In a State of Fragmentation

Ethnic Heterogeneity and Civil War

Torbjørn Graff Hugo

Master Thesis

Peace & Conflict Studies

Faculty of Social Sciences

UNIVERSITY OF OSLO

10 November 2010

In a State of Fragmentation:

Ethnic Heterogeneity and Civil War

2010

In a State of Fragmentation: Ethnic Heterogeneity and Civil War

Torbjørn Graff Hugo

http://www.duo.uio.no/

Trykk: Oslo Kopisten AS

Abstract

In recent years there has been a debate in the statistical civil war literature over the effects of ethnic heterogeneity on the risk of ethnic conflict. The conventional argument has been that the potential for conflict is higher in countries with many ethnic groups than in countries with few ethnic groups. This has been based on the core assumption that more groups generally raises coordination costs, which in turn is conducive to weak institutions, corruption and eventually higher risk of conflict. Recently, this notion has been challenged by those who contend that the risk of conflict is highest when there are only two large groups in a country. Under such conditions the system is likely to become highly polarized, thus increasing the probability of conflict. According to this argument, more groups would in fact reduce interethnic tension. In parallel, a third line of thought has been advanced that departs from the question of numbers and instead focuses on the relative size of the ethnic groups. This argument has suggested that the risk of conflict should be highest in systems where the largest group has a dominant position, yet without complete control. In such situations the majority will be tempted to abuse its power over the minorities, leading eventually to a revolt by one or more of the excluded groups.

All in all, there has been limited statistical support for either of these arguments, prompting the inevitable question of whether ethnic heterogeneity has any particular effect on the risk of conflict at all. As such, the ambiguity of the results has played into the well-known debate over greed versus grievances, and has effectively weakened the case for those who uphold that ethnicity is a relevant determinant of civil war onsets. In this thesis I argue that the reason why the results have been ambiguous is not that ethnic heterogeneity is irrelevant. Rather, it is because the tests conducted have either lacked proper theoretical foundation; been based on over-aggregated data, confounded the different dimensions of ethnic heterogeneity; or generally struggled with the methodological challenges inherent in research involving the concept of ethnicity. The ambition of this thesis is to remedy some of these shortcomings and provide a better understanding of how the landscape of ethnic heterogeneity affects the risk of conflict

in multiethnic states. More specifically, it is to examine whether some multiethnic configurations are more prone to conflict than others.

To answer this question, I propose two new measures by which the landscape of ethnic heterogeneity can be mapped. The first is concerned with the number of groups while the second focuses on with the potential effect of dominance by a large majority. In testing these measures on a new dataset of ethnic groups compiled by Wimmer et.al (2009) I find considerable support for the conventional argument that countries with a fragmented ethnic landscape – i.e. many groups – have a higher risk of ethnic conflict than countries with few ethnic groups. These results are consistent across different datasets of civil wars and ethnic groups. I find less support for the notion that dominance increases the risk of ethnic conflict in general, but when disaggregating the dependent variable into conflicts over territory and conflicts over government I find that dominance is in fact a significant determinant for territorial conflicts, but not for conflicts over government. The opposite seems to be the case for fragmentation; it is significant for government conflicts, but not for territorial ones. I conclude by making the case that the two core dimensions introduced to map the landscape of ethnic heterogeneity – fragmentation and balance – could serve as the basis for a classification of multiethnic states, which in turn can facilitate more qualitatively oriented studies of the effect of ethnic heterogeneity on the risk of civil war.

Acknowledgements

I would like to extend a profound expression of gratitude to my supervisor Håvard Hegre for his support and guidance, and in particular for helping me through the dire methodological straits of the statistical research on civil war. I am also immensely grateful to all the people at PRIO and the Center for the Study of Civil War (CSCW) for making this process both academically and socially inspiring. I felt warmly welcomed from day one, and I hope I'll be able to remain part of this family also in the time to come.

I would also like to thank all my fellow students at the Peace and Conflict Studies master program. Studies are only as exciting as the people around you make it, and you have been a wonderful crowd. As for the work with this thesis, I would especially like to mention Katharina Last, whose inexplicable interest and invaluable insight into the conceptual quagmire of my project made for the perfect sparring partner. Thank you for sharing of your time to argue with me. On the motivational side, I should also include Magnus Løvold, who taught me never to accept the implicit premises of my own arguments – never to accept the truth as truth. One day I am sure I will understand what you meant.

Torbjørn Graff Hugo
Oslo, November 2010

Table of Contents

1 Introduction

> The importance of ethnic conflict, as a force shaping human affairs, as a phenomenon to be understood, as a threat to be controlled, can no longer be denied.
>
> Horowitz (1985:XV)

In May 2009, the world sat back and watched as the Sri Lankan army launched is final devastating attack on the last remaining forces of the Liberation Tigers of Tamil Eelam – the Tamil Tigers. The ferocious campaign egged on by President Mahinda Rajapakse seemingly ended the civil war that had ravaged the country since its violent inception in august 1983 (BBC 2010a; Rotberg 1999:72). In the eyes of many observers, however, the defeat of the Tamil Tigers ended nothing but the actual fighting. A peaceful solution, under which the Sinhala majority and the Tamil minority on the island could live in peaceful coexistence, seems as distant as ever. No viable answers have been found on questions such as regional autonomy, minority rights and perhaps most importantly; language (Horowitz 2000; Polgreen 2009; Reddy 2007).

Many skeptics believe, moreover, that such solutions will not be found, and that Sri Lanka is destined to remain in a state of ethnic quarreling, tension and dispute, and a prime reason for this is attributed to ethnicity and language. Spoken by something close to 20 percent of the population Tamil is an official language, yet few Sinhalese speak it. The Tamils insist that English should be the common language, but the Sinhalese do not see why a country with an overwhelming Sinhala majority should not have Sinhalese as the main administrative language.

The above scenario constitutes a fascinating challenge for a peace mediator, yet for the peace *researcher* the case of Sri Lanka points to a more profound question. Might it be that Sri Lanka represents a typical case; that countries with the same ethnic setup would exhibit similar patterns of tension and political dynamics? And does this constellation of a clear majority facing a relatively large minority represent a particularly conflict-prone type of multiethnic state?

One could plausibly imagine that if Sri Lanka had more than two main languages the question could actually have been easier to solve. In South Africa, with its 11 official languages, it is evident for all parties that all of these cannot be learned by everyone. And a common language of administration is, after all, something of a precondition for a functioning modern state. In South Africa the *lingua franca* has become English, despite the limited number of South Africans who claim English as their mother tongue.[1] A similar situation reigns in India, and even more so in the ethnically atomized country of Papua New Guinea, with its round about 800 ethnic groups (see e.g. Reilly 2000). But does this imply that multiethnic societies with large numbers of groups are always more stable than countries with only two main ethnic groups? One might be eluded to think so, and as I will return to shortly some statistical research has also pointed in this direction.

The case of Sri Lanka indicates another dimension that has been pondered in the academic debate on ethnic heterogeneity and civil war. Sri Lanka is not merely an example of a case where a limited number of groups become an "ethnic witches' brew", to quote Tanja Ellingsen (2000), but also a case in which a dominant majority uses – and some would say abuses – its power over the minority, setting in motion feelings of resentment and frustration among the minority. In short, it is a case of ethnic dominance. Such constellations, it has been argued (see e.g. Horowitz 1985, Collier & Hoeffler 2004), tend to push the minority to the point where it decides to launch some form of retaliatory action; in the name of justice and the right to self-determination.

The examples of Rwanda and Burundi – where large Hutu majorities share their territories with strong and coherent Tutsi minorities – seductively serve to reinforce the impression that two groups in a highly unbalanced ethnic configuration will eventually come to explode (see e.g. Uvin 1999). More recently, the ethnic fallout between the Kyrgyz majority and the Uzbek minority following the ousting of the Kyrgyz president Bakiyev in April 2010 further strengthens the suspicion in the minds

[1] English is the preferred language of business, politics and media, and is generally understood around the country. See http://www.southafrica.info/about/people/language.htm

of the lay observer – the suspicion that a constellation of ethnic groups where a large majority faces a large minority might possibly be the most conflict-prone system of all (BBC 2010b). Ethnic conflict in such countries may almost seem unavoidable.

Fortunately – or unfortunately – the picture is not that simple. First, there are more than enough examples of stable and well functioning states with ethnic configurations similar to Sri Lanka to rebuke such notions of inevitability. Canada works quite well, despite having only two main ethnic communities, one of which is considerably larger than the other. So too does Jordan, Togo, Cuba, Kuwait and Estonia, all with similar majority-minority configurations. Secondly, ethnic conflicts are by no means limited to countries with only two main groups where a dominant majority faces a large minority. When the post-election violence erupted in Kenya in late 2007, analysts were quick to point out that Kenyan politics have been marred by ethnic and tribal infighting precisely because there were so *many* groups (see e.g. Hanson 2008). Kenya, with its 40-odd ethnic groups, thus seems to point in the opposite direction of Sri Lanka; that ethnic fragmentation increases the risk of conflict.[2]

These examples draw the attention to what has been a primary line of disagreement in statistical research on ethnic heterogeneity and conflict; whether more groups increase or reduce the risk of conflict. It was long taken almost for granted that ethnically fragmented societies – understood here and henceforth as countries with a relatively large number of ethnic groups – would be particularly prone to conflict due to increased coordination costs, cooperation problems and a general notion of *the more the messier* (see e.g. Easterly & Levine 1997; Ellingsen 2000; Sambanis 2001; Collier & Hoeffler 2004; Buhaug 2006). Recently, this notion has been contested and it has been argued that the relationship between ethnic fragmentation and conflict is in fact non-monotonic; that the most conflict-prone systems are those where only two major groups exist, as in Rwanda or Eritrea (see e.g. Esteban & Ray 1994, 1999, 2008; Reynal-Querol 2002; Montalvo & Reynal-Querol 2005). Under such conditions the

[2] This notion of an increase in levels of fragmentation is a theoretical concept. It does not suggest, nor suppose, that the number of groups in a country actually changes over night. Rather, it depicts the analytical exercise of studying countries with different numbers of ethnic groups. An *increase* in fragmentation thus implies that you go from studying a country with few groups to studying a country with many groups.

political landscape would become polarized and gradually drive the two groups apart. A third argument in the debate over ethnic heterogeneity has been linked to the notion of dominance and to the assumption that large majorities tend to exclude ethnic minorities from power, leading to frustration and grievances that eventually cause minorities to rebel (see e.g. Collier & Hoeffler 2004). As such, the statistical research on ethnic heterogeneity and conflict has been a debate over numbers, and more specifically, over how the number and relative size of the ethnic groups in a given country affect the potential for armed conflict.

There have been few conclusive results in the tests of these hypotheses. Some findings suggest that more groups does indeed increase the risk of conflict (e.g. Sambanis 2001), while others find that fragmentation is relevant only for conflicts over territory, but not for conflicts over government (e.g. Buhaug 2006). Others again argue that the number of groups affects the intensity of conflict, so that conflict can be expected to occur more often in highly fragmented societies, but these conflicts will be less devastating than if conflict erupts in countries with few groups (Esteban & Ray 2008). A fourth line of argument suggests that the most important factor determining ethnic conflict is the degree of minority exclusion (e.g. Cederman & Girardin 2007).

The only certain conclusion to be drawn from all these contributions is that the debate over ethnic heterogeneity and conflict is itself in a state of fragmentation, for which a lack of clear and unambiguous findings must take part of the blame. There are several possible reasons why the results have not been conclusive. The most obvious of these would be that there is no particular pattern to find – that ethnic heterogeneity has no measurable effect on the risk of conflict. Before coming to terms with such a conclusion, however, there are several lines of critique that should be dealt with, and methodological shortcomings that should be scrutinized. These range from the manner in which ethnicity is coded, via the theoretical foundation upon which the hypotheses are predicated, to the dependent variables with which the tests have been conducted. Herein lays ample scope for improvement, which is also part of the motivation for the undertaking of this project.

The fact is that even though the debate is fragmented, it is by no means irrelevant. There are indeed several reasons why a better understanding of the relationship between ethnic heterogeneity and conflict would be useful. First, it would provide – in combination with other determinants of ethnic conflict – some clue as to whether a given country has a particularly volatile ethnic composition. Special attention could then be given to these countries in terms of monitoring developments that might trigger the mechanisms expected to drive the country towards conflict. This does not imply that the knowledge of some ethnic constellations being more volatile than others would automatically translate into practical measures for peace-building. On the contrary; in and of itself, it is of little use for Kenya or Sri Lanka to know that the ethnic configuration in their particular country would be among the most conflict-prone types of multiethnic systems.

For the research on ethnic heterogeneity to have practical implications on policy it will need go beyond the mere mapping of tendencies, and provide insight into the political dynamics and specific mechanics that cause civil wars to erupt. Only then can policies aimed at mitigating the risk of conflict be devised and implemented with any credible chance of success. Such insight however, can only be obtained through more qualitatively oriented studies that compare systems in order to locate and fully understand the mechanism at play. As Donald Horowitz states it in his seminal work on the topic, *Ethnic Groups in Conflict* (1985:13); "the need for comparative analysis is compelling". And this points to the second reason why statistical mapping of conflict risk is an important exercise.

For comparative studies to yield meaningful results it is a plain precondition that they be guided by some basic parameters of classification that enables the researcher to distinguish between most similar and least similar systems. The units under study must be placed into meaningful categories that indicate what it is that is to be compared, "for the simple reason that comparison requires comparability" (Horowitz 1985:16). And this is where the statistical research on ethnic heterogeneity can serve a purpose, in providing indications as to what these parameters ought to be and to where the thresholds for classification should be placed.

The ambition of this thesis is thus twofold: First, to revisit the questions of how the number of groups and their relative size affects the risk of ethnic conflict, with the composite aims of strengthening the theoretical foundations of the hypotheses, of developing measures that reliably measure the notions laid down in these hypotheses, and of improving the viability of the tests by using new sources of data. The second ambition of the thesis is to use the results of these tests to draw the contours of a classification by which multiethnic systems can be categorized and compared. In sum, this could be contracted into the following core research question:

Are some multiethnic configurations more prone to conflict than others?

1.1 Structure

The thesis is structured into four main parts, in addition to this introduction and a set of concluding remarks. The first of these parts, Chapter 2, introduces the academic debate on ethnic heterogeneity and provides a theoretical foundation upon which a set of hypotheses is developed. Chapter 3 deals with the operationalization of these hypotheses, while Chapter 4 discusses methodological issues related to the statistical study of ethnicity and ethnic conflict. In Chapter 5, the results of the tests are laid out and discussed. Each chapter starts with a short introduction that lays out the main elements and arguments of the chapter.

It should be noted that throughout the paper the concepts of civil war and armed conflict generally refer to *ethnically* motivated civil wars, in which a minimum of 25 people have been killed during the course of one year and the fighting fractions hail from different ethnic groups. More elaborate definitions of ethnic civil wars are provided in Chapter 4.

2 Background & Theoretical Approach

In this chapter the previous literature on the subject of ethnic heterogeneity and conflict is introduced and discussed, with a particular focus on the theoretical underpinnings of the research. Three different arguments are identified: (i) that the risk of ethnic conflict increases with the more groups; (ii) that the risk of ethnic conflict is highest with only two large groups, and; (iii) that the risk of conflict is highest when one group dominates, but not overwhelmingly so. These arguments are then mapped along two main dimensions (fragmentation and balance) that serve as the basis for a framework of classification. The second part of the chapter aims to strengthen the theoretical foundation for this framework by bringing in concepts and ideas both from the field of international relations and from political science. The chapter ends with a summary of the main arguments and the formulation of four concrete hypotheses derived from the theoretical discussions.

Over the last two centuries, the share of wars fought in the name of national liberation or ethnic autonomy has risen steadily; from a mere 20 percent in the 19th century to an average of about 45 percent for the period between 1919 and 1989. Since the end of the Cold War, the share of such wars has reached a staggering 75 percent (Wimmer et.al 2006, 2009). According to the PRIO/Uppsala Armed Conflict Dataset (Gleditsch et.al. 2002) more than 90 percent of the wars since 1990 have also been either internal or so-called internationalized internal conflicts. These developments have prompted a bourgeoning scholarly debate on the causes and consequences of internal ethnic conflict – or ethnic civil war.

2.1 Fractionalization – the more the messier

In the large N literature on civil war, the debate on ethnicity has revolved largely around the concept of ethnic heterogeneity, with the aim of identifying patterns that could indicate whether some varieties of multiethnic states are more prone to conflict than others (see Collier & Hoeffler 1998, 2004; Ellingsen 2000; Doyle & Sambanis

2000; Fearon & Laitin 2003; Esteban & Ray 1999, 2008; Montalvo & Reynal Querol 2005; Cederman & Girardin 2007). Early attempts to test the socio-economic impact of ethnic heterogeneity statistically started from the assumption that an increase in the number of ethnic groups in a given country would, *ceteris paribus*, increase the costs of coordination. Based on a rationality drawn from the field of economics, this was in turn assumed to cause higher levels of corruption, reduced investments and less diffusion of technology, all of which could be associated with a higher risk of ethnic conflict (see Mauro 1995; Easterly & Levine 1997; La Porta et.al. 1999; Alesina et.al 1999, 2003).

The measure employed to test this relationship was an index of ethno-linguistic fractionalization (ELF) that calculated the chance that two randomly selected persons from a given country belong to the same ethnic group. The index was based on data from the *Atlas Narodov Mira (ANM)* – a global overview of ethnic groups compiled by a Soviet research institute in the early '60s (USSR 1964).

Despite its widespread use, the ELF index has not yielded convincing results in terms of explaining the outbreak of civil wars (Fearon & Laitin 2003; Montalvo & Reynal-Querol 2005; Hegre & Sambanis 2006). To some extent, this could be attributed to the fact that the index was developed as a tool for measuring inequality and economic fractionalization, which, it has been argued, has a different and more indirect effect on conflict than do ethnic group constellations. Ethnic groups have a dynamic of their own, which is hard to capture by the methodological individualism associated with economic theory (see e.g. Posner 2004, Cramer 2002). In the words of Cederman & Girardin (2007:173), the ELF-index "has yet to be supported by a convincing set of causal mechanisms that links it to political violence".

2.2 Polarization – less is more

More recently and partly driven by the relatively weak results obtained by the ELF-index, an alternative hypothesis has been advanced. Several authors have suggested that the relationship between the number of ethnic groups and the risk of conflict is 'non-monotonic' (Esteban & Ray 1994, 1999; Reynal-Querol 2000; Montalvo &

Reynal-Querol 2005). More specifically, it is assumed that the risk of conflict is at its highest when two groups are juxtaposed in a bifurcated or so-called bipolar system, and that the risk of conflict falls exponentially with an increase in the number of groups.

In advancing this argument Esteban & Ray (1994) draw largely on Marxist theory. They refer to Karl Deutsch (1971) who explains that as the struggle between the bourgeoisie and the proletariat advances, the relationship becomes more and more antagonistic, a process he labels *polarization*. On this basis, Esteban & Ray (1994) proceed to develop a measure of polarization which is maximized on a scale from 0 to 1 when the two largest groups make up 50% each (½, 0, 0,… ½). From this bipolar peak, the values on the index approach zero either if you move towards homogenization or towards fractionalization. This non-monotonic relationship between the number of groups and the risk of conflict constitutes the most important difference between the fractionalization and the polarization indexes. Montalvo & Reynal-Querol (2005) elaborate and test the differences between the two measures and with a simplified version of the polarization index they show that it is significantly better at explaining conflict than is fractionalization.

Despite these findings, the polarization index does have some precarious shortcomings, the first and most important of which is the rather arbitrary link between the theoretical hypothesis and the empirical evidence. Using Marxist theory to explain ethnic conflict sounds, to some extent, like a contradiction in terms. At the core of Marxist theory lies the idea that economic boundaries can be transcended; that the poor can become richer and the rich become poorer. This view starkly contrasts with the rather static properties associated with ethnicity and ethnic groups. One does not simply change ethnic group, even if, as most sociologist and anthropologists would contend, ethnicity is a fluid concept (Jenkins 1997, Lijphart 1977, Eriksen 1993). I return to this in Chapter 4.

Montalvo and Reynal-Querol (2005) seem to have grasped this dilemma, and in their attempt to link the polarization hypothesis to ethnic conflict theory they turn to Donald

Horowitz (1985) instead of Karl Deutsch (1971). In his seminal work *Ethnic Groups in Conflict* (1985) Horowitz provides a plethora of empirically induced hypotheses. The trouble is that he barely mentions the question of numerical balance between groups, and when he does, it is to illustrate the limitations of democratic government in ethnically dominated societies (Horowitz 1985:629). His contention is that in a system where group A constitute 60 percent of the population and group B make up the remaining 40 percent, there is no prospect for group B of ever gaining electoral majority and government control, regardless of the electoral rules employed. Rather than theoretically underpinning the polarization argument, as was seemingly the aim of Montalvo & Reynal-Querol, this argument points to a third hypothesis advanced in the literature, namely that of ethnic *dominance.*

2.3 Dominance – this land is my land

Collier & Hoeffler (2004) elaborate on this argument of dominance and hypothesize that the propensity of an ethnic group to seek domination should increase in proportion to its share of the population. A group constituting 90 percent of the total will hence be more inclined to dominate, and in effect discriminate, than a group of only 45 percent. That said, they also suggest that "the incentive to exploit the minority diminishes the smaller is the minority, since there is less *potential for exploitation*" (Collier & Hoeffler 2004:13, italics added). On this basis they conclude that the risk of ethnic conflict as a result of exclusion should be highest in countries where the largest ethnic group constitutes "a majority, but not an overwhelming majority" (Ibid.). This notion coincides with the so-called hegemonic stability theory, which assumes that once an actor has gained hegemonic status (i.e. an overwhelming majority), the system will be pacified since no group or coalition of groups will have the resources to challenge its position. The hegemonic power will then turn its attention to providing collective goods, such as security at sea, which in turn reinforces stability. Political scientists often point to the classic cases of the *Pax Romana* and the *Pax Britannica* as examples of this. William Wohlforth (2006) argues that the United States today functions as a hegemonic pacifier, and he believes that this *Pax Americana* is likely to endure for the foreseeable future.

Collier & Hoeffler (2004) go on to test the ethnic dominance hypothesis by employing a binary variable that codes countries as ethnically dominant if the largest group make up between 45 and 90 percent of the population. While they do find some statistical support for the hypothesis, this binary variable is by no means a perfect measure of ethnic dominance. For one, there is a logical flaw in the assumption that a group constituting 45 percent of the population should qualify as dominant. Not because it would not dominate – in most cases it probably would – but because whether or not it does necessarily depends on how the remaining 55 percent of the population is constituted. If only one other group exists (i.e. of 55 percent) it makes little sense to speak of an ethnically dominant society. In fact, such a setup would more closely resemble the bipolar system assumed to be the most volatile under the polarization argument. This notion is seemingly confirmed by Esteban & Ray's (2008:166) rather confusing interpretation of Collier & Hoeffler's dominance variable as "an alternative and considerably cruder specification of polarization". Theoretically, it is necessary to distinguish between these two concepts – *polarization* and *dominance* – since they purport to measure different underlying mechanisms.

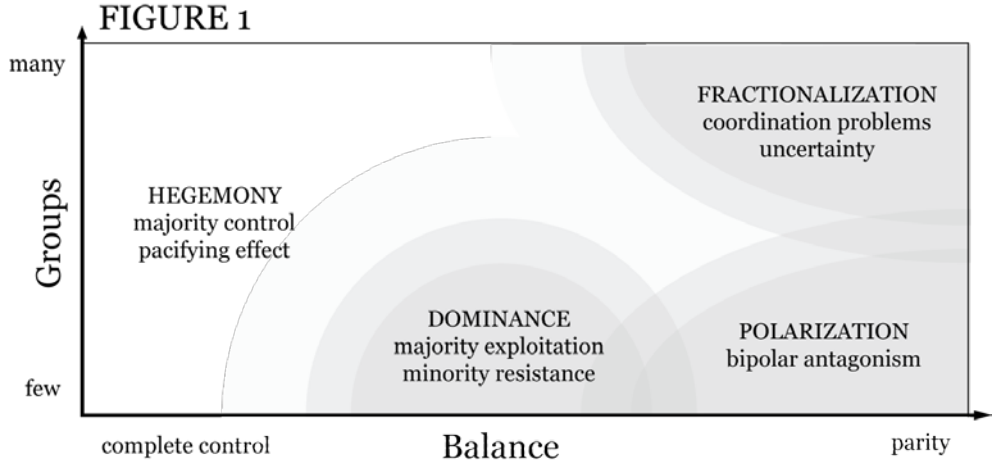
A better way to measure the concept of ethnic dominance would be to create a variable similar and comparable to the fractionalization and polarization indexes that could maximize the value of dominance somewhere between 45 and 90 percent. Such a measure could mitigate the above problems by also taking into consideration the size of the opposition, and thus linking it closer to the theoretical concepts of dominance. Horowitz uses the example of a 60-40 distribution to illustrate the problems of dominance and exclusion, but he does not purport to establish this particular constellation of groups as the most volatile ethnic configuration. He makes no such contention. The 'risk peak' might equally well be situated around 75-25, at which point even constitutional changes can be pushed through democratically by the majority, while the minority remains big enough to retain reasonable prospects of forced resistance. The uncertainty with regards to the point of maximum conflict risk suggests that an ethnic dominance index should allow for the adjustment of the peak point, in order to locate the best fit empirically.

2.4 Three arguments – two dimensions

If we now attempt to summarize the three arguments introduced above, we can identify two main dimensions along which these align themselves; namely the *number of groups* and the *balance* between them. The fractionalization measure assumes that the risk of conflict increases alongside the number of groups in the system, while polarization assumes that the likelihood of conflict is maximized when two equally sized groups compose 50 percent each. Both measures implicitly assume that the risk of conflict is highest when all the groups are relatively equal in size; that is when the system is *balanced*. The third argument, regarding ethnic dominance, breaks with this view and suggests that the risk of conflict will be higher in systems where one group dominates over the others, yet it concurs with the polarization argument in assuming that tension is maximized with only two groups, since the potential for exploitation increases in proportion to the size of the largest minority. I will return to this in the subsequent section. Figure 1 captures these two dimensions and illustrates how the measures and hypotheses are placed accordingly.

The model provides an overview of the hypotheses and measures previously employed in the quest for patterns of ethnic heterogeneity and conflict. Yet, as the above discussion indicates, there are methodological challenges associated with each of these measures. Both *fractionalization* and *polarization* are neatly constructed indexes that reliably capture their intended concepts. The problem is that these concepts lack proper theoretical foundation, and as such one is effectively faced with a validity problem. Validity can generally be understood as the extent to which a measure tests

FIGURE 1



what it is intended to test. Hence, if the theoretical foundation upon which the measure is constructed is uncertain, it becomes difficult to make "accurate inferences based on the test scores" (Weathington et.al 2010:87).

For Collier & Hoeffler's (2004) *dominance* measure, reliability appears to be a bigger concern than validity. The theoretical foundation is sound, but the measure itself is crude. Weathington et.al (2010:84) describes reliability as the "consistency with which observations or measurements are made", and since, as explained above, Collier & Hoeffler's binary variable risks coding fully balanced systems as dominant, there is likely to be considerable measurement error in the results. The methodological challenge is thus threefold; to provide a theoretical basis for the two first measures, and to improve the reliability of the third.

2.5 Towards a framework for comparison

In order to do this, we need a better framework for comparison – one that enables a more comprehensive approach to the theoretical constructs of the debate. Despite extensive research into the causes and nature of ethnic conflicts over the last decades, the theoretical framework for comparing different types of multiethnic systems remains weak. In the words of Esteban & Schneider; "much more theoretical work needs to be developed, seeking convincing explanations for the usage of armed force across different political and social *systems*" (2008:135, italics added). Horowitz (1985:13) similarly contends that it is not a lack of data that inhibits further research on ethnic conflicts, but a lack of what Giovanni Sartori (1970) calls 'data containers' – meaning categories within which to classify the data. Sartori (1970:1039), in turn, asserts that "regardless of whether we rely on quantitative data or on more qualitative information, […] the problem is the same, namely, to construct fact-finding categories that own sufficient discriminating power". In short, there is need for a proper classification of multiethnic states that allows for the exercise of comparison.

In his seminal work on the topic of ethnic conflict, Horowitz (1985) lays out some important distinctions between varieties of multiethnic states, notably the divide between *ranked* and *unranked* systems and between *advanced* and *backward* groups

and regions. He suggests that each of these types will have their own dynamics and will thus require different remedies to mitigate potential conflict. These distinctions undoubtedly serve a purpose, yet Horowitz largely bypasses the task of categorizing multiethnic systems according to numerical criteria – that is the *number* of groups and the *balance* between them. If the ultimate aim of the research on conflict in multiethnic societies is to be able to provide insight into how the inter-ethnic dynamics of say Afghanistan differs from that of Kyrgyzstan, this knowledge gap needs filling. Granted, such a perspective axiomatically presupposes that the numerical distribution of groups in a system actually matters, a notion that in and of itself merits scrutiny. But for the purpose of classification, numbers might be the best criteria available. That said, the exercise of counting ethnic groups should by no means be portrayed as simple, a topic I will return to later. In general, however, the epistemological approach adopted in this paper aligns with the view expressed by Sartori (1990:318) in his study of political party systems, in which he states that "it is not whether the number of parties matters – it does – but whether a numerical criterion of classification enables us to get hold of what matters".

The overview in Figure 1 provides a useful starting point for developing a more comprehensive theoretical framework, with its two basic dimensions; the number of groups and the balance between them. Yet, for the theoretical mapping of the various hypotheses, we might need to look beyond the relatively contracted scope of ethnic conflict theory. Esteban & Schneider (2008) recognize this as they turn to the fields of economics, political science and international relations (IR) in order to underpin the various assumptions of the ethnic heterogeneity debate. In the field of IR the question of 'polarity' has been argued over for decades, or even centuries (Morgenthau 2006), and the most prominent arguments closely resemble those of the ethnic heterogeneity debate.

2.6 The domestic analogy – extending the theoretical toolbox

There is of course a question of whether, and to what extent, theories aimed at explaining behavior in the international system can be projected onto the domain of domestic politics. Esteban & Schneider (2008:135) suggest that from a "universalist perspective, the mechanisms that drive a system of states into war should be identical to those that increase the risk of violence among competing groups". Others would argue that there is a fundamental difference in the degree to which domestic politics reflect the anarchic foundations of the international system (see Bull 1977, Suganami 1989).

This touches upon the long-standing debate in IR over the utility of the 'domestic analogy' - that is the "presumptive reasoning which holds that there are certain similarities between domestic and international phenomena" (Suganami 1989:1). Scholars like Hedley Bull have been critical of this notion. He perceives the world of states is a *sui generis* – a special case – and consequently argues that the individualistic approaches to domestic politics have only very limited power to explain the dynamics of the international system (Ibid; Bull 1977). On the other hand, the study of multiethnic states arguably escapes much of Bull's criticism, since the main units of analysis are the ethnic groups, not the individual citizens of each polity. As such, one could contend that the politics of multiethnic states is in fact more reminiscent of the interaction between states in the international system than of that in an ethnically homogenous nation-state. It is precisely for this reason John Stuart Mill famously concludes that "free institutions are next to impossible in a country made up of different nationalities" (Mill 1926:120), since the dynamics of such a country would make the creation of free institutions as challenging as establishing democracy on an international level.

Such a perspective on multiethnic states – as a hybrid between the non-ethnic domestic level and the international system level – suggests that the theoretical postulates for an analytical framework of multiethnic states can be drawn both from political science and from international relations.

2.7 IR-rationality – strife and stability in a world of states

One of the most basic concepts in the study of international relations – and particularly within the realist school of thought – is that of *balance-of-power* (see e.g. Schelling 1980; Morgenthau 2006). At the heart of this notion lies the recognition of power as a perpetual force for both good and bad, and of deterrence as a necessary strategy for limiting the excessive use and abuse of power. Since the pioneering political juggling and *Raison d'état* of Cardinal Richelieu in the 17th century, via the *Realpolitik* of the Metternich system in the first half of the 19th century, European powers have become accustomed to this manner of thinking (Kissinger 1994). The stability of the system was predicated upon a tacit, and sometimes even explicit, agreement that no state could be allowed to dominate, and that alliances would have to be forged with this fundamental aim in mind.

As such, the balance-of-power theory neatly fits into the first dimension introduced in the framework above, as it assumes that *balanced* systems will be more stable than *unbalanced* systems. The question of whether more actors increase or reduce stability received little attention in the early days, partly because the concepts and strategies were derived from, and adapted to, the political landscape of Europe, in which the number of actors remained relatively stable once the major states of the continent were consolidated. But with the advent of the cold war, and the apparent bipolarization of world politics that ensued, scholars started pondering whether a power-balance with only two major powers might actually be more stable than the multipolar herd of war-mongering states that all played to their own tune during the Concert of Europe.

2.8 Structuring the realist argument

Kenneth Waltz (1964), a founding figure of the IR-school of so-called structural realism, argues that *bipolarity* represents the ultimate form of international stability, since it minimizes uncertainty and allows decision-makers to better assess the preferences of a potential foe. He starts from the assumption that the primary concern of each unit in the system is survival, and that the actions of states can be understood largely in such terms. Somewhat counter-intuitively Waltz induces from this

assumption that the international system, with its generally anarchic features, will be competitive in nature, even if states do not have any specific conflicts of interest (Glaser 2010:20). The explanation for this mechanism lies in the classic perception of the security dilemma, in which the uncertain intentions of other actors makes it rational to act preemptively.

Upon these basic assumptions, John Mearsheimer draws the contours of a typology for the international system. In *The Tragedy of Great Power Politics* (2001) he describes three main types of systems, all linked to the pivotal concept of balance-of-power; (i) bipolarity, (ii) unbalanced multipolarity, and (iii) balanced multipolarity. A *bipolar* international system is understood as a system in which two large actors are juxtaposed and where neither has the power to exercise control over the other. Smaller actors in the system then align themselves with either of the two major powers. The second concept, *unbalanced multipolarity*, implies that one actor is dominant in size and power compared to either of the other groups. But the dominant actor does not necessarily wield power over the whole system, and the smaller actors can potentially align to counter the strength of the dominant power. It can thus be defined as a system of multiple poles in which one state has the potential to become a hegemonic power. In a system of *balanced multipolarity* there is no clear dominant actor with the ambition or power to control the system.

Mearsheimer argues that between these, *bipolarity* should be the most stable, since the relationship between only two main actors reduces uncertainty, fear and the risk of miscalculation. He believes that all states, and in particular 'great powers', will seek hegemony whenever they can, since that is the most effective manner of ensuring survival (Dunne & Schmidt 2005:169-170). This leads him to conclude that *unbalanced multipolarity* is the most unstable type of international system, since the smaller powers will be chronically skeptical of the 'inevitable' hegemonic aspirations of the largest power, and because they will aim, as far as they can, to contain these aspirations. Finally, *balanced multipolarity* falls somewhere in between the two former. It is seen as less stable than bipolarity, but more stable than unbalanced multipolarity, since the fear of one group achieving status as hegemon is limited

(Mearsheimer 2001). In sum, Mearsheimer holds that *balanced* systems are likely to be more stable than *unbalanced* systems, and that *bipolar* systems will be more stable than *multipolar* systems. He thus provides a theoretical basis that underpins both the fractionalization argument and the ethnic dominance argument, deeming the latter as the least stable.

2.9 Unless more is actually merrier

That said there is little agreement among IR scholars regarding these conclusions (Geller & Singer 1998:113). In contrast to the structural realists, Deutsch & Singer (1964:390) make the case for multipolar stability, suggesting that "as the system moves away from bipolarity toward multipolarity, the frequency and intensity of war should be expected to diminish". They base their argument on the notion that stability is contingent on the number of possible relationships in the system – what they refer to as 'interaction opportunities'. When every bilateral relationship is counted, a system of three groups would have three interaction opportunities (A-B, A-C, B-C). With four groups, the number increases to six (A-B, A-C, A-D, B-C, B-D, C-D). More groups thus increase the potential for interaction exponentially. The assertion is that the increased number of possible coalitions dampens the potential for conflict. An underlying assumption for Deutsch & Singer appears to be that multipolar systems are divided by crosscutting cleavages that make some actors natural interaction partners on certain issues, but not on others. This thinking closely mirrors Arendt Lijphart's (1977) view of crosscutting cleavages as stabilizing factors in multiethnic societies. Alliances formed under such conditions will be less stringent, more flexible and more open to compromise than if the system had only two groups – in particular if these two groups are separated by multiple overlapping segmental cleavages. And with flexibility comes stability – presumably. This branch of IR theory thus provides a theoretical basis for the polarization argument, as it assumes that the risk of conflict will diminish with an increase in the number of groups.

2.10 Polarized confusion

A potential source of confusion at this point is the manner in which the IR debate employs the term 'polarity'. As mentioned, Deutsch (1971) perceived of polarization as a process of gradually increasing antagonism between two or more parties. The logical implication of this view is that a system cannot *a priori* be defined as polarized. Horowitz (1985) argues along the same lines when he suggests that there is an inherent tendency in most systems towards bipolarity, regardless of the actual number of groups. He ponders why it is that "despite the plurality of groups in an environment (rarely there are only two), polarity frequently emerges", and he suggests that an 'economy of antipathy' causes pairs of antagonists to emerge as "comparative reference groups" (Horowitz 1985:182, original parentheses). Polarization thus understood is a process, not a static property by which different systems can be categorized. It follows from this that a system of multiple groups could easily become bipolar, as long as the groups align on either side of one main cleavage.

Therefore, if the ambition of a classification of multiethnic societies is to say something about each system type's relative potential for armed conflict, it is important not to confuse the static criteria for classification with the dynamic effects and presumed mechanisms that the units of the different categories are thought to exhibit. The criteria for classification should be limited to factors that remain relatively constant over time, such as the number of groups and their relative size.

2.11 Into the realm of political science

Adherence to this principle has arguably been stronger in the domain of political science than in that of international relations. The previously mentioned Giovanni Sartori is notably famous for his work on a typology of political party systems. One of his primary assertions is that the *a priori* classification based on numbers should be distinguished from the dynamics typically associated with the different classes. He refers to this as the difference between *classes* of systems and *types* of systems – i.e. the difference between a classification and a typology (Sartori 1970, 1990). The former simply amounts to a counting exercise, albeit guided by certain counting rules,

while the latter focuses on the system itself and the behavioral incentives it instills on the different actors.

Sartori (1990) lists a total of seven classes within which democratic party systems can be categorized (see List 1). These range from one-party systems, where there is only one party in existence (i.e. homogenous), via two-party systems with the US as the most famous case, to limited and extreme pluralism. Between the one-party system and the two-party system, Sartori also identifies hegemonic and predominant systems, of which only the latter is deemed competitive. In hegemonic party systems there is no prospect of a second party or coalition gaining more than 50 percent of the seats in parliament. Smaller parties are accepted only as long as they do not threaten the position of the hegemon. In predominant systems the majority party may actually lose power, though this rarely happens. Sartori further includes a class he labels atomized party systems, in which the number of parties exceeds a threshold beyond which the addition of one more party does not affect the overall dynamic of the system.

List 1

Sartori's (1990) classification of party systems

1. One party
2. Hegemonic party
3. Predominant party
4. Two-party
5. Limited pluralism
6. Extreme pluralism
7. Atomized

Sartori contends that some systems will be more conducive to cooperation, while others are relatively more prone to competition. That said, he underlines that it is not the amount of competition that matters, but the direction of it (Sartori 1990). For a system to remain stable the direction of the political competition ought to be centripetal; that is, the political parties should compete over 'floating voters' in the centre of the political spectrum. A two-party system typically exerts such tendencies, with the Republicans and the Democrats in the US as a prime example. In effect, this makes the political landscape look bell-shaped with a single peak in the middle, where the majority of the voters are situated. For limited pluralism, which Sartori defines as having between 3 and 5 relevant parties, the same logic applies, but he notes that as the number of parties increases, so does the risk that any given party dissociates itself from the established norms of the system and becomes what he calls an 'anti-system

party'. Such parties, he warns, will generate a second peak in the party landscape, and consequently reverse the direction of the competition. The effect will be centrifugal competition, or, as Karl Deutsch (1971) calls it; *polarization*.

2.12 Strengths and limitations of Sartori's typology

The benefit of Sartori's typology, in contrast to Mearsheimer's categories, is that it distinguishes between the numerical criteria for classification and the presumed dynamics that each of the system types are prone to. It further provides a set of analytical terms that may come in handy, notably centrifugal and centripetal competition as well as the more nuanced thresholds for categorization. These ideas can undoubtedly serve a purpose also in the debate on ethnic heterogeneity in multiethnic states.

First, it suggests that a country should be classified by the number of relevant groups, not by the number of relevant coalitions. If a country with a large number of groups is found to be polarized into only two main camps, as Horowitz' 'economy of antipathy' stipulates, this should be seen as an effect, not as a predisposition. Secondly, Sartori's typology suggests that such polarization is a result of centrifugal competition – it is not the competition between the groups in itself that causes the polarization, but the direction of the competition. Third, Sartori proposes specific thresholds for categorizing countries into different classes according to *numbers* (fragmentation) and *balance*. Sartori does not distinguish clearly between these two dimensions, though this hardly poses a problem in terms of transposing the concepts to a multiethnic framework.

A more challenging feature in that regard is the fact that Sartori's assumptions of the direction of competition in the different systems are linked to the concept of 'floating voters'; understood as a person who is open for switching loyalty to a different party or group. The problem is that there is a fundamental difference between ethnic cleavages and other political cleavages. Generally speaking, political cleavages such as geographic, economic or ideological are easier to transcend, less stringent and thus

more dynamic. Ethnic cleavages are often deeper and harder to transcend as they are kept together by cultural and social boundaries.

This does not imply that ethnicity is a static property as such, but the fluidity of the concept pertains to the contextual and inter-temporal change in the relevance of the different ethnic cleavages (see e.g. Eriksen 1993). For instance, before the fall of Siyaad Barre in Somalia, the country was largely perceived as ethnically homogenous, but over the last 20 years the remnants of the Somali state has been ravaged by inter-clan rivalries, which, according to the multi-dimensional interpretation of ethnicity applied in this thesis, would constitute a clear-cut case of ethnic fragmentation. Yet, this does not mean that the cleavages between the clans were easily transcendable or irrelevant before 1990. Rather, it indicates that these cleavages were less predominant than other cleavages, for one reason or another. Yugoslavia represents a similar case at a similar point in time. When Tito's rule crumbled, the underlying ethnic cleavages surfaced and became the most important lines of division. I will return to this topic in Chapter 4. For now, it should suffice to say that in ethnically segmented societies the cleavages of the political party system tend to correspond largely to the boundaries of the ethnic groups, and the number of so-called 'floating voters' will therefore be limited (Lijphart 1977). The result will typically be a relatively static party system, divided along ethnic lines, and the strategy of the political parties will tend to focus on intra-ethnic competition since it is virtually impossible to recruit votes from opposing ethnic groups. This causes the political landscape to be multi-peaked and competition to be centrifugal, leading in turn to polarization (Sartori 1970, 1990; Horowitz 1985).

By suggesting that an increase in the number of groups increases the risk of anti-system parties to emerge, Sartori effectively underpins the argument of ethnic fragmentation as a factor that increases the risk of conflict. The mechanism that causes this risk to increase is the centrifugality of the political forces in the system. More groups implies more potential centrifugality, which in turn, and over time, increases the chance that one of the groups in the system decides its interests are no longer served by succumbing to or supporting the political framework in place. The said group could then be expected to employ anti-systemic means in order to change the

setup of the system, be it by orchestrating a classic *coup d'état* or by jumping on the secessionist bandwagon.

2.13 Disaggregating actor aims - territory vs. government

This raises another question that has received only limited attention in the debate on ethnic heterogeneity, namely the question of whether some ethnic constellations are mainly prone to a particular type of ethnic conflict. As such it latches on to the debate on 'disaggregation' that has intensified recently in the civil war literature (see Sambanis 2001, Fearon & Laitin 2003; Buhaug 2006, Wimmer et al 2009). According to Wimmer et.al (2009:318) "quantitative approaches tend to overaggregate the dependent variable and treat ethnic conflicts as though they have uniform causes". In the PRIO/Uppsala conflict dataset, civil wars, or intrastate wars, are coded according to the stated aim of the actors, with two alternative forms of incompatibility; government and territory (UCDP/ PRIO 2009). The former implies that the incompatibility between the parties is focused on the "type of political system, the replacement of the central government, or the change of its composition", while the latter concerns the "status of a territory, e.g. […] secession or autonomy (internal conflict)" (UCDP/PRIO 2009:4).[3]

Halvard Buhaug (2006) elaborates on this distinction and tests whether the relative capability of a rebel group affects the objective chosen. Among a number of interesting findings, herein the indication that democracies are more prone to secessionist conflict than non-democracies, he finds that ethnic fractionalization, as measured by Fearon & Laitin's (2003) updated ELF-index, significantly increases the risk of territorial conflict, while it has little effect on government conflict. The explanation for this goes back to the concept of relative rebel capabilities. Buhaug (2006:694) argues that a basic premise for using rebel capability as a determinant of the type of conflict is the notion that "capturing the state apparatus requires more power and resources than securing limited territorial authority". Moreover, it

[3] Wimmer et.al. (2009) employ the terms *secessionist* and *non-secessionist* to denote the same distinction. The meaning is understood to be the same.

presupposes that if a rebel group is capable of attaining control of the central government, it will inevitably attempt to do so, since such control constitutes "the ultimate price" (ibid.). Essentially, this suggests that the objectives and aspirations of a given ethnic group will be roughly proportional to its size and power. Hence, "weaker dissidents […] may have no other realistic option than to seek autonomy or secession for a limited piece of the country" (Ibid.).

The concepts of size and power are evidently not as simple and straightforward as the above argument alludes to. In particular the latter concept of *power* has been the subject of never-ending debates within practically all branches of the social sciences. This, of course, also applies to the study of multiethnic states. When the Tutsi minority in Burundi manages to maintain control over the Hutu majority for decades, it underlines the simple fact that ethnic relations are shaped by more than the sheer number of individuals belonging to each group. Horowitz (1985) also hints at this when he labels some systems as *ranked* and others as *unranked*, and some groups as *advanced* and others as *backward*.

It could consequently be argued that a classification of multiethnic states will remain hollow without a proper treatment of the concept of power. Yet, it is precisely the near insurmountable challenge of operationalizing the concept of power which makes it unfit for the overall purpose of this thesis. As noted above, the classification of multiethnic states should be based on relatively static criteria that do not change much over time, something the notion of power does not easily comply with. Along the same lines one could argue that the power relations of groups in multiethnic states are better described as effects of other underlying parameters, not as causes in themselves. Without exhausting the arguments on either side of this debate, it should be made clear that throughout this thesis the concept of power has been placed into the realm of the *ceteris paribus* – of the presuppositions that facilitate the exercise of simplification. A number of other group properties also fall into the same category; such as economic development, irredentist links, global outreach, historical ties and more. Groups are hence discussed and measured solely in terms of their numerical strength, not by other parameters that might indicate its relative strength and power in the system.

As for Buhaug (2006), he goes on to contend that the *fractionalization* measure should capture some of this relative capability, yet he does not explicitly deal with the question of balance, which, one could argue, should be just as relevant as the mere number of groups for determining the rebel objective. In fact, it is tempting to suggest that balance should be *more* important than fragmentation in this regard, especially if secession is considered the result of majority dominance and exclusion from power (c.f. Collier & Hoeffler 2004).

This could indicate that the closer a system gets to equilibrium (i.e. balance), the more prone it will be to conflict over government, since the prospects for any given minority to take control of the state (the ultimate price) will then be at its highest. Conversely, in dominant and even more so in hegemonic systems, the likely strategy of a rebel group will be to aim for autonomy or secession, since the prospects for taking over the state and keeping it are slim, to say the least. In hegemonic systems, however, the capability of the minority to initiate any form of rebellion should be lower than in dominant systems, since the discrepancy between majority control and minority capability is higher. This conforms to the dominance argument advanced by Collier & Hoeffler (2004).

The notions of majority control and minority capabilities admittedly rest upon the premise that the majority is in fact in control of the state in question and that it is the minority that rebels. That is of course not always the case, something the Tutsi control in Rwanda and Burundi are examples of.[4] The cases of Rwanda and Burundi also point to another important parameter in determining the type of rebel objective, namely the geographic concentration of the groups. It is commonly taken for granted that for a secessionist conflict to erupt, the ethnic group in question must have a territorial 'homeland' that defines the scope of the secessionist claims (see e.g. Horowitz 1985). Lijphart (1977) refers to this as *incongruence*, which implies that a state is divided into relatively homogenous entities of geographically separated ethnic

[4] The Tutsi make up approximately 10-15 percent of the population in both Rwanda and Burundi. Since independence Burundi has largely been controlled by the Tutsi minority which had full control over the army. In Rwanda, the Hutu were in control until the genocide in 1994, after which President Paul Kagame, a Tutsi, has led the country uninterruptedly (Uvin 1999; Rice 2010)

groups. A pertinent example of this is India, where the borders of the 28 federal states are drawn largely along ethnic lines (Mathew 2005:167).

In contrast, the Tutsi minorities in Rwanda and Burundi are quite evenly distributed around the territory of the two states (Uvin 1999). This *congruence* effectively precludes secessionism as a viable rebel objective, since there is simply no homeland to proclaim as independent. Incongruence can thus be seen as something close to a necessary precondition for secessionist conflicts to erupt, whereas in a congruent system the excluded and frustrated minorities would be forced to aim for government control.

While the face validity of this argument is strong, testing it statistically would require data on the level of ethnic groups that to my knowledge is non-existent as of today. Some groundbreaking work has recently been done on the geo-referencing of ethnic groups (see Weidmann et.al. 2010), yet these datasets use geographic polygons as the unit of analysis, not the groups themselves. This might make it difficult to determine whether a group perceives of this polygon, within which it constitutes a majority, as a homeland. This is work in progress, and it may very well open up a set of new opportunities for testing.

The question of incongruence is undoubtedly important in terms of complementing our understanding of rebel objectives, yet it might be less important for the classification of multiethnic states. Hence, it should suffice here to suggest that territorial conflicts are likely to be most prevalent in dominant systems, while government conflicts should be relatively more prevalent in balanced systems.

2.14 Four main hypotheses

If we now attempt to summarize the theoretical constructs introduced so far, we can discern four main hypotheses. First, the lion's share of the arguments over fragmentation seems to concur with the economic theory that initially made the ELF-index popular; that more groups increases the risk of conflict. Both Mearsheimer (2001) and Sartori (1990) indicate that a country with extensive ethnic fragmentation is likely to be more prone to conflict than a country with low fragmentation. This is because the addition of more groups increases the actor complexity of the system, makes competition largely centrifugal, and raises the costs of coordination – which in turn augments the risk of some form of ethnic conflict. Moreover, on high levels of fragmentation the lack of a clear majority creates uncertainty in terms of actor motives, which in turn is likely to bring about elements of what Sambanis (2001:262) labels an "ethnic security dilemma". The logical strategy for actors in such a situation is to resort to deterrence as the primary strategy of defense, and balance-of-power is likely to be the order of the day. And, as the structural realists contend, balance-of-power will be less stable if there are many actors in the system, simply because this increases the degree of uncertainty.

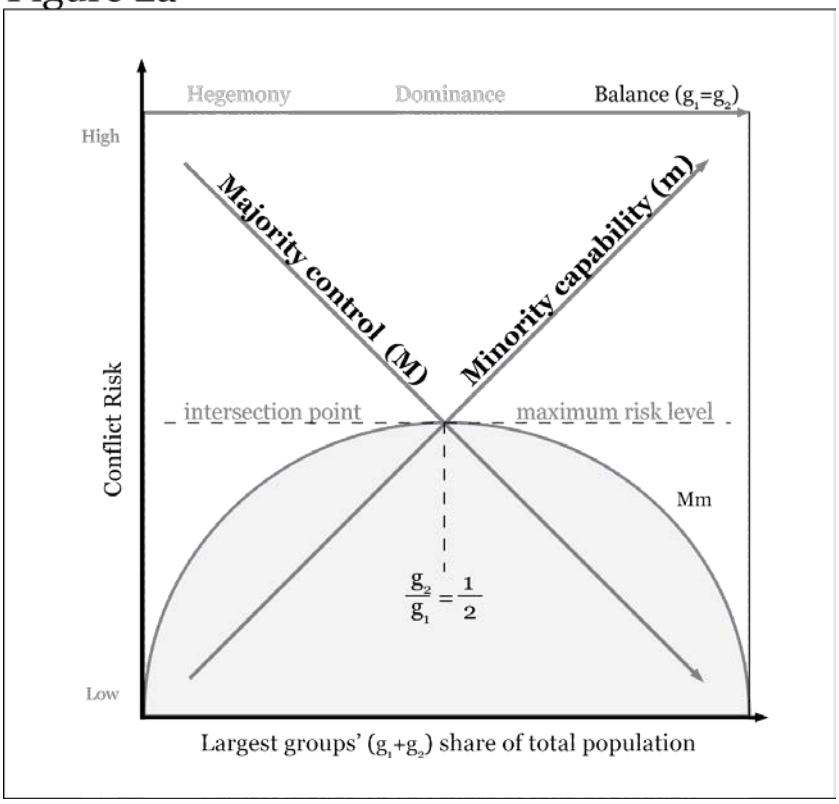
On the other hand, increasing fragmentation should not be expected to have the same effect in all countries, regardless of the balance in the system. In dominant systems, where the largest group is in clear but not overwhelming majority, the effect of fragmentation would presumably be countered by the effect of dominance – understood here as the propensity of the majority to exclude and discriminate against one or more minorities. As Collier and Hoeffler (2004) suggest; the risk from dominance is likely to be highest in systems with only two groups, since, in absolute terms, this would maximize the degree to which the majority is able to discriminate while the minority is able to resist.

Figure 2a provides a graphic overview of how majority control can be conceptualized to interact with minority capability. As the illustration shows, the expected control by the majority (*majority control*) is understood to decrease in proportion to the balance of the system, meaning that the more equal the two largest groups are in size (right

side), the less control the largest group will exercise over the minority, and the less inclined it will be to dominate. Conversely, the capability of the minority to resist this dominance (minority capability) runs in the opposite direction; the larger the minority, the greater the capability to resist exclusion by the majority. As the system approaches balance (i.e. the two largest groups are near equal in size), the majority control effectively disappears. At the other end of the spectrum, where majority control is at its highest, and minority capability is at its lowest (i.e. hegemony), the majority control is overwhelming, and what Collier & Hoeffler (2004) call 'exploitative potential' is reduced to a minimum.

Both the left and the right edges of the model thus represent situations where the risk of conflict resulting from dominance should be limited (Mm = 0). Assuming that these two axes are symmetric and proportional, one could further imagine a point of intersection at which the control of the majority and the resistance capability of the minority intersect. This would then indicate the ratio between the majority and the largest minority at which the effect of dominance should be at its highest, and hence

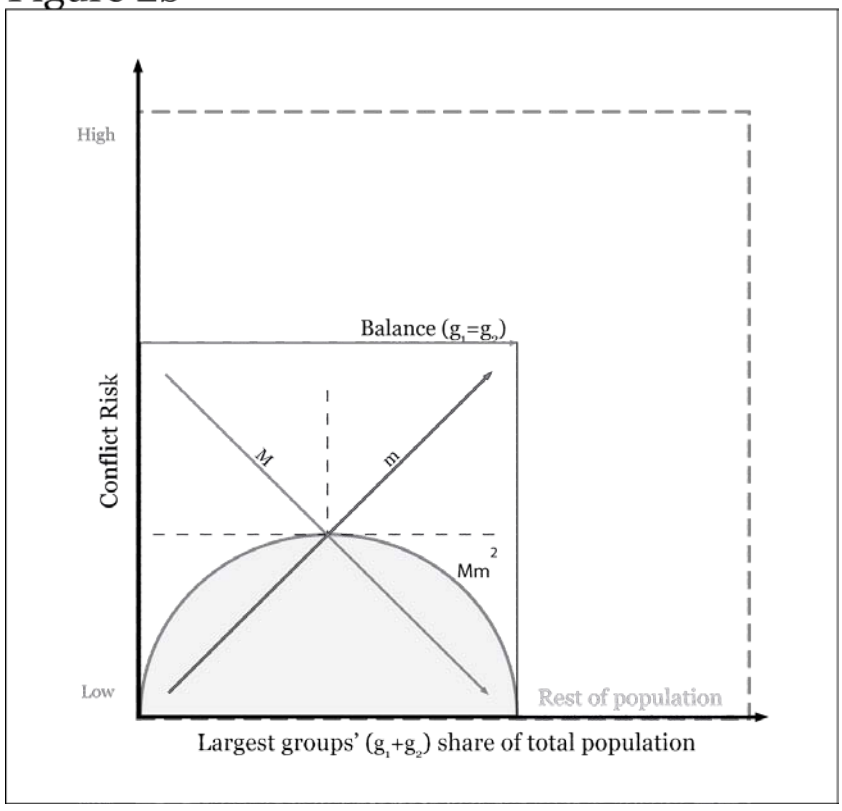
Figure 2a



the risk of conflict resulting from dominance. Given that the outer left perimeter of the model represents a near monopoly situation – where the largest group is close to 100 percent and the second largest group close to 0 – and that the outer right perimeter represents a 50:50 balance, the point of intersection could hence be written as a 2:1 equation: $g_2/g_1=1/2$.

When there are many groups in the system, the exploitative potential of the majority is smaller since the control of the majority and the resistance capability of the largest minority is limited in absolute terms. In Figure 2b this is illustrated by reducing the size of the box that depicts the sum of the two largest groups compared to the rest of the population. This suggests that if the exploitative potential is highest on low levels of fragmentation, while the risk from uncertainty, security dilemmas and centrifugal competition is highest on high levels of fragmentation, these would effectively neutralize each other if the two forces are comparable in strength. The expected impact of fragmentation in dominant systems is therefore hard to predict. By contrast, in hegemonic and balanced systems, where the risk from dominance is lower, higher

Figure 2b



levels of fragmentation should, in line with the arguments laid out above, increase the overall risk of ethnic conflict since there is no force that pulls the risk in the opposite direction.

In addition to providing a qualified interpretation of the effect of fragmentation, the model in Figure 2a and 2b thus provide the basis for the second hypothesis, namely that dominant systems as such are likely to be more prone to conflict than either hegemonic or balanced systems. The basis for this argument resides partly in the above notions of the exploitative potential which is maximized in dominant systems (intersection point), and partly as a deduced implication of the arguments for hegemonic and balanced stability. Provided that Wohlforth (2006) and Mearsheimer (2001) are right in their assumptions regarding bipolar and hegemonic stability, and that their arguments can in effect be transposed onto the domestic domain, the most volatile ethnic constellation would be the unbalanced (i.e. dominant) systems.

Third, there is reason to believe that some ethnic constellations will be more prone to a certain type of ethnic conflict. As noted above, the strategy chosen by an 'anti-system group' can be expected to depend on the opportunity and relative capability of the group, and if a given group has a chance of obtaining the 'ultimate prize' of control over the whole territory, one should expect it to aim for this objective. If, on the other hand, the rebel group is facing a clear majority, whose power and dominance effectively precludes strategies for taking control over the government then one should expect the rebel objective to be focused on autonomy or secession. If this assumption holds true, then balanced systems should be more prone to conflict over government while dominant systems should to be more prone to conflict over territory, which would be the fourth hypothesis. Hegemonic systems should also be associated with territorial conflict, but the reduced exploitative potential of hegemonic systems implies that the grievances and frustration, as well as the increased majority control, will limit the frequency of either type of war.

The four hypotheses can thus be summarized as follows:

Hypothesis 1:

Fragmentation increases the overall risk of ethnic conflict.

Hypothesis 2:

Dominance increases the overall risk of ethnic conflict.

Hypothesis 3:

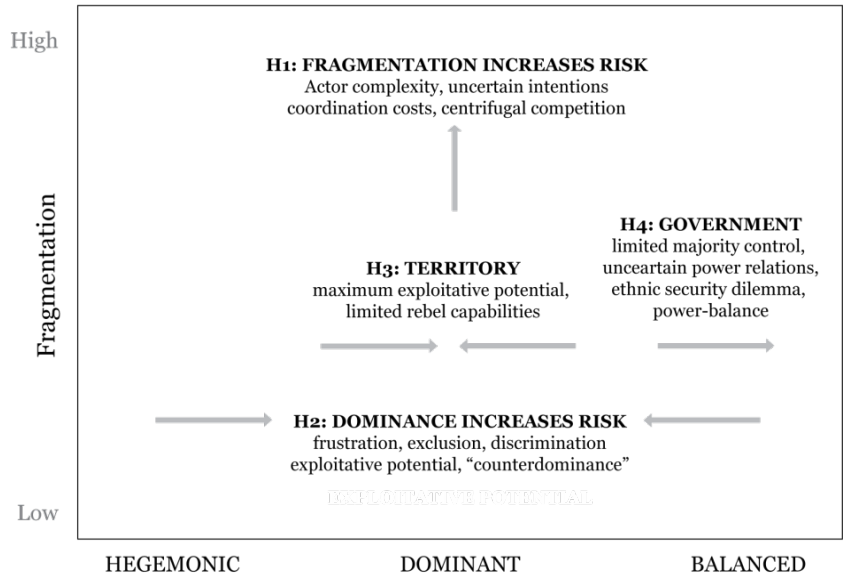
The risk of territorial conflict is highest in dominant systems.

Hypothesis 4:

The risk of government conflict is highest in balanced systems.

Figure 3 provides a graphic overview of the theoretical arguments that underpin the different hypotheses.

Figure 3



3 Operationalization

In this chapter a set of different measures for testing the four hypotheses are discussed. Some are drawn from previous research (such as fractionalization and polarization) while others, notably a new fragmentation measure, are introduced as new alternative measures. The main focus is placed on the two first hypotheses, i.e. on fragmentation and dominance.

In order to test the four hypotheses we need measures that, as accurately as possible, can capture the intended effects on both fragmentation and dominance. As for fragmentation, the measure should minimize the potential inconsistency resulting from individual coding rules of different ethnic conflict datasets. Secondly, the measure should aim to count only the relevant groups of the system. Third, the measure testing fragmentation should minimize the correlation with balance, in order to single out the effects of each. The same criterion logically applies to the dominance measure. In addition, the dominance measure should be adjustable so that different imbalance relations can be tested. While Collier & Hoeffler's (2004) theoretical assumptions, as explained and elaborated in Figures 2a and 2b, suggest that the most conflict prone balance ratio is 2:1, it should also be possible to test other balance ratios, such as 3:1 or 4:1.

3.1 Fragmentation

The most intuitive manner of measuring fragmentation would be to simply count the number of groups in each country, preferably guided by a set of rules for how groups should be included or discarded. The problem with using the plain number of groups is that countries with very high numbers of groups could disproportionately influence the results. To limit this effect, one could calculate the natural logarithm of the number of groups instead, thereby reducing the marginal effect of additional groups at high levels of fragmentation. This would be in line with Sartori's (1990) reasoning regarding

atomized systems, in which the addition of another group no longer alters the overall dynamics of the system.

In order to test the alternative hypothesis – namely that fragmentation *reduces* the risk of conflict – one could also adopt the measure used by Deutsch & Singer (1964) for calculating the number of interaction opportunities. The measure is written in the following form

$$IO = \frac{N(N-1)}{2}$$

where *N* is the number of groups in the system (Deutsch & Singer 1964:393). Figures 4a and 4b show the relationship between these three varities of measures, using the number of groups as the base. Figure 4a displays the interaction opportunities, while Figure 4b the natural logarithm of groups compared to the actual number of groups. Note that *interaction opportunities* are hypothesized to reduce the risk of conflict, not increase it.

Figure 4a

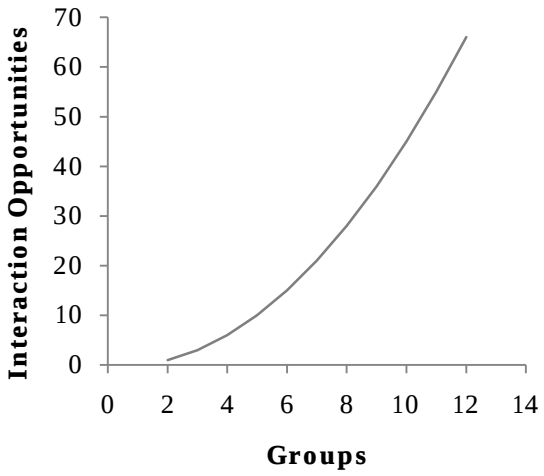
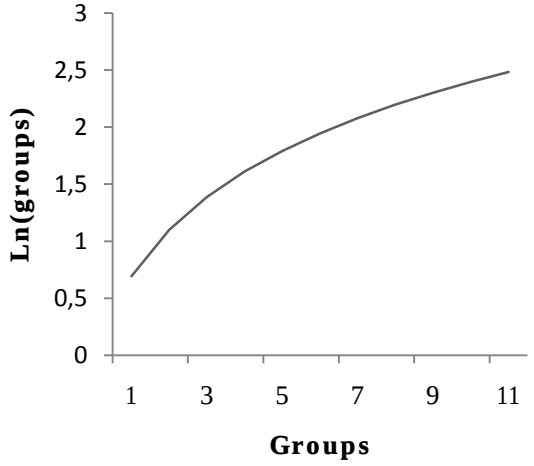


Figure 4b



A fourth way of measuring fragmentation could be to calculate the share of the two largest groups in relation to the rest of the population. This could be written simply as

$$F = 1-g_1-g_2$$

where g_1 is the largest groups' share of the total population, and g_2 is the share of the second largest group. Hence, if the two largest groups together make up nearly 100 percent of the total population, *fragmentation* would be low. Conversely, if the share of the two largest groups constitutes only a small share of the total, then *fragmentation* would be high since the remaining share would necessarily have to be split into entities smaller than the two largest groups.

The advantage of such an approach is that it is relatively insensitive to the individual coding of all the groups, and it would thus allow for easier robustness testing across different datasets of ethnic groups. Moreover, it effectively sidesteps the question of which groups to consider as relevant. It simply suggests that if the share of the two largest groups is small, then fragmentation must necessarily be high. The reason for not using only the largest group is that the largest group alone will be highly correlated with the *balance* of the system.[5] This could, of course be controlled for, but lower correlation between the two dimensions (fragmentation or balance) makes it easier to determine which of the two that has the highest impact on conflict risk. In order to focus the measure on fragmentation, it therefore makes sense to include the two largest groups instead. This does not eliminate the problem, since there will still be a certain amount of correlation between balance and fragmentation, especially on very low levels of the former, yet the effect should be mitigated.

Another drawback of such a measure of fragmentation is that it does not capture the nuances of the distribution beyond the second largest group. A country with a (*50,25,25)* distribution would obtain the same score on the *fragmentation* measure as a country with a (*50,25,1,1,..1)* distribution. Theoretically, this difference could alter the ability of the minorities to unite in opposition against the majority, a property that would be concealed in this fragmentation measure. On the other hand it also shows how the alternative – the mere counting of groups – might say even less about the dynamics of the system, since it does not take into account the actual size of any of the groups.

[5] Fearon & Laitin (2003) find that the size of the largest group has no significant effect on the onset of civil war, neither ethnic nor non-ethnic.

The measures that do take the size of all the groups into account are *fractionalization* and *polarization*, as introduced above. The main drawback with these measures is that they capture both the number of groups and balance at the same time. Nevertheless, both measures can be expected to give an indication of whether the risk of ethnic conflict is highest in systems with few groups or with many groups, and thus provide some clue as to the validity of the hypothesis.

In line with previous research (see e.g. Easterly & Levine 1997; Ellingsen 2000; Sambanis 2001; Collier & Hoeffler 2004; Buhaug 2006) the measure for *fractionalization* can be calculated using a version of the Herfindahl index. It can be written in the following form

$$F = \sum_{i=1}^{N} g_i(1 - g_i)$$

where *g* represents group *i*'s share of the total population, and *N* the number of groups in the system. The measure ranges from 0 to 1 and approaches 1 when all groups are equal in size and the number of groups nears infinity. As such, one should expect the *fractionalization* measure to be closely correlated with the *fragmentation* measure, since both react to increasing number of groups. The main difference is that *fractionalization* is sensitive to the balance between the two largest groups. In a country of only two groups, where the largest is 99 percent and the second is 1 percent of the population, both measures will return values close to 0. But if the two groups are closer to 50 percent each, then fractionalization would increase towards .5, while fragmentation remains close to zero. This implies that for the medium levels of fractionalization, the corresponding score on the fragmentation index would generally be lower. On very high levels of fractionalization, the two measures will converge again.

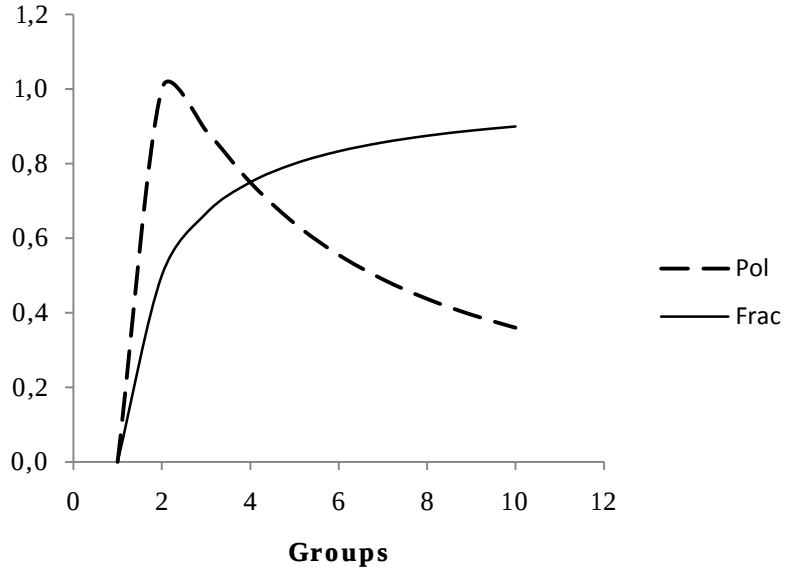
The *polarization* measure was not originally constructed to capture fragmentation as such. On the contrary, it captures the lack of it, since it is maximized when there are only two relevant groups. But precisely for that reason it can also serve it test whether fragmentation has any effect on ethnic conflict, since it would yield a negative

coefficient if Hypothesis 1 is correct. The measure can be calculated using a simplified version of Esteban & Ray's (1994) measure, as employed by Reynal-Querol (2002). It is written in the following form

$$P = 4\sum_{i=1}^{N} g_i^2(1 - g_i)$$

The measure thus calculated is set to maximize on a scale from 0 to 1 when two groups constitute 50 percent each (50,0,0,..50). Moreover, it exponentially favors large groups over small ones, something which makes sense given the fact that the mechanism of bipolar antagonism (as discussed above) is most likely to take effect on high levels of polarization. Figure 5 compares *fractionalization* and *polarization* for fully balanced systems ($g_1=g_2=g_3=\ldots g_N$) with an increasing number of groups.

Figure 5



In total, this leaves us with a range of six different measures that all can be used in different degrees to test the effect of fragmentation on the risk of ethnic conflict:

1. The number of groups: *Groups*
2. The natural logarithm of the number of groups: *Ln(Groups)*
3. The interaction opportunities between the groups: *Interaction Opportunities*

4. The share of the population not belonging to either of the two largest groups: *Fragmentation*
5. The chance that two randomly selected individuals belong to different groups: *Fractionalization*
6. The extent to which the country is split into two equally sized groups: *Polarization*

Of these six measures, the fourth (*fragmentation*) is the one that best meets the criteria introduced above: It minimizes the potential inconsistency resulting from individual coding rules of different ethnic conflict datasets, and it ensures that only relevant groups are counted (since it only takes the size of the two largest groups into account) while minimizing the correlation with the balance dimension. All the six measures are still included in the subsequent regressions in order to test their respective performances.

3.2 Balance and Dominance

Testing this hypothesis first requires a measure for calculating the balance of the system. The most straightforward manner of doing this is to use the ratio between the largest and the second largest group in the system, written as

$$B = \frac{g_2}{g_1}$$

where g_1 is the largest and g_2 the second largest group. This yields a linear index which is maximized when the two largest groups are equal in size (g_2/g_1=.5/.5), and minimized when the largest group approaches 100 percent of the population (g_2/g_1=0/1). This could in turn form the basis for a binary measure for dominance by placing countries into three classes of systems; *hegemonic*, *dominant* and *balanced*. The *dominant* class could theoretically be understood to comprise the center half of the balance spectrum. If the intersection point introduced in Figure 2 represents the maximum conflict risk, and if each end of the spectrum represents minimum risk, then the thresholds could be set half way towards either end of the scale *(B=.25 and*

B=.75). This implies that systems in which the second largest group constitutes less than 25 percent of the largest group would be classified as *hegemonic*. Conversely, all systems in which the same ratio is above 75 percent would be classified as *balanced*. All those in between would be classified as *dominant*. A crude manner of testing the hypothesis could thus be to use a dummy that codes dominant systems as 1, and hegemonic and balanced systems as 0. Yet, this would not remedy the shortcomings of the binary measure for dominance used by Collier & Hoeffler (2004), where a country is coded as dominant if the largest group constitutes between 45 and 90 percent of the population. If the aim is to identify the ratio between the largest and second largest group at which the risk of conflict is at its highest, then a more sophisticated approach is required.

Such a measure can be developed by enabling the above *balance* measure to be adjusted for different 'peak points'. This could be done by introducing a parameter (*c*) that indicates the *balance* value at which the measure will be maximized on a scale from 0 to 1. When *B=c*, the measure should thus yield 1. To ensure symmetry in the measure, the difference between *balance* and *c* could be squared. This also makes the curve parabolic, which corresponds well with the theoretical notion that the most volatile *balance* ratio is not *at* the 'peak point' but *around* it (c.f. Horowitz 1985, Sartori 1990). The measure could be written in the following form

$$D^* = 1 - (B - c)^2$$

Figure 6a shows how the *balance* ratio at which the conflict risk is assumed to be at its highest changes with different values of *c*. The drawback of this equation is that it does not utilize the whole spectrum from 0 to 1, and that different values of *c* span different parts of the spectrum. When *c=.5*, the minimum obtainable value for *D** is *.75*, while the minimum value for *c=.7* and *c=.3* is *.5*. Adjusting this requires a dividend that changes with different values of *c*. It could be devised by taking *c=.5* as the base value, adding the distance from this base (.5) to *c*, and squaring the sum. This ensures that the dividend is minimized at c=.5 while gradually increasing on higher and lower values of *c*. The *D** measure could thus be extended and written as follows

$$D = 1 - \frac{(B - c)^2}{(|c - .5| + .5)^2}$$

With *c=.7*, the dividend becomes $(|.7-.5|+.5)^2 = (.2+.5)^2 = .49$. With *c=.3* the dividend is the same and with *c=.5* it returns $(|.5-.5|+.5)^2 = (0+.5)^2 = .25$. This ensures that each value of *c* returns dominance values between 0 and 1, and it effectively allows for the identification of the hypothesized peak point of ethnic conflict risk empirically. Figure 6b illustrates how the measure now changes with different values of *c*.

Figure 6a

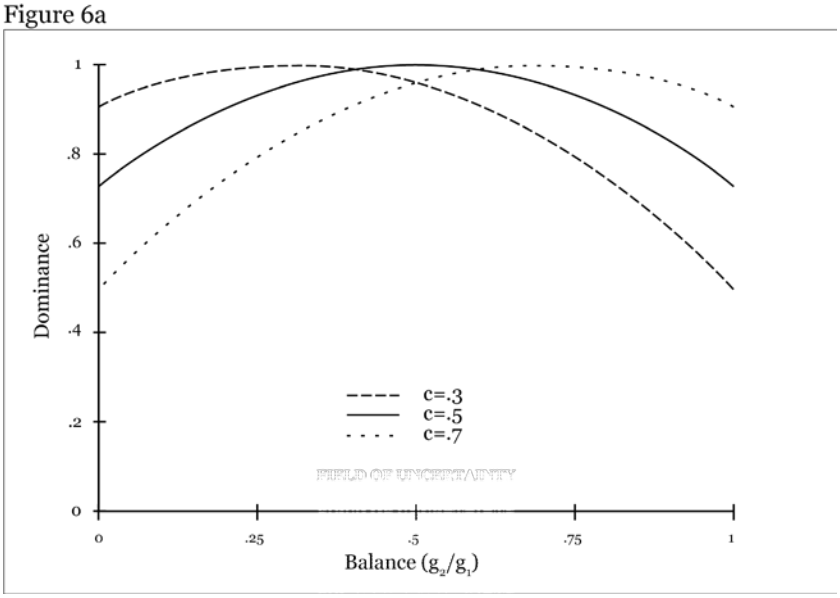
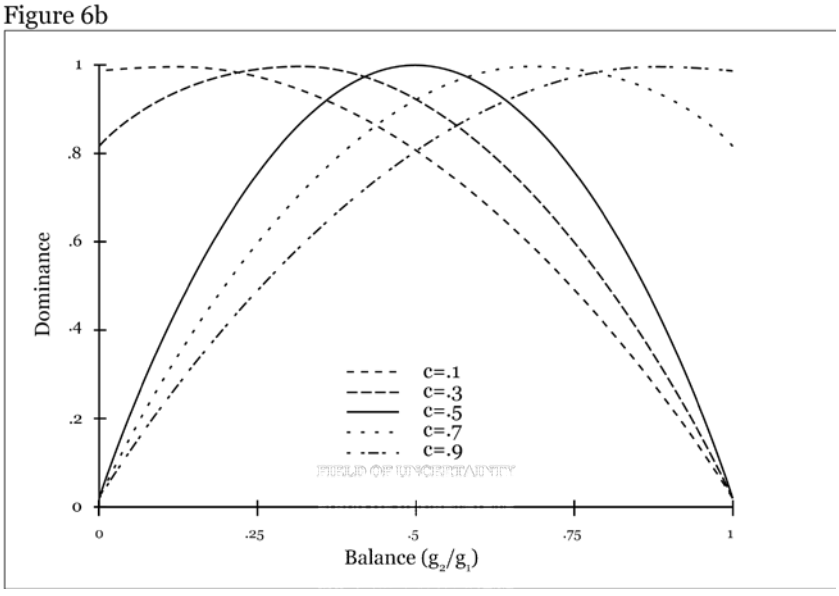


Figure 6b



In order to distinguish this measure of dominance from the binary measure employed by Collier & Hoeffler (2004), the latter is henceforth referred to as $dominance_{CH}$. The reason for not using a different term to describe this measure is that the theoretical construct upon which it is built is drawn largely from Collier and Hoeffler. For clarification, I should perhaps also repeat that fragmentation and balance represent the two main dimensions of the heterogeneity landscape. Each of these now have a measure explicitly aimed at placing countries along these two dimensions – fragmentation and balance. The measures are consitently written in italics in order to distinguish them from the dimensions. The *dominance* measure also aims at the balance dimension, yet it attempts to identify a nonlinear pattern and thus identify the value of *balance* at which the risk of conflict is highest.

Since the effect of dominance is assumed to be at its highest on low levels of fragmentation (c.f. Figure 2b), one could devise an additional test that only includes countries with low values of fragmentation. The idea would be to test the effect of dominance where it is theoretically thought to be at its highest, triggered by the cases such as Sri Lanka and Rwanda. If this test returns insignificant results, the notion that dominance may be most relevant on low levels of fragmentation should be discarded. The interaction between *dominance* and *fragmentation* could also be checked by multiplying the two measures and include this as a separate variable.

3.3 Government and Territory

The theoretical argument that underpins Hypothesis 3 is linked to notion of relative rebel capability, as discussed above. It is further coupled with the concept of exploitative potential, which suggests that the risk of conflict would be highest when the largest group is double the size of the second largest group ($B=.5$).This because $B=.5$ is the center-point between absolute majority control and absolute parity, as derived from Collier & Hoeffler's (2004) notion of 'clear but not overwhelming majority' and explained in Figure 2. In combination this indicates that the risk of territorial conflict will be highest in dominant systems since the relative capability of the minority will be large enough to contemplate armed resistance, but not large enough to aim for full control of the state. An underlying assumption for this

hypothesis is that the conflict will be started by rebels, and not by the majority. This is, of course, not always the case, but it nevertheless remains a viable assumption.

The test of whether this hypothesis holds true could be done in the same manner as for H_2, but with the difference that it singles out only territorial conflicts on the dependent variable. Fragmentation should also be added to these regressions, in order to test the assumption that fragmentation is less relevant for territorial conflicts.

The theoretical foundation for Hypothesis 4 rests on the dual notions of relative rebel capability and the presumed effects of fragmentation. If rebel capability increases in proportion to the size of the group in question, then the risk of government conflict should be at its highest in balanced systems, since the discrepancy between majority and minority would be at its lowest. In addition, the general arguments for increased risk from fragmentation indicate that the risk of government conflict should be at its highest in balanced systems with high fragmentation.

This could be tested by simply employing the linear measure of balance alongside a measure that tests fragmentation. These could also be combined into a new variable to see how they interact. As for H_3 one could also use a simple dummy variable to check whether balanced systems indeed have a higher risk of government conflict. These tests would, of course, have to be conducted on a dependent variable that only includes government conflict onsets. H_3 and H_4 should also be tested together in a binominal regression that captures the possible interaction between the two types of conflict, and the effect each of the explanatory variables have on the outcome.

4 Methodological Issues

In this chapter some of the methodological issues related to the statistical study of ethnic conflict are discussed. The chapter 2s divided into two main parts; input problems and output problems. The first mainly concerns issues related to the coding and interpretation of ethnicity, while the latter focuses on matters related to the coding of the dependent variable. The chapter concludes by pointing to the Ethnic Power Relations dataset and the corresponding dataset of ethnic conflicts (both by Wimmer et.al. 2009) as the best available sources for statistical testing of the relationship between ethnic heterogeneity and conflict.

4.1 Input

I mentioned earlier that the ELF-index has been criticized for its lack of theoretical underpinnings. The theoretical discussion in Chapter 2 aimed, to some extent, to remedy this shortcoming by placing the link between fractionalization and conflict into a broader theoretical framework of IR and political science theory. Yet, this theoretical weakness represents only part of the criticism directed at the ELF-index.

Drawing on constructivist theory Laitin & Posner (2001) point to four fundamental challenges to the ELF index that all revolve around the exercise of counting and coding ethnic groups. First they contend that by using ethnic data from the early ‘60s to explain the outbreak of conflict over a period of half a century, statisticians make the cardinal error of viewing ethnicity as a fixed property that does not change over time. Laitin & Posner (2001:4) liken this to an economist who uses the rate of inflation in 1945 to measure a country’s economic prosperity in 1990. While this analogy might be deemed an exaggeration, most sociologists and anthropologists would insist that ethnicity is a fluid concept; multilayered and multifaceted, contextual and largely negotiable (see Eriksen 1993; Jenkins 1997).

Using the example of Somalia, Laitin & Posner (2001) show how the relevance of the different ethnic or tribal cleavages can change over time. When the *Atlas Narodov Mira* (ANM) was compiled, Somalia was considered one of the most ethnically

homogenous countries in the world, yet towards the end of the cold war and the fall of Siyaad Barre, the country became "severely divided by inter-clan fractionalization" (Laitin & Posner 2001:2). Hence, if the ethnic configuration of a country is to serve as a viable explanatory variable, data on ethnic groups will need to be coded in time-series, similar to variables such as GDP and population.

Laitin & Posner's (2001) second criticism regards the one-dimensional interpretation of ethnicity used in the ANM dataset. Provided an understanding of ethnicity as a fluid and contextual concept, one cannot *a priori* adopt a single criterion for ethnic divisions. In some societies ethnic groups will be defined by religious cleavages, in others it could be linguistic, cultural or even pigmental (as in much of Latin America). Quite often a country would even be divided by multiple types of cleavages at the same time. Iraq constitutes such a case, with the Kurds separated from the rest of the country by a linguistic and cultural cleavage, while the Shiite and Sunni in the south are divided along a confessional cleavage. In the ANM, these latter groups are lumped together simply as 'Iraq Arabs', since the dataset does not include confessional parameters. There is no simple remedy for this problem, besides coding a country's ethnic groups according to the most relevant cleavages, regardless of the nature of these cleavages.[6]

The question of relevance points to a third criticism of the ELF-index. As Laitin & Posner (2001:3) note, "once we have established which dimension of ethnic cleavage is salient, we still need to decide which groups we should include in our count". This idea rests on the assumption that not all groups are equally relevant politically, and those that are deemed irrelevant should be excluded from the statistical measures that aim to explain political outcomes such as armed violence and civil wars. As such, the challenge is analogous to the exercise of counting relevant political parties in order to classify party systems, a question that has received considerable attention from political scientists. Sartori (1990) suggests that a general rule for deciding whether a

[6] There is good reason to believe that some types of ethnic cleavages are more conflict prone than others. Pigmental cleavages, it could be argued, are less 'dangerous' than linguistic ones, but testing this would require data on group level that as of today is not available.

party is relevant is to assess whether it has coalition potential, that is, whether it over time has been considered by the other parties to be a potential coalition partner. A similar criterion could be applied for ethnic groups, though with certain practical adjustments. The inherent challenge in such an approach to coding is that one risks excluding large parts of the populations from the dataset. I return to this below.

The fourth criticism advanced by Laitin & Posner (2001:4) concerns the potential problems of endogeneity involved in the coding of ethnic groups. Quite often the relevance of ethnic cleavages is the *effect* of political change, not the *cause*, which would have the "causal arrows going the wrong direction". That said, one could argue that the potential problems of endogeneity are even more prevalent when ethnic groups are coded according to relevance, in the manner proposed by Laitin & Posner. If ethnic groups are to be counted according to the effect these groups can be expected to have on the outcome variable, then the risk of defining the independent variable by means of the dependent variable is compelling. Posner (2004:856) recognizes this dilemma, but insists that as long as the criterion for relevance is reasonably decoupled from the concepts measured on the outcome variable this should not cause systematic measurement errors. In any case, he contends, the implications of using an index "that captures a constellation of ethnic groups that is in some cases completely unrelated to the process whose effects are being investigated" would be even worse (ibid.). In other words, counting groups by *relevance* is perceived as the least bad option.

There is an alternative approach to this dilemma that could strike a balance between the question of relevance and endogeneity. The *fragmentation* measure escapes some of this criticism since it disregards completely the configuration of the remaining share of the population. Small groups would thus be counted only as far as they contribute to this 'residual share'. The drawback is of course that the nuances of highly fragmented societies are lost, yet if a balance needs to be struck between relevance and endogeneity such a measure of fragmentation might be the most viable manner of doing it.

In recognizing the shortcomings of the ANM – in particular its age and vestigiality – some scholars have turned to others sources of ethnic groups. Alesina et.al. (2003) use data from *Encyclopedia Britannica (EB)*, Fearon & Laitin (2003) resort, among other things, to the *CIA Factbook*, while Montalvo & Reynal-Querol (2002, 2005) rely mainly on the *World Christian Encyclopedia (WCE)*. Despite the application of certain counting rules for relevance, these datasets fall far short of meeting the criticism advanced by Laitin & Posner above. The WCE, for one, fails to distinguish between different Muslim denominations and effectively lumps together the Shiite and the Sunni in Iraq in one large category labeled 'Iraqi Arab'.

In an attempt to improve on these shortcomings, Posner (2004) develops a dataset of Politically Relevant Ethnic Groups (PREG), and, despite the abovementioned problems of endogeneity, this goes a long way in improving the validity of the ethnic source material. The main limitation of this dataset is that it covers only Africa.

Recently a more comprehensive project has been undertaken by Wimmer, Cederman & Min (2009) in a project entitled *Ethnic Power Relations* (EPR). With the aid of nearly 100 different country experts, they code politically relevant ethnic groups in a total of 155 countries for the period from 1946 to 2005. Of these, 128 countries are considered multiethnic, the rest are classified as ethnically homogenous, i.e. having only one relevant ethnic group. By coding the relevant groups as they change over a period of nearly 60 years, this allows for testing the effect of changes in ethnic configurations over time. In order to minimize problems of endogeneity they also make sure that "the coding of ethnic power relations reflects the power constellations before the outbreak of conflict in cases where political changes occurred in the same year as a conflict" (Wimmer et.al. 2009:326).

The main problem with the EPR dataset pertains to the abovementioned problem of exclusion. For certain countries in the dataset, such as Liberia, the total number of relevant groups cover only a small share of the population. Hence, if countries are classified according to the number of relevant ethnic groups, there is no way of

knowing how large the ‘shadow population’ is in each country - meaning the share of the population not included in the relevant groups.

4.2 Output

The second main category of methodological problems concerns the output phase of the statistical testing, and more specifically the question of how to code the dependent variable. Nicholas Sambanis (2001:259) notes that “as we undertake further study of the causes of civil war, it is important to know if our conclusions apply equally to wars of different types”. First of all, one should not expect the ethnic heterogeneity to explain the onset of non-ethnic conflicts. Hence, it makes little sense to include all civil war onsets in the dependent variable. The challenge is, of course, that it is often hard to say what constitutes an ‘ethnic’ civil war. Fearon & Laitin (2003:79) code conflicts as ethnic when the fighters were “mobilized primarily along ethnic lines”, and find that about 50 percent of the civil wars in their dataset were ethnic in nature, while about 20 percent were non-ethnic conflicts. The remaining conflicts were coded as ambiguous.

Wimmer et.al (2009) find a similar distribution in their recently constructed dataset. They define conflicts as ethnic if the identified aim of one or more of the fighting factions is to achieve either “ethnonational self-determination, a more favorable ethnic balance-of-power in government, ethnoregional autonomy, the end of ethnic and racial discrimination, language [or] other cultural rights” (2009:326). Drawing on the PRIO/Uppsala Armed Conflict Dataset (Gleditsch et.al 2002) they identify a total of 215 internal armed conflicts in the period from 1945 to 2005, of which 110 where coded as ethnic (Wimmer et.al 2009:327) and the rest as non-ethnic.

Secondly, not all ethnic conflicts have the same aim. Territorial conflicts are likely to be triggered by somewhat different causes and mechanism than conflicts over government. Testing this evidently requires a dataset that distinguishes between the types of ethnic conflict. Crucially, Wimmer et al (2009) includes such a distinction in their dataset. In line with the above discussion on disaggregation, this allows for the testing of some ethnic configurations are particularly prone to one to of ethnic conflict.

The sum of these attributes makes the EPR dataset on ethnic groups and the corresponding dataset on conflicts, both compiled by Wimmer et.al (2009), the most suitable basis for testing the four main hypotheses of this paper. In the subsequent regression analyses, the EPR dataset is therefore used as the main source of data, unless otherwise indicated.

5 Testing & Results

This chapter starts with an overview of the different variables employed to test the four main hypotheses, including the control variables. I then go on to test the four hypotheses, using logistic regression models. Hypothesis 3 and 4 are tested together, first in separate models and then binominally. After a slight detour on the topic of anocracy, some preliminary conclusions from the test are drawn. I then move on to test the robustness of the results across different datasets for both conflicts and ethnic groups. The results are also discussed in terms of probabilities, before the chapter ends with an outline of a classification for multiethnic states.

The tests are conducted using standard logistic regression models in Stata.[7] In some of the tests binominal regression models are also employed in order to capture the interaction between the two types of ethnic conflicts (*territory* and *government*).

5.1 The main ethnic variables

To test Hypothesis 1, the six different measures expected to give an indication of system fragmentation are employed; namely the number of *groups*, the *natural logarithm (ln) of groups*, the *interaction opportunities*, *fragmentation*, *fractionalization* and *polarization*. All are calculated on the basis of the EPR dataset (Wimmer et.al 2009). As mentioned, the dataset contains annual data for the years 1946 to 2005 for a total of 155 countries, of which 27 are coded as ethnically homogenous. These are dropped in the regressions since the aim of these tests is to identify patterns among ethnically heterogeneous countries, not to test the relationship between multiethnic and monoethnic countries as such.

These variables are all regressed onto a binary dependent variable coded as 1 in the first year of an ethnic civil war and 0 in years without war. This is drawn directly from Wimmer et.al (2009). Ongoing war-years are coded as missing in order to omit wars

[7] Stata 10. See do-files with replication data in Appendix 1.

that start while the first war is still ongoing. Wimmer et.al. (2009:328) find that this reduces the number of observations by about 15 percent, but that it yields almost exactly the same results as when all years are included and ongoing war years added as a separate dummy.

Hypothesis 2 is tested with the basic linear *balance* measure as well as a set of different versions of the *dominance* measure. Collier & Hoeffler's (2001) dominance variable ($dominance_{CH}$) is also added for comparison, and so is a new binary $dominance_{t=1}$ variable coded 1 if *balance* is between .25 and .75. If the ratio between the two largest groups does indeed affect the risk of conflict, then this should be captured by at least one of these measures.

Hypothesis 3 is tested using the best measures for fragmentation and dominance respectively, regressed onto a dependent variable that only captures territorial ethnic conflicts. For Hypothesis 4, the standard *balance* measure is used in combination with a measure of fragmentation. The dependent variable for these latter tests covers only government conflicts. The interaction between these types of conflict is also tested by means of binominal regression models.

5.2 Control variables

Using data from Wimmer et.al (2009) I also include a set of control variables that cover the most salient determinants of civil war, such as population size, GDP per capita, and form of government. Wimmer et.al (2009) use GDP data compiled from the Penn World Table and the World Bank, supplemented with data from Fearon and Laitin (2003). The numbers are in constant 2000 US dollars. Both population and GDP per capita are lagged in order to limit endogeneity problems.

Wimmer et.al. (2009) also use a dummy for anocracy which aims to capture the non-monotonic relationship between democracy and domestic conflict – that is, the tendency of both democracies and autocracies to be relatively stable compared to countries in between (see Muller & Weede 1990; Ellingsen 2000; Hegre et.al. 2001; Sambanis 2001; Reynal-Querol 2002; Mansfield & Snyder 2005; Hegre & Sambanis 2006). Based on Polity IV data, the threshold for coding *anocracy* as 1 is set at -5 and

+5. Countries with higher or lower scores on the democracy index are coded as 0. A variable for political instability is also added, coded as 1 if there has been a change of more than 3 points on the Polity score in the last 3 years.

There has further been a debate over whether or not the availability and 'lootability' of natural resources affects the risk of a either territorial or government conflict (see e.g. Ross 2003; Collier & Hoeffler 2004; Buhaug 2006). To control for this, an oil production per capita variable is included, based on data from Wimmer & Min (2006). Wimmer et.al. (2009) also control for mountainous terrain, which has been hypothesized to impact the ability of rebels to escape and hide (see e.g. Fearon & Laitin 2003). Finally, the number of years since last war is also controlled for with a *peace years* variable. The idea of the latter is that conflict breeds conflict and stability reinforces stability. All control variables are drawn directly from the *EPR Ethnic Armed Conflict dataset* (Wimmer et.al 2009).

5.3 Testing hypothesis 1: the effect of fragmentation

In Table 1, the six variables aimed at measuring fragmentation in different forms are tested separately. The same dataset is used for all the models. The first five are positive, and three of them are also significant on a 0,1 percent level or better. *Groups* and *ln(groups)* seem to yield the best results in terms of significance, while the *fragmentation* variable is the only one that remains significant at below 5 percent when all measure are included in the same model (1.1). *Fractionalization* is also positive, and close to 5 percent significance, while *polarization* is negative but weak. All the models in Table 1 thus seem to point in the same direction and towards general support for Hypothesis 1; that the risk of ethnic conflict is higher in countries with many ethnic groups.

In terms of overall explanatory power the race appears tied between model 1.3 and 1.5, both of which yield a log-likelihood of -352,6. This suggests that *fragmentation* and *ln(groups)* are equally good at measuring the effect of increasing numbers of groups on the risk of ethnic conflict. However, the correlation between the two measures is limited to .48, suggesting that there are considerable discrepancies between what the two measures capture. Part of this could be attributed to the fact that the EPR dataset includes only *relevant* groups, effectively excluding parts of the population from the equation. The extreme case of this is Liberia, as mentioned above, where the five listed groups together constitute only 24 percent of the population.[8] In other words, a whopping 76 percent of the population remains unaccounted for. It could of course be the case that three quarter of the population should in fact considered non-ethnic or

TABLE 1: Logistic regressions on ethnic conflict onsets: Fragmentation

	(1.1)	(1.2)	(1.3)	(1.4)	(1.5)	(1.6)	(1.7)
Groups	0.0704	0.0301***					
	(0.132)	(0.00884)					
Ln(Groups)	0.715		0.692***				
	(0.734)		(0.156)				
Interaction Opportunities	-0.00288			0.000662			
	(0.00347)			(0.000357)			
Fragmentation	2.958*				2.203**		
	(1.296)				(0.677)		
Fractionalization	-2.759					1.249	
	(1.688)					(0.690)	
Polarization	1.611						-0.559
	(1.310)						(0.690)
Peace years	0.00713	0.00300	0.00383	0.00322	0.00568	0.00339	0.00382
	(0.00998)	(0.00996)	(0.0101)	(0.00986)	(0.00975)	(0.00992)	(0.00970)
GDP per capita $_{lag}$	-0.0976*	-0.143**	-0.122**	-0.150**	-0.104*	-0.124*	-0.156**
	(0.0424)	(0.0477)	(0.0441)	(0.0493)	(0.0439)	(0.0502)	(0.0519)
Ln(Population) $_{lag}$	0.0831	0.168	0.0861	0.231*	0.238*	0.256**	0.293**
	(0.100)	(0.109)	(0.102)	(0.108)	(0.0975)	(0.0929)	(0.0971)
Oil production $_{pc}$	0.731**	0.840**	0.809**	0.817**	0.748**	0.748**	0.801**
	(0.280)	(0.288)	(0.277)	(0.288)	(0.259)	(0.265)	(0.269)
Anocracy $_{t=1}$	0.0213	0.0411*	0.0350*	0.0440*	0.0223	0.0353	0.0471*
	(0.0195)	(0.0190)	(0.0173)	(0.0193)	(0.0213)	(0.0191)	(0.0190)
Political instability	0.303	0.268	0.278	0.245	0.268	0.213	0.226
	(0.291)	(0.294)	(0.296)	(0.293)	(0.295)	(0.289)	(0.288)
Ln(Mountains)	-0.0180	-0.0632	-0.0601	-0.0581	0.00843	-0.0444	-0.0380
	(0.109)	(0.102)	(0.104)	(0.102)	(0.106)	(0.103)	(0.101)
Constant	-6.730***	-5.684***	-5.825***	-6.104***	-7.013***	-6.945***	-6.374***
	(1.225)	(1.046)	(0.955)	(1.055)	(0.975)	(1.089)	(1.140)
Observations	5059	5059	5059	5059	5059	5059	5059
Pseudo R^2	0.090	0.064	0.077	0.060	0.077	0.064	0.059
AIC	722.9	732.7	723.2	736.1	723.1	733.1	736.7
Log-likelihood	-347.5	-357.3	-352.6	-359.0	-352.6	-357.5	-359.3

Robust standard errors in parenthesis. Based on the Ethnic Power Relations dataset. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[8] For the year 2000 the EPR dataset lists the following relevant groups in Liberia: Gio (8%), Mano (7%), Krahn/Guere (5%), Americo-Liberians (2%), Mandingo (1,7%)

belonging to ethnically irrelevant groups. It seems more likely, however, that this amounts to either a coding error or to a problem with the coding rules.

The discrepancy between the measures also points to the fact that the number of relevant groups may vary even if the size of the two largest groups stays constant. When compared in the same model, it appears the *fragmentation* measure is in fact the most significant of them, and consequently that the size of the two largest groups is in fact a better indicator of the risk of conflict than the natural logarithm of the number of relevant groups. *Fragmentation* is therefore used to test the effect of group numbers in the subsequent regressions.

In terms of the control variables, *GDP per capita* yields a consistent, negative effect on conflict risk, while *population* is consistently positive. In model 1.2 and 1.3 the effect of *population* seems to be taken over by the group variables (*groups* and *ln(groups)*), which could be explained by the fact that for high levels of *groups* and *population*, the two are likely to correlate closely. *Anocracy* is significant for most of the models, but only at a 5 percent level. Finally, *peaceyears*, *political instability* and *ln(mountains)* remain insignificant throughout.

5.4 Testing hypothesis 2: the effect of dominance

Table 2 presents the results from adding measures of balance and dominance to model 1.5. The results are less conclusive than those of Table 1. In general, the balance dimension seems to have little if any effect on the risk of conflict, whether it is tested with the basic *balance* measure or with different versions of *dominance*. The best log-likelihood is obtained in model 2.2 and 2.3 (-351.1), where dominance is calculated with c=.4 and c=.5, though these models are only marginally better at explaining ethnic conflict onsets than when the balance measures were excluded altogether. In model 2.7 the product of *fragmentation* and $dominance_{c=.5}$ is added as a separate variable in order to test the interaction between the two variables. This does not improve the log-likelihood, suggesting that there is no interaction to speak of between *dominance* and *fragmentation*. This indicates that the effect of fragmentation should be similar at all levels of *balance*, hence contradicting the theoretical notion that

TABLE 2: Logistic regressions on ethnic conflict onsets: Balance & Dominance

	(2.1)	(2.2)	(2.3)	(2.4)	(2.5)	(2.6)	(2.7)	(2.8)
Fragmentation	2.216**	2.236**	1.939**	1.913**	2.173**	2.844***	1.751	
	(0.736)	(0.693)	(0.708)	(0.724)	(0.661)	(0.822)	(1.321)	
Balance $_{g2/g1}$	-0.0240							
	(0.455)							
Dominance $_{c=.4}$		0.962						
		(0.659)						
Dominance $_{c=.5}$			0.681				0.599	-0.0522
			(0.462)				(0.658)	(1.158)
Dominance $_{c=.6}$				0.501				
				(0.462)				
Dominance $_{t=1}$					0.0648			
					(0.347)			
CH dominance $_{t=1}$						0.437		
						(0.368)		
Fragm*Balance$_{c=.5}$							0.292	
							(1.926)	
Peace years	0.00570	0.00627	0.00590	0.00552	0.00575	0.00748	0.00578	0.0329
	(0.00975)	(0.00975)	(0.00979)	(0.00980)	(0.00966)	(0.0101)	(0.0101)	(0.0194)
GDP per capita $_{lag}$	-0.104*	-0.101*	-0.0990*	-0.1000*	-0.104*	-0.106*	-0.0988*	-0.421*
	(0.0437)	(0.0418)	(0.0422)	(0.0432)	(0.0440)	(0.0435)	(0.0422)	(0.186)
Ln(Population) $_{lag}$	0.237*	0.238*	0.250*	0.251**	0.240*	0.202	0.252*	-0.000412
	(0.0972)	(0.101)	(0.0991)	(0.0970)	(0.0988)	(0.105)	(0.0993)	(0.224)
Oil production $_{pc}$	0.0224	0.0214	0.0194	0.0197	0.0222	0.0221	0.0194	0.488**
	(0.0211)	(0.0203)	(0.0216)	(0.0220)	(0.0213)	(0.0213)	(0.0216)	(0.189)
Anocracy $_{t=1}$	0.750**	0.768**	0.730**	0.718**	0.749**	0.771**	0.732**	1.200
	(0.265)	(0.258)	(0.261)	(0.266)	(0.259)	(0.263)	(0.261)	(0.718)
Political instability	0.269	0.259	0.248	0.252	0.267	0.286	0.247	0.0548
	(0.294)	(0.294)	(0.289)	(0.289)	(0.294)	(0.298)	(0.289)	(0.727)
Ln(Mountains)	0.00828	0.0326	0.0376	0.0250	0.0125	0.00197	0.0402	-0.0131
	(0.106)	(0.109)	(0.109)	(0.107)	(0.112)	(0.108)	(0.113)	(0.263)
Constant	-7.001***	-7.823***	-7.551***	-7.416***	-7.054***	-7.069***	-7.523***	-4.393*
	(0.978)	(1.235)	(1.095)	(1.054)	(1.012)	(1.004)	(1.101)	(1.962)
Observations	5059	5059	5059	5059	5059	5059	5059	1565
Pseudo R^2	0.077	0.080	0.081	0.079	0.077	0.079	0.081	0.103
AIC	725.1	722.3	722.2	723.7	725.1	723.6	724.2	161.5
Log-likelihood	-352.6	-351.1	-351.1	-351.8	-352.5	-351.8	-351.1	-71.74

Robust standard errors in parenthesis. Based on the Ethnic Power Relations dataset. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

fragmentation should have the highest impact on low levels of dominance (c.f. discussion in Chapter 2).

Similarly, in model 2.8 countries with three groups or more are omitted from the regression in order to test the marginal effect of dominance at the point where it is hypothesized to have the highest effect. This limits the number of observations to 1565. Not only does this return an insignificant coefficient; it even appears the direction of it has shifted. Hence, insofar as the ratio between the largest and second largest group has any significant effect on the risk of conflict it does not show up when tested on an aggregate level of ethnic conflict. Whether the effect is any stronger when

the dependent variable is disaggregated by actor aim (territory and government) should become clear in Table 3, where the dependent variable is split into the two types of conflict.

5.5 Testing Hypotheses 3 & 4: Disaggregating the DV

In Table 3 the first two models are regressed onto binary variables for territorial and government conflicts respectively. The dataset is the same as in the previous tests (Wimmer et.al 2009) in which ethnic conflicts are also coded as either territorial or governmental. As in the previous tests, ongoing war years are coded as missing. In terms of ethnic variables, *fragmentation* is used in all models, while *dominance* is used for conflicts over territory and *balance* is used for conflicts over government. This corresponds to the last two hypotheses which suggest that territorial conflict should be most prevalent on high levels of dominance (H_3), while the government conflicts should be most prevalent on high levels of balance (H_4). In models 3.3 and 3.4 the same relationships are tested binominally, with territorial conflict onsets coded as 1 on the dependent variable and government conflicts coded as 2. The only difference between model 3.3 and 3.4 is that the former uses *dominance* to supplement fragmentation while the latter uses *balance*. After testing different versions of *dominance* (not shown) it turns out the best results are obtained when the peak point is set at $c=.4$. $Dominance_{c=.4}$ is therefore used instead of $dominance_{c=.5}$.

Interestingly, in Table 3 $dominance_{c=.4}$ becomes significant at a 5 percent level, both in model 3.1 (ordinary regression) and in model 3.3 (binominal). Moreover, it appears that fragmentation has little if any effect on the risk of territorial conflicts; it is far from any relevant statistical significance. The results obtained in models 3.1 and 3.3 thus point in the direction of general support for Hypothesis 3, that the risk of territorial conflict is highest in dominant systems.

For conflicts over government, however, the results are not as expected. Instead of yielding statistical significance for *balance*, models 3.2 and 3.4 restore the significance of *fragmentation*, while rendering *balance* largely irrelevant. Since *fragmentation* did not form part of the hypothesis as such, a separate test was also done excluding

Table 3: Logistic regressions disaggregated on actor aim

	(3.1)	(3.2)	(3.3)		(3.4)	
	Terr.	Gov.	Terr	Gov	Terr	Gov
Fragmentation	1.110	2.682**	1.159	2.915**	1.442	2.696**
	(0.926)	(0.934)	(0.928)	(0.910)	(0.941)	(0.935)
Dominance $_{c=.4}$	2.240*		2.244*	0.282		
	(0.973)		(0.972)	(0.847)		
Balance $_{g2/g1}$		0.393			-0.477	0.386
		(0.567)			(0.645)	(0.569)
Peace years	-0.00809	0.0161	-0.00789	0.0163	-0.00877	0.0160
	(0.0164)	(0.0131)	(0.0163)	(0.0128)	(0.0155)	(0.0131)
GDP per capita $_{lag}$	-0.0519	-0.171**	-0.0525	-0.170**	-0.0519	-0.172**
	(0.0523)	(0.0651)	(0.0524)	(0.0648)	(0.0537)	(0.0652)
Ln(Population) $_{lag}$	0.570***	-0.0288	0.570***	-0.0300	0.532***	-0.0222
	(0.144)	(0.131)	(0.144)	(0.133)	(0.143)	(0.132)
Oil production $_{pc}$	0.0115	0.0409	0.0114	0.0416	0.0143	0.0412
	(0.0277)	(0.0231)	(0.0280)	(0.0225)	(0.0246)	(0.0231)
Anocracy $_{t=1}$	1.022*	0.571	1.031*	0.611	1.040*	0.581
	(0.431)	(0.351)	(0.430)	(0.348)	(0.444)	(0.350)
Political instability	0.0487	0.391	0.0566	0.398	0.102	0.391
	(0.552)	(0.400)	(0.550)	(0.400)	(0.547)	(0.399)
Ln(Mountains)	-0.0302	0.0723	-0.0303	0.0807	-0.0479	0.0712
	(0.143)	(0.147)	(0.143)	(0.162)	(0.142)	(0.146)
Constant	-12.54***	-5.518***	-12.55***	-5.639***	-10.30***	-5.569***
	(1.704)	(1.267)	(1.703)	(1.552)	(1.540)	(1.271)
Observations	5059	5059	5059		5059	
Pseudo R^2	0.101	0.088	0.094		0.089	
AIC	359.7	463.2	822.3		826.5	
Log-likelihood	-169.8	-221.6	-391.1		-393.2	
Countries	124	124	124		124	
Total onsets	31	42	115		115	

Robust standard errors in parenthesis. Based on the Ethnic Power Relations dataset. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

fragmentation. This returned a 5 percent significance level also for *balance*, but the log-likelihood of models 3.2 and 3.4 dropped from -220 to -226. The most important conclusion to be drawn in terms of conflict over government is therefore that fragmentation seems to be much more important than balance to explain the onset of this type of civil war. The support for Hypothesis 4 is therefore limited.

In terms of the other variables in Table 3, *GDP per capita* has a higher and more significant effect on government conflicts than on territorial conflicts, while *population* turns out as highly significant (0,1 % level) for territorial conflicts, but is insignificant for government conflicts. Both of these results support the argument of relative rebel capability, as discussed above, since lower levels of GDP should increase the relative capability of a given ethnic group vis-à-vis the central

government. Conversely, an increase in the size of a population reduces the ability of the government to control the whole territory of the state. This is consistent with the findings of Buhaug (2006), even though he uses geographic size instead of population to measure the size of a country.

Another interesting feature of Table 3 is that *anocracy*, which was consistently significant at a 1 percent level in Table 2, appears to have lost much of its relevance. It is significant at a 5 percent level for territorial conflict, but not for conflicts over government. These results seem to be the reverse of what Buhaug (2006) finds in his tests. One possible explanation for this discrepancy could be that Buhaug includes both ethnic and non-ethnic onsets on the dependent variable, which in itself would alter the results. The fact that anocracy appears relevant when ethnic and non-ethnic government conflicts are tested together (as found by Buhaug) but not for ethnic government conflicts alone could suggest that anocracy primarily affects the non-ethnic onsets of such conflicts. This notion would, however, require further testing before a conclusion could be drawn.

5.6 Into democracy – a slight detour

The question of anocracy points to a related, though slightly different view of the effects of ethnic heterogeneity. As mentioned earlier, John Stuart Mill (1926) is famous for having subscribed to a rather sobering view of the prospects for establishing functioning democracy in multiethnic states. These notions are not the main focus of this thesis, but it may serve to exemplify how the two-dimensional approach to ethnic heterogeneity (i.e. fragmentation and balance) can facilitate testing of other relevant variables besides ethnic conflict onsets. If certain constellations of ethnic groups show signs of being particularly conducive to anocratic forms of government – or rather non-government – then such tendencies merits scrutiny. Theoretically, the most plausible effect of ethnic heterogeneity on anocracy would be that the risk of anocracy increases in proportion to *balance*, that is; balanced systems should be more anarchic and hence more anocratic since no group is in a clear position to exert control over the other groups. This does not automatically translate into higher risk of conflict. Rather, it suggests that stability will be maintained by other means,

TABLE 3b: Logistic regressions on form of government

	(1) Anocracy	(2) Democracy	(3) Autocracy
Fragmentation	-0.451	-0.830	0.715
	(0.595)	(0.903)	(0.646)
$Balance_{g2/g1}$	1.088**	0.150	-0.957*
	(0.345)	(0.579)	(0.432)
$Dominance_{c=.5}$	0.0618	-0.443	0.236
	(0.331)	(0.541)	(0.439)
$GDP\ per\ capita_{lag}$	-0.138***	0.314***	-0.223***
	(0.0311)	(0.0416)	(0.0395)
$Ln(Population)_{lag}$	-0.0987	0.138	-0.0173
	(0.0878)	(0.158)	(0.105)
$Oil\ production_{pc}$	0.0139	-0.203*	0.177*
	(0.0314)	(0.0930)	(0.0762)
Ln(Mountains)	0.250**	-0.102	-0.115
	(0.0810)	(0.110)	(0.0886)
Constant	-0.632	-2.868	0.999
	(0.777)	(1.492)	(0.962)
Observations	5767	5767	5767
pseudo R^2	0.096	0.310	0.147
AIC	5746.4	5099.4	6720.2
Log-likelihood	-2865.2	-2541.7	-3352.1
Countries	127	127	127

Robust standard errors in parenthesis. Based on the Ethnic Power Relations dataset. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. N is limited to 127 due to missing GDP data for one country.

such as through balance-of-power and deterrence. The analogy to the international arena is evident, underlined by the frequent description in IR literature, particularly in the realist tradition, of the international system as 'anarchic' (see e.g. Morgenthau 2006; Waltz 1964; Mearsheimer 2001; Glaser 2010).

In terms of fragmentation one could also expect that more groups would increase the difficulties of establishing a functioning government, be it democratic or autocratic. As such, one could assume that the system properties conducive to anocracy would be much the same as those leading to ethnic conflict; general actor uncertainty, ethnic security dilemmas and high coordination costs. In other words, anocracy could be expected to correlate positively with both *balance* and *fragmentation*, though not necessarily with *dominance*. By extending the logic of this argument one could imagine that democracy should be negatively correlated with the same variables, since countries with low *fragmentation* and low *balance* would effectively approach a state of ethnic homogeneity.

Table 3b shows how the ethnic heterogeneity variables fare in terms of explaining the outcomes of *anocracy*, *democracy* and *autocracy* respectively. Control variables related to political stability (*peaceyears* and *political instability*) are excluded in order to avoid problems of collinearity and endogeneity. The dependent variables are coded in binary form based on the Polity IV democracy index. Countries with scores between -5 and +5 are coded as anocratic, countries with scores above 6 are coded as democratic, while countries with scores from -6 and below are coded as autocratic. All variables are drawn from or calculated on the basis of the EPR dataset.

Among the ethnic variables, only *balance* appears to have any significant effect on the form of government, and as suspected the relationship between *balance* and *anocracy* is positive. Neither *dominance* nor *fragmentation* reveals any particular pattern. The results thus suggest that it is not, in fact, the number of groups that influences the form of government, but the balance between them: The more equal the two largest groups are in size, the more difficult it seems to be to establish a stable form of government (either autocracy or democracy).

That said, by far the strongest variable in all three models is *GDP per capita*, with consistent significance levels of 0,1 percent, though the direction of the relationship is positive only for democracy. These results are compelling, yet at the same time it could be argued that *GDP per capita* is as much the effect of the form of government as the cause. For that reason the models in Table 3b were also tested without *GDP per capita* among the control variables (not shown), and this made *fragmentation* turn out as significant (1% level) for both *democracy* and *autocracy*, negative for the former and positive for the latter.

The aim of Table 3b – or for that matter this thesis – is by no means to provide an exhaustive set of explanations for or determinants of the form of government in a given country. This slight detour is simply an attempt to suggest that ethnic heterogeneity can be expected to have an impact on a range of different social and political outcomes, and that such relationships could be better understood with the aid

of improved statistical tools for testing heterogeneity and of a proper classification of multiethnic states.

5.7 Some preliminary conclusions

With that in mind, we can return to the main question of the thesis; ethnic conflict. From the results of the tests in Tables 1, 2 and 3 some preliminary conclusions can be drawn regarding how different ethnic constellations affect the potential for conflict, and what this implies for the four hypotheses.

First, there is relatively strong support for Hypothesis 1 and the notion that fragmentation increases the risk of ethnic conflict. Linking this back to the theoretical foundation upon which the hypothesis is predicated the results suggest that more ethnic groups in a given country generally tends to; (i) increase the costs of coordination in the system, as the conventional ELF argument runs (c.f. Mauro 1995; Easterly & Levine 1997; La Porta et.al. 1999; Alesina et.al 1999, 2003); (ii) increase actor uncertainty and cause security dilemmas to arise, which in turn will be both countered and reinforced by strategies of deterrence and balance-of-power (c.f. Morgenthau 2006; Waltz 1964; Mearsheimer 2001), and; (iii) increase the centrifugal competition of the system and the probability of anti-system actors emerging (Sartori 1990).

The results thus seem to contradict both the theoretical notion of 'interaction opportunities' (Deutsch & Singer 1964) and the findings of Montalvo & Reynal-Querol (2005) who conclude that the risk of conflict is highest on low levels of fragmentation (i.e. high polarization). One possible explanation for this discrepancy could be that Montalvo & Reynal-Querol, in line with most other contributions in the literature, measure the effect on a dependent variable that comprises both ethnic and non-ethnic conflicts. Another reason why the results are contradicting could be that the datasets they use for testing the effect of ethnic heterogeneity, in their case the *World Christian Encyclopedia*, are coded one-dimensionally (as described in Chapter 4).

A second conclusion from the above tests is that there is only limited support for Hypothesis 2, which states that the risk ethnic conflict should be higher in dominant

systems than in either balanced or hegemonic systems. While stopping short of any viable statistical significance, the results in Table 2 suggest that the risk of ethnic conflict caused by dominance (c.f. Collier & Hoeffler 2004; Buhaug 2006; Cederman & Girardin 2007) is at its highest when the largest group is approximately double the size of the second largest group (c=.4 / c=.5). One example of such a constellation would be Afghanistan, where the Pashtun majority (41%) is about double the size of the second largest group, the Tajiks (25%).[9]

A noteworthy feature of these results is that, when combined, they seem to mirror the theoretical framework laid out by Mearsheimer (2001). As described above, Mearsheimer depicts *bipolarity* (i.e. few actors) as the most stable configuration; *balanced multipolarity* as less stable and; *unbalanced multipolarity* as the least stable. For what it's worth, then, these results also seem to provide empirical support to Mearsheimer's argument for the international arena. Such notions should be treated with a set of qualifications, however. As I have previously alluded to, the number of groups in a system does not directly translate into polarity, as understood by Mearsheimer. The bipolar structure of the Cold War was not a system of low fragmentation. Rather, it was a highly fragmented but relatively balanced system which split into two camps through a process of polarization – in Karl Deutsch's (1971) original meaning of the term. Bipolarity during the Cold War was thus the result of an 'economy of antipathy', as Horowitz (1985:182) labels it, which caused pairs of antagonists to emerge as comparative reference groups, despite the plurality of actors.

The third conclusion from the above is that dominance seems more important than fragmentation as a determinant of territorial conflict onsets. The opposite appears to be the case for government conflicts, where neither balance nor dominance returns any significant results once fragmentation is included. This provides general support for Hypothesis 3, but only limited support for Hypothesis 4.

[9] The percentages listed are drawn from the EPR dataset in the year 2000. See Appendix 1 for a full list of countries in the dataset.

Linking this back to the graphic model presented in Figure 2, it does indeed appear that the risk of conflict is highest at the point where *majority control* and *minority capability* intersect, though this is limited to territorial conflicts. Theoretically this makes sense when it is combined with the notion of relative rebel capability, as described by Buhaug (2006). On the other hand, the assumption that conflicts arising from dominance and minority exclusion should be highest in countries with low fragmentation (Figure 2b) is not supported.

If we return to the examples mentioned in the introduction, it then appears that both Sri Lanka and Kenya may be typical cases, though for different types of conflicts. Sri Lanka, with a fragmentation value of .18 and a balance value of approximately .2, is among the countries with a low risk of government conflict, but a relatively high risk of territorial conflict. [10] Kenya, for its part, has a fragmentation score of .58, which makes it comparatively prone to government conflict.[11] On the other hand, Kenya also has a balance score of .55, suggesting that it may in fact also be prone to territorial conflict. This, however, would depend on a large set of other factors that contribute to the outbreak of civil wars. While Sri Lanka and Kenya may be typical, therefore, they also serve to underline that the overall explanatory power of ethnic heterogeneity for the onset of civil war remains limited. It is by no means a sufficient cause in itself. I will return to this below when discussing probabilities.

5.8 Robustness

None of the above tests have aimed to explain the onset of all civil wars (i.e. non-ethnic), yet a test across a set of different dependent variables (both ethnic and non-ethnic) may provide an indication as to the robustness of the results obtained, and in particular the robustness of the *fragmentation* variable. In Table 4, *dominance* and *fragmentation* are tested on a range of different dependent variables, the first three of

[10] Sri Lanka is listed with four relevant ethnic groups in the EPR dataset (for the year 2000):
Sinhalese (70%), Indian Lankan Tamils (12%), Sri Lankan Tamils (11%) and Moors/Muslims (6%).

[11] Kenya is listed with seven relevant ethnic groups in the EPR dataset (for the year 2000):
Kikuyu-Meru-Emb (27%), Kalenjin-Masai-Turkana-Samburu (15%), Luhya (14%), Luo (12%), Kamba (11%), Kisii (6%) and Mijikenda (5%).

TABLE 4: Robustness across different conflict datasets

	Ethnic onsets			All onsets			
	(4.1) EPR	(4.2) EPR high	(4.3) FL	(4.4) EPR	(4.5) $PRIO_5$	(4.6) FL	(4.7) Sambanis
Fragmentation	1.964**	2.019*	2.295***	1.206*	1.157***	1.329*	0.911
	(0.711)	(0.994)	(0.572)	(0.537)	(0.334)	(0.593)	(0.584)
Dominance $_{c=.5}$	0.716	0.834	0.158	-0.00531	-0.0592	-0.0524	-0.132
	(0.458)	(0.572)	(0.304)	(0.298)	(0.242)	(0.339)	(0.311)
Peace years	0.00618	-0.0163	-0.00676	-0.00573	0.00279	-0.0436**	-0.0188
	(0.00981)	(0.0177)	(0.0135)	(0.00771)	(0.00752)	(0.0147)	(0.0110)
GDP per capita $_{lag}$	-0.0987*	-0.130	-0.109*	-0.118***	-0.0724**	-0.102**	-0.127**
	(0.0423)	(0.0748)	(0.0536)	(0.0318)	(0.0268)	(0.0381)	(0.0494)
Ln(Population) $_{lag}$	0.254*	0.222	0.370***	0.141*	0.233***	0.340***	0.0601
	(0.0992)	(0.117)	(0.0799)	(0.0709)	(0.0637)	(0.0897)	(0.0634)
Oil production $_{pc}$	0.0195	0.0336	0.00496	0.0370**	0.0234*	-0.00197	0.0264
	(0.0216)	(0.0350)	(0.0384)	(0.0123)	(0.0116)	(0.0445)	(0.0368)
Anocracy $_{t=1}$	0.724**	0.771*	1.181***	0.555**	0.552**	0.796**	0.532*
	(0.261)	(0.367)	(0.259)	(0.181)	(0.174)	(0.249)	(0.222)
Political instability	0.248	0.341	-0.0644	0.531**	0.184	0.238	0.253
	(0.290)	(0.405)	(0.332)	(0.194)	(0.169)	(0.272)	(0.244)
Ln(Mountains)	0.0404	0.300*	0.00981	0.0561	0.142*	0.119	0.115
	(0.109)	(0.138)	(0.115)	(0.0618)	(0.0630)	(0.0969)	(0.0804)
Constant	-7.628***	-8.535***	-8.668***	-4.885***	-6.252***	-7.010***	-4.671***
	(1.087)	(1.360)	(0.881)	(0.682)	(0.655)	(0.815)	(0.669)
Observations	5129	5353	5111	4880	5828	4359	5828
Pseudo R^2	0.082	0.105	0.115	0.064	0.051	0.114	0.056
AIC	723.0	424.5	620.9	1280.6	1498.7	816.8	928.3
Log-likelihood	-351.5	-202.2	-300.4	-630.3	-739.3	-398.4	-454.1

Robust standard errors in parenthesis. Based on the Ethnic Power Relations dataset. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
EPR: ethnic power relations; EPR high: same with 1000 deaths threshold; FL: Fearon & Laitin (2003); $PRIO_5$: UCDP/PRIO ACD (5 year peace criterion); Sambanis: Sambanis (2004).

which are ethnically defined, while the last four include all civil war onsets. In terms of *fragmentation*, the results are quite convincing, with significant results for nearly all the models. The effect of *dominance*, on the other hand, remains insignificant throughout.

When introducing the *fragmentation* measure, I suggested that the manner in which it was constructed should allow for easier testing across different sources of ethnic groups since it only takes into account the share of the two largest groups and compares these to the rest of the population. Table 5 shows how fragmentation effectively remains significant across various different ethnic datasets, ranging from the infamous ANM to the web-based *Joshua Project*,[12] which contains detailed information on ethnic groups in just about every country in the world. The *World Christian Encyclopedia* (WCE) is split into two categories (models 5.2 and 5.3), one

[12] See http://www.joshuaproject.net/

TABLE 5: Robustness across different ethnic datasets

	(5.1) EPR	(5.2) WCE	(5.3) WCE culture	(5.4) ANM	(5.5) JOSHUA
Fragmentation	1.939**	1.884**	2.813***	1.428	1.281*
	(0.708)	(0.669)	(0.768)	(0.909)	(0.628)
Dominance $_{c=.5}$	0.681	-0.316	-0.124	1.216**	0.428
	(0.462)	(0.517)	(0.540)	(0.416)	(0.434)
Peace years	0.00590	0.00370	0.00587	0.00885	0.00171
	(0.00979)	(0.00960)	(0.00966)	(0.0109)	(0.00962)
GDP per capita $_{lag}$	-0.0990*	-0.105**	-0.125**	-0.101*	-0.109**
	(0.0422)	(0.0377)	(0.0411)	(0.0409)	(0.0377)
Ln(Population) $_{lag}$	0.250*	0.268**	0.214*	0.269**	0.287**
	(0.0991)	(0.102)	(0.0944)	(0.0993)	(0.0994)
Oil production $_{pc}$	0.0194	0.00762	0.00948	0.0168*	0.0102
	(0.0216)	(0.00898)	(0.00779)	(0.00764)	(0.00837)
Anocracy $_{t=1}$	0.730**	0.718**	0.710*	0.679*	0.726**
	(0.261)	(0.273)	(0.280)	(0.293)	(0.269)
Political instability	0.248	0.211	0.242	0.262	0.192
	(0.289)	(0.297)	(0.286)	(0.291)	(0.288)
Ln(Mountains)	0.0376	0.0572	0.0286	0.0904	0.00572
	(0.109)	(0.114)	(0.110)	(0.116)	(0.105)
Constant	-7.551***	-7.487***	-6.907***	-8.114***	-7.645***
	(1.095)	(1.086)	(0.910)	(1.137)	(0.994)
Observations	5059	6036	6036	5605	6063
Pseudo R^2	0.081	0.080	0.082	0.106	0.081
AIC	722.2	738.4	737.2	669.5	746.5
Log-likelihood	-351.1	-359.2	-358.6	-324.8	-363.3

Robust standard errors in parenthesis. EPR: Ethnic Power Relations, WCE: World Christian Encyclopedia (2009), ANM: Atlas Narodov Mira, JOSHUA: The Joshua Project. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

for the name of the different ethnic group (labeled *autoglossonym* in the dataset), and one for the culture to which these belong. The latter provides for a more generalized level with more groups placed under each culture, and it actually yields better results than the detailed level of the autoglossonym, with a higher coefficient and .1 percent significance. The data for the WCE is pulled from the online version of the encyclopedia and thus contains more updated information than the original from 1982 used by Montalvo & Reynal-Querol (2005).[13]

Atlas Narodov Mira is actually the only dataset in which the fragmentation variable does not reach 5 percent significance.[14] Interestingly, though, *dominance* becomes significant instead. This could perhaps be explained by the fact that the number of observations changes slightly, but it also underlines how sensitive statistical tests of

[13] See http://www.worldchristiandatabase.org/wcd/

[14] Data from the Atlas Narodov Mira is drawn from Weidmann et.al. (2010), whose project on the Geo-referencing of Ethnic Groups is based on ANM data. See http://www.icr.ethz.ch/research/greg.

ethnic heterogeneity are to the actual coding of the groups, despite the use of relatively flexible measures like *fragmentation*. Figures 7 and 8 illustrate the profoundness of this challenge by comparing the distributions of *fragmentation* and *fractionalization* for the ANM and EPR datasets. For *fractionalization* (Figure 7) – which is quite sensitive to individual coding decisions as it is calculated on the basis of all the groups in a given country – the discrepancy between the two datasets is surprisingly wide. The correlation is hardly perfect for *fragmentation* either (Figure 8), but it is possible to discern why this measure yields the best results across different datasets. Concretely, the correlation between the variables in the two datasets is .5572 for *fractionalization* and .7160 for *fragmentation*.

Figure 7: Fractionalization for EPR and ANM

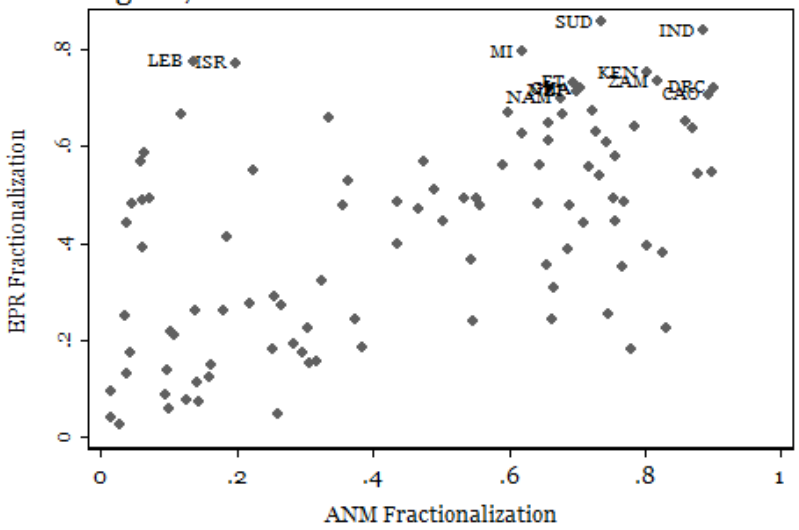
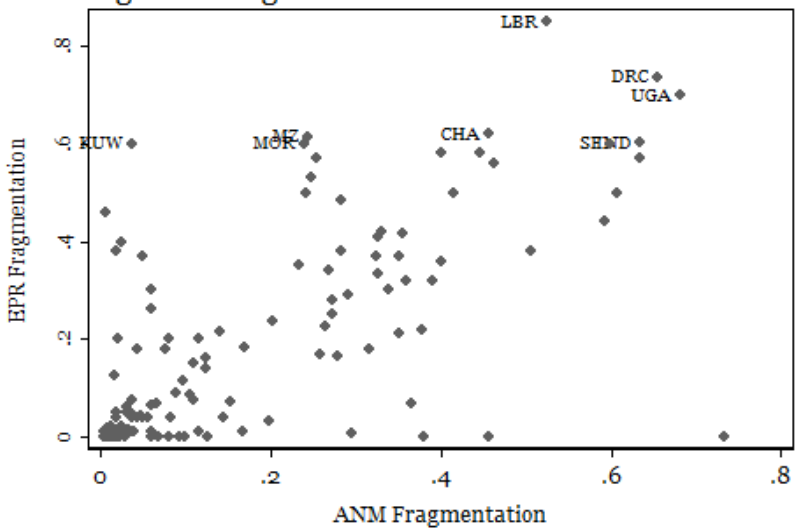


Figure 8: Fragmentation for EPR and ANM



5.9 In terms of probabilities

While the above findings provide some indications as to the direction and significance of the different variables that affect the risk of ethnic conflict, it does not supply much insight into to the actual change in probability. The pseudo R^2 tends to remain below 0.1 in all the tests, which is far from impressive. Moreover, the effect of both *fragmentation* and *dominance* are unlikely to be completely linear, and thus have the same effect on all levels. In Figures 9-12, the probability of territorial and government conflict is calculated for *fragmentation* and *dominance/balance*. The coefficients for these tests are taken from models 3.1 and 3.2.

Figure 9: Probability of Territorial Conflict

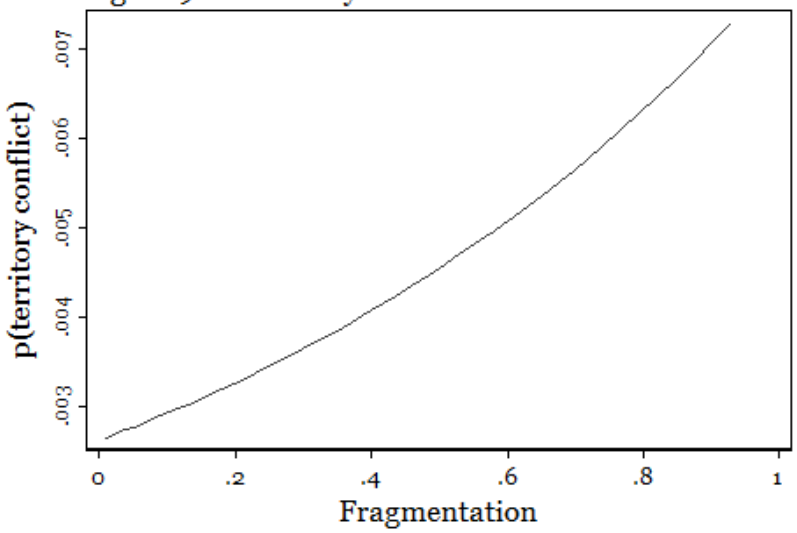


Figure 10: Probability of Territorial Conflict

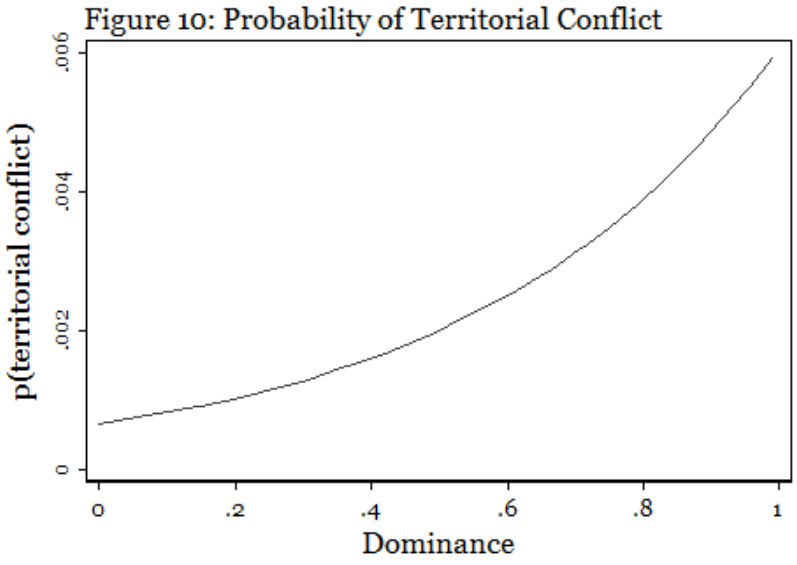


Figure 11: Probability of Government Conflict

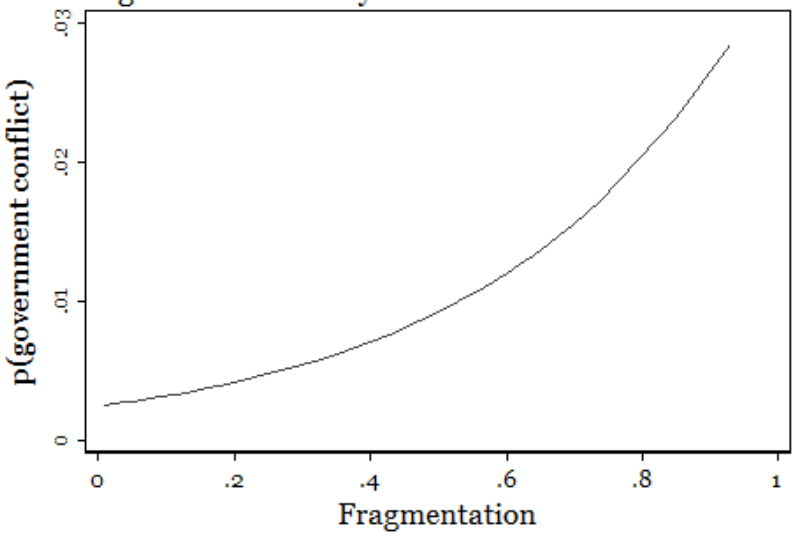


Figure 12: Probability of Government Conflict

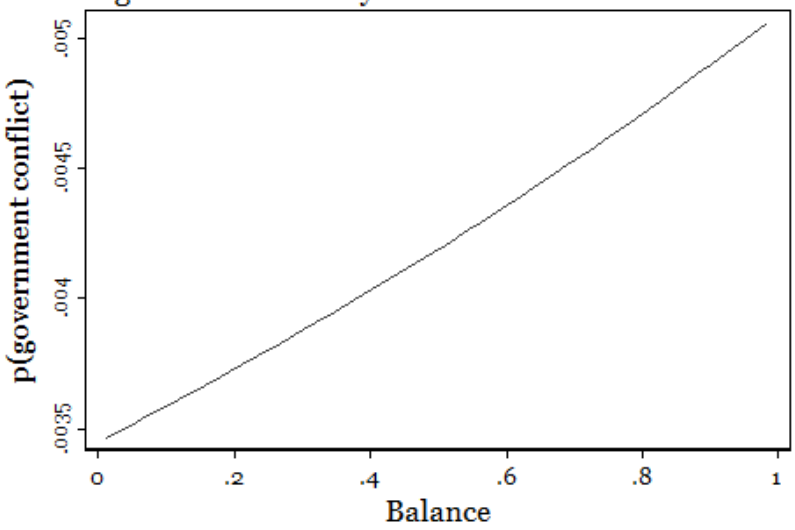


Figure 9 illustrates how increasing levels of fragmentation affects the probability of territorial conflict. When going from minimum to maximum levels of *fragmentation*, the probability increases from approximately .0025 to around .007 – an effective increase of nearly 300 percent. For *dominance* (Figure 10), the corresponding numbers are from below .001 to about .006, an increase of about 600 percent. The actual probability, however, remains below 1 percent, and is similar on maximum levels of both *fragmentation* and *dominance (.007 and .006).*

For conflict over government, the probability appears to be considerably higher in real terms. As Figure 11 shows, the probability of government conflict increases from

approximately .0025 to about .025 when going from the lowest to the highest value of *fragmentation*. This is effectively a near ten-fold increase. On the highest levels of *fragmentation*, this means that in real terms the probability of government conflict reaches about 2.5 percent. The effect of *balance* on government conflict (Figure 12) is much more limited, increasing from .0035 to a mere .005 when moving from minimum to maximum values.

These are of course extreme values, making the probability axes slightly misleading. Very few countries are actually listed with *fragmentation* values much over .6. In fact, Uganda, Liberia and the DRC would be the only three. There is a certain logic to this, of course, since the two largest groups taken together only rarely would constitute less than 40 percent the total population. But quite a few countries are indeed situated on the borderline with fragmentation values of approximately .6, prompting the question of whether a threshold could be set for classifying the countries into main categories.

5.10 Towards a classification of multiethnic states

Based on the theoretical discussions and the empirical indications in this thesis it should now be possible to draw the outline of a classification for multiethnic states. The core of this classification would be the two main dimensions of balance and fragmentation. In order to create meaningful 'data containers', as Sartori (1970) calls it, each of these dimensions further require thresholds that places the countries into comparable classes. With two thresholds for each dimension, this would create nine classes of multiethnic states.

For the balance dimension, the most logical values for these thresholds would be at $B=.25$ and $B=.75$, as discussed during the operationalization of Hypothesis 2 (Chapter 3). If the value of B is below .25, the country would then be classified as 'hegemonic'. If the value is above .75, it could be classified as 'balanced'. Everything in between then falls into the 'dominant' class. This is based on the notion that the peak point of dominance is placed where majority intersects with minority capability (c.f. Figure 2), and that the risk of conflict from dominance decreases as it approaches either edge of

the model. This assumption has also been underpinned by the results obtained in the above regressions, in particular that of Table 3.

For the fragmentation dimension, the issue becomes a bit more complicated. Both logically and theoretically, the easiest approach would be to adopt the categories used by Sartori (1990) for political parties. This would set the thresholds between 2 and 3 groups, and between 5 and 6 groups. His contention is that three parties represent a fundamentally different interaction pattern than two parties, and that systems with more than five parties "tend to produce a different mechanics than the interactions among five-or-less parties". At the same time, Sartori willingly notes that these thresholds are not absolutes and that the "borderline is not *at* five (or at six), but *around* five (or six)" (Sartori 1990:329, original italics). He further asserts that whether a system should be labeled *limited* or *extreme* pluralism eventually depends more on the ideological distance in the landscape than on the exact number of groups.

Since ideological distance – which could also be understood as *degree of polarization* – is a relatively volatile property, it does not conform to the criteria for classification adhered to in this thesis. We are thus left with the *a priori* estimates provided by Sartori (1990), which yields three main categories; (i) dualism (two groups); (ii) limited pluralism (3-5 groups), and; (iii) extreme pluralism (6 groups or more). With two thresholds for each of the two dimensions, we can then attempt to place the countries of the EPR dataset into the framework. A distribution of the results is shown in Figure 13. The thresholds dividing the different classes are marked with dashed lines.

While there is a clear conceptual advantage in using groups as the threshold marker (Figure 13), a question looms as to whether the numerical dimension of the classification should ultimately be based on the *fragmentation* measure or on the number of *groups*. There are sufficient arguments for favoring either approach. The *fragmentation* measure has the frequently mentioned benefit of easing the task of coding. Moreover, it provides an indication of the size of the groups without falling

hostage to the need for detailed information on all the relevant minorities. Equally important, *fragmentation* appears to return the most robust statistical results.

Figure 13: Distribution by Balance & Groups

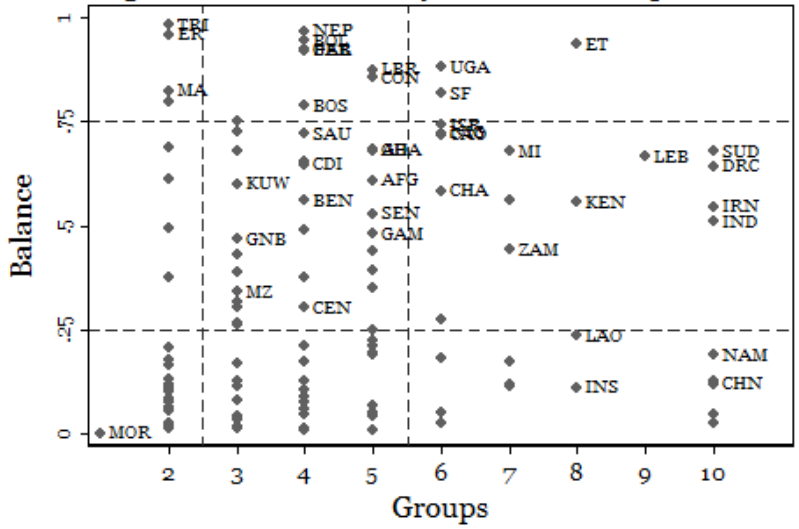
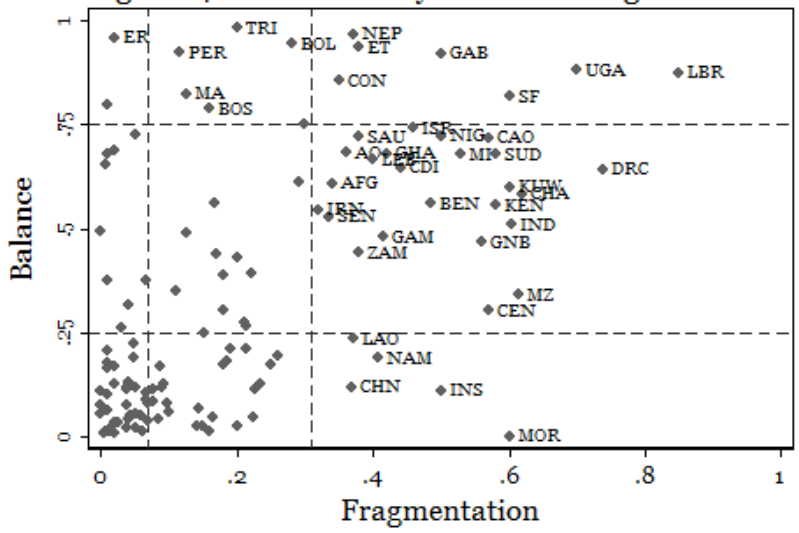


Figure 14: Distribution by Balance & Fragmentation



As for the *groups* variable there is an obvious benefit in the simple interpretation of the measure: Four relevant groups mean four relevant groups. On the other hand, there is no way of knowing whether these four groups are similar in size, or whether some are more relevant than others. Where *fragmentation* takes into consideration the size of

the two largest actors, the *groups* variable simply counts them. The effect is that countries with very similar properties may end up in different categories when only the groups are counted. For instance, Peru is listed with four relevant groups in the EPR dataset, and the two largest groups – the Quechua (46 %) and the so-called Peruvians (42%) – together constitute close to 90 percent of the population. This places it in the class of *balanced limited pluralism* in Figure 13 (top center). Trinidad & Tobago, on the other hand, is listed with only two relevant groups, even though the two largest groups constitute 40 percent each. In other words, 20 percent of the population remains unaccounted for. The *fragmentation* score for Trinidad & Tobago is therefore higher than for Peru, yet there are only two relevant groups, compared to four in Peru. Consequently, when countries are classified by the number of relevant *groups* instead of by *fragmentation*, countries with quite similar properties may easily end up in different categories. If the *fragmentation* had been used instead of *groups* in the framework, Trinidad & Tobago and Peru would both have ended up in the class of *balanced limited pluralism*.

The main drawback of using the *fragmentation* measure instead of *groups* is that the interpretation of the measure is much less intuitive. Theoretically this poses a challenge, and the most plausible solution may be to transform Sartori's thresholds for groups into *fragmentation* scores. This is attempted in Figure 14 where *fragmentation* is replaced by the number of groups on the x-axis. The thresholds in this distribution have been calculated from the mean value of the *fragmentation* scores for the countries with 2-3 groups (.071) and 5-6 groups (.314) respectively. It is evidently possible to devise more cunning manners of setting these thresholds, but I will reserve such exploratory exercises for future projects.

When comparing Figures 13 and 14, it becomes evident that there are considerable differences in the way *groups* and *fragmentation* classifies countries. A country like Gambia ends up in the category of *balanced limited pluralism* in Figure 12 since it is listed with only five relevant groups, while in Figure 14 it falls in the category of *balanced extreme pluralism* since the *fragmentation* value is higher than the number of groups suggests. By contrast, China is listed with a total of 47 relevant groups in the

EPR dataset, yet the *fragmentation* value is only .37 since most of these groups are relatively small compared to the Han majority.[15]

Tables 6 and 7 show the frequency distribution of Figures 13 and 14. The number of relevant groups is used in Table 6 and fragmentation is used in Table 7. The thresholds on the balance dimension are the same for both tables. The main difference between the two is that Table 7 provides a more even distribution between the three classes of fragmentation; dualism, limited pluralism and extreme pluralism. The threshold values are the same as for Figure 14 and are noted in the column headings.

These differences may be marginal, but the point I strive to make is that the criteria for classification of multiethnic states should not just be guided by theoretical assumptions. They should also be judged by their ability to group countries with similar attributes into similar categories. The question of classification is thus also an empirical question. A precondition should be that the thresholds have some substantive meaning. Sartori (1970:1039) points to this when he suggests that the interaction

Table 6: Frequency distribution by *balance* and *groups*

	Dualism (2 groups)	*Limited Pluralism (3-5 groups)*	*Extreme Pluralism (6- groups)*	*N*
Balanced	4	10	3	17
Dominant	4	26	15	45
Hegemonic	21	31	13	65
N	**29**	**67**	**31**	*127*

Table 7: Frequency distribution by *balance* and *fragmentation*

	Dualism (F<.071)	*Limited Pluralism (F>=.071)*	*Extreme Pluralism (F>=.314)*	*N*
Balanced	2	7	8	17
Dominant	10	11	24	45
Hegemonic	35	25	5	65
N	**47**	**43**	**37**	*127*

[15] Note that in Figure 13, a ceiling has been set on the number of groups so that countries with more than 10 groups (such as China) are listed with only 10.

between three to five groups is likely to be *substantively* different from the interaction between six of more groups.

In practical terms, one could ask whether the countries classified under *balanced limited pluralism* share a particular political dynamic that distinguishes these from the other countries in the sample. If so, then the classification of these into one category is useful as it allows for both intra-class and inter-class comparisons. For such is the fundamental rationale of classification; to facilitate the exercise of comparison.

This points to a final check of whether the number of groups or the fragmentation measure should be used to classify the countries along the numerical dimension of the framework, namely whether the properties of the groups in the different classes are indeed similar and comparable. A thorough analysis of this goes beyond the scope of this thesis, but a brief look at one of the classes could prove useful.

Table 8: Countries classified as Balanced Limited Pluralism – by *groups*

Country	Bal.	Gr.	Frag.	Group 1	%	Group 2	%
Bosnia-Herzegovina	.79	4	.16	Bosniaks/Muslims	47	Serbs	37.1
Peru	.92	4	.12	Quechua	46	Peruvians	42.5
Mauritania	.75	3	.30	Haratins (Black Moors)	40	Black Africans	30
Guinea	.75	3	.30	Peul	40	Malinke	30
Bolivia	.95	4	.28	Quechua	37	Bolivians	35
Congo	.86	5	.35	Nibolek (Bembe etc.)	35	Lari/Bakongo	30
Nepal	.97	4	.37	Hill Brahmins/Chetri	32	Madhesi	31
Gabon	.92	4	.50	Fang	26	Eshira/Bapounou	24
Liberia	.88	5	.85	Gio	8	Mano	7

(Source: EPR, classification by *balance* and relevant *groups*)

Table 8 lists the countries that fall under the category of *balanced limited pluralism* when using the *groups* variable as the selection criterion for the numerical dimension. The countries are placed in descending order by the size of the largest group. The general similarity between most of the countries on the list testifies to the comparative potential inherent in the classification. The countries are highly balanced, with the two largest groups making up between 50 and 88 percent of the total population. Yet, there is one peculiar exception: Liberia.

With the largest group (Gio) constituting only 8 percent of the population, one wonders whether the mechanics of the Liberian ethnic relations are at all comparable to those of Bosnia-Herzegovina, where the largest group constitutes 47 percent. What they have in common is a similar number of relevant groups and a relatively high balance score. But it makes little sense to classify the Liberian case as *limited pluralism*. One could argue, of course, that the reason why Liberia has been included in this category stems from the failure to code all the relevant groups in Liberia. There should evidently have been more of them in order to fill up the full share of the population. But this misses the point. It might very well be that Liberia has only five politically relevant ethnic groups. Yet, if the vast majority of the population in a country remains unaccounted for, the result is that the number of groups ceases to function as a viable criterion for comparison.

Table 9: Countries classified as Balanced Limited Pluralism – by *fragmentation*

Country	**Bal.**	**Gr.**	**Frag.**	**Group 1**	**%**	**Group 2**	**%**
Madagascar	.82	2	.13	Côtiers	48	Highlanders	40
Bosnia-Herzegovina	.79	4	.16	Bosniaks/Muslims	47	Serbs	37
Peru	.92	4	.12	Quechua	46	Peruvians	43
Trinidad and Tobago	.98	2	.20	East Indians	40	Blacks	40
Mauritania	.75	3	.30	Haratins (Black Moors)	40	Black Africans	30
Guinea	.75	3	.30	Peul	40	Malinke	30
Bolivia	.95	4	.28	Quechua	37	Bolivians	35

(Source: EPR, classification by *balance* and *fragmentation*)

In Table 9, the same exercise is conducted with *fragmentation* as the numerical criterion. This reduces the number of countries in the class to seven, but in return it escapes the problem of outliers. For easy reading and comparison, the countries that appear in bold are present in both Tables 8 and 9. Note that two of the countries in Table 9 have only two relevant groups, yet together these constitute less than 90 percent of the population. The range in the size of the largest group is also smaller in Table 9 than in Table 8, with a mere 11 percentage point span (compared to 21 in Table 8 when Liberia is excluded).

These brief examples suggest that alongside *balance*, *fragmentation* may be the most viable numerical criterion for the classification of multiethnic states, as it categorizes countries into clusters of relative similarity. This may all in all outweigh the drawback related to the interpretation of the measure.

That said, the eventual utility and applicability of such a classification of multiethnic states will only become apparent once more profound studies into the internal dynamics of countries in the different classes have been undertaken. Until then, the thresholds and criteria for classification will remain indicative.

6 Concluding Remarks

Over the last two decades ethnically motivated civil wars have become a growing concern for researchers as well as for politicians. As the opening quote of this paper suggests, "[t]he importance of ethnic conflict, as a force shaping human affairs, as a phenomenon to be understood, as a threat to be controlled, can no longer be denied" (Horowitz 1985:XV). Yet, despite the frequency and devastation caused by such conflicts, we still know remarkably little about the root causes and mechanisms that lie at the heart of the problem. In the large N literature, considerable effort has been put into identifying factors that might increase the risk of conflict. Some of these contributions have focused on material factors such as economic growth, natural resources or even topography, while others have drawn attention to the cultural and social causes of civil war The focus of this thesis has firmly been placed within the latter strand of scientific discourse, as I have stepped into the debate on ethnic heterogeneity and its effects on the risk of ethnic conflict.

The core research question of the thesis has been whether some multiethnic configurations are more prone to conflict than others. This approach is largely motivated by the frequent notions of inevitability that proliferate during discussions on the causes of ethnic conflict. Sri Lanka, Rwanda and Kenya served as preliminary examples of countries where dominance or fragmentation appears to make ethnic strife predestined.

With this motivational backdrop I started by providing an overview of the previous literature on the subject of ethnic heterogeneity. Three different arguments were identified: (i) that the risk of ethnic conflict increases with the number of groups; (ii) that the risk of ethnic conflict is highest with only two large groups, and; (iii) that the risk of conflict is highest when one group dominates, but not overwhelmingly so. I then suggested that these arguments, as well as the landscape of ethnic heterogeneity, is better comprehended if mapped along two main dimensions; one capturing the

number of groups (fragmentation) and the other capturing the balance between the groups.

I then went on to elaborate the theoretical framework upon which these arguments rest, drawing largely on IR theory and political science. This discussion subsequently formed the basis for four main hypotheses; (i) that fragmentation increases the risk of conflict; (ii) that dominance increases the risk of conflict; (iii) that territorial conflict is most likely in dominant systems, and; (iv) that government conflict is most likely in balanced systems.

In Chapter 3 the operationalization of these hypotheses was discussed and a set of different measure introduced. Some were drawn from previous research while others were introduced as new alternative measures. In particular, I proposed a new measure of *fragmentation* that I argued should be more robust than previously applied measures like *fractionalization* and *polarization*. I further introduced a measure of *dominance* that uses the relative size of the two largest groups to identify the balance point between them at which the risk of conflict is at its highest.

Chapter 4 dealt with some of the methodological issues related to the study of ethnic heterogeneity and conflict, roughly categorized as input problems and output problems. In these discussions I reiterated some of the criticism advanced against previously used datasets, and suggested that the recently compiled EPR datasets by Wimmer et.al (2009) is the most reliable source of data for both ethnic groups and ethnic conflict onsets available.

In Chapter 5 the results from the testing of the different hypotheses were presented and discussed. I found considerable support for the conventional argument that ethnically fragmented countries are more prone to conflict than countries with low fragmentation, but less support for the second hypothesis which states that dominance increases the risk of ethnic conflict. When the dependent variable was disaggregated into territorial and government conflicts, I found that dominance is a significant determinant for territorial conflict but not for conflicts over government. The reverse appears to be the case for fragmentation. This translates into very strong support for Hypothesis 1, only

limited support for Hypothesis 2, relatively strong support for Hypothesis 3, while Hypothesis 4 should be rejected.

These findings contribute towards painting a more nuanced picture of how ethnic heterogeneity affects the risk of civil war. At the same time, it is important to recognize that as long as the research on ethnic heterogeneity is limited to identifying patterns of increasing and falling risk, the substantive implications in terms of peace-building will remain limited. When Kenyans went to the polls on August 5, 2010 to decide on a new constitution, it would have made little difference to know that the country belongs to a class of multiethnic states labeled *dominant extreme pluralism*, and that the risk of conflict therefore should be high. If the research on ethnic conflict is to have any substantive impact on policy, the agenda needs to go beyond the mere mapping of tendencies and towards a better understanding of how the number of groups and their relative size affects the political dynamics and social interaction in a given country. Such insight might very well require more qualitatively and comparatively oriented studies. In the words of Donald Horowitz (1985:13); "the need for comparative analysis is compelling".

With this acknowledgement in mind I have suggested that the two dimensions of balance and fragmentation could form the basis for a classification of multiethnic states that, in turn, could facilitate more comparatively oriented studies. However, the strength and utility of this classification will only become evident once more qualitatively oriented studies have been undertaken, by which the comparative potential of the various categories can be mapped. Until then, the classification should be treated as a preliminary draft – as analytical scaffolding by which the research on ethnic heterogeneity can be further developed. In the meantime, the statistical research on ethnic heterogeneity can give us some indications as to the configurations of ethnic groups that are more and less prone to ethnic conflicts. The findings in this paper suggest that dominant systems are slightly more conflict-prone that hegemonic and balanced systems, but the strongest tendency identified is that high levels of fragmentation increases the risk of conflict – or in lay terms; the more the messier.

7 References

Alesina, A., Baqir, R. & Easterly, W. (1999) “Public goods and ethnic division”, *Quarterly Journal of Economics*, 114, pp 1243-84.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. & Wacziarg, R. (2003) “Fractionalization”, *Journal of Economic Growth*, Vol. 8, No. 2, pp 155–194.

BBC News (2010a) "After the Tigers, a new Tamil struggle", Published: April 5, 2010, at: http://news.bbc.co.uk/go/pr/fr/-/2/hi/south_asia/8603658.stm

BBC News (2010b) "Fear and shock as ethnic Uzbeks flee Kyrgyz clashes", Published June 13, 2010, at: http://www.bbc.co.uk/news/10303838

Buhaug, H. (2006) “Relative Capability and Rebel Objective in Civil War", *Journal of Peace Research*, vol. 43, no. 6, 2006, pp. 691–708.

Bull, H. (1977) *The anarchical society: a study of order in world politics*, Columbia University Press.

Cederman, L. & Girardin, L. (2007) "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies", *American Political Science Review*, 101:1:173-185.

Collier, P & Hoeffler, A. (1998) “On Economic Causes of Civil War”, *Oxford Economic Papers* 50(4): 563–573.

Collier, P & Hoeffler, A. (2004) “Greed and Grievance in Civil War”, *Oxford Economic Papers* 56(4): 563–595.

Deutsch, K. & Singer, D. (1964) "Multipolar Power Systems and International Stability", *World Politics*, Vol. 16, No. 3 (Apr., 1964), pp. 390-406.

Deutsch, M. (1971) "Conflict and its Resolution," in *Conflict Resolution: Contributions of the Behavioral Sciences*, ed. by C. G. Smith. Notre Dame: University of Notre Dame Press.

Doyle, M. & Sambanis, M. (2000) “International Peacebuilding: A Theoretical and Quantitative Analysis”, *American Political Science Review*, Vol. 94, No. 4 (December, 2000), pp. 779-801.

Dunne, T. & Schmidt, B.C. (2005) ”Realims”, in Baylis, J. & Smith, S. (eds.) *The Globalization of World Politics,* 3rd ed, Oxford University Press, pp 161-184.

Easterly, W. & Levine, R. (1997) “Africa’s Growth Tragedy: Policies and Ethnic Divisions”, *Quarterly Journal of Economics* 112(3): 1203–1250.

Ellingsen, T. (2000) "Colorful Community or Ethnic Witches' Brew?: Multiethnicity and Domestic Conflict during and after the Cold War", *Journal of Conflict Resolution*, April 2000, 44: 228-249.

Eriksen, T.H. (1993) *Ethnicity and Nationalism. Anthropological Perspectives,* Pluto Press.

Esteban, J. & Ray, D. (1994) “On the Measurement of Polarization”, *Econometrica,* 62(4): 819–851.

Esteban, J. & Ray, D. (1999) ‘Conflict and Distribution’, *Journal of Economic Theory* 87(2): 379–415.

Esteban, J. & Ray, D. (2008), "Polarization, Fractionalization and Conflict", *Journal of Peace Research,* Vol. 45, No. 163.

Esteban, J. & Schneider, G. (2008) "Polarization and Conflict: Theoretical and Empirical Issues", *Journal of Peace Research*, Vol. 45, No. 131.

Fearon, J.D. & Laitin, D.D. (2003) “Ethnicity, Insurgency, and Civil War”, *American Political Science Review,* Vol. 97, No. 1 (Feb., 2003), pp. 75-90.

Geller, D.S. & Singer, J.D. (1998) *Nations at War: A Scientific Study of International Conflict*, Cambridge, MA: Cambridge University Press.

Glaser, C.L. (2010) “Realism”, in Collins, A. (ed) *Contemporary Security Studies*, 2nd ed., Oxford University Press.

Gleditsch, N.P., Wallensteen, P., Erikson, M., Sollenberg, M & Strand, H. (2002) "Armed conflict 1946-2001: A new dataset." *Journal of Peace Research*, Vol. 39, pp 615-637, version 4-2009.

Hanson, S. (2008) "Understanding Kenya's Politics", Council on Foreign Relations, Published: January 25, 2008, at: http://www.cfr.org/publication/15322/understanding_kenyas_politics.html

Hegre, H. & Sambanis, N. (2006) "Sensitivity Analysis of Empirical Results on Civil War Onset." *Journal of Conflict Resolution*, 50:508–35.

Horowitz, D. (2000) *The deadly ethnic riot*, University of California Press.

Horowitz, D. (1985) *Ethnic Groups in Conflict*, 2nd ed. (2000), University of California Press.

Jenkins, R. (1997) *Rethinking ethnicity*, Thousand Oaks, CA: Sage.

Kissinger, H. (1994) *Diplomacy*, Simon & Schuster Paperbacks, New York.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. & Vishny, R. (1999) "The Quality of Government," *Journal of Law, Economics and Organization*, Vol. 15, No. 1, pp. 222-279.

Laitin, D. & Posner, D. (2001) "The Implications of Constructivism for Constructing Ethnic Fractionalization Indices", *Comparative Politics Newsletter*, 12 (Winter 2001).

Lijphart, A. (1977) *Democracy in Plural Societies. A Comparative Exploration*, Yale University Press.

Mathew, G. (2005) "India (Republic of India)", in Griffiths (ed.*) Handbook of Federal Countries, 2005*, McGill-QUeen's University Press, pp.165-182.

Mauro, P. (1995) "Corruption and Growth", *The Quarterly Journal of Economics* 110:681-712.

Mearsheimer, J.J. (2001) *The Tragedy of Great Power Politics,* New York: Norton.

Mill, J.S. (1861, 1926) *Considerations on Representative Government,* Logmans, Green and Co. Ltd., London.

Montalvo, J G & Reynal-Querol, M. (2002) "Why ethnic fractionalization? Polarization, ethnic conflict and growth", UPF Economics and Business Working Paper No. 660.

Montalvo, J G & Reynal-Querol, M. (2005), "Ethnic Polarization, Potential Conflict, and Civil Wars", *American Economic Review*, Vol. 95, No. 3 (Jun., 2005), pp. 796-816.

Morgenthau, H.J. (1948, 2006) *Politics Among Nations,* 7th ed, revised by Thompson & Clinton, New York: The McGraw-Hill Company.

Polgreen, L. (2009) "Justifying a Costly War in Sri Lanka ", *New York Times*, Published July 19, 2009, at: http://www.nytimes.com/2009/07/19/world/asia/19lanka.html

Posner, D. N. (2004), "Measuring Ethnic Fractionalization in Africa", *American Journal of Political Science*, Vol. 48, No. 4, October 2004, Pp. 849–863.

Reddy, M. (2007) "The great language divide in Sri Lanka", *The Hindu*, Published: May 15, 2007, at: http://www.thehindu.com/2007/05/15/stories/2007051502311100.htm

Reilly, B (2000) "Democracy, Ethnic Fragmentation, and Internal Conflict", *International Security*, Vol. 25, No. 3 (Winter 2000/01), pp. 162-185.

Reynal-Querol, M. (2002) "Ethnicity, Political Systems, and Civil Wars", *Journal of Conflict Resolution*, Vol. 46, No. 1, February 2002, pp 29-54, Sage Publications

Rice, X. (2010) "Rwanda election set to return Kagame as president", *The Guardian*, published August 9, 2010, at: http://www.guardian.co.uk/world/2010/aug/09/rwanda-election-kagame-president

Rotberg, Robert (1999) *Creating peace in Sri Lanka: civil war and reconciliation*, Washington: Brookings Institution Press

Sambanis, N. (2001) "Do Ethnic and Nonethnic Civil Wars Have the Same Causes?: A Theoretical and Empirical Inquiry (Part 1)", *Journal of Conflict Resolution*, Vol. 45, No. 3 (Jun., 2001), pp. 259-282

Sartori, G. (1970), "Concept Misformation in Comparative Politics", *The American Political Science Review*, Vol.64,no.4 (Dec 1970), pp. 1033-1053.

Sartori, G. (1990), "A Typology of Party Systems", in *The West European Party Systems*, Oxford University Press

Schelling, T. (1960, 1980) *The Strategy of Conflict*, Cambridge, Massachusetts.

Suganami, H. (1989) *The domestic analogy and world order proposals*, Cambridge studies in international relations; 6, Cambridge University Press.

UCDP/PRIO (2009) "Armed Conflict Dataset Codebook", Version 4-2009, Uppsala Conflict Data Program (UCDP) and the Centre for the Study of Civil War (CSCW) at the Peace Research Institute Oslo (PRIO), http://www.prio.no/CSCW/Datasets/

USSR (1964) *Atlas Nadorov Mira*, Department of Geodesy and Cartography of the State, Geological Committee of the USSR. Moscow.

Uvin, P. (1999) "Ethnicity and Power in Burundi and Rwanda: Different Paths to Mass Violence", *Comparative Politics*, Vol. 31, No. 3 (Apr., 1999), pp. 253-271.

Waltz, K.N. (1964) "The Stability of a Bipolar World", *Daedalus*, 93: 881–909.

Weathington B., Cunningham, C. & Pittenger D. (2010) *Research Methods for the Behavioral and Social Sciences*, Wiley.

Weidmann, N.B., Rød, J.K. & Cederman, L. (2010) "Representing Ethnic Groups in Space: A New Dataset". *Journal of Peace Research*, in press.

Wimmer, A. & Min, B. (2006) “From Empire to Nation-State: Explaining Wars in the Modern World, 1816–2001”, *American Sociological Review*, 71:867–97.

Wimmer, A., Cederman, L.-E., & Min, B. (2009) “Ethnic Politics and Armed Conflict: A Configurational Analysis”, *American Sociological Review*, 74(2): 316-337.

Wohlforth, W.C. (2006) "The Stability of a Unipolar World", *International Security*, Summer 1999, Vol.24, No. 1, pp. 5-41.

8 Appendix 1

List of countries with main ethnic variables (Source: EPR Dataset)

Country	Bal.	Gr.	Frag.	g_1	g_2
Afghanistan	0,61	9	0,34	41%	25%
Albania	0,03	3	0,02	95%	3%
Algeria	0,38	2	0,01	72%	27%
Angola	0,68	5	0,36	38%	26%
Argentina	0,01	2	0,16	83%	1%
Armenia	0,01	3	0,01	98%	1%
Australia	0,02	2	0,14	84%	2%
Austria	0,01	2	0,06	93%	1%
Azerbaijan	0,04	3	0,07	90%	3%
Bangladesh	0,13	4	0,02	87%	11%
Belarus (Byelorussia)	0,17	3	0,09	78%	13%
Belgium	0,68	3	0,01	59%	40%
Benin	0,56	4	0,49	33%	19%
Bolivia	0,95	4	0,28	37%	35%
Bosnia-Herzegovina	0,79	4	0,16	47%	37%
Brazil	0,80	2	0,01	55%	44%
Bulgaria	0,11	3	0,08	83%	9%
Burundi	0,16	2	0,01	85%	14%
Cambodia (Kampuchea)	0,04	5	0,09	88%	4%
Cameroon	0,72	6	0,57	25%	18%
Canada	0,39	3	0,18	59%	23%
Central African Republic	0,30	4	0,57	33%	10%
Chad	0,58	6	0,62	24%	14%
Chile	0,04	3	0,04	92%	4%
China	0,12	47	0,37	57%	7%
Colombia	0,32	3	0,04	73%	23%
Congo	0,86	5	0,35	35%	30%
Congo, Democratic Republic of (Zaire)	0,64	13	0,74	16%	10%
Costa Rica	0,02	2	0,05	93%	2%
Cote D’Ivoire	0,65	4	0,44	34%	22%
Croatia	0,05	5	0,06	90%	5%
Cuba	0,49	2	0,00	67%	33%
Czech Republic	0,03	3	0,03	94%	3%
Czechoslovakia	0,49	4	0,07	63%	31%
Dominican Republic	0,08	2	0,00	93%	7%
Ecuador	0,73	3	0,05	55%	40%
Egypt	0,10	2	0,01	90%	9%
Eritrea	0,96	2	0,02	50%	48%
Estonia	0,38	4	0,07	68%	26%

Ethiopia	0,94	8	0,38	32%	30%
Finland	0,06	2	0,01	93%	6%
France	0,01	4	0,01	98%	1%
Gabon	0,92	5	0,50	26%	24%
Gambia	0,48	5	0,42	40%	19%
Georgia	0,12	7	0,23	69%	8%
Ghana	0,68	5	0,42	35%	24%
Greece	0,02	3	0,01	97%	2%
Guatemala	0,65	4	0,01	60%	39%
Guinea	0,75	3	0,30	40%	30%
Guinea-Bissau	0,47	3	0,56	30%	14%
Haiti	0,05	2	0,00	95%	5%
Honduras	0,05	6	0,04	91%	5%
Hungary	0,06	2	0,05	90%	5%
India	0,51	19	0,60	26%	13%
Indonesia	0,11	7	0,50	45%	5%
Iran (Persia)	0,55	11	0,32	44%	24%
Iraq	0,30	3	0,18	63%	19%
Israel	0,74	6	0,46	31%	23%
Japan	0,01	4	0,02	97%	1%
Jordan	0,69	2	0,02	58%	40%
Kazakhstan	0,56	7	0,17	53%	30%
Kenya	0,56	7	0,58	27%	15%
Kuwait	0,60	3	0,60	25%	15%
Kyrgyz Republic	0,21	4	0,21	65%	14%
Laos	0,24	8	0,37	51%	12%
Latvia	0,49	4	0,13	59%	29%
Lebanon	0,67	9	0,40	36%	24%
Liberia	0,88	5	0,85	8%	7%
Lithuania	0,08	3	0,10	84%	7%
Macedonia (Former Yugoslav Republic of)	0,35	5	0,11	66%	23%
Madagascar (Malagasy)	0,82	2	0,13	48%	40%
Malawi	0,68	7	0,53	28%	19%
Malaysia	0,44	5	0,17	58%	25%
Mali	0,11	2	0,00	90%	10%
Mauritania	0,75	3	0,30	40%	30%
Mexico	0,13	2	0,04	85%	11%
Moldova	0,21	5	0,19	67%	14%
Mongolia	0,08	2	0,07	86%	7%
Morocco	0,00	1	0,60	40%	0%
Mozambique	0,34	3	0,62	29%	10%
Myanmar (Burma)	0,13	12	0,24	68%	9%
Namibia	0,19	11	0,41	50%	9%
Nepal	0,97	4	0,37	32%	31%
Netherlands	0,12	2	0,05	85%	10%
New Zealand	0,11	3	0,04	87%	10%

Nicaragua	0,11	4	0,07	85%	9%
Niger	0,39	5	0,22	56%	22%
Nigeria	0,72	6	0,50	29%	21%
Pakistan	0,17	7	0,25	64%	11%
Panama	0,08	2	0,08	85%	7%
Paraguay	0,02	2	0,04	94%	2%
Peru	0,92	4	0,12	46%	43%
Philippines	0,09	4	0,07	86%	8%
Poland	0,01	5	0,00	99%	1%
Rumania	0,07	4	0,04	90%	7%
Russia (Soviet Union)	0,05	36	0,16	80%	4%
Rwanda	0,18	2	0,01	84%	15%
Saudi Arabia	0,72	4	0,38	36%	26%
Senegal	0,53	5	0,34	44%	23%
Sierra Leone	0,79	3	0,32	38%	30%
Slovakia	0,13	3	0,09	81%	10%
Slovenia	0,02	6	0,15	83%	2%
South Africa	0,82	6	0,60	22%	18%
Spain	0,25	5	0,15	68%	17%
Sri Lanka (Ceylon)	0,17	4	0,18	70%	12%
Sudan	0,68	13	0,58	25%	17%
Switzerland	0,26	3	0,21	62%	16%
Syria	0,19	5	0,26	62%	12%
Taiwan	0,17	3	0,02	84%	14%
Tajikistan	0,19	5	0,05	80%	15%
Thailand	0,05	4	0,23	74%	4%
Togo	0,61	2	0,29	44%	27%
Trinidad and Tobago	0,98	2	0,20	40%	40%
Turkey/Ottoman Empire	0,21	2	0,01	82%	17%
Turkmenistan	0,06	4	0,10	85%	5%
Uganda	0,88	6	0,70	16%	14%
Ukraine	0,22	5	0,05	78%	17%
United Kingdom	0,12	7	0,09	82%	10%
United States of America	0,18	6	0,18	69%	13%
Uzbekistan	0,07	5	0,15	80%	6%
Venezuela	0,12	2	0,04	86%	10%
Vietnam, Democratic Republic of	0,02	10	0,20	78%	2%
Vietnam, Republic of	0,01	2	0,21	78%	1%
Yemen	0,43	3	0,20	56%	24%
Yugoslavia (Serbia)	0,27	6	0,21	62%	17%
Zambia	0,44	7	0,38	43%	19%
Zimbabwe (Rhodesia)	0,26	3	0,03	77%	20%

9 Appendix 2

DO-FILES (for STATA)

```
clear
set memory 200m

/* ---------------- PREPARING PRIO-DATA -------------------------- */
/* --------------------------------------------------------------- */
/* --------------------------------------------------------------- */

use "UCDP_PRIO_v42009_mainconflicttable.dta", clear
save "temp_ucdp-PRIO.dta", replace


rename location country

keep if type==3 | type==4

keep year country gwnoa type incomp intensity

rename gwnoa gwno
replace gwno = 750 if country == "Hyderabad"

gsort +gwno +year

gen PRIO_cw=0
replace PRIO_cw=1 if intensity==1 | intensity==2

gen PRIO_25=0
replace PRIO_25=1 if intensity==1

gen PRIO_1000=0
replace PRIO_1000=1 if intensity==2

gen PRIO_terr=0
replace PRIO_terr = 1 if incomp==1

gen PRIO_gov=0
replace PRIO_gov = 1 if incomp==2

rename incomp PRIO_incomp

egen yearc = concat(gwno year)
destring yearc, replace

sort yearc

by yearc: egen incompmed = median(PRIO_incomp)
by yearc: keep if PRIO_incomp == incompmed
by yearc: keep if _n==1

sort yearc

save "temp_ucdp-PRIO.dta", replace

/* --------------------------------------------------------------- */
```

```
use "PRIO_onsets.dta"
save "TGRH_countryyears_priobased.dta", replace

gsort +yearc

merge yearc using "temp_ucdp-PRIO.dta", update

replace PRIO_cw=0 if PRIO_cw==.
replace PRIO_25=0 if PRIO_25==.
replace PRIO_1000=0 if PRIO_1000==.
replace PRIO_incomp=0 if PRIO_incomp==.
replace PRIO_terr=0 if PRIO_terr==.
replace PRIO_gov=0 if PRIO_gov==.
replace intensity=0 if intensity==.
replace type=0 if type==.

drop _merge

gsort +country +year

save "TGRH_countryyears_priobased.dta", replace


/* ---------------- ADD SAMBANIS VARS - ONSETS ONLY ----------------- */
/* ---------------- -------------------------- ----------------- */
/* ---------------- -------------------------- ----------------- */

clear
set memory 200m

use "Sambanis.dta", clear
save "Sambanis_TEMP.dta", replace

rename cowcode gwno
replace gwno = 580 if gwno == 511
replace gwno = 260 if gwno == 255 | gwno == 265
replace gwno = 530 if gwno == 529

egen yearc = concat(gwno year)
destring yearc, replace

rename warstns SCW_onset
rename elfo ANM_ethfrac
rename ethfrac FL_ethfrac


keep yearc year country gwno SCW_onset ncontig lnmtn warst1 warst1b warst2
warst2b warst3 warst4 warst5 warst6 warst7 warst8 warst9 warst10 warst11

gsort +yearc

save "Sambanis_TEMP.dta", replace


/* ------------ MERGING SAMBANIS (DSCW_onset) -------- */


use "TGRH_countryyears_priobased.dta", clear

gsort +yearc
```

```
merge yearc using "Sambanis_TEMP.dta", update

replace SCW_onset=0 if missing(SCW_onset)
replace warst7=0 if missing(warst7)

duplicates drop
drop if missing(onset2)
drop _merge

gsort +yearc

save "TGRH_countryyears_priobased.dta", replace


clear
set memory 200m

/* ---------------- MAKING SYSTEM VARS FROM EPR ----------------- */
/* ---------------- ---------------------------- ----------------- */
/* ---------------- ---------------------------- ----------------- */


use "EPR_groupyear_v1.1.dta", clear
save "EPR_groupyear_TEMP.dta", replace

keep cowcode year statename size group stateabb

rename cowcode gwno
replace gwno = 580 if gwno == 511
replace gwno = 260 if gwno == 255 | gwno == 265
replace gwno = 530 if gwno == 529

egen yearc = concat(gwno year)
destring yearc, replace

gsort +year +statename

by year statename, sort: egen systemgroups=count(size)

gen EPR_groups = systemgroups
gen EPR_groupsl = log(systemgroups)
gen EPR_interops = systemgroups*(systemgroups-1)/2

replace systemgroups = 10 if systemgroups>10

gen EPR_groups5peak=((systemgroups-5)^2*(-1)+25)*.04



gsort +year +statename -size

gen propSQ=size*size*(1-size)
by year statename: egen EPR_ethpol = total(propSQ*4)
drop propSQ

gen propSQ=size*(1-size)
by year statename: egen EPR_ethfrac = total(propSQ)
drop propSQ
```

```
by year statename: gen g1 = group
by year statename: gen p1 = size

by year statename: gen g2 = group[_n+1]
by year statename: gen p2 = size[_n+1]

by year statename: gen g3 = group[_n+2]
by year statename: gen p3 = size[_n+2]

by year statename: gen g4 = group[_n+3]
by year statename: gen p4 = size[_n+3]

by year statename: gen g5 = group[_n+4]
by year statename: gen p5 = size[_n+4]

by year statename: gen g6 = group[_n+5]
by year statename: gen p6 = size[_n+5]

by year statename: gen g7 = group[_n+6]
by year statename: gen p7 = size[_n+6]

by year statename: gen g8 = group[_n+7]
by year statename: gen p8 = size[_n+7]

by year statename: gen g9 = group[_n+8]
by year statename: gen p9 = size[_n+8]

by year statename: gen g10 = group[_n+9]
by year statename: gen p10 = size[_n+9]

by year statename: gen g11 = group[_n+10]
by year statename: gen p11 = size[_n+10]

replace p2 = 0 if missing(p2)
replace p3 = 0 if missing(p3)
replace p4 = 0 if missing(p4)
replace p5 = 0 if missing(p5)
replace p6 = 0 if missing(p6)
replace p7 = 0 if missing(p7)
replace p8 = 0 if missing(p8)
replace p9 = 0 if missing(p9)
replace p10 = 0 if missing(p10)
replace p11 = 0 if missing(p11)


by year statename, sort: keep if _n == 1


/* ---------------- BALANCE ------------------------------- */


gen bal=p2/p1
gen EPR_order = 1-p1-p2

gen EPR_balclass = "M"
replace EPR_balclass = "H" if bal>=.1
replace EPR_balclass = "D" if bal>=.33
replace EPR_balclass = "B" if bal>=.66

gen EPR_groupclass = 0
replace EPR_groupclass = 1 if EPR_groups==2
```

```
replace EPR_groupclass = 2 if EPR_groups>=3
replace EPR_groupclass = 3 if EPR_groups>=6

gen EPR_orderclass = 1
replace EPR_orderclass = 2 if EPR_order>=.1
replace EPR_orderclass = 3 if EPR_order>=.33

gen EPR_class_B1 = 0
replace EPR_class_B1 = 1 if EPR_balclass== "B" & EPR_orderclass== 1
gen EPR_class_B2 = 0
replace EPR_class_B2 = 1 if EPR_balclass== "B" & EPR_orderclass== 2
gen EPR_class_B3 = 0
replace EPR_class_B3 = 1 if EPR_balclass== "B" & EPR_orderclass== 3

gen EPR_class_D1 = 0
replace EPR_class_D1 = 1 if EPR_balclass== "D" & EPR_orderclass== 1
gen EPR_class_D2 = 0
replace EPR_class_D2 = 1 if EPR_balclass== "D" & EPR_orderclass== 2
gen EPR_class_D3 = 0
replace EPR_class_D3 = 1 if EPR_balclass== "D" & EPR_orderclass== 3

gen EPR_class_H1 = 0
replace EPR_class_H1 = 1 if EPR_balclass== "H" & EPR_orderclass== 1
gen EPR_class_H2 = 0
replace EPR_class_H2 = 1 if EPR_balclass== "H" & EPR_orderclass== 2
gen EPR_class_H3 = 0
replace EPR_class_H3 = 1 if EPR_balclass== "H" & EPR_orderclass== 3

egen EPR_system = concat(EPR_balclass EPR_orderclass)



/* ---------------- DOMINANCE ---------------------------------------- */

local c=1
gen EPR_balance10 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.9
gen EPR_balance9 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.8
gen EPR_balance8 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.7
gen EPR_balance7 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.6
gen EPR_balance6 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.59
gen EPR_balance59 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.58
gen EPR_balance58 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.57
gen EPR_balance57 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.56
gen EPR_balance56 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.55
gen EPR_balance55 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.54
gen EPR_balance54 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
```

```
local c=.53
gen EPR_balance53 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.52
gen EPR_balance52 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.51
gen EPR_balance51 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.5
gen EPR_balance5 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.4
gen EPR_balance4 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.3
gen EPR_balance3 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.2
gen EPR_balance2 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.1
gen EPR_balance1 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

local c=.0
gen EPR_balance0 =(-(p2/p1-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1

gen EPR_balancelin = (bal*-1)+1
gen EPR_balancexp = ((bal-1)*-(bal-1))*-1



gen EPR_ethdom45 = 0
replace EPR_ethdom45 = 1 if p1<=.9
replace EPR_ethdom45 = 0 if p1<=.45

gen EPR_ethdom55 = 0
replace EPR_ethdom55 = 1 if p1<=.9
replace EPR_ethdom55 = 0 if p1<=.55

rename size largestgroup
move largestgroup statename

gsort +year +statename


save "EPR_groupyear_TEMP.dta", replace

rename p1 EPR_p1
rename p2 EPR_p2
rename bal EPR_bal

gen EPR_BPxG = EPR_bal*EPR_groups
gen EPR_B5xG = EPR_balance5*EPR_groups
gen EPR_B6xG = EPR_balance6*EPR_groups
gen EPR_BPxGl = EPR_bal*EPR_groupsl
gen EPR_B5xGl = EPR_balance5*EPR_groupsl
gen EPR_balance5inv = 1-EPR_balance5
gen EPR_B5xGlneg = EPR_balance5inv*EPR_groupsl
gen EPR_B6xGl = EPR_balance6*EPR_groupsl

gen EPR_balxp1 = EPR_bal*EPR_p1
gen EPR_balxp1_x2 = EPR_bal*(EPR_p1+EPR_p1)
```

```
gen EPR_balxp1neg = EPR_bal*(1-EPR_p1)
gen EPR_bal5xp1 = EPR_balance5*EPR_p1
gen EPR_balxp5neg = EPR_balance5*(1-EPR_p1)
gen EPR_multiheg = (1-EPR_bal)*EPR_order
gen EPR_bal4xorder = EPR_balance4*EPR_order
gen EPR_bal5xorder = EPR_balance5*EPR_order

save "EPR_groupyear_TEMP_full.dta", replace


drop if missing(EPR_system)
rename statename country
gsort +country -year

duplicates drop

gsort +yearc

save "EPR_groupyear_TEMP.dta", replace

clear
/* ---------------- MERGING INTO PRIO-onsets ----------------- */

use "TGRH_countryyears_priobased.dta", clear

gsort +yearc
merge yearc using "EPR_groupyear_TEMP.dta", update

drop _merge
gsort +yearc

save "TGRH_countryyears_priobased.dta", replace

/* ---------------- PREPARING WCE FULL ----------------- */
/* ---------------- ----------------- ----------------- */
/* ---------------- ----------------- ----------------- */

use "WCE2005_full.dta", clear
save "WCE2005_temp.dta", replace

replace country = "Antigua and Barbuda" if country == "Antigua"
replace country = "British Virgin Islands" if country == "British Virgin
Is"
replace country = "Central African Republic" if country == "Central African
Rep"
replace country = "Congo, Republic of the" if country == "Congo"
replace country = "Congo, Democratic Republic of" if country == "DR Congo"
replace country = "Cote d'Ivoire" if country == "Ivory Coast"
replace country = "Micronesia, Federated States of" if country ==
"Micronesia"
replace country = "Myanmar (Burma)" if country == "Myanmar"
replace country = "Korea, North" if country == "North Korea"
replace country = "Korea, South" if country == "South Korea"
replace country = "United Kingdom" if country == "Britain"
replace country = "Vatican City" if country == "Holy See"
replace country = "Malvinas" if country == "Falkland Islands"
replace country = "Northern Mariana Islands" if country == "Northern
Mariana Is"
replace country = "Cyprus" if country == "Northern Cyprus"
replace country = "Saint Kitts and Nevis" if country == "Saint Kitts &
Nevis"
```

```
replace country = "Saint Pierre and Miquelon" if country == "Saint Pierre &
Miquelon"
replace country = "St Vincent and Grenadines" if country == "Saint Vincent"
replace country = "Sao Tome and Principe" if country == "São Tomé &
Príncipe"
replace country = "Timor Leste" if country == "Timor"
replace country = "Trinidad and Tobago" if country == "Trinidad & Tobago"
replace country = "Turks and Caicos Islands" if country == "Turks & Caicos
Is"
replace country = "Virgin Islands (U.S.)" if country == "Virgin Is of the
US"
replace country = "West Sahara" if country == "Sahara"
replace country = "Wallis Futuna Islands" if country == "Wallis Futuna Is"

kountry country, from(other) stuck
drop if missing(_ISO3N_)
sort _ISO3N_

drop country
rename NAMES_STD country
rename _ISO3N_ iso3n

kountry iso3n, from(iso3n) to(cown)
rename _COWN_ gwno

replace gwno = 580 if gwno == 511
replace gwno = 260 if gwno == 255 | gwno == 265
replace gwno = 530 if gwno == 529


/* ---------- ADDING ETHPOL byGROUP VARIABLE (INDEPENDENT) ----------- */

save "WCE2005_temp.dta", replace


rename pop_2005 pop2005
rename people_name group

keep country gwno group pop2005

gsort +country +group

by country: egen total = total(pop2005)
move total pop2005

gen size = (pop2005/total)

gen sizeSQ=size*size*(1-size)
by country: egen WCE_ethpol = total(sizeSQ*4)
drop sizeSQ

gen sizeSQ=size*(1-size)
by country: egen WCE_ethfrac = total(sizeSQ)
drop sizeSQ

drop if missing(gwno)

gsort +country -size

by size, sort: keep if size > .01
```

```
by country, sort: egen systemgroups=count(size)


/* ---------------- MAKING SYSTEM VARS FROM WCE ----------------- */

gen WCE_groups = systemgroups
gen WCE_groupsl = log(systemgroups)


replace systemgroups = 10 if systemgroups>10

gen WCE_groups5peak=((systemgroups-5)^2*(-1)+25)*.04

replace systemgroups = 5 if systemgroups>5

gsort +country -size

by country: gen g1 = group
by country: gen p1 = size

by country: gen g2 = group[_n+1]
by country: gen p2 = size[_n+1]

by country: gen g3 = group[_n+2]
by country: gen p3 = size[_n+2]

by country: gen g4 = group[_n+3]
by country: gen p4 = size[_n+3]

by country: gen g5 = group[_n+4]
by country: gen p5 = size[_n+4]

by country: gen g6 = group[_n+5]
by country: gen p6 = size[_n+5]

by country: gen g7 = group[_n+6]
by country: gen p7 = size[_n+6]

by country: gen g8 = group[_n+7]
by country: gen p8 = size[_n+7]

by country: gen g9 = group[_n+8]
by country: gen p9 = size[_n+8]

by country: gen g10 = group[_n+9]
by country: gen p10 = size[_n+9]

by country: gen g11 = group[_n+10]
by country: gen p11 = size[_n+10]


replace p1 = 0 if missing(p1)
replace p2 = 0 if missing(p2)
replace p3 = 0 if missing(p3)
replace p4 = 0 if missing(p4)
replace p5 = 0 if missing(p5)
replace p6 = 0 if missing(p6)
replace p7 = 0 if missing(p7)
replace p8 = 0 if missing(p8)
replace p9 = 0 if missing(p9)
replace p10 = 0 if missing(p10)
```

```
replace p11 = 0 if missing(p11)


by country, sort: keep if _n == 1

gen bal=p2/p1

gen balclass = "H"
replace balclass = "D" if bal>=.33 & bal<.66
replace balclass = "B" if bal>=.66
replace balclass = "M" if p1>=.9


gen groupclass = 0
replace groupclass = 1 if systemgroups==2
replace groupclass = 2 if systemgroups>=3 & systemgroups<=5
replace groupclass = 3 if systemgroups>=6



gen WCE_balance =(((bal-.25)*(bal-.15)*(.75-bal))+.15)/.18


gen WCE_ethdom45 = 0
replace WCE_ethdom45 = 1 if p1<=.9
replace WCE_ethdom45 = 0 if p1<=.45

gen WCE_ethdom55 = 0
replace WCE_ethdom55 = 1 if p1<=.9
replace WCE_ethdom55 = 0 if p1<=.55


move systemgroups country
rename size largestgroup
move largestgroup country

gsort +country

egen WCE_system = concat(balclass groupclass)


gsort +country

by country, sort: keep if _n == 1

rename p1 WCE_p1
rename p2 WCE_p2
rename bal WCE_bal

replace gwno = 678 if gwno == 679

save "WCE2005_temp_full.dta", replace

keep gwno WCE_p1 WCE_p2 WCE_bal WCE_system WCE_bipol WCE_groups WCE_groupsl
WCE_ethdom45 WCE_ethdom55  WCE_ethpol WCE_ethfrac WCE_balance

gsort +gwno


save "WCE2005_temp.dta", replace
```

```
/* ---------------- MERGING INTO PRIO-onsets ----------------- */

use "TGRH_countryyears_priobased.dta", clear

gsort +gwno
merge gwno using "WCE2005_temp.dta", update

drop _merge
drop if missing(year)
gsort +yearc

save "TGRH_countryyears_priobased.dta", replace



/* ---------------- PREPARING WCE by culture ----------------- */
/* ---------------- ----------------- ----------------- */
/* ---------------- ----------------- ----------------- */

use "WCE2005_full.dta", clear
save "WCEc2005_temp.dta", replace

replace country = "Antigua and Barbuda" if country == "Antigua"
replace country = "British Virgin Islands" if country == "British Virgin
Is"
replace country = "Central African Republic" if country == "Central African
Rep"
replace country = "Congo, Republic of the" if country == "Congo"
replace country = "Congo, Democratic Republic of" if country == "DR Congo"
replace country = "Cote d'Ivoire" if country == "Ivory Coast"
replace country = "Micronesia, Federated States of" if country ==
"Micronesia"
replace country = "Myanmar (Burma)" if country == "Myanmar"
replace country = "Korea, North" if country == "North Korea"
replace country = "Korea, South" if country == "South Korea"
replace country = "United Kingdom" if country == "Britain"
replace country = "Vatican City" if country == "Holy See"
replace country = "Malvinas" if country == "Falkland Islands"
replace country = "Northern Mariana Islands" if country == "Northern
Mariana Is"
replace country = "Cyprus" if country == "Northern Cyprus"
replace country = "Saint Kitts and Nevis" if country == "Saint Kitts &
Nevis"
replace country = "Saint Pierre and Miquelon" if country == "Saint Pierre &
Miquelon"
replace country = "St Vincent and Grenadines" if country == "Saint Vincent"
replace country = "Sao Tome and Principe" if country == "São Tomé &
Príncipe"
replace country = "Timor Leste" if country == "Timor"
replace country = "Trinidad and Tobago" if country == "Trinidad & Tobago"
replace country = "Turks and Caicos Islands" if country == "Turks & Caicos
Is"
replace country = "Virgin Islands (U.S.)" if country == "Virgin Is of the
US"
replace country = "West Sahara" if country == "Sahara"
replace country = "Wallis Futuna Islands" if country == "Wallis Futuna Is"

kountry country, from(other) stuck
drop if missing(_ISO3N_)
sort _ISO3N_
```

```
drop country
rename NAMES_STD country
rename _ISO3N_ iso3n

kountry iso3n, from(iso3n) to(cown)
rename _COWN_ gwno

replace gwno = 580 if gwno == 511
replace gwno = 260 if gwno == 255 | gwno == 265
replace gwno = 530 if gwno == 529

/* -------- ADDING ETHPOL byCULTURE VARIABLE (INDEPENDENT) ------------ */

save "WCEc2005_temp.dta", replace

rename pop_2005 pop2005
keep country gwno culture pop2005

gsort +country +culture

by country culture: egen cult_total = total(pop2005)
move cult_total pop2005

gsort +country -cult_total
by country culture, sort: keep if _n == 1

by country: egen totpop = total(cult_total)
gen size = (cult_total/totpop)

rename culture group

gen sizeSQ=size*size*(1-size)
by country: egen WCEc_ethpol = total(sizeSQ*4)
drop sizeSQ

gen sizeSQ=size*(1-size)
by country: egen WCEc_ethfrac = total(sizeSQ)
drop sizeSQ

drop if missing(gwno)

gsort +country -size

by size, sort: keep if size > .01

by country, sort: egen systemgroups=count(size)


/* ---------------- MAKING SYSTEM VARS FROM WCEc ----------------- */

gen WCEc_groups = systemgroups
gen WCEc_groupsl = log(systemgroups)

replace systemgroups = 10 if systemgroups>10

gen WCEc_groups5peak=((systemgroups-5)^2*(-1)+25)*.04

replace systemgroups = 5 if systemgroups>5

gsort +country -size
```

```
by country: gen g1 = group
by country: gen p1 = size

by country: gen g2 = group[_n+1]
by country: gen p2 = size[_n+1]

by country: gen g3 = group[_n+2]
by country: gen p3 = size[_n+2]

by country: gen g4 = group[_n+3]
by country: gen p4 = size[_n+3]

by country: gen g5 = group[_n+4]
by country: gen p5 = size[_n+4]

by country: gen g6 = group[_n+5]
by country: gen p6 = size[_n+5]

by country: gen g7 = group[_n+6]
by country: gen p7 = size[_n+6]

by country: gen g8 = group[_n+7]
by country: gen p8 = size[_n+7]

by country: gen g9 = group[_n+8]
by country: gen p9 = size[_n+8]

by country: gen g10 = group[_n+9]
by country: gen p10 = size[_n+9]

by country: gen g11 = group[_n+10]
by country: gen p11 = size[_n+10]

replace p1 = 0 if missing(p1)
replace p2 = 0 if missing(p2)
replace p3 = 0 if missing(p3)
replace p4 = 0 if missing(p4)
replace p5 = 0 if missing(p5)
replace p6 = 0 if missing(p6)
replace p7 = 0 if missing(p7)
replace p8 = 0 if missing(p8)
replace p9 = 0 if missing(p9)
replace p10 = 0 if missing(p10)
replace p11 = 0 if missing(p11)


by country, sort: keep if _n == 1

gen bal=p2/p1

gen balclass = "H"
replace balclass = "D" if bal>=.33 & bal<.66
replace balclass = "B" if bal>=.66
replace balclass = "M" if p1>=.9


gen groupclass = 0
replace groupclass = 1 if systemgroups==2
replace groupclass = 2 if systemgroups>=3 & systemgroups<=5
replace groupclass = 3 if systemgroups>=6
```

```
gen WCEc_balance =(((bal-.25)*(bal-.15)*(.75-bal))+.15)/.18



gen WCEc_ethdom45 = 0
replace WCEc_ethdom45 = 1 if p1<=.9
replace WCEc_ethdom45 = 0 if p1<=.45

gen WCEc_ethdom55 = 0
replace WCEc_ethdom55 = 1 if p1<=.9
replace WCEc_ethdom55 = 0 if p1<=.55


move systemgroups country
rename size largestgroup
move largestgroup country

gsort +country

egen WCEc_system = concat(balclass groupclass)


gsort +country

by country, sort: keep if _n == 1


rename p1 WCEc_p1
rename p2 WCEc_p2
rename bal WCEc_bal

replace gwno = 678 if gwno == 679

save "WCEc2005_temp_full.dta", replace

keep gwno WCEc_p1 WCEc_p2 WCEc_bal WCEc_system WCEc_bipol WCEc_groups
WCEc_groupsl WCEc_ethdom45 WCEc_ethdom55 WCEc_balance WCEc_ethpol
WCEc_ethfrac

gsort +gwno


save "WCEc2005_temp.dta", replace

/* ---------------- MERGING INTO PRIO-onsets ----------------- */

use "TGRH_countryyears_priobased.dta", clear

gsort +gwno
merge gwno using "WCEc2005_temp.dta", update

drop _merge
drop if missing(year)
gsort +yearc

save "TGRH_countryyears_priobased.dta", replace


/* ---------------- PREPARING ANM (by GREG) ----------------- */
/* ---------------- ----------------- ----------------- */
```

```
use "ANMGREG.dta", clear
save "ANMGREG_temp.dta", replace


/* ------ ADDING ETHPOL byGROUP VARIABLE (INDEPENDENT) ----------------- */

rename shortname group
rename cowcode gwno

drop if gwno == "n.a."
drop if gwno == "GL"
drop if country == "Korea (North and South)"


destring gwno, replace
replace gwno = 580 if gwno == 511
replace gwno = 260 if gwno == 255 | gwno == 265
replace gwno = 530 if gwno == 529

keep country gwno group size

gsort +country +group

gen sizeSQ=size*size*(1-size)
by country: egen ANM_ethpol = total(sizeSQ*4)
drop sizeSQ

gen sizeSQ=size*(1-size)
by country: egen ANM_ethfrac = total(sizeSQ)
drop sizeSQ

drop if missing(gwno)

gsort +country -size

by size, sort: keep if size > .01

by country, sort: egen systemgroups=count(size)


/* ---------------- MAKING SYSTEM VARS FROM ANM ----------------- */

gen ANM_groups = systemgroups
gen ANM_groupsl = log(systemgroups)

replace systemgroups = 10 if systemgroups>10

gen ANM_groups5peak =((systemgroups-5)^2*(-1)+25)*.04

replace systemgroups = 5 if systemgroups>5

gsort +country -size

by country: gen g1 = group
by country: gen p1 = size

by country: gen g2 = group[_n+1]
by country: gen p2 = size[_n+1]

by country: gen g3 = group[_n+2]
```

```
by country: gen p3 = size[_n+2]

by country: gen g4 = group[_n+3]
by country: gen p4 = size[_n+3]

by country: gen g5 = group[_n+4]
by country: gen p5 = size[_n+4]

by country: gen g6 = group[_n+5]
by country: gen p6 = size[_n+5]

by country: gen g7 = group[_n+6]
by country: gen p7 = size[_n+6]

by country: gen g8 = group[_n+7]
by country: gen p8 = size[_n+7]

by country: gen g9 = group[_n+8]
by country: gen p9 = size[_n+8]

by country: gen g10 = group[_n+9]
by country: gen p10 = size[_n+9]

by country: gen g11 = group[_n+10]
by country: gen p11 = size[_n+10]


replace p1 = 0 if missing(p1)
replace p2 = 0 if missing(p2)
replace p3 = 0 if missing(p3)
replace p4 = 0 if missing(p4)
replace p5 = 0 if missing(p5)
replace p6 = 0 if missing(p6)
replace p7 = 0 if missing(p7)
replace p8 = 0 if missing(p8)
replace p9 = 0 if missing(p9)
replace p10 = 0 if missing(p10)
replace p11 = 0 if missing(p11)


by country, sort: keep if _n == 1

gen bal=p2/p1

gen balclass = "H"
replace balclass = "D" if bal>=.33 & bal<.66
replace balclass = "B" if bal>=.66
replace balclass = "M" if p1>=.9


gen groupclass = 0
replace groupclass = 1 if systemgroups==2
replace groupclass = 2 if systemgroups>=3 & systemgroups<=5
replace groupclass = 3 if systemgroups>=6

gen ANM_balance =(((bal-.25)*(bal-.15)*(.75-bal))+.15)/.18


gen ANM_ethdom45 = 0
replace ANM_ethdom45 = 1 if p1<=.9
replace ANM_ethdom45 = 0 if p1<=.45
```

```
gen ANM_ethdom55 = 0
replace ANM_ethdom55 = 1 if p1<=.9
replace ANM_ethdom55 = 0 if p1<=.55


move systemgroups country
rename size largestgroup
move largestgroup country

gsort +country

egen ANM_system = concat(balclass groupclass)

gsort +country

by country, sort: keep if _n == 1

save "ANMGREG_temp_full.dta", replace

rename p1 ANM_p1
rename p2 ANM_p2
rename bal ANM_bal


keep gwno ANM_p1 ANM_p2 ANM_bal ANM_system ANM_bipol ANM_groups ANM_groupsl
ANM_ethdom45 ANM_ethdom55 ANM_ethpol ANM_ethfrac ANM_balance

gsort +gwno

save "ANM2005_temp.dta", replace

/* ---------------- MERGING INTO COUNTRYYEAR_EGIP ----------------- */

use "TGRH_countryyears_priobased.dta", clear

gsort +gwno +year
merge gwno using "ANM2005_temp.dta", update

drop _merge
gsort +gwno +year

save "TGRH_countryyears_priobased.dta", replace




/* ---------------- PREPARING JOSHUA (by GREG) ----------------- */
/* ---------------- ----------------- ----------------- */
/* ---------------- ----------------- ----------------- */

use "JOSHUA.dta", clear
save "JOSHUA_temp.dta", replace


/* ------ ADDING ETHPOL byGROUP VARIABLE (INDEPENDENT) ----------------- */

rename peopnameacrosscountries group

kountry iso3, from(iso3c) to(cown)
rename _COWN_ gwno
```

```
drop if missing(gwno)
replace gwno = 580 if gwno == 511
replace gwno = 260 if gwno == 255 | gwno == 265
replace gwno = 530 if gwno == 529

keep country gwno group population

gsort +country +group

by country: egen total = total(population)

gen size = (population/total)

gen sizeSQ=size*size*(1-size)
by country: egen JOSHUA_ethpol = total(sizeSQ*4)
drop sizeSQ

gen sizeSQ=size*(1-size)
by country: egen JOSHUA_ethfrac = total(sizeSQ)
drop sizeSQ

drop if missing(gwno)

gsort +country -size

by size, sort: keep if size > .01

by country, sort: egen systemgroups=count(size)


/* ---------------- MAKING SYSTEM VARS FROM JOSHUA ----------------- */

gen JOSHUA_groups = systemgroups
gen JOSHUA_groupsl = log(systemgroups)

replace systemgroups = 10 if systemgroups>10

gen JOSHUA_groups5peak =((systemgroups-5)^2*(-1)+25)*.04

replace systemgroups = 5 if systemgroups>5

gsort +country -size

by country: gen g1 = group
by country: gen p1 = size

by country: gen g2 = group[_n+1]
by country: gen p2 = size[_n+1]

by country: gen g3 = group[_n+2]
by country: gen p3 = size[_n+2]

by country: gen g4 = group[_n+3]
by country: gen p4 = size[_n+3]

by country: gen g5 = group[_n+4]
by country: gen p5 = size[_n+4]

by country: gen g6 = group[_n+5]
by country: gen p6 = size[_n+5]
```

```
by country: gen g7 = group[_n+6]
by country: gen p7 = size[_n+6]

by country: gen g8 = group[_n+7]
by country: gen p8 = size[_n+7]

by country: gen g9 = group[_n+8]
by country: gen p9 = size[_n+8]

by country: gen g10 = group[_n+9]
by country: gen p10 = size[_n+9]

by country: gen g11 = group[_n+10]
by country: gen p11 = size[_n+10]


replace p1 = 0 if missing(p1)
replace p2 = 0 if missing(p2)
replace p3 = 0 if missing(p3)
replace p4 = 0 if missing(p4)
replace p5 = 0 if missing(p5)
replace p6 = 0 if missing(p6)
replace p7 = 0 if missing(p7)
replace p8 = 0 if missing(p8)
replace p9 = 0 if missing(p9)
replace p10 = 0 if missing(p10)
replace p11 = 0 if missing(p11)


by country, sort: keep if _n == 1

gen bal=p2/p1

gen balclass = "H"
replace balclass = "D" if bal>=.33 & bal<.66
replace balclass = "B" if bal>=.66
replace balclass = "M" if p1>=.9


gen groupclass = 0
replace groupclass = 1 if systemgroups==2
replace groupclass = 2 if systemgroups>=3 & systemgroups<=5
replace groupclass = 3 if systemgroups>=6

gen JOSHUA_balance =(((bal-.25)*(bal-.15)*(.75-bal))+.15)/.18


gen JOSHUA_ethdom45 = 0
replace JOSHUA_ethdom45 = 1 if p1<=.9
replace JOSHUA_ethdom45 = 0 if p1<=.45

gen JOSHUA_ethdom55 = 0
replace JOSHUA_ethdom55 = 1 if p1<=.9
replace JOSHUA_ethdom55 = 0 if p1<=.55


move systemgroups country
rename size largestgroup
move largestgroup country

gsort +country
```

```
egen JOSHUA_system = concat(balclass groupclass)


gsort +country

by country, sort: keep if _n == 1

save "JOSHUA_temp_full.dta", replace

rename p1 JOSHUA_p1
rename p2 JOSHUA_p2
rename bal JOSHUA_bal

keep gwno JOSHUA_p1 JOSHUA_p2 JOSHUA_bal JOSHUA_system JOSHUA_bipol
JOSHUA_groups JOSHUA_groupsl JOSHUA_ethdom45 JOSHUA_ethdom55 JOSHUA_ethpol
JOSHUA_ethfrac

JOSHUA_balance

gsort +gwno

save "JOSHUA_temp.dta", replace

/* ---------------- MERGING INTO COUNTRYYEAR_EGIP ----------------- */

use "TGRH_countryyears_priobased.dta", clear

gsort +gwno +year
merge gwno using "JOSHUA_temp.dta", update

drop _merge
gsort +gwno +year

save "TGRH_countryyears_priobased.dta", replace




/* -----PREPARING VARS FROM 'DRIVING DEMOCRACY' DBASE ----------------- */
/* ------------------------------ ----------------- ----------------- */
/* ------------------------------ ----------------- ----------------- */

clear
set memory 200m

use "driving_dem.dta", clear
save "temp_demdev.dta", replace

egen yearc = concat(gwno Year)
destring yearc, replace

generate demsq = Polity1^2
replace demsq = . if demsq>101

rename INST inst
rename Watts2 fed
rename elecfam04 elecsys
rename Polity1 democracy
```

```
label var demsq "Democracy index squared, calc from Polity1 in Driving
Democracy dataset"
label var inst "Institution, (INST) from Driving Democracy dataset"
label var fed "Watts2 definition in EGIP dataset"

keep yearc demsq inst fed elecsys democracy

gsort +yearc

save "temp_demdev.dta", replace

/* ------------ MERGING DEM & INST -------- */

use "TGRH_countryyears_priobased.dta", clear

gsort +yearc

merge yearc using "temp_demdev.dta", update


drop _merge
duplicates drop

gsort +country +year

save "TGRH_countryyears_priobased.dta", replace


/* ------- PREPARING REGIONS VARIABLE FROM 'DRIVING DEM' ------------- */

clear
set memory 200m

use "driving_dem.dta", clear
save "temp2_demdev.dta", replace


rename Region8 region8
rename Africa africa
rename Asia asia
rename CEurope ceurope
rename Nam nam
rename Sam sam
rename Scan scan
rename MEast meast
rename WEuro weuro

keep gwno region8 africa asia ceurope nam sam scan meast weuro

gsort +gwno

save "temp2_demdev.dta", replace

/* ------------ MERGING REGIONS -------- */

use "TGRH_countryyears_priobased.dta", clear

gsort +gwno

merge gwno using "temp2_demdev.dta", update
```

```
drop _merge

gsort +country +year

save "TGRH_countryyears_priobased.dta", replace



/* ------PREPARING LPOP & LGDPC VARIABLEs FROM WDI DBASE -------- */
/* ---------------- ----------------- --------------------------- */
/* ---------------- ----------------- --------------------------- */

use "wdi.dta", clear
save "wdi_TEMP.dta", replace

egen yearc = concat(gwno Year)
destring yearc, replace

gen lgdpc = log(GDPCap)
gen lgdp_ppp = log(GDP_ppp)
gen lpop = log(Pop_total)

label var lgdpc "Log of GDP per capita, calc from WDI dataset"
label var lgdp_ppp "Log of total GDP in Purchasing Power Parity, calc from
WDI dataset"
label var lpop "Log of total population, calc from WDI dataset"

keep yearc lgdpc lgdp_ppp lpop

gsort +yearc

save "wdi_TEMP.dta", replace

/* ------------ MERGING DEM & INST -------- */

use "TGRH_countryyears_priobased.dta", clear

gsort +yearc

merge yearc using "wdi_TEMP.dta", update


drop _merge


replace region8 = 2 if country == "Republic of Vietnam"
replace region8 = 4 if country == "Yemen Arab Republic"
replace region8 = 3 if country == "Czechoslovakia"
replace region8 = 2 if country == "Vietnam"


drop if missing(country)
drop if missing(gwno)
duplicates drop

gsort +yearc

gen coldwar = 0
replace coldwar = 1 if year<1990

gsort +yearc
```

```
save "TGRH_countryyears_priobased.dta", replace


/* ---------------- ADDING EPR CONFLICT DATA ----------------- */
/* ---------------- ----------------------- ----------------- */
/* ---------------- ----------------------- ----------------- */


clear
set memory 200m

use "EPR\EPR_EAC.dta", clear
save "EPR_EAC_temp.dta", replace

replace cowcode = 260 if cowcode == 255
egen yearc = concat(cowcode year)
destring yearc, replace
move yearc country

gen EPR_p1 = .
replace EPR_p1 = .99 if ethrelevant == 1

gen EPR_p2 = .
replace EPR_p2 = .01 if ethrelevant == 1

gen EPR_system = "M1" if ethrelevant == 1
gen EPR_bal = 1/99
local c=.4
gen EPR_balance4 =(-(.01/.99-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.5
gen EPR_balance5 =(-(.01/.99-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
local c=.6
gen EPR_balance6 =(-(.01/.99-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2+1
gen EPR_groups = 2
gen EPR_groupsl = log(2)
gen EPR_interops = 1
gen EPR_order = 0
gen EPR_ethpol = 4*(    (.99*.99*(1-.99)) + (.01*.01*(1-.01)))
gen EPR_ethfrac = (.99*(1-.99)) + (.01*(1-.01))
gen EPR_ethdom45 = 0

keep country EPR_ethdom45 EPR_p1 EPR_p2 EPR_system EPR_bal EPR_balance4
EPR_balance5 EPR_balance6 EPR_groups EPR_groupsl EPR_interops EPR_order
EPR_ethpol EPR_ethfrac ///
yearc lpopl gdpcapl newethwartype2 newonset newonset2 newethonset2
newseconset newhionset2 newethhionset2 newethwartype ///
newethintensity onsetfl2 ongoingdrop ethonsetfl2 oilpc anocl democl regchg3
lmtnest npeaceyears ethrelevant


rename lpopl EPR_lpopl
rename gdpcapl EPR_gdpcapl
rename ongoingdrop EPR_ongoingdrop
rename newonset2 EPR_newonset2
rename newhionset2 EPR_newhionset2
rename newethonset2 EPR_newethonset2
rename newseconset EPR_newseconset
rename newethhionset2 EPR_newethhionset2
rename onsetfl2 EPR_onsetfl2
rename ethonsetfl2 EPR_ethonsetfl2
```

```
rename oilpc EPR_oilpc
rename newethintensity EPR_newethintensity
rename anocl EPR_anocl
rename democl EPR_democl
rename regchg3 EPR_regchg3
rename lmtnest EPR_lmtnest
rename npeaceyears EPR_npeaceyears
rename newethwartype EPR_newethwartype
rename newethwartype2 EPR_newethwartype2
rename ethrelevant EPR_ethrelevant

replace EPR_newethwartype2 = 2 if EPR_newethwartype2 == 3


gsort +yearc

save "EPR_EAC_temp.dta", replace

clear
use "TGRH_countryyears_priobased.dta", clear

gsort +yearc

merge yearc using "EPR_EAC_temp.dta", update

drop if missing(gwno)
drop _merge

save "TGRH_countryyears_priobased.dta", replace


/* -------FINAL TOUCH-UP OF VARBIABLES ---------------- */
/* ---------------- ----------------- ----------------- */
/* ---------------- ----------------- ----------------- */


gen EPR_balxorder = EPR_bal*EPR_order
gen EPR_groupsxorder = EPR_groups*EPR_order

gen WCE_order = 1-WCE_p1-WCE_p2
gen WCE_balxorder = WCE_bal*WCE_order

gen WCEc_order = 1-WCEc_p1-WCEc_p2
gen WCEc_balxorder = WCEc_bal*WCEc_order
gen ANM_order = 1-ANM_p1-ANM_p2
gen ANM_balxorder = ANM_bal*ANM_order

gen JOSHUA_order = 1-JOSHUA_p1-JOSHUA_p2
gen JOSHUA_balxorder = JOSHUA_bal*JOSHUA_order

gen SNLEBCC_order = 1-SNLEBCC_p1-SNLEBCC_p2
gen SNLEBCC_balxorder = SNLEBCC_bal*SNLEBCC_order

local c=.4
gen WCE_balance4=(-(WCE_bal-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2 +1
gen WCEc_balance4=(-(WCEc_bal-`c')^2)    /(sqrt((`c'-.5)^2)+.5)^2 +1
gen ANM_balance4=(-(ANM_bal-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2 +1
gen JOSHUA_balance4=(-(JOSHUA_bal-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2 +1
gen SNLEBCC_balance4=(-(SNLEBCC_bal-`c')^2)    /(sqrt((`c'-.5)^2)+.5)^2 +1

/*------------------ ADDING LABELS--------------------*/
```

```
label var EPR_npeaceyears "Years since last war"
label var EPR_gdpcapl "GDP per capita, lagged"
label var EPR_lpopl "Ln(Population), lagged"
label var EPR_oilpc "Oil production per capita"
label var EPR_anoc "Anocracy, dummy"
label var EPR_regchg3 "Political instability"
label var EPR_lmtnest "Ln(Mountains)"
label var EPR_groups "Groups"
label var EPR_order "Fragmentation(1-g1-g2)"
label var EPR_groupsl "Ln(Groups)
label var EPR_interops "Interaction Opportunities"
label var EPR_npeaceyears "Years since last war"
label var EPR_gdpcapl "GDP per capita, lagged"
label var EPR_lpopl "Ln(Population), lagged"
label var EPR_oilpc "Oil production per capita"
label var EPR_anoc "Anocracy, dummy"
label var EPR_regchg3 "Political instability"
label var EPR_lmtnest "Ln(Mountains)"
label var EPR_groups "Groups"
label var EPR_groupsl "Ln(Groups)
label var EPR_interops "Interaction Opportunities"
label var EPR_bal "Balance (g2/g1)"
label var EPR_balance1 "Dominance (1/10 peak)"
label var EPR_balance2 "Dominance (2/10 peak)"
label var EPR_balance3 "Dominance (3/10 peak)"
label var EPR_balance4 "Dominance (4/10 peak)"
label var EPR_balance5 "Dominance (5/10 peak)"
label var EPR_balance6 "Dominance (6/10 peak)"
label var EPR_balance7 "Dominance (7/10 peak)"
label var EPR_balance8 "Dominance (8/10 peak)"
label var EPR_balance9 "Dominance (9/10 peak)"
label var EPR_balance10 "Dominance (10/10 peak)"
label var EPR_ethpol "Ethnic polarization"
label var EPR_ethfrac "Ethnic fractionalization"
label var EPR_ethheg70 "Hegemony, two-group"
label var EPR_multiheg "Hegemony, multigroup"
label var EPR_B5xGl "Ln(Groups)*Dominance(5/10)"
label var EPR_bal4xorder "Fragm.*Dominance(4/10)"
label var EPR_bal5xorder "Fragm.*Dominance(5/10)"

drop if gwno == 571

save "TGRH_countryyears_priobased.dta", replace

/* --------- CREATING COUNTRIES DATABASE FOR YEAR 2000 --- */
/* ------------------------------------------------------- */
/* ------------------------------------------------------- */

use "TGRH_countryyears_priobased.dta", clear
save "TGRH_countries_priobased.dta", replace

gsort +country +year
by country, sort: egen PRIO_cwtot = total(PRIO_cw)
by country, sort: egen PRIO_25tot = total(PRIO_25)
by country, sort: egen PRIO_1000tot = total(PRIO_1000)
by country, sort: egen PRIO_terrtot = total(PRIO_terr)
by country, sort: egen PRIO_govtot = total(PRIO_gov)
by country, sort: egen PRIO_incompmed = median(PRIO_incomp)
by country, sort: egen SCW_onsettot = total(SCW_onset)
```

```
by country, sort: egen EPR_newonset2tot = total(EPR_newonset2)
by country, sort: egen EPR_newhionset2tot = total(EPR_newhionset2)
by country, sort: egen EPR_newethonset2tot = total(EPR_newethonset2)
by country, sort: egen EPR_newseconsettot = total(EPR_newseconset)
by country, sort: egen EPR_newethhionset2tot = total(EPR_newethhionset2)
by country, sort: egen EPR_onsetfl2tot = total(EPR_onsetfl2)
by country, sort: egen EPR_ethonsetfl2tot = total(EPR_ethonsetfl2)
by country, sort: egen EPR_newethwartype2max = max(EPR_newethwartype2)

by country, sort: egen EPR_multipol3max = max(EPR_multipol3)
by country, sort: egen EPR_multipol4max = max(EPR_multipol4)
by country, sort: egen EPR_multipol5max = max(EPR_multipol5)
by country, sort: egen EPR_multipol6max = max(EPR_multipol6)
by country, sort: egen EPR_lmultipol3max = max(EPR_lmultipol3)
by country, sort: egen EPR_lmultipol4max = max(EPR_lmultipol4)
by country, sort: egen EPR_lmultipol5max = max(EPR_lmultipol5)
by country, sort: egen EPR_lmultipol6max = max(EPR_lmultipol6)
by country, sort: egen EPR_ethdom45mean = mean(EPR_ethdom45)
by country, sort: egen EPR_ethdom55mean = mean(EPR_ethdom55)

by country, sort: egen instmedian = median(inst)
by country, sort: egen elecsysmedian = median(elecsys)
by country, sort: egen fedmedian = median(fed)
by country, sort: egen demsqmedian = median(demsq)
by country, sort: egen lgdpcmean = mean(lgdpc)
by country, sort: egen lgdp_pppmean = mean(lgdp_ppp)
by country, sort: egen lpopmean = mean(lpop)

by country, sort: egen ANM_ethfracmean = mean(ANM_ethfrac)
by country, sort: egen ncontigmed = median(ncontig)
by country, sort: egen lnmtnmean = mean(lnmtn)
by country, sort: egen region8med = median(region8)

by country, sort: egen africamed = median(africa)
by country, sort: egen asiamed = median(asia)
by country, sort: egen ceuropemed = median(ceurope)
by country, sort: egen sammed = median(sam)
by country, sort: egen nammed = median(nam)
by country, sort: egen scanmed = median(scan)
by country, sort: egen meastmed = median(meast)
by country, sort: egen weuromed = median(weuro)

replace PRIO_incompmed = 0 if missing(PRIO_incompmed)

gsort +country -year
by country, sort: keep if _n == 5


gen PRIO_cwdum = 0
replace PRIO_cwdum = 1 if PRIO_cwtot>=1 | PRIO_25tot>=1
replace PRIO_incompmed = 0 if PRIO_cwdum==0

gen EPR_ethwardummy = 0
replace EPR_ethwardummy = 1 if EPR_newethonset2tot>=1

gen EPR_ethterrdummy = 0
replace EPR_ethterrdummy = 1 if EPR_newethwartype2max==1

gen EPR_ethgovdummy = 0
replace EPR_ethgovdummy = 1 if EPR_newethwartype2max==2
```

```
label var PRIO_incompmed "1=Territory, 2=Government, 3=Both"


gsort +country

save "TGRH_countries_priobased.dta", replace

/* ---------------- PREPARING TABLE FOR COUNTRY COMPARISON ------------ */

keep  country stateabb EPR_system PRIO_cwtot EPR_ethwardummy
EPR_newethonset2tot EPR_ethgovdummy EPR_ethterrdummy gwno

save "TGRH_SYSTEM_comparisons_with-war-defs.dta", replace




*/------------------------ TABLE 1 -------------------------------------*/
*/--------------------------------------------------------------------------*/
*/--------------------------------------------------------------------------*/

clear
set memory 200m

use "TGRH_countryyears_priobased.dta", clear


eststo clear

eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_anocl EPR_oilpc EPR_regchg3 EPR_lmtnest EPR_groups EPR_groupsl
EPR_interops EPR_order

EPR_ethfrac EPR_ethpol, cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_anocl EPR_oilpc EPR_regchg3 EPR_lmtnest EPR_groups, cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_anocl EPR_oilpc EPR_regchg3 EPR_lmtnest EPR_groupsl, cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_anocl EPR_oilpc EPR_regchg3 EPR_lmtnest EPR_interops, cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_anocl EPR_oilpc EPR_regchg3 EPR_lmtnest EPR_order, cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_anocl EPR_oilpc EPR_regchg3 EPR_lmtnest EPR_ethfrac, cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_anocl EPR_oilpc EPR_regchg3 EPR_lmtnest EPR_ethpol, cluster(country)

esttab using "TABLE1_whether-groups-matter.rtf", replace se unstack onecell
label nogaps nomtitles compress pr2 aic sca("ll Log-likelihood") ///
addnotes("Robust standard errors in parenthesis. Based on the Ethnic Power
Relations dataset. ") ///
title("TABLE 1: Logistic regressions on ethnic conflict onsets: The effect
of fragmentation")


*/------------------------ TABLE 2 -------------------------------------*/
*/--------------------------------------------------------------------------*/
*/--------------------------------------------------------------------------*/
```

```
clear
set memory 200m

use "TGRH_countryyears_priobased.dta", clear



eststo clear

eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_order EPR_bal,
cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_order EPR_balance4 ,
cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5 ,
cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_order EPR_balance6 ,
cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_order EPR_dominant,
cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_order EPR_ethdom45,
cluster(country)


eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5
EPR_bal5xorder, cluster

(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_balance5 if EPR_groups==2,
cluster(country)

esttab using "TABLE2_whether-balance-matters.rtf", replace se unstack
onecell nogaps nonotes compress pr2 aic sca("ll Log-likelihood") ///
addnotes("Robust standard errors in parenthesis. Based on the Ethnic Power
Relations dataset.") ///
title("TABLE 2: Logistic regressions on ethnic conflict onsets: The effect
of Balance")


*/------------------------ TABLE 3 -------------------------------------*/
*/----------------------------------------------------------------------*/
*/----------------------------------------------------------------------*/

clear
set memory 200m

use "TGRH_countryyears_priobased.dta", clear


replace EPR_newethwartype2 = 0 if EPR_newethonset2 == 0

gen EPR_balanced = .
replace EPR_balanced = 0 if EPR_bal<=1
replace EPR_balanced = 1 if EPR_bal>=.75
```

```
gen EPR_hegemonic = .
replace EPR_hegemonic = 0 if EPR_bal<=1
replace EPR_hegemonic = 1 if EPR_bal<=.25

gen EPR_dominant = .
replace EPR_dominant = 0 if EPR_bal<=1
replace EPR_dominant = 1 if EPR_bal>=.25 & EPR_bal<=.75

label var EPR_dominant "Dominant system(dummy)"
label var EPR_balanced "Balanced system(dummy)"
label var EPR_hegemonic "Hegemonic system(dummy)"

gen EPR_terronset = 0
replace EPR_terronset = . if missing(EPR_newethonset2)
replace EPR_terronset = 1 if EPR_newethwartype2 == 1

gen EPR_govonset = 0
replace EPR_govonset = . if missing(EPR_newethonset2)
replace EPR_govonset = 1 if EPR_newethwartype2 == 2 | EPR_newethwartype2 ==
3

gen EPR_autocl = .
replace EPR_autocl = 1 if EPR_democl==0 | EPR_anocl==0
replace EPR_autocl = 0 if EPR_democl==1 | EPR_anocl==1



eststo clear


eststo: logit EPR_terronset EPR_npeaceyears EPR_gdpcapl EPR_lpopl EPR_oilpc
EPR_anocl EPR_regchg3 EPR_lmtnest EPR_groups EPR_balance5, cluster(country)
estadd ysumm, sum
eststo: logit EPR_govonset EPR_npeaceyears EPR_gdpcapl EPR_lpopl EPR_oilpc
EPR_anocl EPR_regchg3 EPR_lmtnest EPR_groups EPR_bal , cluster(country)
estadd ysumm, sum
eststo: mlogit EPR_newethwartype2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_groups EPR_balance5 ,
cluster(country)
estadd ysumm, sum
eststo: mlogit EPR_newethwartype2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anocl EPR_regchg3 EPR_lmtnest EPR_groups EPR_bal ,
cluster(country)
estadd ysumm, sum


esttab using "TABLE3_territory-government.rtf", replace se unstack onecell
label nogaps nonotes compress pr2 aic sca("ll Log-likelihood" "N_clust
Countries" "ysum

Total onsets") ///
addnotes("Robust standard errors in parenthesis. Based on the Ethnic Power
Relations dataset. ") ///
mtitles("Territory" "Government" "Binominal 1" "Binominal 2") ///
title("Table 3: Logistic and binominal regressions on territorial and
government onsets")

*/----------------------- TABLE 3b ------------------------------------*/
*/-------------------------------------------------------------------*/
*/-------------------------------------------------------------------*/
```

```
clear
set memory 200m

use " TGRH_countryyears_priobased.dta", clear

drop if EPR_p1 >= .99

replace EPR_newethwartype2 = 0 if EPR_newethonset2 == 0

gen EPR_balanced = .
replace EPR_balanced = 0 if EPR_bal<=1
replace EPR_balanced = 1 if EPR_bal>=.75

gen EPR_hegemonic = .
replace EPR_hegemonic = 0 if EPR_bal<=1
replace EPR_hegemonic = 1 if EPR_bal<=.25

gen EPR_dominant = .
replace EPR_dominant = 0 if EPR_bal<=1
replace EPR_dominant = 1 if EPR_bal>=.25 & EPR_bal<=.75

label var EPR_dominant "Dominant system(dummy)"
label var EPR_balanced "Balanced system(dummy)"
label var EPR_hegemonic "Hegemonic system(dummy)"

gen EPR_terronset = 0
replace EPR_terronset = . if missing(EPR_newethonset2)
replace EPR_terronset = 1 if EPR_newethwartype2 == 1

gen EPR_govonset = 0
replace EPR_govonset = . if missing(EPR_newethonset2)
replace EPR_govonset = 1 if EPR_newethwartype2 == 2 | EPR_newethwartype2 ==
3

gen EPR_autocl = .
replace EPR_autocl = 1 if EPR_democl==0 | EPR_anocl==0
replace EPR_autocl = 0 if EPR_democl==1 | EPR_anocl==1


eststo clear


eststo: logit EPR_anocl EPR_order EPR_bal EPR_balance5 EPR_gdpcapl
EPR_lpopl EPR_oilpc EPR_lmtnest, cluster(country)
eststo: logit EPR_democl EPR_order EPR_bal EPR_balance5 EPR_gdpcapl
EPR_lpopl EPR_oilpc EPR_lmtnest, cluster(country)
eststo: logit EPR_autocl EPR_order EPR_bal EPR_balance5 EPR_gdpcapl
EPR_lpopl EPR_oilpc EPR_lmtnest, cluster(country)

esttab using "TABLE3b_quality-of-government.rtf", replace se unstack
onecell label nogaps nonotes compress pr2 aic sca("ll Log-likelihood"
"N_clust Countries" "ysum Total onsets") ///
addnotes("Robust standard errors in parenthesis. Based on the Ethnic Power
Relations dataset. ") ///
mtitles("Anocracy" "Democracy" "Autocracy") ///
title("Table 3b: Logistic regressions on form of government")



*/------------------------ TABLE 4 -------------------------------------*/
```

```
*/-------------------------------------------------------------------------*/
*/-------------------------------------------------------------------------*/

clear
set memory 200m

use "TGRH_countryyears_priobased.dta", clear

eststo clear

eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anoc EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5,
cluster(country)
eststo: logit EPR_newethhionset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anoc EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5,
cluster(country)
eststo: logit EPR_ethonsetfl2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anoc EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5,
cluster(country)

eststo: logit EPR_newonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl EPR_oilpc
EPR_anoc EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5, cluster(country)
eststo: logit onset5 EPR_npeaceyears EPR_gdpcapl EPR_lpopl EPR_oilpc
EPR_anoc EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5, cluster(country)
eststo: logit EPR_onsetfl2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl EPR_oilpc
EPR_anoc EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5, cluster(country)
eststo: logit SCW_onset EPR_npeaceyears EPR_gdpcapl EPR_lpopl EPR_oilpc
EPR_anoc EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5, cluster(country)


esttab using "TABLE4_different-dep-vars.rtf", replace se unstack onecell
label nogaps nonotes nomtitles compress pr2 aic sca("ll Log-likelihood")
///
addnotes("Robust standard errors in parenthesis.") ///
title("TABLE 4: Logistic regressions on different dependent variables")



*/------------------------ TABLE 5 -------------------------------------*/
*/-------------------------------------------------------------------------*/
*/-------------------------------------------------------------------------*/

clear
set memory 200m

use "TGRH_countryyears_priobased.dta", clear


local c=.5
gen WCE_balance5=(-(WCE_bal-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2     +1
gen WCEc_balance5=(-(WCEc_bal-`c')^2)    /(sqrt((`c'-.5)^2)+.5)^2     +1
gen ANM_balance5=(-(ANM_bal-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2     +1
gen JOSHUA_balance5=(-(JOSHUA_bal-`c')^2)/(sqrt((`c'-.5)^2)+.5)^2     +1

eststo clear

eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anoc EPR_regchg3 EPR_lmtnest EPR_order EPR_balance5,
cluster(country)
```

```
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anoc EPR_regchg3 EPR_lmtnest WCE_order WCE_balance5,
cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anoc EPR_regchg3 EPR_lmtnest WCEc_order WCEc_balance5,
cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anoc EPR_regchg3 EPR_lmtnest ANM_order ANM_balance5,
cluster(country)
eststo: logit EPR_newethonset2 EPR_npeaceyears EPR_gdpcapl EPR_lpopl
EPR_oilpc EPR_anoc EPR_regchg3 EPR_lmtnest JOSHUA_order JOSHUA_balance5,
cluster(country)

esttab using "TABLE5_robustness_different-ethnic-datasets.rtf", replace se
unstack onecell nogaps label nonotes compress pr2 aic sca("ll Log-
likelihood") ///
addnotes("Robust standard errors in parenthesis. Based on the Ethnic Power
Relations dataset. Ethnic monopoly (g2/g1<1/10) excluded.") ///
title("TABLE 5: Robustness across different ethnic datasets")



*/------------------------ TABLES 6 & 7 -------------------------------*/
*/---------------------------------------------------------------------*/
*/---------------------------------------------------------------------*/

clear
set memory 200m

use " TGRH_countries_priobased.dta", clear

drop if EPR_p1 >= .99

gen frag_kvant = 1
replace frag_kvant = 2 if EPR_order>=.071
replace frag_kvant = 3 if EPR_order>=.314

gen bal_kvant = 1
replace bal_kvant = 2 if EPR_bal>=.25
replace bal_kvant = 3 if EPR_bal>=.75

table bal_kvant frag_kvant , row col scol

gen group_kvant = 1
replace group_kvant = 2 if EPR_groups>=3
replace group_kvant = 3 if EPR_groups>=6

table bal_kvant group_kvant , row col scol




*/------------------------ FIGURES 7 & 8 ------------------------------*/
*/---------------------------------------------------------------------*/
*/---------------------------------------------------------------------*/



clear
set memory 200m
```

```
use "TGRH_countries_priobased.dta", clear

replace EPR_groups = 10 if EPR_groups>10

set scheme s1mono, perm

gen stateabb_top = ""
replace stateabb_top = stateabb if EPR_order>.6

twoway (scatter EPR_order ANM_order, mlabel(stateabb_top) ytitle("EPR
Fragmentation", height(7)) xtitle("ANM Fragmentation", height(7))
mlabposition(9) title("Figure 8: Fragmentation for EPR and ANM",
placement(nw)))
graph export EPR_frag-ANM_frag_distribution.png, replace

replace stateabb_top = ""
replace stateabb_top = stateabb if EPR_ethfrac>.7

twoway (scatter EPR_ethfrac ANM_ethfrac, mlabel(stateabb_top) ytitle("EPR
Fractionalization", height(7)) xtitle("ANM Fractionalization", height(7))
mlabposition(9) title("Figure 7: Fractionalization for EPR and ANM",
placement(nw)))
graph export EPR_frac-ANM_frac_distribution.png, replace




*/------------------------ FIGURES 9 to 12 - PROBABILITIES --------------*/
*/-----------------------------------------------------------------------*/
*/-----------------------------------------------------------------------*/

clear
set memory 200m

use "TGRH_countryyears_priobased.dta", clear

drop if EPR_p1 >= .99

*** FIGURE 2
* means: fragmentation: .1732317
* means: dominance(4): .7050623
* means: balance: .294015
* predicted probabilities (Model 3.1 and 3.2)


gen pterr_dom= 1/(1+2.71^(-( ///
-12.54 /// constant
+(-0.00809*15.54) /// peaceyears
+(-0.0519*6.056) /// GDP per capita
+(0.570*9.20229) /// population
+(0.0115*2.078) /// oil per capita
+(1.022*0.2242) /// anocracy
+(0.0487*0.1219) /// political instability
+(-0.0302*2.2026) /// mountains
+(1.110*0.173231) /// fragmentation
+(2.240*EPR_balance4) /// DOMINANCE4
)))

gen pterr_frag= 1/(1+2.71^(-( ///
-12.54 /// constant
+(-0.00809*15.54) /// peaceyears
```

```
+(-0.0519*6.056) /// GDP per capita
+(0.570*9.20229) /// population
+(0.0115*2.078) /// oil per capita
+(1.022*0.2242) /// anocracy
+(0.0487*0.1219) /// political instability
+(-0.0302*2.2026) /// mountains
+(2.240*0.7050623) /// dominance
+(1.110*EPR_order) /// FRAGMENTATION
)))


gen pgov_bal= 1/(1+2.71^(-( ///
-5.518 /// constant
+(0.0161*15.54) /// peaceyears
+(-0.171*6.056) /// GDP per capita
+(-0.0288*9.20229) /// population
+(0.0409*2.078) /// oil per capita
+(0.571*0.2242) /// anocracy
+(0.391*0.1219) /// political instability
+(0.0723*2.2026) /// mountains
+(2.682*0.173231 ) /// fragmentation
+(0.393*EPR_bal) /// BALANCE
)))

gen pgov_frag= 1/(1+2.71^(-( ///
-5.518 /// constant
+(0.0161*15.54) /// peaceyears
+(-0.171*6.056) /// GDP per capita
+(-0.0288*9.20229) /// population
+(0.0409*2.078) /// oil per capita
+(0.571*0.2242) /// anocracy
+(0.391*0.1219) /// political instability
+(0.0723*2.2026) /// mountains
+(0.393*0.294015 ) /// balance
+(2.682*EPR_order) /// FRAGMENTATION
)))

twoway (mband pterr_frag EPR_order, clcolor(gs4) clpat(solid)) ,
ytitle(p(territory conflict), size(large) height(7)) ///
title("Figure 9: Probability of Territorial Conflict", placement(nw)) ///
xtitle(Fragmentation, size(large) height(7)) scheme(s1mono)
graph export Probabilities_TERR_frag.png, replace

twoway (mband pterr_dom EPR_balance4, clcolor(gs4) clpat(solid)) ,
ytitle(p(territorial conflict), size(large) height(7)) ///
title("Figure 10: Probability of Territorial Conflict", placement(nw)) ///
xtitle(Dominance, size(large) height(7)) scheme(s1mono)
graph export Probabilities_TERR_dom.png, replace

twoway (mband pgov_frag EPR_order, clcolor(gs4) clpat(solid)) ,
ytitle(p(government conflict), size(large) height(7)) ///
title("Figure 11: Probability of Government Conflict", placement(nw)) ///
xtitle(Fragmentation, size(large) height(7)) scheme(s1mono)
graph export Probabilities_GOV_frag.png, replace

twoway (mband pgov_bal EPR_bal, clcolor(gs4) clpat(solid)) ,
ytitle(p(government conflict), size(large) height(7)) ///
title("Figure 12: Probability of Government Conflict", placement(nw)) ///
xtitle(Balance, size(large) height(7)) scheme(s1mono)
graph export Probabilities_GOV_bal.png, replace
```

```
*/------------------------ FIGURES 11 & 12 - DISTRIBUTIONS --------------*/
*/-----------------------------------------------------------------------*/
*/-----------------------------------------------------------------------*/


clear
set memory 200m

use "TGRH_countryyears_priobased.dta", clear

keep if year == 2000

drop if EPR_groups == 1
replace EPR_groups = 10 if EPR_groups>10

gen stateabb_top = ""
replace stateabb_top = stateabb if EPR_order>.31
replace stateabb_top = stateabb if EPR_bal>.75 & EPR_order>.01


twoway (scatter EPR_bal EPR_order, mlabel(stateabb_top) ytitle("Balance",
size(large) height(7)) ylabel(0(.25)1) xlabel(0(.2)1)
xtitle("Fragmentation", size(large) height(7)) ///
xscale(range(0 1)) xline(.07 .31, lpattern(dash)) yline(.25 .75,
lpattern(dash)) mlabposition(3) title("Figure 11: Distribution of Countries
by Fragmentation", placement(nw)))
graph export distribution_bal-frag.png, replace

twoway (scatter EPR_bal EPR_groups, mlabel(stateabb_top) ytitle("Balance",
size(large) height(7)) ylabel(0(.25)1) xlabel(2(1)10) xtitle("Groups",
size(large) height(7)) ///
xscale(range(1 11)) xline(2.5 5.5, lpattern(dash)) yline(.25 .75,
lpattern(dash)) mlabposition(3) title("Figure 12: Distribution of Countries
by Groups", placement(nw)))
graph export distribution_bal-groups.png, replace
```