Patterns of Ethnic Group Segregation and Civil Conflict*

Janina Matuszeski† Harvard University

Frank Schneider‡ Harvard University

First draft: September 28, 2006
This draft: November 15, 2006
(Job Market Paper - Matuszeski)

Abstract

We present a new index of ethnic geography, the Ethnic Diversity and Clustering (EDC) index, which measures the clustering of ethnic groups within a country, as well as the overall ethnic diversity of the country. Using digital map data for over 7000 linguistic groups around the world, we construct the EDC index for 189 countries. We also calculate the traditional Ethno-Linguistic Fractionalization (ELF) index of ethnic diversity for 189 countries, including 186 countries for which we also have the EDC index. In cross-country regressions, our EDC and ELF indices are significantly correlated with measures of civil war, including the number of conflicts, total time spent in war, and total combatant deaths. Evidence from regressions using both indices indicates that civil war is more frequent and severe in countries where citizens of a given ethnic group tend to be more clustered together. Results for the average duration of conflicts are weaker for both indices. In addition, higher levels of ethnic diversity and clustering are associated with an increased incidence of civil conflict for countries with the straighter borders typical of artificial states, but not for other countries. Our results are robust to the inclusion of controls for former colonial status, continent, and climate.

JEL Classification: J15, F50, F51
Keywords: Civil War, Ethnicity, Fractionalization, Diversity, Segregation

*The authors would like to thank Alberto Alesina, Aron Balas, Robert Barro, Albert Bravo-Biosca, Jeffrey Friedman, Guoping Huang, Michael Kremer, Michael Oltmans, Phillip Schnabl and participants at the Harvard Economics Department Macroeconomics and Development Workshops for useful comments and suggestions. We thank Global Mapping International for granting us access to their World Language Mapping System database. All remaining errors are our own. This paper is accessible from `http://www.people.fas.harvard.edu/ matuszes/papers.html`.

†**Corresponding Author:** Janina Matuszeski. National Bureau of Economic Research (NBER), 1050 Massachusetts Avenue, Cambridge MA 02138. matuszes@fas.harvard.edu

‡Frank Schneider. National Bureau of Economic Research (NBER), 1050 Massachusetts Avenue, Cambridge MA 02138. schneid2@fas.harvard.edu

1 Introduction

From the wars in Yugoslavia in the 1990's to the current strife in Iraq, Sudan and Russia, divisions between ethnic groups are frequently seen as a basis for civil wars, whether by sparking conflict or prolonging it. However, despite extensive anecdotal evidence, empirical analysis of the impact of ethnic diversity on civil conflict has been largely inconclusive. In order to better understand the connections between ethnicity and conflict, it is important to consider not just the ethnic diversity of the country as a whole, but also the pattern of the distribution of ethnic groups within a country. This research takes a geographic approach to this question, using digital map data to construct several new indices of ethnic geography. Importantly, by using map-based data, we are able to measure the pattern of the distribution of ethnic groups within a country's borders.

We design and construct a new index, the Ethnic Diversity and Clustering (EDC) index, which measures ethnic segregation in the form of clustering, as well as the overall ethnic diversity of a country. Specifically, the measure reflects the degree to which citizens belonging to each ethnic group are clustered together or dispersed throughout the country.[1] In addition to this index, we consider ethnic diversity and ethnic clustering separately. We construct a new version of the widely-used Ethno-Linguistic Fractionalization (ELF) index as a measure of ethnic diversity. We also present preliminary evidence on a third index, the Ethnic Clustering (EC) index, which measures only the clustering of ethnic groups.[2] We calculate the ELF, EDC and EC indices for 182 countries. In cross-country regressions, we find all three indices to be significantly correlated with several measures of civil conflict.

Previous work on the relationship between ethno-linguistic fractionalization[3] and civil conflict has led to no consensus as to even the direction of the effect of ethnic diversity on civil conflict. Researchers have also considered ethnic dominance, which is defined as either one large ethnic group or two to three main ethnic groups.[4] Studies which have found no effect of ethnic fractionalization

[1] An example of a country with highly-clustered ethnic groups is Belgium, with Flemish speakers almost exclusively in the north of the country and French speakers almost exclusively in the south. By contrast, Senegal has more interspersion of its ethnic groups, with Woluf speakers and other ethnic group members located in diverse regions of the country.

[2] This index corresponds to the "H" index detailed in Reardon and O'Sullivan (2004). The EC index has the advantage in that it is a pure measure of clustering, whereas our EDC index reflects ethnic diversity as well as clustering. However, the EDC measure has a useful intuitive interpretation as the ethnic diversity experienced by the average citizen, and is also more robust than the EC index for ethnically homogeneous countries. The construction of all of our indices is described further in Sections 2 and 3.

[3] The literature has focused mainly on fractionalization and not on segregation.

[4] With regard to the relationship between ethnic dominance and ethnic fractionalization, one main ethnic group corresponds to a high value of ELF, close to 1; whereas two to three main groups corresponds to values of ELF closer to 0.5, which is in the middle of the range of the index. If having two to three main ethnic groups in a country is significantly correlated with civil war, then the relationship between ELF and conflict will be non-monotonic, with

on the onset of civil wars include Fearon and Laitin (2003); Lujala, Gleditsch and Gilmore (2005), who focus on the impact of the availability of natural resources, especially diamonds; and Miguel, Satyanath and Sergenti (2004) who estimate the effect of GDP growth on civil war in African countries, instrumenting for changes in GDP using rainfall data. Other research has found that higher ethnic fractionalization is associated with a *lower* incidence of civil war, including work by Collier (2001), Collier and Hoeffler (2004) and Fearon (2005). The first two of these studies also show that ethnic dominance increases the chance of civil war, while Fearon (2005) finds no effect of ethnic dominance. Collier and Hoeffler (2002) find no effect of ethnic dominance on civil war onset in Africa. Finally, Sambanis (2004) estimates twelve models using different inclusion criteria for civil wars from previous studies. He finds a significant positive effect of fractionalization on civil war onset in one model out of twelve, and marginal significance in two additional models, suggesting that using different criteria for including civil wars can affect the empirical results.

Other studies have focused on the duration of civil war. Lujala et al. (2005) find a positive effect of fractionalization on civil war duration. Collier and Hoeffler (1998) also find that fractionalization predicts longer civil wars, and that the effect is non-monotonic. Collier, Hoeffler and Söderbom (2004) show a non-monotonic relationship between ethnic fractionalization and civil war duration. They find that civil war duration is longest for middle levels of fractionalization (ELF=0.5), and is shorter for both very low and very high fractionalization. Finally, Fearon (2004) finds no effect of ethnic fractionalization on civil war duration.

There are several possible reasons for this lack of conclusive evidence concerning the effect of fractionalization on civil conflict. First, as outlined above, the pattern of ethnic groups within a country has received little attention to date. In our research, we examine one possible aspect of these patterns, namely the segregation or clustering of ethnic groups. Second, the literature has so far used a country-wide measure of ethnic diversity instead of focusing on the diversity in the conflict location. This is important if the area of a conflict is quite different from the country as a whole. While we do not address the issue in this paper, our map-based methodology can easily be extended to calculate sub-national indices, and we are currently constructing conflict-area indices as part of related research.

Third, with regard to measures of civil conflict, the effects of ethnic diversity have been shown to vary significantly depending on the choice of the left hand side variable, which might include the incidence, duration, or intensity of civil war. In our research, we examine all three of these aspects of civil war. Finally, results of previous studies depend crucially on the definition of what

middle values of ELF associated with the highest levels of conflict.

constitutes a civil war. Numerous databases have been used in the literature and most authors have used their own inclusion criteria.[5] Our aim is not to devise a "correct" or "new" definition of civil war. Instead we base our analysis on the UPPSALA/PRIO version of the Correlates of War (COW) database, a widely used source. We use the Major Episodes of Political Violence (MEPV) database as a robustness check, and also apply several additional inclusion criteria for civil wars, which are typical of the other principle studies in this area.

Our research also relates to the question of the proper measurement of segregation. Massey and Denton (1988) propose five dimensions along which segregation can be measured: evenness, exposure, concentration, centralization, and clustering.[6] Our analysis focuses on the clustering aspect of segregation. To date, the vast majority of work on segregation indices has been based on the premise that data is available for certain distinct sub-areas such as census block groups, but not at the level of an individual person or location. However, any such segregation index is highly dependent on the precise designation of the sub-areas.[7] Problems with this approach include the fact that sub-areas are most often designed to facilitate data collection, not the accuracy of segregation measures. Sub-areas may also completely ignore social and cultural geography, introducing noise to the measure. Or, sub-areas may explicitly group similar populations together, causing the measure to be biased.

Several recent papers describe new indices which can be constructed using individual level data, avoiding these problems with sub-area definitions. Reardon and O'Sullivan (2004) describe the construction of several measures of two aspects of segregation: clustering/evenness and exposure/isolation. Also, work by Echenique and Fryer (2006) uses information on the interactions between individuals to create a new, spatially-based index of segregation, the Spectral Segregation Index. Our research is in this spirit, as we use data on many individual locations to construct our indices. In addition, the index which we refer to as the EC or Ethnic Clustering index corresponds to the "H" measure of clustering from Reardon and O'Sullivan (2004). To the best of our knowledge, this index has not been used to measure clustering of ethnic groups in countries around the world.[8]

[5]The most widely known database is the Correlates of War (COW) database. A popular version of that database in the UPPSALA/PRIO version. For a detailed review of approximately 60 databases on civil war we refer to Eck (2005). Sambanis (2004) offers an overview of the inclusion criteria of various authors.

[6]More recent work has argued that several of these dimensions should be combined, for example evenness and clustering can be seen as opposite ends of the same spectrum. (Reardon and Firebaugh (2002)).

[7]See Reardon and O'Sullivan (2004) for a discussion of these critiques.

[8]By contrast, while we could theoretically construct the Spectral Segregation Index in Echenique and Fryer (2006), in practice the task would be too large from a computational stand-point since it is based on interactions among all individual data points.

Ethnicity can be described by several different dimensions including language, religion, cultural traditions, and visual characteristics. In our research we focus on the linguistic aspect of ethnicity, specifically the primary language spoken, which we consider to be one of the most important factors. Previous research has used tabulated statistical data to explore other aspects of ethnicity. For example, Alesina, Devleeschauwer, Easterly, Kurlat and Wacziarg (2003) consider religion and visual racial characteristics, in addition to language differences. Depending on the availability of digital map data, we hope to explore some of these additional aspects of ethnicity in the future.

In order to calculate the ELF, EDC and EC indices, we require information on the location of each language group, and also population density data. For the linguistics information, we rely on a proprietary dataset from Global Mapping International showing the location of over 7000 language groups around the world.[9] We also use digital map data on population density from Columbia University's Gridded Population of the World.[10] Both data sets are based on data for the early 1990's. We calculate the ELF, EDC and EC measures for 189, 189 and 185 countries, respectively, including 182 countries for which we have all three indices.[11]

With regard to theoretical predictions for the relationship between ethnic geography measures and civil conflict, we draw on the existing literature to construct four hypotheses which are described in detail in Section 4. Several hypotheses rely on the idea that ethnicity can be used to coordinate or enforce the coalitions that engage in conflict. Our theories predict that higher levels of ethnic diversity and higher levels of ethnic clustering will both be associated with more civil conflicts. However predictions for the effect on the duration of the typical civil war are more ambiguous. We predict that higher ethnic diversity can lead to either shorter or longer conflicts, while higher ethnic clustering should be associated with shorter conflicts.

In Section 5, we show that both the ELF and EDC measures prove to be significantly correlated with the incidence of civil conflicts, the total years spent in civil war, and the total casualties. The results are less strong for the average duration and the average casualties for each conflict. Evidence for the EC index also shows it to be correlated with these civil war outcomes. However, for several reasons detailed in Sections 2 and 3, we consider the evidence using the EC index to be preliminary. Based on these findings, we find empirical support for three of our four hypotheses. However, in

[9] World Language Mapping System Version 3.2, from Global Mapping International (www.gmi.org). This data is based on the 15th Edition of the Ethnologue linguistics database.

[10] Gridded Population of the World database Version 3, from the Center for International Earth Science Information Network (CIESIN) Socio-Economic Data Center (SEDAC), Columbia University, New York; http://sedac.ciesin.columbia.edu/gpw/index.jsp.

[11] Technical difficulties, which are described further in Section 3.1 currently prevent our calculating the indices for a small subset of countries.

contradiction to our last hypothesis, we find that more ethnic clustering is associated with longer wars.

Our results are robust to excluding small conflicts; excluding small countries; using different criteria for civil war; and controlling for former colonial status, climate, and continent dummy variables. We do not currently control for GDP, although it is an important factor, because GDP is also affected by ethnic diversity and is thus endogenous. In future work, we plan to use a panel data format to accurately account for the effect of GDP. To address potential reverse causality from civil conflicts to ethnic group geography, we show that our results are largely robust to including only conflicts that began in the 1990's or later.[12]

Finally, we also consider the impact of measures of artificial states. Based on data from Alesina, Easterly and Matuszeski (2006) we construct a dummy variable for artificial states using the median of their "fractal" variable as a cutoff. We find evidence that more diversity and more clustering of ethnic groups are associated with more civil conflict for artificial states, but not for other countries.

2 Ethnic Geography Indices

We compute three separate indices in order to measure the two aspects of ethnic geography that we choose to examine, ethnic diversity and ethnic clustering. One index, the Ethno-Linguistic Fractionalization (ELF) index, measures only ethnic diversity. A second index, the Ethnic Diversity and Clustering (EDC) index measures both ethnic diversity and clustering, while a third index, the Ethnic Clustering (EC) index measures only ethnic clustering The ELF index constructed using earlier data from the 1960's[13] has been used extensively in the literature to date. The EDC index was designed by the authors and we know of no other instances of the use of this index. Finally, the EC index is based on the "H" index described by Reardon and O'Sullivan (2004). All three indices have a range of zero to one.

While both the EDC and EC indices measure clustering, each index has its advantages and disadvantages. As we show below, the EDC index has an intuitive interpretation as the average diversity in local areas across the country, while the EC index has no equivalent interpretation. Also, the EC index is not defined for perfectly homogenous countries (those with only one ethnic group) and, based on our data, cannot be reliably computed for countries that are mostly homo-

[12]Our ethnic geography data corresponds to the early 1990s time period.

[13]The most commonly used ELF index to date is based on ethnic groups described in the *Atlas Narodov Mira [Atlas Of The Peoples Of The World]* (1964).

geneous (with one big ethnic group and a few very small ethnic groups). By contrast, the EDC index is computable for all countries, both in theory and in practice. On the other hand, the EC index has the obvious advantage of measuring only clustering while the EDC index is measuring diversity as well as clustering. Because of difficulties interpreting the EC index for very or perfectly homogeneous countries, we present only preliminary results for this index in Section 5, and we rely on the EDC index for our main conclusions regarding ethnic clustering. Reassuringly, the EDC and EC indices produce similar results regarding the relationship between clustering and civil conflict.

A final difference among the three indices is that the EDC and EC indices are constructed using measures of ethnic diversity in each local area of the country, while the ELF index only considers the country-wide populations of each ethnic group. Because of the local component of the EDC and EC indices, use of digital map data greatly aids in the construction of these indices.

We now proceed to describe each index in detail. In the formulas below, l indexes the languages within a country, L is the total number of languages in the country, and p indexes points in an even grid across the country, each approximately 1.5 kilometers apart. Finally, "$\widetilde{\ }_p$" designates a variable that relates to the local area around point p, for example the area within 50 kilometers of point p. Thus, we define:

n_{lp} Population of language l at point p

$\widetilde{n}_{lp}$ Population of language l in region of point p

$\widetilde{N}_p$ Total population in region of point p

n_l Total population of language l in country

N Total population of country

The ELF index is described in previous literature and is calculated based on the population of each ethnic group in the country as a whole. It is constructed using a Herfindahl index of the shares of each ethnic group in the total population $\left(\frac{n_l}{N_l}\right)$. The ELF index is equal to the probability that two citizens picked at random from the country's population will be from different ethnic groups.

$$ELF = 1 - \sum_{l=1}^{L} \left(\frac{n_l}{N}\right)^2 \tag{1}$$

We also construct our new Ethnic Diversity and Clustering (EDC) index which is related to the

ELF index in that it uses the formula for the ELF index to calculate a measure of ethnic diversity in a local region. We then average this local elf index over the entire country to get the EDC index.

In general, as in Reardon and O'Sullivan (2004), we can define the population of a given ethnic group in a local area using the population at nearby points (n_q) and a weighting proximity factor ($\phi(p,q)$) which gives greater weight to a point q if it is close to point p, and less weight for a point that's far from point p. The general formula is given by:

$$\widetilde{n}_{lp} = \frac{\int \phi(p,q) n_{ql} dq}{\int \phi(p,q) dq} \tag{2}$$

For the EDC index, we use a weighting factor which is 1 inside of a circle of 50 kilometers and zero outside. Thus our formula for the number of people speaking a given language l in a local region is simply the number of people within a 50 kilometer radius that speak that language. The formula is given below, where R_p is a region of radius 50 kilometers around point p and A is normalized to one.[14]

$$\widetilde{n}_{lp} = \frac{\int\limits_{q \in R_p} n_{lq} dq}{A} \tag{3}$$

We consider a series of points p, which are spaced in a grid across each country. The points are approximately 1.5 kilometers from each other, and we calculate $\widetilde{n}_{lp}$ at each of these points. Given the definition for $\widetilde{n}_{lp}$, we also define the total population in a local region of each point and a value for the ELF index in the local region of each point.

$$\widetilde{N}_p = \sum_{l=1}^{L} \widetilde{n}_{lp} \tag{4}$$

$$elf_p = 1 - \sum_{l=1}^{L} \left(\frac{\widetilde{n}_{lp}}{\widetilde{N}_p} \right)^2 \tag{5}$$

[14] In terms of the size of the radius used, there is no *a priori* reason to choose a 50 kilometer radius. However, introspection based on the authors' personal experiences with the distances that people tend travel on a regular basis in developed and developing countries, led us to use a 50 kilometer benchmark. Having created the software architecture and toolset for calculating one version of the EDC index, it is easy to adapt the process to using different radii. Future research will explore the impact of changing the radius to smaller and larger values. We do use a more general Gaussian proximity weighting factor when calculating the EC index. The index was calculated at a later stage in this research, when additional software tools became available to the authors. Future research will extend the use of the Gaussian weighting factor to the EDC index.

Finally, we construct our EDC index by averaging this local elf variable over the whole country, weighting by the total population at each point.

$$EDC = \frac{\sum_p N_p elf_p}{\sum_p N_p} \tag{6}$$

In interpreting the EDC index, there are three main characteristics. First, the index has an intuitive interpretation as the average value of the local elf index for the country as a whole. Since the typical citizen experiences the ethnic diversity of his or her local area, the EDC index reflects the average diversity that the typical citizen experiences.

Second, the index also reflects the clustering of ethnic groups within a country. For two countries with the same value of the ELF index, the EDC index will vary depending on how clustered or dispersed members of the ethnic group are. The country with more clustering will have a lower EDC value because the typical citizen in that country will live in an area that is relatively more homogeneous. For that country, the average local elf value at points within the country will be lower, reflecting the clustering of ethnic groups into local areas. But this average of the local elf value is just the EDC index. So, for a given value of the ELF (diversity) index, the EDC index will be lower for countries with clustered ethnic groups. Likewise, if a country has citizens of different ethnic groups dispersed throughout the country, then the typical citizen will live in a relatively ethnically diverse area, the local elf index will be higher on average, and the EDC index will be higher. A final characteristic is that the ELF index serves as an upper bound on the EDC index, with the two being equal when the radius of the "local area" is equal to infinity.[15]

We also construct a third measure, which we refer to as the Ethnic Clustering or EC index. This index corresponds to the entropy-based "H" measure described by Reardon and O'Sullivan (2004), which is their preferred measure of clustering according to several criteria. It is given by

[15]Intuitively, if the radius is larger than the size of the country, then all parts of the country are included in the "local area" for each point in the country. So the local elf index will be equal to the country-wide ELF index at each point in the country and the EDC index will be equal to the ELF index. For radii smaller than the country as a whole, some, but not all of the country will be in each "local area". Although one particular area can be more diverse than the country as a whole, on average the local areas will be less diverse than the country as a whole, and the EDC index will be less than the ELF index. One exception to this is the population of each ethnic group is spread completely evenly throughout the country, in which case the EDC index is equal to the the ELF index.

the formulas below.[16]

$$\widetilde{E}_p = -\sum_{l=1}^{L} \left(\frac{\widetilde{n}_{lp}}{\widetilde{N}_p} \right) \log_L \left(\frac{\widetilde{n}_{lp}}{\widetilde{N}_p} \right) \tag{7}$$

$$E = -\sum_{l=1}^{L} \left(\frac{n_l}{N} \right) \log_L \left(\frac{n_l}{N} \right) \tag{8}$$

$$EC = \widetilde{H} = 1 - \frac{1}{NE} \sum_{p} \widetilde{E}_p N_p \tag{9}$$

E and $\widetilde{E}_p$ refer to the entropy of the national and local environments, respectively. Entropy can be thought of as the noisiness or chaos of a particular system, where the system in this case is the set of numbers that reflect the population share of each ethnic group. The EC index compares the typical entropy of the local environments with the entropy of the country as a whole. The key concept is that, if the local environments are the same as the national environment, with the same population share for each ethnic group locally and nationally, then the local entropies will all be identical to the national entropy. This is the case if the ethnic groups are perfectly interspersed. In this situation, the average of the local entropies will be equal to the country-wide entropy, so the second term in Equation (9) will be one and the EC will be equal to zero.

However, if the local environment is very different from the national environment, the local and national entropies will also differ. This is the case if ethnic groups are clustered together. Here, the average local entropy will be very different from the national entropy and the second term will be closer to zero, causing the EC index to be closer to one. Thus the higher the value of the EC index, the greater the clustering. (Note that this is the opposite direction from that of the EDC index; a higher EDC index is associated with *less* clustering.)

As with our EDC measure, we calculate the EC measure for the special case where the "local area" is designated by a 50 kilometer circle around the point in question. However, for the EC measure, we also add a Gaussian weighting function so that more weight is given to points closer to the center of the circle and less weight is given to points that are further away. Thus $\widetilde{n}_{lp}$ is given by Equation 2, where $\phi(p, q)$ is a Gaussian function that falls to zero at a radius of 50 kilometers. In future research, we will also use this Gaussian function for calculating the EDC index. (See

[16] Reardon and O'Sullivan (2004) discuss their measure in the context of racial segregation, but the index is also appropriate for groups that differ on ethnic or linguistic dimensions, as in our research.

footnote on page 9.)

The EC measure is preferable to the EDC index because it measures the clustering of ethnic groups within a country, irrespective of the overall diversity. This is useful for separating out the impact of ethnic diversity from the impact of ethnic clustering. However, the interpretation of the EC measure is less intuitive than that of the EDC measure, which is the average of the elf index diversity for all local areas.

In addition, the EC measure is undefined for perfectly homogeneous countries, that is countries with one ethnic group, which are also countries with an ELF index equal to zero. For these countries, the value of EC is equal to (1 - 0/0), which is undefined. In practice, the EC index also appears to be highly sensitive to measurement errors when the country is close to homogeneous, namely when the country has one large ethnic group and a few, very small ethnic groups. Thus, when considering the impact of the EC index on civil conflict, we exclude those countries with low or zero value for ELF. We are still exploring this aspect of the index, and our results for the EC index should be considered preliminary.

3 Data Description and Methodology

3.1 Construction of Ethnic Variables

Calculation of the ELF, EDC and EC measures requires information on the location of ethnic groups and information on population density. As was discussed earlier, we use linguistic differences as a proxy for ethnic differences. Data on the locations of language groups is obtained from Global Mapping International's World Language Mapping System.[17] This dataset consists of polygons covering most of the world, for each language spoken today. The language group locations are accurate for the approximate years of 1990-1995.[18] The data are based on SIL International's 15th edition of the Ethnologue linguistics database of languages around the world.[19] Figure 2 shows an

[17] World Language Mapping System Version 3.2, from Global Mapping International (GMI), www.gmi.org.

[18] Small portions of the map are designated as areas with a mix of languages, with the location polygons for several groups overlapping. We use information on the population of each langauge group, available in the Ethnologue/GMI database, to apportion population in a mixed language area between the two or more designated ethnic groups. The GMI database also contains information on widespread languages within a country, but since these languages are not explicitly mapped, their location is not clear. Some may have speakers in every corner of a country, while other widespread languages may only be in certain areas such as large cities. Thus, we do not currently consider the widespread languages in our calculations. The GMI database does not contain information on migration, or explicit information on language group populations in urban areas, which may be more mixed.

[19] Gordon, Raymond G., Jr. (ed.), 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International (formerly the Summer Institute of Linguistics). Online version: http://www.ethnologue.com/.

example of this data, mapping the location of currently-spoken language groups in Senegal. Each language is shown in a different shade of grey.

Data on population is provided by the Gridded Population of the World (GPW) population map data for 1990, from Columbia University.[20] The GPW dataset takes the form of a grid covering the whole world. Each square of the grid contains information on the population in that square. The population data is based on carefully-compiled information from censuses around the world, at the smallest administrative level possible for each country. The grid itself is designated by degrees longitude and latitude and one square measures approximately 5 kilometers on a side.[21] Combining these two sources of data, our estimated ethnic group populations approximately apply to the countries around the world in the early 1990's.

As a first step in the process of calculating the indices, the ethnic language data and the population data are combined to create a grid of dots covering the entire world. The dots are located at the center of each small square on the grid of data from the Gridded Population of the World dataset, and are approximately 1.5 kilometers on a side (or 2.25 km squared).[22] Each dot represents a group of citizens who are assigned a population based on the population data from the underlying square of the GPW data. Each dot also has an ethnic group identity, derived from the ethnic group for the area in which the dot is located.[23]

Figure 1 illustrates this for a small portion of Senegal. Three language areas are designated, using light grey and medium grey shading. The area in the upper left corner is Wolof-speaking, while the area in the middle is Serer-Sine-speaking and the bottom portion is an area where both languages are spoken. Super-imposed on this are small squares, each representing one of the "dots" of information. The dots that correspond to higher population density are colored darker. For the mixed language area, only one of the two dots at each point is shown. At Example Point 2, which is in the Serer-Sine region, the population at that point is designated as 4288, while the ethnic group

[20]Gridded Population of the World database Version 3, from the Center for International Earth Science Information Network (CIESIN) Socio-Economic Data Center (SEDAC), Columbia University, New York; http://sedac.ciesin.columbia.edu/gpw/index.jsp. There is some data on the population of each ethnic group available in the GMI dataset. However, the information is incomplete and is listed for widely varying census years.

[21]Prior to our analysis, we project the population and ethnic data using Albers Equal Area Conic and Lambert Conformal Conic projections for each continent.

[22]In practice, we create a grid that is three times smaller than the GPW data, so that each original square of about 5km on a side is turned into nine smaller squares. We do this so as to have a smaller-scale fit between the population and ethnic group data. Because of this division, the resulting population information is essentially increased by a factor of nine. However all indices that we compute rely on ratios of populations, so this factor of nine cancels out across the board.

[23]For mixed langauge areas, two or more dots are created on top of one another, one for each language in that area. The population at each spot is divided among the language dots in proportion to the share of each language group in the total population of the country.

tagged to that point is Serer-Sine. Example Point 1 is slightly more complicated, as it is in an area of mixed languages. Here the population at that location is divided between the two ethnic groups based on their relative populations in the country as a whole. As a result, at Example Point 1 there are two dots, a dot representing 2211 Serer-Sine-speaking people and a dot representing 6833 Wolof speakers.

The collection of these dots is the basis for the calculation of the ELF, EDC and EC indices. The ELF index is constructed by considering all the points in a country. We use the population and ethnic group information from the dots to calculate estimates of the country-wide population for each ethnic group. Based on these estimates, we calculate the ELF index, which is based solely on the ratio of each group's estimated population in the whole country to the total estimated population of the country.

For the EDC index, we require detailed information on the ethnic population in the surrounding area of each location. We divide the country into a grid of the same size and location as the population data grid. For each square in that grid, we designated a circular area of 50 kilometers in radius as the "local area" within which the typical citizen would travel frequently and experience the ethnic diversity of the area. For each square, we calculated the total number of people in each ethnic group, and the fraction of each ethnic group in the surrounding 50km circle. From these fractions of ethnic groups for the local area around each point, we calculate a local elf index for each point. We then generate a map which has the local elf index for each grid square in the country. The EDC index is calculated by taking a population-weighted average of the local elf index across the entire country.

Figures 2 through 5 show the language maps for Senegal and Zimbabwe and the resulting maps of the local elf index for both countries. Although Zimbabwe and Senegal have reasonably similar values for the country-wide ELF index (0.633 and 0.659, respectively), there is more interspersion of the ethnic groups in Senegal. For example, in Senegal there are a dozen or so smaller areas of people of a different ethnic group within some of the larger ethnic group areas (Figure 2). With less clustering (more interspersion) in Senegal, the local elf index is higher on average, so the EDC index, which is the average value of the local elf variable, should also be also higher. In fact, the EDC index for Senegal is 0.349 as opposed to a value of 0.160 for Zimbabwe. Thus the EDC index reflects the relative clustering of ethnic groups in a country, as well as the overall ethnic diversity of that country.

Although the EDC and ELF indices are the main focus of our research, it is worth pointing out

that the maps of the local elf index are useful in and of themselves, as they indicate areas of high and low ethnic diversity within a country (Figures 3 and 5). Visually, areas of the country that are more diverse are lighter, while less diverse areas are darker. For example, the places in Senegal where several smaller groups are located in the middle of a larger ethnic group area are colored lighter, indicating high ethnic diversity in those areas. Places where one or more language groups meet are areas of higher diversity, while places in the middle of a language group's area have low diversity.

The EC index is computed in a similar manner to the EDC index. For each point in the country we calculate the population of each ethnic group in the local area around each point. Then, using the shares of each language group in these local areas, we compute the EC index.[24] The EC index also requires information about the share of each ethnic group in the country as a whole. But this information is the basis for the ELF index, so it is readily available.

We calculate the ELF, EDC and EC indices for 189, 189 and 185 countries, respectively; there is an overlap of 182 countries for which we have all three indices.[25] The raw correlation between the ELF and EDC indices is 0.86 (Table 2, Panel A). This reflects the fact that the ELF index measures the diversity of a country (specifically the chance that two randomly drawn citizens will be of different ethnic groups) while the EDC index reflects both overall diversity and the clustering of ethnic groups. The EC index is not at all correlated with the ELF index, which is not surprising since they measure different aspects of ethnic geography, namely clustering and diversity. There is a small negative correlation between the EDC and EC indices.

Encouragingly, the ELF index which we construct is highly correlated with the principle previous calculation of the ELF index (referred to here as the ELF60 index), which was based on the tabulated populations of each ethnic group from the *Atlas Narodov Mira [Atlas Of The Peoples Of The World] (1964)*. The correlation is 0.79 between the ELF60 index and our ELF index (Table 2,

[24] See Section 2 for details on the EC formula. One difference is that, for the EC index we use a Gaussian weighting function, so that populations close to the point in question received higher weight than those further out.

[25] The calculation of these indices is conceptually quite straightforward, but requires considerable computing power and attention to the exact manner in which the data is divided up to be processed. The software used to manipulate the digital map data, ArcGIS, is immensely powerful and versatile, but it is Windows-based (limiting the possible use of servers), and uses only limited computer memory (1GB of RAM). Even the typical medium-sized country has 200,000 to 1,000,000 points, each with ethnic and population data. In computing our EDC index, we have developed a specialized toolset for ArcGIS which is designed to meet the needs of our project. Thanks to this toolset, we anticipate that future versions of the indices will be much more straightforward to compute. With regard to specific countries, idiosyncratic technical difficulties with the calculation method and software prevented us from including a dozen or so countries in one or more of the indices. In addition, six large countries (Australia, Canada, China, Kazakhstan, Russia, and the United States) are omitted because their files are too large to process with the current toolset. We are currently working on modifications to our algorithms so as to be able to calculate the indices for all omitted countries.

Panel B). The two measures are from different time periods, the 1960's versus the 1990's, and are constructed based on different datasets. The ELF60 index uses lists of ethnic group populations for each country. By contrast, our ELF measure is based on combining the ethnic group areas with population density data, on a geographic basis, and estimating ethnic populations from the resulting data. Finally, the definitions of ethnic groups in the Atlas Narodov Mira are mostly based on language but also include some aspects of religion and more loosely-defined "culture." Our index is based solely on ethnic groups as determined by linguistically-defined languages.[26]

Given these significant differences, it is reassuring that the correlation is as high as it is. We are encouraged by this strong correlation between our measure and the traditional measure and see this as a reflection of the validity of our map-based methodology to estimate ethnic populations. Finally, a major advantage of our method is that data is available for essentially the entire world, allowing us to create the indices for many more countries than are typically available for cross-country analysis.

Summary statistics for the ethnic variables we construct are given in Table 3. All four indices have values between 0 and 1 and the ELF and ELF60 indices have similar characteristics. Note that ELF60 is available for only 113 countries compared to 189 countries for our version of the ELF index. The mean and median for the EDC index are lower than the corresponding values for the ELF index, reflecting the fact that the ELF index for each country is an upper bound for the EDC index for that country.

3.2 Civil War Variables

We construct our civil war variables based on two alternative sources. Our main data come from the Correlates of War (COW) database, specifically the PRIO/Uppsala Version 3.0.[27] The COW data are widely used in the literature (for example, Collier et al. (2004) and Fearon and Laitin (2003)) and hence we use this data for our baseline regressions. The dataset includes information on the location, start date, end date and total battle deaths per year for each conflict. As a robustness check, we also use the Major Episodes of Political Violence (MEPV) dataset, which includes a slightly different set of conflicts.[28] The MEPV dataset also has some information, albeit

[26] In order to explore further the differences between the ELF60 and our ELF index, we are currently constructing an "intermediate" version of the ELF index, in which we use the (somewhat incomplete) population data in the GMI linguistics database, but rely on the older, non-map-based method to calculate an ELF index. Thus we combine the method of the ELF60 index and the data from our ELF index.

[27] http://new.prio.no/CSCW-Datasets/Data-on-Armed-Conflict/. We update the number of battle deaths through 2005 using updated data from Uppsala/Prio. Data on battle deaths are based on Gleditsch, Wallensteen, Eriksson, Sollenberg and Strand (2002) and Lacina and Gleditsch (2006).

[28] See http://members.aol.com/cspmgm/warlist.htm for more details.

incomplete, on civilian deaths, while the COW data only considers battle combatants.[29]

There has been extensive debate in the literature concerning the definition of civil conflict, with many authors relying on different definitions and hence using a different number of conflicts in their analysis.[30] We abstain from entering into this debate, and instead use the broadest definition of civil conflict for our analysis, specifically the entire COW database. We then perform several robustness checks with alternative definitions. First, we use the MEPV database. Next, for each of the COW and MEPV databases, we restrict the sample to conflicts with more than 100 battle dead per year and more than 1000 total battle dead during the entire conflict, a common threshold in the literature. For the MEPV data, we also run regressions in which we only include conflicts which are labeled as having an "ethnic" component. Finally we also restrict our COW database civil conflicts to those also considered by Fearon and Laitin (2003), whose dataset is widely used in the civil war literature.

Using data from the COW database, we construct eight civil war variables using data on the conflicts in each country.[31] First, we sum the number of conflicts in the country that begin between 1945 and 2005. This variable is our primary measure of civil war incidence. We also create a dummy variable that indicates the presence or absence of civil war during the sample period of 1945-2005.

To measure the average duration of war, we calculate the length of each conflict and then average that number across all the conflicts in each country. This measure reflects the propensity of a country to have long wars versus short wars and can more closely address the question of whether certain patterns of ethnic groups, such as diversity or clustering/interspersion, lead to longer wars. We also calculate the average casualties per conflict for each country and report this variable in absolute terms and as a proportion of the population of the country (per 1000 inhabitants).

While the preceding two sets of variables, conflict incidence and conflict duration/intensity, relate more directly to our hypotheses outlined in Section 4, several additional variables measure the total burden placed on the country by civil war and may also be of interest. Total duration is the total time spent in war and is measured as the fraction of years between 1945 and 2005 that the country had a civil war. Similarly, we calculate the total number of casualties due to civil war, and the total casualties normalized by the country's population. Finally we make use of the COW variable describing the intensity of civil wars. This intensity variable takes a value of 1, 2,

[29]The MEPV conflicts include approximately 16 million total deaths, while the COW conflicts include approximately 5 million total battle deaths.

[30]See Sambanis (2004) for an in-depth discussion.

[31]We construct similar measures using the MEPV data.

or 3, depending on the number of casualties in the conflict. When combining information on many conflicts for one country, we use the highest value of the COW intensity variable for that country.

Since a country cannot have a negative number of conflicts, years spent in conflict, or battle deaths, all civil war variables are censored at zero. In fact, the median country in our sample has a zero for all of its civil war variables. Because of this censoring of the civil war data, we use a Tobit specification in our regression analysis.

Summary statistics for the COW civil war variables, as well as for our main control variables, are described in Table 3. The average country had a little over one conflict and spent over five years in civil war. However, as reflected by the zero median for each civil war variable, the typical country has no civil war. Among countries with conflict (not shown in the table), the mean number of conflicts is 2.6 and the mean total time spent in war is 11.6 years.

Raw correlations between the ethnic and civil war variables are shown in Table 2, Panel C. Civil war outcomes are only somewhat correlated with the ELF index and even less correlated with the EDC index. However, the EDC index is picking up on diversity as well as clustering, so regression results in which we control for diversity using the ELF index should be more helpful. Finally, note that the six civil war variables listed (number of conflicts, total duration, average duration, total casualties, average casualties, and the COW intensity measure) are somewhat correlated with one another, but they are not identical. These correlations are all between 0.24 and 0.84.

4 Theoretical Relationship between Ethnicity and Conflict

In this section we consider several hypotheses concerning the impact of ethnic diversity and clustering on civil conflict. Theories of ethnic identity can be divided into two categories. Primordialist theories describe situations in which an agent's affinity for his or her ethnic group members enters directly into his or her utility function. By contrast, instrumentalist theories describe agents who do not care about ethnicity itself, but who pay attention to ethnicity because of its potential strategic role. While we feel that primordialist theories are important, for the most part we focus on instrumentalist theories so as to make the case that ethnicity can play a role even outside of any natural affinity that members of an ethnic group have for one another. In many cases, primordialist arguments can serve to strengthen the theories we describe below. In Section 5, we test our hypotheses empirically using our newly constructed ethnic indices and the data on civil wars.

4.1 Ethnic Diversity and Civil War

We begin by considering the relationship between ethnic diversity and the incidence of civil war. In our empirical work, ethnic diversity is measured by the ELF index. Caselli and Coleman (2006) describe a model in which ethnicity allows coalitions to be enforced *ex post*, because non-coalition members can be excluded from a winning coalition based on their ethnicity. Since potential warring coalitions will be stronger with ethnicity as an enforcement mechanism, more ethnic diversity should be associated with more civil conflict. Other authors (Chandra (2003), Hardin (1995)) have shown that ethnicity can be used to coordinate on coalitions, while Fearon (1999) writes that ethnicity can help create and enforce coalitions for pork projects. Glaeser (2005) describes a model in which politicians may actually incite ethnic hatred towards members of racial groups in opposing coalitions, so as to weaken their opponents. Thus ethnicity can play a role in creating and strengthening coalitions. These findings lead to our first hypothesis concerning the impact of ethnic diversity on the incidence of civil conflict.

Hypothesis 1: *A higher degree of ethnic diversity is associated with a higher incidence of civil conflict.*

More specifically, Caselli and Coleman (2006) and others propose a U-shaped relationship between ethnic diversity and conflict incidence. They argue that ethnicity is most likely to be helpful in forming or enforcing coalitions when there are two or three main ethnic groups in a country, which is a situation associated with an intermediate level of overall ethnic diversity. We test a corollary of Hypothesis 1, which predicts that the relationship between ethnic diversity and conflict is non-monotonic.

We now turn to the question of which factors affect the average duration of conflicts in a country. Two crucial concepts are the cohesion of the ethnic group and the enforceability of post-conflict settlements. These two factors affect the duration of conflict in opposite directions. We posit that the more cohesive the coalition, the less likely it is that the conflict will "fade away" over time in the face of random shocks to other factors that affect the coalition's success, such as available resources or outside support. This channel is more along the lines of primordialist theories since it depends on intrinsic affinity among members of an ethnic group. It predicts that higher ethnic diversity should be associated with *longer* wars.

However, other factors may cause higher diversity to be associated with shorter wars. As mentioned in the discussion of Hypothesis 1, ethnicity can be used to help enforce coalitions after

a war has come to a settlement. If this is the case, then it should be easier to come to a settlement between the warring parties if ethnicity plays a role in the coalition. Collier and Hoeffler (2006) discuss the importance of the ability to lock-in post-conflict settlements, in ending a civil conflict. Thus, greater ethnic diversity and, by extension, greater ethnic involvement in the formation of coalitions, could lead to *shorter* wars by allowing the warring parties to more easily come to an enforceable, negotiated settlement. Our second hypothesis details this theoretical ambiguity in the relationship between diversity and the duration of conflict.

Hypothesis 2: *A higher degree of ethnic diversity is associated with a* ***longer*** *average duration of civil conflict if ethnic group cohesion is important, and a* ***shorter*** *average duration if post-settlement enforceability is important.*

Again, non-linearities may be relevant in this relationship. Collier et al. (2004) discuss the fact that having many ethnic groups in a rebel group or coalition can reduce the cohesion within that group. (This is also a primordialist theory.) This situation is more likely in countries with very high diversity and thus many small ethnic groups from which to form coalitions. If the effect is important, we should actually observe shorter conflicts in extremely diverse countries, leading to a predicted non-monotonic relationship between ethnic diversity and civil conflict duration.

4.2 Ethnic Clustering and Civil War

A second dimension that we study is the relationship between ethnic clustering and civil conflict. Ethnic clustering will be examined empirically using the EDC index with the ELF index also included in the regression, to control for the ethnic diversity portion of the EDC index.

We begin by considering the incidence of conflict and then discuss factors affecting the duration of conflict. Theory suggests that higher ethnic clustering should be associated with a higher incidence of civil conflict. Bates (1983) points out that many public goods have a spatial aspect, because citizens located near one another benefit from the same local public goods, such as roads, schools and infrastructure. We add to this observation by postulating that coalitions of neighbors have potentially more to gain from civil conflict than coalitions of non-neighbors. Since ethnicity can be used to enforce coalitions, the clustering of people of one ethnic group is associated with a higher likelihood of feasible coalitions, due to the joint advantages of enforceability and a high potential return to members. Thus more clustering should be associated with more civil conflict.

Moreover a second channel reinforces this effect. A clustered ethnic group allows a coalition

based on that ethnic group to have access to a "home base" area in which to create a secure base of operations for a military force supporting the coalition. Hence, we propose the following hypothesis.

Hypothesis 3: *A higher degree of ethnic clustering is associated with a higher incidence of civil conflict.*

Finally we look at how the clustering of ethnic groups might affect conflict duration. If the conflict quiets down or becomes easier to resolve once each ethnic group's army is in control of its own ethnic group territory, then settlement is likely to happen more rapidly for clustered rather than interspersed ethnic groups. By contrast, interspersed ethnic groups could be associated with protracted wars as competing claims to territory require more time to be sorted out. Civilian casualties are also likely to be lower for clustered ethnic groups than for areas with many ethnic groups interspersed, as it is easier for a sympathetic army to protect civilians in one large area, than to protect civilians in scattered settlements.[32] Our final hypothesis reflects this relationship.

Hypothesis 4: *A higher degree of ethnic clustering is associated with a shorter duration of civil conflict, and fewer average civilian casualties.*

Thus, we have predictions that higher ethnic diversity should be associated with a higher incidence of civil conflict, and either shorter or longer average duration. Ethnic clustering is predicted to be associated with a higher incidence but a shorter duration of ethnic conflict and fewer civilian casualties.

5 Empirical Results

5.1 Ethnic Diversity Results based on the ELF Index

We use cross-country regressions to test Hypotheses 1 and 2 which are outlined in the preceding section. Both hypotheses describe the correlation between civil war outcomes and ethnic diversity, as measured by the ELF index.

Hypothesis 1 concerns the incidence of civil war, which is measured by our "number of conflicts"

[32]Opposing this idea are Fearon (2004)'s findings that "sons of the soil" conflicts tend to last for a long time. Here, Fearon describes a situation in which a resource- or land-poor majority is encouraged to move into a well-off minority group's home territory. Fearon finds that these conflicts tend to last for a long time. However, this situation of a majority group infiltrating a minority area is only one particular case of ethnic interspersion (or low levels of ethnic clustering). Furthermore, our data on linguistics tends to be biased towards reflecting the primary language of indigenous people in an area, so we may not be able to measure this type of interspersion very accurately with our current data from Global Mapping International.

variable. Results for this variable are presented in Table 4. Because of the censored nature of our civil war data, our preferred specification is the Tobit specification shown in Columns 5 through 8. For comparison purposes, we also present OLS results in Columns 1 through 4. For each specification, we show the ELF and EDC indices separately and together. We also include regressions with two basic controls, a dummy for whether a country is a former colony, and a measure of climate, specifically the percentage of the country's land area that has a hot and rainy climate (Koppen-Geiger climate zone A).

As shown in Table 4, the ELF index is significantly and positively correlated with the number of conflicts, in both the OLS and Tobit specifications. Addition of the EDC variable and the two controls, colonial and climate, does not affect the significance of this result. This offers strong support for Hypothesis 1 and we can conclude that ethnic diversity is associated with an increase in the number of civil conflicts. Since this hypothesis is based on theories of ethnicity increasing the strength of warring coalitions, we also find support for these theories of ethnic-based coalitions.

In terms of the magnitude of this effect, consider a country which goes from the 25th percentile of the ELF index to the 75th percentile, an increase of 0.26 in the ELF index. By multiplying this number by the coefficient in Table 4, Column 8, we find that this increase in ethnic diversity is associated with an increase of 1.3 in the number of conflicts for the country. This is a substantial change given that the mean number of civil conflicts among countries with any civil war is 2.6 conflicts.

One possible concern is that the definition of each separate conflict may be somewhat arbitrary in countries with many years of overlapping conflicts. Table 5 shows a probit regression in which the dependent variable reflects whether or not a country had any civil conflict during the period of study, 1945-2005. This variable measures incidence of civil war, but is not subject to the questions concerning the definitions of conflicts and sub-conflicts. The ELF index continues to be significant in this specification, even with the addition of the colonial and climate controls (Column 4).

Regarding the relationship between ethnic diversity and average conflict, Hypothesis 2 provides an ambiguous prediction for the direction of this correlation. Results for the average duration outcome variable are shown in Table 6, Panel B. Here, we also show OLS and Tobit results. While the ELF index is significant when entered by itself or with just the EDC index (Columns 5-7), the ELF index loses significance when we include the two main controls, colonial and climate. From this we can conclude that the evidence is weaker for a correlation between ethnic diversity and the average duration of civil conflict. This is in accordance with the ambiguous predictions

of Hypothesis 2, for either shorter or longer wars in countries with higher ethnic diversity. This evidence cannot be taken as conclusive since it relies on a null result. Still, it is possible that the two competing factors of greater cohesion and greater ability to make binding settlements are canceling each other out to a certain degree when it comes to the net effect of ethnic diversity on the average duration of conflicts.

Average casualties per conflict is another measure of the intensity of each civil conflict. These results are reported in Table 7, Panel B. Here, the coefficient for the ELF index is still significant and positive, even when the main controls are included (Column 8). However, when we consider average casualties as normalized by the population of the country (Table 8, Panel B), the coefficient on the ELF index is again insignificant. Thus, there is still little support for a strong correlation between ethnic diversity and the average duration or intensity of conflicts.

Additional civil war variables detail the overall impact or burden that civil war places on a country. Results for total duration, total casualties, and total casualties per 1000 inhabitants are shown in Panel A of Tables 6 through 8. Considering the Tobit regressions that include colonial and climate controls (Column 8 in each table), we find that ethnic diversity as measured by the ELF index is significantly correlated, at the 5% level or higher, with total duration and total casualties. The index is significant at the 10% level for the casualties per 1000 inhabitants variable. Thus, higher ethnic diversity is associated with a greater overall burden of civil conflict.[33] In terms of the magnitude of this effect, for countries already affected by civil war (that is, the uncensored part of our database), increasing ethnic diversity from the 25th percentile rank among countries to the 75th percentile rank is associated with an increase of 5.7 years in the total time spent in war. This compares to an average of 11.6 years spent in conflict among this group of countries, which is a large effect.

Results for the equivalent MEPV variables are shown in Table 9 and broadly confirm the results from the COW data. The major exception is that average duration of conflicts is now significantly correlated with the ELF index.

Finally, we consider the possibility of a non-monotonic relationship between ethnicity and the civil war outcome variables, which is described further in Section 4. We consider a non-linear relationship between the ELF index and the civil war variables by including a square term for the ELF index. In results not reported here, we find that the coefficient on the ELF squared term is

[33]We also find a significant impact on the COW intensity variable. These results are available at http://www.people.fas.harvard.edu/ matuszes/papers.html.

insignificant for all the civil war outcome variables.[34] Thus, we do not find strong support for a non-linear relationship between ethnic diversity and civil conflict outcome variables.

To conclude, for our first two hypotheses, we find stronger support for an impact of ethnicity on incidence of conflict than for an impact on duration. We also find little evidence of a non-linear effect of ethnic diversity on civil conflict.

5.2 Ethnic Clustering Results Based on the EDC Index

Next, we consider the relationship between ethnic clustering and civil war outcomes, and use these findings to test Hypotheses 3 and 4. Our principle variable for measuring clustering is the Ethnic Diversity and Clustering (EDC) index. We report preliminary evidence from our Ethnic Clustering (EC) index in Section 5.3. Because the EDC index measures both diversity and clustering, we consider regressions in which the ELF index is also included, so that the variation in diversity can be absorbed by the ELF index, leaving the EDC coefficient to reflect the effect of clustering.

Hypothesis 3 predicts that more ethnic clustering will be associated with a higher number of ethnic conflicts per country. Results using the number of conflicts as an outcome variable are shown in Table 4. Again, we include our preferred Tobit specification as well as OLS results. Although the sign on the EDC index is positive and significant when this variable is included by itself (Column 6), it is always negative and significant when the ELF index is included (Columns 7 and 8). This indicates that, when controlling for ethnic diversity, the residual impact of the EDC index is to decrease the incidence of civil conflict. A higher EDC index is associated with higher ethnic diversity but also with a more even distribution of ethnic groups within the country, that is with less clustering. Thus, this negative coefficient on the EDC index in the joint regression implies that countries with more clustered (more segregated) ethnic groups are associated with more civil conflict. This conclusion offers support for Hypothesis 3, which predicts that higher clustering is associated with a higher incidence of civil war. Finally, unlike the ELF index, the EDC index is not robust to using a probit model for the presence or absence of conflict (Table 5). When the two main controls are included (Column 4), the EDC index no longer has a significant effect.

In considering Column 6, where we include only the EDC index, two competing factors are at work. A higher EDC is associated with more ethnic diversity, which is tied to more civil conflicts. However, a higher EDC is also associated with less clustering, which is tied to fewer civil conflicts. Since the coefficient on the EDC index by and large remains positive when ELF is not included in

[34] These results are available from the authors by request.

the regression, there seems to be evidence that the ethnic diversity effect on conflict incidence is stronger than the ethnic clustering effect. An alternative explanation is that the EDC index may, by construction, give more weight to ethnic diversity than to ethnic clustering. Further analysis using the EDC and EC indices may be able to shed light on this issue.

To test Hypothesis 4, we consider the evidence on the average duration of civil conflicts, as shown in Table 6, Panel B. Here, the EDC index is insignificant when the ELF index and two basic controls are included (Column 8). Thus we find no support either way for Hypothesis 4, which predicted that higher clustering would be associated with shorter civil wars. For the other two civil war variables measuring the average intensity of wars, average casualties per conflict (Table 7, Panel A) and average casualties per conflict normalized by population (Table 8, Panel A), we find similar results as for the ELF index. The EDC index is significant for the average casualties outcome, but not for the average casualties per 1000 inhabitants outcome.

Regarding the effect of ethnic clustering on the remaining civil war outcome variables which measure the overall impact of civil war on a country, we find that higher clustering is significantly correlated with greater total time spent in war (total duration), higher total casualties, and higher total casualties per 1000 inhabitants (Tables 6 through 8, Panel A).

Results for the MEPV data are broadly in accordance with our results for the COW data (Table 9). One important difference is that the EDC now has a significant, negative coefficient for the average casualties regression (Column 5). Recall that the MEPV casualties data include civilian as well as combatant deaths. So, this result seems to indicate that higher clustering is associated with higher casualties among combatants and civilians combined. This contradicts Hypothesis 4, which predicted that clustering would lower civilian casualties.

To summarize, we find support for Hypothesis 3, which says that higher clustering is associated with a higher incidence of civil conflict. However, we do not find strong support for Hypothesis 4, which predicts that higher clustering would be associated with shorter and less bloody wars. In fact, using the MEPV data on civilian casualties, we find some evidence that more clustering is associated with higher average civilian casualties per war, which contradicts Hypothesis 4.

Finally, we consider wars which are known to have an explicit ethnic component. The MEPV dataset includes a flag for whether the combatants on either side of a conflict were organized in part along ethnic lines. Table 10 shows results when the sample is restricted to these "ethnic wars." Interestingly, the results turn out to be much less strong than when we include all conflicts. When the colonial and climate controls are included, EDC is never significant and ELF is significant only

for the number of conflicts and the total duration of conflict.

5.3 Preliminary EC Index Results

Table 11 presents preliminary results using the EC index as a measure of ethnic clustering. This index has two caveats. First, it is not defined for homogeneous countries. Second, due to the logarithmic function in the formula, the index is very sensitive to measurement errors when there is one very large ethnic group and a few small ethnic groups. These countries also have a low ELF, since they are relatively un-diverse. Thus, we present preliminary results for the EC index using only countries with an ELF index greater than 0.1. This reduces our sample by between 40-80 countries, depending on the specification.[35]

Using this restricted sample, the coefficients on the ELF index often lose significance, possibly as a consequence of the lower variation in the ELF index. However coefficients on the EC index are very significant for all civil war outcomes, often at the 1 percent level, with the exception of average duration. The coefficients on the EC index are all positive, meaning that more clustering is associated with more civil conflict. Since more clustering is associated with a lower EDC index, but a higher EC index, these results are consistent with our conclusions from regressions using the EDC and ELF indices. We intend to analyze this index more closely in the future.

5.4 Artificial States and Political Outcomes

Next, we consider whether ethnic diversity and clustering is more likely to lead to civil conflict in countries which were created "artificially." Alesina et al. (2006) construct two measures of the degree to which a state is artificial. Their "partitioned" variable measures the fraction of a country's population that belongs to a partitioned ethnic group, which is an ethnic group that is split into two or more different countries. They also create a measure of the straightness of a country's border, a "fractal" index based on the fractal dimension of the border. We control for both of these variables and find that neither fractal nor partitioned is significant.[34] However, all four ethnic variables are noisy and the number of observations is considerably reduced by the inclusion of all the controls.

We construct a variable, artificial, which is a dummy variable which takes the value of one when the fractal variable is below its median value; lower values of fractal correspond to more artificial states. We then interact artificial with our two indices, ELF and EDC (Table 12). For two of

[35]Several zero- or low-ELF countries drop out of the sample when we include the control variables.

our outcome variables, the number of conflicts and the COW intensity variable (Columns 1, 2, 7 and 8), ELF and EDC lose significance when the interaction terms are included, and three of the four interaction terms become significant. The first two coefficients, on the level terms, reflect the effect of ELF and EDC for the non-artificial countries, while the coefficients on the interaction terms reflect the differential impact on artificial countries. Thus having high diversity or high clustering seems to be problematic only for artificial states.[36] The coefficients are not significant for the average duration and average casualties outcomes. Hence these results should be considered indications, but not as strong evidence one way or the other.

Finally, we briefly consider the impact of ethnic clustering on several economic, political and public goods outcome variables (Table 13). Using the same outcomes as Alesina et al. (2006), originally derived from Kaufmann, Kraay and Mastruzzi (2004), we find significant effects of the ELF and EDC indices on six political variables, checks on power and accountability, political stability, government effectiveness, regulatory quality, the rule of law, and corruption. When controlling for colonial status and climate, ELF is significant at the 1% level for all six variables. EDC is not as robust, but is still significant at the 5% level or higher for regulatory quality and marginally significant at the 10% level for political stability and corruption. In all cases, the signs of the coefficients reflect the fact that higher diversity (higher ELF) and more clustering (lower EDC) are associated with worse political outcomes. The impact of ELF and EDC on several economic and public goods variables is largely insignificant. These results are not shown but are available upon request.

To conclude, based on evidence from our EDC and ELF indices, we find support for Hypotheses 1 and 3, which state that higher diversity and higher clustering are associated with a higher incidence of civil conflict. We find less support for Hypotheses 2 and 4, which concern the average duration of civil war. We find that higher ethnic diversity and clustering are associated with worse overall civil war outcomes, particularly in terms of the total time spent in war and the total civilian casualties. Preliminary evidence from our EC index supports these finding. Artificial states appear to be more strongly affected by ethnic diversity and clustering in terms of the number of civil wars; while the impact on non-artificial states is not significant. Finally, we show evidence that higher clustering, and especially higher ethnic diversity, is associated with worse political outcomes.

[36] Interestingly, the coefficient on the artificial variable is positive, indicating that less artificial states tend to have more civil war. However, this coefficient is much smaller in magnitude than the coefficients on the interaction terms, so that in ethnically diverse countries the net effect should still be that an artificially constructed (relatively straight) border is associated with more civil war.

6 Robustness Tests

In keeping with the large variation across the literature on ethnic diversity and civil conflict, we perform several robustness checks with our data. For the first three robustness check, we present results for both the COW data (Panel A in each table) and the MEPV data (Panel B in each table). First, we include continent dummy variables and our results if anything become stronger (Table 14). ELF and EDC still have no impact on conflict duration in the COW data, but do have an impact in the MEPV data.

In regressions not included in this paper, we restrict our data to big countries and big conflicts.[37] We limit our sample to countries with over 500,000 inhabitants, dropping 33 countries in the process, and our results are essentially unaffected. We also limit our sample of civil wars to big conflicts, as measured by the number of battle deaths. We include only conflicts and conflict-years in which there were 100 or more battle deaths per year and 1000 or more total battle deaths during the span of the conflict. This reduces our sample of conflicts considerably, from 264 to 126. Our results for both ELF and EDC remain robust. However, the ELF index becomes significant for the average duration outcome, which provides additional evidence regarding the predictions of Hypothesis 2. The result suggests that ethnic diversity can lengthen big conflicts, but has no significant effect on overall conflict.

We consider alternative definitions of conflict, and include only conflicts in the COW database that are also in the database used by Fearon and Laitin (2003). In addition to the COW database, this data is one of the most widely-used sources in the literature. Although this sample of conflicts is restricted to the overlap between the COW and Fearon-Laitin databases, over two thirds of the conflicts in the Fearon-Laitin database are included. Results are shown in Table 15. Interestingly, using this group of conflicts, the ELF index has a significant impact on the average duration of ethnic conflict.

One control that does have a slight impact on the EDC measure is country size. When we include this variable, the coefficients on the ELF index remain significant. However, results for our EDC index are no longer significant at the 5% level, although they remain significant at the 10% level. Given a constant degree of ethnic diversity, a larger country will tend to have ethnic groups living further apart on average. Hence the EDC may also be picking up on this effect, in addition to the effects of diversity and clustering. Still, our results for the EDC index remain weakly significant even when including this country size factor. By contrast, coefficients for both the ELF and EDC

[37] These results are available at http://www.people.fas.harvard.edu/ matuszes/papers.html.

indices remain significant when we include population as a control.[34]

Finally, we consider the crucial concept of reverse causality from civil conflict to the ethnic geography make-up. Members of an ethnic group may migrate or attempt to switch ethnicities in response to civil war, particularly for ethnically-based civil wars. Since our main civil war data is from 1945 to 2005, and our ethnic data is from the 1990's this is a potentially serious problem.[38] In Table 16, we show results from restricting the sample to only conflicts beginning in 1990 or later. This reduces our sample of conflicts considerably. The result for the EDC index is no longer significant, although the sign is still negative. Results for the ELF index are robust to using only the post-1990 conflicts.

In future work, we hope to instrument for the ELF index using measure of genetic variety within a country. Genetic data is available for many populations all over the world. However, to date it appears that much work has focused on single genetic mutations, or on classifying people around the world in to groups, for example 32 or 44 distinct groups. However, we are interested in a measure of genetic diversity within a country, which requires more information than this. We hope to be able to present instrumental variables results along these lines at some point in the future.

To conclude, our results are essentially robust to considering only large countries or large conflicts, to including continent dummy variables, to alternative definitions of conflict, and to the addition of other ethnic controls. Country land area does have an impact on the results for the EDC index, reducing the significance level to 10%. Our results for the ELF index are unaffected when we consider only post-1990 conflicts, while results for the EDC index become insignificant in this smaller sample.

7 Conclusions

We construct a new index of ethnic diversity and clustering (EDC) using digital map data for the location of language groups around the world and population density data. We also replicate the traditional ethno-linguistic fractionalization (ELF) index using our data and methodology and find that it is highly correlated with previous measures of the ELF index.

Both the ELF index, which measures diversity, and the new EDC index, which measures both diversity and clustering, are shown to be significantly correlated with the incidence and the overall

[38] It seems likely that migration is more likely to cause reverse causality problems, than ethnic group switching. Changes to the mother tongue spoken by a family often occurs over many generations, so these changes are less likely to be observed in the time frame we study.

impact of conflict on a country, as measured by total duration of conflict and total casualties. When the ELF and EDC indices are included together in regressions, the ELF index captures the ethnic diversity aspect of the EDC index. The coefficient on the EDC index then provides an estimate of the effect of the clustering of ethnic groups. Based on these regression results, we find support for Hypotheses 1 and 3, which posit that higher diversity and higher clustering, respectively, should be associated with higher incidence of civil war.

With regard to the average duration of conflicts, we find that the EDC and ELF indices both have an insignificant effect on this variable. Hypothesis 2 provides an ambiguous prediction for the impact of ELF on duration, while Hypothesis 4 predicts that higher clustering should be associated with shorter wars and fewer civilian casualties. Thus, our results are in accordance with Hypothesis 2, and seem to contradict Hypothesis 4. In addition, using civilian and combatant battle deaths from the MEPV database we find that higher clustering is associated with more overall casualties, directly contradicting Hypothesis 4. Future work should explore additional theoretical reasons for this relationship.

Our results are robust to including climate, continent and colonial history as controls; limiting the sample to large countries or large conflicts; and several other robustness checks. To address a potential reverse causality between conflict and ethnic geography, we restrict our sample to wars beginning 1990 or later. Our results for the EDC index become insignificant, but results for the ELF index are robust to this restriction.

We also examine interactions with measures of artificial states from Alesina et al. (2006). We create a dummy variable using their fractal measure of the straightness of a country's border, which reflects the degree to which a country was constructed artificially. When we consider interactions between this measure of artificiality and the EDC and ELF indices, the results indicate that higher diversity, as measured by the ELF index, and more clustering, as measured by the EDC index are both associated with a higher incidence of civil conflict in artificial states. However neither variable has a sizeable impact for non-artificial, or more "natural" states.

Finally, our new indices are significantly correlated with indicators of political stability and freedom within countries. More diversity and more clustering of ethnic groups are associated with worse political outcomes. Correlations between our measures and economic and public goods indicators are not significant.

We feel this paper provides a significant methodological contribution by introducing the technique of using digital map data to calculate new variables. Here, we calculate several new variables

of ethnic geography, but the approach can be applied to creating data for many different areas of interest. One advantage of this technique is the ability to consider sub-national areas, such as the conflict zones themselves. Another advantage is the ability to generate data for large numbers of countries.

In future work, we will examine the effect of including a GDP control in a panel data setting, a necessary framework since GDP is strongly affected by ethnic diversity and is therefore endogenous. We also plan to examine the duration of conflicts using hazard models to estimate the chance that particular conflict will end in a given time frame, in conjunction with work focusing on conflict-area measures of ethnic diversity and clustering.

To help identify a causal relationship between ethnicity and civil conflict, as opposed to the correlations that we report in the paper, we are exploring the possibility of using genetic diversity as an instrumental variable. We also plan to compare the ethnic diversity and segregation of conflict zones to see if there is a correlation between certain patterns of ethnic geography and the duration or severity of the conflicts. Additional possible extensions include regional (sub-national) analysis and more extensive analysis of political and economic outcome variables.

References

Alesina, Alberto and Eliana La Ferra, "Ethnic Diversity and Economic Performance," November 2004. Working Paper.

___ **and Enrico Spolaore**, "War, peace and the size of countries," *Journal of Public Economics*, 2005, *89*, 1333–1354.

___ **, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg**, "Fractionalization," *Journal of Economic Growth*, 2003, *8*, 155–194.

___ **, William Easterly, and Janina Matuszeski**, "Artificial States," June 2006. NBER Working Paper No. 12328.

Atlas Narodov Mira [Atlas Of The Peoples Of The World], Moscow: Glavnoe Upravlenie Geodezii i Kartografii, 1964.

Azam, Jean-Paul and Anke Hoeffler, "Violence Against Civilians in Civil Wars: Looting or Terror?," *Journal of Peace Research*, 2002, *39*, 461–485.

Bates, Robert, *Essays on the Political Economy of Rural Africa*, Cambridge, UK: Cambridge University Press, 1983.

Buhaug, Halvard and Jan Rød, "Local determinants of African Civil Wars, 1970-2001," March 2005. Working Paper.

___ **and Päivi Lujala**, "Accounting for scale: Measuring geography in quantitative studies of civil war," *Political Geography*, 2005, *24*, 399–418.

___ **and Scott Gates**, "The Geography of Civil War," *Journal of Peace Research*, 2002, *39*, 417–433.

Carment, David and Dane Rowlands, "Vengeance and Intervention: Can Third Parties Bring Peace without Separation?," *Security Studies*, 2004, *13*, 366–393.

Caselli, Francesco and Wilbur John Coleman, "On the Theory of Ethnic Conflict," March 2006. NBER Working Paper No. 12125.

Cederman, Lars-Erik and Luc Girardin, "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies," February 2005. Working Paper.

___ **, Jan Rød, and Nils Weidemann**, "Geo-Referencing of Ethnic Groups: Creating a New Dataset," February 2006. Working Paper.

Chandra, Kanchan, *Why Ethnic Parties Succeed: Patronage and Ethnic Headcounts in India*, Cambridge, UK: Cambridge University Press, 2003.

Collier, Paul, "Ethnic diversity: An economic analysis of its implications," *Economic Policy*, 2001, *32*, 129–66.

___ **and Anke Hoeffler**, "On the economic causes of civil war," *Oxford Economic Papers*, 1998, *50*, 563–73.

___ **and** ___ , "On the Incidence of Civil War in Africa," *Journal of Conflict Resolution*, 2002, *46*, 13–28.

___ **and** ___ , "Greed and grievance in civil war," *Oxford Economic Papers*, 2004, *56*, 563–595.

___ **and** ___ , "Civil War," March 2006. Forthcoming in the Handbook of Defense Economics.

___ **and Nicholas Sambanis**, "Understanding Civil War: A New Agenda," *Journal of Conflict Resolution*, 2002, *46*, 3–12.

___ **, Anke Hoeffler, and Måns Söderbom**, "On the duration of civil war," *Journal of Peace Research*, 2004, *41*, 253–73.

___ **, V.L. Elliott, Håvard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis**, *Breaking the Conflict Trap: Civil War and Development Policy*, Washington, DC: World Bank and Oxford University Press, 2003.

Downes, Alexander, "The Problem with Negotiated Settlements to Ethnic Civil Wars," *Security Studies*, 2004, *13*, 230–279.

Duclos, J., J. Esteban, and D. Ray, "Polarization: Concepts, Measurement, Estimation," *Econometrica*, 2004, *72*, 1737.

Easterly, William and Ross Levine, "Africa's Growth Tragedy: policies and Ethnic Divisions," *The Quarterly Journal of Economics*, 1997, *112*, 1203–1250.

Echenique, Federico and Roland Fryer, "A Measure of Segregation Based on Social Interactions," *forthcoming in The Quarterly Journal of Economics*, 2006.

Eck, Kristine, "A Beginner's Guide To Conflict Data: Finding and Using the Right Dataset," December 2005. UCDP Paper #1.

Esteban, J. and D. Ray, "On the Measurement of Polarization," *Econometrica*, 1994, *62*, 819.

Fearon, James, "Why Ethnic Politics and "Pork" Tend to go Together," June 1999. Working Paper.

___ , "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 2003, *8*, 195–222.

___ , "Why Do Some Civil Wars Last So Much Longer Than Others?," *Journal of Peace Research*, 2004, *41*, 275–301.

___ , "Primary Commodity Exports and Civil War," *Journal of Conflict Resolution*, 2005, *49*, 483–507.

___ **and David Laitin**, "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 2003, *97*, 75–90.

Glaeser, Edward, "The Political Economy of Hatred," *Quarterly Journal of Economics*, 2005, *120*, 45–86.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand, "Armed Conflict 19462001: A New Dataset," *Journal of Peace Research*, 2002, *39*, 615.

Gurr, Ted Robert, "Peoples Against States: Ethnopolitical Conflict and the Changing World System: 1994 Presidential Address," *International Studies Quarterly*, 1994, *38*, 347–377.

___ , "Ethnic Warfare on the Wane," *Foreign Affairs*, June 2000.

Hardin, Russel, *One for All: The Logic of Group Conflict*, Princeton, NJ: Princeton University Press, 1995.

Hartzell, Caroline, "Explaining the Stability of Negotiated Settlements to Intrastate Wars," *Journal of Conflict Resolution*, 1999, *43*, 3–22.

___ **, Matthew Hoddie, and Donald Rothchild**, "Stabilizing the Peace after Civil War: An Investigation of Some Key Variables," *International Organization*, 2001, *55*, 183–208.

Hegre, Håvard, "The Duration and Termination of Civil War," *Journal of Peace Research*, 2004, *41*, 243–252.

___ **and Nicholas Sambanis**, "Sensitivity Analysis of Empirical Results on Civil War Onset," April 2006. Working Paper.

Kalyvas, Stathis N., *The Logic of violence in Civil War*, New York: Cambridge University Press, 2006.

Kaufmann, Chaim, "Possible and Impossible Solutions to Ethnic Civil Wars," *International Security*, 1996, *20*, 136–175.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi, "Governance Matters III: Governance Indicators for 1996-2002," April 2004. Worldbank Working Paper.

Lacina, Bethany Ann and Nils Petter Gleditsch, "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths," *European Journal of Population*, 2006, *22*, 145.

Laitin, David, "Ethnic Unmixing and Civil War," *Security Studies*, 2004, *13*, 350–365.

Licklider, Roy, "The Consequences of Negotiated Settlements in Civil Wars, 1945-1993," *The American Political Science Review*, 1994, *98*, 681–690.

Lujala, Päiva, Nils Peter Gleditsch, and Elisabeth Gilmore, "A Diamond Curse? Civil War and a Lootable Resource," *Journal of Conflict Resolution*, 2005, *49*, 538–562.

Mason, David and Patrick Fett, "How Civil Wars End: A Rational Choice Approach," *Journal of Conflict Resolution*, 1996, *40*, 546–568.

Massey, Douglass and Nancy Denton, "The Dimensions of Residential Segregation," *Social Forces*, 1988, *62*, 281–315.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, *112*, 725–753.

Raleigh, Clionadh and Håvard Hegre, "Introducing ACLED: An Armed Conflict Location and Event Dataset," March 2005. Working Paper.

Reardon, S. F. and D. O'Sullivan, "Measures of spatial segregation," *Sociological Methodology*, 2004, *34*, 121–162.

—— **and G. Firebaugh**, "Measures of multi-group segregation," *Sociological Methodology*, 2002, *32*, 33–67.

Sambanis, Nicholas, "What Is Civil War?," *Journal of Conflict Resolution*, 2004, *48*, 814–858.

Toft, Monica, "Indivisible Territory, Geographic Concentration, and Ethnic War," *Security Studies*, 2003, *12*, 82–119.

Walter, Barbera, "The Critical Barrier to Civil War Settlement," *International Organization*, 1997, *51*, 335–364.

Wooldridge, Jeffrey, *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press, 2003.

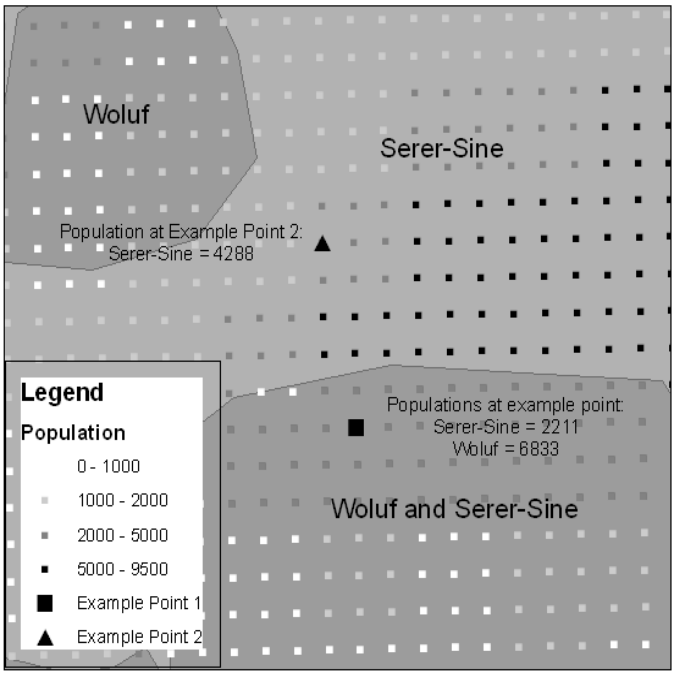

Figure 1: Senegal Points Closeup

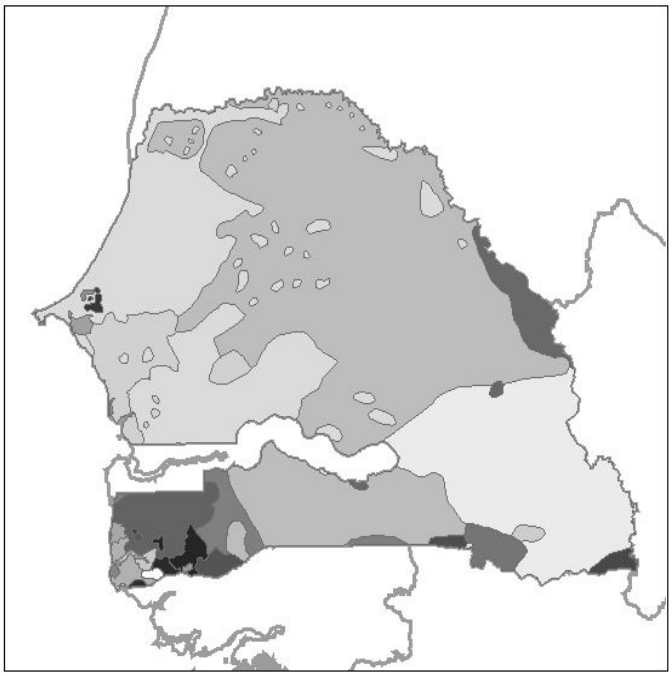

Figure 2: Senegal Languages - Each language is designated with a different shade of grey.

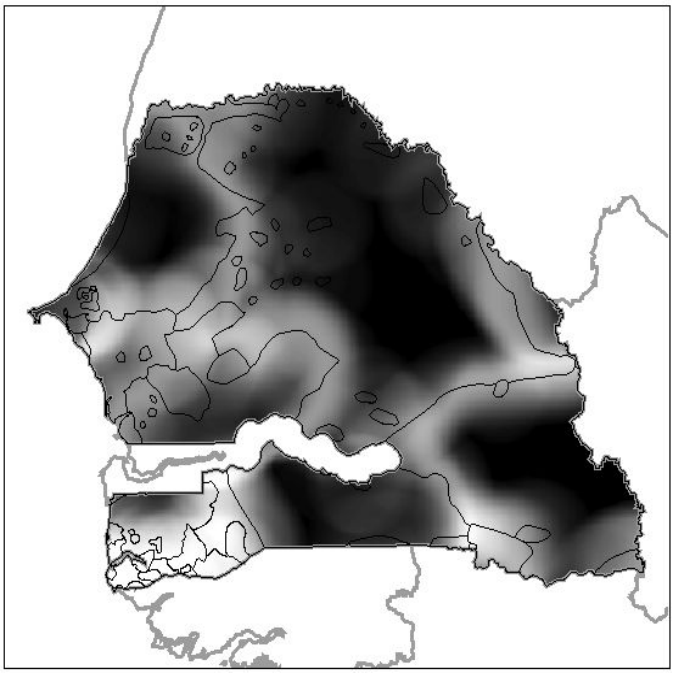

Figure 3: Senegal Local Elf Index - Lighter areas correspond to higher diversity areas, while darker areas have less diversity. Black outlines of the different language areas have been superimposed on this map of the local elf index.

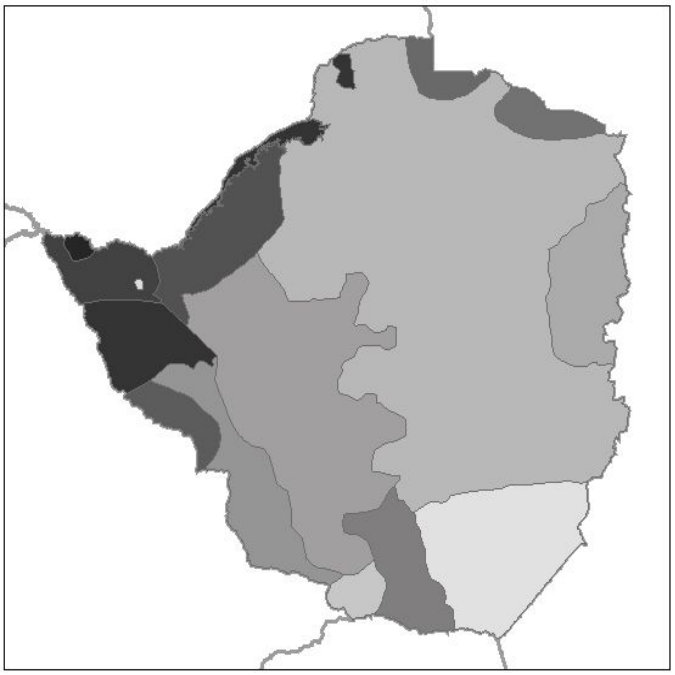

Figure 4: Zimbabwe Languages - Each language is designated with a different shade of grey

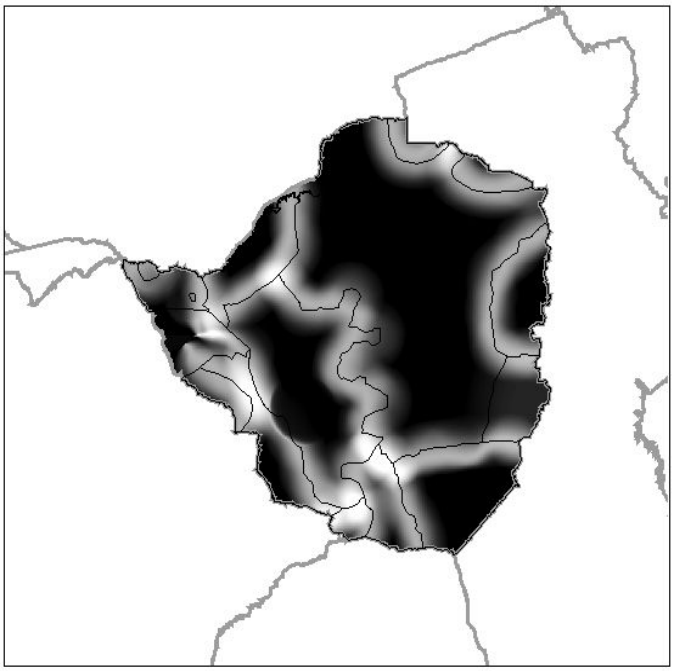

Figure 5: Zimbabwe Local Elf Index - Lighter areas correspond to higher diversity areas, while darker areas have less diversity. Black outlines of the different language areas have been superimposed on this map of the local elf index.

Table 1: **Variable Description and Data Sources**

Variable Name	**Description**	**Source**
ELF	A measure of ethnic DIVERSITY, specifically, the ethno-linguistic fractionalization index. Equal to the probability that two citizens selected at random will be of different ethnic groups. Higher values are associated with higher ethnic diversity.	Computed by the Authors based on Global Mapping Internationals' World Language Mapping System (WLMS), and Columbia University's Gridded Population of the World (GPW).
EDC	A measure of ethnic DIVERSITY and CLUSTERING. Equal to the average diversity (as measured by the ELF index) that the typical citizen experiences in the immediate local area. Higher values are associated with higher ethnic diversity and/or lower levels of ethnic clustering.	Computed by the Authors based on GMI's WLMS and Columbia GPW data.
EC	A measure of ethnic CLUSTERING. Related to the average entropy of the language shares of each area of the conntry, as normalized by the entropy of the language shares of the country as a whole. Higher values are associated with higher levels of ethnic clustering.	Computed by the Authors based on GMI's WLMS and Columbia GPW data.
ELF60	Earlier calculation of the ELF index based on 1960's data from the Atlas Narodov Mira.	Alesina Easterly and Matuszeski (2006)
Number of conflicts (Cow and MEPV)	Total number of conflicts beginning in the country between 1945 and 2005	Calculated based on the COW/PRIO Uppsala Version 3.0 and the Major Episodes of Political Violence databases
Total Duration (Cow and MEPV)	Total number of years spent in civil war between 1945 and 2005	Calculated based on the COW/PRIO Uppsala Version 3.0 and the Major Episodes of Political Violence databases
Average Duration (Cow and MEPV)	Average duration of each conflict in the country, between 1945 and 2005	Calculated based on the COW/PRIO Uppsala Version 3.0 and the Major Episodes of Political Violence databases
Total Casualties (Cow and MEPV)	Total number of battle dead (COW data) or total number of civilian casualties and battle deaths (MEPV data) occurring due to civil wars between 1945 and 2005	Calculated based on the COW/PRIO Uppsala Version 3.0 and the Major Episodes of Political Violence databases
Intensity	Maximum value of the "intensity" code from the COW dataset for the conflicts in the country between 1945 and 2005. Value ranges between 1 and 3	Calculated based on the COW/PRIO Upsala Version 3.0
Colonial	Dummy=1 if ever colonized by European power	Alesina Easterly and Matuszeski (2006)
Climate zone A	Fraction of country that is Koppen-Geiger Climate Zone A (hot, wet climate)	Alesina Easterly and Matuszeski (2006)
Climate zone B	Fraction of country that is Koppen-Geiger Climate Zone B (hot, dry climate)	Alesina Easterly and Matuszeski (2006)
Partitioned	Fraction of country's population that belongs to "partitioned" ethnic groups. A partitioned ethnic group has co-ethnics in at least one neighboring country.	Alesina Easterly and Matuszeski (2006)
Artificial (Fractal)	Dummy variable based on the "fractal" variable in Alesina et al. Equal to 1 if the country has a border that is straighter than the median country's border. Straightness/squiggliness is measured by the fractal dimension of the border.	Alesina Easterly and Matuszeski (2006)
Average Population (1960-2000)	Average population of the country between 1960 and 2000	Penn World Tables
Voice and democracy	Checks on power, accountability to population index	Kaufman Kray (2004)
Political stability	Political stability and violence index	Kaufman Kray (2004)
Government effectiveness	Government effectiveness index	Kaufman Kray (2004)
Regulatory quality	Reglatory quality index	Kaufman Kray (2004)
Rule of law	Rule of law index	Kaufman Kray (2004)
Corruption	Corruption index	Kaufman Kray (2004)

Table 2: **Correlations of Ethnic Indices**

Panel A: Correlation between our calculations of the ELF and EDC indices

(Obs=182)	**ELF**	**EDC**	**EC**
ELF	1		
EDC	0.8578	1	
EC	0.0002	-0.2720	1

Panel B: Correlation of our measures with traditional ELF measure

(Obs=96)	**ELF (1960)**	**ELF**	**EDC**	**EC**
ELF (1960 calculation)	1			
ELF	0.7935	1		
EDC	0.7310	0.8467	1	
EC	0.1955	0.2434	-0.0508	1

Panel C: Correlations between ethnic variables and civil war outcomes (COW/PRIO data, all conflicts)

(Obs=186)	**ELF**	**EDC**	**Number of conflicts**	**Total duration (years in war)**	**Average duration (per war)**	**Total casualties**	**Average casualties (per war)**	**Intensity (COW indicator)**
ELF	1							
EDC	0.8554	1						
Number of conflicts	0.4540	0.2830	1					
Total duration (years in war)	0.3792	0.2566	0.7208	1				
Average duration (per war)	0.1984	0.1530	0.352	0.8351	1			
Total casualties	0.2938	0.1433	0.5299	0.5083	0.2455	1		
Average casualties (per war)	0.1734	0.0791	0.2738	0.3608	0.2929	0.7521	1	
Intensity (COW indicator)	0.4438	0.2757	0.7389	0.6867	0.5216	0.4952	0.4955	1

Table 3: **Summary Statistics**

Variable	**Num Obs**	**Mean**	**Std Dev**	**Min**	**Median**	**Max**
ELF	189	0.32	0.33	0	0.20	0.98
EDC	189	0.16	0.18	0	0.10	0.72
EC	185	0.70	0.27	0	0.75	1
ELF60	113	0.41	0.30	0	0.42	0.93
Number of conflicts	225	1.17	2.04	0	0	15
Total duration (years in war)	225	5.20	10.46	0	0	56
Average duration (per war)	225	2	6	0	0	56
Total casualties	225	23,292	100,778	0	0	1,277,000
Average casualties (per war)	225	7,979	34,091	0	0	425,667
Intensity (COW indicator)	225	0.98	1.26	0	0	3
Colonial (former colony)	208	0.70	0.46	0	1	1
Climate zone A	156	0.32	0.41	0	0	1
Climate zone B	156	0.18	0.30	0	0	1
Partitioned	131	28.23	28.76	0	19.40	100.00
Artificial (Fractal)	143	0	0	-0.01	0	0
Voice and democracy	206	0.01	1.00	-2.19	0.12	1.59
Political stability	206	0.01	1.00	-2.87	0.08	1.77
Government effectiveness	208	0.01	1.00	-2.32	-0.19	2.25
Regulatory quality	203	0.00	1.00	-2.63	-0.06	2.02
Rule of law	207	0.00	1.00	-2.31	-0.11	2.01
Corruption	203	0.00	1.00	-1.65	-0.24	2.53

Table 4: **Number of Conflicts (based on COW/Prio)**

Dependent variable: Number of conflicts in each country between 1945 and 2005. Robust standard errors in parentheses. a, b, and c indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) **OLS**	**(2)** **OLS**	**(3)** **OLS**	**(4)** **OLS**	**(5)** **Tobit**	**(6)** **Tobit**	**(7)** **Tobit**	**(8)** **Tobit**
ELF	**2.451**^{a}		**4.188**^{a}	**3.504**^{a}	**4.928**^{a}		**7.728**^{a}	**4.914**^{a}
	(0.381)		**(0.771)**	**(1.003)**	**(0.845)**		**(1.428)**	**(1.438)**
EDC		**3.158**^{a}	**-3.896**^{a}	**-3.977**^{a}		**7.108**^{a}	**-6.619**^{a}	**-5.655**^{b}
		(0.779)	**(1.137)**	**(1.475)**		**(1.773)**	**(2.087)**	**(2.263)**
Colonial				0.248				0.973
				(0.401)				(0.688)
Climate				0.467				0.816
				(0.372)				(0.572)
Constant	0.329^a	0.689^a	0.365^a	0.486^b	-1.886^a	-1.582^a	-1.728^a	-1.129^b
	(0.097)	(0.117)	(0.096)	(0.186)	(0.513)	(0.526)	(0.510)	(0.566)
Observations	189	189	186	138	189	189	186	138
R-squared	0.21	0.07	0.25	0.18				

Table 5: **Probit Regressions: Presence of Civil Conflict**

Dependent variable: Dummy variable for presence of civil war in a country at any point between 1945-2005. Marginal effects are reported. Robust standard errors in parentheses. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1)	**(2)**	**(3)**	**(4)**
ELF	**0.788**^{a}		**1.258**^{a}	**0.642**^{b}
	(6.27)		**(4.32)**	**(2.21)**
EDC		**0.943**^{a}	**-1.039**^{b}	-0.853
		(3.81)	**(2.15)**	(1.61)
Colonial				**0.277**^{b}
				(2.50)
Climate				0.197
				(1.46)
Observations	189	189	186	138

Table 6: **Duration of Conflicts (Based on COW/Prio)**

Dependent variable: Duration of conflicts in each country between 1945 and 2005. Robust standard errors in parentheses. a, b, and c indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

Panel A: Total Time Spent in Civil War between 1945 an 2005								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	OLS	**OLS**	**OLS**	**OLS**	**Tobit**	**Tobit**	**Tobit**	**Tobit**
ELF	**11.739**^{a}		**18.313**^{a}	**13.291**^{b}	**26.509**^{a}		**39.378**^{a}	**21.409**^{b}
	(2.299)		**(4.575)**	**(6.279)**	**(4.293)**		**(7.396)**	**(8.523)**
EDC		**16.268**^{a}	**-14.546**^{c}	-11.837		**37.380**^{a}	**-29.888**^{b}	-21.100
		(4.525)	**(7.485)**	(10.389)		**(8.729)**	**(12.546)**	(14.527)
Colonial				3.685				**8.911**^{b}
				(2.521)				**(4.291)**
Climate				-0.263				1.544
				(2.791)				(3.753)
Constant	1.449^{b}	2.932^{a}	1.545^{b}	1.439	-11.881^{a}	-9.164^{a}	-11.308^{a}	-9.051^{a}
	(0.651)	(0.661)	(0.646)	(0.977)	(2.357)	(2.223)	(2.343)	(2.855)
Observations	189	189	186	138	189	189	186	138
R-squared	0.14	0.07	0.16	0.12				
Panel B: Average Duration per Civil War between 1945 an 2005								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	OLS	**OLS**	**OLS**	**OLS**	**Tobit**	**Tobit**	**Tobit**	**Tobit**
ELF	**3.651**^{a}		**4.689**^{b}	0.142	**12.427**^{a}		**17.576**^{a}	4.426
	(0.960)		**(2.148)**	(3.683)	**(2.438)**		**(3.834)**	(4.588)
EDC		**5.595**^{a}	-2.165	0.937		**17.324**^{a}	**-11.375**^{c}	-4.118
		(1.860)	(3.792)	(5.576)		**(4.715)**	**(6.779)**	(7.947)
Colonial				**3.111**^{c}				**6.663**^{b}
				(1.825)				**(3.300)**
Climate				-0.596				0.475
				(1.942)				(2.425)
Constant	1.435^{b}	1.788^{a}	1.455^{a}	1.459^{b}	-6.758^{a}	-5.188^{a}	-6.714^{a}	-5.165^{b}
	(0.555)	(0.468)	(0.546)	(0.693)	(1.796)	(1.613)	(1.827)	(2.054)
Observations	189	189	186	138	189	189	186	138
R-squared	0.04	0.03	0.04	0.04				

Table 7: **Casualties in Conflicts (Based on COW/Prio)**

Dependent variable: Casualties in conflicts in each country between 1945 and 2005. Robust standard errors in parentheses. a, b, and c indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

Panel A: Total Casualties in all Civil War between 1945 an 2005								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	OLS	**OLS**	**OLS**	**OLS**	**Tobit**	**Tobit**	**Tobit**	**Tobit**
ELF	**47,920**^{a}		**105,500**^{a}	**107,900**^{b}	**128,500**^{a}		**228,600**^{a}	**161,500**^{a}
	(15,180)		**(36,080)**	**(41,750)**	**(37,780)**		**(71,280)**	**(60,080)**
EDC		**46,710**^{b}	**-124,500**^{b}	**-137,500**^{b}		**152,100**^{a}	**-229,300**^{a}	**-204,500**^{b}
		(18,190)	**(49,690)**	**(57,590)**		**(50,560)**	**(88,960)**	**(84,830)**
Colonial				4,455				**31,000**
				(14,270)				**(26,200)**
Climate				-8,338				2,590
				(16,040)				(18,920)
Constant	1,822	10,290^a	2,831	4,987	-71,350^a	-52,010^a	-67,150^a	-53,490^b
	(2,749)	(3,097)	(2,637)	(5,559)	(22,670)	(17,250)	(21,240)	(22,410)
Observations	189	189	186	138	189	189	186	138
R-squared	0.09	0.02	0.13	0.11				
Panel B: Average Casualties per Civil War between 1945 an 2005								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	OLS	**OLS**	**OLS**	**OLS**	**Tobit**	**Tobit**	**Tobit**	**Tobit**
ELF	**10,550**^{b}		**24,610**^{a}	**19,730**^{c}	**40,230**^{a}		**70,530**^{a}	**37,880**^{b}
	(4,372)		**(8,827)**	**(10,610)**	**(9,535)**		**(16,910)**	**(15,070)**
EDC		9,479	**-30,150**^{b}	**-30,450**^{c}		**47,570**^{a}	**-68,250**^{a}	**-53,950**^{b}
		(6,431)	**(12,570)**	**(15,700)**		**(15,290)**	**(24,350)**	**(24,660)**
Colonial				4,170				**14,960**^{c}
				(4,481)				**(8,136)**
Climate				-3,088				997.7
				(4,016)				(6,060)
Constant	3,120^c	5,078^a	3,393^c	4,847	-24,580^a	-18,190^a	-23,740^a	-17,470^a
	(1,838)	(1,747)	(1,850)	(3,852)	(6,003)	(5,006)	(5,939)	(6,659)
Observations	189	189	186	138	189	189	186	138
R-squared	0.03	0.01	0.05	0.03				

Table 8: **Casualties per 1000 Inhabitants (Based on COW/Prio)**

Dependent variable: Total casualties per 1000 inhabitants in each country between 1945 and 2005. Robust standard errors in parenthesis. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

Panel A: Total Casualties per 1000 Population in all Civil War between 1945 and 2005								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	OLS	**OLS**	**OLS**	**OLS**	**Tobit**	**Tobit**	**Tobit**	**Tobit**
ELF	2.451		**6.480**^{c}	4.525	**9.093**^{a}		**16.377**^{a}	**9.659**^{c}
	(1.592)		**(3.479)**	(4.051)	**(2.429)**		**(5.217)**	**(5.105)**
EDC		2.018	-8.843	-8.092		**11.280**^{a}	**-16.339**^{b}	**-15.060**^{c}
		(2.400)	(5.355)	(5.462)		**(3.841)**	**(8.313)**	**(8.027)**
Colonial				2.325				5.655
				(1.989)				(3.452)
Climate				-1.371				-0.123
				(1.843)				(2.165)
Constant	1.273	1.776^{b}	1.333^{c}	1.141	-5.155^{a}	-3.717^{a}	-5.068^{a}	-5.328^{a}
	(0.778)	(0.722)	(0.779)	(0.749)	(1.432)	(1.172)	(1.394)	(1.966)
Observations	162	163	160	136	162	163	160	136
R-squared	0.02	0.00	0.03	0.04				
Panel B: Average Casualties per 1000 Population per Civil War between 1945 and 2005								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	OLS	**OLS**	**OLS**	**OLS**	**Tobit**	**Tobit**	**Tobit**	**Tobit**
ELF	0.281		1.133	0.034	**3.160**^{a}		**5.430**^{a}	2.107
	(0.683)		(1.080)	(1.512)	**(0.824)**		**(1.706)**	(1.811)
EDC		0.053	-1.832	-1.443		**4.092**^{a}	-5.021	-4.328
		(1.150)	(1.576)	(1.684)		**(1.516)**	(3.053)	(2.919)
Colonial				1.156				2.683
				(0.957)				(1.635)
Climate				-0.558				0.009
				(0.784)				(0.927)
Constant	0.802^{c}	0.884^{b}	0.814^{c}	0.750	-2.052^{a}	-1.567^{a}	-2.054^{a}	-2.134^{b}
	(0.417)	(0.376)	(0.419)	(0.466)	(0.604)	(0.534)	(0.607)	(0.873)
Observations	162	163	160	136	162	163	160	136
R-squared	0.00	0.00	0.00	0.02				

Table 9: **Summary of Results for MEPV Data**

Dependent variables are indicated at the head of each column. All regressions are Tobit regressions. Robust standard errors in parentheses. a, b, and c indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	Number of Conflicts	**Total Duration**	**Average Duration**	**Total Casualties**	**Average Casualties**	**Total Casualties per 1000**	**Average Casualties per 1001**
ELF	**4.215**^{a}	**29.143**^{a}	**14.394**^{a}	**575,700**^{a}	**299,500**^{b}	**35.676**^{b}	**49.594**^{b}
	(1.111)	**(7.766)**	**(4.952)**	**(200,900)**	**(123,200)**	**(15.071)**	**(22.282)**
EDC	**-5.149**^{a}	**-28.748**^{b}	-12.469	**-704,600**^{b}	**-392,600**^{b}	**-56.576**^{b}	**-85.106**^{b}
	(1.910)	**(14.456)**	(9.287)	**(277,800)**	**(182,300)**	**(27.411)**	**(40.213)**
Colonial	0.644	**6.750**^{c}	**4.687**^{c}	52,480	46,420	**7.580**^{c}	**11.996**^{c}
	(0.492)	**(3.676)**	**(2.532)**	(61,670)	(39,300)	**(4.573)**	**(6.606)**
Climate	0.597	1.452	-0.689	94,600	10,760	8.883	**16.068**^{c}
	(0.521)	(3.716)	(2.425)	(70,160)	(36,880)	(8.156)	**(9.540)**
Constant	-0.790^b	-8.886^a	-5.299^a	-213,960^a	-111,080^b	-16.392^b	-21.873^a
	(0.390)	(2.834)	(1.881)	(75,760)	(45,200)	(6.712)	(7.163)
Observations	138	138	138	138	138	136	136

Table 10: **Only Ethnic Wars**

Tobit regressions on ethnic wars. Only conflicts coded as ethnic conflicts by MEPV database are included. Robust standard errors in parenthesis. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) **Number of Conflicts**	**(2)** **Number of Conflicts**	**(3)** **Total Duration**	**(4)** **Total Duration**	**(5)** **Average Duration**	**(6)** **Average Duration**	**(7)** **Total Casualties**	**(8)** **Total Casualties**	**(9)** **Average Casualties**	**(10)** **Average Casualties**
ELF	**5.982**^{a}	**4.226**^{b}	**48.604**^{a}	**31.431**^{b}	**32.510**^{a}	17.376	**394,400**^{b}	277,500	**126,300**^{a}	75,970
	(1.506)	**(1.672)**	**(13.225)**	**(14.952)**	**(10.632)**	(11.751)	**(160,900)**	(171,400)	**(44,250)**	(51,530)
EDC	**-4.375**^{c}	-3.288	-30.152	-18.807	-14.012	-2.564	-300,900	-247,300	-75,730	-42,130
	(2.569)	(2.868)	(22.789)	(26.311)	(18.382)	(21.802)	(212,200)	(229,200)	(79,580)	(90,850)
Colonial		0.131		3.505		3.643		-27,970		-5,346
		(0.789)		(7.173)		(5.701)		(78,790)		(25,930)
Climate		-0.019		-1.463		-2.364		47,940		7,728
		(0.889)		(7.756)		(6.044)		(65,850)		(22,770)
Constant	-3.892^{a}	-2.941^{a}	-35.001^{a}	-27.308^{a}	-28.180^{a}	-21.841^{a}	-309,800^{a}	-229,800^{a}	-111,400^{a}	-80,370^{a}
	(0.674)	(0.724)	(6.620)	(6.865)	(6.500)	(6.374)	(88,720)	(66,860)	(24,240)	(19,800)
Observations	186	138	186	138	186	138	186	138	186	138
R-squared										

Table 11: **Preliminary Analysis Using the EC Index**

All regressions are Tobit regressions. Robust standard errors in parentheses. a, b, and c indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) Number of Conflicts	**(2) Number of Conflicts**	**(3) Total Duration**	**(4) Total Duration**	**(5) Average Duration**	**(6) Average Duration**	**(7) Total Casualties**	**(8) Total Casualties**	**(9) Average Casualties**	**(10) Average Casualties**
Panel A: COW Data										
ELF	**2.954**^{a}	1.486	**15.150**^{b}	6.283	4.080	-2.081	**92,170**^{b}	66,170	**20,040**^{b}	5,254
	(1.007)	(1.229)	**(6.503)**	(8.241)	(3.567)	(5.182)	**(37,270)**	(40,550)	**(9,728)**	(12,310)
EDC	**6.776**^{a}	**7.283**^{a}	**33.401**^{a}	**29.729**^{b}	**14.687**^{a}	8.844	**190,000**^{a}	**190,800**^{a}	**58,490**^{a}	**55,030**^{a}
	(1.685)	**(2.138)**	**(8.156)**	**(12.918)**	**(3.920)**	(6.782)	**(62,330)**	**(64,540)**	**(19,270)**	**(21,230)**
Colonial		0.419		7.969		6.910		17,440		12,440
		(0.994)		(6.394)		(4.638)		(33,930)		(9,095)
Climate		0.549		-0.434		-0.431		-6,949		91.04
		(0.705)		(4.657)		(2.629)		(27,190)		(7,450)
Constant	-5.464^a	-5.358^a	-28.567^a	-25.784^a	-11.659^a	-8.337^c	-186,900^a	-179,800^a	-52,480^a	-49,570^a
	(1.370)	(1.624)	(6.647)	(8.900)	(2.839)	(4.266)	(61,300)	(61,330)	(15,860)	(17,500)
Observations	107	96	107	96	107	96	107	96	107	96
Panel B: MEPV Data										
	(1) Number of Conflicts	**(2) Number of Conflicts**	**(3) Total Duration**	**(4) Total Duration**	**(5) Average Duration**	**(6) Average Duration**	**(7) Total Casualties**	**(8) Total Casualties**	**(9) Average Casualties**	**(10) Average Casualties**
ELF	**3.079**^{a}	**2.670**^{a}	**22.861**^{a}	**17.035**^{b}	**12.233**^{a}	8.377	**478,150**^{a}	**391,400**^{b}	**212,200**^{a}	**180,400**^{b}
	(0.873)	**(0.944)**	**(7.095)**	**(8.271)**	**(4.492)**	(5.653)	**(173,300)**	**(175,100)**	**(81,700)**	**(89,810)**
EDC	**4.018**^{a}	**3.717**^{b}	**29.496**^{a}	**24.089**^{c}	**14.085**^{a}	7.247	**642,100**^{b}	**626,500**^{b}	**277,100**^{b}	**238,600**^{b}
	(1.206)	**(1.694)**	**(8.951)**	**(12.881)**	**(4.796)**	(7.201)	**(262,700)**	**(315,300)**	**(113,100)**	**(119,900)**
Colonial		0.398		7.667		**6.533**^{c}		76,940		66,730
		(0.629)		(5.042)		**(3.499)**		(96,260)		(49,430)
Climate		-0.458		-3.448		-2.747		-27,720		-48,380
		(0.560)		(4.670)		(3.186)		(87,970)		(61,520)
Constant	-3.828^a	-3.426^a	-29.675^a	-26.172^a	-14.970^a	-11.174^b	-770,900^a	-748,100^a	-341,100^a	-323,700^a
	(0.943)	(1.132)	(7.029)	(8.666)	(4.191)	(5.287)	(243,600)	(271,700)	(118,900)	(122,000)
Observations	107	96	107	96	107	96	107	96	107	96

Table 12: **Interactions with Artificial**

All regressions are Tobit regressions and all countries and conflicts are included. Robust standard errors in parentheses. a, b, and c indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) COW Number of Conflicts	**(2) COW Number of Conflicts**	**(3) COW Average Duration**	**(4) COW Average Duration**	**(5) COW Average Casualties**	**(6) COW Average Casualties**	**(7) COW Intensity**	**(8) COW Intensity**
ELF	**5.430**^{a}	1.715	3.591	-3.273	**37,590**^{b}	21,320	**2.780**^{a}	0.176
	(1.696)	(2.144)	(5.431)	(7.379)	**(18,480)**	(35,070)	**(1.064)**	(1.473)
EDC	**-6.470**^{b}	-0.646	-4.545	4.011	**-54,950**^{c}	-42,420	**-3.765**^{b}	-0.086
	(2.554)	(3.820)	(8.752)	(13.432)	**(28,350)**	(54,350)	**(1.901)**	(2.953)
Artificial	-0.797	**-1.596**^{b}	-2.646	-5.304	-12,170	**-23,230**^{c}	-0.523	**-1.274**^{b}
	(0.572)	**(0.779)**	(2.413)	(4.263)	(8,150)	**(12,490)**	(0.370)	**(0.621)**
Artificial*ELF		**6.366**^{b}		12.053		31,320		**4.489**^{b}
		(2.890)		(10.490)		(42,010)		**(1.964)**
Artificial*EDC		**-9.649**^{b}		-12.837		-14,970		-5.845
		(4.912)		(17.941)		(67,180)		(3.708)
Colonial	0.959	0.831	**7.132**^{b}	**6.961**^{b}	**16,740**^{c}	**16,240**^{c}	**1.016**^{b}	**0.922**^{b}
	(0.762)	(0.748)	**(3.587)**	**(3.524)**	**(9,145)**	**(9,356)**	**(0.487)**	**(0.469)**
Climate	0.870	1.126	0.380	0.683	-840.4	-699.5	0.476	0.622
	(0.657)	(0.686)	(3.082)	(3.032)	(7,398)	(8,408)	(0.474)	(0.477)
Constant	-0.690	-0.420	-3.355^c	-2.499	-11,010^c	-7,347	-0.048	0.222
	(0.537)	(0.580)	(1.939)	(2.167)	(6,314.)	(7,795)	(0.395)	(0.420)
Observations	122	122	122	122	122	122	122	122

Table 13: **Political Outcome Variables**

Ordinary least squares regressions using political variables from Kaufmann et al. (2004). Robust standard errors in parentheses. a, b, and c indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) **Checks on Power Accountability**	(2) **Checks on Power Accountability**	(3) **Political Stability**	(4) **Political Stability**	(5) **Government Effectiveness**	(6) **Government Effectiveness**
ELF	**-1.829**^{a}	**-1.246**^{a}	**-2.413**^{a}	**-1.590**^{a}	**-2.051**^{a}	**-1.352**^{a}
	(4.28)	**(2.67)**	**(6.17)**	**(3.94)**	**(5.53)**	**(3.79)**
EDC	1.014	0.696	**1.714**^{b}	**1.303**^{c}	**1.571**^{b}	0.939
	(1.26)	(0.84)	**(2.38)**	**(1.76)**	**(2.22)**	(1.43)
Colonial		**-0.496**^{b}		**-0.442**^{b}		**-0.524**^{a}
		(2.21)		**(2.34)**		**(2.69)**
Climate		0.070		-0.004		-0.170
		(0.31)		(0.02)		(0.91)
Constant	0.434^a	0.498^a	0.539^a	0.431^a	0.447^a	0.627^a
	(4.41)	(3.12)	(6.35)	(3.43)	(4.66)	(3.97)
Observations	175	138	175	138	175	138
R-squared	0.21	0.20	0.33	0.26	0.23	0.28
	(7) **Regulatory Quality**	(8) **Regulatory Quality**	(9) **Rule of Law**	(10) **Rule of Law**	(11) **Corruption**	(12) **Corruption**
ELF	**-2.248**^{a}	**-1.585**^{a}	**-2.163**^{a}	**-1.340**^{a}	**-2.083**^{a}	**-1.329**^{a}
	(5.27)	**(3.61)**	**(5.68)**	**(3.60)**	**(5.82)**	**(3.93)**
EDC	**2.091**^{a}	**1.581**^{b}	**1.480**^{b}	1.060	**1.575**^{b}	**1.002**^{c}
	(2.77)	**(2.20)**	**(2.08)**	(1.62)	**(2.36)**	**(1.70)**
Colonial		**-0.560**^{b}		**-0.536**^{a}		**-0.467**^{b}
		(2.59)		**(2.75)**		**(2.42)**
Climate		-0.037		-0.257		-0.243
		(0.18)		(1.44)		(1.48)
Constant	0.417^a	0.560^a	0.481^a	0.548^a	0.454^a	0.560^a
	(4.26)	(3.54)	(5.13)	(3.45)	(4.56)	(3.25)
Observations	171	138	175	138	171	138
R-squared	0.22	0.25	0.27	0.29	0.25	0.27

Table 14: **Robustness I: Continent Dummies**

All countries and conflicts are included in these Tobit regressions, but in addition we include continent dummies. Robust standard errors in parentheses. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

Panel A: Using COW data										
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
	Number of Conflicts	**Number of Conflicts**	**Total Duration**	**Total Duration**	**Average Duration**	**Average Duration**	**Total Casualties**	**Total Casualties**	**Average Casualties**	**Average Casualties**
ELF	**6.906**^{a}	**4.381**^{a}	**34.392**^{a}	**18.921**^{b}	**13.169**^{a}	1.833	**208,500**^{a}	**148,600**^{b}	**57,740**^{a}	**28,050**^{c}
	(1.440)	**(1.314)**	**(7.842)**	**(8.130)**	**(3.919)**	(4.936)	**(72,540)**	**(62,470)**	**(16,700)**	**(16,570)**
EDC	**-6.055**^{a}	**-5.962**^{a}	**-26.163**^{b}	**-23.255**^{c}	-7.724	-4.103	**-221,800**^{b}	**-221,000**^{b}	**-61,750**^{a}	**-58,790**^{b}
	(1.993)	**(2.086)**	**(12.396)**	**(13.746)**	(7.005)	(7.869)	**(88,740)**	**(88,530)**	**(23,840)**	**(25,140)**
Colonial		0.479		8.348		6.031		19,080		9,079
		(1.071)		(6.076)		(3.768)		(39,170)		(12,750)
Climate		0.970		3.379		1.865		7,522		4,578
		(0.675)		(3.964)		(2.209)		(17,450)		(6,884)
Constant	-1.477^{b}	-0.508	-10.279^{a}	-9.266	-5.271^{b}	-4.581	-60,080^{b}	-37,740	-18,593.320^{b}	-8,225.034
	(0.688)	(1.166)	(3.756)	(6.120)	(2.152)	(3.143)	(28,470)	(46,010)	(7,617)	(15,610)
Observations	186	138	186	138	186	138	186	138	186	138
Panel B: Using MEPV data										
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
	Number of Conflicts	**Number of Conflicts**	**Total Duration**	**Total Duration**	**Average Duration**	**Average Duration**	**Total Casualties**	**Total Casualties**	**Average Casualties**	**Average Casualties**
ELF	**6.021**^{a}	**3.908**^{a}	**44.246**^{a}	**29.581**^{a}	**24.888**^{a}	**15.256**^{a}	**699,200**^{a}	**523,700**^{a}	**396,700**^{b}	**306,500**^{b}
	(1.071)	**(1.099)**	**(7.214)**	**(7.236)**	**(4.831)**	**(5.112)**	**(216,000)**	**(198,300)**	**(158,100)**	**(137,400)**
EDC	**-5.868**^{a}	**-5.978**^{a}	**-37.030**^{a}	**-36.280**^{a}	**-19.533**^{b}	**-17.386**^{c}	**-717,100**^{a}	**-764,900**^{a}	**-431,100**^{b}	**-448,900**^{b}
	(1.673)	**(1.738)**	**(12.042)**	**(13.264)**	**(7.782)**	**(9.102)**	**(261,800)**	**(285,900)**	**(196,000)**	**(204,600)**
Colonial		0.467		**9.390**^{c}		**7.073**^{b}		11,470		46,160
		(0.620)		**(5.081)**		**(3.124)**		(91,870)		(55,990)
Climate		0.892		4.666		1.299		115,700		20,700
		(0.578)		(3.979)		(2.349)		(77,150)		(34,850)
Constant	-1.257^{b}	-0.460	-13.531^{a}	-13.951^{b}	-8.590^{a}	-9.433^{a}	-201,700^{b}	-116,200	-133,700^{b}	-114,800
	(0.606)	(0.758)	(4.273)	(5.793)	(2.344)	(3.259)	(87,170)	(116,200)	(64,720)	(86,090)
Observations	186	138	186	138	186	138	186	138	186	138

Table 15: **Robustness II: Only Conflicts from Fearon and Laitin (2003)**

All regressions are Tobit regressions. This table only considers those conflicts in the COW database that are also included in Fearon and Laitin (2003). Robust standard errors in parentheses. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) Number of Conflicts	(2) Number of Conflicts	(3) Average Duration (Fearon def)	(4) Average Duration (Fearon def)	(5) Average Duration (COW def)	(6) Average Duration (Cow def)	(7) Total Casualties	(8) Total Casualties	(9) Average Casualties	(10) Average Casualties
ELF	**5.002**^{a}	**3.273**^{a}	**33.273**^{a}	**17.329**^{b}	**24.967**^{a}	**12.110**^{c}	**253,100**^{a}	**173,600**^{b}	**123,700**^{a}	**77,250**^{b}
	(1.091)	**(1.211)**	**(7.901)**	**(8.508)**	**(6.581)**	**(6.866)**	**(93,970)**	**(77,170)**	**(33,410)**	**(31,350)**
EDC	**-5.369**^{a}	**-4.705**^{b}	**-26.813**^{c}	-18.855	-19.057	-12.814	**-282,200**^{b}	**-249,000**^{b}	**-129,100**^{b}	**-108,800**^{c}
	(1.884)	**(2.144)**	**(14.905)**	(17.288)	(12.242)	(14.159)	**(128,900)**	**(124,600)**	**(53,600)**	**(56,070)**
Colonial		0.607		**6.413**^{c}		4.676		40,360		19,680
		(0.522)		**(3.801)**		(3.161)		(30,370)		(14,320)
Climate		0.238		2.811		3.170		-3,984		2,481
		(0.515)		(3.924)		(3.187)		(26,820)		(13,800)
Constant	-1.921^{a}	-1.398^{a}	-16.978^{a}	-13.726^{a}	-13.809^{a}	-11.078^{a}	-109,700^{a}	-92,530^{b}	-57,460^{a}	-46,090^{a}
	(0.414)	(0.473)	(3.555)	(3.880)	(3.243)	(3.397)	(37,310)	(36,710)	(12,540)	(13,400)
Observations	186	138	186	138	186	138	186	138	186	138

Table 16: **Robustness III: Post 1990 Conflicts**

All regressions are Tobit regressions. This table only considers those conflicts that began after 1990 so as to address the potential endogeneity of our ethnic variables. Robust standard errors in parentheses. a, b, and c indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) **Number of Conflicts**	**(2)** **Number of Conflicts**	**(3)** **Total Duration**	**(4)** **Total Duration**	**(5)** **Average Duration**	**(6)** **Average Duration**	**(7)** **Total Casualties**	**(8)** **Total Casualties**	**(9)** **Average Casualties**	**(10)** **Average Casualties**
ELF	**5.717**^{a}	**3.360**^{b}	**18.793**^{a}	**10.659**^{b}	**11.613**^{a}	5.364	**72,330**^{a}	**48,620**^{b}	**31,690**^{a}	**19,940**^{b}
	(1.464)	**(1.540)**	**(4.238)**	**(4.704)**	**(2.887)**	(3.270)	**(23,700)**	**(21,420)**	**(11,270)**	**(9,740)**
EDC	**-4.052**^{c}	-2.695	**-13.551**^{c}	-9.267	-6.751	-3.223	**-61,680**^{b}	-49,430	**-24,810**^{c}	-18,880
	(2.278)	(2.510)	**(7.660)**	(8.596)	(5.426)	(6.335)	**(30,830)**	(32,440)	**(13,760)**	(14,410)
Colonial		0.449		2.919		2.370		7,268		4,153
		(0.771)		(2.268)		(1.646)		(8,363)		(3,786)
Climate		0.263		0.720		0.651		558.9		984.6
		(0.682)		(2.313)		(1.693)		(8,828)		(4,137)
Constant	-2.926^a	-2.025^a	-9.219^a	-7.067^a	-6.626^a	-5.193^a	-38,100^a	-30,480^a	-17,590^a	-14,510^b
	(0.686)	(0.707)	(1.540)	(1.788)	(1.211)	(1.325)	(10,940)	(11,150)	(5,725)	(5,670)
Observations	186	138	186	138	186	138	186	138	186	138