Civil conflict and three dimensions of ethnic inequality

John D. Huber*

Laura Mayoral†

November 7, 2012

Abstract

THE ABSTRACT COULD BE REWRITTEN EMPHASIZING THE TRADITIONAL DIFFICULTIES IN FINDING A RELATION BETWEEN INEQU. AND CONFLICT AND OUR EXPLANATION FOR THAT. Most empirical studies on civil conflict are not able to find a significant relationship between interpersonal-measures of economic inequality and the likelihood of conflict. When individuals belong to groups, general inequality (measured by the Gini) can be decomposed into three components: between-group inequality (BGI), within-group inequality (WGI), and 'Overlap' (which is inversely related to the economic segregation of groups). This paper shows that is possible to establish a robust empirical relation between group-based measures of income differences and conflict. Drawing on over 200 individual-level surveys from 89 countries, we create a new data set that allows us to measure these three components and to examine their empirical relationship with civil conflict. Consistent with Esteban and Ray's (2011) argument about the need for labor and capital to fight civil wars, we find a strong, robust positive association between WGI and civil conflict. And consistent with the "contact hypothesis" in sociology, we find that the economic segregation of groups (as measured by a lower Overlap component) is often associated with more civil conflict. Since some components of inequality are associated with more civil conflict but others are associated with less, the analysis helps explain why it has been difficult to identify a relationship between general inequality and civil war. And the strong finding for WGI underscores the value of developing clear theories about how the internal characteristics of groups influence the incidence of civil conflict.

*Department of Political Science, Columbia University, jdh39@columbia.edu.

†Institute for Economic Analysis, Barcelona, Spain, laura.mayoral@iae.csic.es.

1 Introduction

Economic inequality has long been posited as a central driver of social conflict (e.g., Sen 1973). Cross-national empirical research has not, however, established a robust empirical connection between conflict and inequality, leading scholars to often conclude that economic grievances are not important to explaining civil conflict (e.g., Fearon and Laitin 2003 and Collier and Hoeffler 2004).[1] As a consequence, considerable research has focused away from *class* and toward *ethnicity* as key determinant of civil war.

Most internal conflicts since WWII appear to be ethnic in nature (Sambanis 2001), and considerable empirical research examines the connection between ethnicity and conflict (e.g., Montalvo and Reynal-Querol 2005). Recently, Esteban and Ray (2011) and Esteban, Mayoral and Ray (2012) have shown that ethnic divisions matter for conflict, not intrinsically as *primordialist* arguments claim, but rather *instrumentally*, when ethnic markers are used as a means of restricting political power or economic benefits to a subset of the population. However, the channels through which ethnicity and conflict are connected need further examination. An important question concerns how the interaction of inequality and ethnicity affects the incidence of civil war.

Some scholars have linked inequality and ethnicity by arguing that inequality *between* ethnic groups leads to conflict (e.g., Wintrobe 1995, Stewart 2002, Acemoglu and Robinson 2005, and Stewart 2008). Cederman et al. (2011) present evidence supporting this idea: using a novel global dataset combining geocoded data on politically relevant ethnic groups' settlement areas and spatial wealth measures, they show that both advanced and backward ethnic groups are more likely to experience civil wars than groups whose wealth lies closer to the national average (after controlling for political power access). They interpret this result as evidence that economic grievances between groups contribute to civil war.

Our goal is to probe more deeply how the intersection of group identity and inequality are related empirically to civil conflict. As noted, scholars have invoked arguments about economic grievances to posit that inequality between groups should be associated with more ethnic conflict. But recent theoretical research points to another aspect of inequality and group identity that de-

[1] In a detailed survey of the many attempts to empirically link income inequality and social conflict, Lichbach (1989) concludes that the evidence is thoroughly mixed and he cites a variety of studies that support positive, negative or no relationship at all. More recent empirical studies such as Fearon and Laitin (2003) and Collier and Hoeffler (2004) point in the same direction.

serves empirical scrutiny: inequality *within* groups. Specifically, Esteban and Ray (2011) present an innovative theoretical model that is based on the assumption that waging conflict requires both labor and capital. Since poor individuals typically provide the labor and rich individuals typically provide the necessary economic resources, groups that have both – i.e., groups with higher levels of within-group inequality – should be best positioned to wage conflict. Thus, countries with higher levels of within group-inequality should have higher levels of conflict.

AND OVERLAP?

We use 232 individual-level surveys from 89 countries to test arguments about inequality, ethnicity and conflict. The surveys make it possible to estimate the three components of the Gini decomposition: between-group inequality, within-group inequality, and a residual, often called 'Overlap.' The new data set makes it possible to explore empirically the relationship between the various components of inequality and civil war. Our results indicate that there is a strong and robust relationship between within-group inequality and conflict. The empirical results also reveal that the Overlap component of inequality – which is negatively related to the social stratification of groups – is often associated with less civil conflict. In contrast to previous research, we cannot find a significant asociation between between-group inequality and conflict.

Our findings regarding within-group inequality and overlap help explain why it should be difficult to find a relationship between overall inequality and conflict. As the Overlap term of the Gini increases, overall inequality increases, but the economic segregation of groups decreases, which is associated with less conflict. By contrast, as within-group inequality increases, the Gini increases, and this increase is associated with more conflict. The empirical findings therefore suggest that the Gini coefficient masks the conflicting effects of two substantively important dimensions of inequality.

The paper is organized as follows. The next section describes in more detail the three elements of the Gini decomposition and the theoretical arguments that can be used to link each to civil conflict. Section 3 describes our inequality data set, discusses problems of combining data from diverse types of surveys and lays out our solution to these problems. Section 4 presents our empirical results, and we conclude with a discussion of their implications.

2 Three dimensions of inequality

The Gini coefficient (G) is often depicted using the familiar Lorenz curve, the right most (solid) curve in Figure 1. Assume individuals are arrayed from poorest to richest on the x-axis and let p represent a percentile. For example, at $p = .4$, forty percent of the population is to the left of p and is poorer than the individual at p, and sixty percent of the population is to the right of p and is richer than the individual at p. The y-axis depicts $L(p)$, the proportion of income in society held by all individuals to the left of any given point p. The graph of p against $L(p)$ traces the Lorenz curve. If there were perfect equality, the Lorenz curve would be a 45-degree line (i.e., $L(p) = p$ for all p), and as inequality increases, the Lorenz curve bends to the southeast, away from the 45-degree line (so that $L(p) < p$ for all $p < 1$). The area between the 45-degree line and the Lorenz curve can be directly related to the Gini, and as this area increases, inequality increases. We can write the Gini as

$$G = 2\int_0^1 [p - L(p)]dp. \tag{1}$$

WE NEED THE DISCRETE VERSION AS WELL, SO THAT WE CAN COMPARE THIS DEFINITION TO THOSE OF BGI AND WGI.

If individuals can be assigned to groups, then at least since Pyatt (1976), scholars have observed that this Gini coefficient is the sum of three components: BGI (between-group inequality), WGI (within-group inequality) and O (overlap). BGI is a calculation of the society's Gini based on the assumption that each member of a group has the group's average income (with a weighting of groups by their size and a normalization for average income in society). Using discrete data, it can be written as

$$BGI = \frac{1}{2\bar{y}}(\sum_{m=1}^{k}\sum_{n=1}^{k} p_m p_n \mid \bar{y}_m - \bar{y}_n \mid), \tag{2}$$

where k denotes the number of groups, p_m is the proportion of the population in group m, $\bar{y}$ is the mean income in society, and $\bar{y}_m$ is the average income of group m.

WGI is determined by calculating the Gini coefficient for each group, and then summing these across all groups, weighting by group size (so unequal small groups have less weight than

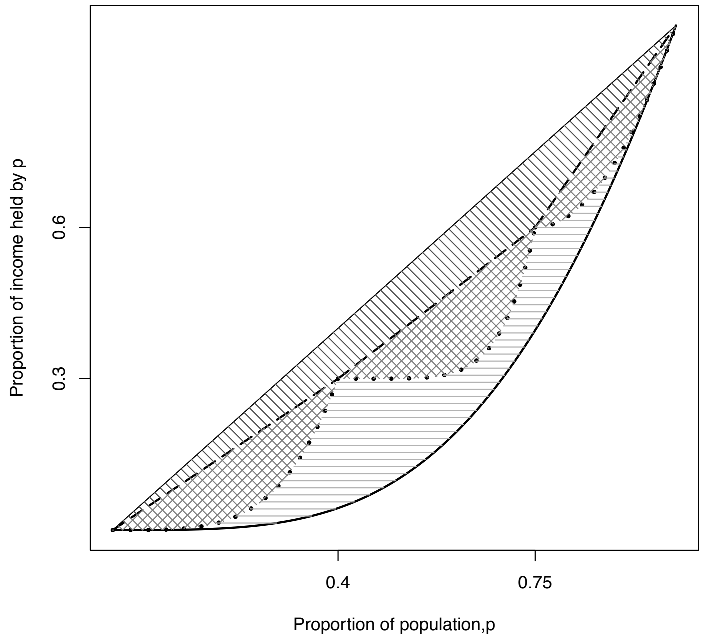

Figure 1: The Gini decomposition

unequal large groups) and by the proportion of income controlled by groups (so that holding group size constant, high inequality in a group controlling a small proportion of resources in society will contribute less to WGI than will high inequality in a group controlling a large proportion of resources). Using discrete data, WGI can be written as

$$WGI = \sum_{i=1}^{k} G_i p_i \pi_i, \tag{3}$$

where G_i is the Gini coefficient for group i and π_i is the proportion of total income going to group i. Overlap is the residual that remains when BGI and WGI are subtracted from the Gini (i.e., 0=G-WGI-BGI).

Figure 1 graphically depicts the three components of the Gini coefficient.[2] In Figure 1 there are three groups. The between-group inequality component of the Gini is the area with diagonal shading that is between the 45-degree line and the dashed line. To create this dashed line, we assume that all individuals have the average income of their group. Thus, the individuals in the

[2]The figure and discussion borrows from Huber, Orgozalek and Gore (2012). The original figure and exposition of the graphical interpretation of the Gini decomposition is by Lambert and Aronson (1993).

poorest group, which has 40 percent of the population, are arrayed in the space on the x-axis to the left of 0.4. The individuals in the middle-income group (which has 35 percent of the population) are arrayed between the position marked 0.4 and 0.75 on the x-axis. And the individuals in the richest group, 25 percent of the population, are all arrayed to the right of the point 0.75. In this example, the poorest group controls 30 percent of societal income. Thus, the straight line from (0,0) to the point (0.4, 0.3) represents this poor group's contribution to the between-group Lorenz curve. The middle income group also controls 30 percent of income, so the line segment from (0.4, 0.3) to (0.75, 0.6) represents their contribution to the between-group Lorenz curve, and the rich group, which controls 40 percent of income, makes a contribution to the group Gini represented by the segment from (0.75, 0.6) to the point (1,1). Thus, the area between the 45-degree line and the dashed line segments is the inequality in society that is due exclusively to differences in the average income of groups (weighting by group size).

To depict within-group inequality, for each group we can array the individuals from poorest to richest, constraining them to be located in the regions on the x-axis used to calculate BGI (so there is no overlap in the incomes of group members). For the poorest group, for example, we would arrange the individuals from poorest to richest in the interval 0, 0.4 on the x-axis, and we can then calculate the inequality within the poor group. The Lorenz curve depicting inequality within the poor group is therefore the cross-hatched area in the lobe below the BGI line segment between 0 and 0.4. The Gini coefficients for the other two groups, depicted by their Lorenz curves, are depicted in a similar fashion. WGI, then, is the sum of these three cross-hatched areas.

One can see in Figure 1 that BGI and WGI do not capture total income inequality in a society. The area shaded by the horizontal lines represents the third component of the Gini coefficient, Overlap. Recall that in depicting BGI and WGI, we array the individuals from each group so that there is no overlap in the incomes of individuals from different groups. The richest individual in the poorest group, for example, is assumed to have a lower income than the poorest individual in the middle income group. While this assumption does not interfere with accurately calculating WGI and BGI, overlap typically exists in the incomes of individuals from different groups. The extent to which such overlap in group incomes exists determines the magnitude of this residual component. In the extreme case, where in fact there is no overlap in group based incomes, the overlap term is zero. As overlap in group incomes increases, the Overlap term increases.[3]

Scholars have interpreted the Overlap term as a measure that is inversely related to the income stratification of groups (e.g., Yitzhaki and Lerman 1991, Yitzhaki 1994, Lambert and Aronson 1993 and Lambert and Decoster 2005) – the greater is O, the less stratified is society. If individuals from particular groups tend to have incomes that are different than members of other groups, then Overlap will be small (and thus will contribute little to the Gini). As the number of individuals from different groups who have the same income increases, the Overlap term increases, decreasing the economic segregation of groups from each other. Overlap can be related to WGI and BGI. For example, assume there are two groups and that there is some overlap in their income distributions. Holding constant the distribution of groups' incomes around their means, as the average incomes move closer together, BGI declines and Overlap increases. Alternatively, assume that the average incomes of the two groups stay the same, but that the dispersion of incomes around the means grows larger. Then WGI would increase and so would Overlap.

But changes to BGI or WGI need not lead to changes in Overlap. For example, suppose that in the two group example the richest people in the rich group become better off. This would increase the dispersion in the rich group, increasing WGI, it would increase the mean of the rich group, raising BGI, but it would leave Overlap essentially unchanged (so long as the rich people in the rich group (whose incomes increase) already have a higher income than the rich people in the poor group). Similarly, suppose the poorest in the poor group increase their income, but not so much that there is more overlap in the incomes of the poor and rich group. Then WGI would decrease (because there would be more equality in the poor group), BGI would decrease (because the mean of the poor group would increase), but Overlap would be unaffected. Thus, while the Overlap term is often related to WGI and BGI, it is distinct and it has substantively useful interpretation related to how segregated groups are by income. Such segregation, it is important to note, could be high (or low) with high or low levels of BGI, or with high or low levels of WGI. As we show below, empirically these components are in fact very weakly correlated with each other.

[3] Since the Gini does not decompose neatly into within-group and between-group components, scholars have at times turned to general entropy measures like the Theil index, which cleanly decompose into within- and between-group components. General entropy measures, however, cannot be used to make the sort of cross-national comparisons we are making because the upper bound on the measures is sensitive to the number of groups, making the measures incomparable across countries where the number or size of groups vary considerably. For this reason, the components of the Theil index are most useful in making comparisons where the number of groups across comparison units is constant (such as when comparing inequality between urban and rural areas across states).

2.1 Inequality, ethnicity and civil conflict

As noted in the Introduction, most empirical studies on civil conflict do not find a significant relationship between economic inequality and the likelihood of conflict (e.g., Fearon and Laitin 2003 and Collier and Hoeffler 2004). Several scholars have argued that these studies fail to find such a connection because they rely on interpersonal rather than group-based measures of income differences. For instance, Acemoglu and Robinson (2005) argue that

> [E]verything else equal, greater inter-group inequality makes revolution more attractive for the citizens: with revolution, they get a chance to share the whole income of the economy (minus what's destroyed in revolution)[...]. However, these predictions about inter-group inequality may not translate into statements about standard measures of inequality and income distribution (such as the labor share or the Gini coefficient). This is particularly so when political conflict is not rich versus poor, but along other lines, perhaps between ethnic or religious groups.

BGI and WGI have been connected theoretically to civil conflict. Between-group inequality is a direct measure of the economic differences between groups, a variable that has also been linked to economic grievances. The basic idea, which is summarized in Cederman et al. (2011), stems from relative deprivation theory – inequality between groups leads to frustrated expectations by group members, leading to a propensity for conflict. Their empirical results show that both advanced and backward ethnic groups are more likely to experience civil wars than groups whose wealth lies closer to the national average. Although highly intuitive, this basic premise is not without ambiguity. Condra (2012) provides an alternative explanation to account for seemingly contradictory result that both poor and rich groups are involved in conflict. He argues that different kinds of insurgency might have different requirements. Whereas territorial rebellions are most likely to involve peripheral groups that usually tend to be poorer, rebellions aiming to take over the state require wealthier and better organized ethnic groups. Using data for Africa, he presents some evidence supporting these claims.

Esteban and Ray (2011) highlight basic theoretical ambiguity. On one hand, if the winning group can expropriate the rival's resources, the larger is the income gap between the groups, the greater the potential prize, and hence the greater the incentive for conflict by the poorer group (see Acemoglu and Robinson 2005, Wintrobe 1995, Stewart 2002 and Cramer 2003). On the other hand, we might expect decreases in between-group inequality to be associated with increased civil conflict. When a poor group is very poor, it may have incentive to fight, but it will have little ability to vanquish its richer foe. As the poor group becomes richer, it has more resources to fight, and thus should be more likely to engage in conflict. Thus, decreases in BGI could lead to increases in civil conflict.

Within-group inequality is a direct measure of income heterogeneity within groups. Esteban and Ray (2011) argue that this variable should be unambiguously related to conflict when we consider both the human and financial opportunity costs of engaging in conflict. When within-group inequality is high, the poor individuals in the group will be rather poor, lowering their opportunity costs of fighting the conflict. Thus, the supply of combatants will be highest in groups with a large number of poor individuals. This alone, however, cannot lead to conflict as the combatants need resources. If the group also contains rich individuals, such resources exist and the opportunity cost of funding combatants decreases. Thus, within-group inequality should be associated with higher levels of ethnic conflict because economically diverse groups have lower opportunity costs of providing combatants and a higher ability to pay them. Esteban and Ray provide a number of examples of how economic diversity within groups is linked with ethnic conflict, but to our knowledge their theory has not been previously tested.

Unlike BGI and WGI, which have been analyzed directly in the literature on civil conflict, to the best of our knowledge the Overlap term in the Gini decomposition has not been discussed in previous research. As mention before, O is (negatively) related to the economic segregation of groups and therefore there are reasons to suspect it could be related to conflict. As O decreases, there is less overlap in the distribution of group incomes – that is, individuals in one group are less likely to find individuals in another group with the same income. Moreover, Income and spatial segregation often go together. The correlation between O and the several indices of spatial segregation presented in Alesina and Zhuravskaya (2011) is around 40%. Thus, it is reasonable to expect that a decrease in O will tend to reduce the number of contacts among people belonging to different groups. As pointed out by Alesina and Zhuravskaya (2011), Glaeser (2005) and Uslaner (2008), by means of reducing contacts between groups, segregation may reduce trust, reinforce negative stereotypes and increase hate. The contact hypothesis from sociology points in a similar direction. Forbes (2004, 69–70; see also Forbes 1997) describes the logic as follows:

> The contact hypothesis is a broad generalization about the effects of personal contact between the members of different ethnic or racial groups on their prejudiced opinions and discriminatory behavior. The basic idea is that more contact between individuals belonging to antagonistic social groups (defined by culture, language, beliefs, skin color, nationality, etc.) tends to undermine the negative stereotypes they have of each other and to reduce their mutual antipathies, thus improving intergroup relations by making people more willing to deal with each other as equals. In short, more contact means less ethnic or cultural conflict, other things being equal.

This hypothesis is controversial and scholars have expended considerable energy understanding the circumstances under which it might be true, and when it might be the case that more inter-group contact is associated with more prejudice, discrimination and conflict. Pettigrew and Tropp's (2006) review of the literature, however, finds that on balance, there is considerable support for the contact hypothesis. Alesina and Zhuravskaya (2011) provide evidence supporting a negative relationship between indices of spatial segregation and the quality of the government. Matuszeski and Schneider (2006) find a positive association between spatial segregation and civil conflict. Thus, to the extent that the economic segregation of groups is related to social segregation, we should expect conflict to decline with O.

3 Measuring the components of the Gini decomposition.

To measure the three components of the Gini, we need data on the group identity of individuals and their incomes. Such data are available only in individual-level surveys, and the challenge we face is to identify surveys containing this information from a wide variety of countries. Ideally, such surveys would have fine-grained income or household expenditure data, but unfortunately the number of surveys with such information is quite small, and in fact in many poorer countries there are large regions where cash transactions are relatively rare. Our strategy is therefore to cast a wide net to include as many countries as possible, and then to take advantage of the highest quality inequality data we can identify to adjust the measures we generate using "income" data that varies in quality. This section describes our approach and provides descriptive information about the Gini decomposition in 89 countries.

We have identified three types of surveys with relevant group and income variables. The first category, which we refer to as HES (for "Household Expenditure Data") includes the best surveys available in the world for calculating inequality. These include surveys like the Luxembourg Income Study, the Living Standards Monitoring Surveys, other similar household expenditure surveys, as well as national censuses. The second type of survey has household income data, but in a form that is much less precise than that of HES surveys. These include the World Values Surveys (WVS), which typically has about 10 household income categories per country, and the Comparative Study of Elections Surveys (CSES), which reports income in quintiles. The third type of survey does not have household income data, but rather has information on various assets that individuals possess. Such surveys are typically used in countries where there are many poor individuals, and thus where cash incomes are often non-existent. In such cases, social scientists have often used an array of asset indicators (such as the type of housing, flooring, water, toilet facilities, transportation, or electronic equipment the household possesses) to determine the relative economic well-being of the household. We follow Filmer and Pritchett (2001) and McKenzie (2005) and run a factor analysis on these asset variables to determine the weights of the various assets in distinguishing household well-being. We then use the factor scores, and the responses to the asset questions, to measure the "wealth" of the respondent. The surveys of this type include the Demographic Health Surveys (DHS) and the Afrobarometer Surveys (Rounds 2-4). Of these two, the DHS surveys allow a more precise measure of "well-being" because they contain a larger number of asset indicators than do the Afrobarometer surveys.

The "income" data from these surveys, along with a group identity measure, make it possible to measure inequality within and across groups. We rely on the list of groups from Fearon (2003) to identify the group membership of survey respondents. Fearon provides a set of clear and reasonable criteria for identifying the relevant groups across a wide range of countries, and his list is widely used in the literature.[4] Following Huber, Orgazalek and Gore (2012), we omit surveys that do not adequately represent the groups in the Fearon list.[5] Our surveys do an excellent job of representing the ethnic divisions in society: the correlation between the ELF from our surveys and the ELF from Fearon's data is .93.

In total, we have 234 surveys from 89 countries, and they are depicted in the map in Figure 2. The surveys were conducted from 1992 to 2008, with 224 of the 234 surveys in the period 1995-

[4] See Fearon (2003) and Baldwin and Huber (2010) for a discussion of the Fearon group categories.

[5] Specifically, if the groups that Fearon identifies and that cannot be identified in our surveys represent more than 10 percent of the population (per Fearon's data), we omit the survey.

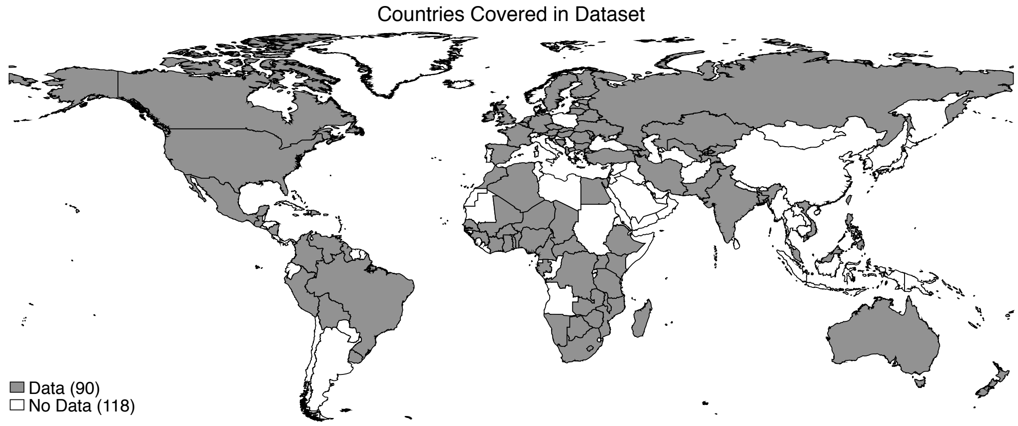

Figure 2: Countries included in data set

2008 (and 10 surveys in the period 1992-94).[6] For 28 countries, we have only one survey, whereas in others we have multiple surveys, at most 7. In countries where several surveys are available, the inequality variables display very limited within-country variation in the period considered. Thus, in our empirical analysis we don't exploit that source of variation (we do not include fixed effects) but we focus on exploiting the cross-country one. We use all the data available to create our best estimate of WGI and BGI for each country (see below), and then we focus on a time period, 1995-2009, which is most proximate to the dates for the preponderance of the surveys we utilize. As a robustness check, we also extend the data set backwards in time to 1960.

3.1 Survey heterogeneity

An important issue we face is that the surveys themselves vary in the nature and quality of their "income" measures. These differences, which are discussed in some detail in Huber, Orgazalek and Gore (2012), make it difficult to compare measures based on different surveys. The DHS, for example, measures income based on the asset variables, which will focus attention on differences at the low end of the income distribution, underestimating differences due to richer individuals in society. The CSES, to take another example, measures household income, but in a crude fashion that forces all individuals into income quintiles. Thus, it is necessary to control or correct for these differences prior to carry out any cross-country analysis of the data.

The tradeoff between obtaining a sample with broad coverage and having inequality mea-

[6] A list of these surveys is provided in the Appendix.

sures that are comparable across countries is not new. In fact, similar limitations have plagued most inequality datasets. For instance, the observations in Deininger and Squire's (1996) dataset differ in many respects (most significantly, in their income definitions and their reference units), so they are rarely comparable across countries or even over time within a single country.[7] Following Deininger and Squire's advice, scholars have typically followed two strategies to overcome this problem. The first is to restrict the sample so that only observations based on the same type of underlying data are employed. However, this leads to a dramatic reduction of the number of observations available for analysis. The second strategy is to calculate the average difference in inequality between observations that vary in their income definition and then adjust observations by this difference. This route leads to applying a constant adjustment across all countries and years that share the same income definition. This is clearly problematic because it fails to capture the variation across countries (and over time) and, therefore, will lead to underestimation of inequality for some observations and overestimation for others.

Our approach to addressing this tradeoff draws on the technique described in Solt (2009). He has recently proposed an alternative methodology that permits to obtain a dataset that maximizes the comparability of income inequality data while maintaining the widest possible coverage across countries and over time. Solt (2009) considers a wide variety of data sources, the most relevant ones being the WIID and the Luxembourg Income Studies (LIS).[8] The goal is to make all other sources "comparable" to LIS data, since the LIS data have earned a reputation as the best available for making cross-national comparisons of income inequality. The main idea of Solt's procedure is to exploit the information of the "duplicates," that is, inequality observations belonging to different categories (i.e., different income definitions and/or reference units) that are available for the same country and year. Whenever available, ratios of these duplicates are computed to obtain "conversion" factors that make it possible to transform inequality data from one category to another. As pointed out by Solt (2009), these ratios can only be computed for those pairs of categories in those countries and years for which they are not particularly useful because data from the two categories already exist. However, he argues that

[7]Its successor, the World Income Inequality Database (WIID), which provides the most comprehensive dataset of income inequality, presents identical shortcomings.

[8]Other data sources are the World Bank's Povcalnet, the Socio-Economic Database for Latin America, Branko Milanovic's World Income Distribution data, and the ILOs Household Income and Expenditure Statistics, as well as data from other national statistical offices, see Solt (2009) for details.

> [T]hose ratios that are directly calculable are valuable nevertheless because they provide information about what the ratios that are missing are likely to be. Because the factors that affect these ratios – redistributive policies, patterns of consumption, and so on– tend to change only slowly over time within a given country, the best prediction for a missing ratio will be based on available data on the same ratio in the same country in proximate years, thereby minimizing any differences in these factors. (p. 236)

Thus, Solt's algorithm treats the unknown ratios as missing data and they are predicted from the results of a series of models (see Solt 2009 for details). The output of this procedure is the Standardized World Income Inequality Data (SWIID), that provides comparable Gini indices of gross and net income inequality for more than 4500 observations, corresponding to 173 countries from 1960 to the present.

3.2 Adjusting the survey data

As mentioned above, the surveys employed to compute our inequality dataset differ in their methodology and income definitions. The impact of these differences might be twofold: they can bias the *level* of the overall Gini coefficient and can bias the *proportions* of the Gini assigned to each of its three components. In the next sections, we describe how the heterogeneity in our inequality measures has been controlled for.

To check whether our data need adjustment, we have first explored whether such biases exist. We have found that there are significant differences across surveys in the level of the Gini coefficient: with the exception of the HES, all other surveys seem to underestimate inequality. Thus, we have applied a procedure similar to that in Solt (2009) to make the Gini observations comparable. On the other hand, the proportions assigned to each of the Gini components seem to be much more stable across surveys and, in general, do not seem to display systematic biases. In what follows, we provide more details about the adjustments made to obtain comparable measures. These "adjusted" data will be used for our empirical analysis.

3.2.1 Adjusting the level of the Gini coefficient

Since data on the Gini coefficient is available from external sources, we have checked how the Ginis obtained from the different surveys relate to existing datasets. To establish the comparison, we have focused on data from the SWIID since, as mentioned above, this dataset addresses the comparability issue in a more convincing way. Our survey Ginis are on average smaller than SWIID net Ginis (averages equal 0.30 and 0.38, respectively) and the correlation coefficient is around 40%. However, the relation between these datasets varies considerably across surveys. With the exception of the HES observations, all other surveys underestimate inequality on average. The largest discrepancies are found for the Afrobarometer (whose observations are 40% smaller on average than those in SWIID), but they are also large for the remaining surveys (around 20% smaller on average). It is remarkable, however, that the mean and the standard deviation corresponding to HES and SWIID observations are basically identical and the correlation coefficient relating both set of observations is 0.9, confirming the good quality of both sources.[9]

To control for survey heterogeneity, we have followed an approach similar to that employed by Solt (2009). The procedure works as follows: Our goal is to make the survey Ginis comparable to the series of (net income) Gini from SWIID (which are constructed so they are comparable to LIS Ginis). Whenever the SWIID Gini and a Gini from our data are available for the same country and year, we compute their ratio. These allows us to compute 211 ratios. Since the total number of surveys is 234, 23 ratios are missing. Next, as in Solt (2009), ratios are predicted by regressing the available ratios on country, time and regional dummies. These regressions provide both estimated ratios and their associated standard errors. Next, the "adjusted" Ginis are obtained as the product of the original survey Ginis and the predicted ratios obtained before. Standard errors for the adjusted data are also obtained. The resulting series of adjusted Ginis is very similar to the original ones from SWIID, with a correlation coefficient between the two datasets equal to 0.97.

3.2.2 Adjusting the components of the Gini coefficient

The next step is to explore whether the shares of the Gini assigned to each of its components are comparable across surveys. Unfortunately, in this case we don't have external data to establish a comparison. However, we do have observations for the same country and year (and/or neighboring years) coming from different surveys.[10] Since proportions are likely to move slowly over time, it is possible to exploit this overlap of surveys to check whether proportions differ systematically across

[9]Additionally, we have regressed the survey Ginis on survey dummies (excluding HES) and some additional variables –time and regional dummies, and time and country dummies–. In all cases, the survey dummies are negative and highly significant, which confirms that observations coming from surveys different from HES systematically underestimate inequality.

[10]More than half of the countries in our sample have inequality measures coming from more than one survey

them.

Correlations between proportions from different surveys for the same country are very high. For instance, the correlation between WGI (BGI) shares from HES and WVS shares is 0.94 (0.93). Correlations are also high between WVS and the more 'problematic' surveys (the Afrobarometer and DHS), with correlations above .80 in all cases.[11]

The average values of the shares differ significantly across surveys. For instance, the average WGI share of Afrobarometer observations (0.32) is half that of HES or WVS (0.64 in both cases). However, these differences are due to regional rather than survey effects. To see this, we have regressed WGI (BGI) shares on survey and time dummies; survey, time and regional dummies; and survey, time and country dummies. Although in the first set of regressions the survey dummies are significant, they cease to be so when regional or country dummies are introduced in the regression. This suggests that the proportions of WGI and BGI do not present systematic upward or downward biases related to the survey type.

A final exercise that we have carried out is as follows. We know that some of our surveys provide high quality income inequality data (HES and, to a lesser extent, WVS) whereas others are more problematic (Afrobarometer, DHS). So, we have first excluded from the sample the "good" surveys and have predicted them by means of a regression of the remaining Gini shares on the fractionalization index, time, survey and regional dummies. The results are encouraging: the predicted shares and the true ones are very similar. For instance, when HES (WVS) observations are excluded the correlation between the predicted and true HES (WVS) WGI proportions is 0.94 (0.89). Since this method seems to produce reliable predictions, we have next predicted the observations from DGS and AFRO using the proportions corresponding to the other surveys and a procedure analogous to the one described above. Interestingly, the resulting predicted proportions for DGS and the Afrobarometer are also very similar to the ones obtained from the surveys, with correlation coefficients larger than 0.9.

The analysis above therefore suggests that the proportions are quite stable across surveys and do not suffer from systematic biases. That is, while a number of the surveys produce estimates of the overall level of inequality that tend to underestimate the true Gini, they seem to produce

[11]Correlations between HES and DHS or Afrobarometer observations cannot be computed because there are not enough observations.

Table 1: Descriptive statistics for components of the Gini decomposition in 89 countries

	Mean	Std. Dev.	Min.	Max.
BGI	.062	.063	.000	.374
WGI	.201	.089	.030	.424
Overlap	.122	.093	.004	.377

reasonable estimates of the proportion of inequality that can be attributed to each component.

Thus, in our baseline analysis, we have computed the 'adjusted" WGI, BGI and O by multiplying the original proportions (i.e, the original values of WGI, BGI and O over the original Ginis) by the "adjusted" Ginis obtained in the previous section. For countries with more than one survey, we take the average of the adjusted values to obtain a single measure of the Gini decomposition for each country. For robustness, Section (?) presents results where the inequality data has been

3.3 The Gini decomposition in 89 countries.

Table 1 presents basic descriptive statistics for the three components of the Gini using the adjusted data. The survey data, unlike data using geographic location of groups, allows us to understand what proportion of inequality is attributable on average to the three different aspects of inequality. We find that within-group economic differences are, on average, the largest component of the Gini. Indeed, the average of WGI is larger than the sum of the BGI and Overlap averages, and WGI is the largest component of the Gini in 70 percent of countries. Overlap is the next largest component of the Gini, but is on average only slightly larger than half the average of WGI. Overlap is the largest component in 29 percent of countries. Between-group inequality is the smallest component of the Gini, and on average is less than one-third the size of WGI. It is the largest component of the Gini in only one country, Cameroon.

Next consider the relationship between each of the components of Gini. These components will obviously be correlated with ELF, the definition of which is essentially a part of BGI and WGI. When there is one large dominant group, for example, almost all inequality must be within-group. Thus, to understand the relationship between the components of Gini, we need to control for ELF. To this end, we regressed each of the three components on ELF and calculated the residual. The plots of these residuals are presented in Figure 3. The top panel presents the relationship between BGI and WGI, controlling for ELF. We can see there is essentially no relationship between the two variables (the correlation is .14), and that at any level of WGI there is considerable variation in BGI (or vice versa). The same null relationship is found in the middle panel, which plots WGI against Overlap (controlling for ELF). Here the correlation is only .09. The strongest relationship we find is in the bottom panel, where Overlap and BGI are correlated at .37. But this correlation is largely influenced by South Africa, which is the strong outlier in the top right corner. With South Africa removed, the correlation is a more modest .25. Thus, though the three components of the Gini could be inter-related conceptually, they need not be, and we find that empirically they are not. We can therefore estimate the effects of each component of the Gini on civil conflict without concern that the estimated effect for any particular component is actually capturing the effect of some other component.

4 Baseline results

We now turn to the main purpose of our paper, which is to assess the relationship between the three components of the Gini and civil conflict. Our data on civil conflict is taken from the UCDP/PRIO dataset.[12] For the tests with the short time period, we consider three variables that tap the incidence of conflict. PRIO25 is an indicator variable that takes the value 1 in a country-year if there was a conflict with 25 or more battle deaths in that year.[13] Since the threshold of conflict is rather low, this measure contains conflicts of quite heterogeneous intensities, from low intensity ones to full scale civil wars. Using data on battle-related deaths (Lacina and Gleditsch, 2005), we have constructed a variable that reflects conflict intensity. PRIO-INT takes 6 values (from 0 to 5), where 0 and 5 correspond to country/year observations where battle deaths are less than 25 and larger than 2000 respectively.[14] PRIOCW is a measure of intermediate conflict that takes the value 1 in

[12]This is a joint dataset of the Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University, and the Centre for the Study of Civil War at the International Peace Research Institute, Oslo (PRIO). It is available at http://www.prio.no/Data/. See Gleditsch et al. (2002) for a presentation of the dataset and the relevant definitions.

[13]More specifically, PRIO's armed conflict definition is as follows: it is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths per year and per incompatibility. We consider only types 3 and 4 (internal armed conflict)

[14]The intermediate values are assigned as follows: PRIO-INT is 1 if battle deaths in a given country/year are more than 25 but less than 100, 2 if deaths are in (100, 500) interval, 3 if deaths are between 500 and 1000 and 4 if battle

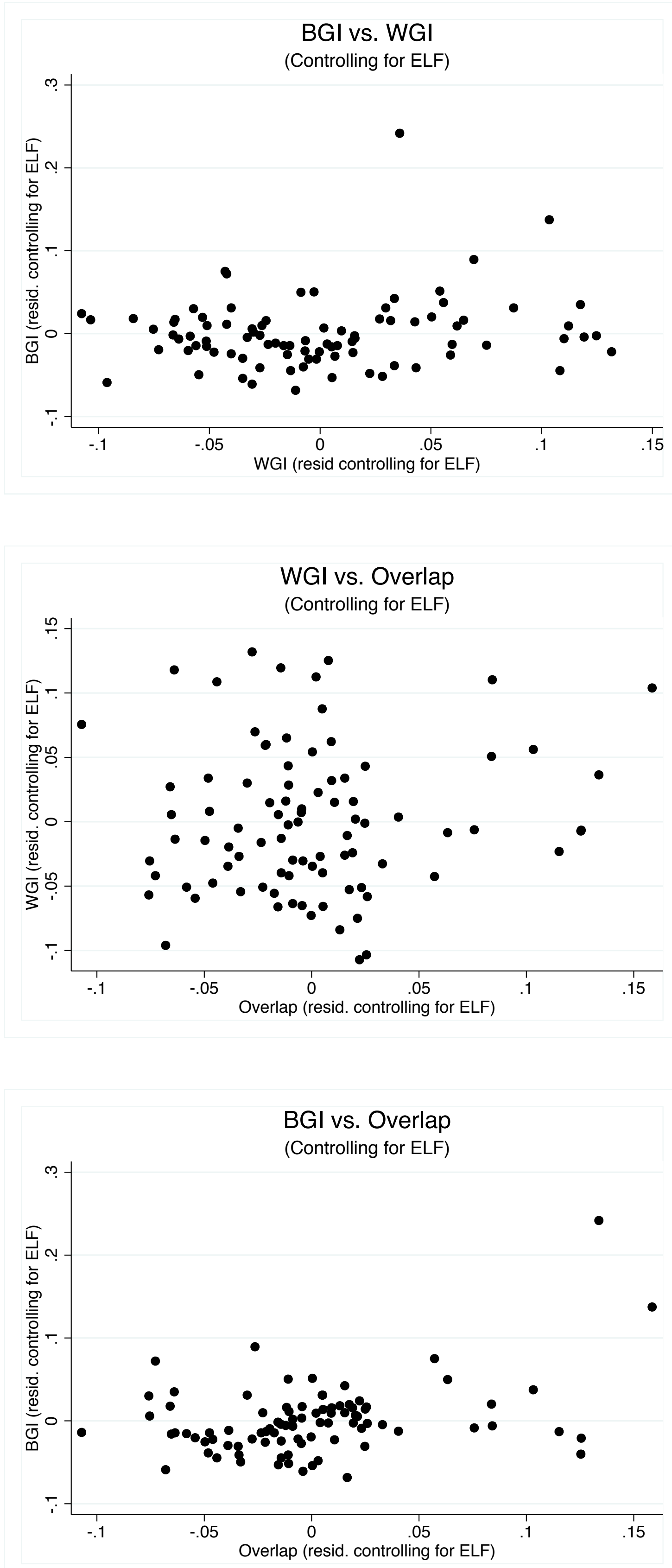

Figure 3: Relationships between the three components of the Gini

a country-year if there are at least 25 deaths and if the aggregate level of deaths for the conflict exceeds 1,000. For the longer time period, we add PRIO1000, which is a measure of high intensity conflict or war. It takes the value 1 if there are at least 1,000 conflict-related deaths in a year. There are too few such conflicts in the shorter time period to make possible meaningful analysis. We estimate logit (or ordered logit for PRIO-INT) regressions with robust standard errors clustered by country. All models include a lagged dependent variable, regional indicator variables and year indicator variables.

The statistical models include a number of standard controls. A table containing summary statistics of all the variables is presented in Appendix B.

- ELF is the standard measure of ethno-linguistic fractionalization, as measured by Fearon (2003).
- EP is the Esteban and Ray (1994) polarization index with binary distances (Reynal-Querol, 2002). It is defined as $EP = 4\sum_{n=i}^{N} ni^2(1-ni)$, where N is the total number of ethnic groups and n_i is the relative size of group i. Data on n_i comes from Fearon (2003).
- NON-CONTIG is an indicator variable taking the value 1 in countries with territory holding at least 10,000 people and separated from the land area containing the capital city either by land or by 100 kilometers of water, as measured in Fearon and Laitin (2003b).
- MOUNTAINOUS is the percent of the country that is mountainous terrain, as measured by Fearon and Laitin (2003b), who use the codings of geographer A. J. Gerard.
- GDP, LAG is the log of real GDP per capita, lagged one year. It is taken from the Penn World Tables (2011).
- POPULATION, LAG is the log of the population in millions, lagged one year, as reported by the Penn World Tables (2011).
- POLITY is the democracy score from Polity IV (2011). It ranges from -10 (full autocracy) to 10 (full democracy).

deaths are larger than 1000 but smaller than 2000.

- NAT. RESOURCES is an indicator variable that takes the value 1 if the country is rich in oil or produces (any positive quantity of) diamonds. A country is rich in oil if the average value of its oil production in a period is larger than 100 US dollars in 2000 constant dollars. The source is Ross (2011).

Table 2 presents results from the post-1994 period (which most closely corresponds to the period in which the surveys were conducted) when PRIO25 is the dependent variable. Columns 1 to 4 contain the control variables discussed above, including the regional and year indicator variables (which are not reported due to space constraints). Our discussion of the results will focus on the inequality variables. In column 1, the Gini coefficient is the only inequality variable that is included. Though it has a positive coefficient, consistent with much previous research, it is estimated with considerable error (with a standard error almost as large as the coefficient). Column 2 considers the effect of between-group inequality. Since BGI is a component of the Gini coefficient, we can estimate the coefficient and standard error of BGI by subtracting BGI from Gini and including this new variable,"Gini-BGI," on the right-hand side (instead of Gini, itself). The coefficient on Gini-BGI therefore estimates the effect of all inequality unrelated to BGI on conflict, and the coefficient on BGI estimates the effect only of BGI (and not of BGI through Gini). We find that the effect of Gini (unrelated to BGI) remains positive but insignificant, and that the coefficient on BGI is positive but also non-significant since it is measured with substantial error. Column 3 includes WGI on the right-hand side (and includes Gini-WGI as the control for non-WGI inequality). We find that WGI has a positive and significant coefficient. The coefficient on Gini-WGI has a negative sign but it is not precisely estimated. Column 4 includes all three components of the Gini separately. The coefficient of WGI remains very similar as in the previous case. Interestingly, Overlap has a negative and significant coefficient whereas BGI still presents a positive but non-significant one.

Column 4 provides an explanation for the lack of statistical relation between inequality and conflict. It highlights the fact that the different components of the Gini are related to conflict in opposite ways. Whereas more inequality *within* the group is positively related to conflict, more economically segregated groups (i.e., those with lower Overlap) tend to fight more. These effects cancel out when considered aggregated in the Gini coefficient. This explains why, despite the fact that inequality and conflict are so intuitively related, it is difficult to find a robust empirical relation between them.

Columns 5-6 explore some alternative specifications of the control variables. In each of models 1-4, the coefficient for Polity is estimated with considerable error. Fearon and Laitin (2003) have argued against using a linear scale for regime, and instead for dividing countries into autocracies (Polity$<$-5), anocracies (-5$\leq$Polity$\leq$5) and democracies (Polity $>$ 5). Model 5 therefore drops the Polity variable and instead includes indicator variables for anocracy and democracy. This specification is more informative than simply including Polity: we find that the coefficients for both variables are positive and estimated with reasonable precision, suggesting civil conflict is more likely in anocracies and democracies than in autocracies. The results for the inequality variables are unaffected by this change. Finally, the results for one control variable – Natural Resources – are consistently at odds with existing research (with a negative but very imprecisely estimated coefficient). Model 6 therefore examines whether the results for the inequality variables are robust to the exclusion of this variable. We find this to be the case.

ADD EXPLANATION ABOUT HOW TO INTERPRET THE RESULTS: WE ARE LOOKING AT THE CROSS-COUNTRY DISTRIBUTION, NOT AT CHANGES WITHIN COUNTRIES. THEREFORE, WE CAN MOVE ONE OF THE COMPONENTS, LET'S SAY WITHIN, WITHOUT MOVING THE REST...WE MIGHT NEED EXAMPLES OF PARTICULAR COuNTRIES HERE IN LINE WITH OUR EMPIRICAL RESULTS.

The models in Table 2 show nonsignificant results for BGI and robust and significant results for the other two components of the Gini coefficient, within-group inequality and Overall. Interestingly, all these findings are consistent with the theoretical predictions, as explained in Section **??**. How large are these effects?[15] To address this question, we focus on Model 5.[16] Our estimated coefficients imply that an increase in WGI of one standard deviation from its mean (i.e. from 0.18 to 0.26) is associated with an overall increase in the probability of conflict of almost 10 percentages points (from 10.8% to 20%) which amounts to an increase of 86% in the initial probability of

[15]To interpret the size of these effects notice that we are identifying the coefficients by using the cross-country distribution of the inequality components. Thus, although for a particular country it is difficult to move one of the components leaving the other two constant (since changes in the income distribution of one of the groups will most likely affect the three components), it is perfectly possible to do so when comparing these measures across countries. We have many examples in our dataset where countries possess very similar values for two of the inequality components and a very different one for the third. For instance, Estonia and Peru present similar values for Overlap and WGI but the value of BGI is 10 times larger in Peru.

[16]Similar results are obtained if other specifications are employed.

Table 2: Baseline

	(1)	(2)	(3)	(4)	(5)	(6)
Gini	2.108					
	(1.693)					
Gini-BGI		0.152				
		(2.305)				
Gini-WGI			-2.661			
			(2.698)			
BGI		5.247		1.688	2.295	2.040
		(3.497)		(3.551)	(3.567)	(3.503)
WGI			12.918**	13.732**	13.721***	13.439***
			(5.190)	(5.340)	(5.281)	(5.111)
Overlap				-8.195*	-8.299*	-8.282**
				(4.256)	(4.344)	(4.147)
ELF	2.505**	1.703	8.620***	9.604***	9.186***	9.076***
	(1.086)	(1.076)	(2.940)	(3.054)	(2.957)	(2.798)
EP	1.265	1.609*	1.396	1.979**	2.093**	2.076**
	(0.900)	(0.888)	(0.916)	(0.882)	(0.851)	(0.884)
Non-contig.	0.973	1.319**	0.847	1.606***	1.679***	1.657***
	(0.614)	(0.608)	(0.635)	(0.613)	(0.629)	(0.610)
Mountainous	0.011	0.010	0.011	0.009	0.008	0.012
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)
GDP, lag	-0.398*	-0.468**	-0.290	-0.358	-0.365	-0.427**
	(0.231)	(0.227)	(0.231)	(0.222)	(0.223)	(0.213)
Population, lag	0.444***	0.429***	0.437***	0.368***	0.381***	0.339***
	(0.124)	(0.128)	(0.128)	(0.135)	(0.136)	(0.127)
Polity	0.015	0.020	0.015	0.019		
	(0.029)	(0.029)	(0.029)	(0.029)		
Anocracy					0.956*	0.743
					(0.500)	(0.479)
Democracy					0.884*	0.789*
					(0.497)	(0.477)
Nat. Resources	-0.271	-0.232	-0.323	-0.251	-0.292	
	(0.308)	(0.305)	(0.327)	(0.332)	(0.323)	
PRIO25(lag)	4.448***	4.406***	4.372***	4.268***	4.220***	4.289***
	(0.503)	(0.496)	(0.482)	(0.467)	(0.472)	(0.467)
Constant	-8.368***	-6.928***	-13.182***	-12.193***	-12.883***	-11.918***
	(2.187)	(2.159)	(3.302)	(3.196)	(3.144)	(2.866)
Pseudo R-squared	0.615	0.616	0.619	0.623	0.625	0.628
N	1309	1309	1309	1309	1309	1335

Note: Dependent variable is PRIO25. There are 89 countries. Logit regressions for the period 1995-2009 have been estimated. Robust standard errors (clustered by country) are in parentheses. Regional indicator variables and year indicator variables are included in all models.
* p<.10, ** p<.05, *** p<.01

conflict.[17] This number, however, masks important regional variation. The same increase in WGI is associated with a 21, 18 and 15 percentage point increase in the probability of conflict in Africa, South Asia and Middle East, respectively. However, the increase is small (around 2-3 percentage points) in Latin America or in Europe. As for Overall, its impact on conflict is somewhat smaller. An increase in Overlap of one standard deviation from its mean (i.e. from 0.15 to 0.24) is associated with a decrease in the probability of conflict from 10.9% to 6.51% (a decrease of 40%). Also in this case there is considerable regional variation. The decrease in the probability of conflict is specially remarkable in South Asia and the Middle East (with a decrease of 18 percentage points) but it is small in Europe and Latin America (where the decrease is smaller than 1%)

5 Robustness checks

This section discusses a number of variations we have carried out in order to examine the robustness of the results presented in Tables 2 and 3. For the sake of brevity, all tables in this section are presented in Appendix A and here we only describe the different exercises and the main results.

5.1 Alternative dependent variables

As noted above, PRIO25, the measure of civil conflict used in Table 2, classifies very heterogeneous conflicts within the same category. Since small and large-scale conflicts are conceptually different, we have dropped the former to make sure that they alone do not drive our results. Additionally, we have used a measure that explicitly captures the different intensities of conflict. Table 3 presents similar regressors as Table 2 using PRIOCW (columns 1-3) and PRIO-INT (columns 4-6) as dependent variables. The main conclusions from Table 2 carry through to this table: there is not a significant effect of inequality (as measured by the Gini coefficient) on conflict. However, when the components of the Gini are considered individually, WGI has a positive and significant impact on conflict and Overlap has a negative and significant one. The coefficient of BGI is generally insignificant. (THE TABLE NEEDS TO BE UPDATED, INCLUDE GINI ALONE).

[17] All variables are held at their means but for the regional dummies. For each region, we have computed the rise in the probability of conflict due to the increase of one standard deviation in WGI. The overall probability is computed as the weighted average of the regional probabilities, where the weights are given by the relative frequency of each region.

5.2 Alternative ways of dealing with missing values in inequality measures

In our baseline analysis we use averages of the available inequality observations for all the years from 1995 onwards, which implies that inequality measures are considered to be time-invariant. The reason for doing so is that data on the inequality components has a lot of missing observations, which makes constructing an annual panel difficult. Since inequality observations are likely to evolve slowly in the 15 year period analyzed in this period, this strategy seems reasonable.

For robustness, this section considers an alternative way of constructing the inequality measures. We have generated time-varying components of the Gini coefficient by combining information on time-varying data on the Gini itself, for which there is wide coverage, and the current (or, if missing, the predicted) values of the shares corresponding to each of its components. More specifically, we have proceeded as follows. For each country/year for which a survey is available, we have first obtained the proportion of BGI and WGI in the overall survey Gini, computed as the ratio of BGI (or WGI) over Gini. Next, we have predicted the missing proportions of BGI and WGI by regressing the available ones on the fractionalization index, regional indicator variables and year dummies.[18] Proportions for Overlap have been obtained by sustracting BGI and WGI shares from 1. The fit for WGI is very good, with a correlation coefficient between the predicted and the available shares for WGI of 0.99. This correlation is somewhat smaller for BGI and Overlap (0.74 and 0.88 respectively), suggesting that the predictions might be noisier for these measures. Finally, the time-varying Gini components have been generated by multiplying the Gini coefficient from the SWIID (based on net income) and the available (or, if missing, the predicted) shares of WGI, BGI and Overlap.

Results are presented in Table 4. All inequality measures have been lagged 4 years to avoid reverse causality as much as possible (similar results are obtained if different lags are employed). In column 1 only the time-varying Gini is introduced and, as in Table 2, it has a positive but insignificant coefficient. Column 2 decomposes the Gini coefficient in two components, WGI and the rest. Again, the conclusions are identical as before: GINI-WGI has a negative but insignificant coefficient while WGI has a positive and significant one. Moreover, the size of its coefficient is very similar to the ones obtained in Table 2 where the constant inequality measures were employed.

[18] We have used generalized linear regression and a bernoulli distribution for the dependent variable to make sure that our predicted proportions are between 0 and 1.

Column 3 introduces the three components of the Gini separetely. In this case, it is found that both Overlap and BGI present a negative coefficient, which is significant only for the latter variable. This last result has to be interpret with caution, given that the predictions for BGI shares are the noisiest ones. However, the qualitative conclusions remain intact: the components of the Gini coefficient have effects on conflict that go in opposite directions and thus, they cancel out when considered aggregated.

TWO COMMENTS:

INTRODUCE A COMMENT —WOOLDRIDGE ON GENERATED REGRESSORS. WE COULD INTRODUCE ANOTHER COLUMN WITH GINI AND BETWEEN ONLY

5.3 Dealing with survey heterogeneity

As discussed in Section 3, the surveys used to obtain the inequality measures vary considerably in quality. The most problematic surveys are the Afrobarometer and DHS, since they do not provide household income data, but only information on assets that individuals possess. To check whether our results depend on these observations, we have run similar regressions as those in Table 2, dropping all inequality observations obtained from either DHS or Afrobarometer surveys. As a consequence, 29 countries, mostly African, are dropped from the sample. Table 5 presents the results. Column 1 shows that the Gini coefficient has a positive and significant coefficient in this reduced sample. When the different components of the Gini are introduced (columns 2 and 3), it is seen that the three of them have a positive coefficient, although only that of WGI is significant. To see whether the change in the sign the coefficient of Overlap stems from the fact that we are dropping asset-based inequality measures or from the fact that we are studying a quite different sample of countries, we have predicted the shares of BGI, WGI and Overlap for the 89 countries of our original sample in a similar way as in the previous exercise (Table 4), this time dropping observations from the DHS and Afrobarometer surveys. The results obtained in this case are very similar to those obtained in Table 4 when observations from all the surveys where used to carry out the predictions. While the sign and the significance of WGI is unaffected, the signs of Overlap and BGI become negative (and significant for BGI). Thus, dropping observations from DHS and Afrobarometer surveys doesn't modify our conclusions in a substantive way.

5.4 Extending the time period

Another way to gauge the robustness of these results is to expand the time frame for analysis. The obvious disadvantage of doing this is that we must assume that the components of inequality are stable over a long period when in fact we should expect these components to change, and perhaps to change in response to the outcomes of civil conflict occurring during the extended time period. This is a problem that has similarly plagued the study of other variables for which we do not have good time-varying measures, including ELF and polarization. The advantage is that civil conflict is relatively rare, even using the lowest thresholds in the literature regarding the number of deaths from the violence. Expanding the time frame therefore substantially expands the amount of variation in civil conflict that appears in our data. In addition, using the extended data set allows us to employ PRIO1000 as a dependent variable. The incidence of such conflicts is too rare for meaningful analysis in our post-1995 data.

Table 6 presents results from models using the period 1960-2009. The models use the same specification as those in Table 3. Models 1-3 use PRIO25 as the dependent variable, models 4-6 use PRIOCW as the dependent variable, and models 7-9 use PRIO1000 as the dependent variable. Looking across the models, the coefficient for WGI is always positive and precisely estimated while the coefficient of Overlap is always negative and significant. but is only precisely estimated when PRIOCW is the dependent variable. The coefficient for BGI is positive and precisely estimated only when overlap is not introduced in the regression, but it becomes insignificant when the three inequality components are separately introduced in the regression.

5.5 Other estimation strategies

One major concern of any cross-country analysis is that risk of committing omitted variable bias. Unfortunately,

6 Conclusion

Our data yields several findings regarding inequality, ethnicity and conflict. Most importantly, we have found robust evidence of a positive and substantively large association between within-group inequality and civil war. This relationship is present in different time periods, using different model specifications of the inequality variables, using different measures of civil war, and using different data on group size. The result provides support for Esteban and Ray's (2011) model arguing that the economic diversity of groups should be positively related to conflict. More generally, the result underscores the potential usefulness of developing and testing theories about how the internal characteristics of groups affect the opportunities and incentives they have to engage in conflict. Although ethnic groups obviously play a central role in the civil conflict literature, their internal characteristics are often absent from the debate, particularly in cross-national research.

We have also found evidence of a negative association between Overlap and civil conflict. This result is robust when we use PRIOCW as the measure of conflict, but is less robust when other measures are used. Since Overlap is negatively correlated with the economic segregation of groups, this finding suggests that holding WGI and BGI constant, societies with groups that are economically segregated from each other are most likely to engage in conflict.

The results for Overlap help explain why it has been difficult to find a relationship between general inequality and ethnic conflict. General inequality is the sum of three conceptually distinct dimensions of inequality. Some dimensions – BGI and WGI – are in fact associated with more conflict. But others – Overlap, which increases as groups become more economically integrated – are associated with less conflict. It is therefore not at all clear that the sum of these three components should be related to more conflict, or that we should find a relationship between overall inequality and conflict. This paper therefore underscores the value of examining different dimensions of inequality rather than the sum of these dimensions.

Finally, the results for BGI are the weakest of the three group-based inequality variables. While the coefficient for BGI is positive and significant in a handful of regressions, it is generally estimated with considerable error. This stands in contrast to the findings of Cederman et al. (2011), who find a significant relationship between group inequality and conflict using geo-coded measures of group well-being.

There are a number of potential explanations for the different results in the two studies. One, argued above, is that "grievance theory" is itself conceptually incomplete and unconvincing – that is, there are good reasons to expect the null relationship that we find. A second possibility is that as an empirical matter, between-group economic differences are a small enough proportion of total inequality that they seldom create sufficient incentives for civil war. A third possibility is that the differences in the results stem from different methods, particularly in the measurement of income. The Cederman et al. paper assumes that individuals from different groups live in different areas, and that the economic well-being of these areas can thus be attributed to groups. We use surveys. There are trade-offs associated with both approaches, and an important task for future research is to understand and explain them. A fourth explanation might be that arguments about inter-group grievances are largely correct, but that they are better tested using the economic segregation of groups rather than the differences in group mean incomes. If this is true, then the results for the Overlap component of inequality might be interpreted as support for grievance-based arguments. This possibility underscores the need for more precise theorizing about the circumstances under which particular types of economic differences across groups lead to conflict.

7 References

Acemoglu, Daron, and James Robison. 2005. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

Baldwin, Kate, and John D. Huber. 2010. "Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision." *American Political Science Review*, 104(4):644–662.

Cederman, Lars-Erik, Nils B. Weidmann, and Kristian S. Gleditsch. 2011. "Horizontal Inequalities and Ethnonationalist Civil War: A Global Comparison." *American Political Science Review*, 105(3): 478–495.

Collier, Paul, and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economics Papers*, 56(4): 563–595.

Condra, Luke. 2012. "Perils of the Periphery Explaining African Ethnic Group Rebellion, 1980-2006." *Mimeo*.

Deininger, Klaus, and Lyn Squire. 1996. "A New Data Set Measuring Income Inequality." *The World Bank Economic Review*, 10(3): 565–591.

Esteban, Joan, and Debraj Ray. 2011. "A Model of Ethnic Conflict." *Journal of the European Economic Association*, 9(3):496–521.

Esteban, Joan, Laura Mayoral, and Debraj Ray. 2012. " Ethnicity and Conflict: An empirical Study." *American Economic Review*, 102(4): 1310–42.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8(2): 195–222.

Fearon, James D., and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review*, 97(1): 75–90.

Filmer, Deon, and Lant H. Pritchett. 2001. "Estimating Wealth Effects Without Expenditure Data-Or Tears: An Application to Educational Enrollments in States of India." *Demography*, 38(1):115–132.

Forbes, H.D. 1997. *Ethnic Conflict: Commerce, Culture, and the Contact Hypothesis*. New Haven, CT: Yale University Press.

Forbes, H.D. 2004. "Ethnic conflict and the contact hypothesis." In *The Psychology of Ethnic and Cultural Conflict*, edited by Yueh-Ting Lee, Clark McCauley, Fathali Moghaddam and Stephen

Worchel. Westport, CT: Praeger Publishers.

Glaeser, Edward L. 2005. “The Political Economy of Hatred.” *Quarterly Journal of Economics*, 120(1): 4586.

Gleditsch, Nils P., Peter Wallensteen, Mikael Eriksson, Margareta Sollenber, and Håvard Strand. 2002. “Armed Conflict 1946-2001: A New Dataset.” *Journal of Peace Research*, 39(5): 615–637 (accessed October 1, 2010).

Huber, John D., Thomas K. Orgazalek, and Radhika Gore. 2012. “Democracy, targeted redistribution and ethnic inequality.” Typescript, Columbia University.

Lacina, Bethany and Nils P. Gleditsch, 2005. 2012. “Monitoring Trends in Global Combat: A New Dataset of Battle Deaths.” *European Journal of Population*, 21(23): 145166.

Lambert, Peter J., and J. Richard Aronson. 1993. “Inequality Decomposition Analysis and the Gini Coefficient Revisited.” *The Economic Journal*, 103(420): 1221–1227.

Lambert, Peter J., and André Decoster. 2005. “The Gini coefficient reveals more.” *Metron*, LXIII(3): 373–400.

Lichbach, Mark I. 1989. “An Evaluation of ‘Does Economic Inequality Breed Political Conflict?’ Studies.” *World Politics*, 41(4): 431–470.

McKenzie, David J. 2005. “Measuring Inequality with Asset Indicators.” *Journal of Population Economics*, 18(2):229–260.

Montalvo, José G., and Marta Reynal-Querol. 2005. “Ethnic Polarization, Potential Conflict and Civil War.” *American Economic Review*, 95(3): 796–816.

Pettigrew , Thomas F., and Linda R. Tropp. 2006. “A Meta-Analytic Test of Intergroup Contact Theory.” *Journal of Personality and Social Psychology*, 90(5): 751–783.

Polity IV. “Polity IV Project: Political Regime Characteristics and Transitions, 1800-2009, ” (accessed October 1, 2011).

http://www.systemicpeace.org/polity/polity4.htm.

Pyatt, Graham. 1976. “On the Interpretation and Disaggregation of Gini Coefficients.” *The Economic Journal*, 86(342): 243–255.

Reynal-Querol, Marta. 2002. “Ethnicity, Political Systems,and Civil Wars.” *Journal of Conflict Resolution*, 46(1), pp. 2954.

Ross, Michael. 2006. “A Closer Look at Oil, Diamonds and Civil War.” *Annual Review of Political*

Science, 9(1): 265–300.

Ross, Michael. 2011. "Replication data for: Oil and Gas Production and Value, 1932-2009." (accessed May 1, 2011).

http://hdl.handle.net/1902.1/15828UNF:5:Hwe3jAjxG7fgOMzpGQXOxw==V4[Version].

Sambanis, Nicholas. 2001. "What Is An Ethnic War? Organization and Interests in Insurgencies." *Mimeo*.

Sen, Amartya. 1973. *On Economic Inequality*. Clarendon Press, Oxford.

Solt, Frederik. 2009. "Standardizing the World Income Inequality Database." *Social Science Quarterly*, 90(2): 231–242.

Stewart, Frances. 2002. "Horizontal Inequalities: A Neglected Dimension of Development." *Annual Lecture No. 5, UNU World Institute for Development Economics Research.*

Stewart, Frances. 2008. "Horizontal Inequalities and Conflict: An Introduction and Some Hypotheses." In *Horizontal Inequalities and Conflict*, ed. Frances Stewart. New York: Palgrave Macmillan.

Wintrobe, Ronald. 1995. "Some Economics of Ethnic Capital Formation and Conflict." In *Nationalism and Rationality*, edited by Albert Breton, Gianluigi Galeotti, and Ronald Wintrobe: 43–70.

Yitzhaki, Shlomo, and Robert I. Lerman. 1991. 'Income Stratification and Income Inequality." *Review of Income and Wealth*, 37(3): 313–329.

Yitzhaki, Shlomo. 1994. "Economic Distance and Overlapping of Distributions." *Journal of Econometrics*, 61(1): 147–159.

8 Appendix A

This section contains the output of several variations aimed to check the robustness of our main results. Detailed descriptions are provided in Section **??**

9 Appendix B

This Appendix presents the summary statistics of the data employed in our empirical exercise as well as a list of all the surveys employed in the analysis.

Table 3: Alternative dependent variables

	(1)	(2)	(3)	(4)	(5)	(6)
Gini-BGI	-2.200			-0.529		
	(2.906)			(1.946)		
Gini-WGI		-10.829***			-2.039	
		(3.714)			(2.830)	
BGI	6.128		-2.791	7.360**		3.683
	(3.889)		(4.355)	(3.151)		(3.305)
WGI		22.984***	27.400***		10.234**	9.897**
		(7.023)	(7.055)		(4.899)	(5.016)
Overlap			-24.877***			-7.021*
			(5.769)			(4.045)
ELF	1.649	16.069***	20.186***	0.022	5.875**	6.017**
	(1.367)	(3.986)	(4.111)	(0.844)	(2.885)	(2.951)
EP	1.067	0.703	1.847	2.334***	1.798**	2.381***
	(1.030)	(1.135)	(1.131)	(0.862)	(0.844)	(0.803)
Non-contig.	1.480**	0.722	2.729***	0.978*	0.300	1.164**
	(0.680)	(0.713)	(0.802)	(0.561)	(0.558)	(0.540)
Mountainous	0.009	0.013	0.010	0.005	0.006	0.004
	(0.010)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
GDP, lag	-0.239	0.133	0.008	-0.411**	-0.261	-0.299
	(0.264)	(0.220)	(0.212)	(0.207)	(0.210)	(0.199)
Population, lag	0.621***	0.637***	0.437**	0.419***	0.441***	0.347**
	(0.170)	(0.173)	(0.171)	(0.145)	(0.143)	(0.145)
Anocracy	1.076	1.354	1.530	0.807**	0.795*	0.785*
	(0.915)	(1.016)	(1.152)	(0.410)	(0.427)	(0.425)
Democracy	0.880	1.022	1.320	0.402	0.302	0.341
	(0.787)	(0.899)	(1.078)	(0.385)	(0.414)	(0.404)
Nat. Resources	-0.534	-0.752*	-0.716*	-0.600**	-0.679**	-0.621**
	(0.338)	(0.384)	(0.365)	(0.284)	(0.302)	(0.310)
PRIOCW(lag)	5.621***	5.462***	5.202***			
	(0.567)	(0.553)	(0.555)			
Num. Deaths(lag)				1.832***	1.819***	1.794***
				(0.203)	(0.202)	(0.199)
Constant	-11.035***	-23.632***	-22.538***			
	(2.851)	(4.315)	(4.132)			
Pseudo R-squared	0.742	0.749	0.757	0.456	0.456	0.460
N	1309	1309	1309	1309	1309	1309

Note: NEEDS UPDATE!! Dependent variable is PRIOCW for columns 1–3 and PRIO-INT for columns 4–6. There are 89 countries. Logit and Ordered logit regressions were estimated for columns 1–3 and 4–6 respectively.
Robust standard errors (clustered by country) are in parentheses. Regional and year indicator variables are included in all models.
* p<.10, ** p<.05, *** p<.01

Table 4: Time-Varying inequality measures

	(1)	(2)	(3)
Gini_{t-4}	3.286		
	(2.958)		
Gini-WGI_{t-4}		-4.938	
		(3.397)	
WGI_{t-4}		13.231***	14.770***
		(4.609)	(5.286)
BGI_{t-4}			-10.594*
			(6.263)
Overlap_{t-4}			-2.983
			(3.601)
ELF	2.783*	9.107***	10.247***
	(1.463)	(2.526)	(3.036)
EP	0.948	1.983	1.553
	(1.231)	(1.212)	(1.266)
Non-contig.	0.856	1.235	1.410*
	(0.826)	(0.795)	(0.779)
Mountainous	0.011	0.012	0.012
	(0.012)	(0.010)	(0.010)
GDP, lag	-0.369	-0.248	-0.182
	(0.357)	(0.349)	(0.361)
Population, lag	0.547***	0.369**	0.364**
	(0.161)	(0.154)	(0.153)
Anocracy	0.772	0.664	0.652
	(0.719)	(0.689)	(0.685)
Democracy	0.959	1.107*	1.051
	(0.632)	(0.655)	(0.658)
Nat. Resources	0.376	-0.011	-0.027
	(0.411)	(0.452)	(0.450)
PRIO25(lag)	4.863***	4.844***	4.879***
	(0.731)	(0.597)	(0.601)
Constant	-11.108***	-14.485***	-15.485***
	(3.634)	(3.918)	(4.413)
Pseudo R-squared	0.666	0.668	0.669
N	726	886	886

Note: Dependent variable is PRIO25. Robust standard errors clustered at the country level are in parentheses. Regional and year indicator variables are included in all models. Gini_t is (net income) Gini from SWIID. Missing observations of WGI_t, BGI_t and Overlap_t have been predicted as described in the text. All inequality measures are lagged four years.
* p<.10, ** p<.05, *** p<.01

Table 5: Dropping Afrobarometer and DHS observations

	(1)	(2)	(3)	(4)	(5)
Gini	13.123**				
	(6.516)				
Gini-WGI		9.196			
		(11.111)			
WGI		16.921**	16.971**		
		(7.556)	(7.677)		
BGI			8.419		
			(10.582)		
Overlap			5.741		
			(13.797)		
Gini$_{t-4}$				3.286	
				(2.958)	
Gini-WGI$_{t-4}$			-8.588*		
				(4.544)	
WGI$_{t-4}$			21.953**	20.471**	
				(9.415)	(9.057)
BGI$_{t-4}$					-15.298*
					(8.127)
Overlap$_{t-4}$					-3.403
					(4.720)
ELF	2.374	5.496	6.480	13.802***	13.560***
	(2.016)	(6.637)	(7.096)	(5.248)	(5.079)
EP	2.905*	2.560	2.721	2.255*	1.297
	(1.612)	(1.864)	(2.071)	(1.217)	(1.373)
Non-contig.	0.633	0.621	0.856	1.484*	1.341*
	(0.792)	(0.797)	(0.861)	(0.806)	(0.806)
Mountainous	0.015	0.016	0.015	0.011	0.011
	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
GDP, lag	0.477	0.570*	0.535*	-0.208	-0.181
	(0.291)	(0.311)	(0.324)	(0.336)	(0.334)
Population, lag	0.388*	0.371*	0.368*	0.343**	0.355**
	(0.207)	(0.195)	(0.199)	(0.154)	(0.156)
Anocracy	0.652	0.716	0.759	0.610	0.668
	(0.705)	(0.727)	(0.751)	(0.680)	(0.690)
Democracy	0.237	0.302	0.349	1.038	1.127*
	(0.674)	(0.677)	(0.699)	(0.650)	(0.671)
Nat. Resources	-0.449	-0.529	-0.560	0.002	-0.056
	(0.454)	(0.448)	(0.440)	(0.443)	(0.463)
PRIO25(lag)	4.911***	4.915***	4.905***	4.789***	4.863***
	(0.660)	(0.664)	(0.667)	(0.580)	(0.609)
Constant	-23.090***	-24.561***	-24.082***	-17.871***	-16.966***
	(5.377)	(5.086)	(4.934)	(5.896)	(5.708)
Pseudo R-squared	0.723	0.723	0.724	0.671	0.673
Num. Countries	60	60	60	89	89
N	884	884	884	886	886

Note: Dependent variable is PRIO25. Robust standard errors clustered at the country level are in parentheses. Regional and year indicator variables are included in all models. Afrobarometer and DHS observations have been dropped in Columns (1)-(3). Time varying inequality measures, lagged 4 years, are used in columns(4)-(5). Afrobarometer and DHS observations have been dropped prior to predicting the missing observations of WGI$_t$, BGI$_t$ and Overlap$_t$, see the text for details. * p<.10, ** p<.05, *** p<.01

Table 6: Alternative time period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gini-BGI	0.835			-0.623			-0.487		
	(1.858)			(2.233)			(2.396)		
Gini-WGI		-1.519			-5.061**			-1.256	
		(1.955)			(2.338)			(2.404)	
BGI	6.009***		2.844	6.889**		1.192	3.421		1.468
	(2.270)		(2.123)	(2.954)		(2.629)	(2.625)		(2.947)
WGI		13.815***	14.421***		19.784***	21.631***		7.958*	8.985*
		(4.121)	(4.057)		(5.147)	(5.033)		(4.698)	(4.793)
Overlap			-8.442**			-16.820***			-6.706*
			(3.281)			(3.922)			(3.492)
ELF	0.333	6.893***	8.057***	0.559	11.076***	13.619***	-0.163	3.966	5.283**
	(0.657)	(2.249)	(2.232)	(0.787)	(2.721)	(2.709)	(0.962)	(2.549)	(2.565)
Eth. Polarization	1.479**	1.136	1.548**	1.545*	0.994	1.724**	-0.426	-0.678	-0.161
	(0.723)	(0.692)	(0.670)	(0.882)	(0.851)	(0.861)	(1.105)	(1.033)	(1.082)
Non-contig.	1.060**	0.670	1.370***	1.113***	0.542	1.803***	1.654***	1.333***	1.884***
	(0.428)	(0.491)	(0.421)	(0.424)	(0.505)	(0.453)	(0.436)	(0.361)	(0.462)
Mountainous	0.007	0.009	0.005	0.007	0.012	0.004	0.008	0.011	0.006
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)
GDP, lag	-0.443***	-0.279*	-0.310*	-0.357*	-0.077	-0.112	-0.209	-0.097	-0.104
	(0.164)	(0.156)	(0.159)	(0.209)	(0.192)	(0.193)	(0.250)	(0.260)	(0.267)
Population, lag	0.298***	0.326***	0.265**	0.351***	0.391***	0.301**	0.560***	0.577***	0.529***
	(0.106)	(0.104)	(0.110)	(0.131)	(0.130)	(0.136)	(0.168)	(0.165)	(0.175)
Anocracy	0.699***	0.849***	0.752***	0.753***	0.976***	0.845***	0.731*	0.818*	0.726*
	(0.183)	(0.182)	(0.181)	(0.218)	(0.222)	(0.224)	(0.424)	(0.436)	(0.421)
Democracy	0.593**	0.597**	0.577**	0.516	0.456	0.401	-0.253	-0.257	-0.285
	(0.287)	(0.284)	(0.294)	(0.363)	(0.340)	(0.363)	(0.568)	(0.575)	(0.553)
Nat. Resources	-0.663***	-0.774***	-0.734***	-0.859**	-1.047***	-1.107***	-0.890**	-0.969**	-1.016**
	(0.248)	(0.246)	(0.255)	(0.347)	(0.354)	(0.359)	(0.409)	(0.420)	(0.405)
PRIO25(lag)	4.690***	4.661***	4.550***						
	(0.316)	(0.319)	(0.303)						
PRIOCW(lag)				6.195***	6.121***	5.967***			
				(0.416)	(0.423)	(0.403)			
PRIO1000(lag)							4.777***	4.767***	4.739***
							(0.395)	(0.388)	(0.400)
Constant	-5.479***	-11.800***	-10.733***	-7.179***	-17.221***	-16.143***	-9.312***	-13.503***	-13.100***
	(1.482)	(2.100)	(2.046)	(1.800)	(2.693)	(2.567)	(2.275)	(2.896)	(2.864)
Pseudo R-squared	0.620	0.623	0.627	0.738	0.742	0.748	0.561	0.562	0.564
N	3512	3512	3512	3512	3512	3512	3186	3186	3186

Note: Dependent variables are PRIO25 in columns 1-3, PRIOCW in columns 4-6 and PRIO1000 in columns 7-9. There are 89 countries. Time period is 1960-2009. Robust standard errors (clustered by country) are in parentheses. Regional indicator variables and year indicator variables are included in all columns. * p<.10, ** p<.05, *** p<.01

Table 7: Summary statistics

	Mean	Std Dev.	Min	Max
PRIO25	0.17	0.375	0	1
PRIOCW	0.141	0.348	0	1
PRIO1000	0.031	0.174	0	1
PRIO-INT	0.405	1.026	0	5
Gini	0.390	0.115	0.189	0.889
BGI	0.062	0.06	0.001	0.383
WGI	0.178	0.082	0.011	0.378
Overlap	0.151	0.094	0.011	0.403
Gini$_t$	0.376	0.093	0.186	0.666
BGI$_t$	0.058	0.053	0	0.509
WGI$_t$	0.187	0.076	0.023	0.391
Overlap$_t$	0.132	0.082	0.01	0.409
ELF	0.508	0.243	0.077	0.953
Eth. Polarization	0.579	0.201	0.154	0.986
Non-contig.	0.144	0.352	0	1
Mountainous	16.155	19.754	0	81
log(GDP)	8.432	1.375	4.764	10.82
log(Population)	9.579	1.38	6.624	13.961
Polity	4.554	5.627	-9	10
Anocracy	0.304	0.46	0	1
Democracy	0.604	0.489	0	1
Nat. Resources	0.286	0.452	0	1

newpage

Table 8: Inequality Surveys

EUROPE			
Albania	2002(WVS) 2005(HES)	Ireland	1999(WVS)
Austria	2000(LIS)	Latvia	1996(WVS) 1999(WVS)
Belarus	1996(WVS) 2001(CSES)	Lithuania	1997(CSES,WVS)
Belgium	1999(CSES,WVS)	Macedonia	1998(WVS) 2001(WVS)
Bosnia	1998(WVS) 2001(WVS) 2004(HES)	Moldova	1996(WVS) 1999(WVS) 2005(DHS) 2006(WVS)
Bulgaria	1995(HES) 1997(WVS) 2001(CSES) 2006(WVS)	Netherlands	1999(WVS)
Cyprus	2006(WVS)	Romania	1996(CSES,WVS) 1997(HES) 2005(WVS)
Czech Rep	1996(CSES)	Slovakia	1998(WVS)
Estonia	1996(WVS) 1999(WVS) 2000(HES)	Slovenia	1996(CSES)
Finland	2003(CSES) 2004(HES) 2005(WVS)	Spain	1995(WVS) 1996(CSES) 2000(CSES,WVS) 2004(CSES) 2007(WVS)
France	1999(WVS) 2002(CSES) 2006(WVS)	Sweden	2005(HES) 2006(WVS)
Germany	1999(WVS) 2004(HES) 2006(WVS)	UK	2004(HES)
Hungary	2002(CSES)	Ukraine	1996(WVS) 1998(CSES) 2006(WVS)
ASIA			
Armenia	1997(WVS) 2000(DHS)	Pakistan	2001(WVS)
Azerbaijan	1995(HES) 1997(WVS) 2006(DHS)	Philippines	1993(DHS) 1998(DHS) 2003(DHS) 2008(DHS)
Bangladesh	1996(WVS) 1997(DHS) 2000(DHS) 2002(WVS) 2004(DHS) 2007(DHS)	Russia	1995(WVS) 1999(CSES) 2000(HES,CSES) 2006(WVS)
Georgia	1996(WVS)	Singapore	2002(WVS)
India	1995(WVS) 2001(WVS) 2006(WVS)	Taiwan	1995(WVS) 1996(CSES) 2004(CSES)
Iran	2007(WVS)	Tajikistan	1996(HES)
Israel	1995(HES) 2005(HES)	Turkey	1993(DHS) 2007(WVS)
Kazakhstan	1995(DHS) 1999(DHS)	Uzbekistan	1996(DHS)
Kyrgyz Rep	1997(DHS) 2003(WVS)	Vietnam	1997(DHS) 2002(DHS) 2005(DHS)
Malaysia	2006(WVS)		
AFRICA			
Algeria	2002(WVS)	Madagascar	2005(AFRO)
Benin	1996(DHS) 2001(DHS) 2005(AFRO) 2006(DHS)	Malawi	2000(DHS) 2003(AFRO) 2004(DHS) 2005(AFRO)
Botswana	2003(AFRO) 2005(AFRO)	Mali	1994(HES) 1995(DHS) 2001(DHS) 2002(AFRO) 2005(AFRO) 2006(DHS)
Burkina Faso	1992(DHS) 1998(DHS,HES) 2003(DHS)	Morocco	2001(WVS) 2007(WVS)
Cameroon	1998(DHS) 2004(DHS)	Mozambique	2002(AFRO) 2005(AFRO)
Central African Rep	1994(DHS)	Namibia	2000(DHS) 2003(AFRO) 2006(AFRO)
Chad	1997(DHS) 2004(DHS)	Niger	1992(DHS) 1998(DHS) 2006(DHS)
Cote d'Ivoire	1998(DHS)	Nigeria	2000(WVS) 2005(AFRO)
DRC	2007(DHS)	Senegal	1992(DHS) 2002(AFRO) 2005(DHS) 2005(AFRO)
Egypt	1995(DHS) 2000(WVS) 2005(DHS) 2008(DHS)	South Africa	1996(WVS) 1998(DHS) 2001(HES,WVS) 2002(AFRO) 2006(AFRO) 2007(WVS)
Ethiopia	2000(DHS) 2005(DHS)	Tanzania	1993(HES)
Gabon	2000(DHS)	Togo	1998(DHS)
Ghana	1993(DHS) 1998(DHS) 2003(DHS) 2008(DHS)	Uganda	1995(DHS) 2005(AFRO)
Guinea	1999(DHS) 2005(DHS)	Zambia	1996(DHS) 2001(DHS) 2003(AFRO) 2005(AFRO) 2007(DHS,WVS)
Kenya	1993(DHS) 1998(DHS) 2003(AFRO,DHS) 2005(AFRO) 2008(DHS)	Zimbabwe	2001(WVS) 2004(AFRO) 2005(AFRO)
AMERICAS			
Bolivia	2002(HES) 2003(DHS)	Mexico	1997(CSES) 2000(WVS,CSES) 2003(CSES)
Brazil	1996(DHS) 1997(WVS) 2002(HES,CSES) 2006(HES,WVS)	Nicaragua	2001(HES)
Canada	1997(HES,CSES) 2000(WVS) 2001(HES) 2006(WVS)	Peru	2000(DHS) 2004(HES,DHS) 2008(WVS)
Colombia	1998(WVS)	United States	1996(CSES) 1997(HES) 2000(WVS) 2004(CSES) 2005(HES) 2006(WVS)
Dominican Rep	1996(WVS)	Uruguay	1996(WVS) 2006(WVS)
Guatemala	1995(DHS) 1998(DHS) 2000(HES) 2005(WVS) 2006(HES)	Venezuela	1996(WVS) 2000(WVS)
Guyana	2005(DHS)		
OCEANIA			
Australia	1995(WVS) 1996(CSES) 2004(CSES) 2005(WVS)	New Zealand	1996(CSES) 1998(WVS) 2002(CSES)