HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division

Department Economics

Committee

have examined a dissertation entitled

"Three Essays in Ethnicity, Conflict and the Political Economy of Development"

presented by ***Janina Matuszeski***

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

*Signature*____________________

Typed name Alberto Alesina, Chairman

*Signature*____________________

Typed name Michael Kremer

*Signature*____________________

Typed name Robert Barro

Date: April 25, 2007

Three Essays in Ethnicity, Conflict and the Political Economy of Development

A dissertation presented

by

Janina Matuszeski

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

May 2007

Dissertation Advisors: Professor Alberto Alesina
Professor Robert Barro
Professor Michael Kremer

Author: Janina Matuszeski

Three Essays in Ethnicity, Conflict and the Political Economy of Development

Abstract

This thesis presents three chapters concerning the political economy of developing nations, including economic, political and conflict factors. The first two chapters focus on the role of ethnic diversity and ethnic geography in countries' development and civil war tendencies, while the third chapter considers the potential impact of foreign aid inflows.

In the first chapter, 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. Results for the ELF index are robust to a panel regression format, in which we control for GDP per capita.

In the second chapter, we consider the issue of artificial states, which are countries in which the political borders do not coincide with a division of nationalities desired by the people on the ground. We propose and compute for all countries in the world two new measures of the degree to which states are artificial. One index measures how borders split ethnic groups into two separate adjacent countries. The other index measures the straightness of land borders, under the assumption that the straight land borders are more likely to be artificial. We show that these two measures are highly correlated with several measures of political and economic success.

In the final chapter, we provide empirical evidence that the correlation between U.S. foreign aid and anti-U.S. terrorism is very small in magnitude. The correlation is significant and positive, and is stronger for military aid than for economic aid. Since military aid can strengthen a recipient country's government, this result lends credence to mechanisms in which support for unpopular governments leads to anti-U.S. sentiments. Our results are robust to several specifications and the use of instrumental variables.

Contents

[1]This chapter is co-authored with Frank Schneider We would like to thank Global Mapping International for granting us access to their World Language Mapping System database.

[2]This chapter is co-authored with Alberto Alesina and William Easterly.

[3]This chapter is co-authored with Jennifer Stack.

List of Tables

List of Figures

Acknowledgments

I thank my advisor and thesis committee chair, Professor Alberto Alesina, for his continuous support for my research efforts and his excellent advice. His humor, upbeat encouragement and focused interest at crucial stages in my research helped me to persevere and enjoy the process. Without him, this thesis would not have been possible. I would also like to thank Professor Michael Kremer, for excellent guidance regarding my research, and advice concerning the field of international development. I greatly appreciate his attention to my specific situation and interests. Finally, I wish to thank Professor Robert Barro for his thoughtful advice and his support of my research. His work on religion and growth has been an inspiration for significant portions of my research.

I thank William Easterly and Alberto Alesina for excellent guidance and advice during our work together, crafting a paper on Artificial States. I have learned a lot about finding research topics that are interesting and relevant to the world today, while at the same time drawing on the lessons and examples of history.

Albert Bravo-Biosca and Mario Gamboa-Cavazos join my co-authors Frank Schneider and Jennifer Stack in forming a core friendship group throughout my PhD experience. I deeply appreciate all the good times and the humor, sympathy, support and caring. These true friends are amazing people! Both Frank and Jen have been excellent co-authors. And while we had our disagreements about which's and that's or asterisk's and letter's, the big issues were always fun to discuss. It was extremely rewarding to work with such professional and entertaining co-authors and it was a wonderful thing to work closely with good friends. Frank, thanks for being an inspiration to us all, in hanging in through thick and thin, and for being a great friend throughout. Jen, it is a special thing to find a super-smart, hysterically funny girl pal. I truly treasure our friendship.

Albert Bravo-Biosca offered cruelly tough editing advice that was necessary and very helpful. He also kept us all well aware that everything was tastier, lovelier and cheaper in Barcelona. Albert, thanks for being my late-night buddy and for always showing up to events even later than me. Mario Gamboa-Cavazos kept us all laughing and helped me keep perspective in the middle of difficult times. Thank you all very, very much.

I thank Global Mapping International for granting us access to their excellent World Language Mapping System database. For much needed help with map data and GIS mapping software, I thank Bonnie Burns, Patrick Florance, Guoping Huang, Kimberly Karish, Michael Oltmans and the staff at the Harvard Map Collection.

In addition to the above, I would like to thank Siva Anantham, Silvia Ardagna, Jean Marie Baland, Aron Balas, Ernesto dal Bo, Gary Chamberlain, David Clingingsmith, Dave Evans, Martin Feldstein, Jeffrey Friedman, Bryan Graham, Jason Hwang, Ashley Lester, Marcelo Moriera, and Phillip Schnabl for useful feedback on earlier versions of these papers. I would also like to thank participants at the National Bureau of Economic Research Political Economy Workshop and the Harvard Economics Department Workshops (Macroeconomics, Development Economics, and Economics of National Security) for useful comments and suggestions.

Support from Harvard University, the National Bureau of Economic Research, and the Oppenheimer Graduate Fellowship Fund is gratefully acknowledged.

I would like to thank my boyfriend Michael Oltmans. This has been an amazing journey and I am so happy to have spent the last part of it with you. Thanks for all the advice and the invaluable computer assistance. It's been great fun trading stories and helping each other with research, even from very different fields. Thanks most of all for being there, for supporting me through thick and thin, and for being my companion on this journey. I love you very much.

Finally, I would like to thank my family, including my mother Mary, father Bill, brothers Thaddeus and Adam, and sisters-in-law Gargi and Margaret for loving support throughout my graduate career. I turned to you for advice and sympathy, and you came through with insights, humor, and loving support. I admire and love each of you very much.

To my father and mother,

who gave me an appreciation for living life to the fullest

and always supported me in all my endeavors.

Chapter 1

Patterns of Ethnic Group Segregation and Civil Conflict[1]

1.1 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. But in spite of extensive anecdotal evidence, the 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 be-

[1]This chapter is co-authored with Frank Schneider We would like to thank Global Mapping International for granting us access to their World Language Mapping System database.

longing to each ethnic group are clustered together or dispersed throughout the country.[2] 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.[3] 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.

The literature in this area has focused mainly on the interaction of ethno-linguistic fractionalization and civil war, and has not considered the segregation of ethnic groups. This work on the relationship between fractionalization and civil conflict has not led to a consensus as to the direction of the effect of ethnic diversity on civil conflict, let alone the magnitude. Researchers have also considered ethnic dominance, which is defined as the size of the largest ethnic group or the largest two to three ethnic groups.[4] Studies which have found no effect of ethnic fractionalization 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

[2]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.

[3]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 1.2 and 1.3.

[4]With regard to the relationship between ethnic dominance and ethnic fractionalization, one large ethnic group corresponds to a high value of ELF, close to 1; whereas two to three large groups corresponds to values of ELF closer to 0.6, which is in the middle 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 middle values of ELF associated with the highest levels of conflict.

no effect of ethnic dominance. Collier and Hoeffler (2002b) also 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 inclusion criteria for 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. Our map-based methodology can easily be extended to calculate sub-national indices, and we present preliminary evidence concerning the importance of the level of ethnic diversity in the area of the conflict.

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 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 check if our results hold when we apply several different 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

[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.

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]

Ethnicity can be described along 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 datasets 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 1.4. Several hypotheses rely on the idea that

[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.

[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 1.3.1 currently prevent our calculating the indices for a small subset of countries.

ethnicity can be used to coordinate or enforce coalitions which then 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 1.5, 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 1.2 and 1.3, we consider the evidence using the EC index to be preliminary. Based on these results, we find empirical support for three of our four hypotheses, including that higher levels of ethnic diversity and higher levels of ethnic clustering are both associated with more civil conflicts. We also find that more diversity is associated with longer wars. However, in contradiction to our fourth 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, continent dummy variables, and the level of religious tension. 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] To control for GDP, which is endogenous, we use a decade, panel data format and include GDP at the beginning of each decade as a control. The ELF index is robust to this specification, indicating that our results for ethnic diversity are robust to controlling for the country's GDP level, or in other words its level of economic development. However, the coefficient for the EDC index loses significance when we control for GDP in this manner.

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

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

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.

1.2 Ethnic Geography Indices

We compute three separate indices which measure combinations of two aspects of ethnic geography, 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. An ELF index constructed using earlier data from the 1960's[13] has been used extensively in the literature to date. However we compute the ELF index for many more countries. 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). To the best of our knowledge, we are the first to calculate this index for ethnic groups within countries. 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 homogeneous (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. To interpret our results for the EDC index as measuring the effect of clustering, we also include the ELF index in the regressions to control for the ethnic diversity aspect of the EDC index. Because

[13]The most commonly used ELF index to date is based on ethnic groups described in the Atlas Narodov Mira (Bruk and Apenchenko (1964)).

of difficulties interpreting the EC index for very or perfectly homogeneous countries, we present only preliminary results for this index in Section 1.5, 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 next 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 has been described and utilized extensively in the literature. Specifically, an earlier version of the ELF index was calculated based on tabular data for the 1960's, from the Russian Atlas Narodov Mira (Bruk and Apenchenko (1964)). The ELF index is calculated based on the population of each ethnic group in the country as a whole and 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.

$$ELF = 1 - \sum_{l=1}^{L} \left(\frac{n_l}{N}\right)^2 \tag{1.1}$$

We also construct a 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 given local area by using the population at nearby points (n_q) and a weighting proximity factor ($\phi(p,q)$) that gives greater weight to a point q if it is close to point p, and less weight if it is far from point p. The general formula is given by:

$$\widetilde{n}_{lp} = \frac{\int \phi(p,q) n_{ql} dq}{\int \phi(p,q) dq} \tag{1.2}$$

where ($\phi(p,q)$) is the proximity function that takes a higher value if the points p and q are closer to one another.

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 around point p 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]

$$\widetilde{n}_{lp} = \frac{\int_{q \in R_p} n_{lq} dq}{A} \tag{1.3}$$

[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 choose a 50 kilometer benchmark. Having created the software architecture and the 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.

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. In a similar manner, we define and calculate the total population in a local region of each point (1.4) and a value for the ELF index in the local region of each point (1.5).

$$\widetilde{N}_p = \sum_{l=1}^{L} \widetilde{n}_{lp} \tag{1.4}$$

$$elf_p = 1 - \sum_{l=1}^{L} \left(\frac{\widetilde{n}_{lp}}{\widetilde{N}_p} \right)^2 \tag{1.5}$$

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

$$EDC = \frac{\sum_p N_p elf_p}{\sum_p N_p} \tag{1.6}$$

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, as measured by this local ELF index, the EDC index reflects the average diversity that citizens in the country experience.

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. 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 more 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 the formulas below.[16]

$$\widetilde{E}_p = -\sum_{l=1}^{L} \left(\frac{\widetilde{n}_{lp}}{\widetilde{N}_p}\right) \log_L \left(\frac{\widetilde{n}_{lp}}{\widetilde{N}_p}\right) \tag{1.7}$$

$$E = -\sum_{l=1}^{L} \left(\frac{n_l}{N}\right) \log_L \left(\frac{n_l}{N}\right) \tag{1.8}$$

$$EC = \widetilde{H} = 1 - \frac{1}{NE} \sum_{p} \widetilde{E}_p N_p \tag{1.9}$$

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 (completely un-segregated). In this situation, the average of the local entropies will be equal to the country-wide entropy,

[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 ELF index.

[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.

so the second term in Equation (9) will be one and the EC will be equal to zero. Thus, an EC index of zero is associated with a completely integrated (non-segregated) country.

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, for a highly segregated country, the EC index will be very close to one. In general, the higher the value of the EC index, the greater the clustering (or segregation). Note that this is the opposite direction from that of the EDC index, where a higher value of the 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 1.2, 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 footnote 14.)

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 index is less intuitive than the interpretation of the EDC index, which is the average of the local level of diversity.

In addition, the EC measure is undefined for perfectly homogeneous countries, that is countries with one ethnic group, which are also countries that have an ELF index of 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.

1.3 Data Description and Methodology

1.3.1 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 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

[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/.

[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.

measures approximately 5 kilometers on a side.[21] Combining these two sources of data, our estimated language group populations for countries around the world approximately apply to the period of 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 and is assigned a population number based on the population data from the underlying square of the GPW data. Each dot is also assigned a language group identity, derived from the language group for the area in which the dot is located.[23]

Figure 1.1 illustrates this for a small portion of Senegal. Three language areas are shown, with the borders between the three areas designated by a heavy black line. The area in the upper left corner is Woluf-speaking, while the area in the middle is Serer-Sine-speaking and the bottom portion is an area where both languages are spoken. The figure also shows a grid of small squares. These squares represent the population density data; squares which contain a higher population are colored darker. Next, a grid of dots is superimposed on the population grid data and the language area data. Each dot has two pieces of information tagged to it: (1) a population based on the population of the underlying square, and (2) an ethnic group for that population, based on the language region in which the dot is located. Thus the information contained in the dots is based on the information in the underlying population and language maps. For example, at Example Point 1 (triangle), which is in the Serer-Sine region, the population at that point is designated as 4288 based on the square in

[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 multiple language dots in proportion to the share of each language in the *total* population of the country.

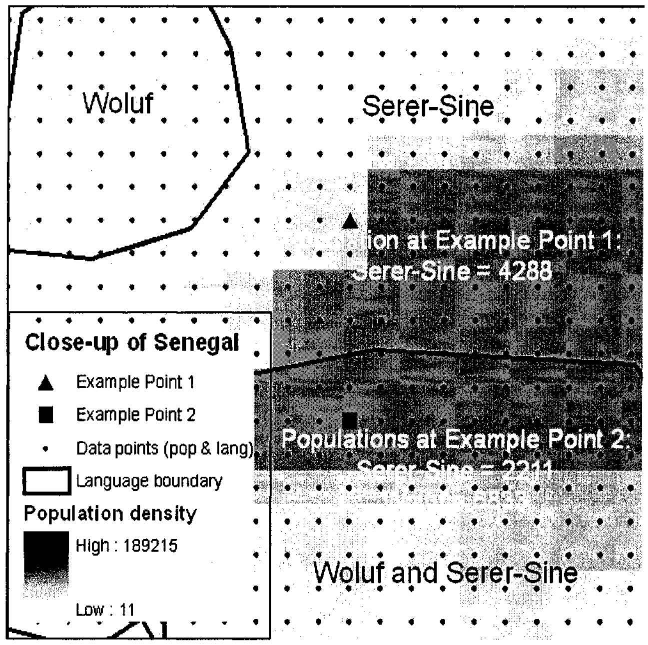


Figure 1.1: Senegal Points Closeup

which the dot is located, while the ethnic group tagged to that point is Serer-Sine. Example Point 2 (square) is slightly more complicated, as it is in an area of mixed languages. Here the total population at that location, 9044 people, is divided between the two language groups based on their relative populations in the country as a whole. As a result, at Example Point 2 there are two dots, one dot representing 2211 Serer-Sine-speaking people and one dot representing 6833 Woluf speakers.

(For the mixed language area, only one of the two dots at each point is shown, since they have the same location.) The collection of dots for a whole country serves as the estimated population/ethnic data on which the remainder of our analysis is based.

Table 1.1: **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 country, 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 Uppsala 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	Regulatory quality index	Kaufman Kray (2004)
Rule of law	Rule of law index	Kaufman Kray (2004)
Corruption	Corruption index	Kaufman Kray (2004)

Because we consider every one living at a location to be of the language group (or groups) that are designated in our language map data, and we ignore any minority populations that don't appear in the language map at this location, the dots we construct represent an estimate of the language group(s) and population of the actual citizens living at each location. Still, the language and population map data sources are remarkably detailed, so we are confident that there is useful information contained in the map of dots that we construct.

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

For the EDC index, we require detailed information on the language population in the surrounding area of each location. We divide the country into a grid of the same size and location as our map of dots. (So, each square of our new grid has one of our dots in its center.) For each square in the grid, we designate a circular area of 50 kilometers in radius as the "local area" within which the typical citizen will travel frequently and experience the ethnic diversity of the area. For each local area around a square, we consider the dots located in this local area, and calculate the total number of people from each language group and the fraction of each language group in the total population of the area. From this, we calculate the local elf index for each point. By performing this calculation repeatedly, we generate a map which has a local elf index value for each grid square in the country. The EDC index is calculated by taking a population-weighted average of these local elf index values, across the entire country.

Figures 1.2 through 1.5 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 language groups in Senegal. For example, in Senegal there are about a dozen smaller areas where a different language group is contained within a larger language group areas (Figure 1.2). With less segregation (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 (segregation) 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 1.3 and 1.5). 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.

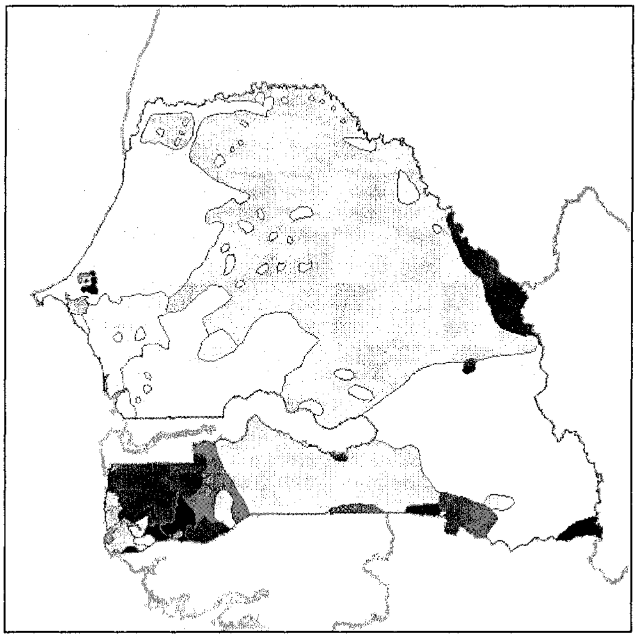


Figure 1.2: Senegal Languages - Each language is designated with a different shade of grey.

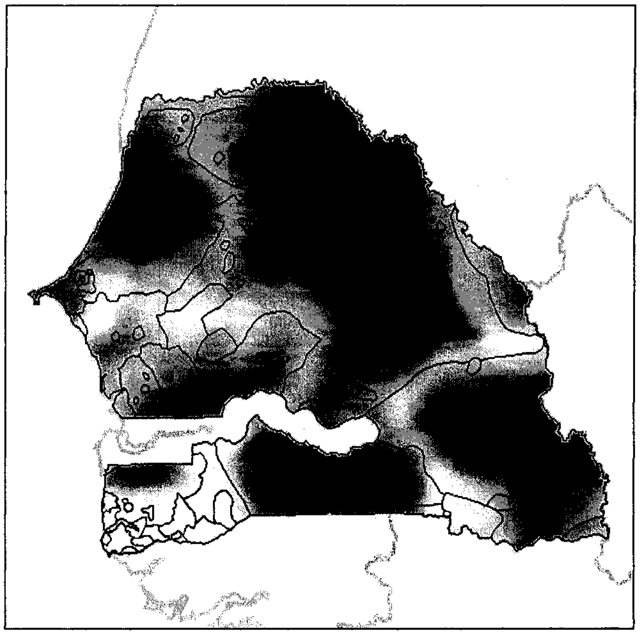


Figure 1.3: Senegal Local Elf Index - Lighter areas correspond to higher diversity areas, while darker areas have less diversity. Black outlines of the different language areas have been superimposed on this map of the local elf index.

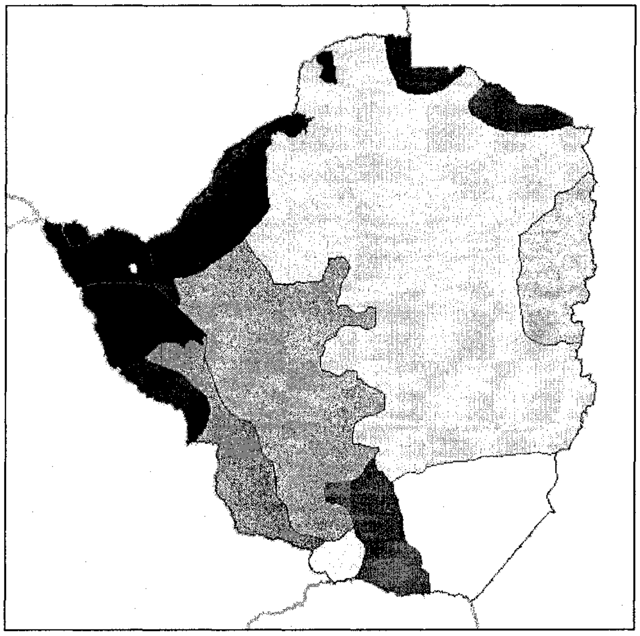


Figure 1.4: Zimbabwe Languages - Each language is designated with a different shade of grey

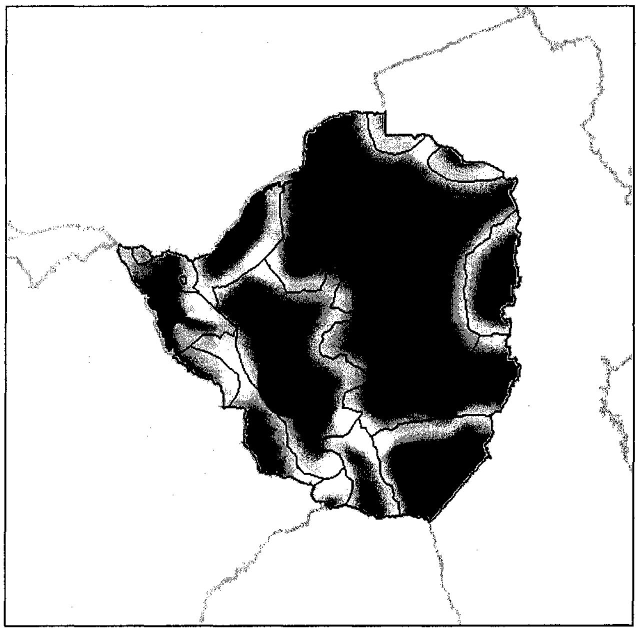


Figure 1.5: Zimbabwe Local Elf Index - Lighter areas correspond to higher diversity areas, while darker areas have less diversity. Black outlines of the different language areas have been superimposed on this map of the local elf index.

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 language group in the local area around each point. Then, using the shares of each language group in each local area, we compute the EC index.[24] The EC index also requires information about the share of each language 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 1.2, 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.

[24]See Section 1.2 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 is principally 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 the omitted countries.

Table 1.2: **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

A major advantage of our map-based approach is that we are not constrained to perform our analysis at the country-level. Thus, in addition to the country-wide indices, we are also able to calculate the ELF index for the area of each conflict.[26] The COW database includes information on the latitude and longitude of the center of each conflict, as well as the approximate radius of the conflict. We use this information to construct an estimate of the conflict area which consists of a circle centered at the reported center of the conflict, with a radius equal to the reported radius of the conflict. We then clip this circle shape to include only the area of the circle that is also within the country in which the conflict takes place. In practice, smaller conflict areas are perfectly circular while larger conflict areas typically have one edge that follows the country's border. We then clip our dots map using these estimated conflict areas and calculate the ELF index of ethnic diversity for each of these conflict areas. This ELF index can then be used in conflict-level analysis and comparisons.

Encouragingly, the country-wide 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 is based on the tabulated populations of each ethnic group from the Atlas Narodov Mira (Bruk and Apenchenko (1964)). The correlation is 0.79 between the ELF60 index and our ELF index (Table 1.2, 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 from the Atlas of the 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.[27]

Given these significant differences, it is reassuring that the correlation is as high as it

[26]In principle we can also calculate the EDC and EC indices for each conflict area. We hope to explore these additional extensions in future work.

[27]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.

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 using a 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 1.3. 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.

Table 1.3: **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

1.3.2 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.[28] 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.[29] The MEPV dataset also has some information, albeit incomplete, on civilian deaths, while the COW data only considers battle combatants.[30]

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 set of conflicts in their analysis.[31] 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

[28] 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).

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

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

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

on the conflicts in each country.[32] 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 group geography, such as high levels of diversity or clustering/interspersion, lead to longer wars. We also calculate the average casualties per conflict for each country and report this variable in level terms and also normalized by 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 1.4, 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, or 3, depending on the number of casualties in the conflict. When combining intensity information across many conflicts for one country, we use the highest value of the COW intensity variable for that country. Thus, we construct eight civil war variables for each country: the number of conflicts; the average duration of conflicts; the total time spent in war (total duration); the average casualties per war, both in levels and normalized by population; the total casualties, both in levels and normalized by population; and the maximum COW intensity rating.

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 had no civil war and thus has a zero for all of its civil war variables. Because of

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

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 1.3. 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 had 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 1.2, Panel C. Civil war outcomes are only somewhat correlated with the ELF index and are 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.

In addition to our country-level cross-sectional analysis, we also consider two other specifications, a conflict-level cross-sectional analysis, and a country-level panel data specification. For these analyses, we construct similar civil war variables to the ones described earlier. For the conflict-level cross-sectional analysis, we calculate three variables for each conflict: the total years of the conflict, the total casualties (battle deaths) for the conflict, and the average casualties per year of the conflict. For the panel data specification, we calculate several civil war variables for each country in each decade. Specifically, we calculate the total years in the decade that the country spent in war; the number of wars that began in that decade; the total number of casualties from civil war in that decade; and a dummy variable for if the country had any civil war in the *previous* decade. We also construct a panel of data for the country's GDP per capita at the beginning of each decade.

1.4 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 primordiality 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 1.5, we test our hypotheses empirically using our newly-constructed ethnic indices and the data on civil wars.

1.4.1 Ethnic Diversity and Civil War Incidence

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.

1.4.2 Ethnic Diversity and Civil War Duration

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 important ethnicity is, the more cohesive the coalition will be over time, and the less likely it is that the conflict will "fade away" 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 the warring parties may come to a settlement more rapidly if ethnicity plays a role in the coalition(s) involved. Collier and Hoeffler (2006) discuss the important role in ending a civil conflict of the ability to lock-in post-conflict settlements. Thus, greater ethnic diversity and, by extension, greater ethnic involvement in the formation of coalitions, could also 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.

1.4.3 Ethnic Clustering and Civil War Incidence

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. This is because coalitions can be doubly strengthened if neighbors are also from the same ethnic group. First, 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. Thus, coalitions of neighbors have potentially more to gain from civil conflict than coalitions of non-neighbors. Second, if neighbors are all of the same ethnic group, then ethnicity can be used to enforce coalitions. Thus, the clustering together of people of one ethnic group is associated with a higher likelihood of feasible coalitions, due to the joint advantages of coalition enforceability and a high potential return to coalition members in the form of local public goods. So, we predict that 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.*

1.4.4 Ethnic Clustering and Civil War Duration

Finally, we look at how the clustering of ethnic groups might affect conflict duration. If a 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 (segregated) ethnic groups. By contrast, interspersed ethnic groups could be associated with protracted wars, as the competing claims to territory might 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.[33] 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 civilian casualties per conflict.*

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 conflicts and fewer civilian casualties per war.

1.5 Empirical Results

1.5.1 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.

[33]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.

Hypothesis 1 concerns the incidence of civil war, which is measured by our "number of conflicts" variable. Results for this variable are presented in Table 1.4. 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. In Columns 4 and 8, we also include regressions using 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 1.4, 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 higher ethnic diversity is associated with an increase in the number of civil conflicts. Since Hypothesis 1 is based on theories of ethnicity increasing the strength of warring coalitions, these empirical results also support the theory 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 1.4, 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. The number of conflicts variable could be biased and/or noisy due to these potentially arbitrary decisions. To address this issue, Table 1.5 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). Thus, our earlier conclusions that increased ethnic diversity is associated with an increase in the incidence of civil war seem to be justified.

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 1.6, 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 opposing factors for ethnically-based coalitions (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 1.7, 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 1.8, 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.

Table 1.4: **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 1.5: **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 1.6: **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) OLS	(2) OLS	(3) OLS	(4) OLS	(5) Tobit	(6) Tobit	(7) Tobit	(8) 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) OLS	(2) OLS	(3) OLS	(4) OLS	(5) Tobit	(6) Tobit	(7) Tobit	(8) 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 1.7: **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 1.8: **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 1.9: **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) Number of Conflicts	(2) Total Duration	(3) Average Duration	(4) Total Casualties	(5) Average Casualties	(6) Total Casualties per 1000	(7) 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

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 1.6 through 1.8. 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.[34] 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.

Results for the equivalent MEPV variables are shown in Table 1.9 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 1.4. 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 insignificant for all the civil war outcome variables.[35] Thus, we do not find strong support for a non-linear relationship between ethnic diversity and civil conflict outcome variables. Although we have chosen not to focus on the non-linear relationship in this research, given that we have estimated population shares for all the language groups in each country, in future work we will be able to directly calculate the ethnic dominance variable (the relative size of the largest ethnic group or groups) and compare our results to those in the literature.

To conclude, for our first two hypotheses, we find strong support for a positive correlation

[34]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.

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

between higher ethnic diversity and a higher incidence of civil conflict, and less support for a correlation with civil war duration. We also find little evidence of a non-linear effect of ethnic diversity on civil conflict.

1.5.2 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 1.5.3. 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 1.4. 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 1.5). When the two main controls are included (Column 4), the EDC index no longer has a significant effect.

In interpreting the result shown in Column 6, where we include only the EDC index, it is helpful to think of two competing factors 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 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 evidence regarding the average duration of civil conflicts, as shown in Table 1.6, 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 civil war (average casualties per conflict - Table 1.7, Panel A; and average casualties per conflict normalized by population - Table 1.8, 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.

Finally, we consider 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 1.6 through 1.8, Panel A).

Results for the MEPV data are broadly in accord with our results for the COW data (Table 1.9). One important difference is that the EDC index 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 (segregation) would be associated with 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, contradicting 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 1.10 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. This result is puzzling and we hope to explore it in future work.

1.5.3 Preliminary EC Index Results

We present preliminary results using the EC index as a measure of ethnic clustering. This index has two caveats. First, it theoretically un-defined for homogeneous countries. (Think what segregation or interspersion even means for a homogeneous country; it is not clear that the concepts are defined.) Homogeneous countries have an ELF index of zero. Second, due to the logarithmic function in the formula, the index is very sensitive to measurement errors when a country has one large ethnic group and a few small ethnic groups. These countries have a very low ELF, since they are relatively un-diverse. Thus, we present preliminary results for the EC index using countries which have an ELF index greater than 0.1. This reduces our sample by between 40-80 countries, depending on the specification, leaving us with a sample of 96-107 countries.[36]

Using this restricted sample in regressions in which we include both the ELF and EC indices (Table 1.11), the coefficients on the ELF index often lose significance, possibly as a consequence of the lower variation in the ELF index. However, the coefficients on the EC index are significant for all of the civil war outcomes, often at the 1 percent level, with the exception of the average duration variable. The coefficients on the EC index are all positive, meaning that more clustering is associated with more civil conflict. Since more clustering is

[36]Several zero- or low-ELF countries drop out of the sample in any case when we include the control variables in the full-sample regressions.

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.

1.5.4 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.[35]

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 that 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 1.12). For two of 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 indices themselves, reflect the effect of ELF and EDC for the non-artificial countries, while the coefficients on the interaction terms reflect the differential impact of these two indices on artificial countries. Since only the interaction terms are significant, having high diversity or high clustering seems to be problematic only for artificial states.[37] However, the coefficients are not significant for the average duration and average casualties outcomes (Columns 3 through 6). Hence these results should be considered indications, but

[37] 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.

Table 1.10: **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 1.11: **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 1.12: **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 1.13: **Political Outcome Variables**

Ordinary least squares regressions using political variables from Kaufmann, Kraay and Mastruzzi (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

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 variables (Table 1.13). Using the same outcomes as Alesina et al. (2006), originally derived from Kaufmann et al. (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.[38] 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 from the authors upon request.

1.5.5 Panel Regression Results

In Sections 1.5.1 and 1.5.2, we control for several important factors in the basic country-level cross-country regressions, including former colonial status and climate. However, we do not include a control for GDP per capita because this variable is endogenous; specifically it can be affected by civil war which affects the standard errors in the regression. To account for GDP per capita, an important factor, we turn to a panel regression format, and include GDP per capita at the beginning of each decade so that it cannot be said to have been influenced by the level of civil conflict in that decade.[39] The panel specification also allows us control for the fact that a country which has had civil conflict in the past may be more likely to have civil conflict again.

We construct several civil war variables including: the total years in the decade that

[38]These results are broadly consistent with the findings of Porta, de Silanes, Shleifer and Vishny (1999).

[39]We also include GDP in the entire preceding decade as a control. Since some countries have missing values early in some decades, and the decade GDP per capita average is calculated over all the years in the prior decade for which there is data, we feel it is more consistent across countries to use the GDP per capita in the final year of the prior decade. However, the results are the same for both methods of controlling for GDP per capita.

the country spent in war; the number of wars that began in that decade; the total number of casualties from civil war in that decade; and a dummy variable for if the country had any civil war in the *previous* decade. We also construct a panel of data for the country's GDP per capita at the beginning of each decade. We present regression results for these three variables in Tables 1.14 through 1.16. In each table, we show Tobit panel random effects regressions for the civil war variable of interest on the ELF and EDC indices as well as several control variables. The ELF and EDC indices, as well as the colonial and climate controls are all non-time-varying, while the GDP per capita variable and the previous-war variable both vary by decade as well as by country.[40]

Results from this panel data specification broadly match the patterns we find in the country-level cross-sectional analysis. Table 1.14 shows results for the number of wars that began in each decade in each country, a variable that roughly corresponds to the "number of conflicts" in the cross-section analysis. Here, both the ELF and EDC indices are significant, with the same sign as in the cross-section regressions. The indices remain significant when we control for former colonial status, climate, GDP per capita, and civil war in the previous decade. Thus it appears that higher ethnic diversity and higher ethnic clustering are both associated with a higher incidence of civil conflicts in a country.

By contrast, the results for the total years in the decade for which there was civil war (Table 1.15) are not always significant for the EDC index, although the ELF index coefficients continue to be significant across the board. This dependent variable broadly corresponds to the "total duration" variable from the cross-section analysis. So the result here is weaker for the EDC index in the panel specification than in the cross-section specification. In particular, the EDC index is no longer significant when we control for GDP per capita. The ELF index is robust to the inclusion of all the controls.

Finally, Table 1.16 shows the regression results when the dependent variable is the total number of battle deaths due to civil conflict in each decade. Here, both the ELF and EDC have a significant effect, implying that higher diversity and higher clustering are associated with a higher number of total casualties due to civil war. These results are robust to the

[40]We rely on a random effects specification so that we can examine the effect of our non-time-varying variables, particularly the ELF and EDC indices.

Table 1.14: **Tobit Panel Random Effects Regressions - Total Number of Wars Beginning in Each Decade (COW)**

Dependent variable: Total number of wars beginning in each decade, in each country, for decades between 1940's and 2000's (2000-2004). GDP data for year prior to decade (eg. 1959) for decades between 1960's and 2000's. All regressions are Tobit, random effects panel regressions. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) Tobit	(2) Tobit	(3) Tobit	(4) Tobit	(5) Tobit	(6) Tobit	(7) Tobit	(8) Tobit
ELF	**4.435**^{a}	**2.855**^{a}	**2.220**^{a}	**1.857**^{a}	**4.155**^{a}	**2.769**^{a}	**2.248**^{a}	**1.901**^{a}
	(0.667)	**(0.678)**	**(0.687)**	**(0.696)**	**(0.690)**	**(0.698)**	**(0.696)**	**(0.701)**
EDC	**-3.613**^{a}	**-3.106**^{b}	**-2.805**^{b}	**-2.892**^{b}	**-3.373**^{a}	**-2.971**^{b}	**-2.814**^{b}	**-2.905**^{b}
	(1.183)	**(1.261)**	**(1.269)**	**(1.271)**	**(1.215)**	**(1.289)**	**(1.269)**	**(1.268)**
Colonial		0.651^{b}		-0.055		0.622^{c}		-0.047
		(0.331)		(0.373)		(0.340)		(0.372)
Climate		0.356		0.437		0.366		0.435
		(0.329)		(0.336)		(0.335)		(0.336)
GDP in prev year			-0.000^{a}	-0.000^{a}			-0.000^{a}	-0.000^{a}
			(0.000)	(0.000)			(0.000)	(0.000)
War in prev decade					0.609^{b}	0.386	-0.071	-0.135
					(0.290)	(0.281)	(0.273)	(0.268)
Constant	-4.130^{a}	-3.753^{a}	-1.377^{a}	-1.204^{a}	-3.971^{a}	-3.630^{a}	-1.366^{a}	-1.183^{a}
	(0.372)	(0.394)	(0.341)	(0.421)	(0.375)	(0.403)	(0.343)	(0.421)
Observations	1302	966	571	494	1116	828	571	494
Countries	186	138	158	134	186	138	158	134

Table 1.15: **Tobit Panel Random Effects Regressions - Years in each Decade Spent in War (COW)**

Dependent variable: Number of years spent in war in each decade, in each country, for decades between 1940's and 2000's (2000-2004). GDP data for year prior to decade (eg. 1959) for decades between 1960's and 2000's. All regressions are Tobit, random effects panel regressions. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) Tobit	(2) Tobit	(3) Tobit	(4) Tobit	(5) Tobit	(6) Tobit	(7) Tobit	(8) Tobit
ELF	**12.781^a**	**7.710^a**	**8.618^a**	**6.396^a**	**9.947^a**	**6.198^a**	**6.651^a**	**5.150^b**
	(1.784)	**(1.990)**	**(2.319)**	**(2.454)**	**(1.725)**	**(1.883)**	**(2.236)**	**(2.339)**
EDC	**-8.538^a**	**-6.579^c**	-6.272	-6.199	**-6.855^b**	-5.302	-5.979	-5.946
	(3.275)	**(3.803)**	(4.347)	(4.571)	**(3.179)**	(3.536)	(4.135)	(4.288)
Colonial		3.063^a		2.192^c		2.363^b		1.443
		(0.949)		(1.201)		(0.924)		(1.200)
Climate		0.450		0.216		0.403		0.461
		(0.989)		(1.191)		(0.922)		(1.125)
GDP in prev			-0.000^a	-0.000^a			-0.000^a	-0.000^a
year			(0.000)	(0.000)			(0.000)	(0.000)
War in prev					7.027^a	6.217^a	5.063^a	4.520^a
decade					(0.754)	(0.760)	(0.859)	(0.861)
Constant	-9.259^a	-8.934^a	-4.600^a	-5.088^a	-9.954^a	-9.499^a	-5.751^a	-5.904^a
	(0.751)	(0.911)	(0.906)	(1.215)	(0.791)	(0.936)	(0.984)	(1.273)
Observations	1302	966	571	494	1116	828	571	494
Countries	186	138	158	134	186	138	158	134

Table 1.16: **Tobit Panel Random Effects Regressions - Total Civil War Casualties in each Decade (COW)**
Dependent variable: Total number of battle deaths due to civil war, in each decade and in each country, for decades between 1940's and 2000's (2000-2004). GDP data for year prior to decade (eg. 1959) for decades between 1960's and 2000's. All regressions are Tobit, random effects panel regressions. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) Tobit	(2) Tobit	(3) Tobit	(4) Tobit	(5) Tobit	(6) Tobit	(7) Tobit	(8) Tobit
ELF	**106,6**[a]	**66,560**[a]	**87,210**[a]	**70,970**[a]	**77,830**[a]	**55,650**[a]	**62,070**[a]	**56,95**[a]
	(17,820)	**(17,800)**	**(21,390)**	**(21,980)**	**(13,080)**	**(13,470)**	**(16,230)**	**(18,09)**
EDC	**-85,750**[a]	**-68,390**[b]	**-78,140**[b]	**-76,870**[c]	**-65,590**[a]	**-57,780**[b]	**-66,040**[b]	**-68,240**[b]
	(32,570)	**(33,510)**	**(39,190)**	**(40,210)**	**(23,690)**	**(25,260)**	**(29,550)**	**(32,820)**
Colonial		3,616		2,806		1,057		2,819
		(8,759)		(10,570)		(6,685)		(8,706)
Climate		19,300[b]		16,490		13,870[b]		7,603
		(8,572)		(11,560)		(6,796)		(9,866)
GDP in prev			-3.147[a]	-2.594[b]			-3.266[a]	-2.966[a]
year			(0.976)	(1.036)			(0.981)	(1.062)
War in prev					44,540[a]	41,530[a]	39,800[a]	33,270[a]
decade					(6,060)	(5,980)	(7,107)	(7,704)
Constant	-85,37[a]	-79,090[a]	-56,270[a]	-61,230[a]	-78,890[a]	-76,270[a]	-58,110[a]	-59,110[a]
	(7,030)	(8,046)	(8,724)	(11,827.541)	(5,765)	(6,875)	(7,763)	(10,520)
Observations	1302	966	571	494	1116	828	571	494
Countries	186	138	158	134	186	138	158	134

Table 1.17: **OLS Conflict Level Regressions Including Conflict-Area ELF Index**

OLS regressions at the conflict level. Dependent variables include: (1) Duration of the conflict, (2) Total number of battle deaths during the conflict and (3) Average number of battle deaths per year of the conflict. *a*, *b*, and *c* indicate statistical significance at the 1%, 5%, and 10% confidence level, respectively.

	(1) Duration of Conflict	(2) Duration of Conflict	(3) Duration of Conflict	(4) Total Casual-ties	(5) Total Casual-ties	(6) Total Casual-ties	(7) Average Casualties per Year	(8) Average Casualties per Year	(9) Average Casualties per Year
Conflict Area ELF	**9.661^a**	**9.529^a**	**7.709^a**	**34,180^a**	**36,690^a**	**30,130^a**	**3,380^b**	**4,096^a**	**3,519^b**
	(2.589)	**(2.585)**	**(2.478)**	**(11,140)**	**(11,17)**	**(9,093)**	**(1,457)**	**(1,481)**	**(1,449)**
Country ELF	**-4.837^c**	**-4.732^c**	-1.483	**-15,630^c**	**-14,45^c**	**-18,850^b**	**-2,765^b**	**-2,622^b**	**-3,917^a**
	(2.616)	**(2.637)**	(2.678)	**(8,452)**	**(7,923)**	**(8,848)**	**(1,396)**	**(1,310)**	**(1,457)**
Colonial (Ctry)		1.248	-0.756		-1,354	-6,322		-1,575	-1,459
		(1.812)	(1.818)		(9,754)	(8,871)		(1,246)	(1,357)
Climate (Ctry)		-1.560	-0.468		-12,030	-1,239		-1,129	-391.1
		(1.837)	(1.738)		(10,260)	(5,110)		(1,103)	(559.1)
GDP in 2002			1.599^b			-2,361			-874.1
(Ctry)			(0.795)			(2,690)			(625.5)
Constant	5.349^a	5.054^a	-8.141	9,484^b	14,130^b	35,720	2,308^b	3,697^b	10,940^b
	(1.274)	(1.423)	(6.510)	(4,795)	(7000)	(21,720)	(935.1)	(1,531)	(4,519)
Num of Confl	206	204	170	207	205	170	207	205	170
R-Squared	0.05	0.05	0.07	0.03	0.05	0.05	0.02	0.04	0.06

inclusion of the GDP per capita and previous war control variables.

In general, the conclusions of the panel regressions broadly mirror the results from the country-level cross section analysis. Correlations of ELF and EDC with civil war incidence and total casualties are robust to the inclusion of GDP per capita and a control for civil war in the prior decade. Results for the total number of years in a decade with war are less robust for EDC, although the ELF index is still significant.

1.5.6 Conflict Level Analysis

Taking advantage of our map-based approach, we calculate an ELF index for the area of each conflict.[41] This allows us to compare various aspects of the conflicts and relate them to the ELF index for the conflict area and the country-wide ELF index. In particular, we consider three variables: the total duration of the conflict; the total casualties (battle deaths) due to the conflict; and the average casualties per year of the conflict.

The results for this analysis are shown in Table 1.17. We show results for the conflict area ELF and the country-wide ELF, as well as controls for former colonial status and climate. We also include GDP per capita in 2002 as a control, although this variable is endogenous, so results for these regressions should be interpreted with caution.[42]

Interestingly, the conflict-level ELF index seems to matter more than the country-level ELF index. Although both variables are significant, the magnitude of the conflict-level ELF coefficients are larger. Also, when the conflict-level ELF index is included, the sign on the country-level ELF index is reversed from our usual results. The interpretation is that a higher ethnic diversity in the conflict area is associated with worse civil war outcomes, but that given this level of diversity in the conflict area, a more diverse country may decrease civil conflict. This effect may be due to the fact that the two ELF variables are highly correlated (0.738). However, future research will explore this result further.

[41]It is straightforward to also calculate the EDC and EC indices for each conflict area. We hope to explore these additional extensions in future work.

[42]Unlike with the country-level regressions, we cannot use a panel data specification with the conflict-level data, and so we cannot properly control for GDP per capita.

1.5.7 Results Summary

To conclude, based on evidence from our country-level cross-section regressions, 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. In terms of the total impact of civil war on a country, 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. We also show evidence that higher clustering, and especially higher ethnic diversity, is associated with worse political outcomes.

Finally, our basic results from our country-level cross-sectional analysis are broadly confirmed by two additional specifications, a country-level panel regression, and a conflict level cross-section regression. Based on results from the former specification, GDP per capita does not appear to change our results for the ELF index, while results for the EDC index are unchanged for two of the three civil war outcome variables.

1.6 Robustness Tests

In keeping with the large variation in the set of conflicts included, across the literature on ethnic diversity and civil conflict, we perform several robustness checks with our data. For the first three robustness checks, 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 1.18). ELF and EDC still have no impact on conflict duration in the COW data, but do have an impact in the MEPV data.

Table 1.18: **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.

	(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: Using COW data										
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) 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	**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 1.19: **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 1.20: **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

Table 1.21: **Robustness IV: Controlling for Religious Tension**

All regressions are Tobit regressions. Dependent variables are various measures of civil war outcomes. 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	(4) Average Duration	(5) Total Duration	(6) Total Duration	(7) Average Casualties	(8) Average Casualties	(9) Total Casualties	(10) Total Casualties
ELF	**5.669^a**	**4.537^a**	6.650	0.511	**24.756^a**	**16.992^c**	**40,490^b**	**26,880^c**	**128,100^a**	**109,800^b**
	(1.512)	**(1.513)**	(5.273)	(6.153)	**(9.086)**	**(9.688)**	**(15,750)**	**(15,410)**	**(40,180)**	**(43,500)**
EDC	**-4.119^c**	**-5.341^b**	0.773	-0.693	-14.035	-17.928	-33,580	**-40,260^c**	**-116,100^b**	**-133,000^b**
	(2.256)	**(2.079)**	(9.744)	(9.349)	(16.186)	(15.875)	(25,010)	**(23,910)**	**(58,710)**	**(59,260)**
Colonial		1.298		6.997^b		9.111^c		16,200^c		15,970
		(0.846)		(3.345)		(4.864)		(9,627)		(20,120)
Climate		1.201^c		2.041		4.201		6,486		20,830
		(0.615)		(2.543)		(3.989)		(5,667)		(13,450)
Religious	-0.094	-0.135	-1.498	-1.637	-1.651	-1.884	-752.4	-1,117	-2,777	-4,343
Tension 1999	(0.182)	(0.178)	(1.264)	(1.215)	(1.532)	(1.489)	(2,709)	(2,546)	(4,799)	(4,548)
Constant	-0.514	-0.811	4.334	2.888	1.935	0.023	-9,486	-13,050	-17,810	-15,780
	(1.191)	(1.345)	(6.321)	(5.941)	(8.535)	(8.704)	(16,670)	(17,870)	(28,830)	(27,720)
Observations	121	113	121	113	121	113	121	113	121	113

In regressions not included in this paper, we restrict our data to big countries and big conflicts.[35] 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. This result suggests that ethnic diversity can lengthen big conflicts, but has no significant effect on the set of all conflicts.

We consider alternative definitions of conflict, and include only the conflicts in the COW database which 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 1.19. 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 indices remain significant when we include population as a control.[35]

We consider the crucial issue of a potential 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.[43] In Table 1.20, 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. However, results for the ELF index are robust to using only these post-1990 conflicts.

A different concern with our analysis is the potential of omitted variables bias. In particular, there could be some factor, such as xenophobia or combatitiveness, that is typical of a country or area. If this omitted factor affects both the pattern of ethnic group geography (particularly segregation or clustering) and also the tendency towards civil conflict, then our results will be biased. To attempt to control for this omitted variable, we include the level of religious tension in the country as a control in the cross section regressions. We use the ICRG's measure of religious tension in 1999, although our results are similar (while our sample is smaller) when we use the level of religious tension in 1985, the first year for which the index is available. The results for these regressions are shown in Table 1.21 and are broadly similar to our basic regression results. Thus it does not appear that the omitted variable of religious tension is an important factor in this analysis.

To address both the potential reverse causality and omitted variables concerns, in future work, we hope to instrument for the ELF index using measure of genetic variety within a country.[44] 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. Thus, much work must be done to make this rich genetic data useable for our purposes. 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

[43]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.

[44]The use of genetic variation as an instrument is inspired in part by Spolaore and Wacziarg (2006).

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.

1.7 Conclusions

We construct a new index of ethnic diversity and clustering, the EDC index, using digital map data for the location of language groups around the world and for population density. We 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. We also construct a preliminary version of an EC index which measures only ethnic clustering.

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 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, while the coefficient on the EDC index provides an estimate of the effect of the clustering of ethnic groups. Based on the 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.

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.

Finally, we consider several different specifications including a country-level panel specification which allows us to control for GDP per capita. Results for the ELF index are analogous to those for the cross-country regression, while results for the EDC index are mostly analogous. We also compare the ELF index for the conflict area alone and find that it is significantly correlated with the duration of a conflict and the number of casualties.

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 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. Additional possible extensions include regional (sub-national) analysis and more extensive analysis of political and economic outcome variables.

Chapter 2

Artificial States[1]

2.1 Introduction

Artificial states are those in which political borders do not coincide with a division of nationalities desired by the people on the ground. Former colonizers or post war agreements among winners regarding borders have often created monstrosities in which ethnic, religious or linguistic groups were thrown together or separated without any respect for those groups' aspirations. Eighty per cent of African borders follow latitudinal and longitudinal lines, and many scholars believe that such artificial (unnatural) borders, which create ethnically fragmented countries or, conversely, separate the same people into bordering countries, are at the roots of Africa's economic tragedy.[2] Not only in Africa, but around the globe including Iraq and the Middle East, failed states, conflict and economic misery are often very visible near borders left over by former colonizers, borders which bore little resemblance to the natural division of peoples.

There are three ways in which those who drew borders created problems. First they gave territories to one group ignoring the fact that another group had already claimed the same territory. Second, they drew boundary lines which split ethnic (or religious or linguistic)

[1]This chapter is co-authored with Alberto Alesina and William Easterly.

[2]See Easterly and Levine (1997) for early econometric work on this point. Herbs (2000) and especially Englebert, Tarango and Carter (2002) focus on the arbitrariness of African borders as an explanation of political and economic failures in this region. At the time of decolonization, new rulers in Africa made the decision to keep the borders drawn by former colonizers to avoid disruptive conflicts among themselves.

groups into different countries, frustrating the national ambitions of various groups and creating unrest in the countries formed. Third, they combined into a single country groups that wanted independence. The results can be disastrous. Artificial borders increase the motivation to safeguard or advance nationalist agendas at the expense of economic and political development. When states represent combinations of peoples put together by outsiders, these peoples may find it more difficult to reach consensus on public goods delivery and the creation of institutions that facilitate economic development, compared to states that emerged in a homegrown way. Peoples may find their allegiance to various collective agendas more divided in artificial states than in non-artificial states. As George Bernard Shaw eloquently put it "A healthy nation is as unconscious of its nationality as a healthy man is unconscious of his health. But if you break a nation's nationality it will think of nothing else but getting it set again."

While the nature of borders has been mentioned in the political science (especially) and economic literature, we are not aware of systematic work relating the nature of country borders to the economic success of countries. Our goal is to provide measures which proxy for the degree to which borders are natural or artificial, and relate these measures to economic and political development. By "artificial", we mean a political border drawn by individuals not living in the areas divided by these borders, normally former colonizers. All other borders can be considered "natural", as they were drawn by people on the ground. Needless to say, often borders may start as artificial and then be modified by people on the ground. Of course, these adjustments on the ground may or may not reflect the desire of a majority of the people living there especially if dictatorial regimes make the adjustments.

We provide two measures never before used in econometric analysis of comparative development. One is relatively simple and captures whether or not an ethnic group is "cut" by a political border line. That is, we measure situations in which the same ethnic group is present in two bordering countries. This measure accounts fairly precisely for one of the ways in which borders may be "wrong", that is when they cut through groups and leave them in separate countries. But it does not capture other ways in which borders may be undesirable; for instance situations in which two ethnic groups are forced into the same country. We then provide a second measure, based upon the assumption that if a land border is close to a straight line it is more likely to be drawn artificially, for example by former colonizers. However, if it is relatively squiggly it is more likely to represent geographic features (rivers, mountains etc.) and/or divisions carved out in time to separate different people. This second measure probably comes closer to capturing instances in which lines drawn at former colonizers' tables have remained in place. (The first measure may also capture adjustments of borders on the ground that do not reflect an appropriate division of people on the ground.) Needless to say, the straight-border measure is not perfect, but much of our paper concerns precisely discussing this measure and its alternatives. It turns out that these two new measures are in fact not highly correlated, implying that they do capture different aspects of artificiality. Thus, we define artificial states as those that have straight borders and/or a large fraction of their population belonging to a group(s) split with a neighboring country.

In many ways the main contribution of the paper is to provide two new measures of borders and divisions of people that can be used for many other purposes. Here we take a first pass at examining whether our measures are correlated with factors important for understanding politico-economic success. After constructing our measures, we explore how they are correlated with various standard measures of economic development, such as per capita GDP; institutional success such as freedom and corruption; and quality of life and public services such as infant mortality and education. Both measures of "artificiality" are correlated with several of these measures of political and economic development. Artificial states, as measured by these two proxies, function much less well than non-artificial states. The correlations of our measures with measures of political and economic success across countries are fairly robust to controlling for climate, colonial past and a traditional measure of ethno-linguistic fractionalization. Although our measure of the straightness or "squiggliness" of borders is more innovative and has attracted much of the discussion in preliminary presentations of this paper, it is less robust than our measure of partitioned ethnic groups; the latter results are much stronger and much more robust.

We also checked our measures' relationship to the occurrence of wars, domestic or international. A measure of political instability and violence is indeed correlated with our measure of artificial states; however we do not find evidence of correlations between the number and intensity of wars fought within a country, with our measures of artificial borders.[3] We consider our results to be a first step towards further research, which will address these questions using data on bilateral conflicts between neighbors, as well as data on civil conflict.

Because borders can be changed, as Alesina and Spolaore (1997) emphasized, citizens can rearrange the borders of artificial states. Indeed this happens, for example during the breakdown of the Soviet Union. In fact it is quite possible that as time goes by many currently straight borders will become squiggly as they are rearranged. Relatively newly independent countries have had "less time" than countries which were never colonized to carve their borders as a result of an equilibrium reflecting how different people want to organize themselves. With specific reference to Africa, Englebert et al. (2002) document several instances of border instability in Africa due to the artificial original borders. Even among never-colonized countries, tensions remain, for example with the Basque independence movement in Spain.

We are not aware of other papers that have attempted to consider formally (as opposed to narratively) the relationship of the shape of countries to economic development. However our paper is related to three strands of the literature. One strand is the recent work on the size of countries and its relationship to economic growth, as in Alesina and Spolaore (2003), Alesina, Spolaore and Wacziarg (2000), and Alcala and Ciccone (2004), among others. Second, our work builds on the literature concerning the relationship between ethno-linguistic fractionalization and economic growth, as in Easterly and Levine (1997), Alesina et al. (2003), and several others. Our paper discusses one historical phenomenon that may have led to excess ethnic fractionalization.[4] Third, the role of former colonizers has also been widely studied (see Porta et al. (1999), Acemoglu, Johnson and Robinson (2001), Glaeser, Porta, de Silanes and Shleifer (2004)) but not specifically with regard to the importance of borders. Our paper specifies a new mechanism by which colonizers affected subsequent development. In many ways we bridge these three strands because we

[3]Other authors as well have not identified a simple way of relating ethnic conflicts and civil wars, see for instance Easterly and Levine (1997) and Fearon and Laitin (2003).

[4]For a recent survey of this literature see Alesina and Ferra (2004).

focus on how colonizers have created fragmented societies by drawing artificial borders.

The paper is organized as follows. In Section 2 we provide historical examples of the artificial border-drawing. Section 3 describes our basic hypothesis, presents our measures of artificial borders, and discusses the properties of these measures. Section 4 investigates whether artificial states indeed perform less well than other states, by relating our measures of borders to various indicators of economic and political development. The last section concludes.

2.2 Examples of Problematic Borders

Examples of problematic borders abound. Millan (2003) in her analysis of the post First War meeting at Versailles describes how the redrawing of borders around the world was decided based on compromises between the winning powers, often with little regards for preserving nationalities. American President Woodrow Wilson spoke often and eloquently in favor of a nationality principle, namely that political borders had to respect ethnic boundaries and respect nationality, but that principle was often ignored, including by Woodrow Wilson himself. The book by Mc Millan clearly documents, sometimes even in hilarious ways, how borders were drawn on maps with strikes of a pencil by the leaders of England, France and the US, ignoring the leg work of their experts and without even knowing the names of the ethnicities involved. Historians agree that the Treaty of Versailles created many problematic borders that set the seeds for a very large number of future conflicts.

The past and current trouble in the Middle East at least in part originated from this kind of agreement between Western powers. Under the Sykes-Picot agreement between Britain and France during WWI, Northern Palestine would go to the French, Southern Palestine to the British, and Central Palestine including Jerusalem would be an allied Condominium shared by the two. After the war, the French agreed to give up any claims to Palestine in return for control over Syria. The British abandoned their protegee (Faisal) in Syria and offered him Iraq, cobbling together three different Ottoman provinces containing Kurds, Shiites and Sunnis. This set the stage for instability and the military coups that led to Saddam Hussein. In Lebanon, the French added Tripoli, Beirut and Sidon to the traditional Maronite area around Mount Lebanon, giving their Maronite Christian allies control to what were originally Muslim areas.

The partition of India and Pakistan is another famous example of artificial borders. The burning issue in the partition of 1947 was whether and how to award separate rights of national self-determination to Hindus and Muslims (the British ignored the national aspirations of smaller groups such as the Sikhs, which would bring its own bitter consequences). The Congress Party of Gandhi and Nehru campaigned for independence for one unitary Indian state, including Hindus, Muslims, and Sikhs from Peshawar to Dhaka. Mohammed Ali Jinnah founded the Muslim League, which called for a separate state for Muslims: Pakistan. But since Hindus and Muslims were mixed together all over the subcontinent, how could you come up with a plan to carve a Muslim nation out of India?

This intermixing was the result of a complex history that included the Muslim Mughal dynasty that the British Raj replaced. Until the last days of the Raj, there were Muslim princes ruling over majority Hindu princedoms and Hindu princes ruling over majority Muslim princedoms. The only areas with a Muslim majority were in the extreme northwest and the extreme northeast, separated by a thousand miles, and still containing large minority Sikh and Hindu communities.

In the Muslim Northwest Frontier Province (NWFP), ethnic Pathans were separated from their fellow Pathans in Afghanistan by the Durand Line, an arbitrary boundary between Afghanistan and British India laid down by a previous British bureaucrat. Peshawar, the capital of NWFP, was the traditional winter home of the Afghan kings. The Pathans preferred either an independent Pukhtoonwa uniting all Pathans or a Pathan-led Greater Afghanistan. At the time of partition, NWFP had a Congress-allied government led by a charismatic advocate of nonviolence, Khan Abdul Ghaffar Khan (the "Frontier Gandhi").

Back in British India, two other provinces of the future Pakistan were Sindh and Balochistan. Sindhi feudal landowners initially opposed the Pakistan idea and only later gave their grudging support under the naive hope that Sindh would be largely autonomous. Balochi tribesmen (also divided from ethnic compatriots by a colonial boundary with Iran) preferred an independent Balochistan, which would lead to a secessionist attempt in the 1970s, met with murderous repression by the Pakistani state. As far as Punjab and Bengal, Congress leaders would not consent to hand them over to the Muslims. This meant that the British would partition the mosaic of Hindus and Moslems in each state (and Sikhs in the Punjab, which was a Sikh state at one point). The Unionist government in Punjab prior to partition backed neither the Muslim League nor Congress.

The unhappiest heir of the partition of 1947 is Pakistan. Jinnah complained that he got a "moth-eaten" Pakistan, with missing halves of Bengal and Punjab, little of Kashmir, some frontier territory, and two disjointed areas of West and East Pakistan. As late as 1981, only 7 percent of the Pakistani population were primary speakers of the supposed national language, Urdu. So to sum up, Pakistan wound up as a collection of Balochistan, NWFP, Sindh (all of whom entertained secession at various times), East Bengal (which successfully seceded in 1971 to become Bangladesh, although only after a genocidal repression by West Pakistani troops), mohajir migrants from India (many of whom regretted the whole thing), and West Punjab (which had its own micro-secessionist movement by the Seraiki linguistic minority).[5]

Besides the examples above, artificial borders were drawn during the colonial period and few borders changed after decolonization. Africa is the region most notorious for arbitrary borders. Historian VanDerVeen (2004) points out that prior to the era of decolonization, states had to prove their control of a territory before being recognized by the international system. Virtually all new African states would have failed this test. With decolonization in Africa (and to some extent in other regions), the leading international powers changed this rule to recognize nations that existed principally on paper as the heir to a former colonial demarcation. As Van Der Veen put its, "letterbox sovereignty" was conferred upon whatever capital and whichever ruler the letters from the UN, the IMF, and the World Bank were addressed to. This left the new rulers more accountable to international organizations and leading industrial powers than to their purported citizens.[6] States consisted of little more than a few former independence agitators, the indigenous remnant of the colonial army, and a foreign aid budget. The new rulers of African states had no incentive to change a system

[5]These examples are from Easterly (2006).

[6]VanDerVeen (2004), p29

of which they were the main beneficiaries, and hence the Organization of African Unity adopted a convention in the 1960s to treat colonial boundaries as sacrosanct (only rarely violated since). We refer to Englebert et al. (2002) for many more examples of problematic borders in Africa that led to disputes, political instability and economic failures.

Latin America is a lesser known (and much earlier) example of artificial borders drawn by a colonial power, in this case Spain. The Spanish created administrative units (vice royalties, captaincies, audiencias, etc.) in the Americas that had virtually nothing to do with indigenous groups on the ground. For example, the various Mayan groups in southern Mexico, Guatemala, and what became other Central American states were split between units. The province of Upper Peru, which later became Bolivia, split the Quechuas between Bolivia and Peru, and combined the Quechuas with the Aymaras in Bolivia. When independence arrived in the early 19th century, the new states were controlled by the European elites who formed states based on these colonial demarcations. In the words of one historian, " the new 'sovereign' states were often little more than a loose collection of courts, custom houses, and military units." (Winn (1992), p. 83). Although there were some wars that altered a few borders, today's Latin American states still correspond closely to Spanish colonial divisions.

2.3 Artificial States: Hypotheses and Measures

Our main hypothesis is that artificial states perform less well than non-artificial ones. Measure of performance may include indicators of economic and political development, education, health, public goods delivery, political instability and violence. Basically our goal is to provide some statistical content to the widely-held view that countries which do not match nationalities well and are a mix of ethnic or religious group thrown together (or separated) artificially by former colonizers do not perform well.

The main difficulty is of course, to provide a measure of artificial states which is as much as possible based upon objective criteria rather than judgement calls. We will use two measures. The first measures the degree to which ethnic groups were split by borders, based upon a calculation for each pair of adjacent nations using detailed data of ethnic groups within nations from Alesina et al. (2003). The second measure is completely new, and the construction of this measure per se is, we hope, a significant contribution in itself; this is the fractal measure described below.

2.3.1 The Fractal Measure

The basic idea is to compare the borders of a country to a geometric figure. If a country looks like a perfect square with borders drawn with straight lines, the chances are these borders were drawn artificially. On the contrary, borders which are coast lines or squiggly lines (perhaps meant to capture geographic features and/or ethnicities) are less likely to be artificial. Squiggly geographic lines (such as mountains) are likely to separate ethnic groups, for reasons of patterns of communication and migration.

But how can we measure squiggliness? We first present the measure and then we discuss its properties and alternatives.

Fractal dimension is analogous to the typical concept of the dimension of an object, although, unlike the simple definition of dimension, the fractal dimension can be a fractional number. A point has a fractal dimension of zero, a straight line a fractal dimension of one, and a plane a fractal dimension of two. However, unlike the traditional definition of dimension, as a line stops being perfectly straight and begins to meanders more and more, i.e. to become more and more squiggly, the fractal dimension increases. In the limit that a curve meanders so much that it essentially fills a whole page, then the fractal dimension becomes much closer to 2 than to 1. This is because the "line" is behaving more like a "plane".

Our measure is meant to capture how close a border is to a straight line which would have a fractal dimension of 1, versus a line so squiggly that fills a plane and has a fractal dimension of 2. In practice the fractal measure of actual borders is much closer to 1 than to 2 but there is variation. Figures 2.1 and 2.2 show two countries, Sudan and France. Visually, they are quite different, as the borders of Sudan are very straight and those of France are quite squiggly. It will turn out that the fractal dimension for France is 1.0429 and that of Sudan is 1.0245, reflecting the fact that Sudan's borders are much closer to being straight lines (dimension 1.0000) than France's borders.

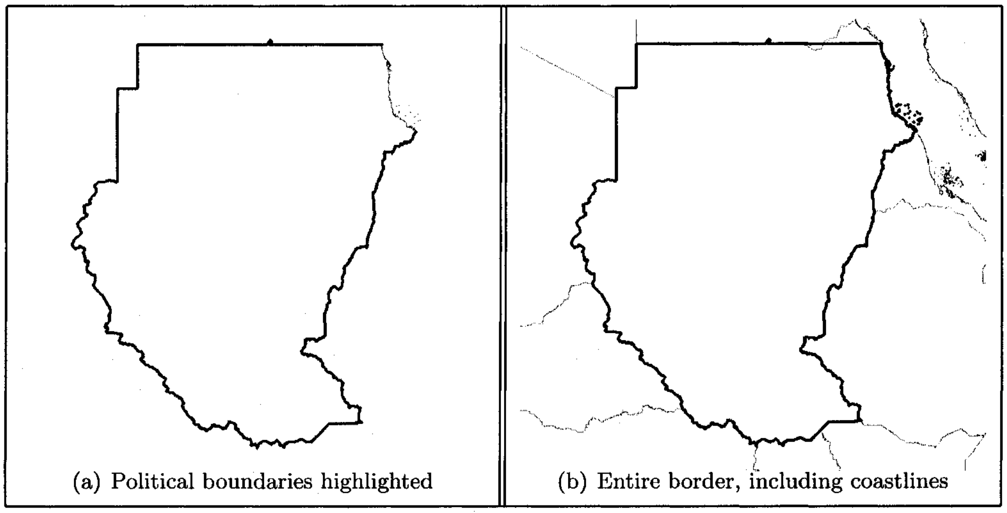


Figure 2.1: Sudan has Straight Borders and a Fractal Index of 1.0245

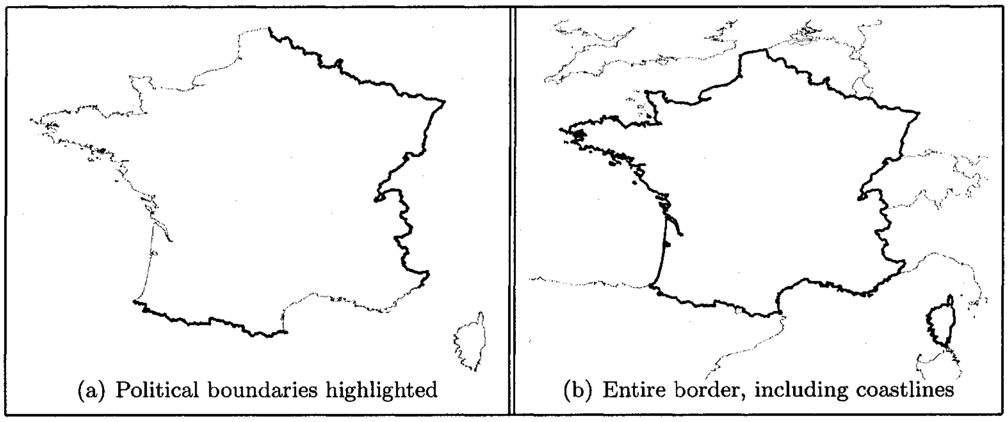


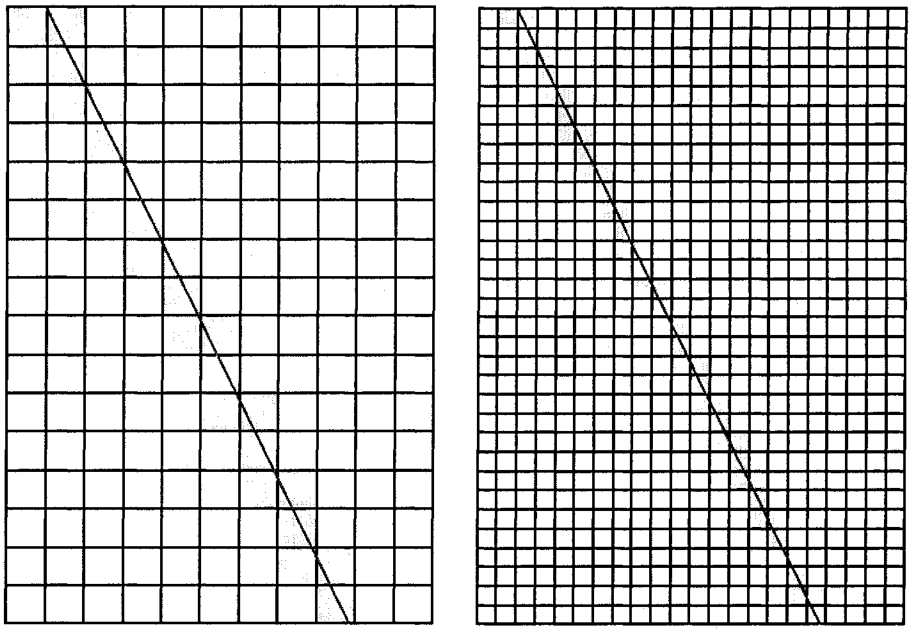
Figure 2.2: France Has Squiggly Borders and a Fractal Index of 1.0429

The fractal dimension can be calculated in several ways. We use the box-count method which is the most straightforward; (Peitgen, Jurgens and Saupe (1992), p 218-219). For this method, a grid of a certain size/scale is projected onto the border and the number of boxes which the border crosses is tallied. The scale of this grid is also recorded, as measured by the length of a side of a box in the grid. This gives a pair of numbers: box-count and box-size. The process is then repeated using grids with different box-sizes, each time recording both the box-size and the number of boxes that the border crosses. Given the pairs of data, box-size and box-count, the log-log plot of this data gives the fractal dimension as follows, where the negative of the slope (b) is the fractal dimension of the line:

$$ln(boxcount) = a - -b * ln(boxsize) \tag{2.1}$$

Some intuition for this method can be gained by considering two extreme cases, a perfectly straight line and a line so wiggly that it covers a whole page (Figures 2.3 and 2.4). Figure 2.3 shows two different grids projected onto a perfectly straight line. The length of the side of a box or the "box size" in Figure 2.3a is twice that of Figure 2.3b and we can normalize the box sizes to 2 and 1, respectively. Counting the number of squares that the line crosses in each case, we get a box count of 24 for Figure 2.3a when the box size is 2, and a box count of 48 for Figure 2.3b when the box size is 1. Thus, for the straight line, the box count doubles (or increases by a factor of 2^1) when the box size is halved (or "increases" by a factor of 2^{-1}). Plotting ln(box count) versus ln(box size) yields a downward-sloping line with a slope of –1 (Figure 2.6 and Table 2.1). Thus the fractal dimension for the straight line depicted in Figure 2.3 is determined to be 1. This makes sense because the normal notion of dimension for a perfectly straight line is one.

Next consider Figure 2.4, which shows a line so squiggly that it covers the whole page. Here the box count is 176 when the box size is 2 (Figure 2.4a) and the box count is 704 when the box size is 1 (Figure 2.4b). Thus the box count quadruples (increases by a factor of 2^2) when the box size is halved ("increases" by a factor of 2^{-1}). In this case, the plot of ln(box count) versus ln(box size) yields a downward-sloping line with a slope of 2 / –1 = –2 (Figure 2g and Table 2.1). Consequently, for this line, which is so squiggly that it fills the

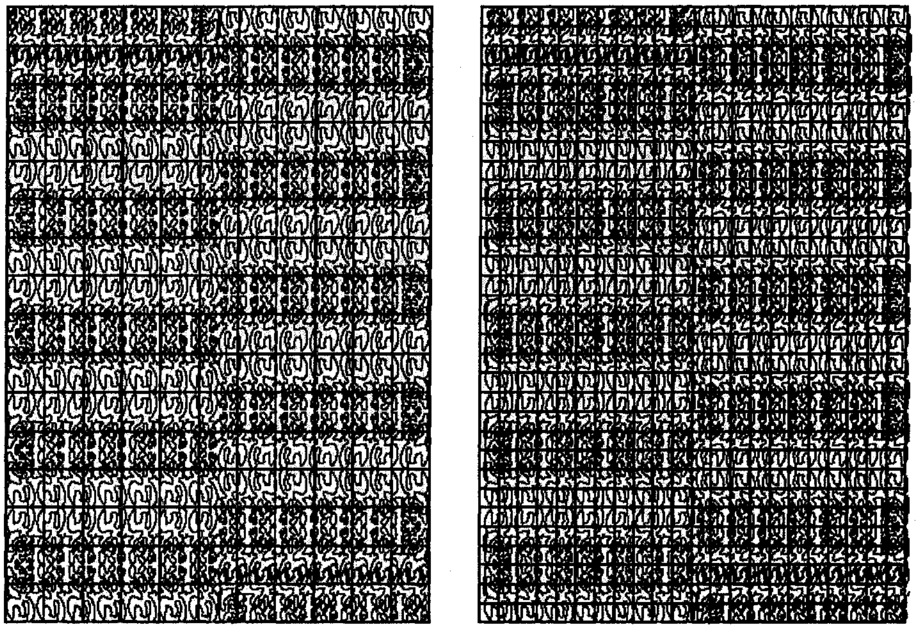


(a) Box size = 2; Box count = 24 (b) Box size = 1; Box count = 48

Figure 2.3: Box Counting Method - Straight Line (Fractal Dimension = 1)

whole page, the fractal dimension is 2. This is also in agreement with the standard notion of dimension in which a plane or a page has two dimensions.

The borders of countries will be in between these two extremes of a perfectly straight line with fractal dimension 1 and a very squiggly line which fills a whole page and has a fractal dimension of 2. Consider the somewhat less squiggly line in Figure 2.5. Here, when we calculate the fractal dimension using the box counting method, we find that the box count increases from 54 (Figure 2.5a) to 130 (Figure 2.5b) when the box size is reduced from 2 to 1, respectively. Thus the box count is more than doubling when the box size is halved. But yet the box count is not quadrupling, as was the case with the very squiggly line (Figure 2.4). We would thus expect that a plot of ln(box count) versus ln(box size) would have a slope that is steeper than -1 but not quite a steep as -2. In fact, when we do the calculation for this example, the slope is -1.267 (Figure 2.6 and Table 2.1). Based on this result, we would a sign a fractal number of 1.267 to this squiggly line. In practice the fractal dimension of most country borders is between 1.000 and 1.100. Squiggly borders have fractal dimensions closer to 1.100, while straighter borders have fractal dimensions



(a) Box size = 2; Box count = 176 (b) Box size = 1; Box count = 704

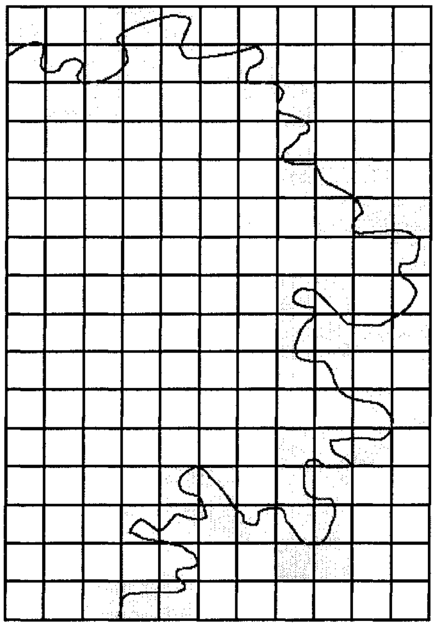
Figure 2.4: Box Counting Method - Very Squiggly Line (Fractal Dimension = 2)

closer to 1.000.

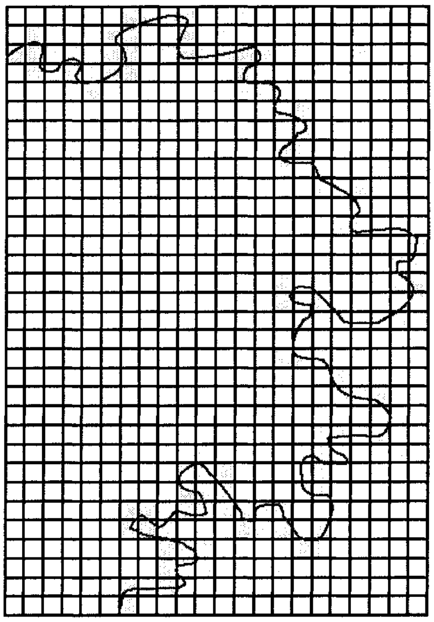
These examples use only two data points to determine the fractal dimension of a line form. In practice, when calculating the fractal dimension of country borders, we use twelve different box sizes. The smallest box size is the smallest possible, given the digital nature of our data. This smallest box size corresponds to about 0.001 of a degree latitude or longitude. In addition to this box size, which we normalize to 1, we also use grids with box sizes of 2, 3, 4, 6, 8, 16, 31, 64, 128, 256, and 512. As in the examples above, for each box size, we project a grid with that box size onto our country border. We then count the number of boxes that the border crosses, resulting in a data point of box count and box size. Using all twelve box sizes gives us twelve data points with which to regress ln(box count) on ln(box size). Recall that the general formula for the fractal dimension is given by

$$ln(boxcount) = (constantintercept) - (fractaldimension) * ln(boxsize) \tag{2.2}$$

Thus, we take the negative of the slope of the regression of ln(box count) on ln(box size)



(a) Box size = 2; Box count = 54



(b) Box size = 1; Box count = 130

Figure 2.5: Box Counting Method - Somewhat Squiggly Line (Fractal Dimension = 1.267)

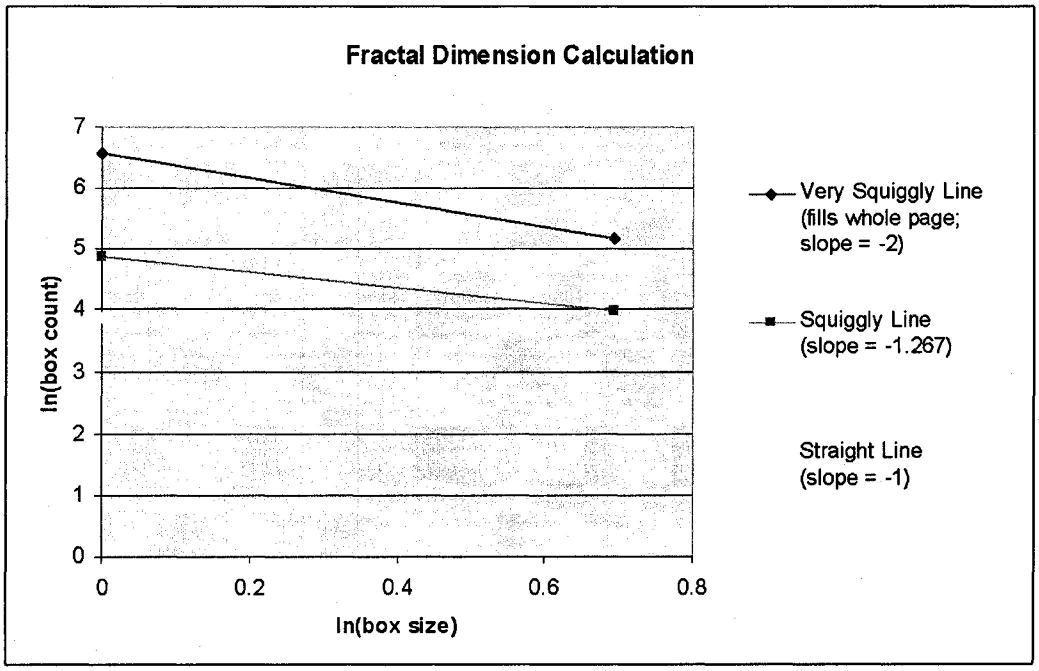


Figure 2.6: Fractal Dimension Calculation

Table 2.1: **Fractal Dimension Calculation**

ln (box count) = a – fractal dimension * ln(box size)

Straight Line (Figure 2.3)

box size	box count	ln (box size)	ln (box count)
1	48	0	3.871
2	24	0.693	3.178

Regression coeff:	Fractal Number:
-1	1

Very Squiggly Line (Figure 2.4)

box size	box count	ln (box size)	ln (box count)
1	704	0	6.557
2	176	0.693	5.17

Regression coeff:	Fractal Number:
-2	2

Squiggly Line (Figure 2.5)

box size	box count	ln (box size)	ln (box count)
1	130	0	4.868
2	54	0.693	3.989

Regression coeff:	Fractal Number:
-1.267	1.267

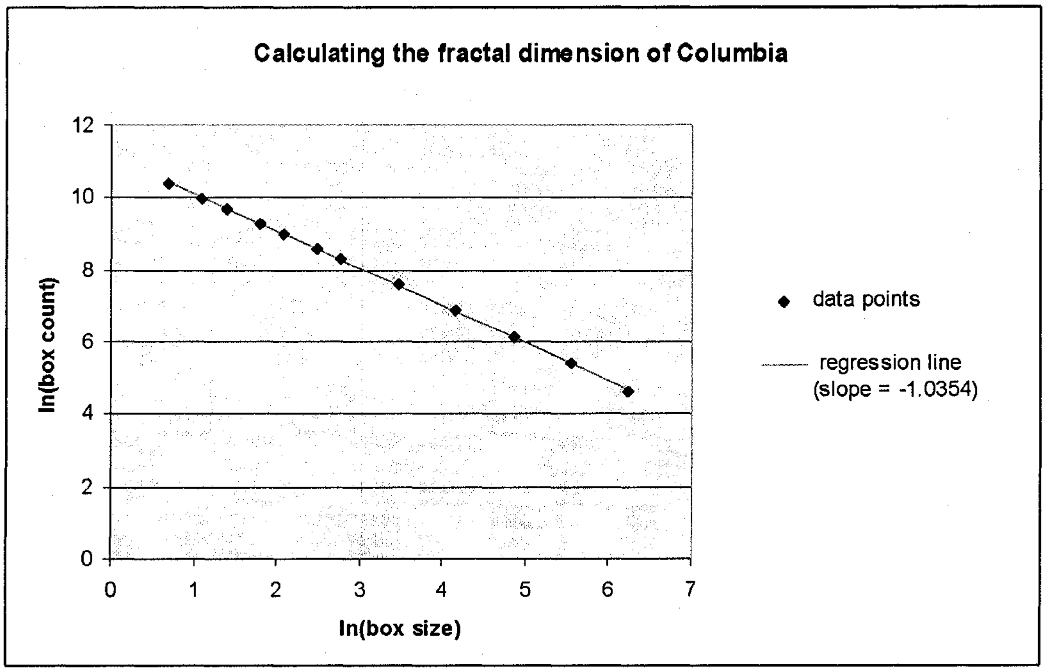


Figure 2.7: Calculating the Fractal Dimension of Columbia

as the fractal dimension for the country.

It is useful to present an example, using the case of Colombia. Figure 2.7 shows our method for determining the fractal dimension for Colombia. The graph plots ln(box count) versus ln(box size) and has twelve points, corresponding to the twelve different box sizes. For each box size, we have projected a grid of that size onto the border for Colombia and counted the number of boxes that the border crosses. Taking logs of this data, we arrive at our twelve data points, representing the pairs of data, ln(box size) and ln(box count). Regressing ln(box count) on ln(box size) using these twelve data points gives the straight line pictured on the graph. This line has a slope of -1.0354. Using the equation above, we take the negative of the slope of the regression line as the fractal dimension. Thus the fractal dimension for Colombia is 1.0354. Finally, for the purposes of our analysis, we calculate a fractal index for each country, which is the log of the fractal dimension. Returning to our example, since the fractal dimension of Colombia is 1.0354, the fractal index for Colombia is $\ln(1.0354) = 0.0348$.

2.3.2 Properties

A measure of the straightness or squiggliness of country borders ideally exhibits several properties. One desirable property is scale-invariance, meaning the ideal measure should not differ systematically for large or small countries. Scale-invariance also means we should be able to apply our measure to a particular country and get consistent results regardless of the scale of the analysis for that country. Our measure is indeed scale invariant.[7]

A second desirable property of a "squiggliness" measure is the degree to which it measures larger-scale irregularities as opposed to smaller-scale irregularities. Small-scale deviations from a smooth curve or line may well be the result of how ethnic considerations or other local politics determined whether a particular parcel of land should be on one side of a border or another. Since we are interested in comparing borders where local and ethnic considerations were taken into account, with more "artificial" borders, we prefer our measure to focus on these small-scale irregularities, rather than measuring the overall shape of a country. Unlike measures such as this circumscribed/inscribed circle ratio, the fractal measure emphasizes the small-scale variation that we are interested in measuring.

Finally, and most importantly, we would like a measure which allows us to consider only part of the border at a time. In particular, we disregard coastlines, since they are determined by nature and not by politics, and may be highly non-compact. The fractal measure can be applied to selected portions of the border, such as just the political boundaries. Most other measures of compactness must use the entire boundary, including coastlines. For instance other common compactness measures include: the ratio of the longest axis to the maximum perpendicular length; the ratio of the minimum shape diameter to the maximum diameter; various ratios among the area of the shape, the area of an inscribing circle and the area of a circumscribing circle; the moment of inertia of the shape; and the ratio of the area of

[7] To be precise our measure is not 100 percent scale invariant, but it is close to scale invariant. Analyzing a country when at differing degrees of being " zoomed in" or "zoomed out" may yield slightly different values for the fractal dimension. However, these numbers do not vary greatly for each county and the relative rankings of countries are maintained. More importantly, our measure allows us to consistently compare large and small countries. By using the same set of 12 box-sizes (as measured in degrees latitude and longitude) for each country, our analysis for each country is on the same "human" scale as for the other countries. By contrast other measures of compactness, such as the ratio of the area of a circumscribed and an inscribed circle for the country border, may differ systematically for large and small countries.

the shape to the area of a circle with the same perimeter.[8] All of these measures require a closed shape in order to be calculated.

2.3.3 Partitioned Groups and Other Measures

Our second new measure focus on the specific issue of borders cutting across an ethnic group and dividing it into two adjacent countries. This variable is defined as the percent of the population of a country that belongs to a partitioned group. In turn, a partitioned group is one that appear in two or more adjacent countries. One possible objection to this variable is mobility of people. If members of the same ethnic groups wanted to be together they could move into the same country. However mobility of people is often not free and many countries may prevent entry (or in some cases exit). We calculate the fractal variable for 144 non-island countries. Islands have no political boundaries, so they cannot have a political boundary fractal dimension. The partitioned variable is calculated for 131 countries, including 117 countries for which both indices are available.

The literature of ethno-linguistic fractionalization has normally focused on one index of fractionalization, the Herfindhal index which captures the probability that two randomly drawn individuals from the population of the country belong to different groups.[9] The original index was based on a linguistic classification of groups from a Soviet source (the Atlas Narodov Mira by Bruk and Apenchenko (1964)). It was originally used in the economic development literature by Mauro (1995) and Easterly and Levine (1997), and it is if often referred to as Elf (Ethno-linguistic fractionalization) index. Alesina et al. (2003) proposed another index that in addition to linguistic differences includes differences based on other characteristic such as skin color. They label it Fract but to avoid confusion we label is in the present paper Elfnl(See Alesina et al. (2003) for more discussion about the construction of this variable)

How do our new measures, FRACTAL and PARTITIONED, relate to each other and to the previously used index of fractionalization? Our fractal measure is meant to capture a

[8] For more on this, see Niemi, Grofman, Carlucci and Hofeller (1990) and Flaherty and Crumplin (1992).

[9] Another index frequently used is a polarization index.

much broader idea than ethnic fractionalization. However, artificial states as proxied by our measure may end up including different ethnic groups within the same political borders, and therefore there should be some correlation between the Herfindhal index of fractionalization and our fractal measure. Similar consideration apply for the portioned variable.

Table 2.2 displays the correlation coefficients between the two measures of artificial borders and the more traditional measure of ethno-linguistic fractionalization. Several comments are in order. First note how the partition variable is positively correlated with the index of ethnic fractionalization, but the correlation is in the order of 0.5 so clearly these are "different variables". Given the way the two variables are constructed it is not surprising that they are positively correlated but they indeed capture different things. Second the fractal variable is correlated with the ELF and ELF2 measures (with the appropriate negative sign, less curvy borders is associated with more fractionalization), but the correlation is not very high especially with ELF, while it is -0.22 with ELF2. Third the correlation between our partitioned variable and our fractal variable is basically zero. This was frankly a surprise to us. It suggests artificial states are not easy to summarize with one measure. (For example, the partitioned variable captures only one of the problematic features of artificial states mentioned in the introduction.) We use both measures as providing independent information on "artificiality." Finally, ELF and ELF2 are highly correlated but are not statistically identical. In summary are two new measure are different from each other and are not very highly correlated with other measures previously used in the literature of ethnic fractionalization.

2.3.4 Data and Sources

Data for determining the fractal dimension for each country's political boundary comes from the GIS (Geographic Information Systems) format data set World Vector Shoreline. This data set is the largest scale digital data set of political boundaries available today. The data is based on work done by the U.S. military in the early 1990's. The non-coastline borders for each country are isolated using ArcGIS software[10]. This data is then changed to a raster

[10] ArcGIS 9.0 Desktop software from ESRI; www.esri.com

Table 2.2: **Correlations of Various Ethnic and Artificial State Measures**
This table shows the correlations between several ethnic variables and our two measures of artificial states, Partitioned and Fractal.

	Partitioned index	Fractal index	Ethno-linguistic fractionalization (ELF) index	Alesina-Easterly fractionalization index (ELF2)
Partitioned index	1			
Fractal index	0.0554	1		
Ethno-linguistic fractionalization (ELF) index - 1960	0.5245	-0.1001	1	
Alesina-Easterly fractionalization index (ELF2)	0.5152	-0.2168	0.766	1

(digitized) format and then to a " tif" format. With a few minor modifications, the software program ImageJ[11] calculates the box-count/ box-size data for twelve different box-sizes; the smallest box-size corresponds to the smallest scale of the raster data exported from GIS (approximately 0.001 degrees latitude or longitude). A fractal dimension is calculated for each country using this data, ranging from 1.000 to 1.100. Finally, we take logs of the fractal dimension to achieve a fractal index, which ranges from 0 to 0.10.

2.4 Empirical Results

2.4.1 Which States are "Artificial"?

To illustrate which states are most artificial according to both measures, we took countries that were in the top third of PARTITIONED and in the bottom third of FRACTAL (the straightest borders). Given the weak correlation between the two measures, there were not that many countries in both – 13 to be exact. These " most artificial" states are Chad, Ecuador, Equatorial Guinea, Eritrea, Guatemala, Jordan, Mali, Morocco, Namibia, Niger, Pakistan, Sudan, and Zimbabwe. These examples accord with what we know of the historical process that led to formation of these states (some of it described above).

What about the US and Canada? Their border is a straight line most of the way, are

[11] Available online at http://rsb.info.nih.gov/ij/download.html and at http://rsb.info.nih.gov/ij/developer/index.html

they artificial states? According to our measures yes they do score relatively in terms of how artificial they are, which is certainly not consistent with a view of artificial as failed states, One may notice that this a case in which borders were drawn before many people actually moved in. In many ways the same applies to US states: in the west, their borders drawn when they were close to deserted are often straight lines. On the contrary borders of East coast states, drawn earlier with more population are less straight.[12]

2.4.2 Economic and Political Success

We now turn to verifying whether these new measures of artificial states are correlated with economic and institutional success. We consider three groups of variables as left hand side variables. (See Table 2.3) for variable definitions and sources). First, the variables that measures economic or economic policy success: (log of) per capita income in 2002; an index of economic freedom in 2005 that measures adherence to a free market economic system; and an alternative index of economic freedom averaged over 1970-2002.[13] Second, we look at poltico-institutional variables: voice and accountability (which measure checks on power), political stability and violence, government effectiveness, regulatory quality, rule of law, and corruption. Third, we use quality of life and public goods delivery-related measures: infant mortality in 2001, literacy rate averaged over the period 1995-2002; measles immunization rate in 2002; immunization rate against DPT in 2002, percent of population with access to clean water, in 2000.[14] We choose these variables as representative of state performance in the core public goods areas of health, education, and infrastructure, selecting particular measures based on which ones have data available for a large sample of countries. All of these variable are clearly correlated with each other. Obviously rich country have lower

[12]Needless to say US and Canada are included in our regressions below.

[13]We use the second measure as a robustness check on the first measure of economic freedom, since each is based on a complicated mix of indicators and may reflect some subjectivity. Given the uncertainty surrounding this measure, we also check robustness with respect to using a long period average of the second measure rather than just a single year, which may average out data errors and noise (while sacrificing our preferred approach of using the most recent datapoint available).

[14]Data on literacy is spotty, with different countries reporting different years over 1995-2002, so we average all available data over this period. Otherwise, the year given is the most recent for which data are widely available.

infant mortality, more clean water etc. Table 2.4 reports a correlation chart between all of these variables: the correlations are not all very close to 1 (or -1 depending on the variable definition). That is, this set of variables do capture different aspects of political and economic development that are different from each other, so there is information provided by considering all of them.

Table 2.5 presents the basic univariate regressions of our measures of artificial states. Consider line one: the left hand side variable is the log of per capita GDP in 2002, and we report only the coefficient and the p value of the single right hand side variable. (Obviously we include also a constant in the regression). Each line represents the same regressions with a different left hand side variable which is listed in the first column. We use all the observations available, and their number varies (from 84 to 144) in different regressions because of data availability on the left hand side variable. The dependent variables are divided in three blocs: economic variables, institutional variables and quality of life/public goods variables. Notice that because of how the right hand side variables are constructed, we expect the opposite sign in the first and second column. So for instance in the first line we expect a negative correlation of economic success measured as income per capita in countries where the partition variable assumes a lower value, and in countries where the measure of how straight borders are assumes a higher value. The coefficient in bold represents all the cases in which statistical significance (with the expected sign of course) is 5 per cent or better; marginally significant coefficient at the 10 per cent level or better are indicated with a "+" sign . Of the 28 coefficients in the first two columns, 20 are statistically significant (5 per cent or better) and there are borderline (p value 0.10 or better). Our two measures are not highly correlated with each other and in fact as discussed above, they capture different aspects of the nature of borders. For this reason there is no reason why they could not be used in the same regressions. In the third column, we use them both. In all regressions at least one is significant at the 5 per cent level or better and in almost all regressions they are either both statistically significant at the five per cent level or one is and the other is borderline.

Table 2.3: **Data Sources**

VARIABLE	DATA SOURCE
Code	3-letter World Bank country code
ETHNIC VARIABLES	
Partitioned	Percent of population belonging to groups partitioned by a border
Fractal	Log of basic fractal index (latest revision as of September 2005) based on the World Vector Shoreline Dataset (GIS format)
SmallFractal	Log of small country fractal index (used only in robustness checks)
ELF	Ethno-linguistic fractionalization, 1960 (as used in Easterly and Levine 1997)
ELF2	Ethnic fractionalization (from Alesina et al. 2003)
POLITICAL VARIABLES	
Noncolonial	dummy =1 if never colonized by European power
Kaufmann-Kray indices of institutions for 2004 (increase means better institutions):	
Voice/democracy	Checks on power, accountability to population
Political stability	Political stability and violence
Govt. Effectiveness	Government effectiveness
Regulatory Quality	Regulatory quality
Rule of Law	Rule of law
Corruption	Corruption
ECONOMIC VARIABLES	
Log GDP per capita	Log per capita income in 2002 (Summers-Heston updated with World Bank per capita growth rates)
Index Econ Freedom	Index of Economic Freedom, 2005 (increase means less freedom) from the Heritage Foundation
EFW index	Economic Freedom in the World, average 1970-2002 from the Fraser Institute
QUALITY OF LIFE AND PUBLIC GOODS VARIABLES	
Infant Mortality	Infant mortality rate in 2001 (WDI)
Literacy	Literacy rate averaged over available data 1995-2002 (EDI)
Measles immun.	Measles immunization rate, 2002 (WDI)
DPT immun.	Immunization rate against DPT, 2002 (WDI)
Access to Water	Percent of population with access to clean water, 2000 (WDI)
GEOGRAPHY VARIABLES Source: Sachs (199X), Center for International Development, Harvard	
Climate	Percent of cultivated land in Koppen-Geiger climate zone A (humid climate with no winter)
Climate (cultcb)	Percent of cultivated land in Koppen-Geiger climate zone B (dry climate with no winter) Note: cultca and cultcb included separately as controls
Desert	Percent of total land in Koppen-Geiger climate zone BW (desert)
Land area	Total land area in kilometers squared
Population density	Population density experienced by the typical citizen (population density of many small regions is averaged, using the population of each region as a weight)

Table 2.4: **Correlations Among the Principle Dependent Variables**

Economic Variables	Log GDP per capita	Index of Econ Freedom	Economic Freedom in the World
Log GDP per capita, 2002	1		
Index of Econ Freedom, 2005 (higher = less free)	-0.7078	1	
Economic Freedom in the World, avg 1970-2002	0.7431	-0.7494	1

Quality of Life Variables	Literacy rate	Percent access to clean water	Infant mortality	Measles immuniz. rate	DPT immuniz. rate
Literacy rate, avg of available data 1995-2002	1				
Percent pop with access to clean water, 2000	0.5105	1			
Infant mortality, 2001	-0.7074	-0.6835	1		
Measles immunization rate, 2002	0.6743	0.5771	-0.6975	1	
DPT immunization rate, 2002	0.6666	0.6079	-0.7461	0.8956	1

Political Variables	Voice-checks on and power	Political stability violence	Govt effectiveness	Regulatory quality	Rule of law	Corruption
Voice - checks on power	1					
Political stability and violence	0.7306	1				
Government effectiveness	0.7197	0.7858	1			
Regulatory quality	0.8147	0.8034	0.9092	1		
Rule of law	0.8035	0.8791	0.9315	0.9244	1	
Corruption	0.7412	0.7993	0.9564	0.8898	0.9496	1

Table 2.5: **OLS Regressions with No Controls**

OLS regressions using our two measures of artificial states. No controls are included in these basic regressions. Each column and each band (eg. for Log GDP per capita) represents a different regression.

Dependent variables:		Coefficient on:	1	2	3
Economic variables	Log GDP per capita, 2002	Partitioned	**-0.021****		**-0.019****
			0		**0**
		Fractal		**11.49***	10.23+
				-0.041	-0.083
	Index of Econ Freedom, 2005 (higher = less free)	Partitioned	**0.006***		**0.005***
			-0.013		**-0.028**
		Fractal		-6.12+	**-7.54***
				-0.08	**-0.031**
	Economic Freedom in the World, avg 1970-2002	Partitioned	**-0.009***		**-0.008***
			-0.029		**-0.037**
		Fractal		5.7	9.8
				-0.369	-0.142
Political and governance variables	Voice - checks on power	Partitioned	**-0.01****		**-0.01****
			-0.003		**-0.002**
		Fractal		**13.16****	**14.66****
				-0.002	**-0.002**
	Political stability and violence	Partitioned	**-0.009****		**-0.01****
			-0.008		**-0.001**
		Fractal		5.79	7.26
				-0.199	-0.151
	Government effectiveness	Partitioned	**-0.01****		**-0.011****
			-0.002		**0**
		Fractal		9.57+	**11.46***
				-0.062	**-0.042**
	Regulatory quality	Partitioned	**-0.01****		**-0.011****
			-0.003		**0**
		Fractal		**11.23***	**12.93***
				-0.016	**-0.011**
	Rule of law	Partitioned	**-0.011****		**-0.012****
			-0.001		**0**
		Fractal		8.33+	9.43+
				-0.099	-0.092
	Corruption	Partitioned	**-0.011****		**-0.011****
			-0.001		**-0.001**
		Fractal		8.53	10.22+
				-0.106	-0.079
Quality of life variables	Literacy rate, avg of available data 1995-2002	Partitioned	**-0.442****		**-0.441****
			0		**0**
		Fractal		**290.6***	**393.5****
				-0.029	**0**
	Percent pop with access to clean water, 2000	Partitioned	**-0.261****		**-0.267****
			0		**0**
		Fractal		**238.0***	168.3+
				-0.021	-0.1

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.5: continued

Dependent variables:	Coefficient on:	1	2	3
Infant mortality, 2001	Partitioned	**0.702****		**0.774****
		0		**0**
	Fractal		**-548.0****	**-556.5****
			-0.001	**-0.002**
Measles immunization rate, 2002	Partitioned	**-0.317****		**-0.379****
		0		**0**
	Fractal		94.8	110.2+
			-0.13	-0.061
DPT immunization rate, 2002	Partitioned	**-0.323****		**-0.375****
		0		**0**
	Fractal		**190.5****	**214.5****
			-0.009	**-0.006**

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.6 displays information on the size of the impact of these measures of artificial states, which is considerable. For the partitioned variable, going from the 75th most partitioned country to the 25th most partitioned country is associated with an increase of 83% in GDP per capita (0.832 log-points; Table 2.6, Column 2). Many of the other variables are also strongly affected, by around half of a standard deviation (Column 3). The impact of the fractal variable is smaller but still significant in size. Moving from the 75th most squiggly border to the 25th most squiggly border is associated with a 37% increase in GDP per capita. The other dependent variables are also affected by about a third of a standard deviation.

We now check whether these strong univariate correlations survive adding other exogenous variables to the right hand side. We begin with ethnic fractionalization to see whether our new measure add anything to traditional and already used measures of ethnic fractionalization. In Table 2.7 we add as a control in the right hand side the variable ELF, the "traditional" ethno-linguistic fractionalization variable used by Easterly and Levine (1997) and by many after them. In the case of our FRACTAL measure, the result suggests that in about half the regressions (6 out of 14) both variables are statistically significant, in another one FRACTAL is marginal at the 10 per cent level. In particular, for the institutional regressions, FRACTAL remains significant when controlling for ELF. For the other regressions, ELF is significant but FRACTAL is not. Consider now column 1. Here the variable PARTITIONED remains significant in 7 out of 14 regressions. For GDP per capita, PARTITIONED remains significant when controlling for ELF. Column 3 shows our results when we include both variables and control for ELF. Of the 28 coefficients on our artificial states variables (from the 14 regressions), 16 are significant at a the 5 per cent level or greater and 9 are

Table 2.6: **Impact of Partitioned and Fractal Variables**

PARTITIONED index (high value = artificial state):
25th %tile rank = Vietnam: 2.8
75th %tile rank = Latvia: 42.4

FRACTAL index (low value = artificial state):
25th %tile rank = Israel (incl WB border): 0.0498
75th %tile rank = DR of Congo (Zaire): 0.0241

Dependent variable	Standard deviation of dependent variable	Impact of going from 25th to 75th percentile in the PARTITIONED Index (coeff * 39.6)	**Impact of PARTITIONED/ std dev of dep variable**	Impact of going from 25th to 75th percentile in the FRACTAL Index (coeff * 0.0257)	**Impact of FRACTAL/ std dev of dep. variable**
Log GDP per capita, 2002	1.141	0.832	**0.73**	0.374	**0.33**
Index of Econ Freedom, 2005 (higher = less free)	0.72	-0.238	**0.33**	-0.222	**0.31**
Economic Freedom in the World, avg 1970-2002	0.998	0.356	**0.36**	0.272	**0.27**
Voice - checks on power	1	0.396	**0.4**	0.393	**0.39**
Political stability and violence	1	0.356	**0.36**	0.181	**0.18**
Government effectiveness	1	0.396	**0.4**	0.324	**0.32**
Regulatory quality	1	0.396	**0.4**	0.355	**0.36**
Rule of law	1	0.436	**0.44**	0.282	**0.28**
Corruption	1	0.436	**0.44**	0.297	**0.3**
Literacy rate, avg of available data 1995-2002	21.18	17.503	**0.83**	7.132	**0.34**
Percent pop with access to clean water, 2000	20.601	10.336	**0.5**	7.012	**0.34**
Infant mortality, 2001	41.825	-27.799	**0.66**	-16.023	**0.38**
Measles immunization rate, 2002	17.049	12.553	**0.74**	2.851	**0.17**
DPT immunization rate, 2002	18.445	12.791	**0.69**	5.654	**0.31**

borderline at the 10 per cent level.

The next experiment is about former colonial status. As we discussed in section 2 above, much of the problem of artificial states has to do with colonizers drawing borders which did not respect indigenous divisions. In fact, the FRACTAL index for former colonies is lower than for non-former colonies, with the index averages equal to 0.0335 and 0.0435 for these two groups respectively. This difference is significant at the 1% level. The overall standard deviation for the fractal index is about 0.02, so this is an important difference of about half a standard deviation between former colonies and non-colonies. Likewise, for the PARTITION variable, former colonies and non-colonies differ by 13.6 out of the 100 point scale; former colonies have higher proportions citizens from "partitioned" ethnic groups. This difference is also significant at the 1% level. But having been a colony or not may influence political and economic outcomes in many different ways, so it is important to check that controlling for colonial status does not change all the significance of our variables of interest. We do that in Table 2.8 where we add a dummy variable that assumes the value of 1 if the country has never been a colony. In column 1 note how 11 out of the 14 coefficients on the partition variable are now significant at the 5 per cent level and all the others except one are borderline. For the fractal measure, however, only 1 out of 14 is and one is borderline. This show that it is difficult to identify separately the effect of colonial status and straight-line borders, since one led to the other. For the regressions with both variables, about half of the 28 coefficients are significant.

Another important exogenous factor that can explain economic and political success is geography and climate. Many geographic variables have been suggested in the literature. One of the most precise in capturing weather pattern is the variable climate defined as the percentage of a country's cultivatable land that is in the Koppen-Geiger Climate Zone A, which is a humid climate with no winter. This is a classical definition of what constitutes a tropical area. In Table 2.9 we add this variable to our regression. Our variables are generally quite robust much more so than the ELF variable. Perhaps most important is that both of our measures of artificial states are significant in the single most central and most complete regression, that for per capita income controlling for ethnic fractionalization, colonial status, and climate all at the same time.

Table 2.7: **Controlling for Ethno-Linguistic Fractionalization (ELF60)**

OLS regressions using our two measures of artificial states, and a control for ethno-linguistic fractionalization (ELF60). Each column and each band (eg. for Log GDP per capita) represents a different regression.

Dependent variables:		Coefficient on:	1	2	3
Economic variables	Log GDP per capita, 2002	Partitioned	**-0.016****		**-0.018****
			-0.003		**0**
		Fractal		10.555	**16.284***
				-0.112	**-0.01**
		ELF60	**-0.013****	**-0.021****	**-0.01***
			-0.01	**0**	**-0.046**
	Index of Econ Freedom, 2005 (higher = less free)	Partitioned	0.003		0.003
			-0.358		-0.365
		Fractal		-6.47	-7.927+
				-0.121	-0.056
		ELF60	**0.007***	**0.008****	**0.007***
			-0.018	**0**	**-0.048**
	Economic Freedom in the World, avg. 1970-2002	Partitioned	-0.007		-0.008
			-0.196		-0.106
		Fractal		8.916	**15.517***
				-0.176	**-0.018**
		ELF60	-0.006	**-0.009****	-0.004
			-0.19	**-0.002**	-0.377
Political and governance variables	Voice - checks on power	Partitioned	-0.008+		**-0.01***
			-0.095		**-0.029**
		Fractal		**14.21****	**17.728****
				-0.008	**-0.003**
		ELF60	**-0.008***	**-0.01****	-0.006
			-0.046	**-0.002**	-0.168
	Political stability and violence	Partitioned	-0.007		-0.009+
			-0.146		-0.06
		Fractal		**11.88***	**16.01****
				-0.032	**-0.009**
		ELF60	**-0.009***	**-0.012****	-0.007
			-0.047	**-0.001**	-0.128
	Government effectiveness	Partitioned	-0.008		-0.009+
			-0.121		-0.051
		Fractal		**13.52***	**17.59***
				-0.04	**-0.012**
		ELF60	**-0.011***	**-0.013****	-0.009+
			-0.02	**-0.001**	-0.066
	Regulatory quality	Partitioned	-0.008		**-0.01***
			-0.101		**-0.043**
		Fractal		**14.55****	**17.94****
				-0.009	**-0.005**
		ELF60	**-0.009***	**-0.012****	-0.007
			-0.035	**0**	-0.117

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.7: continued

Dependent variables:		Coefficient on:	1	2	3
	Rule of law	Partitioned	-0.008		-0.009+
			-0.143		-0.074
		Fractal		**13.40***	**16.74***
				-0.035	**-0.013**
		ELF60	**-0.012****	**-0.014****	**-0.011***
			-0.008	**0**	**-0.034**
	Corruption	Partitioned	-0.008		-0.008+
			-0.125		-0.062
		Fractal		**13.81***	**17.09****
				-0.026	**-0.007**
		ELF60	**-0.013****	**-0.014****	**-0.011***
			-0.004	**0**	**-0.015**
Quality of life variables and public goods delivery	Literacy rate, avg of available data 1995-2002	Partitioned	**-0.38****		**-0.396****
			0		**-0.001**
		Fractal		204.7	325.6+
				-0.254	-0.051
		ELF60	**-0.154***	**-0.271****	-0.128
			-0.032	**-0.001**	-0.122
	Percent pop with access to clean water, 2000	Partitioned	-0.173+		-0.152+
			-0.052		-0.067
		Fractal		88.79	8.58
				-0.538	-0.948
		ELF60	**-0.226****	**-0.27****	**-0.224****
			-0.006	**0**	**-0.005**
	Infant mortality, 2001	Partitioned	**0.426***		**0.452***
			-0.032		**-0.024**
		Fractal		-380.3+	-497.8+
				-0.068	-0.06
		ELF60	**0.752****	**0.861****	**0.687****
			0	**0**	**-0.002**
	Measles immunization rate, 2002	Partitioned	**-0.27****		**-0.288****
			0		**0**
		Fractal		-41.841	42.992
				-0.602	-0.634
		ELF60	**-0.157****	**-0.297****	**-0.166***
			-0.009	**0**	**-0.012**
	DPT immunization rate, 2002	Partitioned	**-0.206****		**-0.242****
			-0.003		**-0.001**
		Fractal		99.039	188.871+
				-0.278	-0.084
		ELF60	**-0.276****	**-0.332****	**-0.255****
			0	**0**	**-0.002**

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.8: **Controlling for ELF60 and Former Colonial Status**

OLS regressions using our two measures of artificial states, and controls for ethno-linguistic fractionalization (ELF60) and former colonial status. Each column and each band (eg. for Log GDP per capita) represents a different regression.

Dependent variables:		Coefficient on:	1	2	3
Economic variables	Log GDP per capita, 2002	Partitioned	-0.016**		-0.016**
			0		0
		Fractal		5.918	7.406+
				-0.324	-0.095
		ELF60	-0.003	-0.014**	-0.002
			-0.481	-0.001	-0.654
		NON-COLONIAL	1.426**	1.099**	1.474**
			0	-0.001	0
	Index of Econ Freedom, 2005 (higher = less free)	Partitioned	0.004		0.003
			-0.165		-0.367
		Fractal		-3.965	-3.546
				-0.269	-0.305
		ELF60	0.002	0.004+	0.002
			-0.57	-0.085	-0.587
		NON-COLONIAL	-0.739**	-0.607**	-0.785**
			0	-0.001	0
	Economic Freedom in the World, avg. 1970-2002	Partitioned	-0.007+		-0.006
			-0.086		-0.15
		Fractal		2.042	6.819
				-0.678	-0.162
		ELF60	0.002	-0.003	0.003
			-0.587	-0.276	-0.53
		NON-COLONIAL	1.133**	1.145**	1.228**
			0	0	0
Political and governance variables	Voice - checks on power	Partitioned	-0.007*		-0.008*
			-0.032		(0.0360
		Fractal		10.40+	10.67+
				-0.053	(0.0930
		ELF60	0	-0.005	0
			-0.988	-0.124	-0.952
		NON-COLONIAL	1.255**	0.847**	1.149**
			0	-0.003	0
	Political stability and violence	Partitioned	-0.007+		-0.007+
			-0.065		-0.081
		Fractal		7.572	8.161
				-0.121	-0.121
		ELF60	-0.001	-0.006	0
			-0.849	-0.115	-0.923
		NON-COLONIAL	1.272**	0.959**	1.277**
			0	-0.001	0

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.8: continued

Dependent variables:		Coefficient on:	1	2	3
	Government effectiveness	Partitioned	**-0.008***		-0.007+
			-0.043		-0.073
		Fractal		7.68	7.796
				-0.153	-0.161
		ELF60	-0.001	-0.005	-0.001
			-0.816	-0.152	-0.891
		NON-COLONIAL	**1.526****	**1.30****	**1.594****
			0	**0**	**0**
	Regulatory quality	Partitioned	**-0.008***		-0.007+
			-0.044		-0.067
		Fractal		**9.977***	9.814+
				-0.04	-0.077
		ELF60	0	-0.006+	0
			-0.912	-0.075	-0.954
		NON-COLONIAL	**1.343****	**1.018****	**1.322****
			0	**0**	**0**
	Rule of law	Partitioned	**-0.007***		-0.006+
			-0.048		-0.099
		Fractal		7.295	6.553
				-0.148	-0.192
		ELF60	-0.002	-0.006+	-0.002
			-0.611	-0.098	-0.706
		NON-COLONIAL	**1.576****	**1.358****	**1.658****
			0	**0**	**0**
	Corruption	Partitioned	**-0.007***		-0.006+
			-0.045		-0.089
		Fractal		7.76	7.214
				-0.12	-0.148
		ELF60	-0.003	**-0.006***	-0.002
			-0.437	**-0.045**	-0.504
		NON-COLONIAL	**1.533****	**1.346****	**1.608****
			0	**0**	**0**
Quality of life variables and public goods delivery	Literacy rate, avg of available data 1995-2002	Partitioned	**-0.373****		**-0.388****
			0		**-0.001**
		Fractal		202.8	318.8+
				-0.259	-0.06
		ELF60	-0.135+	**-0.273****	-0.112
			-0.08	**-0.002**	-0.2
		NON-COLONIAL	5.712	-0.6	5.518
			-0.261	-0.933	-0.269
	Percent pop with access to clean water, 2000	Partitioned	**-0.171***		-0.144+
			-0.039		-0.069
		Fractal		71.76	-12.69
				-0.608	-0.911
		ELF60	**-0.173***	**-0.234****	**-0.171***
			-0.044	**-0.003**	**-0.043**
		NON-COLONIAL	**10.19***	7.853	**11.105***
			-0.018	-0.251	**-0.013**

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.8: continued

Dependent variables:	Coefficient on:	1	2	3
Infant mortality, 2001	Partitioned	**0.421***		**0.404***
		-0.022		**-0.035**
	Fractal		-270.4	-325.6
			-0.185	-0.201
	ELF60	**0.566****	**0.719****	**0.536***
		-0.009	**0**	**-0.015**
	NON-COLONIAL	**-28.477****	**-24.45****	**-28.02****
		-0.002	**-0.007**	**-0.001**
Measles immunization rate, 2002	Partitioned	**-0.27****		**-0.287****
		0		**0**
	Fractal		-50	38.08
			-0.549	-0.66
	ELF60	**-0.159***	**-0.286****	**-0.162***
		-0.018	**0**	**-0.022**
	NON-COLONIAL	-0.346	1.846	0.799
		-0.924	-0.611	-0.802
DPT immunization rate, 2002	Partitioned	**-0.205****		**-0.23****
		-0.002		**-0.002**
	Fractal		62.04	146.4
			-0.489	-0.157
	ELF60	**-0.225****	**-0.284****	**-0.217***
		-0.008	**0**	**-0.012**
	NON-COLONIAL	**7.713***	**8.368***	**6.915***
		-0.046	**-0.03**	**-0.049**

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.9: **Controlling for ELF60, Colonial Status and Climate**

OLS regressions using our two measures of artificial states, and controls for ethno-linguistic fractionalization (ELF60), former colonial status and climate, as measured by the percentage of the country's land area in Koppen-Geiger climate zone A (rainy and hot). Each column and each band (eg. for Log GDP per capita) represents a different regression.

Dependent variables:		Coefficient on:	1	2	3
Economic variables	Log GDP per capita, 2002	Partitioned	**-0.015****		**-0.016****
			0		**0**
		Fractal		10.53+	**10.76***
				-0.074	**-0.013**
		ELF60	-0.002	**-0.013****	-0.001
			-0.562	**-0.001**	-0.835
		NON-COLONIAL	**1.323****	0.694+	**1.226****
			0	-0.053	**0**
		CLIMATE	-0.456+	**-0.813***	-0.531+
			-0.064	**-0.016**	-0.062
	Index of Econ Freedom, 2005 (higher = less free)	Partitioned	0.004		0.004
			-0.193		-0.282
		Fractal		-6.333+	-5.122
				-0.069	-0.157
		ELF60	0.001	0.003	0.001
			-0.722	-0.149	-0.799
		NON-COLONIAL	**-0.709****	**-0.411***	**-0.68****
			-0.001	**-0.049**	**-0.001**
		CLIMATE	0.196	**0.463***	0.239
			-0.318	**-0.029**	-0.28
	Economic Freedom in the World, avg. 1970-2002	Partitioned	-0.006		-0.006
			-0.103		-0.115
		Fractal		3.982	8.549+
				-0.416	-0.069
		ELF60	0.002	-0.003	0.003
			-0.574	-0.317	-0.453
		NON-COLONIAL	**1.195****	**0.98****	**1.103****
			0	**0**	**0**
		CLIMATE	-0.116	-0.273	-0.24
			-0.669	-0.3	-0.377
Political and governance variables	Voice - checks on power	Partitioned	**-0.007***		**-0.009***
			-0.032		**-0.018**
		Fractal		**12.59***	**13.83***
				-0.023	**-0.038**
		ELF60	0.001	-0.004	0.001
			-0.871	-0.216	-0.681
		NON-COLONIAL	**1.184****	**0.642***	**0.942****
			0	**-0.046**	**-0.006**
		CLIMATE	-0.193	**-0.56***	-0.448
			-0.464	**-0.048**	-0.119

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.9: continued

Dependent variables:	Coefficient on:	1	2	3
Political stability and violence	Partitioned	-0.006+		**-0.008***
		-0.078		**-0.028**
	Fractal		**11.51***	**12.24***
			-0.012	**-0.015**
	ELF60	0	-0.005	0.001
		-0.988	-0.203	-0.766
	NON-COLONIAL	**1.17****	**0.637***	**1.009****
		0	**-0.039**	**-0.001**
	CLIMATE	-0.326	**-0.707***	-0.579+
		-0.274	**-0.016**	-0.06
Government effectiveness	Partitioned	**-0.007***		**-0.008***
		-0.05		**-0.026**
	Fractal		**12.49***	**12.45***
			-0.011	**-0.019**
	ELF60	0	-0.004	0.001
		-0.933	-0.299	-0.759
	NON-COLONIAL	**1.42****	**0.898****	**1.289****
		0	**-0.005**	**0**
	CLIMATE	**-0.515***	**-0.916****	**-0.66***
		-0.048	**-0.001**	**-0.014**
Regulatory quality	Partitioned	-0.007+		**-0.008***
		-0.061		**-0.042**
	Fractal		**12.95****	**12.53***
			-0.006	**-0.021**
	ELF60	0	-0.005	0.001
		-0.99	-0.134	-0.844
	NON-COLONIAL	**1.314****	**0.762***	**1.144****
		0	**-0.012**	**0**
	CLIMATE	-0.197	**-0.618***	-0.385
		-0.469	**-0.028**	-0.171
Rule of law	Partitioned	**-0.007***		**-0.007***
		-0.048		**-0.031**
	Fractal		**12.22****	**11.65***
			-0.009	**-0.014**
	ELF60	-0.001	-0.004	0
		-0.879	-0.201	-0.913
	NON-COLONIAL	**1.428****	**0.951****	**1.324****
		0	**-0.002**	**0**
	CLIMATE	**-0.581***	**-0.911****	**-0.722****
		-0.022	**0**	**-0.005**

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.9: continued

Dependent variables:		Coefficient on:	1	2	3
	Corruption	Partitioned	-0.007+		**-0.007***
			-0.053		**-0.034**
		Fractal		**12.28****	**11.65***
				-0.009	**-0.019**
		ELF60	-0.001	-0.005	-0.001
			-0.678	-0.108	-0.829
		NON-COLONIAL	**1.418****	**0.97****	**1.317****
			0	**-0.002**	**0**
		CLIMATE	**-0.531***	**-0.853****	**-0.628****
			-0.026	**-0.001**	**-0.009**
Quality of life variables and public goods delivery	Literacy rate, avg of available data 1995-2002	Partitioned	**-0.364****		**-0.39****
			-0.001		**-0.001**
		Fractal		212.6	**333.1***
				-0.268	**-0.037**
		ELF60	-0.144+	**-0.272****	-0.109
			-0.068	**-0.003**	-0.197
		NON-COLONIAL	8.2	-1.326	5.061
			-0.197	-0.873	-0.371
		CLIMATE	4.773	0.359	-1.119
			-0.41	-0.962	-0.837
	Percent pop with access to clean water, 2000	Partitioned	**-0.163***		**-0.160***
			-0.039		**-0.037**
		Fractal		170.6	59.51
				-0.196	-0.594
		ELF60	**-0.163***	**-0.217****	-0.156+
			-0.05	**-0.003**	-0.057
		NON-COLONIAL	7.994+	1.206	7.727+
			-0.086	-0.869	-0.079
		CLIMATE	-6.676	**-14.72***	-7.662
			-0.124	**-0.014**	-0.124
	Infant mortality, 2001	Partitioned	**0.409***		**0.419***
			-0.026		**-0.026**
		Fractal		-368.9+	-385.2
				-0.051	-0.1
		ELF60	**0.555****	**0.69****	**0.512***
			-0.009	**0**	**-0.016**
		NON-COLONIAL	**-28.23****	-16.84	**-24.11****
			-0.008	-0.109	**-0.009**
		CLIMATE	4.644	14.81	8.447
			-0.7	-0.243	-0.501

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

Table 2.9: continued

Dependent variables:		Coefficient on:	1	2	3
	Measles immunization rate, 2002	Partitioned	**-0.272****		**-0.295****
			0		**0**
		Fractal		0.707	72.339
				-0.993	-0.384
		ELF60	**-0.153***	**-0.272****	**-0.149***
			-0.017	**0**	**-0.031**
		NON-COLONIAL	0.142	-2.092	-1.449
			-0.971	-0.63	-0.7
		CLIMATE	-3.197	-7.155	-4.856
			-0.465	-0.193	-0.283
	DPT immunization rate, 2002	Partitioned	**-0.202****		**-0.249****
			-0.004		**-0.001**
		Fractal		149.8+	**224.2***
				-0.065	**-0.03**
		ELF60	**-0.21***	**-0.256****	**-0.187***
			-0.012	**0**	**-0.025**
		NON-COLONIAL	6.08	1.18	1.807
			-0.169	-0.78	-0.638
		CLIMATE	-5.949	**-14.54***	**-11.04***
			-0.274	**-0.018**	**-0.047**

P values in parenthesis.
**, * and + refer to significant results at the 1%, 5% and 10% levels, respectively.

2.4.3 Other Robustness Checks

We consider a number of other possible explanations for our results, adding further controls that might otherwise have introduced a spurious correlation with our measures of artificiality of states. In order to keep the length of this paper manageable, we simply summarize the results here in the text. A separate appendix with the full results will be available on our web sites.

First, we include the index of ethnic fractionalization ELF1, from Alesina et al. (2003), in place of the control variable ELF. The results are slightly less strong, especially for the fractal measure, but the results for GDP and several health indicators remain strong. We then control for the percent of a country's land area that is desert. Borders may be more likely to be straight in deserts, and desert itself might influence our dependent variables of interest. However, controlling for desert leaves our results basically unaffected.

Another possible concern is to what extent our results reflect outcomes mainly in Africa. We have mixed feelings about introducing an African dummy variable into our regressions. On one hand, we are concerned that the Africa dummy is not truly exogenous because the decision to introduce an African dummy is influenced by the knowledge of poor outcomes in the endogenous variables in Africa (even the conventional definition of Africa as being countries below the Sahara has likely been influenced by the differing outcomes in North Africa and sub-Saharan Africa). On the other hand, it is clearly of interest to see whether our results are heavily influenced by the sub-Saharan African observations of very artificial borders and very poor outcomes. The results are definitely weakened by including the Africa dummy, which is always significant. The only result to survive with FRACTAL is for democracy (still significant at the 5 percent level). More of the results on PARTITIONED survive, with the result on per capita income level, literacy, measles immunization, and DPT immunization still significant at the 5 percent level, and corruption, clean water, and infant mortality still significant at the 10 percent level.

Finally, we control for two other important characteristics of countries that might be related to the nature of the borders (and thus possibly causing a spurious correlation with artificial borders): population density and the land area of the country. Population density is sometimes significant in our regressions, but leaves the results on PARTITIONED and FRACTAL basically unchanged. Land area is often significant and has some effect on the FRACTAL results, but little effect on the PARTITIONED results.

2.4.4 Borders and Failed States

In recent years, the phenomenon of "state failure," in which a nominal state fails to perform one or more of the core functions of governments, has received increasing attention from economists and policymakers. One obvious question for our paper is whether our measures of artificiality predict which states will become failed states. We use the classification of failed states developed by The Brookings Institute and the Center for Global Development (although we found similar results with using other measures such as Foreign Policy magazine's classification of failed states). Our probit regressions (Table 2.10) failed to confirm an effect of the "squiggliness" of borders, but the other measure of artificiality, partitioned, was a robust predictor of state failure. The result held when controlling for ethnic fractionalization and tropical climate (we could not control for colonial status, as there are no examples of non-colonial states that failed). We conclude that at least one of our measures of artificiality predicts whether the state will succeed in even being a state in the long run.

2.4.5 Borders and Wars

One type of variable is conspicuously missing in our analysis: wars, both international and civil. Our reason for not discussing it at length is that we found no effects of artificial borders on war. We did find an effect of artificial borders on a subjective measure of political instability and violence,

Table 2.10: **State Failure and Artificial States**
This table shows results for regressions in which the dependent variable is whether the state is a failed state, as defined by the CGA. Robust z statistics in parentheses. * significant at 5%; ** significant at 1%

Dependent variable:	Coefficient on:	1	2	3
Dummy variable for failed state (CGA)	PARTITIONED	**0.024**	**0.024**	**0.024**
		(5.08)**	**(3.27)****	**(3.33)****
	FRACTAL	-6.952	-14.483	-18.77
		-0.94	-1.2	-1.43
	ELF60		**0.017**	**0.014**
			(2.62)**	**(2.30)***
	CLIMATE			0.515
				-1.13
	Constant	**-1.026**	**-1.54**	**-1.492**
		(3.26)**	**(2.96)****	**(2.93)****
	Observations	117	76	76

as described above, but clearly it would be desirable to study the objective outbreaks of wars in addition to this variable.

The lack of an immediate and strong evidence of a correlation between borders and wars surprised us (although it echoes similar non-results in the literature on ethnic diversity and war). We are not ready to conclude that ethnic rivalries and border disputes are unrelated to wars: we believe that more work is needed. For international war, there is first of all the international system (mentioned for Africa in the introduction) that has tended to support existing borders no matter how artificial. These international conventions are more binding in some regions than others. Second, to study international wars properly, we need to study pairs of countries and to study to what extent the probability of war between them depends on whether the border dividing adjacent ones is artificial. There are clearly some examples of border wars arising from partition, such as Israel and its neighbors, India and Pakistan, and Eritrea and Ethiopia. To what extent these examples are validated by a systematic association requires a study that uses pairwise data on war outbreaks that is beyond the scope of this paper. For civil wars, a more detailed analysis would also require some attention to the nature of artificial states, especially finding some objective way of measuring whether previously hostile groups were combined into one state. The level of further work required for both civil and international war would unduly extend the length of this paper, so we plan a subsequent paper (not yet done) in which we focus exclusively on artificial states and war.

2.5 Conclusions

The idea of "failed states" is a recurrent them both in newspapers and within academia. The borders of many countries have been the result of processes that have little to do with the desire of people to be together or not. In some cases groups who wanted to be separate have been thrown into the same political unit; others have been divided by artificial borders. Former colonizers have been mainly responsible for such mistakes, but the botched agreements after the two major wars of the last century have also played a role.

The main contribution of this paper is to provide two new measures meant to capture how "artificial" political borders are. One measure considers how straight land borders are, under the assumption that straight borders are more likely to be artificial and less likely to follow geographic features or the evolution of hundreds of years of border design. The second measure focuses on ethnic or linguistic groups separated by borders. We have then investigated whether these variables are correlated with the political and economic success of various countries, and we found that indeed they are. The general patterns of correlations that we presented in a battery of tables suggest that these two new measures do quite well in cross-country regressions in which other exogenous measures of geography, ethnic fragmentation and colonial status are controlled for. We have also explored the correlation of our measures of artificial borders with the occurrence of civil and international wars and our results are inconclusive. While we find correlations of our variables with measure of political instability and lack of democracy, we do not find a clear pattern of correlations with wars. Further research is needed on this point looking at bilateral data on wars, namely which country engaged in war with whom.

Probably the single most important issue that we have not addressed is that of migrations. One consequences of artificial borders is that people may want to move, if they can. Often movement of peoples is not permitted by various government but migration certainly occur. In some cases migrations that respond to artificial borders may be partly responsible for economic costs, wars, dislocation of people, refugee crises and a host of undesirable circumstances. Thus, the need to migrate, created by the wrong borders may be one reason why artificial borders are inefficient. But sometimes the movement of people may correct for the artificial nature of borders. This dynamic aspects of movement of people and migrations, and changes of borders for that matter is not considered in this paper in which we consider a static picture of the world.

The bottom line in this paper is that the artificial borders bequeathed by colonizers are a significant hindrance to the political and economic development of the independent states that followed the colonies.

Chapter 3

Smoothing the Way or Stirring the Pot? Evidence Concerning the Impact of Foreign Aid on Anti-U.S. Sentiments and Terrorism [1]

3.1 Introduction

We consider the question of the degree to which foreign aid may lead to positive sentiments towards the donating nation among the general population in the aid-receiving country. Conceptually, if foreign aid is seen as supporting the growth and well-being of the country as a whole, then it is likely to lead to friendly feelings towards the aid donor. However foreign aid, particularly military aid, might also be seen as propping-up unpopular, authoritarian, or ethnically-divisive elites, leading the general population or an important segment of the population to have negative sentiments towards the donating nation. We consider one measure of anti-donor sentiment, namely anti-donor terrorism, and examine the correlation between foreign aid and terrorist incidents perpetrated by citizens of the aid-receiving country against citizens of the donor country.

Examining this correlation empirically, using data for the United States, we find that the correlation of U.S. foreign aid with anti-U.S. terrorism is statistically significant and very small; so small,

[1] This chapter is co-authored with Jennifer Stack.

in fact, as to be economically insignificant. For example a $1 increase in annual aid per capita (a 25-30 percentage increase in aid for the typical recipient country) for each year over a period of 20 years is associated with a difference of 0.03 in the number of U.S. citizens killed in terrorist incidents perpetrated by citizens of the aid-receiving nation. Interesting, this small, significant correlation is positive, with more foreign aid associated with more anti-U.S. terrorism, leading us to conclude that the mechanisms by which foreign aid can have a negative impact on sentiments towards the donor should not be lightly dismissed. We also find that U.S. military aid is associated with a larger increase in anti-U.S. terrorism than is U.S. economic aid. This result also supports the idea that mechanisms in which donor-nation support for an unpopular government can lead to anti-donor sentiments, since military aid is more likely than economic aid to directly increase the power of the central government.

Our research constitutes part of a growing literature on the political economy of aid, both its determinants and its effects. Alesina and Dollar (1998) find that Egypt and Israel receive by far the largest amounts of U.S. aid. But controlling for these two countries, U.S. aid tends to be given to poor, democratic, and free-trade countries. Alesina and Dollar (1999) show that democracies, corrupt countries, and U.S. allies tend to receive more U.S. foreign aid per capita.[2]

Other authors have explored the links among foreign aid, corruption, democracy, and growth. Burnside and Dollar (2000) show that aid only leads to growth for developing countries with "good" government policies[3], while countries with "bad" policies experience no additional growth. Recent work has both challenged and confirmed the Burnside and Dollar result, with the overall conclusions still outstanding. [Boone (1995); Easterly (2003); Easterly, Levine and Roodman (2003); Easterly, Levine and Roodman (2004); Economides, Kalyvitis and Philippopoulos (2004); and Burnside and Dollar (2004)].

The literature on foreign aid and corruption is also extensive [Boone (1995); Alesina and Dollar (1999); Knack (2001); Tavares (2003); and Economides et al. (2004)]. On balance, there appears to be evidence that aid increases corruption. Regarding other measurements of other political factors, evidence on the impact of foreign aid on democracy [Knack (2001)] and civil conflict [Collier and Hoeffler (2002a)] is inconclusive.

Krueger and Maleckova (2002) consider the link between poverty and terrorism. They show that historical hate crimes in the United States are largely independent of the economic situation of the

[2]The authors defined U.S. allies to be nations whose votes in the United Nations General Assembly correlate with U.S. votes.

[3]According to Burnside and Dollar [2000], these good policies include openness to trade, low inflation, a government budget surplus, and low government consumption

perpetrator. They also find that support for terrorism among residents of Middle Eastern countries may be the same or higher among the wealthy and educated as among the poor and uneducated.

Our analysis of the impact of aid on terrorism sheds more light on the relationships among aid, growth, corruption, and democracy, since terrorism is related to issues of governance in the aid-receiving countries. To address potential endogeneity problems, we instrument for U.S. foreign aid and also consider the within-country variation in aid using a fixed effects panel. Our main results are robust to the instrumented estimation and to other variations on our main specification.

3.2 Mechanisms

In conducting our empirical analysis, we have in mind several possible mechanisms by which U.S. foreign aid might impact the likelihood of an anti-U.S. terrorist incident. Certain mechanisms are likely to be more salient for military aid, and others more relevant for economic aid.

Several possible mechanisms could lead to more favorable sentiment towards the United States. Foreign aid could cause the economy to grow and improve living standards for the general population. If this were attributed to generosity from the United States, this could lead to a decline in anti-U.S. terrorism. The United States could also benefit from the perception of generosity even if the aid turns out to be unsuccessful at promoting growth. In reasonably stable countries, both of these mechanisms are more likely to occur with economic aid rather than military aid. In highly unstable countries, it might be the case the military aid also causes growth, through increased stability of the business climate.

Scenarios also exist, under which U.S. foreign aid causes an increase in anti-U.S. sentiment and anti-U.S. terrorism perpetrated by citizens of the recipient nations. United States support for corrupt governments or for governments that favor one group over other groups could cause citizens who are dissatisfied with their government to also resent the United States for supporting it. As marginalized groups often have little power to express their grievances, terrorism might be a particularly attractive way to try to have minority concerns addressed. Both mechanisms (corrupt governments and biased governments) in which U.S. support for the government is resented are more likely to be relevant for military aid, since it directly strengthens the recipient government. Economic aid may also affect views of the U.S. via these mechanisms, but to a lesser degree.

Alternatively, it could be the case that the recipient country's government is simply incompetent, and so the aid money is wasted with no consequent benefits from the financial assistance. To the extent that conditions have been attached to the aid by the United States, the recipient country might blame the lack of growth on these stringent requirements attached and thereby create animosity.

As these various mechanisms elucidate, whether and how U.S. foreign aid affects anti-U.S. terrorism is theoretically ambiguous. Indeed, some of these mechanisms may be at work in certain recipient countries, while other mechanisms may be relevant in other countries. Consequently, the question of the overall impact of foreign aid on terrorism is best answered empirically.

3.3 Data Description

Our data cover the 78 non-OECD countries[4] listed in Tables 3.1 and 3.2, and include information on U.S. foreign aid disbursements, terrorist attacks against U.S. citizens, and political and economic attributes for the recipient countries. Most data are available from 1960 to 2002; however several variables, including data on U.S. foreign aid, are available from 1940. Specific variables and data sources are described in Tables 3.3 and 3.4.

We construct several different indicators of anti-U.S. terrorism from the International Terrorism Attributes of Terrorist Events dataset (ITERATE), which contains information on over 12,000 separate international or transnational terrorist incidents from 1968-2002.[5] Our variables use information about these incidents to record the number of anti-U.S. terrorist attacks each year associated with a given country, as well as various measures of the severity of these attacks.

We select incidents based on whether the aid-receiving country's citizens are involved in perpetrating the incident and whether the incident affects the United States. A country's citizens are considered perpetrators of the incident if one or more terrorists are of that nationality, or if the incident occurs in that country. We count the incident as affecting the United States if the attack occurs on U.S. soil or if there is at least one U.S. victim.[6] We create a yearly panel for each terror-

[4]Since OECD countries rarely receive foreign aid from the U.S. we choose to exclude them from our analysis.

[5]The International Terrorism: Attributes of Terrorist Events (ITERATE) database [2003]. Edward Mickolus, Todd Sandler, Jean Murdock and Peter Flemming. This database defines international/transnational terrorism as "the use, or threat of use, of anxiety-inducing, extra-normal violence for political purposes by any individual or group, whether acting for or in opposition to established governmental authority, when such action is intended to influence the attitudes and behavior of a target group wider than the immediate victims and when, through the nationality or foreign ties of its perpetrators, its location, the nature of its institutional or human victims, or the mechanics of its resolution, its ramifications transcend national boundaries." International and transnational terrorism are differentiated, respectively, as being carried out by individuals or groups controlled by a sovereign state; or by autonomous non-state actors with or without some support from sympathetic states. We do not differentiate these two forms of terrorism when constructing our terrorism indicators

[6]As a robustness check, additional terrorism variables are constructed using selection criteria based solely on the nationalities of the victims and terrorists and not on the location of the event. For these variables, an incident was considered to have been perpetrated by the citizens of a particular country if at least one of the terrorists was of that nationality. The incident was considered to have affected the U.S. if at least one of the victims was a U.S. citizen. Our conclusions are not altered by the use of these terrorism variables

Table 3.1: **Countries for the Cross-Section Regressions**

	Regions				
	Latin America	**Asia E. Europe**	**Middle East N. Africa**	**Sub-Saharan Africa**	
	Bolivia/Brazil	China	Morocco	Angola/Botswana	Kenya
	Dominican Rep.	India	Syria	Burkina Faso	Madagascar
Countries with	Ecuador/Guyana	Malaysia	Tunisia	Cote d'Ivoire	Malawi/ Mali
ZERO	Jamaica/Mexico	Singapore		Cameroon	Mozambique
terrorism	Paraguay			Congo, Dem. Rep.	Niger/Senegal
(1980-2002)	Trin. & Tobago	Cyprus		Congo, Rep.	Sierra Leone
43 countries	Uruguay			Gabon/The Gambia	Tanzania/Togo
	Venezuela			Ghana/Guinea	Zambia/Zimbabwe
				Guinea-Bissau	
	Argentina	Indonesia	Algeria	Nigeria	
	Chile/Colombia	Korea, Rep.	Egypt	Uganda	
Countries with	Costa Rica	Philippines	Iran		
NON-ZERO	El Salvador	Sri Lanka	Jordan		
terrorism	Guatemala	Thailand	Turkey		
(1980- 2002)	Haiti/Honduras				
23 countries	Nicaragua				
	Panama/Peru				

Table 3.2: **Countries for the Panel Regressions**

	Regions				
	Latin America	**Asia E. Europe**	**Middle East N. Africa**	**Sub-Saharan Africa**	
	Ecuador	China	Morocco	Benin/Botswana	Kenya/Lesotho
	Paraguay	Fiji	Tunisia	Burkina Faso	Madagascar
Countries with	Trin. & Tobago	Nepal		Cameroon	Malawi/Mali
ZERO		N. Guinea		Central Afr. Rep/Chad	Mauritania
terrorism		Singapore		Comoros/ Congo, Rep	Mauritius
(1980-2002)				Cote d'Ivoire	Mozambique
40 countries		Cyprus		Eq. Guinea/Ethiopia	Niger/Senegal
				Gabon/Gambia	Sierra Leone/Togo
				Ghana/Guinea-Bissau	Zambia/Zimbabwe
	Argentina	Bangladesh	Algeria	Burundi	
	Bolivia/Brazil	India	Egypt	Congo, Dem. Rep.	
Countries with	Chile/Colombia	Indonesia	Iran	Nigeria	
NON-ZERO	Costa Rica/Dom. Rep.	Malaysia	Jordan	Rwanda	
terrorism	El Salvador	Pakistan	Syria	Uganda	
(1980- 2002)	Guatemala/Guyana	Philippines	Turkey		
38 countries	Haiti/Honduras	Sri Lanka			
	Jamaica/Mexico	Thailand			
	Nicaragua/Panama				
	Peru/Uraguay/Venezuala				

Table 3.3: **Data Sources: Terrorism and U.S. Foreign Aid Variables**

Database Name and Source	**Data Type**	**Time Period**	**Variables**
ITERATE - (2003) Sandler et al.	terrorism data	1968-2002	usvictims, uswounded, uscasualties, totcasualtiesnt, damage, incidentcnt,statespns, usgovt
OECD DAC database - OECD	U.S. foreign aid data	1960-2002	usaid_net usaid_grants
USAID Greenbook	U.S. foreign aid data	1946-2002	econassistg, econassistlg, econmilassist, foodaid,militassist, nonprlfantiterr, peacecorps

Table 3.4: **Data Sources: Control Variables**

Database Name and Source	Data Type	Time Period	Variables
World Development Indicators: World Bank Group	time varying	1960-1999	Age_dep_ratio, aid_fracgni, aidpc, gov_debt, claims_on_govs, claims_on_private_sector, cpi, crops, current_revenue_no_grant, x_capacity_m, exports, external_debt, financing_from_abroad, fdi, fdi_netinflow, gdp, gdp_growth, gdp_pc, gov_cons, illit, life_exp, net_income_from_abroad, official_aid, exchange_rate, pop_growth, pop, taxrev, debt_service_total, trade_percent_gdp, urban_pop
Freedom House	time varying	1972-2001	political rights, civil liberties
Polity IV	time varying	1940-2002	polity2
International Country Risk Group	time varying	1980-2002	burqual, corruption, demaccount, ethnictension, externalconflict, internalconflict, militaryinpol, religioustension
Easterly-Levine	time varying	1960, 1970, 1980	anti_gov_demo, assassin, blk_mkt_prem, genocide, coups, log_schooling, purges, revolutions, war, civil_war
UN	time varying	1950-2000	un_pop
Penn World Tables	time varying	1950-2000	Gdp_pc
La Porta R., A. Shleifer, and R.Vishny (2000)	non-time varying	N/A	Bur_delay, infrast, tax_compliance, school_attainment, ethno_frac, catholic, muslim, political, democ, lattitude_capital, property_rights, soe
Ethno-linguistic fractionalization	non-time varying	N/A	Elf61, elf85
CIA factbook	non-time varying	N/A	year of independence

ism variable (U.S. casualties, total casualties, etc.), by summing the attribute for all of the selected incidents associated with a particular country in a particular year.

Our principle measure of terrorism, "US casualties" reflects the total number of U.S. citizens killed or wounded. Our "US victims" variable is based on the wider definition of a U.S. "victim" (anyone directly affected by the event: casualties, kidnapped, and/or losing property). Our "total casualties" variable measures the overall size of the incident. This variable may be biased by the inclusion of attacks which wound U.S. citizens but are not directly targeted at the United States. Consequently, we prefer to use "US casualties" in our main regressions. However, we do use "total casualties" for our panel regressions because it is available over a longer time period.[7] For our Poisson regressions, we use the "incident count" variable, which tallies the number of incidents; and we use a binary incident dummy variable, "incident dummy," for our Logistic regressions.[8] As a robustness check, we repeat our main regressions with all of these terrorism variables.[9]

The principle characteristic of our data is that the majority of countries do not perpetrate terrorism against the United States. Summary statistics for our terrorism data (Table 3.5) reveal that over half of our data is zeros, for both the cross-section and panel specifications. We use Tobit regressions in our principle analyses because of this censoring.[10]

Data for United States foreign aid come from the USAID Greenbook and are available from 1946 to 2002. We focus on three data series: economic aid in the form of loans and grants; military aid; and total economic plus military aid. We also consider Greenbook data on economic aid in the form

[7] For example, if a terrorist attack targeted at country X happens to wound five hundred of country X's citizens and only one American, this event will still be recorded as 500 in the "totcasualties" variable.

[8] Due to characteristics of the ITERATE database, we also use two different criteria for determining whether any U.S. citizens were "affected" by a particular incident. One criterion we use is whether there were any "U. S. victims" of the attack. However, because of the aforementioned ambiguity in the definition of a victim, this is not entirely satisfactory. We prefer to use the criteria of whether there were any U.S. citizens wounded in the incident, which is a more precise identification. Unfortunately, this second criterion cannot be used for the years from 1968 to 1977, because there is no information in the database on the number of U.S. citizens wounded for these years. Still, regressions run using the years for which data on both criteria are available produce similar results.

[9] Information on monetary damages is available in the ITERATE database, but the coding method leads too much noise for the information to be useful. Information about damage was coded in ranges (none; $0 = $10,000; $10,000 = $100,000; $100,000 = $1 million; and = $1 million). To combine this information across incidents, we assumed that the damage from a given incident was the midpoint of each range, or equal to $1 million in the case of the last group. Because of this averaging method, our "damage" variable turned out to be quite noisy.

[10] With reference to the Tobit model, one could also interpret these data as reflecting the degree of friendliness towards the United States. Because there is no possibility of a "negative terrorism event," we cannot distinguish between those countries which are moderately friendly towards the United States from those that are extremely friendly, since both types of countries do not promote terrorism against the United States

Table 3.5: **Summary statistics for selected terrorism and foreign aid variables**

Variable	# of Non-Zero Obs	Max Value	Mean	Std. Dev.
Cross-section regressions (66 countries)				
Terrorism variables (total 1980-2002)				
Uscasualties	21	66	4.61	11.3
Usvictims	22	79	5.44	14.6
Totcasualties	21	520	24.5	75.9
Incidentcnt	22	14	1.2	2.33
Aid variables (average annual per capita amount 1960-80)				
economic aid	65	46	4.06	6.89
military aid	57	20.4	0.89	3.54
economic plus military aid	65	66.4	4.95	9.51
Panel regressions (79 countries, 254 obs)				
Terrorism variables (decade totals)				
Uscasualties	47	65	1.35	5.74
Usvictims	57	125	2.57	11.9
Totcasualties	66	235	7.26	27.1
Incidentcnt	82	162	3.14	13.1
Aid variables (decade averages of annual per capita aid)				
economic aid	247	$56.40	4.16	7.19
military aid	211	$25.20	0.93	3.24
economic plus military aid	249	$76.30	5.09	9.2

of grants only; OECD DAC data on U.S. foreign aid (available from 1960 onwards); food aid; and funding for the Peace Corps.[11] All data on United States foreign aid are converted into per capita terms for the recipient country.

Political and cultural controls include the PolityIV index of democracy (Polity2), Freedom House indices on political rights and civil liberties, as well as indices on corruption and bureaucratic quality from the International Country Risk Group. Economic control variables from the World Bank's World Development Indicators, the United Nations, and the Penn World Tables include GDP per capita, population, total foreign aid received, life expectancy, urban population, trade, and FDI measures. Further information on all variables and their sources can be found in Tables 3.15 and ??. Due to noise in the yearly panel, we consider five, ten or twenty year averages of the variables and we scale the terrorism variables to reflect total terrorism for the period rather than an average.

Before running our regressions, we drop two countries (in addition to the OECD members) from our dataset. Israel is dropped because, although including it in our sample results in a stronger positive correlation between our measures of anti-U.S. terrorism and U.S. foreign aid, this is due to the specific way in which we construct our terrorism measure. Specifically, Israel receives a large amount of aid from the United States, and a large number of U.S. citizens are injured in terrorist attacks which are primarily targeted at Israel, not the United States.[12] We also drop Saudi Arabia from the dataset, since it is a significant outlier and only strengthens our result.[13] Additional potential outlier countries are dropped as a robustness check.

3.4 Results

Our main specification is a cross-section Tobit regression in which we regress terrorism from 1980-2002 on U.S. foreign aid from 1960-1979 and country-specific control variables. We first discuss this specification in detail and then compare the effects of economic aid, military aid, and foreign direct investment. Later in this section, we address the endogeneity of foreign aid and terrorism, as well

[11] One issue for both the Greenbook and the OECD measures of aid concerns the treatment of loans. To the extent that a loan is offered on concessional terms. Chang, Fernandez-Arias and Serven (1999) develop a dataset which calculates the grant equivalent of all loans, using information about concurrent interest rates. Unfortunately, this calculation is only available for the *total* foreign aid to each developing nation

[12] We do include Israel for our robustness check that considers terrorist events only based on the nationalities of the terrorists and victims and our results are not affected

[13] The fact that Saudi Arabia is an outlier for our results is not dependent on the terrorist incidents of September 11, 2001. Because of the nature of the coding in the terrorism database, the September 11 attacks are top-coded at 999 in several categories and thus these incidents are automatically dropped from our analysis

as potential omitted variable bias.

3.4.1 Tobit Cross-Section Results

We begin by considering the impact of average U.S. foreign aid from 1960 to 1979 on terrorism from 1980 to 2002, as measured by total U.S. casualties. One advantage of this specification is that, by using foreign aid data that predates the terrorism data, we address some of the potential endogeneity issues as it is less likely that foreign aid would be given in anticipation of terrorism several decades in the future.

Results from our Tobit cross-section regression are shown in Table 3.6, Column 1. Table TAB:Aid3 also presents OLS, Poisson, and Logit regressions for comparison purposes.[14] In all four specifications there exists a positive relationship between total U.S. economic and military aid given to countries in the 1960s and 1970s and terrorism originating from those countries in the 1980s and 1990s.

For the Tobit regression, the reported coefficients reflect the conditional marginal effect, which is the increase in U.S. casualties, for each additional dollar of U.S. foreign aid, among those countries that already perpetrate terrorism against the United States. This effect is extremely small in magnitude: *A one dollar increase in annual per capita foreign aid, given over a twenty-year period, is associated with a 3% increase in the probability of one additional U.S. citizen being killed or wounded in a terrorist attack associated with the recipient country, over the subsequent twenty years.* A typical developing country receives approximately three to four dollars per capita in foreign aid each year from the United States, and so an increase of a dollar per capita per year represents a 25-30% increase in the amount of U.S. foreign aid. Although very small, since this correlation is statistically significant, we can say with some confidence that there is no large effect of foreign aid *decreasing* terrorism, at least in the case of the United States.

Our main specification includes controls which are contemporaneous with the terrorism variables. Also, to control for omitted variable bias, we include controls that are contemporaneous with the foreign aid variables. We use only one lag of the highly auto-correlated control variables, allowing us to add as many factors as possible while avoiding potential co-linearity problems. Specifically, we control for democracy (Polity2), civil liberties, percent Muslim, ethno-linguistic fractionalization, corruption, bureaucratic quality, life expectancy, GDP per capita, total non-U.S. foreign aid, trade, population (level and squared terms), and urban population. Of these twelve controls, seven are

[14]The Poisson regression can be interpreted as the impact of aid on the number of terrorism incidents. The Logit regression reflects aid's impact on whether a country's citizens perpetrate any terrorist incidents against the U.S.

Table 3.6: **Effect of US Foreign Aid on Anti-US Terrorism**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. For the Tobit specification, the coefficient given is the marginal effect, conditional on being uncensored. Also controlling for population, urban population, life expectancy, and trade as a percentage of GDP. The terrorism variables used are from the 1980 to 2002 period. For the Tobit, OLS, and Poisson specifications we use the U.S. casualties variable. For the Logit specification we use the incident dummy variable. Each regression has 66 observations.

		Tobit	OLS	Poisson	Logit
US foreign aid in the	Economic and	0.031	1.052	0.126	0.575
1960s and 1970s	military aid	(0.007)***	(0.224)***	(0.039)***	(0.268)**
Controls from the	GDP per capita	-0.0001	-0.001	0	0
1980s and 1990s		(0.000)***	(-0.001)	(0.000)**	(0)
	Total non-US	-0.002	-0.012	-0.023	-0.031
	foreign aid	(-0.002)	(-0.063)	(0.013)*	(-0.041)
	Polity2	-0.009	-0.062	0.195	-0.14
	(democ/autoc)	(-0.01)	(-0.398)	(0.092)**	(-0.196)
	Civil liberties	0.002	1.892	1.083	1.329
		(-0.042)	(-1.82)	(0.381)***	(0.742)*
	Percent Muslim	-0.001	0.019	-0.001	-0.021
		(-0.001)	(-0.037)	(-0.006)	(-0.02)
	Corruption	0.074	0.977	0.329	0.597
		(0.034)**	(-1.501)	(-0.248)	(-0.757)
	Bureaucratic	0.058	-0.016	0.623	-0.875
	quality	(-0.065)	(-2.434)	(-0.516)	(-1.257)
Controls from the	GDP per capita	0.00005	0.001	0	0
1960s and 1970s		(0.000)**	(-0.001)	(0)	(0)
	Total non-US	-0.005	-0.055	-0.022	-0.153
	foreign aid	(0.002)**	(-0.069)	(0.010)**	(0.080)*
	Polity2	0.003	-0.122	-0.199	0.164
	(democ/autoc)	(-0.008)	(-0.355)	(0.074)***	(-0.164)
	Civil liberties	0.054	-0.05	-0.739	0.976
		(-0.039)	(-1.576)	(0.345)**	(-0.852)
	Ethnoling	-0.325	-0.14	-2.825	-1.728
	frac	(0.143)**	(-6.5)	(0.963)***	(-2.791)

significant at the 10% level or higher in our main specification. In research available from the authors, we consider these control variables in detail, alone and also interacted with the foreign aid variables.

3.4.2 Economic Aid, Military Aid, and Foreign Direct Investment

We compare the relative impact of economic aid, military aid, and foreign direct investment in Tables 3.7 and 3.8. Interestingly, we find that military aid has a larger and more robust correlation with terrorism than economic aid. When either economic aid or military aid is included in the regression by itself, the coefficient is positive and significant (Table ??, regressions 1 and 2) and the magnitude for military aid is larger. However, when both economic and military aid data are included in the same regression, military aid remains significant, while economic aid becomes insignificant (Table ??, regression 3). The magnitude of both coefficients is smaller in this case, although the drop appears to be much larger for economic aid.[15]

The positive correlation of terrorism with foreign aid indicates that we should seriously consider the possible mechanisms detailed in Section 3.2 by which foreign aid can be associated with an increase in terrorism. The two mechanisms in which U.S. support for a corrupt or a biased government led to anti-U.S. sentiment are also the mechanisms which are more likely to be relevant for military aid, rather than economic aid, as military aid directly strengthens the central government. Thus, the result that military aid may have a more important detrimental impact than economic aid also supports these two mechanisms.

One possible explanation of our results is that it is the presence of the United States in the aid-receiving country that matters, putting U.S. citizens within easy reach of those wishing to commit terrorist acts against the U.S. It is therefore interesting to consider the correlation of foreign direct investment (FDI), another indicator of the presence of the U.S., with terrorism. We use total world FDI, which should be highly correlated with United States FDI. The coefficient on world FDI is negative, although it is not significant when FDI is included by itself (Table ??, regression 5). When economic and/or military aid are added to the regression, the coefficient on FDI becomes significant and remains negative (Table ??, regressions 7-10). In all of these regressions, the order of magnitude of the coefficient on FDI is the same as for economic and military aid.

Thus FDI, which along with foreign aid is associated with an increased profile for the United

[15] One possible cause of this result is that economic and military aid are collinear, with military aid picking up the significance when both variables are in the regression. We argue that collinearity is not driving this result, because many countries (12%–15% in our cross-section sample) receive economic aid from the United States, but no military aid.

Table 3.7: **Comparison of Foreign Aid with FDI**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Also controlling for GDP per capita, total non-US foreign aid, Polity2 (democracy/autocracy), civil liberties, percent Muslim, life expectancy, trade as a percent of GDP, percent urban population, corruption, bureaucratic quality, ethno-linguistic fractionalization, population and population squared, as in our main specification. For comparison purposes, six countries with a missing value for 1960s and 1970s FDI were dropped from all the regressions. Coefficients reflect the marginal effect, conditional on being uncensored. Each regression has 59 observations.

		Specification: Tobit				
		Effect on terrorism from 1980 to 2002, as measured by US casualties:				
		(1)	(2)	(3)	(4)	(5)
1960s and 1970s	Economic aid from the US	0.039 (0.010)***		0.004 (-0.004)		
	Military aid from the US		0.062 (0.014)***	0.055 (0.016)***		
	Combined economic and military aid from the US				0.03 (0.007)***	
	Total world FDI for the RC					-0.1 (-0.14)

Table 3.8: **Continued: Comparison of Foreign Aid with FDI**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Also controlling for GDP per capita, total non-US foreign aid, Polity2 (democracy/autocracy), civil liberties, percent Muslim, life expectancy, trade as a percent of GDP, percent urban population, corruption, bureaucratic quality, ethno-linguistic fractionalization, population and population squared, as in our main specification. For comparison purposes, six countries with a missing value for 1960s and 1970s FDI were dropped from all the regressions. Coefficients reflect the marginal effect, conditional on being uncensored. Each regression has 59 observations.

Specification: Tobit					
Effect on terrorism from 1980 to 2002, as measured by US casualties:					
		(6)	(7)	(8)	(9)
1960s and 1970s	Economic aid from the US	0.048 (0.011)***		0.006 (-0.004)	
	Military aid from the US		0.058 (0.012)***	0.045 (0.014)***	
	Combined economic and military aid from the US				0.035 (0.007)***
	Total world FDI for the RC	-0.126 (0.054)**	-0.028 (0.014)**	-0.034 (0.015)**	-0.117 (0.045)***

States, appears to decrease terrorism. One implication is that it may not be simply the presence of the United States in the aid-receiving country which is leading to increased anti-U.S. sentiment. Instead it appears to be another aspect of foreign aid, perhaps related to the direct interaction of the government of the recipient country with the U.S. government, which causes anti-U.S. sentiment to increase.[16]

3.4.3 Endogeneity and Omitted Variable Bias

We address the important issue of endogeneity between foreign aid and terrorism using several techniques. We begin by instrumenting for foreign aid in our cross-section Tobit regression. Next, because omitted characteristics of the aid-receiving countries might be simultaneously influencing both foreign aid disbursements and anti-U.S. terrorism, we use a fixed-effects, panel specification. In this specification, country-specific omitted variables are subsumed into a fixed effect for each country. Since coefficient estimates using a Tobit fixed effects panel are not consistent, we use an OLS specification for these regressions. Finally, we instrument for U.S. foreign aid in this OLS fixed-effects panel specification, addressing endogeneity and omitted variable bias simultaneously.

One common instrument used in growth regressions and in the literature on foreign aid is lagged foreign aid. Foreign aid is reasonably correlated over time, with a correlation coefficient of between 0.6 and 0.8 from half-decade to half-decade.[17] Thus, past foreign aid is a good predictor of current foreign aid. In addition, correlations between terrorism and distant lags of aid should be much smaller than between terrorism and more recent lags of aid.

We also instrument for U.S. foreign aid with contemporaneous foreign aid from the European Union. Countries that receive large aid disbursements from the United States may also receive aid from the European Union if, for example, they are in great need of economic development assistance. This correlation between E.U. and U.S. aid may have little to do with the determinants of terrorism. Furthermore, E.U. aid is much less likely to be given in an attempt to prevent future anti-U.S. terrorism, although it may be trying to prevent anti-E.U. terrorism, and so may have a second-order effect of reducing anti-U.S. terrorism.

Our results are consistent with our main specification when we use lagged U.S. foreign aid or E.U. foreign aid as an instrument for current U.S. foreign aid (Table 3.9; Panels A, B). In all cases,

[16]One caution with regard to our FDI results is a potential selection problem which would bias our results in the direction observed: companies typically select to invest in countries that are economically and politically stable, and so are not likely to be sources of terrorist events.

[17]By contrast our terrorism data are not highly auto-correlated, with correlation coefficients between -0.1 and 0.3

the coefficients remain positive, and in all but one case they are also significant. Furthermore, four of the five significant coefficients on foreign aid are larger in magnitude than in the non-instrumented specifications.[18]

In addition to endogeneity, we are also concerned that certain unobserved aspects of the recipient countries are simultaneously causing them to generate more terrorism and to receive more foreign aid. To address this issue, we turn to an OLS fixed-effects panel specification (Table 3.10). Here, we use decade averages for both foreign aid and terrorism. We only consider terrorism incidents from the 1970s through the 1990s because of a lack of terrorism data for the 1960s. In addition, we use the "totcasualties" variable because data limitations prevent us from constructing our "uscasualties" variable for years prior to 1978.[19] In these regressions, our military aid result is robust, with the coefficient remaining positive and significant. This is also true for the total aid variable. The coefficient on economic aid is no longer significant, although it is still positive when entered in the regression by itself.

We address both the endogeneity and omitted variable bias issues by instrumenting for U.S. foreign aid in the fixed effects OLS panel specification (Tables 3.11 and 3.12). We examine the effect of three different instruments: a one decade lag of U.S. foreign aid, a two decade lag of U.S. foreign aid, and contemporaneous E.U. foreign aid. For each instrument we consider the effect of foreign aid in a given decade on terrorism in that decade. Our results for economic aid and total aid are no longer significant when we instrument in this panel setting. However, the coefficient for military aid remains significant and positive, even increasing in size, when we use lagged U.S. foreign aid as the instrument (Table 3.11).[20] The failure of our economic aid and total aid results to be robust to the panel IV specification may be due in part to our use of an OLS model instead of the more appropriate Tobit model.

[18]We also consider two completely exogenous instruments, population and physical distance to the United States. Empirically, a smaller population is associated with a higher level of per capita foreign aid, possibly because of the "fixed costs" of establishing an aid mission, or a desire by developed countries to "plant their flag" in many countries around the world. Alesina and Dollar (1998) Physical distance is also a plausible instrument on the grounds that more distant countries might be less likely to receive aid, either because of transportation costs, or because of less public attention to and knowledge of the plight of citizens of that country. Unfortunately, it turns out that both of these instruments are weak, producing large standard errors

[19]No information is available in the ITERATE database on the number of U.S. wounded or killed, for incidents that occurred prior to 1978.

[20]None of our results remain when we use E.U. foreign aid as an instrument in this setting, however this may be due to our use of an OLS instead of a Tobit specification

Table 3.9: **Instrumental Variables with Tobit Cross-section Regressions**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Also controlling for GDP per capita, total non-US foreign aid, Polity2 (democracy/autocracy), civil liberties, percent Muslim, life expectancy, trade as a percent of GDP, percent urban population, corruption, bureaucratic quality, ethno-linguistic fractionalization, population and population squared, as in our main specification. Each regression uses 66 observations. Terrorism in the 1980s and 1990s is measured by the U.S. casualties variable. Tobit and Tobit Instrumental Variables (IV) specifications are shown.

Panel A - Instrument for aid: Lagged US foreign aid (1950-54 average)

		Regression Number					
		1	1 IV	2	2 IV	3	3 IV
US foreign aid	Economic aid	3.514	7.836				
(1960s and		(0.815)***	(2.679)***				
1970s)	Military aid			8.283	7.989		
				(1.893)***	(3.793)**		
	Economic and					2.77	3.943
	military aid					(0.591)***	(1.160)***

Panel B - Instrument for aid: EU foreign aid in the 1960s and 1970s

		Regression Number					
		1	1 IV	2	2 IV	3	3 IV
US foreign aid	Economic aid	3.514	10.043				
(1960s and		(0.815)***	(-12.67)				
1970s)	Military aid			8.283	14.863		
				(1.893)***	(4.915)***		
	Economic and					2.77	5.794
	military aid					(0.591)***	(3.069)*

Table 3.10: **Instrumental Variables with Tobit Cross-section Regressions**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Also controlling for life expectancy, trade as a percentage of GDP, percent urban population, population and population squared. All regressions use an OLS fixed effects panel specification. Each sample includes 104 countries and 254 observations. Terrorism is measured by our total casualties variable for each regression.

		Regression Number			
		1	2	3	4
US foreign aid	Economic aid	0.417		-0.286	
(contemporaneous		(-0.345)		(-0.388)	
decade averages)	Military aid		3.169	3.559	
			(0.870)***	(1.019)***	
	Economic and				0.569
	military aid				(0.275)**
Controls	GDP per capita	0.003	0.003	0.003	0.003
(contemporaneous		(-0.002)	(0.002)*	(-0.002)	(0.002)*
decade averages)	Total non-US	0.199	0.187	0.188	0.195
	foreign aid	(0.055)***	(0.053)***	(0.053)***	(0.054)***
	Polity2	-0.761	-0.504	-0.424	-0.762
	(democ/autoc)	(-0.534)	(-0.511)	(-0.523)	(-0.526)

Table 3.11: **Instrumental Variables with Panel Fixed Effect Regressions: Panels A and B**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Also controlling for life expectancy, trade as a percentage of GDP, percent urban population, population and population squared. Our independent variable is the decade sums of terrorism ('70s through '90s) as measured by the total casualties variable. We use the contemporaneous decade averages of U.S. foreign aid as our independent variables. The samples for Panels A and B include 106 countries and 254 observations.

Panel A - Instrument for aid: US foreign aid lagged by one decade						
	OLS	OLS IV	OLS	OLS IV	OLS	OLS IV
Economic aid	0.417	-0.331				
	(-0.345)	(-1.222)				
Military aid			3.169	11.08		
			(0.870)***	(6.355)*		
Economic and					0.569	-1.162
military aid					(0.275)**	(-1.493)
Panel B - Instrument for aid: US foreign aid lagged by two decades						
OLS	OLS IV	OLS	OLS IV	OLS	OLS IV	
Economic aid	0.417	4.436				
	(-0.345)	(-5.675)				
Military aid			3.169	16.87		
			(0.870)***	(5.389)***		
Economic and					0.569	-9.673
military aid					(0.275)**	(-11.53)

Table 3.12: **Instrumental Variables with Panel Fixed Effect Regressions: Panel C**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Also controlling for life expectancy, trade as a percentage of GDP, percent urban population, population and population squared. Our independent variable is the decade sums of terrorism ('70s through '90s) as measured by the total casualties variable. We use the contemporaneous decade averages of U.S. foreign aid as our independent variables. The sample for Panels C includes 95 countries and 243 observations.

Panel C - Instrument for aid: Contemporaneous EU foreign aid						
	OLS	OLS IV	OLS	OLS IV	OLS	OLS IV
Economic aid	0.417	-2.967				
	(-0.345)	(-2.713)				
Military aid			3.169	-142.661		
			(0.870)***	(-1370)		
Economic and					0.569	-2.9
military aid					(0.275)**	(-2.955)

3.4.4 Political and Economic Control Variables

In order to shed light on possible mechanisms for our general findings on foreign aid, we consider our basic country-specific controls, as well as additional interaction terms. Of the twelve controls that are included in our main Tobit cross-section specification (Table 3.6, regression 1), seven are significant at the 10% level or higher: GDP per capita, life expectancy, trade as a percentage of GDP, urban population, corruption, ethno-linguistic fractionalization, and population. While our basic specification only contains these control variables in level terms, Tables 3.13 and 3.14 display regressions including the interaction terms for each of these controls.

Table 3.13 contains results for the political control variables.[21] Regression 1 is our main specification and subsequent columns add interaction terms one variable at a time, both during the time period of the foreign aid data (1960-70s) and during the time period of the terrorism data (1980-90s). Our measure of democracy, Polity2, is not significant by itself, however the interaction between foreign aid in the 1960-70s and Polity2 in the 1980-90s is negative and significant. A higher value for Polity2 reflects a more democratic government, implying that more foreign aid given in the past to countries which are currently less democratic, results in relatively more current terrorism. This result lends support to the theory that funding autocratic regimes increases anti-U.S. terrorism. By itself, the percentage of the recipient country that is Muslim does not have a significant coefficient, but the sign is negative. When we include the interaction of the percent Muslim variable with U.S. foreign aid, the coefficient on the level term is negative and significant, while the coefficient on the interaction term is positive and significant. Thus a higher percentage of Muslims makes the recipient country's citizens less likely to cause terrorism, but giving higher amounts of aid to countries with a higher percentage of Muslims increases terrorism generated by that country.

Our results for corruption and bureaucratic quality are surprising. For both of those variables, a higher value indicates a more favorable policy environment. For corruption, in levels, less corrupt governments are associated with higher levels of terrorism. Bureaucratic quality by itself is not significant. However, the interaction terms for corruption and bureaucratic quality are both positive and significant, implying that giving more foreign aid to "better" governments results in more terrorism.

Ethno-linguistic fractionalization is robustly significant across all of the regressions displayed in Table 3.13, except the last regression, when an interaction term for this variable is included (regression 7). The coefficient is always negative, implying that the more ethnically fragmented a

[21] All seven regressions also include the economic controls in levels, as in the basic specification, although these coefficients are not shown.

Table 3.13: **Interaction Terms for our Main Controls: Political controls** Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions also control for GDP per capita, total foreign aid, life expectancy, trade as a percent of GDP, and the percent urban population, population and population squared, as in our basic Tobit cross-section specification. "Decades" refers to the time period of the control variables; economic and military aid is always from the 1960s and 1970s. N/A indicates that the data we have for these variables is non-time varying. All samples include 66 observations.

		LHS variable: US casualties from 1980-2002						
Variable	Decades	1	2	3	4	5	6	7
Economic and	60-70s	2.77	4.599	0.946	1.282	-2.097	-4.952	3.089
military aid		(0.591)***	(1.851)**	(-4.213)	(-0.887)	(-2.724)	(2.645)*	(1.264)**
Polity2	60-70s	0.232	1.329	-0.038	0.108	-0.059	-0.628	0.228
(democracy)		(-0.705)	(-0.998)	(-0.75)	(-0.725)	(-0.738)	(-0.692)	(-0.705)
	80-90s	-0.751	0.055	-0.23	-1.79	-0.335	-0.853	-0.726
		(-0.896)	(-0.895)	(-1.076)	(1.062)*	(-0.951)	(-0.846)	(-0.899)
Civil liberties	60-70s	4.718	7.56	0.529	4.217	5.115	1.449	4.716
		(-3.474)	(-4.559)	(-4.406)	(-3.619)	(-3.600)	(-3.332)	(-3.478)
	80-90s	0.206	-7.083	1.298	-3.557	-0.783	-1.594	0.275
		(-3.739)	(-5.267)	(-6.295)	(-4.158)	(-3.951)	(-3.735)	(-3.733)
Percent Muslim	N/A	-0.128	-0.018	-0.045	-0.293	-0.056	-0.126	-0.129
		(-0.088)	(-0.103)	(-0.103)	(0.121)**	(-0.098)	(-0.088)	(-0.088)
Corruption	80-90s	6.497	7.197	6.815	7.679	-2.428	1.772	6.153
		(3.001)**	(2.954)**	(3.066)**	(3.066)**	(-6.119)	(-3.002)	(3.216)*
Bureaucratic	80-90s	5.138	4.701	7.083	4.471	11.39	-1.132	4.923
quality		(-5.746)	(-6.039)	(-5.77)	(-5.962)	(6.715)*	(-5.899)	(-5.792)
Ethno-linguistic	N/A	-28.61	-38.307	-35.708	-34.475	-38.209	-35.9	-24.707
fractionalization		(12.58)**	(13.30)***	(13.68)**	(13.10)**	(14.19)***	(11.91)***	(-18.500)
				Interactions:				
Polity2 *	60-70s		-0.053					
Foreign aid			(-0.135)					
	80-90s		-0.535					
			(0.281)*					
Civil libs *	60-70s			0.714				
Foreign aid				(-0.57)				
	80-90s			-0.613				
				(-0.955)				
Percent Muslim *	N/A				0.045			
Foreign aid					(0.022)**			
Corruption *	80-90s					1.553		
Foreign aid						(0.897)*		
Bureaucratic qual *	80-90s						3.652	
Foreign aid							(1.282)***	
Ethnolinq frac *	N/A							-0.735
Foreign aid								(-2.562)

Table 3.14: **Interaction Terms for our Main Controls: Economic controls** Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions also control for Polity2, civil liberties, percent Muslim, corruption, bureaucratic quality, ethno-linguistic fractionalization, population and population squared, as in our basic Tobit cross-section specification. "Decades" refers to the time period of the control variables; economic and military aid is always from the 1960s and 1970s. All samples include 66 observations.

		LHS variable: US casualties from 1980-2002					
Variable	Decades	1	2	3	4	5	6
Economic and	60-70s	2.77	2.987	0.592	-27.89	-10.095	2.783
military aid		(0.591)***	(-2.586)	(-0.725)	(-16.731)	(3.145)***	(0.718)***
GDP per capita	60-70s	-0.421	-0.428	-2.977	-0.412	-1.935	-0.424
		(0.183)**	(0.232)*	(0.640)***	(0.184)**	(0.470)***	(0.204)**
	80-90s	-0.207	-0.206	0.022	0.005	-0.155	-0.206
		(-0.163)	(-0.170)	(-0.227)	(-0.198)	(-0.163)	(-0.163)
Total non-US	60-70s	1.251	1.256	1.81	0.694	1.842	1.242
foreign aid		(0.579)**	(0.633)*	(0.539)***	(-0.5800)	(0.639)***	(0.644)*
	80-90s	0.004	0.005	0.001	0.004	-0.003	0.004
		(0.002)*	(-0.003)	(-0.002)	(0.002)*	-0.003	(0.002)*
Life expectancy	80-90s	-0.008	-0.008	-0.011	-0.007	-0.01	-0.008
		(0.002)***	(0.003)**	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Trade as a	80-90s	-0.217	-0.215	-0.59	-0.306	-1.292	-0.218
percent of GDP		(-0.132)	(-0.141)	(0.162)***	(0.132)**	(0.359)***	(-0.133)
Percent urban	80-90s	-0.879	-0.89	-1.814	-0.937	-1.612	-0.876
population		(0.349)**	(0.367)**	(0.415)***	(0.327)***	(0.384)***	(0.359)**
			Interactions:				
GDP per capita *	60-70s		0				
US Foreign aid			(-0.001)				
	80-90s		0				
			(-0.001)				
Total non-US	60-70s			0.049			
foreign aid *				(0.020)**			
US Foreign aid	80-90s			-0.006			
				(-0.027)			
Life expectncy*	80-90s				0.43		
US Foreign aid					(0.235)*		
Trade % GDP *	80-90s					0.153	
US Foreign aid						(0.039)***	
Urban pop *	80-90s						-0.001
US Foreign aid							(-0.043)

country is, the less likely it is to be a source of terrorism. This result might be explained by the fact that more ethnically fragmented countries tend to have more internal conflict, and consequently may direct fewer resources to harming outsiders. Our results for the political controls help to distinguish among possible mechanisms for how foreign aid impacts terrorism. Among our three proposed mechanisms, our regression analyses do not support the mechanism in which foreign aid increases terrorism via increases in corruption or bureaucratic inefficiency in the recipient country. Likewise, the coefficients for ethno-linguistic fractionalization might be taken as evidence against a mechanism in which a disillusioned minority becomes angry at the central government, and by extension against the United States. However, since ethnic fractionalization may differ from the concept of the isolation of a minority group, we feel that further research in this area is warranted. Our third proposed mechanism involves support for undemocratic governments in the form of foreign aid resulting in resentment against the United States. This mechanism is supported by our results for the Polity2 control variable.

Table 3.14 considers the economic control variables from our main regression.[22] In our baseline specification (regression 1), the level term for 1980-90s GDP per capita indicates that countries which are currently poor are more likely to be sources of terrorism. However, the GDP per capita term for the earlier time period implies that countries that were rich in the 1960-70s are more likely to be sources of terrorism today. When interaction terms are included (regression 2), only the coefficient on the 1980-90s level term remains significant and the interaction terms are not significant. Overall, it seems that poorer countries are more likely to be sources of terrorism.

For non-U.S. foreign aid (regressions 1, 3), only the 1960-70s time period appears to matter. By itself, the coefficient on the 1960-70s level term (regression 1) is positive and significant. However, when the interaction terms are included (regression 3), the coefficient on the 1960-70s non-U.S. foreign aid level term becomes negative, while the coefficient on the interaction with foreign aid is positive and significant. This evidence suggests that countries that received more non-U.S. foreign aid in the 1960-70s are less likely to be sources of terrorism. However, when the United States gave more money to countries that were receiving a lot of non-U.S. foreign aid, anti-U.S. terrorism increased. One interpretation is that more overall assistance will improve the situation in a recipient country, but that U.S. foreign aid may have a unique impact that generates anti-U.S. sentiments.

Life expectancy can be considered a proxy measure for the level of inequality in a country, with higher values associated with lower levels of inequality, controlling for GDP. The coefficient on the

[22] All five regressions also include the political controls in levels, as in the basic specification, although these coefficients are not shown.

levels term for life expectancy is significant and positive in all of the regressions (except the fourth regression where the interaction term for this variable is added) implying that more equal countries have more anti-U.S. terrorism. This result is evidence against a mechanism involving a disenfranchised lower class becoming angry with the United States for supporting the country's elite. However, these results are in accordance with our earlier findings for ethno-linguistic fractionalization, which indicate that more homogenous societies may be larger sources of terrorism. The interaction term is also positive (regression 4), implying that anti-U.S. terrorism increases when the United States gives more aid to relatively equal countries.[23]

The percentage of the population that is urban has a significant and negative coefficient in all five regressions, implying that countries with a smaller fraction of their citizens living in urban areas have higher levels of terrorism. This is surprising, in that one might imagine urban areas to be more supportive environments for terrorist organizations. One possibility is that urban population is proxying for other omitted variables. The interaction term for this variable is insignificant (regression 5).

In summary, results from the economic variables provide evidence that poorer countries, more equal countries, and countries with a smaller urban population are more likely to be sources of terrorism. We also find evidence that countries which receive little aid from countries other than the United States have more anti-U.S. terrorism.

3.5 Robustness Checks

In this section, we detail several robustness checks. Unless otherwise noted, the robustness checks are variations of our basic cross-section Tobit regression (Tables 3.6, 3.7).[24]

We consider several different measures of terrorism, all of which are constructed from the ITERATE database. Our main conclusions of a very small, significant, positive correlation continue to hold when we use "usvictims," "totcasualties," and "incidentcnt" instead of "uscasualties" as the dependent variable. In addition, the finding that military aid has a larger and more robust effect on terrorism than economic aid also holds with these alternative terrorism measures. Furthermore, these conclusions hold when we select terrorism incidents based solely on the nationalities of the

[23]We also consider the effect of an interaction between trade as a percentage of GDP and foreign aid. While this regression converges, we are suspicious of the coefficients and standard errors, and thus do not report these results.

[24]Results in this section that are not included in the following tables are available from the authors upon request.

Table 3.15: **Changes in Foreign Aid**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Also controlling for GDP per capita, total non-US foreign aid, Polity2 (democracy/autocracy), civil liberties, percent Muslim, life expectancy, trade as a percent of GDP, percent urban population, corruption, bureaucratic quality, ethno-linguistic fractionalization, population and population squared, as in our main specification. All regressions use the Tobit specification and measure terrorism using our U.S. casualties variable. Sample includes 66 observations for each regression.

		Regression Number			
		1	2	3	4
CHANGE in the average level of US foreign aid from the 1960s to the 1970s	Economic aid	-2.34		-1.388	
		(1.334)*		(-1.375)	
	Military aid		6.959	7.514	
			(2.917)**	(2.986)**	
	Economic and military aid				1.003
					(-0.861)

victims and terrorists and not on the location of the incidents.

Our results are also robust to using other measures of foreign aid, including the Greenbook measure of total economic aid in grants, the OECD DAC measure of total economic aid in grants, and the OECD DAC measure of total "net" aid (grants plus new loans, minus loan repayments). Results for food aid and total Peace Corps spending are not significant.

Our main regressions examine the effect of the level of U.S. foreign aid on anti-U.S. terrorism. Table 3.15 shows regressions analogous to those in Table 3.7, but uses changes in U.S. foreign aid per capita as the principle exogenous variable. Here an interesting difference arises; an increase in economic aid alone appears to decrease terrorism, while an increase in military aid alone appears to increase terrorism. Only the effect of military aid retains significance when changes in both aid variables are entered together in the regression. The change in total aid (economic plus military aid) does not have a significant coefficient.[25]

Similar results occur when we use the fraction of total foreign aid that is donated by the United States (Table ??). This variable reflects the relative impact of U.S. foreign aid compared to other foreign aid. The coefficients on economic and military aid are again positive, but insignificant when entered separately into the regression. When both military and economic aid are included in

[25]These results are for the impact on terrorism from 1980-2002, of the change in the average level of foreign aid between the 1960s and the 1970s. There is no significant impact of the change in foreign aid levels between the 1970s and 1980s on terrorism

Table 3.16: **U.S. Foreign Aid as a Fraction of Total Foreign Aid**

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Also controlling for GDP per capita, total non-US foreign aid, Polity2 (democracy/autocracy), civil liberties, percent Muslim, life expectancy, trade as a percent of GDP, percent urban population, corruption, bureaucratic quality, ethno-linguistic fractionalization, population and population squared, as in our main specification. All regressions use the Tobit specification and measure terrorism using our U.S. casualties variable. Sample includes 65 observations for each regression.

		Regression Number			
		1	2	3	4
US foreign aid per capita in the 1960s and 1970s as a FRACTION OF TOTAL FOREIGN AID per capita in the 1960s and 70s	Economic aid	0.131		-4.862	
		(-0.241)		(1.969)**	
	Military aid		0.501	12.833	
			(-0.613)	(5.046)**	
	Economic and military aid				0.108
					(-0.173)

the regression, however, both coefficients are significant, with military aid increasing terrorism and economic aid decreasing terrorism. The coefficient on total U.S. foreign aid as a fraction of total foreign aid is positive, but insignificant. For both changes in foreign aid and the fraction of U.S. foreign aid in total aid, the coefficients are many times larger than for our main results, however these effects remain small enough as to be insignificant from a practical standpoint.

We address potential non-linearities in the effect of foreign aid on terrorism by adding a square term for U.S. foreign aid (results not shown). The coefficient remains positive and significant for the total aid variable, but becomes insignificant for the separate measures of economic and military aid. Several of the square terms are significant, suggesting potentially important non-linearities in this relationship.

We also investigate whether our results are driven by countries with extreme levels of corruption or terrorism.[26] The impact of total U.S. foreign aid on terrorism remains positive for countries with high amounts of corruption and for countries with low amounts of corruption. When we restrict our sample to contain only countries with non-zero terrorism (as measured by the sum of "uscasualties" from 1980-2002), all measures of foreign aid (economic, military, and total) have positive and significant coefficients. To verify that none of our results are being driven by countries

[26]For the case of corruption, we run the regression for countries with corruption greater than 2.5 on a six point scale and for countries with corruption less than 3.0 on the same scale. The small amount of overlap in the measure is necessary to obtain a sufficient number of observations. Our results are not driven by the countries in this overlap

with extreme amounts of terrorism, we also drop countries with large values for the "uscasualties" variables and find that our basic conclusions remain.

Additional robustness checks include: adding dummies for region; including a variable for percent Catholic; using the ethno-linguistic fractionalization measure from 1985 instead of 1961; using only terrorism from 1985 to 2002 (to see if there was a special aspect of the early 1980s); and normalizing the terrorism variables by the population of each recipient country. In all cases, the main results discussed in Section 3.4.1 remain.

3.6 Conclusions and Future Research

Our research shows that U.S. foreign aid is associated with a very small, statically significant increase in anti-U.S. terrorism. The effect is robust to the use of several specifications in cross-section (including Tobit, Poisson, OLS and Logit specifications). Our results also remain when we instrument for U.S. foreign aid in the Tobit cross-section specification and when we consider a fixed-effects OLS panel framework. When we use instrumental variables in conjunction with the fixed-effects OLS panel, only the result for military aid remains. We find that military aid consistently has a larger coefficient than economic aid. This finding lends support to the mechanism in which assistance to unpopular governments leads to anti-U.S. sentiments.

Future analysis of country-specific control variables should help to clarify the ways in which U.S. foreign aid leads to an increase in anti-U.S. terrorism. Our analysis thus far points towards mechanisms involving democracy in the recipient country, rather than corruption or inequality. It would also be interesting to examine the relationship between foreign aid and terrorism among other donor nations, such as the United Kingdom or France.

Bibliography

Acemoglu, Daron, Simon Johnson, and James Robinson, "The Colonial Origin of Comparative Development: An Empirical Investigation," *American Economic Review*, 2001, *91*, 1369–401.

Alcala, Francisco and Antonio Ciccone, "Trade and Productivity," *Quarterly Journal of Economics*, 2004, *119*, p613–46.

Alesina, Alberto and David Dollar, "Who Gives Foreign Aid to Whom and Why?," 1998. NBER Working Paper No. 6612.

___ and ___ , "Do Corrupt Governments Receive Less Foreign Aid?," 1999. NBER Working Paper No. 7108.

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

___ and Enrico Spolaore, "On the Number and Size of Nations," *Quarterly Journal of Economics*, 1997, *112*, 1027–56.

___ and ___ , *The Size of Nations*, Cambridge, MA: MIT Press, 2003.

___ and ___ , "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.

___ , Enrico Spolaore, and Romain Wacziarg, "Economic Integration and Political Disintegration," *American Economic Review*, 2000, *90*, 1276–96.

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

Armijo, Leslie, "Mixed Blessing: Foreign Capital Flows and Democracy in Emerging Markets," 1998. Northeastern University, mimeo.

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.

Boone, Peter, "Politics and the Effectiveness of Foreign Aid," 1995. NBER Working Paper No. 5308.

Bourke, P, "Fractal Dimension Calculator User Manual, Online," 1993. http://astronomy.swin.edu.au/~/.

Bruk, S. I. and V.S. Apenchenko, *Atlas Narodov Mira [Atlas Of The Peoples Of The World]*, Moscow: Glavnoe Upravlenie Geodezii i Kartografii, 1964.

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.

Burnside, Craig and David Dollar, "Aid, Policies and Growth," *American Economic Review*, 2000, *XC*, 847–868.

___ **and** ___ , "Aid, Policies and Growth: Reply," *American Economic Review*, 2004, *XCIV*, 781–784.

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

Casella, Alessandra and Barry Eichengreen, "Can Foreign Aid Accelerate Stabilization?," 1994. NBER Working Paper No. 4694.

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.

Chang, Charles C., Eduardo Fernandez-Arias, and Luis Serven, "Measuring Aid Flows: A New Approach," 1999. World Bank Development Research Group Working Paper No. 2050.

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** ——, "Aid, Policy and Peace: Reducing the Risks of Civil Conflict," *Defense and Peace Economics*, 2002, *XIII*, 435–450.

—— **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.

Dalgaard, Carl-Johan, Henrik Hansen, and Finn Tarp, "On the Empirics of Foreign Aid and Growth," *Econ Journal*, 2004, *CXIV*, 191–216.

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*, 1737–1772.

Easterly, William, "Can Foreign Aid Buy Growth?," *Journal of Economic Perspectives*, 2003, *XVII*, 23–48.

——, *The White Man's Burden: Why the West's Efforts to Aid the Rest Did So Much Harm and So Little Good*, New York: The Penguin Press, 2006.

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

——, ——, **and David Roodman**, "New Data, New Doubts: A Comment on Burnside and Dollar's Aid, Policies and Growth' (2000)," 2003. NBER Working Paper No. 9846.

——, ——, **and** ——, "Aid, Policies and Growth: Comment," *American Economic Review*, 2004.

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.

Economides, George, Sarantis Kalyvitis, and Apostolis Philippopoulos, "Do Foreign Aid Transfers Distort Incentives and Hurt Growth? Theory and Evidence from 75 Aid-Recipient Countries," 2004. CESIFO Working Paper No. 1156.

Englebert, Peter, Stacy Tarango, and Matther Carter, "Dismemberment and Suffocation: A Contribution to the Debate on African Boundaries," *Comparative Political Studies*, 2002, *35*, 1093–1118.

Esteban, J. and D. Ray, "On the Measurement of Polarization," *Econometrica*, 1994, *62*, 819–851.

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.

Flaherty, Mark S. and William W. Crumplin, "Compactness and Electoral Boundary Adjustment: An Assessment of Alternative Measures," *The Canadian Geographer*, 1992, *36*, 159–171.

for International Development, United States Agency, "Foreign Aid in the National Interest: Promoting Freedom, Security, and Opportunity," 2002. USAID website, http://www.usaid.gov/fani/.

Gang, Ira and Haider Ali Khan, "Foreign Aid and Fiscal Behavior in a Bounded Rationality Model: Different Policy Regimes," *Empirical Economics*, 1999, *XXIV*, 121–134.

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

——, **Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer**, "Do Institutions Cause Growth?," *Journal of Economic Growth*, 2004, pp. 271–303.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand, "Armed Conflict 19462001: A New Dataset," *Journal of Peace Research*, 2002, *39*, 615637.

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.

Halonen-Akatwijuka, Maija, "Coordination Failure in Foreign Aid," 2004. World Bank Policy Research Working Paper No. 3223.

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

Harms, Philipp and Matthias Lutz, "Aid, Governance, and Private Foreign Investment: Some Puzzling Findings and a Possible Explanation," 2003. HWWA Discussion Paper No. 246, Hamburg Institute of International Economics.

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.

Herbst, Jeffrey, *State and Power in Africa*, Princeton, NJ: Princeton University Press, 2002.

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.

Knack, Stephen, "Does Foreign Aid Promote Democracy?," 2000. Center for Institutional Reform and the Informal Sector Working Paper No. 238, University of Maryland.

——, "Aid Dependence and the Quality of Governance: Cross-Country Empirical Tests," *Southern Economic Journal*, 2001, *LXVIII*, 310–329.

Krueger, Alan B. and Jitka Maleckova, "Education, Poverty, Political Violence and Terrorism: Is There a Causal Connection?," 2002. NBER Working Paper No. 9074.

Lacina, Bethany Ann and Nils Petter Gleditsch, "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths," *European Journal of Population*, 2006, *22*, 145–165.

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.

Mandelbrot, Beniot, "How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension," *Science*, 1967, *156*, 636–638.

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.

Mauro, Paolo, "Corruption and Growth," *Quarterly Journal of Economics*, 1995, *110*, 681–712.

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

Millan, Margaret Mc, *Paris 1919* 2003.

Niemi, Richard G., Bernard Grofman, Carl Carlucci, and Thomas Hofeller, "Measuring Compactness and the Role of a Compactness Standard in a Test for Partisan and Racial Gerrymandering," *Journal of Politics*, 1990, *52*, 1155–1181.

Peitgen, Heinz-Otto, Hartmut Jurgens, and Dietmar Saupe, *Chaos and Fractals; New Frontiers of Science*, New York: Springer-Verlag, 1992.

Porta, Rafael La, Andrei Shleifer, and Robert Vishny, "Investor Protection and Corporate Governance," *Journal of Financial Economics*, 2000, *LVIII*, 1–25.

——, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny, "The Quality of Government," *Journal of Law Economics and Organization*, 1999, pp. 315–388.

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.

Spolaore, Enrico and Romain Wacziarg, "The Diffusion of Development," December 2006. Working Paper.

Tavares, Jose, "Does foreign aid corrupt?," *Economic Letters*, 2003, *LXXIX*, 99–106.

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

VanDerVeen, Roel, *What Went Wrong With Africa: A Contemporary History*, Amsterdam: KIT Publishers, 2004.

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

Winn, Peter, *Americas: The Changing Face of Latin America and the Caribbean*, New York: Pantheon Books, 1992.

Wooldridge, Jeffrey, *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press, 2002.

?