Ca' Foscari University of Venice

Department of Economics

Working Paper

Matija Kovacic
and
Claudio Zoli

Ethnic Distribution, Effective Power and Conflict

ISSN: 1827-3580
No. 22/WP/2018

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 22/WP/2018
ISSN 1827-3580

Ethnic Distribution, Effective Power and Conflict

Matija Kovacic
Ca' Foscari University of Venice

and

Claudio Zoli
University of Verona

Abstract
This paper highlights the fact that different distributional aspects of ethnicity matter for conflict. We axiomatically derive a parametric class of indices of conflict potential obtained as the sum of each group relative power weighted by the probability of across group interactions. The power component of an extreme element of this class of indices is given by the Penrose-Banzhaf measure of relative power. This index combines in a non-linear way fractionalization, polarization and dominance. The empirical analysis verifies that it outperforms the existing indices of ethnic diversity in explaining ethnic conflict onset.

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

JEL Codes
D63, D74, O57

Address for correspondence:
Matija Kovacic
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: matija.kovacic@unive.it

The Working Paper Series
is available only on line
(http://www.unive.it/pag/16882/)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

Ethnic Distribution, Effective Power and Conflict

Matija Kovacic and Claudio Zoli[*]

First Version: April, 2013, **This Version**: September, 2018

Abstract

This paper highlights the fact that different distributional aspects of ethnicity matter for conflict. We axiomatically derive a parametric class of indices of conflict potential obtained as the sum of each group relative power weighted by the probability of across group interactions. The power component of an extreme element of this class of indices is given by the Penrose-Banzhaf measure of relative power. This index combines in a non-linear way fractionalization, polarization and dominance. The empirical analysis verifies that it outperforms the existing indices of ethnic diversity in explaining ethnic conflict onset. (*JEL* D63, D74, O57)

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

In this paper we provide a novel investigation of the relation between different distributional aspects of ethnicity and occurrence of ethnic conflicts. The empirical evidence on the association between ethnic diversity and conflict is generally ambiguous and still no broad consensus is reached on which distributional aspect of diversity is an important correlate to conflict. Another strand of literature develop several theoretical models that investigate the relationship between diversity and conflict. Esteban and Ray (1999) derive from a rent-seeking model a relation between polarization and conflict. Along similar lines Montalvo

[*]**Kovacic**: Ca' Foscari University of Venice, Department of Economics, Canareggio 873, 30121 Venice; matija.kovacic@unive.it. **Zoli**: University of Verona, Department of Economics, Via Cantarane, 24, 37129 Verona; claudio.zoli@univr.it. We are grateful for comments and suggestions received on an earlier version of the paper presented at the International Economic Association World Congress (2014), Jan Tinbergen European Peace Science Conference (2014), European Econometric Society Meeting (2013), and Meeting of the Society for the Study of Economic Inequality (2013).

and Reynal-Querol (2005, 2008) motivate the use of discrete ethnic polarization as a correlate for conflict. Esteban and Ray (2011) propose a behavioral model according to which the equilibrium level of conflict can be approximately described by a weighted average of a Gini's inequality index, the fractionalization index, and a specific polarization index from the class of indices axiomatically derived in Esteban and Ray (1994). Esteban, Mayoral and Ray (2012) implement the above measures in an empirical exercise and confirm the predictions of the model. Caselli and Coleman (2013) provide a model of social distributive conflict in which ethnic boundaries are not fixed and immutable, and relate the incidence of ethnic conflict to groups relative size and to the share of expropriable resources in overall wealth.

In this paper we show that the probability of conflict outbreak can be related to fractionalization, polarization and dominance. The relative importance of these three aspects of diversity in the determination of conflict depends on the characteristics of the underlying population distribution across groups. We start from the basic specification of the Esteban and Ray's (1994) [ER henceforth] model of social antagonism, and characterise a parameterized index of diversity which we refer to as the *P Index of Conflict Potential* that combines the groups' effective power and the between-group interaction. Our approach departs from ER by *three* specific features. First, like in Montalvo and Reynal-Querol (2005), we define the distances between groups using a discrete rather than a continuous metric. Second, we assume that the 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. As a consequence, a power of a group is not necessarily proportional to its size. Third, we do not treat each group as an independent actor but we assume that groups can either act individually or form alliances with other groups in order to exploit potential increasing returns to coalition formation.

We show that for some parameter values, the P index reduces to the existing diversity indices (fractionalization and discrete polarization). For high values of the parameter, the index is able to capture the presence of dominance, where all the power goes to the majoritarian group. The effective power component of the index in these cases approaches the Penrose (1946) - Banzhaf (1965) and Shapley - Shubik (1954) measure of voting power in a simple majority game if applied to the distribution of the population shares of the different groups.

In general, the index emphasizes differently the overall effects of power and between groups interaction according to the features of the underlying population distribution across groups. If the power component results more evenly distributed across groups the interaction component becomes predominant and the index highlights the fractionalization as the relevant aspect of diversity, while for unequal distributions of the powers the emphasis is given to the combined effect of dominance and interaction.

We present an empirical exercise in which we test the performance of the indices of conflict potential we have derived against the commonly used distributional indices of ethnic diversity. Since our measures link the features of the population distribution across groups to the probability of conflict outbreak, we consider conflict onset rather than incidence as the dependent variable along the lines of the empirical specifications in Wimmer, Cederman and Min (2009), Cederman and Girardin (2007) and Fearon (2003). Using the data on ethnic groupings from the Ethnic Power Relations data set (Wimmer and Duhart, 2014) we show that, when compared to the existing and widely used indices of ethnic diversity, the index based on the Penrose - Banzhaf measure of relative power results a strong and significant correlate of ethnic conflict onset even after inclusion of an additional set of regressors and under alternative model specifications. Our results highlight the fact that the different aspects of diversity should be combined in order to investigate their relation with conflict onset. The derived index provides a specification of the relative relevance of these aspects across different distributions.

The paper is organized as follows. In Section 1 we axiomatically derive the P index of conflict potential. Section 2 analyses the shape of the P index for different parameter values and different population distributions, and explores the differences between the derived indices and the existing distributional measures using the data on ethnic distribution for a large set of countries. Section 3 presents our main empirical results and Section 4 concludes.

1 The P Index of Conflict Potential

Consider a population partitioned into $n \geq 2$ non-overlapping groups. Let π_i be the relative population size of group i, where $i = 1, 2, ..., n$, and $\Pi = (\pi_1, \pi_2, ..., \pi_n)$ denote the vector of groups' population shares. As in ER we 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 another is a function of the distance between them. Since, by assumption, individuals within each group are all alike, 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 extend the ER approach and assume that the power of a group depends also on the relative sizes of all the groups listed in Π. As in ER, we specify a function Φ that combines the group's power, that depends on π_i and Π, with the alienation felt towards other groups. Following Montalvo and Reynal-Querol (2005), we define the distance D_{ij} between individuals belonging to two groups i and j, using a discrete metric, *i.e.*,[1]

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

The potential of conflict in a society then derives from the interaction between power and alienation aggregated over all pairwise comparisons:

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

[1]As pointed out in Montalvo and Reynal-Querol (2005), when the population is partitioned according to some categorical attribute like ethnicity, language or religion, identifying groups according to the so-called "*belong - does not belong to*" criterion is less controversial than defining the distances between them simply because it reduces significantly the measurement error.

We assume that $\Phi(\pi_i, \Pi, 0) = 0$ and let $\Phi(\pi_i, \Pi, 1) := K\phi^n(\pi_i, \Pi)$ for some constant $K > 0$. We add a superscript n to ϕ to distinguish between distributions characterised by a different number of groups. Since $\sum \pi_i = 1$, the index defined in (1) can be written as:

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

and will be called the *P Index of Conflict Potential*. The function $\phi^n(\pi_i, \Pi)$ for $i = 1, 2, ..., n$, will be referred to as the *effective power* associated with group i. The $P(\Pi)$ index, hence, is obtained as a combination between two different elements, namely the groups' effective power and the between groups' interaction, $\pi_i(1-\pi_i)$, measuring the probability of randomly selecting an individual from group i that interacts with an individual from another group. The sum of these components gives the probability that two individuals randomly selected from a population belong to different groups. Special cases of (2) are the Montalvo and Reynal-Querol (2005) discrete polarization index, $RQ = 4 \sum \pi_i^2(1 - \pi_i)$, where each group's power is equal to its relative population size, and the fractionalization index, $FRAC = \sum \pi_i(1 - \pi_i)$, which derives from (2) when each group's power is constant and equal to 1.

The most interesting specification of the P Index in (2) is obtained when each group effective power is neither constant nor proportional to the group relative size but may also depend on the distribution of the relative sizes of the other groups. Here we assume that groups are allowed to form coalitions that generate bipartitions of the population, and that the effective power of each group depends on all the potential contributions of that group to the worth of all the coalitions that it can theoretically belong to. With only two groups, the power of both groups is $1/2$ when their sizes are equal, and is non-decreasing in the groups' size. The extreme case is the one where the power of the bigger group is 1, while the power of the smaller one equals 0. Applying this latter rule to any bipartition of the population when the number of groups is larger than two, the marginal contribution of any group to the worth of a coalition equals 1 whenever the sum of the relative sizes of all the groups forming the coalition exceeds 1/2 and becomes smaller than 1/2 if that particular group leaves the coalition. The relative weight of the marginal contributions of any group with respect to the sum of the contributions of all the groups in the population, represents then a measure of that group's effective power. In this particular case where the extreme relevance is given to inequality between the bipartitions, the effective power coincides with the relative Penrose-Banzhaf index of voting power in a simple majority game. The associated index corresponds to the extreme element of a parametric family that will be discussed in detail in the following sections.

1.1 Axiomatic Derivation of the Effective Power Function

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

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

and satisfies the next properties.

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

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

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

Normalization requires that the powers of all groups sum to 1, and implies that the effective power of each group is bounded in the interval $[0, 1]$.

[2]For ease of exposition here we consider $\Pi \in \Delta^n$ even though for a given π_i only a subset of Δ^n is consistent with having one element associated with group i equal to π_i, we also include the extreme values $\pi_i \in \{0, 1\}$ within the domain.

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 the effective power of a larger group *cannot* be lower than the effective power of a smaller group. This property implies the Symmetry of ϕ^n which requires that, if two groups are of equal size, then their effective power has to be the same. The reverse, however, is *not* necessarily true: the effective power could still be equal for groups of different relative size. Monotonicity in combination with Normalization implies that if all groups have identical relative size, each one of them has an effective power equal to $1/n$. This result will provide a reference point for all the indices that we will obtain from the axiomatization. In fact, a common feature of these indices is that they all exhibit the same value for distributions where all the groups are of equal size. Moreover, this value will be proportional to the fractionalization index divided by n.

We now introduce two crucial assumptions: **i)** groups can either act individually or through a coalition, and **ii)** if any two or more groups form a coalition, the remaining groups belong to the "opponent" block. So we consider only *bipartitions* of the population.

What is the *rationale* behind these two assumptions? Suppose that there are 3 groups involved in a contest with only one strategic endowment, namely human resources. A relatively smaller group that is interested in winning the contest may find profitable to join the forces with some other group 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 large enough to ensure the victory alone will act as an independent actor. Hence, one block or coalition may be formed in order to contrast or challenge the other block. 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 a coalition against the third if the formation of the alliance generates *synergies* that enhance the winning probability of the coalition.[3]

[3] Skaperdas (1998) argues that this tendency is not only theoretical but also frequent in many real life situations and provide an example of the "... on and off alliance of the Bosnian and Croat forces against the more (strategically) well endowed Serb forces in Bosnia during the recent past ..."

Here we do not model any endogenous mechanism of coalition formation nor we are interested in which coalition is more likely to form. The probability distribution over coalitions, hence, is assumed to be symmetric. Symmetry is a plausible assumption in our case since we do not have or use information about the differences among groups and we use a discrete metric to define these distances.[4] Considering only the bipartitions of the population, we rule out the possibility that more groups run on their own against the rest.[5] As we will show later, even under these simplifying assumptions the distribution of the effective power between 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 diversity indices based on the assumption of groups as independent actors.[6]

In order to characterise the effective power for any arbitrary number of groups we first consider the simpler case of a distribution with only two groups. The results that we obtain will then be used to generalize the analysis for any arbitrary number of groups.

Consider a population divided into two different groups ($n = 2$) with population shares π and $1 - \pi$. Denoting with $\phi^2(\pi)$ and $\phi^2(1 - \pi)$ the effective power of the groups, we define the *relative effective power* between them as:

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

Thus, the relative effective power between groups is a function $r(\cdot)$ of the groups relative population size ρ that coincides with the population shares *odds ratio*. From Monotonicity it follows that whenever $\pi = 1/2$, hence $\rho = 1$, the groups will equally share the power, that

[4]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 s/he possesses, we could define the probability of any coalition in terms of the similarity between the groups income or wealth attributes.

[5]We do not rule out any coalition between two or more groups. This assumption is not unrealistic, since in many situations that involve coalition formations in conflicts, even unmatchable parties sometimes coordinate their interests in order to contrast the opponent, even when they are aware that the coalition is temporary (Esteban and Sakovics, 2003).

[6]For instance, the RQ index and the fractionalization index assume that there is no interaction between groups.

is, $r(1) = 1$.[7] The relative effective power is supposed to satisfy the following property:

Axiom 3 ***Two Groups Relative Power Homogeneity (2GRPH)*** *Given* Π *and* Π', *let* $\rho, \rho' \leq 1 (i.e., \pi, \pi' \leq 1/2)$. *If* $r(\rho), r(\rho') \neq 0$ *then:*

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

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 larger group (*i.e.*, $\rho = 0.4$). Now imagine that a portion of the population from the second group migrates in a neighboring country such that the size of the smaller group is now 80% of that of the larger group (*i.e.*, $\rho = 0.8$). Thus ρ 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 where ρ doubles but the relative size of one group moves from 30% to 60%. The 2GRPH axiom 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 of the same proportion across the two distributions.

We can now state the first result proved in the Appendix.

Lemma 1.1 *Let* $n = 2$, *the effective power of a group with population share* π *satisfies Axioms N, M and 2GRPH if and only if* $\phi^2(\pi) = \phi^2_\alpha(\pi)$ *for* $\alpha \in \Re_+ \cup \infty$ *where*

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

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

[7] Given a functional form for ϕ^2, then $r(\rho)$ derives directly by recalling that $\pi = \rho/(1+\rho)$ and thus $r(\rho) := \frac{\phi^2(\rho/(1+\rho))}{\phi^2(1/(1+\rho))}$.

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.[8] The coefficient α represents the elasticity of the relative effective power with respect to the relative population size. When $\alpha = 0$ the relative effective power equals 1, for $\alpha = 1$ each group's power equals its population share, while for $\alpha \to \infty$ the majoritarian group holds the absolute power. We will call *dominant* such group with $\pi > 1/2$.

Consider now $n > 2$. Groups are allowed to form coalitions that generate the bipartitions of the population. A coalition is defined as any subset of the set N of all groups (including the empty set). In particular the *grand* coalition contains all the groups; an *individual* coalition contains only one group; and the *empty* coalition 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 a measure of effective power has to consider all the potential contributions of a group to all the coalitions that it can possibly belong to.

Denote with C_i the set of all coalitions c that include group i. This set contains both the grand coalition and the $i's$ individual coalition. The power of any coalition c is obtained by Lemma 1.1 as $\phi^2(\sum_{j\in c}\pi_j)$, where 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 power of any coalition $c \in C_i$ as (Shapley, 1953):

$$m_i(c) := \phi^2(\sum_{j\in c}\pi_j) - \phi^2(\sum_{j\in c}\pi_j - \pi_i). \tag{4}$$

[8] The results achieved *cannot* be obtained from the properties underlying a standard contest success function. For instance, Skaperdas (1996, 1998) assumes homogeneity of the relevant variables that are unbounded, which is not the case in our problem where π is bounded between 0 and 1.

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

$$M_i = \sum_{c \in C_i} m_i(c). \tag{5}$$

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

Axiom 4 ***Relative Effective Power (REP)*** *For any $i, j \in N$, $i \neq j$ and $n \geq 2$; $\exists\, g : \Re_+ \to \Re_+$, such that for $\phi^n(\pi_j, \Pi) > 0$ we have*

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

The REP axiom states that the *relative* effective power between any two groups $i, j \in N$ depends on their sum of marginal contributions to all the coalitions that they can theoretically belong to. 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 will be determined *exclusively* by a ratio of a transformation $g(\cdot)$ of their own Ms. It follows that the effective power of groups with same M has to be the same.

The relationship between the ratio of marginal contributions and the relative effective power is clarified by the following axiom where comparisons are extended to groups belonging to different distributions.

Axiom 5 ***n Groups Relative Power Invariance (nGRPI)*** *Given two distributions,* Π *and* Π' *with the same number of groups $n \geq 2$, if $\phi^n(\pi_j, \Pi) > 0$ and $\phi^n(\pi'_j, \Pi') > 0$ then*

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

According to $nGRPI$ if we compare two population 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. That is, the relative effective power is *invariant* with respect to the distribution of the groups population shares for groups with the same sum of marginal contributions.

Next theorem, proved in the Appendix, provides a role for the sum M_i^α of the marginal contribution to all coalitions of group i obtained as in (5) making use in (4) of the functional form ϕ_α^2 derived in Lemma 1.1.

Theorem 1.2 *The effective power of group i satisfies Axioms N, M, 2GRPH, REP and nGRPI if and only if:*

$$\phi_\alpha^n(\pi_i, \Pi) = \frac{M_i^\alpha}{\sum_{j=1}^n M_j^\alpha}, \quad for\ i \in N; \quad \alpha \in \Re_+ \cup \infty. \tag{6}$$

Group i's effective power, hence, is defined as the relative sum of the marginal contributions of this group to all possible coalitions, valued according to ϕ_α^2. Given (6), the effective power of a group can be a function of the relative size of all the groups in the population. For $n > 2$ and $\alpha \notin \{0, 1\}$, the effective power of any group i depends on *both* π_i and Π_{-i}. As a consequence, the effective power of a group with a fixed population share π_i may vary significantly across different population distributions in response to the variation of the relative size of the other groups Π_{-i}.

The previous result suggests that the effective power is *not* necessarily proportional to the groups' relative size. This is in line with the literature on *voting power*.[9] For instance, when $\alpha \to \infty$, the group $i's$ effective power, $\phi_\infty^n(\pi_i, \Pi)$, coincides with its relative Penrose - Banzhaf index of voting power in a simple majority game (Felsenthal and Machover, 1998).[10]

[9] In his famous critique of the practice of assigning voting weights proportional to the number of citizens in different legislative bodies ("one man, one vote" requirement), Banzhaf (1965, p.318) argues that the number of votes is not even a rough measure of the voting power of the individual legislator.

[10] A simple majority game is a voting game in which an actor (or a coalition of actors) wins if the number of votes s/he possesses exceeds 50% of the total number of votes. In this case s/he is attributed the value of 1 and 0 otherwise. The Penrose - Banzhaf Index of voting power measures the ability of an actor to influence the outcome of voting in a collectivity and it is defined as the relative number of times s/he can switch a coalition from losing to winning relative to the total number of swings of all the other actors. In our case, according to $\phi_\infty^2(\pi)$ in Lemma 1.1, in addition it is considered that the power of a coalition covering exactly 50% of the population is 1/2.

2 Properties of the P Index of Conflict Potential

With the effective power function specified in (6), the P *index of conflict potential* becomes:

$$P_\alpha^n(\Pi) = K \sum_{i=1}^{n} \frac{M_i^\alpha}{\sum_{j=1}^{n} M_j^\alpha} \; \pi_i(1-\pi_i); \quad \alpha \in \Re_+ \cup \infty. \tag{7}$$

We set $K = 4$ so that the index ranges between 0 and 1.[11]

Within the index formulation, the relative importance of *groups' power* and *between groups' interaction* depends on the features of the population distribution across groups, and crucially on the parameter α.

In the next subsection we analyse the properties of the P index for different values of the coefficient α and for different population distributions. We show that for the case of two groups the parameter α plays no role and the P index reduces to the RQ index of discrete polarization which is twice the fractionalization index. When the population is partitioned into more than two groups, the shape of the index depends on the choice of the parameter α.

2.1 The Role of the Coefficient α

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 equals $1/n$. The P index becomes:

$$P_0^n(\Pi) = 4 \sum_i \frac{1}{n} \pi_i(1-\pi_i) = 4\frac{1}{n} \cdot FRAC. \tag{8}$$

[11] When $K = 4$, for any $\alpha \neq 0$, the supremum of the index is 1 for any n. If $\alpha = 0$, the supremum is 1 if $n = 2$ and decreases as n increases.

This is *not* exactly the fractionalization index because it is scaled by $4/n$. 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 groups. For $P_0^n(\Pi)$ the interaction component is combined with the effective power assigned to each group, which is *decreasing* in n. For a given n, the P_0^n index and the fractionalization index provide the same ranking. However, they significantly differ over distributions with different n. This aspect can be made evident when all the groups have the same size. In this particular case, P_0^n and $FRAC$ move in opposite directions as n increases. In fact, when the relative size of each group is $1/n$, the P_0^n index becomes $4\frac{1}{n}\frac{n-1}{n}$ while $FRAC = \frac{n-1}{n}$.[12]

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 may not necessarily lead to conflict because groups become smaller, hence their chances to mobilize efficiently may decrease. 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 both these aspects into account. For the P_0^n index, on the other hand, as n increases the contribution of interaction increases but it is rescaled by the power component, which decreases at a higher rate. With n equally sized groups, the maximum of conflict potential is reached for two groups, as happens with the discrete polarization index.

Case 2. When $\alpha = 1$, the effective power of each group equals its relative population size. With $\phi_1^n(\pi_i, \Pi) = \pi_i$ for all i, the P index reduces to the RQ index of discrete polarization:

$$P_1^n(\Pi) = 4\sum_{i=1}^{n} \pi_i^2(1-\pi_i) = RQ. \tag{9}$$

The larger is a group, the proportionally higher is its effective power to translate alienation

[12]It follows that as n increases P_0^n converges to 0 while $FRAC$ increases and converges to 1.

into effective voicing.

For $\alpha = 0$ and $\alpha = 1$, hence, a group i's effective power depends only on n and π_i. In both cases 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 group i is a function of *both* π_i and Π_{-i}. If we denote by π^* the relative size of the largest group in the population and with γ_i the relative Penrose-Banzhaf Index of voting power associated to group i, the P^n_∞ index can be written as:

$$P^n_\infty(\Pi) = \begin{cases} 4\pi^*(1-\pi^*) & \text{if} \quad \pi^* > 1/2, \\ 1-\theta_n(1-P^n_0(\Pi)) & \text{if} \quad \pi^* = 1/2, \\ 4\sum_i \gamma_i \pi_i(1-\pi_i) & \text{if} \quad \pi^* < 1/2. \end{cases} \tag{10}$$

where $\theta_n = n/(2^{n-1}+n-2)$.

When one group is dominant, *i.e.,* its relative size exceeds 1/2, the potential of conflict is determined only by that group's relative size. In this case the P index coincides with the interaction component associated with this group. As the relative size of a dominant group approaches 1/2 the value of the index converges to 1. Similarly, when the size of a dominant group increases, the overall interaction decreases, and the index moves downward.[13] 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 a convex combination between its maximum value 1 and $P^n_0(\Pi)$.

[13]When $\pi^* > 1/2$ the P_∞ index is equivalent to the RQ index with only two groups (which is proportional to the fractionalization index) and measures the degree of bi-polarization, with the majoritarian group at one extreme and the "opponent" block at the other.

2.2 P Index for Two and Three Groups

In the case of two groups the interaction component of each group is symmetric. As a result the P index is proportional to $FRAC$ irrespective of α:

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

Thus, the simplest way to analyse 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 $\sum \pi_i = 1$). For expositional purposes, we fix the size of one group (here π_2) to $1/3$ because we want to compare alternative population distributions with the uniform distribution, and we express the indices in terms of π_1.

When all the groups have the same size, the P index yields the same value $4\frac{n-1}{n^2}$ for any α. The P index with $\alpha = 1$ (*i.e.*, the RQ index) is invariant to population transfers between groups when the relative size of one group is set to $1/3$. The shape of P_0^3 is identical to the shape of the fractionalization index that is quadratic and concave with respect to π_1 with the maximum for $\pi_1 = 1/3$.

For α different from 0 and 1 the index becomes non-monotonic in π_1 for $\pi_1 > 1/3$. As α approaches infinity, the shape of the index becomes particularly interesting. With $n = 3$, $\pi_2 = 1/3$ and $\alpha \to \infty$, the P index is:

$$P_\infty^3(\Pi) = \begin{cases} 4\pi_1(1-\pi_1) & \text{if} \quad \pi_1 > 1/2, \\ \frac{2}{5} + \frac{3}{5}P_0^3(\Pi) & \text{if} \quad \pi_1 = 1/2, \\ P_0^3(\Pi) & \text{if} \quad \pi_1 < 1/2, \end{cases} \tag{11}$$

where $\pi_1 \geq 1/3$ denotes the population share of the larger group.

Figure 1 shows the P^3 index for $\alpha = 0$ (dashed curve), $\alpha = 1$ (dot-dashed line) and $\alpha > 1$ (solid curves) expressed in terms of π_1.[14] In the limit as α approaches infinity, the P^3 index assumes a particular shape characterised by a discontinuity at $\pi_1 = 1/2$ (solid curve in the

[14]The graph is symmetric by construction around $\pi_1 = 1/3$.

right hand-side of the figure).

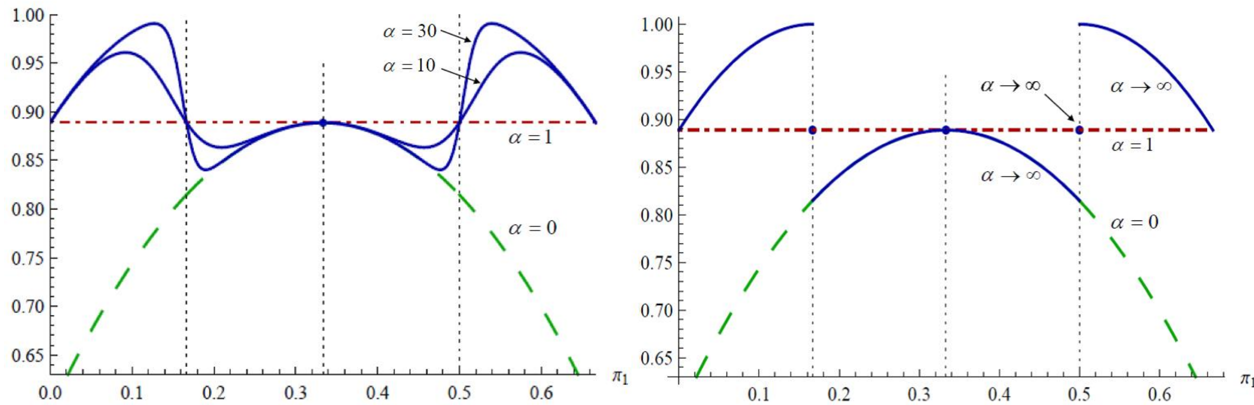

Figure 1: P Index for $\alpha = 0, 1, 10, 30$ (left), and $\alpha = 0, 1, \infty$ (right).

Starting from a uniform distribution, *i.e.*, when $\pi_1 = 1/3$, the P^3_∞ index follows the shape of the fractionalization index. As π_1 increases, the population becomes less fragmented and the index decreases. When the relative size of group 1 reaches $1/2$, the index "jumps" to $8/9$, the constant value obtained for $\alpha = 1$. Once π_1 exceeds $1/2$, the index reaches almost 1 and then decreases. The P^3_∞ index reaches its maximum when the relative size of one group becomes scarcely higher than $1/2$ because in that case this group gains the absolute power and the "opposition" is powerless. This fact is in line with the notion of dominance of one group over the other(s).[15] It is worth noting here that the P^3_∞ index combines dominance (and, hence power) and interaction. It follows that, as the size of the dominant group increases, the probability of interaction decreases and so also the potential of conflict. As a result for a very large dominant group the index tends to 0. With no dominance, *i.e.*, if the relative size of all groups is lower than $1/2$, the conflict potential is entirely determined by the interaction component - the shape of P^3_∞ follows the shape of the fractionalization index.

[15]For instance, Collier and Hoeffler (2004) reason in terms of minority exploitation in ethnically heterogeneous societies, 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 legal channels for achieving political change, use of arms or some other kind of conflict technology is regarded a plausible alternative strategy.

2.3 P_∞ Index for more than Two Groups

For any arbitrary number of groups, the value of the P_∞^n index in the presence of dominance coincides solely with the interaction component of the dominant group. In the absence of dominance the index is either proportional to fractionalization or to a combination of fractionalization and the interaction component of either the largest or the smallest group as long as the number of groups in the population is not too large.

Consider for instance population distributions with $6 > n \geq 3$ such that $\pi_1 > \pi_2 > ... > \pi_n$ and $\pi_1 < 1/2$. For $n = 3$ the P_∞^3 index is given by the formulation in (11), thus if $\pi_1 < 1/2$ it coincides with P_0^3 and is proportional to fractionalization.[16]

For $n = 4$ and population distributions characterised by $\pi_1 + \pi_4 < 1/2$, the groups' relative power distribution is $(1/3, 1/3, 1/3, 0)$, and for a given π_4, the value of the P_∞^4 index is entirely determined by the interaction component. When the population becomes more fragmented, the interaction increases and the P_∞^4 index follows the shape of P_0^4. Similarly, for all those distributions where $\pi_1 + \pi_4 > 1/2$, the groups' relative power is then constant and its distribution is given by $(1/2, 1/6, 1/6, 1/6)$. Also in these cases, now for a given π_1, the P_∞^4 index results linearly correlated with the P_0^4 index.[17] Thus, in the absence of dominance, when $n = 4$, the P_∞^4 index combines fractionalization with the interaction component either of the largest group or of the smallest one, or of both.

With $n = 5$, the P_∞^5 index coincides with P_0^5 when $\pi_1 + \pi_2 < 1/2$, and for a given π_1 is linearly related to P_0^5 when $\pi_1 + \pi_5 > 1/2$. In fact, in the former case, all the groups in the population have the same relative power of $1/5$, while in the latter case the groups' relative power distribution is constant and given by $(7/11, 1/11, 1/11, 1/11, 1/11)$, which makes the P_∞^5 index linearly correlated with the P_0^5 index for a given value of π_1.[18]

[16]With three groups, when $\alpha \rightarrow \infty$, if the larger group is not dominant, the relative Penrose - Banzhaf index of power of each group is $1/3$ irrespective of their relative sizes.

[17]The P_∞^4 index in this case coincides with $-\frac{1}{3} \cdot 4\pi_4(1-\pi_4) + \frac{4}{3}P_0^4$ for distributions characterised by $\pi_1 + \pi_4 < 1/2$, and with $\frac{1}{3} \cdot 4\pi_1(1-\pi_1) + \frac{2}{3}P_0^4$ for distributions where $\pi_1 + \pi_4 > 1/2$. While for all distributions where $\pi_1 + \pi_4 = 1/2$, the P_∞^4 index is $\frac{1}{6} \cdot 4\pi_1(1-\pi_1) + P_0^4 - \frac{1}{6} \cdot 4\pi_4(1-\pi_4)$.

[18]The P_∞^5 index in the latter case is given by $\frac{6}{11} \cdot 4\pi_1(1-\pi_1) + \frac{5}{11}P_0^5$.

As the number of groups increases, even a very small variation in the groups' relative size may significantly alter their relative power. As a consequence, the correlation between the P^n_∞ index and fractionalization becomes less clear. For instance, when $n = 6$, starting from a distribution with a very low dispersion of population across groups, and increasing the relative size of the largest group, the population becomes less fragmented and the P^6_0 index decreases. However, the relative power shifts towards the larger groups and the P^6_∞ index moves in the opposite direction with respect to the P^6_0 index.

2.4 Comparison between Indices: a first insight into the data

In the previous section we have shown how the choice of the parameter α and the features of the population distribution across groups determine the shape of the P index. In this section we analyse graphically the relationship between the P index for different values of α using the data on ethnic distribution for 146 countries from the Ethnic Power Relations data set (Wimmer and Duhart, 2014). We discuss the features of the data in the next section.

Figure 2 shows the relationship between P^n_0, P^n_4, P^n_{10} and P^n_∞, versus P^n_1 (RQ polarization index). As α increases, the correlation between RQ and the P index decreases, especially for high values of the indices. For instance, the correlation between RQ and P for $RQ \in [0.7, 0.9]$ is 0.5 in the case of $\alpha = 4$, is 0.27 for $\alpha = 10$ and boils down to 0.19 for $\alpha \to \infty$.

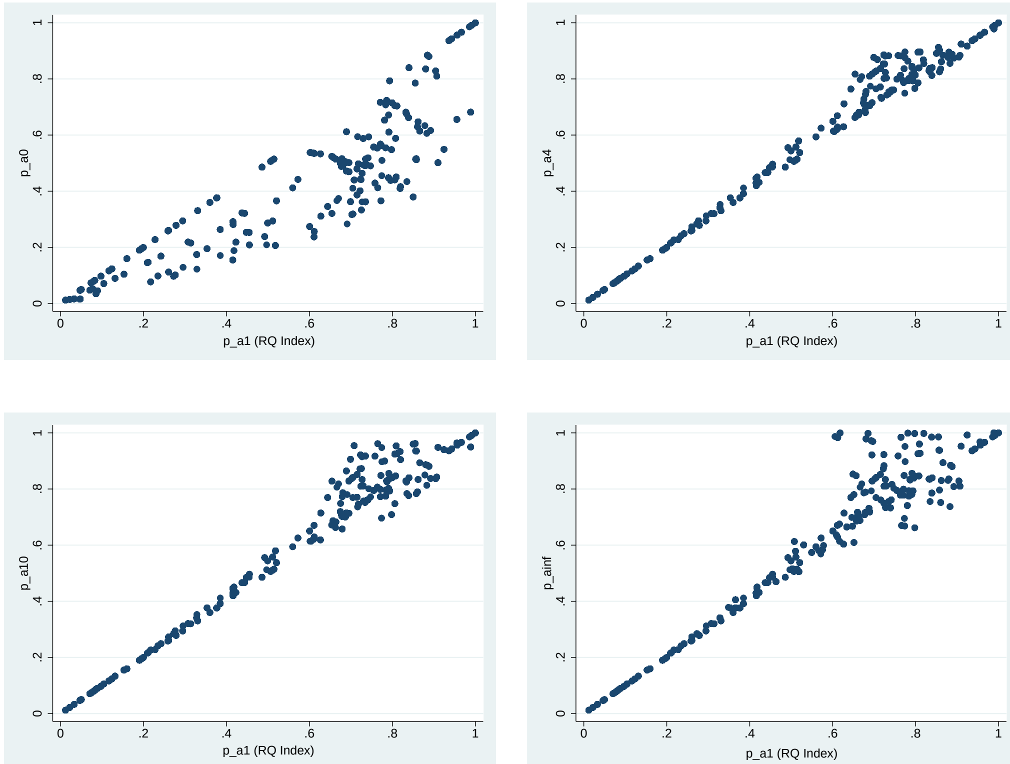

Figure 2: P index with $\alpha = 0, 4, 10$ and $\alpha \to \infty$ versus RQ polarization index. Source: *Ethnic Power Relations (EPR3) data set, Wimmer and Duhart (2010)*

.

In order to analyse the relationship between the probability of conflict outbreak and different distributional aspects of ethnicity, we preliminarily verify how the indices correlate with conflict outcomes. Figure 3 shows P_∞^n versus RQ with the labels for the frequency of ethnic conflict onsets [EC] in a time range from 1946 and 2005. The horizontal and the vertical lines represent respectively the mean values of P_∞^n (0.5809) and RQ (0.5408).[19] The two indices differ most when they are both larger than their respective means. This range of values is associated with 76% of all conflict episodes as shown in detail in the right-hand side frame of Figure 3.

In the next section we test the empirical performance of the derived indices and we show

[19]The mean value of P_∞^n in conflict and peace episodes is 0.74 and 0.57 respectively, while for the RQ index the mean values associated to conflict and peace episodes are respectively 0.65 and 0.54.

that the predictive power of P^n_∞ is significantly higher than the predictive power of the other indices of ethnic diversity in the explanation of ethnic conflict onset.

Figure 3: P^n_∞ versus RQ with EC Label. Source: *Ethnic Power Relations (EPR3) data set, Wimmer and Duhart (2010).*

3 Empirical Relevance of the P Index of Conflict Potential

In this section we investigate the relationship between the P index and conflict behaviour. The measures of conflict potential relate the features of the population distribution across groups to the probability of conflict onset rather than to the incidence or intensity of a conflict. Our empirical exercise hence relies on a logistic model specified in Wimmer, Cederman and Min (2009) and Cederman and Girardin (2007) 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 a conflict and 0 otherwise. The data on ethnic distributions and main explanatory and control variables come from the Ethnic Power Relations (Version 3.01) dataset [EPR3 henceforth] provided by Wimmer and Duhart (2014) [WD henceforth] which extends and improves the original Version 1.0 of the Ethnic Power Relations dataset

(Wimmer, Cederman and Min, 2009) [WCM henceforth].[20] As for the ethnic groups coding, the EPR3 data set has several advantages with respect to other data sources commonly used in the empirical literature such as the Minority at Risk dataset (Gurr et al. 1993; Gurr 2000), the Atlas Narodov Mira (1964) dataset and the Fearon (2003) dataset. First, it identifies all politically relevant ethnic groups and records changes in politically relevant categories over time. Second, the coding of ethnic groups does not limit the possibilities to any existing ethnic group list. Third, the EPR3 dataset assesses formal and informal degrees of political participation and exclusion along ethnic lines.[21] As a robustness check, however, we also test the empirical performance of the P index using the Fearon's (2003) groupings.

Regarding the conflict data, EPR3 extends the Armed Conflict Data Set [ACD henceforth] by coding each conflict for whether rebel organizations pursued ethno-nationalist aims and recruited along ethnic lines.[22] We consider *ethnic* conflicts for several reasons. First, the majority of the conflicts after the Second World War were ethnic in nature. Second, there is a substantial difference in the nature and the determinants of ethnic and non-ethnic conflicts (Sambanis, 2001, 2004). There is no reason to believe that ethnic diversity is an important determinant of non ethnic conflicts, such as revolutions or any other form of anti-governmental protest. Third, ethnic conflicts are closely related to cultural and political

[20]The version 3.01 of the Ethnic Power Relations dataset improves the previous coding of much of Latin America, Central Asia, Eastern Europe, and many other countries. See: `http://www.epr.ucla.edu/`

[21]The list of politically relevant ethnic categories, in some cases, was changing from one sub-period to another, either because certain categories ceased to be or became relevant for the first time, or because higher or lower level of ethnic differentiation became salient (WCM, p.326). The coders were asked not to exclude groups based only on their relative population size, since even small groups can be significant at the national or regional level. An ethnic category is politically relevant if is represented by at least one significant political actor or if the members of the groups are "[...] systematically and intentionally discriminated against in the domain of public politics [...]" (WCM, p.325). By "significant" political actor it is meant a political organization (not necessarily a party), that is *active* in the national political arena. 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.

[22]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). The authors identify as *ethnic* "[...] the aims of achieving ethno-national self-determination, a more favourable ethnic balance of power in government, ethno-regional autonomy, the end of ethnic and racial discrimination, language and other cultural rights [...]" (WCM, p.326). All other wars are defined as *non-ethnic*.

identity - in ethnically heterogeneous societies political mobilization occurs mostly along ethnic lines.

The list of explanatory and control variables considered in our empirical analysis is the one commonly used in conflict research[23] (Fearon and Laitin, 2003; Collier and Hoeffler, 2004; Montalvo and Reynal-Querol, 2005; Sambanis, 2001; Hegre and Sambanis, 2006; Wimmer, Cederman and Min, 2009), it includes: GDP per Capita, Population Size, Oil Production per Capita, Mountainous Terrain, Noncontiguous Territory and New State, Democracy and Anocracy, Instability (regime change). In order to take into account the political dimension of ethnic conflicts, we follow WCM and consider 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. 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, and regional dummies.[24] In order to account for the variation in within-region ethnic conflict onset due to factors that are region-specific over time, we also construct a regional time trend dummy variables.[25]

Together with the indices of fractionalization, discrete polarization, and several dominance dummies, we calculate the P index of conflict potential for different values of the parameter α using the groups' relative shares calculated in relation to total population. Given a particular structure of the effective power function and the related computational complexities, in order to calculate the P index for $\alpha \geq 2$, we consider all the countries with no more than 6 ethnic groups as well as all those countries for which the number of groups was reduced to 6 according to the following criteria: the sum of the population sizes of all the groups ranked below the sixth largest group could not exceed 8% of the total population, and the relative population size of the biggest excluded group could not be larger than 5%. In such a way we consider 21 out of 27 countries originally characterised by more than 6

[23]The illustration of the explanatory variables is given in the Appendix.

[24]The world is divided in 6 geographical regions: Asia, Sub-Saharan Africa, Latin America, East Europe, North Africa and Middle East, and Western.

[25]The regional time trend dummy variables are defined as: $RegTrend = RegDummy + RegDummy * Year$.

ethnic groups. The average (median) size of the largest eliminated group is 2% (2%), while the average (median) number of eliminated groups across countries is 2.612 (2). The average (median) sum of the size of the eliminated groups, on the other hand, is 3.3% (3%). The remaining 6 countries[26] were excluded from the analysis since they don't meet one or both the above mentioned criteria. Regarding the P index for $\alpha = 1$ and $\alpha \to \infty$, as well as the fractionalization index, for our main empirical specifications we consider the original dataset with no group or country excluded, containing 146 countries for which complete ethnic grouping is available.

3.1 Explaining Ethnic War Onset

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 risk of 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 mitigating effects of social fractionalization on conflict, however, disappear in the presence of ethnic dominance: with one ethnic group covering between 45% and 90% of the overall population, the risk of conflict is almost doubled. 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 Wimmer, Cederman and Min (2009) show that once we account for the political exclusion and competition along ethnic lines, ethnic diversity has no effect on the probability of conflict outbreak. 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 homo-

[26] The list of countries not considered in the reduced sample contains: Russia, Democratic Republic of Congo, India, China, Namibia and Sudan.

geneous and highly heterogeneous societies are less conflict prone with respect to societies divided into few prominent ethnic groups. Following this logic, Montalvo and Reynal-Querol (2005) apply their index of discrete ethnic polarization and find a positive and statistically significant association between ethnic polarization and the incidence of conflict.[27] 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 a dependent variable while Collier and Hoeffler (2004) find no statistically significant relationship between ethnic polarization and the risk of conflict outbreak.

From an empirical point of view, hence, the relationship between ethnic diversity and conflict is quite ambiguous and still no broad consensus is reached on which distributional aspect of diversity is an important correlate to conflict onset. The particular feature of the P index of conflict potential that combines different aspects of diversity depending on the characteristics of the underlying population distribution across groups, may make a difference. In order to assess the relative performance of the P index with respect to different α, we first estimate our models of ethnic conflict onset using the reduced sample of countries. We then reestimate our models using the entire sample with no group or country excluded from the analysis, and consider the P index with the best performance in terms of the statistical significance and the goodness of fit.

Table 1 presents the results of our estimations based on the reduced sample of countries.[28] Regarding the parameter α, only the coefficients associated to the P index with $\alpha \geq 4$ are significantly different from zero. The magnitude and the significance of the coefficient associated to the index increases with α and reaches its maximum for $\alpha \to \infty$. Similarly, the goodness of fit measured by the Pseudo R^2 is also increasing in α. Indeed, by comparing the outcome of various estimations based on P^n_α, it results that the highest value of the Pseudo

[27] Collier and Hoeffler (1998) obtain a similar result by considering the square of ethnic fractionalization.

[28] In all model specifications we correct for error correlation over time for a given country by calculating cluster - robust standard errors. For the sake of space we do not show time controls variables and regional dummies in the regression results tables.

Table 1: *Logit Model - Ethnic Conflict Onset. Models with* P^n_∞ *for different* α.

Variable	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
C	-8.888***	-9.380***	-9.795***	-9.554***	-9.652***	-9.676***	-9.687***	-9.706***	-9.807***
	(1.324)	(1.345)	(1.297)	(1.385)	(1.411)	(1.413)	(1.413)	(1.416)	(1.447)
GDP per capita	-0.090**	-0.088**	-0.081**	-0.087**	-0.086**	-0.086**	-0.087**	-0.087**	-0.087**
	(0.038)	(0.040)	(0.037)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Population size	0.219**	0.234**	0.228***	0.230**	0.227**	0.227**	0.227**	0.227**	0.223**
	(0.096)	(0.099)	(0.077)	(0.101)	(0.102)	(0.102)	(0.102)	(0.102)	(0.104)
Excluded population	0.397***	0.350***	0.318***	0.328***	0.320***	0.318***	0.316***	0.314***	0.315***
	(0.122)	(0.118)	(0.100)	(0.117)	(0.117)	(0.116)	(0.116)	(0.116)	(0.114)
Imperial rule	-0.123	-0.137	-0.146	-0.183	-0.209	-0.204	-0.198	-0.193	-0.217
	(0.615)	(0.641)	(0.574)	(0.636)	(0.638)	(0.636)	(0.633)	(0.628)	(0.625)
Centre segmentation	0.149	0.125	0.134***	0.113	0.109	0.107	0.107	0.106	0.115
	(0.089)	(0.092)	(0.042)	(0.093)	(0.094)	(0.094)	(0.094)	(0.094)	(0.092)
Democracy	0.047	0.019	0.006	0.001	-0.005	-0.004	-0.003	-0.000	0.012
	(0.435)	(0.433)	(0.409)	(0.431)	(0.428)	(0.427)	(0.427)	(0.427)	(0.423)
Anocracy	0.220	0.193	0.282	0.181	0.180	0.180	0.180	0.179	0.189
	(0.238)	(0.239)	(0.230)	(0.241)	(0.242)	(0.243)	(0.243)	(0.244)	(0.246)
Oil production	0.011	0.009	0.009	0.008	0.008	0.008	0.009	0.009	0.009
	(0.026)	(0.029)	(0.027)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.032)
Mountains	0.125	0.116	0.110	0.118	0.121	0.123	0.124	0.124	0.130
	(0.085)	(0.090)	(0.089)	(0.094)	(0.095)	(0.095)	(0.095)	(0.095)	(0.098)
Regime change	0.156	0.147	0.166	0.149	0.153	0.156	0.158	0.160	0.165
	(0.262)	(0.262)	(0.237)	(0.261)	(0.260)	(0.260)	(0.260)	(0.260)	(0.257)
NC State	0.470	0.451	0.520	0.433	0.424	0.421	0.421	0.420	0.429
	(0.433)	(0.467)	(0.396)	(0.485)	(0.493)	(0.496)	(0.497)	(0.499)	(0.494)
New State	2.294***	2.239***	2.238***	2.238***	2.245***	2.251***	2.254***	2.257***	2.246***
	(0.686)	(0.685)	(0.635)	(0.679)	(0.676)	(0.676)	(0.676)	(0.676)	(0.668)
$p_{\alpha=0}$	-0.253								
	(0.783)								
$p_{\alpha=0.5}$		0.878							
		(0.698)							
$p_{\alpha=1}$			0.968						
			(0.587)						
$p_{\alpha=2}$				1.273					
				(0.689)					
$p_{\alpha=4}$					1.416**				
					(0.673)				
$p_{\alpha=7}$						1.436**			
						(0.652)			
$p_{\alpha=10}$							1.455**		
							(0.645)		
$p_{\alpha=15}$								1.491**	
								(0.645)	
$p_{\alpha\to\infty}$									1.639**
									(0.654)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6276	6276	6558	6276	6276	6276	6276	6276	6276
Pseudo R^2	0.143	0.144	0.150	0.147	0.148	0.149	0.149	0.149	0.151
Chi2	175.829***	186.177***	285.777***	189.565***	193.502***	196.215***	197.262***	197.889***	205.689***
BIC	1076.412	1074.682	1191.073	1072.357	1071.051	1070.563	1070.253	1069.839	1067.950
AIC	880.822	879.092	994.208	876.767	875.461	874.973	874.663	874.249	872.360

Notes: *The sample includes 140 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high intensity ethnic conflict. The method of estimation is Logit. Standard errors, clustered by country, in parentheses. Significance levels:* ** $p < 0.05$, *** $p < 0.01$.

R^2 is obtained for $\alpha \to \infty$.[29] The results remain robust even when we consider only countries with no more than 6 groups without relying on sample selection criteria described so far.

Table 2 reports the estimation results related to the entire set of countries with no group/country excluded from the analysis. Together with the RQ and the fractionalization index, we consider only the P index with $\alpha \to \infty$ since it yields the best fit with respect to any other value of α (Table 1). The estimated coefficients in Models 1-3 show that among the three distributional indices of ethnic diversity, only the P index with $\alpha \to \infty$ is significantly different from zero. The level of GDP per capita is negatively correlated with the probability of conflict outbreak while the size of the population and a dummy for the first two years of independence, are positively related to ethnic conflict onset. This is in line with Doyle and Sambanis (2000), Fearon and Laitin (2003), Collier and Hoeffler (2001, 2004), and Wimmer, Cederman and Min (2009), among others. In contrast to Fearon and Laitin's (2003) insurgency model, previous regime change, oil production per capita, and mountainous terrain receive limited support here. Although the coefficients associated to democracy and anocracy have the expected sign they do not reach a significance at the 0.05 level. 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. In line with Wimmer, Cederman and Min (2009), the degree of ethnic exclusion and centre segmentation (number of ethnic groups in power) are significant with the expected sign in all model specifications.

[29] 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^n_∞ index. We have also calculated the Somers' D statistic which provides an estimate of the rank correlation of the observed binary response variable (ethnic war onset) and the predicted probabilities. Since it can be used as an alternative indicator of model fit, we compared its value for the P index with different α. The results are in line with the previous conclusions based on the Pseudo R^2 and on the other informational criteria.

Table 2: *Logit Model - Ethnic Conflict Onset.*

Variable	Model1	Model2	Model3	Model4	Model5	Model6
C	-9.795***	-9.893***	-9.209***	-9.573***	-9.856***	-9.285***
	(1.297)	(1.323)	(1.317)	(1.288)	(1.322)	(1.292)
GDP per capita	-0.081**	-0.084**	-0.083**	-0.082**	-0.084**	-0.089**
	(0.037)	(0.037)	(0.034)	(0.035)	(0.037)	(0.036)
Pop. size	0.228***	0.207**	0.184**	0.205***	0.203**	0.159
	(0.077)	(0.081)	(0.079)	(0.077)	(0.081)	(0.086)
Excl. population	0.318***	0.307***	0.291***	0.293***	0.301***	0.316***
	(0.100)	(0.098)	(0.100)	(0.100)	(0.098)	(0.098)
Imperial rule	-0.146	-0.266	-0.116	-0.136	-0.257	-0.383
	(0.574)	(0.561)	(0.576)	(0.567)	(0.556)	(0.550)
Centre segm.	0.134***	0.148***	0.084	0.102**	0.139***	0.161***
	(0.042)	(0.042)	(0.047)	(0.049)	(0.052)	(0.043)
Democracy	0.006	0.034	0.073	0.044	0.042	0.081
	(0.409)	(0.406)	(0.401)	(0.402)	(0.402)	(0.405)
Anocracy	0.282	0.294	0.337	0.308	0.298	0.359
	(0.230)	(0.240)	(0.228)	(0.236)	(0.242)	(0.250)
Oil production	0.009	0.011	0.007	0.007	0.010	0.016
	(0.027)	(0.027)	(0.025)	(0.027)	(0.028)	(0.024)
Mountains	0.110	0.107	0.114	0.110	0.107	0.111
	(0.089)	(0.092)	(0.095)	(0.093)	(0.093)	(0.091)
Regime change	0.166	0.181	0.147	0.152	0.176	0.204
	(0.237)	(0.234)	(0.238)	(0.237)	(0.235)	(0.231)
NC State	0.520	0.585	0.605	0.567	0.592	0.704
	(0.396)	(0.388)	(0.388)	(0.410)	(0.393)	(0.387)
New State	2.238***	2.226***	2.234***	2.226***	2.222***	2.243***
	(0.635)	(0.630)	(0.626)	(0.630)	(0.628)	(0.629)
$p_{\alpha=1}$	0.968			0.617		-2.527
	(0.587)			(0.737)		(1.538)
$p_{\alpha\to\infty}$		1.359***			1.287**	3.204***
		(0.502)			(0.625)	(1.176)
$frac$			1.081	0.707	0.186	
			(0.566)	(0.703)	(0.752)	
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6558	6558	6558	6558	6558	6558
Pseudo R^2	0.150	0.154	0.150	0.151	0.154	0.157
Chi2	285.777***	251.707***	242.656***	261.443***	246.780***	251.617***
BIC	1191.073	1186.354	1191.045	1199.072	1195.089	1191.680
AIC	994.208	989.489	994.180	995.419	991.436	988.027

Notes: *The sample includes 146 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high intensity ethnic conflict. The method of estimation is Logit. Standard errors, clustered by country, in parentheses. Significance levels:* ** $p < 0.05$, *** $p < 0.01$.

Finally, in Models 5 and 6 we check the relative strength of P_∞^n versus RQ and fractionalization. The coefficients on ethnic polarization and fractionalization are not significantly different from zero in combination with P_∞^n which remains positive and statistically significant. More interestingly, since the coefficient on P_∞^n and the goodness of fit of the model that includes both P_∞^n and fractionalization are very similar to those in Model 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_∞^n index is able to "extract" the impact of the interaction between groups on the probability of conflict outbreak. The features of ethnic distribution as measured with the P_∞^n index are, hence, an important correlate of ethnic conflict outbreak, even after controlling for several economic, structural and geographical characteristic, as well as for political exclusion and competition along ethnic lines.

Since the P_{∞}^{n} index combines interaction and dominance, Table 3 checks its relative strength with respect to several dominance dummy variables commonly used in the empirical literature together with the fractionalization index.

Table 3: *Logit Model - Ethnic Conflict Onset.*

Variable	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8
C	-9.932***	-10.172***	-10.185***	-10.191***	-9.563***	-10.185***	-10.335***	-9.939***
	(1.304)	(1.317)	(1.321)	(1.349)	(1.332)	(1.340)	(1.391)	(1.443)
GDP per capita	-0.087**	-0.087**	-0.088**	-0.088**	-0.080**	-0.080**	-0.083**	-0.081**
	(0.040)	(0.041)	(0.040)	(0.039)	(0.037)	(0.039)	(0.039)	(0.037)
Pop. size	0.224***	0.231***	0.217***	0.187**	0.219***	0.238***	0.219***	0.183**
	(0.084)	(0.083)	(0.084)	(0.083)	(0.082)	(0.082)	(0.084)	(0.083)
Excl. population	0.353***	0.342***	0.329***	0.287***	0.345***	0.315***	0.304***	0.246**
	(0.101)	(0.103)	(0.100)	(0.100)	(0.101)	(0.104)	(0.102)	(0.110)
Imperial rule	-0.398	-0.406	-0.438	-0.469	-0.160	-0.219	-0.339	-0.244
	(0.565)	(0.559)	(0.551)	(0.529)	(0.624)	(0.616)	(0.600)	(0.645)
Centre segm.	0.189***	0.186***	0.184***	0.131***	0.150***	0.151***	0.165***	0.076
	(0.047)	(0.047)	(0.046)	(0.048)	(0.045)	(0.045)	(0.044)	(0.053)
Democracy	-0.037	-0.039	-0.014	0.028	-0.037	-0.058	-0.034	-0.014
	(0.426)	(0.423)	(0.420)	(0.415)	(0.404)	(0.401)	(0.397)	(0.389)
Anocracy	0.290	0.275	0.281	0.313	0.302	0.263	0.273	0.320
	(0.246)	(0.246)	(0.250)	(0.249)	(0.229)	(0.232)	(0.242)	(0.228)
Oil production	0.012	0.011	0.012	0.005	0.014	0.012	0.014	0.009
	(0.029)	(0.030)	(0.030)	(0.032)	(0.023)	(0.027)	(0.027)	(0.026)
Mountains	0.101	0.101	0.102	0.102	0.116	0.110	0.109	0.110
	(0.084)	(0.085)	(0.088)	(0.096)	(0.083)	(0.086)	(0.089)	(0.097)
Regime change	0.210	0.203	0.203	0.182	0.179	0.170	0.182	0.132
	(0.238)	(0.237)	(0.235)	(0.234)	(0.237)	(0.236)	(0.233)	(0.235)
NC State	0.689*	0.673*	0.683*	0.816**	0.570	0.558	0.622	0.718*
	(0.400)	(0.407)	(0.396)	(0.405)	(0.386)	(0.402)	(0.388)	(0.400)
New State	2.269***	2.255***	2.243***	2.218***	2.279***	2.244***	2.236***	2.214***
	(0.622)	(0.625)	(0.623)	(0.615)	(0.626)	(0.628)	(0.623)	(0.611)
$d4590$	0.766***	0.703***	0.567**	0.933***				
	(0.251)	(0.257)	(0.263)	(0.281)				
$p_{\alpha=1}$		0.474				1.025		
		(0.622)				(0.628)		
$p_{\alpha\to\infty}$			0.849*				1.408***	
			(0.507)				(0.540)	
$frac$				1.649***				2.015**
				(0.618)				(0.873)
$d6090$					0.338	0.354	0.358	0.699**
					(0.234)	(0.238)	(0.241)	(0.343)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations	6558	6558	6558	6558	6558	6558	6558	6558
Pseudo R^2	0.155	0.156	0.158	0.161	0.149	0.152	0.156	0.156
Chi2	318.356***	305.815***	270.394***	228.171***	318.722***	290.512***	246.723***	212.456***
BIC	1184.862	1193.092	1191.339	1188.001	1191.767	1197.621	1192.859	1193.589
AIC	987.997	989.438	987.686	984.348	994.903	993.968	989.205	989.935

Notes: *The sample includes 146 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high intensity ethnic conflict. The method of estimation is Logit. Standard errors, clustered by country, in parentheses. Significance levels:* * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Model 1 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 the expected sign. The RQ polarization index is not significant when we control for dominance (Model 2). This evidence is in line with Collier and Hoeffler (2004). Only the P_{∞}^{n} index and the fractionalization index remain significant in combination with dominance (at the 0.09 and 0.01 level respectively). 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%) (Models 5 - 8). When combined with a "pure" dominance (*i.e.*, defined as 1 if the relative size of the biggest group in the population is larger than 50%), the P_{∞}^{n} and the fractionalization index result significant at the 0.01 level.[30]

In addition to the Collier and Hoeffler's and the Schneider and Wiesehomeier's ethnic dominance dummies, Montalvo and Reynal-Querol (2005) propose another indicator of conflict potential, namely the size of the largest ethnic minority. From Table 4 we see that the coefficient on this variable is not significantly different from zero in any model specification, while the P_{∞}^{n} index remains highly significant even in the presence of this variable.

[30]The estimation of this latter model is available upon request.

Table 4: *Logit Model - Ethnic Conflict Onset.*

Variable	Model1	Model2	Model3	Model4
C	-9.176***	-9.831***	-9.580***	-8.851***
	(1.333)	(1.367)	(1.343)	(1.337)
GDP per capita	-0.082**	-0.081**	-0.086**	-0.086**
	(0.036)	(0.038)	(0.037)	(0.034)
Pop. size	0.200**	0.189**	0.166	0.150
	(0.083)	(0.084)	(0.086)	(0.086)
Excl. population	0.346***	0.321***	0.319***	0.297***
	(0.103)	(0.102)	(0.100)	(0.100)
Imperial rule	-0.018	-0.273	-0.341	-0.069
	(0.585)	(0.641)	(0.598)	(0.610)
Centre segm.	0.133***	0.157***	0.168***	0.086
	(0.043)	(0.046)	(0.046)	(0.047)
Democracy	0.031	-0.009	0.042	0.097
	(0.414)	(0.396)	(0.399)	(0.407)
Anocracy	0.328	0.293	0.330	0.370
	(0.227)	(0.239)	(0.249)	(0.234)
Oil production	0.012	0.009	0.013	0.008
	(0.023)	(0.027)	(0.025)	(0.023)
Mountains	0.123	0.127	0.121	0.125
	(0.085)	(0.090)	(0.091)	(0.095)
Regime change	0.166	0.179	0.190	0.142
	(0.239)	(0.232)	(0.231)	(0.238)
NC State	0.466	0.587	0.643	0.588
	(0.391)	(0.391)	(0.389)	(0.402)
New State	2.238***	2.171***	2.191***	2.192***
	(0.630)	(0.613)	(0.617)	(0.613)
largestmin	0.471	-3.293	-1.815	-0.669
	(1.318)	(2.322)	(1.667)	(1.735)
$p_{\alpha=1}$		2.456**		
		(1.080)		
$p_{\alpha\to\infty}$			1.777***	
			(0.589)	
frac				1.292
				(0.723)
Time Controls	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes
N. Observations	6535	6535	6535	6535
Pseudo R^2	0.148	0.153	0.156	0.151
Chi2	320.025***	242.438***	235.427***	242.209***
BIC	1192.042	1195.139	1191.869	1197.809
AIC	995.279	991.591	988.321	994.261

Notes: *The sample includes 146 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high intensity ethnic conflict. The method of estimation is Logit. Standard errors, clustered by country, in parentheses. Significance levels:* ** $p < 0.05$, *** $p < 0.01$.

Finally, Table 5 considers intermediate intensity conflicts only (Models 1-3) and the P^n_∞ index calculated by using the Fearon's (2003) classification of ethnic groups (Models 4-5).[31] The P^n_∞ index remains highly significant in the model for intermediate intensity conflicts, as well as for Fearon's (2003) ethnic grouping.

[31]Models 4 and 5 do not consider the ethnic politics variable since the share of the excluded population and the number of included ethnic groups are defined over the EPR grouping.

Table 5: *Logit Model - Ethnic Conflict Onset: Low Intensity Conflicts only (1-3) and Fearon (2003) grouping (4-5).*

Variable	Model1 EPR3 Intermed.	Model2 EPR3 Intermed.	Model3 EPR3 Intermed.	Model4 Fgroup All	Model5 Fgroup Intermed.
C	-11.796***	-11.927***	-10.631***	-10.615***	-12.676***
	(1.838)	(1.952)	(1.797)	(1.474)	(1.924)
GDP per capita	-0.067	-0.070	-0.067	-0.079**	-0.068
	(0.042)	(0.043)	(0.040)	(0.037)	(0.038)
Population size	0.227	0.197	0.179	0.340***	0.340***
	(0.117)	(0.124)	(0.116)	(0.081)	(0.109)
Excl. population	0.212	0.197	0.237**		
	(0.113)	(0.110)	(0.120)		
Imperial rule	-0.169	-0.354	-0.023		
	(0.826)	(0.796)	(0.858)		
Centre segm.	0.125***	0.148***	0.099		
	(0.043)	(0.048)	(0.054)		
Democracy	0.215	0.263	0.277	0.017	0.263
	(0.438)	(0.452)	(0.438)	(0.315)	(0.347)
Anocracy	0.121	0.158	0.205	0.401	0.262
	(0.371)	(0.373)	(0.374)	(0.219)	(0.353)
Oil production	-0.009	-0.005	-0.003	0.008	0.003
	(0.039)	(0.041)	(0.033)	(0.027)	(0.032)
Mountains	-0.044	-0.048	-0.039	0.169**	0.015
	(0.114)	(0.119)	(0.114)	(0.084)	(0.113)
Regime change	-0.107	-0.101	-0.100	0.104	-0.171
	(0.383)	(0.373)	(0.387)	(0.228)	(0.386)
NC State	0.818	0.916	0.856	0.376	0.599
	(0.587)	(0.612)	(0.554)	(0.444)	(0.670)
New State	2.981***	2.945***	2.986***	2.129***	2.934***
	(0.827)	(0.793)	(0.813)	(0.637)	(0.825)
$p_{\alpha=1}$	1.860**				
	(0.779)				
$p_{\alpha\to\infty}$		2.297***			
		(0.777)			
$frac$			0.489		
			(0.802)		
$p_{\alpha\to\infty}$ (F group)				1.397***	1.974***
				(0.500)	(0.703)
Time Controls	Yes	Yes	Yes	Yes	Yes
Reg. Dummies	Yes	Yes	Yes	Yes	Yes
N. Observations	6558	6558	6558	6558	6558
Pseudo R^2	0.147	0.156	0.139	0.140	0.143
Chi2	332.772***	271.332***	324.517***	203.694***	153.005***
BIC	804.884	799.387	809.958	1175.364	781.331
AIC	608.019	602.522	613.094	998.864	604.831

Notes: *The sample includes 146 countries for the period 1946-2005. The dependent variable is the onset of the intermediate and high intensity ethnic conflict. The method of estimation is Logit. Standard errors, clustered by country, in parentheses. Significance levels:* ** $p < 0.05$, *** $p < 0.01$.

Given the fact that ethnic conflict is a rare event and that the standard logistic regression can underestimate the probability of such events, we also perform a rare event logit estimation (King and Zeng, 2001). The results are similar to those obtained by using the traditional logistic model. Collier and Hoeffler (2004), Sambanis (2001) and Wimmer, Cederman and Min (2009) report similar findings. In addition, we also check whether the results are driven by particular geographical regions that might be considered more or less conflict prone by eliminating one region at a time in our baseline regression models. The results do not change significantly and the relevance of the P^n_∞ index is unaltered. In addition to the clustering on country, we also control for the non-independence of observation over countries and over time. We do not find any substantial difference in the results. The test of the correlation coefficient 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 logistic estimation method. For the sake of space we do not report the estimation coefficients from these additional robustness checks.[32]

4 Concluding Remarks

In this paper we show how the relative importance of fractionalization, polarization and dominance in the determination of conflict potential may depend on the characteristics of the underlying population distribution across groups. We axiomatically derive a parameterized class of indices of conflict potential that combines the groups power and between-groups interaction. Conflict potential is obtained 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 effects similar to the presence of an extreme form of 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 main driver of conflict is fractionalization while for the intermediate case, what matters is the combination between the two. It is not important how large a group is but rather how *decisive* it can be in a hypothetical competition between all the groups in the population. 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 measures could differ from the existing diversity indices. We 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 extreme member of our class of

[32] All the additional regression tables are available upon request.

indices, the P_∞^n index, 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] A. Alesina, A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg. Fractionalization. *Journal of Economic Growth*, 8(2):155–194, 2003.

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

[3] F. Caselli and W. J. Coleman, II. On the theory of ethnic conflict. *Journal of the European Economic Association*, 11:161–192, 2013.

[4] L.E. Cederman and L. Girardin. Beyond fractionalization: Mapping ethnicity onto nationalist insurgencies. *American Political Science Review*, 101:173–185, 2007.

[5] L.E. Cederman, A. Wimmer, and B. Min. Why do ethnic groups rebel? new data and analysis. *World Politics*, 62(1):87–119, 2010.

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

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

[8] P. Collier and A. Hoeffler. Greed and grievance in civil war. *Oxford Economic Papers*, 56(4):563–595, 2004.

[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] M.W. Doyle and N. Sambanis. International peacebuilding: A theoretical and quantitative analysis. *American Political Science Review*, 94(4):779 – 801, 2000.

[11] W. Eichhorn. *Functional Equations in Economics*. Applied mathematics and computation. Addison-Wesley Publishing Company, Advanced Book Program, 1978.

[12] J. Esteban. Olson vs. coase: Coalitional worth in conflict. *Theory and Decision*, 55(4):339–357, 2003.

[13] J. Esteban, L. Mayoral, and D. Ray. Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4):1310–1342, 2012.

[14] J. Esteban and D. Ray. On the measurement of polarization. *Econometrica*, 62(4):819–851, 1994.

[15] J. Esteban and D. Ray. Conflict and distribution. *Journal of Economic Theory*, 87(2):379–415, 1999.

[16] J. Esteban and D. Ray. Polarization, fractionalization and conflict. *Journal of Peace Research*, 45(2):163–182, 2008.

[17] J. Esteban and D. Ray. Linking conflict to inequality and polarization. *American Economic Review*, 101(4):1345–74, 2011.

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

[19] J. Fearon, K. Kasara, and D. Laitin. Ethnic minority rule and civil war onset. *American Political Science Review*, 101:187–193, 2007.

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

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

[22] D.S. Felsenthal and M. Machover. *The measurement of voting power: theory and practice, problems and paradoxes.* Edward Elgar, 1998.

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

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

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

[26] G. King and L. Zeng. Logistic regression in rare events data. *Political Analysis*, 9:137–163, 2001.

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

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

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

[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*, 45(3):259–282, 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, 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, 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 1.1

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

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

We first investigate the implications arising from this axiom combined with all the other axioms, when $r(\rho), r(\rho^{'}) \neq 0$, 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 *non-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, holds also for all the other values $\rho' \neq \rho$ in the interval. As a result either $r(\rho)$ is constant and different from 0 for all $\rho \leq 1$ or it is strictly increasing, that is $r(\rho) < r(\rho')$ if $\rho < \rho'$. We focus first on the latter case.

If $r(\rho), r(\rho^{'}) \neq 0$ then assume that 2GRPH holds. Let $\rho_0 := \lambda\rho \in (0, 1)$, that is $\lambda = \rho_0/\rho$. It follows that $r(\lambda\rho) = r(\rho_0)$ and $r(\lambda\rho') = r(\rho^{'} \cdot \rho_0/\rho)$, thus 2GRPH requires 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 $g(\cdot)$. Note that if we set $\lambda < 1$ (we will discuss the implication of $\lambda > 1$ afterwards) then $\rho_0/\rho < 1$, it then follows that $r(\rho_0) = g(\rho_0/\rho) \cdot r(\rho)$ for all $\rho_0, \rho < 1$ and

$\rho_0/\rho < 1$. This functional equation therefore holds also if we swap ρ_0/ρ with ρ on its r.h.s., 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 holds 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 $r(\rho_0) > 0$ and $r(\rho_0/\rho) > 0$, 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 obtained equation 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 leads to:

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

Moreover, the case analysed earlier where $r(\rho) > 0$ 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 $\rho \in (0, 1)$ where $\beta = 1/k$. Note that this solution implies that 2GRPH holds also for all $\lambda > 1$. Moreover, according to 2GRPH the functional equation should hold also for $\rho = 1$ and therefore its solution extends from $\rho \in (0, 1)$ to $\rho \in (0, 1]$.

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 where 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 extending the admissible values for β considering also $\beta = 0$ in (A.2).

We now move to consider the implications of the remaining axioms. Consider first the solution (A.2) for $r(\rho) > 0$ extended to hold for $\rho \in (0, 1]$. Recall that by Normalization $\phi^2(1 - \pi) = 1 - \phi^2(\pi)$. Then by definition:

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

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

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

for all $\alpha \geq 0$, $\beta > 0$, where $\pi \leq 1/2$. 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 $\phi^2(1/2) = \frac{\beta}{\beta+1} = 1/2$ requires that $\beta = 1$, which gives the desired result for ϕ^2_α in Lemma 1.1. The values for $\phi^2(\pi)$ for $\pi > 1/2$ are obtained by setting $\phi^2(\pi) = 1 - \phi^2(1 - \pi)$ where $1 - \pi < 1/2$. To obtain ϕ^2_∞ consider the solution for $r(\rho) = 0$ for all $\rho \in (0, 1)$, that is $\phi^2(\pi) = 0$ for $\pi < 1/2$. Combine this solution with the condition $\phi^2(1/2) = 1/2$ and derive $\phi^2(\pi)$ for $\pi > 1/2$ by setting $\phi^2(\pi) = 1 - \phi^2(1 - \pi)$ where $1 - \pi < 1/2$, that is $\phi^2(\pi) = 1$ for $\pi > 1/2$.$\blacksquare$

Proof of Theorem 1.2

Sufficiency part. Note that the obtained specification for ϕ^n_α satisfies the axioms considered.

Necessity part. Consider axiom REP. We first check the restrictions that make it consistent with the specification of ϕ^2 obtained in Lemma 1.1 applying 2GRPH, we will then extend the analysis to the case where $n > 2$.

For $n = 2$, the axiom REP requires that $\frac{\phi^2(\pi)}{\phi^2(1-\pi)} = \frac{g(M_i)}{g(M_j)}$ for some function $g : \Re_+ \to \Re_+$ if $\phi^2(1-\pi) > 0$, where M_i is associated to the group with the population share π and M_j is associated to the other group with share $1-\pi$. Note that by definition $M_i = 1-\phi^2(1-\pi)+\phi^2(\pi)$, 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 $\phi^2(\pi) > 0$, that

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

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 $\phi^2 \leq 1/2$, with $h(1/2) = 2f(1/2)$, for some function $h : (0, 1] \to \Re_+$, and $\frac{f(\phi^2)}{\phi^2} = h(1-\phi^2)$ for $\phi^2 > 1/2$.

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

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

Note that the domain of $g(\cdot)$ is $\Re_+$ while the above condition provides a restriction only for the domain interval $(0, 2)$. More generally $g(\cdot)$ may depend on the distribution $\Pi_{-i,-j}$ of all the population groups except i and j and thus it can be written as related to a function $h(\cdot)$ that could depend on $\Pi_{-i,-j}$ if $n > 2$. We denote such function as $h_\Pi(\cdot)$.

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 $M > 1$.

It follows that for the case where $M \in (0, 2)$ with $M_j > 1$ and $M_i < 1$, according to REP it holds $\frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)} = \frac{g(M_i)}{g(M_j)} = \frac{h_\Pi(M_i/2)\cdot M_i}{h_\Pi(2-M_j/2)\cdot M_j}$.

By applying nGRPI one obtains also 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 depend on Π.

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

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, note that this is not a general condition that holds for all $\frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)}$, in fact this holds only if $M_j, M_i \in (0,2)$. That is, this is the case whenever π_i, π_j are sufficiently small.

However, the derived proportionality of $\frac{\phi^n(\pi_i,\Pi)}{\phi^n(\pi_j,\Pi)}$ holds 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, $i, j \in N$, all $n \geq 2$, and all Π (that are consistent with π_i, π_j).

The specification of the two axioms, nGRPI 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 $\phi^n(\pi_i,\Pi) = M_i \cdot w(M)$ where M denotes the distribution of all aggregated marginal contributions of each group, and $w(\cdot)$ is a generic function, identical for all groups (that may also depend on Π).

If this is the case then, by Normalization, $\sum_j \phi^n(\pi_j,\Pi) = \sum_j M_j \cdot w(M) = w(M) \cdot \sum_j M_j = 1$ thus, $w(M) = 1/\sum_j M_j$, thereby leading to:

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

where the M_i components are obtained making use of the function ϕ^2 in Lemma 1.1.

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 a group j whose ϕ^n is 0 irrespective of the value of M_j. Note however that M_j is non-decreasing w.r.t. π_j, and thus by Monotonicity we should have that ϕ^n is 0 also for all groups i whose size is below π_j or whose M_i is lower than M_j. But according to REP what is relevant is the ratio $\frac{M_i}{M_j}$ so, taking two groups one of which has $M_i > 0$ but $\phi_i^n = 0$ with $\frac{M_i}{M_j} \neq 0$ and $\phi_j^n > 0$, one can consider groups distribution such that $\frac{M_i}{M_j}$ is appropriately set at a desired positive value and therefore for all pairs i, j then $\frac{M_i^{'}}{M_j^{'}} < \frac{M_i}{M_j} \Rightarrow \frac{\phi^n(\pi_i^{'}, \Pi')}{\phi^n(\pi_j^{'}, \Pi')} \leq \frac{\phi^n(\pi_i, \Pi)}{\phi^n(\pi_j, \Pi)} = 0$, 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 then follows that $\phi_i^n = 0$ only if $M_i = 0$, making the result consistent with (A.4). ■

Appendix II

EXPLANATORY AND CONTROL VARIABLES

GDP per Capita The GDP per Capita data come 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 kilometres 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 characterise 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 centre segmentation*.

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

Largest Minority The relative size of the largest minority in the population. The data come from EPR3.

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.