we correct for error correlation over time for a given country by calculating cluster - robust standard errors. We also compare the results with a random effects panel estimator and test for the independence of observations. Since ethnic war is a rare event and since the standard logistic regression can underestimate the probability of such events, we also perform a rare event logit estimation (King and Zeng, 2001). Following Cederman, Min and Wimmer's (2009), we control for possible time trends by including the *number of peace years* since the outbreak of the previous conflict, 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 (significant) regional time trends. Regional dummies[10] are included in all model specifications, even though we do not show them for the sake of space.

5.1 Data Sources and Econometric Issues

Our empirical analysis relies on the "Ethnic Power Relations" [EPR henceforth] data set recently provided by Cederman, Min and Wimmer (2009) [CMW henceforth]. The *EPR* data set identifies all *politically relevant* ethnic groups in a time range from 1946 to 2005. The data set 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 addition to the ethnic coding, the authors also define the degree of access to central level state power for representatives of each group for each time period.

The authors define ethnicity as "... a subjectively experienced sense of commonality based on a belief in common ancestry and shared culture ..." (CMW, p.325). This definition of ethnicity includes ethno-linguistic, racial and ethno-religious groups but *not* tribes and clans "... that conceive of ancestry in genealogical terms ..." (CMW, p.325). 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 systematically and intentionally discriminated against in the domain of public politics ..." (CMW, p.325). By "significant" political actor they mean a political organization (not necessarily a party), that is *active* in the national political arena. Given this definition, an ethnic group is included in the dataset if it politically relevant *at least once* in the sample period. One group is discriminated 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. Since politically relevant ethnic categories may change over time, the authors divided the time period and provided separate codings for each sub-period.

Regarding the conflict data, CMW extend the *Armed Conflict Data Set*[11] by coding

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

[11] The ACD data set includes intermediate and high intensity conflicts. The definition of a conflict depends on the "battle death threshold", *i.e.* the number of killed people in a year. The ACD data set considers all conflict with at least 25 battle deaths a year (where high intensity conflicts are those with more than 1000 battle deaths a year.)

each conflict for whether rebel organizations pursued ethno-nationalist aims and recruited along ethnic lines. The authors identify as *ethnic* "... the aims of achieving ethno-national self-determination, a more favorable ethnic balance of power in government, ethno-regional autonomy, the end of ethnic and racial discrimination, language and other cultural rights ..." (CMW, p.326). All other wars are defined as *non-ethnic*.[12] 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.

We restrict the sample to 138 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 a particular structure of the effective power function and the related computational complexities, we were forced to consider countries with no more than 6 ethnic groups[13]. 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 8% of the population, were excluded from the analysis. Moreover, the relative population size of each potentially excludable ethnic category could not have been substantial (not more than 5%). In such a way the marginal impact of the excluded groups on the value of our indices is minimized. Moreover, the number of countries for which we were forced to "eliminate" some ethnic groups is low.

The list of explanatory variables that we consider in our regression models is the one commonly used in conflict research[14] (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, Population Size, Oil Production per Capita, Mountainous Terrain, Noncontiguous Territory and New State, Democracy and Anocracy, Instability, Share of the Excluded Population, Number of Power Sharing Partners, Past Imperial History and Ethnic Diversity Indices* (fractionalization, polarization, dominance dummies and the P index of conflict potential for different values of the coefficient α). In order to calculate the indices of ethnic diversity we make use of the groups' relative shares calculated in relation to total population.

[12] We consider only *ethnic* conflicts for several reasons. First, there is a substantial difference in the nature and the determinants of ethnic and non-ethnic conflicts (Sambanis, 2001, 2004). Second, ethnic conflicts are closely related to political and cultural identity. Moreover, in ethnically heterogeneous societies political mobilization occurs mostly along ethnic lines. Third, the EPR data set offers a precise coding of all ethnic conflicts and identifies all politically relevant ethnic groups as well as their political status.

[13] The average and the median number of groups in Cederman, Min and Wimmer (2009) is 4.5 and 3 respectively.

[14] The list of explanatory variables in Appendix.

5.2 Explaining Ethnic War Onset

Table 1 presents the basic specification of our model of ethnic conflict onset.[15] Five variables are always significant with the correct expected 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 neither 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), and Cederman, Min and Wimmer (2009), among others.

Table 1: *Logit Regression: Ethnic War* **Onset** *- EPR (2009) Data Set: 138 Countries*

EW Onset	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
GDP per Capita	-0.128***	-0.129***	-0.132***	-0.121***	-0.125***	-0.128***	-0.133***
	0.043	0.043	0.044	0.043	0.044	0.044	0.045
Population Size	0.389***	0.388***	0.383***	0.361***	0.375***	0.383***	0.388***
	0.085	0.086	0.087	0.082	0.085	0.087	0.091
Democracy (d)	0.227	0.245	0.272	0.182	0.213	0.240	0.278
	0.322	0.321	0.321	0.327	0.326	0.327	0.327
Anocracy (d)	0.576**	0.579**	0.573**	0.551**	0.561**	0.573**	0.579**
	0.234	0.234	0.239	0.227	0.228	0.229	0.235
Oil per Capita	0.013	0.013	0.012	0.011	0.012	0.013	0.012
	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Mountains	0.173**	0.172	0.173	0.189**	0.176**	0.173**	0.172
	0.088	0.088	0.089	0.088	0.089	0.088	0.088
Instability	0.135	0.135	0.152	0.160	0.145	0.138	0.149
	0.256	0.256	0.254	0.258	0.260	0.259	0.255
NC State	0.325	0.306	0.272	0.334	0.321	0.305	0.272
	0.504	0.506	0.525	0.511	0.510	0.507	0.523
New State	2.175***	2.173***	2.204***	2.217***	2.187***	2.177***	2.203***
	0.713	0.706	0.699	0.719	0.713	0.705	0.698
RQ	**1.321****				1.059		
	0.641				0.885		
P($\alpha = 2$)		**1.560****				1.477	
		0.652				0.881	
P($\alpha \to \infty$)			**1.802*****				**1.875****
			0.638				0.797
FRAC				1.087	0.418	0.144	-0.153
				0.570	0.837	0.865	0.858
Constant	-10.718***	-10.865***	-10.993***	-10.194***	-10.589***	-10.820***	-11.032***
	1.428	1.423	1.392	1.413	1.419	1.428	1.405
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6200	6200	6200	6200	6200	6200	6200
N. Countries	138	138	138	138	138	138	138
Pseudo R^2	0.136	0.138	**0.142**	0.134	0.136	0.138	**0.142**
Wald Chi2	171.089***	165.457***	164.556***	202.261***	186.615***	171.184***	161.839***
Bic	1039.846	1037.533	1033.336	1041.405	1048.293	1046.232	1042.028
Aic	898.468	896.155	891.957	900.026	900.183	898.121	893.917

Notes: *The sample includes 138 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. ** $p < 0.05$, *** $p < 0.01$.*

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

[15]The models were implemented in Stata 12.1. *Mathematica* was used for the computation of P for different α and $FRAC$.

the RQ index and the P index with $\alpha = 2$ and $\alpha \to \infty$ are significantly different from zero. However, the P index with Penrose-Banzhaf relative scores (P_∞) 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 Column 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. When we include P_∞ together with the RQ index of ethnic polarization, the coefficient on P_∞ is positive and highly significant while the coefficient of RQ is not statistically different from zero with a negative sign.

Given that the RQ coincides with the P index with $\alpha = 1$, the goodness of fit measured by the Pseudo R^2 is increasing in α. Indeed, by comparing the outcome of various estimations based on P_∞, it results that the highest value of the Pseudo R^2 is obtained for $\alpha \to \infty$. 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.[16]

5.2.1 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 2 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 1). 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

[16]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.

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.

Table 2: *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 per Capita	-0.151***	-0.147***	-0.144***	-0.148***	-0.143***	-0.144***	-0.144***	-0.142***
	0.049	0.048	0.048	0.047	0.051	0.051	0.050	0.050
Population Size	0.455***	0.432***	0.413***	0.398***	0.431***	0.411***	0.400***	0.355***
	0.097	0.097	0.097	0.093	0.091	0.094	0.096	0.088
Democracy (d)	0.271	0.271	0.297	0.268	0.239	0.261	0.302	0.226
	0.336	0.332	0.330	0.327	0.332	0.327	0.327	0.321
Anocracy (d)	0.640***	0.612**	0.601**	0.573**	0.648***	0.620***	0.617**	0.584**
	0.244	0.242	0.242	0.240	0.242	0.239	0.242	0.234
Oil per Capita	0.016**	0.015	0.013	0.012	0.017**	0.017**	0.015	0.015**
	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Mountains	0.150	0.146	0.155	0.140	0.186**	0.172**	0.177**	0.186**
	0.080	0.082	0.085	0.083	0.080	0.081	0.083	0.083
Instability (d)	0.080	0.099	0.125	0.118	0.093	0.106	0.126	0.131
	0.249	0.250	0.250	0.249	0.255	0.256	0.254	0.255
NC State	0.368	0.335	0.281	0.306	0.362	0.313	0.257	0.275
	0.503	0.505	0.520	0.531	0.486	0.496	0.517	0.518
New State	2.154***	2.147***	2.181***	2.169***	2.178***	2.159***	2.194***	2.195***
	0.707	0.700	0.692	0.689	0.722	0.706	0.694	0.703
CH Dominance (d)	**0.657****	0.537	0.375	**0.865*****				
	0.277	0.278	0.291	0.310				
SW Dominance (d)					0.439	0.475	0.397	**0.858****
					0.255	0.265	0.260	0.362
RQ		0.899				**1.422****		
		0.627				0.678		
P($\alpha \to \infty$)			**1.447****				**1.767*****	
			0.624				0.651	
FRAC				**1.757****				**2.163****
				0.694				0.851
Constant	-10.953***	-11.096***	-11.219***	-11.040***	-10.711***	-11.170***	-11.348***	-10.839***
	1.466	1.455	1.422	1.381	1.479	1.457	1.430	1.406
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6200	6200	6200	6200	6200	6200	6200	6200
N. Countries	138	138	138	138	138	138	138	138
Pseudo R^2	0.138	0.140	0.144	0.145	0.134	0.140	0.145	0.144
Wald Chi2	128.102***	133.944***	142.793***	141.082***	139.191***	148.709***	151.764***	164.707***
Bic	1037.608	1044.294	1040.176	1039.028	1041.318	1044.549	1039.264	1040.624
Aic	896.229	896.184	892.065	890.917	899.940	896.438	891.153	892.513

Notes: *The sample includes 138 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.* ** $p < 0.05$, *** $p < 0.01$.

5.2.2 *Ethnic Conflict Onset, Regional Time trends and Ethnic Politics*

The frequency of ethnic conflict episodes between 1946 and 2005 is more or less stable. 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 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 take the following form:

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

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 CMW 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 shows the results of our estimation. In line with CMW, the degree of ethnic exclusion[17] 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 challenge the government ..." (CMW, 2009, p.322), a high degree of ethnic exclusion increases the likelihood of ethnic conflict. Center segmentation is also significant with the correct expected 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 ..." (CMW, 2009, p.322). 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. In contrast to the Fearon and Laitin's (2003) insurgency model, instability and mountainous terrain receive limited support here. Oil production per capita, on the other hand, receives a full support. 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 1) 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 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 (Balkans and the former Soviet Union) that 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, that is positive and significant at the 0.05 level. CMW 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

[17]Excluded Population is calculated in relation to ethno-politically relevant population in each sub-period. The results do not change significantly when we consider the share of the excluded population calculated in relation to total population.

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.

Table 3: *Logit Model: Ethnic War* **Onset** *- EPR (2009) Data Set:* **Additional Regressors***: Share of the population excluded from central government, Center Segmentation and Imperial Past and Regional Time Trends.*

EW Onset	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
GDP per Capita	-0.140***	-0.139***	-0.140***	-0.140***	-0.138***	-0.137***	-0.137***
	0.045	0.045	0.045	0.045	0.045	0.045	0.046
Population Size	0.355***	0.359***	0.373***	0.350***	0.357***	0.364***	0.388***
	0.085	0.086	0.089	0.085	0.086	0.088	0.094
Excluded Population	0.336***	0.319***	0.302***	0.356***	0.350***	0.349***	0.358***
	0.122	0.118	0.112	0.125	0.124	0.122	0.120
Imperial Past	0.517	0.509	0.552	0.529	0.504	0.477	0.533
	0.571	0.563	0.552	0.580	0.568	0.568	0.573
Center Segmentation	0.117***	0.115***	0.111***	0.112**	0.137***	0.158***	0.180***
	0.045	0.044	0.043	0.053	0.050	0.050	0.050
Democracy (d)	-0.037	-0.035	-0.023	-0.022	-0.050	-0.060	-0.055
	0.441	0.438	0.431	0.440	0.447	0.446	0.442
Anocracy (d)	0.349	0.347	0.334	0.351	0.362	0.376	0.378
	0.233	0.234	0.240	0.229	0.230	0.232	0.242
Oil per Capita	0.017**	0.016**	0.016**	0.017**	0.018**	0.019**	0.019**
	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Mountains	0.151	0.152	0.158	0.154	0.150	0.151	0.159**
	0.088	0.089	0.090	0.088	0.085	0.083	0.080
Instability	0.221	0.222	0.243	0.220	0.225	0.233	0.271
	0.269	0.269	0.266	0.269	0.270	0.268	0.266
NC State	0.098	0.077	0.014	0.124	0.085	0.048	-0.045
	0.470	0.484	0.515	0.454	0.473	0.487	0.512
New State	2.344***	2.336***	2.336***	2.366***	2.338***	2.327***	2.334***
	0.716	0.714	0.711	0.713	0.715	0.709	0.704
EEurope*Y	0.099***	0.100***	0.103***	0.098***	0.098***	0.098***	0.101***
	0.019	0.019	0.020	0.019	0.019	0.019	0.020
RQ	0.751				1.053		
	0.709				0.899		
P$(\alpha = 2)$		1.071				1.695	
		0.706				0.910	
P$(\alpha \to \infty)$			**1.530****				**2.349*****
			0.689				0.857
Frac				0.334	-0.510	-1.089	-1.740
				0.679	0.864	0.917	0.998
Constant	-10.381***	-10.591***	-11.037***	-10.109***	-10.451***	-10.774***	-11.339***
	1.381	1.404	1.455	1.380	1.383	1.421	1.502
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	6190	6190	6190	6190	6190	6190	6190
N. Countries	138	138	138	138	138	138	138
Pseudo R^2	0.160	0.161	**0.165**	0.159	0.160	0.162	**0.168**
Wald Chi2	489.386***	469.836***	438.814***	446.725***	637.714***	623.794***	581.756***
Bic	1085.270	1083.817	1079.889	1086.363	1093.768	1091.531	1085.831
Aic	890.080	888.627	884.699	891.173	891.847	889.610	883.910

Notes: *The sample includes 138 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.* ** $p < 0.05$, *** $p < 0.01$.

5.2.3 Rare Event Logistic Estimation

One of the problems related to conflict data is the relative rareness of events (Gates 2002; 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 zeros (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. 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 event, in this case the probability of the rare events. Table 4 shows the results of the "Rare Events Logit Estimation".

Table 4: *Logit Model: Ethnic War* **Onset** - *EPR (2009) Data Set:* **Robustness Check: Rare Events Logit.**

EW Onset	Model1	Model2	Model3	Model4	Model5	Model6	Model7
GDP per Capita	-0.144***	-0.145***	-0.146***	-0.144***	-0.143***	-0.141***	-0.141***
	0.045	0.045	0.045	0.045	0.045	0.045	0.046
Population Size	0.347***	0.351***	0.365***	0.341***	0.348***	0.355***	0.376***
	0.085	0.086	0.088	0.084	0.086	0.088	0.094
Excluded Population	0.320***	0.303***	0.286**	0.342***	0.334***	0.333***	0.341***
	0.122	0.118	0.111	0.124	0.123	0.122	0.120
Imperial Past	0.512	0.505	0.546	0.519	0.494	0.467	0.518
	0.568	0.561	0.549	0.577	0.566	0.565	0.570
Center Segmentation	0.110**	0.108**	0.104**	0.107**	0.130***	0.150***	0.171***
	0.045	0.044	0.043	0.052	0.050	0.049	0.050
Democracy (d)	0.025	0.027	0.037	0.040	0.012	0.001	0.004
	0.439	0.436	0.429	0.438	0.445	0.444	0.440
Anocracy (d)	0.366	0.364	0.351	0.369	0.379	0.392	0.393
	0.232	0.233	0.239	0.228	0.228	0.231	0.241
Oil per Capita	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	0.155	0.156	0.162	0.158	0.155	0.155	0.163**
	0.087	0.089	0.090	0.087	0.084	0.083	0.080
Instability	0.231	0.232	0.253	0.230	0.233	0.241	0.280
	0.268	0.267	0.265	0.268	0.268	0.267	0.264
NC State	0.076	0.057	-0.002	0.101	0.065	0.031	-0.054
	0.468	0.482	0.512	0.452	0.470	0.485	0.510
New State	2.299***	2.290***	2.290***	2.321***	2.291***	2.280***	2.286***
	0.713	0.710	0.708	0.709	0.712	0.706	0.700
EEurope*Y	0.084***	0.085***	0.089***	0.084***	0.083***	0.083***	0.087***
	0.019	0.019	0.020	0.019	0.019	0.019	0.020
RQ	0.716				1.022		
	0.705				0.894		
P($\alpha = 2$)		1.020				1.632	
		0.702				0.906	
P($\alpha \to \infty$)			**1.468****				**2.264*****
			0.686				0.853
Frac				0.304	-0.520	-1.074	-1.702
				0.676	0.860	0.913	0.993
Constant	-9.750***	-9.941***	-10.363***	-9.504***	-9.802***	-10.101***	-10.642***
	1.375	1.398	1.448	1.373	1.376	1.414	1.495

Notes: *The sample includes 138 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. ** $p < 0.05$, ***$p < 0.01$.*

The sign and the level of significance of our covariates are similar to those from Table 3. 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 can conclude that accounting for rareness makes 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. 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 that is not significantly different from zero for the random effects logistic estimation.

6 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 correlation between ethnic and/or religious 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. 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 between-groups interaction. Although we were not interested in modeling 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 scarcely higher than one half of the population. When the interaction component 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 competition between all the groups in the population. We show that a group can be powerless even when its size is not negligible, which is in line with the literature on voting power in simple majority games.

Our index is radically different from the existing indices of ethnic diversity even though it preserves a uni-dimensional nature. We also show that when we apply our indices to the empirical analyses of the correlates of ethnic conflict onset, this difference is not only theoretical but also empirical. 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.

References

[1] Alberto Alesina, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. Fractionalization. *Journal of Economic Growth*, 8(2):155–94, June 2003.

[2] John F. Banzhaf. Weighting voting doesn't work: A mathematical analysis. *Rutgers Law Rev*, 19:317–343, 1965.

[3] Lars-Erik Cederman and Luc Girardin. Beyond fractionalization: Mapping ethnicity onto nationalist insurgencies. *American Political Science Review*, 101:173–185, 2007.

[4] Lars-Erik Cederman, Andreas Wimmer, and Brian Min. Why do ethnic groups rebel? new data and analysis. *World Politics*, 62:87–119, 2009.

[5] Paul Collier. Implications of ethnic diversity. *Economic Policy*, 16(32):127–166, 04 2001.

[6] Paul Collier and Anke Hoeffler. On economic causes of civil war. *Oxford Economic Papers*, 50(4):563–573, 1998.

[7] Paul Collier and Anke Hoeffler. Greed and grievance in civil wars. *Oxford Economic Papers*, pages 663–595, 2004.

[8] Moshe' Machover Dan S. Felsenthal. *The measurement of voting power: theory and practice, problems and paradoxes*. Edward Elgar, 1998.

[9] K. de Rouen and D. Sobek. The dynamics of civil war duration and outcome. *Journal of Peace Research*, 41(3):303–320, 2004.

[10] Michael W. Doyle and Nicholas Sambanis. International peacebuilding: A theoretical and quantitative analysis, 2000.

[11] Wolfgang Eichhorn. *Functional equations in economics*. Number 11 in Applied mathematics and computation. Addison-Wesley, Reading, Mass. [u.a.], 1978.

[12] Joan Esteban, Laura Mayoral, and Debraj Ray. Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4):1310–1342, June 2012.

[13] Joan Esteban and Debraj Ray. On the measurement of polarization. *Econometrica*, 62(4):819–851, July 1994.

[14] Joan Esteban and Debraj Ray. Polarization, fractionalization and conflict. *Journal of Peace Research*, 45(2):163–182, March 2008.

[15] Joan Esteban and Jozsef Sakovics. Olson vs. coase: Coalitional worth in conflict. Working Papers 3, Barcelona Graduate School of Economics, October 2003.

[16] Joan Esteban and G. Schneider. Polarization and conflict: Theoretical and empirical issues. *Journal of Peace Research*, 45(2):131–141, 2008.

[17] D. Fearon, James. Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8:195–222, 2003.

[18] D. Fearon, James, Kimuli Kasara, and David D. Laitin. Ethnic minority rule and civil war onset. *American Political Science Review*, 101:187–193, 2007.

[19] D. Fearon, James and David D. Laitin. Ethnicity, insurgency, and civil war. *American Political Science Review*, 97:75–90, 2003.

[20] Scott Gates. Empirically assessing the causes of civil war. March 2002.

[21] Scott Gates and H. Strand. Modeling the duration of civil wars: Measurement and estimation issues. In *presentation at the Joint Session of Workshops of the ECPR, Uppsala*, 2004.

[22] H. Hegre and N. Sambanis. Sensitivity analysis of empirical results on civil war onset. *Journal of Conflict Resolution*, 50(4):508, 2006.

[23] J. Hirchleifer. Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. *Public Choice (1986-1998)*, 63 (2):101–112, 1989.

[24] D. L. Horrowitz. *Ethnic Groups in Conflict*. University of California Press, 2000.

[25] G. King and L. Zeng. Logistic regression in rare events data. 2001.

[26] J. G. Montalvo and M. Reynal-Querol. Why ethnic fractionalization? polarization, ethnic conflict and growth. Economics working papers, Department of Economics and Business, Universitat Pompeu Fabra, 2002.

[27] J. G. Montalvo and M. Reynal-Querol. Ethnic polarization, potential conflict and civil wars. Economics working papers, Department of Economics and Business, Universitat Pompeu Fabra, 2005.

[28] J. G. Montalvo and M. Reynal-Querol. Ethnic polarization and the duration of civil wars. *Economics of Governance*, 11(2):123–143, 2010.

[29] JoseG. Montalvo and Marta Reynal-Querol. Discrete polarisation with an application to the determinants of genocides. *Economic Journal*, 118(533):1835–1865, November 2008.

[30] S. Nitzan. Collective rent dissipation. *Economic Journal*, 101(409):1522–1534, 1991.

[31] L. S. Penrose. The elementary statistics of majority voting. *Journal of the Royal Statistical Society*, 109:53–57, 1946.

[32] N. Sambanis. Do ethnic and nonethnic civil wars have the same causes? a theoretical and empirical inquiry (part 1). *Journal of Conflict Resolution*, 2001.

[33] N. Sambanis. What is civil war?: Conceptual and empirical complexities of an operational definition. *Journal of Conflict Resolution*, 48(6):814–858, 2004.

[34] G. Schenider and N. Wiesehomeier. Rules that matter: Political institutions and the diversity - conflict nexus. *Journal of Peace Research*, 45(2):183–203, 2008.

[35] G. Schenider and N. Wiesehomeier. Diversity, conflict and growth: Theory and evidence. *Diversity*, 2(9):1097–1117, 2010.

[36] L. S. Shapley. A value for n-person games. Contribution to the Theory of Games. *Annals of Mathematics Studies*, 2:28, 1953.

[37] L. S. Shapley and M. Shubik. A method for evaluating the distribution of power in a committee system. *American Political Science Review*, 48:787–792, 1954.

[38] S. Skaperdas. Contest success functions. *Economic Theory*, 7(2):283–290, 1996.

[39] S. Skaperdas. On the formation of alliances in conflict and contests. *Public Choice*, 96(1-2):25–42, July 1998.

[40] G. Tan and R. Wang. Coalition formation in the presence of continuing conflict. *International Journal of Game Theory*, 39(1):273–299, 2010.

[41] G. Tullock. Efficient rent seeking. *In Toward a Theory of the Rent Seeking Society*, pages 97–112, 1980.

[42] J. Van der Waal, P. Achterberg, and D. Houtman. Class is not dead it has been buried alive: Class voting and cultural voting in postwar western societies (1956-1990). *Politics Society*, 35(3):403–426, September 2007.

[43] T. Vanhanen. Domestic ethnic conflict and ethnic nepotism: A comparative analysis. *Journal of Peace Research*, 36(1):55, 1999.

[44] A. Wimmer, L-E. Cederman, and B. Min. Ethnic politics and armed conflict: A configurational analysis of a new global data set. *American Sociological Review*, 74(2):316, 2009.

Appendix I

Proof of Lemma 2.1

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

Necessity part. Consider axiom 2GRPH, requering 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 exists 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 then $\phi^2(1-\pi) = 1 - \phi^2(\pi)$. 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 analysed 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 analysing 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) holds. 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:

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

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

Proof of Theorem 2.2

<u>*Sufficiency part.*</u> Note that the obtained specification for $\phi^n_{\alpha,\beta}$ satisfies the axioms considered.

<u>*Necessity part.*</u> 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 with the population 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 $h : (0, 1] \to \Re_+$.

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 $M \in (0, 2]$ 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 it should be that $M_j = 0$. If this is not the case then there exists group 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 $\phi^n_i = 0$ with $\frac{M_i}{M_j} \neq 0$ and $\phi^n_j > 0$, 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 except the largest one in all possible distributions have $\phi^n = 0$. This, however, is not consistent with the Normalization axiom. It follows that $\phi^n_i = 0$ only if $M_i = 0$, making the result consistent with (A.4). ■

Proof of Proposition 3.1

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_i^α, if $\phi^2_\alpha(\pi)$ is continuous, also M_i^α 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 $\lim_{\pi_n \to 0} M_i^\alpha = 0$. In general, considering the remaining $n-1$ groups, by denoting with $\tilde{M}_i^\alpha$ the sum of marginal contributions for group i under $\tilde{\Pi}$ and M_i^α the one under Π, we have that $\lim_{\Pi \to (\tilde{\Pi},0)} M_i^\alpha = \tilde{M}_i^\alpha$ for all $i = 1,2,3,...,n-1$. Given the continuous functional form for $P^n_\alpha(\Pi)$ w.r.t. M_i^α and the distribution 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 $M_n^0 = 1$. 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 $\phi^2_\infty(\pi)$ 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 groups (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 the limit of $M_3^\infty / \sum_j M_j^\infty$ 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. ■

Appendix II

EXPLANATORY AND CONTROL VARIABLES

GDP per Capita

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.

Population Size

In order to account for the size of the country, we include the natural logarithm of the first lag of population.

Oil Production per Capita

The data for oil production per capita (in barells) come from Wimmer and Min (2006) and Cederman, Min and Wimmer's (2009) data sets.

Mountainous Terrain, Noncontiguous Territory and New State

The data on mountains terrain are taken from the A.J.Gerrard's (1990) project on mountains environment. Countries with the territory holding at least 10 000 people and separated from the land area containing the capital city either by land or by 100 kilometers of water are coded as "Noncontiguous". A dummy variable for "New State" is coded as 1 for the first two years of independence.

Democracy and Autocracy

In order to characterize the political system we use the Polity IV data set (*PIV*). The *PIV* is based on a 21-point scale: "autocracies" (-10 to -6), "anocracies" (regimes that are nor autocratic nor democratic) (-5 to +5), and "democracies" (+6 to +10).

Instability

By instability we intend the "previous regime change". The regime change is defined as any change in the Polity Score of at least 3 points over the prior three years. The data are taken from the *EPR* data set and are based on *PIV*.

Share of the Excluded Population

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.

Number of Power Sharing Partners

We include the number of power sharing groups represented by ethnic elites at the central government. This variable is termed as the *degree of center segmentation*.

Past Imperial History

This variable is given by the percentage of years spent under imperial rule between 1816 and independence. The data comes from Min and Wimmer (2006).

Ethnic Diversity Indices

Regarding the P index of conflict potential, we consider three different values for the coefficient α, namely $\alpha = 1$ (actually the RQ discrete polarization index), $\alpha = 2$ and $\alpha \to \infty$. We also construct the Collier and Hoeffler's (2004) and Schneider and Wiesehomeier's (2008) ethnic dominance dummy variables as well as the fractionalization index.