Inequality, Ethnicity and Civil Conflict

John D. Huber
Laura Mayoral

October 2013

Barcelona GSE Working Paper Series

Working Paper nº 744

Inequality, Ethnicity and Civil Conflict

John D. Huber and Laura Mayoral*

October 18, 2013

Abstract

Although economic inequality has long been viewed as a cause of civil conflict, existing research has not found robust empirical support for this relationship. This study explores the connections between inequality and civil conflict by focusing on the mediating role of ethnic identity. Using over 200 individual-level surveys from 89 countries, we provide a new data set with country- and group-level measures of inequality within and across ethnic groups. We then show that consistent with Esteban and Ray's (2011) argument about the need for labor and capital to fight civil wars, at both the country and group level, there is a strong positive association between within-group inequality and civil conflict. We do not, however, find support for previous arguments that inequality across ethnic groups should be associated with the incidence or intensity of civil conflict. By breaking down the measures of inequality into group-level components, the analysis helps explain why it is difficult to identify a relationship between general inequality and conflict. More generally, it highlights the limitations in cross-national research associated with drawing substantive conclusions by relying on measures of overall inequality, like the Gini.

Keywords: Ethnicity, inequality, civil conflict, gini decomposition, within-group inequality, between-group inequality, fractionalization.

JEL: D63, D74, J15, O15

*Huber: Department of Political Science, Columbia University, jdh39@columbia.edu; Mayoral: Institut d'Anàlisi Económica, CSIC and Barcelona GSE; laura.mayoral@iae.csic.es. John Huber is grateful for financial support from the National Science Foundation (SES-0818381). Laura Mayoral gratefully acknowledges financial support from the CICYT project ECO2011-25293, the AXA research fund and Recercaixa. We received helpful comments from Lars-Erik Cederman, Joan Esteban, Debraj Ray and seminar participants at various venues where this paper was presented. We also thank Sabine Flamand and Andrew Gianou for superb research assistance.

1 Introduction

Intra-state civil conflicts have replaced inter-state wars as the nexus for large scale violence in the world. Gleditsch et al. (2002), for example, find that since WWII, there were 22 interstate conflicts with more than 25 battle-related deaths per year, 9 of which have killed at least 1,000 over the entire history of the conflict. Over the same period, there were 240 civil conflicts with more than 25 battle-related deaths per year, and almost half of them have killed more than 1,000 people. Economic inequality has long been posited as a central driver of civil conflict.[1] However, cross-national empirical research has not found robust empirical support for this conjecture (e.g., Lichbach 1989, Fearon and Laitin 2003 and Collier and Hoeffler 2004). Our main purpose is to revisit this relationship by focusing on how group identity and economic inequality interact to precipitate civil conflict.

Most internal conflicts since WWII have been largely ethnic or religious in nature, while outright class struggle seems to be rare (Doyle and Sambanis 2006).[2] If group identity plays a central role in conflict, then the lack of any empirical relationship between standard inequality measures that ignore group membership and conflict should be unsurprising since such measures fail to capture the economic conditions of relevant groups. Instead, the effect of economic inequality on conflict should work through these (ethnic or religious) groups. Large economic differences across groups may lead to grievances that spark civil wars, for instance, and inequality within groups may affect the ability of groups to sustain civil violence. Thus, understanding the empirical relationship between economic inequality and civil conflict requires one to take into account how inequality manifests itself within and across groups.

This study makes three contributions to this end. The first is to provide a new data set measuring both "horizontal inequality" (i.e., economic differences across groups) and "vertical inequality" (i.e., inequality within groups). Existing large-scale cross-national data sets on group-based inequality rely on indirect measures of ethnic affiliation and income status (e.g., spatial measures of ethnic settlement areas and spatial estimates of income).[3] Thus, not only are these data sets based

[1] Influenced by the writings of Karl Marx, Dahrendorf (1959), Gurr (1970, 1980) and Tilly (1978) are some representatives of this literature.

[2] See Montalvo and Reynal-Querol (2005) and Esteban, Mayoral and Ray (2012) for recent evidence on the connection between ethnic structure and conflict.

[3] See Cederman et al. (2011) and Alesina et al. (2013).

on strong assumptions about group identity and income, they cannot measure inequality within groups.[4] In contrast, our data are based on mass surveys in which individuals state their ethnic affiliation and income status. Individual level surveys are likely to provide more reliable measures than those based on the geo-coded approach. In addition, by using individual-level data on group identity and income, we can relate between and within group inequality to overall inequality, and thus examine how different dimensions of inequality affect our ability to make inferences about the relationship between overall inequality and conflict.

Our second contribution is to test theories about the relation between ethnic inequality and conflict. The first theory focuses on the ability of groups to sustain violence. Esteban and Ray (2008 and 2011) argue that within-group inequality should be a key driver of conflict because it facilitates the mobilization of combatants. 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. The second theory focuses on group grievances. Some authors have argued that inequality across groups aggravates group grievances, which in turn may lead to conflict (Stewart 2002, Cederman et al., 2011). As we discuss below, however, this argument is not unambiguous. If one group is particularly poor, for example, it may lack the means to wage violence. And recent empirical research has found that an increase in the income of poorer groups is associated with an intensification of conflict. Using cross-country as well as within-country regressions, we find strong support at the country and group level for the hypothesis that within-group inequality and conflict are positively related. We do not find a significant association between indices of horizontal inequality and conflict, casting doubt on arguments about grievances.

Our third contribution is to shed light on why it should be difficult to find a relationship between measures of overall inequality, such as the Gini coefficient, and conflict. It is well-known that when individuals belong to groups, the Gini coefficient can be decomposed into three terms: between-group inequality, within-group inequality, and a residual, often called overlap, which is negatively related to the economic segregation of groups. Only the coefficient of within-group inequality is significantly associated with conflict, while those of between-group inequality and overlap are not. In addition, although the within-group component is the largest on average, its

[4] An exception is a recent working paper by Kuhn and Weidmann (2013), which we discuss below.

variability is considerably smaller than that of the other two components which causes its correlation with the Gini coefficient to be basically zero. It follows that the "noise" introduced by overlap and the between-group inequality components makes it difficult to find any significant relationship between the Gini coefficient and conflict.

The paper is organized as follows. Section 2 describes the relevant existing theoretical and empirical literature on inequality, group identity and civil conflict. We then describe the survey-based inequality data in section 3, including the methods we employ to address heterogeneity in the types of surveys used in constructing the data set. Section 4 describes the inequality data in the countries included in our empirical analysis. Sections 5 (country-level analysis) and 6 (group-level analysis) provide our main empirical results, and section 7 concludes.

2 Background

2.1 Theory

As noted in the Introduction, most empirical studies of civil conflict do not find a significant relationship between economic inequality and the likelihood of conflict. These papers typically rely on country-aggregate measures of individual (or household) inequality – such as the Gini coefficient – in their empirical analysis. It seems premature, however, to dismiss the possibility that inequality and conflict are related (Cramer 2003, Sambanis 2005, Acemoglu and Robinson 2005). Civil conflicts are often fought between groups defined by non-economic markers. In fact, ethnic or religious conflicts account for 50-75% of civil wars after WWII (Doyle and Sambanis 2006, Fearon and Laitin 2003). It is hardly surprising, then, that measures that fail to capture group aspects of inequality are unrelated to conflict. To the extent that most internal conflicts seem to be fought across ethnic lines, it seems natural to focus on inequality that is related to group identity.

Esteban and Ray, henceforth ER, (2008, 2011) develop a theory arguing that one of the keys to mobilizing combatants is economic heterogeneity within a group. The main idea is highly intuitive: effectiveness in conflict requires various inputs, most notably, financial support and labor (i.e, fighters). Typically, the rich provide funds and the poor provide combat labor. Conflict, therefore, has at least two opportunity costs: the cost of contributing resources and the cost of contributing one's labor to fight. Economic inequality within a group simultaneously decreases both opportunity costs: when the poor within a group are particularly poor, they will require a relatively small compensation for fighting, and when the rich within a group are particularly rich the opportunity cost of resources to fund fighters will be relatively low. Thus, groups with high income inequality should have the greatest propensity to engage in civil conflict. ER do not model group decisions to enter conflict, but rather assume that society is in a state of (greater or lesser) turmoil, with intra-group inequality influencing whether conflict can be sustained.[5]

The potent nature of within-group inequality as a driver of conflict can account not only for conflict intensity but also for the salience of ethnicity (versus class) in conflict. In a model of coalition formation, ER (2008) show that in the absence of bias favoring either type of conflict, ethnicity will be more salient than class. This is because a class division creates groups with strong economic homogeneity. Thus, while the poor may have the incentives to start a revolution, conflict might be extremely difficult for the poor to sustain because of the high cost of resources. But even if the poor are able to overcome these constraints, class conflict may not start. When the rich foresee a class alliance that can threaten their status, they can propose an ethnic alliance (to avoid the class one) that will be accepted by the poor ethnic majority, planting the seeds of ethnic conflict.

The theoretical connection between horizontal inequality and conflict is more ambiguous. On the one hand, if the winning group can expropriate the rival's resources, the larger the income gap between the groups, the greater the potential prize, and hence the greater the incentive for conflict by the poorer group (Acemoglu and Robinson 2005, Wintrobe 1995, Stewart 2002, Cramer 2003). Additionally, theories of "relative deprivation" suggest that if inequality coincides with identity cleavages, it can enhance group grievances and facilitate solutions to the collective action problem associated with waging civil conflict (Stewart 2000, 2002). On the other hand, especially poor groups might find it particularly difficult to wage conflict, and an increase in the income of a poorer group might enhance the group's capacity to fund militants. Thus, the closing of the income gap between groups – rather than its widening – should be associated with higher levels of inter-group conflict. There is empirical evidence supporting this possibility. Morelli and Rohner (2013), for example, find in cross-national analysis that when oil is discovered in the territory of a

[5]It has also been argued that heterogeneity in incomes might create resentment among the poor and reduce group cohesiveness (Sambanis and Milanovic, 2011). ER (2008) argue that this effect is dwarfed by the within-group specialization that such heterogeneity provides. The direction of the relation between within-group inequality and conflict is ultimately an empirical question.

poor group, the probability of civil war increases substantially. And Mitra and Ray (2013) present evidence from the Muslim-Hindu conflict in India (where Muslims are poorer on average), showing that an increase in Muslim well-being generates a significant increase in future religious conflict, whereas an increase in Hindu well-being has a negative or no effect on conflict. Finally, at least since Tilly (1978), scholars argue that grievance factors such as inequality are, for the most part, omnipresent in societies, depriving the variable of explanatory value. According to this approach, the critical factors that foster civil unrest are those that facilitate the mobilization of activists.

2.2 Existing empirical evidence

Testing the relation between ethnic inequality and conflict has been traditionally hampered by the difficulty of obtaining data on within group inequality for a large number of countries. Thus empirical research on this topic is limited. Ostby et al. (2009) have found a positive and significant relation between *within-region* inequalities and conflict onset using data from the Demographic and Health surveys for a sample of 22 Sub-Saharan African countries. Developed in parallel to our paper, Kuhn and Weidmann (KW, 2013) introduce a new global data set on inequality within-group using nightlight emissions and find that higher income heterogeneity at the group level is positively associated with the likelihood of conflict *onset*. Our contribution differs from theirs in several respects. First, in addition to group-level evidence, we also provide country-level regressions that help to clarify why the connection between overall inequality and conflict has been so difficult to establish. Second, although KH claim to be testing the ER (2011) theory, their main dependent variable is conflict onset. As mentioned before, ER explicitly state that they do not model the decision of groups to enter into conflict since it can ignite for a wide variety of reasons; instead, their theory describes why the income-heterogeneity of groups should affect the ability to sustain conflict. Thus, we use measures of conflict incidence/intensity as a more appropriate way of conducting the test. Finally, as we describe in more detail below, the use of nightlight emissions to measure within-group inequality is plagued by serious problems, making it difficult to draw inferences from their results.

With respect to horizontal inequality, Stewart (2002) use case studies to document a positive connection between horizontal inequality and conflict. Ostby et al. (2009) use surveys from

Africa on regional inequality, as noted above, and find that regional inequalities do matter for civil conflict. And in the only large-scale cross-national analysis, Cederman et al. (2011) find that both relatively rich and relatively poor ethnic groups are more likely to be involved in civil wars than groups whose wealth lies closer to the national average.

Some illustrations. Focusing on the connection between within-group inequality and conflict, ER (2011) provide case studies from Africa, Asia and Europe to illustrate the causal mechanisms in their theory. The case of the Rwandan genocide is also suggestive. In the spring of 1994, the Hutu majority carried out a massacre against the Tutsi minority where 500,000 to 800,000 Tutsi and moderate Hutus that opposed the killing campaign were assassinated. In the years immediately prior to the genocide, Rwanda suffered a severe economic crisis motivated by draughts, the collapse of coffee prices, and a civil war. Verwimp (2005) documents an increase in within-group inequality among the Hutu population prior to the genocide: on the one hand, a sizeable number of households that used to be middle-sized farmers lost their land and became wage workers in agriculture or low skilled jobs. On the other, rich farmers with access to off-farm labor were able to keep and expand their land. This new configuration encouraged the Northern Hutu elites to use their power to instigate violence. Backed by the Hutu government, these elites used the radio (particularly RTLM) and other media to begin a propaganda campaign aimed at fomenting hatred of the Tutsis by Hutus (Yanagizawa-Drott, 2012). The campaign had a disproportionate effect on the behavior of the unemployed and on delinquent gang thugs in the militia throughout the country (Melvern 2000), individuals who had the most to gain from engaging in conflict (and the least to lose from not doing so). Importantly, the campaign made it clear that individuals who engaged in the ethnic-cleansing campaign would have access to the property of the murdered Tutsi (Verwimp, 2005). Thus, the rich elites "bought" the services of the recently empoverished population by paying them with the spoils of victory, something that was more difficult to undertake prior to the economic crisis.

Fearon and Laitin (2000) also provide several examples where the elites promote ethnic conflict and combatants are recruited from the lower class to carry out the killings. In the words of Fearon and Laitin (2000),

> [O]ne might conjecture that a necessary condition for sustained *ethnic violence* is the

> availability of thugs (in most cases young men who are ill-educated, unemployed or underemployed, and from small towns) who can be mobilized by nationalist ideologues, who themselves, university educated, would shy away from killing their neighbors with machetes. (p. 869).

Examples of this behavior were found in Bosnia (the "weekend warriors," a lost generation who sustained the violence by fighting during the weekends and going back to their poor-paid jobs in Serbia during the week), in Sri Lanka (where the ethnic war on the ground was fought on the Sinhalese side by gang members), and in Burundi.[6]

3 A new data set on ethnic inequality

Empirical research on the impact of ethnic inequality on several socio-economic outcomes (conflict, redistribution, growth, etc.) has traditionally been hindered by the lack of the appropriate data. Nevertheless, in the last couple of years new data sets with ample country coverage have started to appear. To our knowledge there exist two data sets with near global coverage that measure income differences *across* groups (Cederman et al. 2011 and Alesina et al. 2013) and one that focuses on inequality *within* groups (Kuhn and Weidmann 2013, KW henceforth).

These data sets are similar in construction. They combine geo-referenced data on the geographic location of ethnic groups with geo-referenced estimates of income, which can be measured using the Nordhaus (2006) G-Econ data set (the approach taken by Cederman et al. 2011) or satellite images of light density at night (the approach taken by Alesina et al. 2013 and KH).[7] A significant advantage of these data sets is that they cover the vast majority of countries in the world.

Although these data sets represent a clear improvement over what previously existed, they also have important limitations. One problem is that fine-grained information about where group members live typically do not exist. The geo-referencing methodology therefore relies on expert estimates of the spatial location of groups, an approach that suffers measurement error since the experts themselves often do not have data on which to base their estimates of group locations. A second problem is that the methodology requires one to assume either that particular geo-coded areas are occupied by only one group, or that individuals from different groups in the same geo-coded

[6]See Fearon and Laitin (2000) and the references therein for more complete accounts.

[7]For limitations using satellite images to measure economic development, see Letu et al. (2012).

area have the same income. Neither assumption is attractive. We know that there is substantial variation in the regional segregation of groups, and Morelli and Rohner (2013) link this segregation itself to civil conflict. And if one assumes that individuals from different groups occupy the same geo-coded area, one also has to assume that individuals from these different groups all have the same income – that is, to assume what one is trying to measure. Thus, the geo-coded approach to measuring group-based inequality suffers from considerable measurement error. Finally, these datasets only focus on one dimension of ethnic inequality (either between- or within-group inequality) and, therefore, the magnitudes of these measures are not directly comparable across datasets.

The problems are particularly severe when one uses geo-coded data to measure within-group inequality. KW use data on ethnic settlement regions (GeoEPR) that is divided up into cells of equal size (about 10 km), discarding cells from urban areas (where the rich in particular groups might be especially likely to live). For each cell, KW compute nightlight emissions per capita. Then all cells occupied by a group are used as inputs to calculate the group's Gini coefficient. In addition to the problems described in the previous paragraph, the resulting measure will be very sensitive to cell size, since the larger the size of the cell, the smaller the resulting within-group measure – in the limit, if the whole territory is assigned to one cell, within-group inequality would be zero. But the choice of cell size is arbitrary. In addition, it is difficult to deal with work versus residential areas in the same territory. The former will have high light emissions per capita (since population will be low), which will artificially contribute to the size of within-group inequality measures. Finally, with over half the world's population living in urban areas (Angel 2012), the fact that urban cells are discarded is likely to have a large impact on the estimates, since a huge source of within-group inequality (rural-urban inequality) is dismissed. This seems particularly problematic in countries with high urbanization rates. KW matched their within-group inequality data to similar measures computed using Demographic and Health surveys for 17 sub-Saharan countries finding a correlation of 0.42. This figure is worrisomely low: sub-Saharan Africa has very low urbanization rates and so the correlation is likely much lower in other parts of the world where urbanization is higher.

Our data set addresses the main problems outlined above by using mass surveys where individuals state their ethnic affiliation and their income status. Although the survey-based approach is far from perfect, the type of information it provides is potentially much more reliable than that obtained from the geo-referencing approach. We do not need to rely on ethnic settlement areas, which can be an important source of noise, especially in countries where ethnic segregation is low. And the information about income status of individuals is likely much more accurate than night-light emissions or spatial wealth measures. Finally, we are able to compute (comparable) measures of across and within group inequality which allow us to determine the relative size of the between and within group inequality components in overall inequality measures and, thus, examine whether our ability to establish an association between general inequality and conflict improves when ethnic identity is considered. As a drawback, our data set includes 89 countries, a smaller number than one obtains from the geo-coded approach. Below, however, we show that our sample of countries appears to be representative in key respects.

3.1 The surveys used to create the data set

To accurately measure the components of the Gini decomposition one needs 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 adjust the resulting inequality measures to account for survey heterogeneity.

We have three categories 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 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 households possess. Such surveys are typically used in countries where there are many very poor individuals, and thus where individual income cannot be used to meaningfully distinguish the economic well-being of most individuals from each other. In such cases, social scientist 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 households. The surveys of this type include the Demographic Health Surveys (DHS) and the Afrobarometer Surveys.[8]

The "income" data from these surveys, along with the group identity variables, make it possible to measure inequality within and across groups. We rely on the list of groups from Fearon (2003) to identify the relevant groups in each country. Fearon provides a set of clear and reasonable criteria for identifying the relevant ethnic, religious, racial and/or linguistic groups across a wide range of countries, and his list is widely used in the literature.[9] We match the groups in the surveys to the groups in Fearon (2003) and omit surveys that do not adequately represent the groups in the Fearon list.[10]

In total, we have 233 surveys from 89 countries depicted in the map in Figure 1.[11] The surveys were conducted from 1992 to 2008, with 223 of the 233 surveys in the period 1995-2008 (and 10 surveys in the period 1992-94).[12] Most of the survey data comes from the WVS (79 surveys) and the DHS (71). The number of surveys in the remaining categories are 30, 29 and 24 for HES, CSES and Afrobarometer, respectively. For 29 countries, we have only one survey, whereas in others we have multiple surveys, at most 7.

[8]For the DHS surveys, which contain a large number of asset indicators (typically around 13), we follow Filmer and Pritchett (2001) and McKenzie (2005) and run a factor analysis on the 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 Afrobarometer surveys have a much smaller number of asset questions, typically 5 or less, and so we simply sum the assets.

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

[10]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.

[11]32 out of these 233 surveys correspond to observations for the same country/year coming from different surveys. Thus we have 217 surveys corresponding to different country/year observations

[12]A list of these surveys is provided in Table A.6 in the appendix.

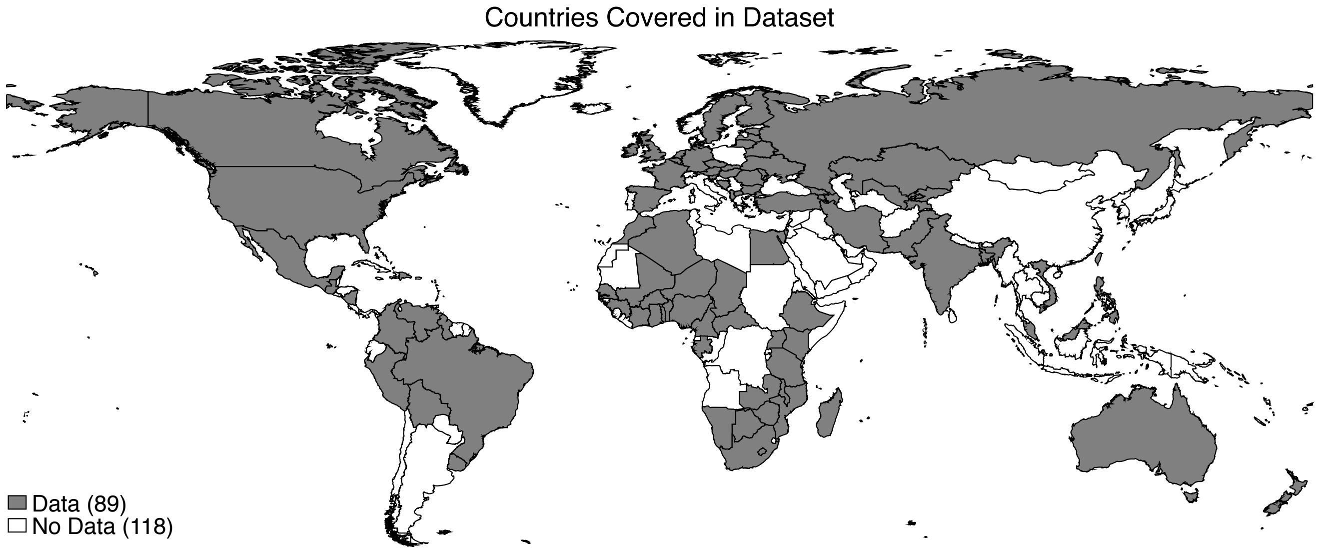


Figure 1: Countries included in data set

3.2 Measuring ethnic inequality

The Gini coefficient is often employed to study the relation between overall inequality and conflict. However, if individuals can be assigned to groups, then at least since Pyatt (1976), scholars have observed that this coefficient is the sum of three components: BGI (between-group inequality), WGI (within-group inequality) and OV (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

$$BGI = \frac{1}{2\bar{y}}(\sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j \mid \bar{y}_i - \bar{y}_j \mid), \tag{1}$$

where m denotes the number of groups, n_i is the size of group i, $\bar{y}$ is the mean income in society, and $\bar{y}_i$ is the average income of group i.

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

$$WGI = \sum_{i=1}^{m} G_i n_i \pi_i, \tag{2}$$

where G_i and π_i are the Gini coefficient and the proportion of total income going to group i, respectively.

In contrast to other inequality indices, the Gini coefficient does not decompose neatly into the BGI and WGI components.[13] Overlap is the residual that remains when BGI and WGI are subtracted from the Gini (G)

$$OV = G - WGI - BGI.$$

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 OV, 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.

For the country-level analysis, we compute the inequality measures by using the surveys to obtain the income and Fearon (2003) to measure the size of groups. For the group-level analysis, we use the surveys to compute the Gini coefficient for each group, as well as a measure of how a group's income differs from the average income in society. We define these variables precisely below, but before doing so, it is important to address the issue of heterogeneity that exists in the measure of "income" across surveys.

[13]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 units is constant (such as when comparing inequality between urban and rural areas across states).

3.2.1 Addressing survey heterogeneity

Since the surveys vary in their measures of "income," we face a trade-off. If we keep only those surveys that contain comparable "income" measures, we are left with a very small sample size. But if we use all of the surveys, we must compare surveys that have different types of income measures. Our approach to this problem is to include as many as countries as possible and to use information from within and outside the surveys to adjust the measures to achieve comparability across survey types.

The tradeoff between obtaining a sample with broad coverage and having inequality measures that are comparable across countries is not new. For instance, the observations in Deininger and Squire's (1996) data set 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. Its successor, the World Income Inequality Database (WIID), which provides the most comprehensive data set of income inequality, presents identical shortcomings. 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. The second is to calculate the average difference in inequality between observations that vary in their income definition and then adjust observations by this difference. This leads to applying a constant adjustment across all countries and years that share the same income definition, which is clearly problematic because it fails to capture the variation across countries (and over time).

We follow a different approach to address this issue. Heterogeneous income measures create two types of problems for constructing measures of the Gini decomposition. One is bias in estimated *levels* of the overall Gini coefficient. The other is bias in estimated *proportions* of the Gini assigned to each of its three components. We consider the two in turn.

Adjusting the *level* of the Gini coefficient. Since measures of the Gini coefficient are available from external sources, we can examine how the Ginis obtained from the surveys relate to existing data sets. We feel that the best data set for comparison purposes comes from Solt (2009). Solt proposes a methodology that makes it possible to maximize the comparability of various income inequality data while maintaining the widest possible coverage across countries and over time. He uses a wide variety of inequality data sources, the most relevant ones being the WIID and the Luxembourg Income Studies (LIS).[14] Since the LIS data are regarded as the best available for making cross-national comparisons of income inequality, Solt's goal is to make non-LIS measures "comparable" to LIS data.

To this end, Solt exploits information from "duplicates" – that is, from 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. These ratios can only be computed for those pairs of categories in those countries and years for which they are least useful because data from the two categories already exist. However, they provide valuable information about what the missing ratios are likely to be because within a given country the factors that affect these ratios tend to change slowly over time. Thus, the best prediction for a missing ratio will be based on available data on the same ratio in the same country in proximate years. Solt's algorithm treats the unknown ratios as missing data, which are then predicted using regression models (see Solt 2009 for details). The output of this procedure is the Standardized World Income Inequality Data (SWIID), which provides comparable Gini indices of gross and net income inequality for more than 4500 observations, corresponding to 173 countries from 1960 to the present.

We compare the Ginis derived from our surveys to Ginis from SWIID (using SWIID's net income variable). The survey Ginis are on average smaller than SWIID Ginis (averages equal 0.30 and 0.38, respectively) and the correlation coefficient is around 40% However, the correlations vary considerably across surveys. With the exception of the HES observations, all other surveys underestimate inequality on average. The largest underestimates are found using 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). However, the mean and the standard deviation corresponding to HES – the best surveys we are able to find – and SWIID observations are basically identical, with a correlation coefficient of 0.9. Additionally, we have regressed the survey

[14] 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 ILO's Household Income and Expenditure Statistics, as well as data from other national statistical offices; see Solt (2009) for details.

Ginis on survey dummies (excluding HES), along with time and regional dummies. In all cases, the survey dummies are negative and highly significant, indicating that observations coming from surveys other than HES are on average smaller.

To correct for these biases, we follow an approach similar to that used by Solt (2009). The goal is to make the survey Ginis comparable to the net income Gini from SWIID (which is constructed to achieve comparability with the LIS Gini). Whenever the SWIID Gini and a Gini from our data are available for the same country and year, we compute their ratio. This allows us to compute 211 ratios. Since the total number of surveys is 233, 22 ratios are missing. Next, missing ratios are predicted by regressing the available ratios on country, time and regional dummies. Finally, the "adjusted" survey Ginis are obtained by taking the product of the original survey Ginis and the predicted ratios. The resulting Gini measures are very similar to the original ones from SWIID, with a correlation coefficient between the two data sets equal to 0.97.

Adjusting the *proportions* of the Gini's components. Next we explore whether the shares of the Gini assigned to each of its components are comparable across surveys. Unlike for the Ginis, there are no appropriate external data for establishing a comparison, but there are observations for the same country and year (and/or neighboring years) coming from different surveys.[15] Since proportions of the Gini that are attributable to each of its three components are likely to move slowly over time, it is possible to exploit this overlap to check whether the proportions differ systematically across survey types.

In fact, the correlations of the proportions of the Gini components using different surveys for the same country are very high. For instance, the correlation between the proportion of the Gini due to WGI and the proportion due to WGI in the HES and WVS surveys is .94. For BGI, this correlation is 0.93. Correlations are also high between WVS and the surveys based on asset data (the Afrobarometer and DHS), with correlations above .80 for all comparisons.[16]

The average values of the shares do 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 the fact that particular surveys are correlated with

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

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

regions, rather than to survey effects. To see this, we regressed WGI (BGI) shares on survey and time dummies; survey, time and regional dummies; and survey, time and country dummies. In the first set of regressions (with no regional or country dummies) the survey dummies are significant, but 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 that are systematically associated with survey type.

As a final check on the reliability of the proportions data for BGI, WGI and OV, we rely on the fact that some of our surveys provide relatively high quality income data (HES and, to a lesser extent, WVS) while others are potentially biased by the fact that they use assets rather than income (Afrobarometer, DHS), or because they report income in quintiles (CSES). We therefore have first excluded from the sample the HES and WVS surveys and have predicted them by regressing the remaining proportions of the Gini 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 DHS 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 DHS and the Afrobarometer are also very similar to the ones obtained from the surveys, with correlation coefficients larger than 0.9.

The analysis thus shows that while a number of the surveys underestimate the true Gini, they produce reasonable estimates of the proportion of inequality that can be attributed to each component. We can therefore use the "adjusted" Gini coefficients and the unadjusted proportions to calculate the Gini decomposition. At the country-level, we compute the "adjusted" WGI, BGI and OV by multiplying their original values from the surveys by the same adjustment factor used to adjust the Gini coefficient. At the group level, we multiply the above-described adjustment factors by the Gini coefficient of each of the groups. Thus, by aggregating the within group level data one obtains the country-level one. More details about the group-level adjustment are provided in the next section.

3.2.2 Country- and group-level measures of ethnic inequality.

In the country-level empirical analysis we use both time-invariant as well as time-varying inequality measures. To compute time-invariant measures, for each country we take the average from all the available years for which surveys are available and assign these average values to all years, beginning with the first year for which a survey exists in the country. Thus, data are missing in years prior to the first available survey year. The variables produced using this procedure are:

- $\overline{\text{GINI}}$: Country's average of the "adjusted" Gini based on income survey data and Fearon (2003) group shares, beginning in first year for which a survey exists.
- $\overline{\text{WGI}}$: Country's average "adjusted" within-group inequality based on income survey data and Fearon (2003) group shares, beginning in first year for which a survey exists.
- $\overline{\text{BGI}}$: Country's average "adjusted" between-group inequality based on income survey data and Fearon (2003) group shares, beginning in first year for which a survey exists.
- $\overline{\text{OV}}$: Country average "adjusted" Overlap based on income survey data and Fearon (2003) group shares, beginning in first year for which a survey exists.

In addition, we consider a time-varying construction of the adjusted inequality measures. WGI$_t$, for example, takes the value of WGI for each country/year for which a survey is available and is missing for those country/year observations for which there is no survey data. BGI$_t$ and OV$_t$ are constructed in the same way. For robustness, we also examine other ways of adjusting the data (using an approach similar to that suggested by Deininger and Squire; see Section 5.3 and the Appendix for details and results).

The group level data are constructed in a similar fashion. GROUPGINI is the adjusted Gini coefficient for the group. The Gini for each group is first calculated from the surveys and then adjusted using the same ratios that are used to correct the country Ginis. As in the country-level analysis, we average the available observations for a given group and project them forward, starting from the first non-missing year. To measure the relative economic well-being of a group, we use the surveys to estimate HORINEQ, a measure of horizontal inequality employed in Cederman et al. (2011) and Kuhn and Weidmann (2013). It is defined as HORINEQ= $log(I_i/\bar{I})^2$, where I_i is income per capita of group i and $\bar{I}$ is the average income per capita of all groups in the country. This definition captures deviations from the country average and is zero for groups at the country average. We average the available observations for a given group and project them forward, starting from the first non-missing year. Since HORINEQ is unrelated to the Gini decomposition, we cannot adjust it in the same way as GROUPGINI to control for survey heterogeneity. We leave this variable unadjusted in the models presented in the main text, but in Section 5.3, we show that the results are robust when we adjust the measure for survey heterogeneity using an approach similar to that of Deininger and Squire.

4 The Gini decomposition in 88 countries

Before describing the inequality data used in our analysis, it is important to recognize that one limitation of the survey data approach to measuring inequality is that it is available in fewer countries than one obtains when using geo-coded data. This would be of particular concern if there were systematic biases in the types of countries for which the survey data is available. Table 1 examines this issue empirically. In this analysis, we focus on 88 countries since data on some key controls are missing for one of the countries in our dataset (Bosnia) and, therefore, it never enters our regressions.

The top half of Table 1 describes the distribution of countries around the world using the SWIID and our survey data, focusing on the post-1994 time period for which most of our survey data exists. There are 137 countries available in SWIID (taking into account that there are some countries in this data set for which conflict or other control variables do not exist) and 88 countries – or 64 percent of the SWIID – for which we have useful surveys. The table shows a slightly higher proportion of the countries in the survey data are from Central Europe, and a slightly higher proportion of the Solt countries are from Latin America, but the distributions of countries across the regions are quite similar. Thus, there is little in the way of regional bias in the survey data.

The bottom half of the table provides descriptive data on key variables in the two data sets. The SWIID countries and the survey countries are quite similar with respect to GDP/capita, ethnic diversity (fractionalization, F, and polarization, P), level of democracy (xPolity), level of inequality,

Table 1: Sample representativeness

	SWIID	Survey data
Number of countries	137	88
Percentage of countries in:		
Central Europe	19.0	25.0
Latin America	16.1	12.5
Middle East	7.3	5.7
Africa	27.7	29.6
Neo-Europe	16.1	18.2
East Asia	8.8	5.7
South Asia	4.4	3.4
Pacific Islands	0.7	0
Average Real GDP/capita	\$9,644	\$10,166
Average F	.46	.50
Average P	.55	.58
Average xPolity	3.4	3.6
Average Gini (SWIID)	.38	.38
Percent of years with Prio25 civil conflict	.15	.17

and the incidence of civil conflict.[17] Thus, although there are limits on the number of countries we can analyze using surveys, these limits do not seem to create biases with respect to the variables of substantive interest in our analysis.

Table 2 presents basic descriptive statistics for Gini and its three components defined in Section 3.2.2. The measures in the table are at the country level, and are based on the average values of all surveys in a country in cases where we have more than one survey. Within-group economic differences ($\overline{\text{WGI}}$) are, on average, the largest component of the Gini (with an average $\overline{\text{WGI}}$=.18). The smallest component is $\overline{\text{BGI}}$, with an average one-third that of $\overline{\text{WGI}}$. Overlap is the second largest component of the Gini, and is only slightly smaller on average than WGI.

Although WGI is the largest component of the Gini, the survey-based data allow us to see that the variability of $\overline{\text{WGI}}$ is considerably smaller than that of $\overline{\text{BGI}}$ and $\overline{\text{OV}}$: the coefficients of variation are 0.91, 0.44 and 0.60 for $\overline{\text{BGI}}$, $\overline{\text{WGI}}$ and $\overline{\text{OV}}$, respectively). And variation in the Gini coefficient is strongly related to variation in BGI and Overlap, but not to variation in WGI. The cor-

[17] Precise variable definitions and sources are provided below.

Table 2: Descriptive statistics for components of the Gini decomposition in 88 countries

	Mean	Std. Dev.	Min.	Max.	Coeff. of Variation
$\overline{\text{BGI}}$	.06	.05	.001	.31	0.91
$\overline{\text{WGI}}$	.18	.08	.01	.31	0.44
$\overline{\text{OV}}$	.15	.09	.01	.38	0.60
$\overline{\text{GINI}}$	.39	.10	.23	.65	0.25

relation between Gini and WGI is non-existent (r=-0.02), but there is a strong correlation between Gini and the other two components (0.71 with BGI and 0.67 with Overlap). This finding has important implications for how social scientists interpret results for Gini coefficients from cross-national regressions because it suggests that variability in this core variable is largely capturing variability to the way that economic well-being is distributed across groups rather than how economic well-being is distributed within groups.

Figure A.2 in the Appendix depicts the distribution of the country-level inequality measures for those countries for which we have data. Latin America, Africa and South Asia are the most unequal regions (with average Gini coefficients of 0.49, 0.47 and 0.40, respectively), while East Asia and the Western countries are the least so (with average Ginis of 0.29 and 0.31, respectively). Focusing on the subcomponents, South Asia and Latin America have the largest values of WGI (0.26 and 0.24) while South-Saharan Africa presents the lowest one (0.11) due to the huge ethnic diversity of this region. Finally, Latin America and Africa are the regions with highest BGI (0.11 in both cases), while this component is lowest in the Western countries (0.02).

5 Country-level analysis of civil conflict

5.1 Country-level variables and definitions

This section provides definitions of the dependent variables and the controls employed in the country-level empirical analysis (see Section 3.2.2 for a description of the country-level inequality variables). A table containing summary statistics of all the variables is presented in the Appendix.

Conflict. Empirical research on civil war distinguishes between conflict onset (the year a civil conflict begins) and conflict incidence (the presence and intensity of conflict in a given year). As Esteban and Ray argue, a civil conflict can break out for a wide variety of reasons, but whether the conflict can be sustained depends on a group's access to both labor and capital. Thus, their theory gives no clear rationale for expecting within-group inequality to be associated with the initiation of conflict, but it should be associated with the ability of groups to sustain violence. We therefore focus on conflict incidence.

Our data is taken from the UCDP/PRIO data set.[18] 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.[19] 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. To capture conflict intensity, we construct a new variable, PRIOINT, that takes the value 0 if there is peace in a given year, the value 1 if there are events satisfying PRIO25 but the total number of battle related deaths that year is smaller than 1,000, and the value 2 if the number of battle-related deaths exceeds 1,000. For robustness, we also consider PRIOCW, a measure of intermediate conflict that takes the value 1 in a country-year if there are at least 25 deaths and if the aggregate level of deaths from the conflict exceeds 1,000.

Controls. We use standard controls, similar to those in Esteban et al. (2012).

- F is the standard measure of ethno-linguistic fractionalization, as measured by Fearon (2003). It is defined as F$= \sum_{n=i}^{m} n_i(1 - n_i)$, where m is the total number of ethnic groups and n_i is the relative size of group i.

- P is the Esteban and Ray (1994) polarization index with binary distances (Reynal-Querol, 2002). It is defined as P$= 4\sum_{n=i}^{m} ni^2(1 - ni)$. Data on n_i comes from Fearon (2003).

- NCONT 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

[18]This is a joint data set 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 description of the data set.

[19]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 types 3 and 4 (internal armed conflict).

by 100 kilometers of water, as measured in Fearon and Laitin (2003).

- MOUNT is the percent of the country that is mountainous terrain, as measured by Fearon and Laitin (2003), who use the codings of geographer A. J. Gerard.

- GDP is the log of real GDP per capita, lagged one year. The source is the Penn World Tables (2011).

- POP is the log of the population in millions, lagged one year, as reported by the Penn World Tables (2011).

- XPOL is a democracy score based on Polity IV, lagged one year. It combines 3 out of the 5 components of Polity IV (XCONST, XRCOMP, XROPEN) and leaves out the two components (PARCOMP and PARREG) that are related to political violence, and hence are likely to be endogeneous. It ranges from -6 (maximum level of autocracy) to 7 (maximum level of democracy). See Vreeland (2008) for details.

- ANOC is a dummy for anocracies that takes a value equal to 1 if the Polity IV index is between -5 and 5 (see Fearon and Laitin 2003 for details).

- DEMOC is a dummy for democracies that takes a value equal to 1 if the Polity IV index is between 5 and 10.

- OIL/DIAM 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 per person in 2000 constant dollars. The source is Ross (2011).

5.2 Country-level results

We now turn to the main purpose of our paper. We begin by examining the relationship between ethnic inequality and conflict at the country level. As noted above, previous research has failed to find a robust connection between general inequality (typically measured by the Gini coefficient) and civil conflict. One reason for this lack of a relationship might be that the quality of the inequality data typically employed in those analyses is low, with observations that cannot be easily compared across countries.[20] To investigate this possibility, we first explore the connection between conflict and overall indices of inequality using improved inequality data. It might also be the case, however, that the relationship between inequality and conflict works through specific subcomponents of the Gini, such as BGI or WGI. If this is true, then the nature of the relationship between the subcomponents and conflict will determine whether we should expect *any* general effect of Gini on conflict. We address this possibility using the survey-based measures of the subcomponents.

To explore the connection between overall indices of inequality and conflict, we employ some the best available estimates of Gini: the Gini based on net and gross income from SWIID and the Gini from Povcalnet (World Bank).[21] Cross-country as well as within-country variation is used to identify the parameters. The time period considered is similar to that in our main analysis below but data are available for a wider range of countries (up to 138).[22] The regression results for Gini, which we will not discuss in detail, are reported in Table A.1 in the Appendix. Consistent with previous research, we find no statistically significant relationship between general inequality and civil conflict. The table also presents the results where the Gini is measured using our surveys ($\overline{\text{GINI}}$ as defined in Section 3.2.2). It is useful to note that the coefficient for our survey-based (average) Gini in column 7 is very similar to that obtained in the first column of Table A.1, where the time-varying net Gini from SWIID is used for a larger number of countries.

Next consider the empirical relationship between inequality and conflict using the subcomponents of the Gini. Table 3 presents the results using PRIO25 as the dependent variable. Models containing regional dummies or country fixed effects are considered (columns 1-6 and 7-8, respectively). The models contain the control variables discussed in the previous section, including the year indicator variables and the regional effects (or country fixed effects), but our discussion will focus on the inequality variables. For convenience, column 1 reproduces column 7 in Table A.1 where $\overline{\text{GINI}}$ is the only inequality variable on the right-hand side. Its coefficient is positive but not significant. Column 2 introduces $\overline{\text{BGI}}$. Since $\overline{\text{BGI}}$ is also a component of the Gini coefficient, we compute a new variable by subtracting $\overline{\text{BGI}}$ from $\overline{\text{GINI}}$ and including this new variable, $\overline{\text{GINI-BGI}}$ on the right-hand side (instead of $\overline{\text{GINI}}$, itself). The coefficient on $\overline{\text{GINI-BGI}}$ therefore estimates

[20] For instance, Collier and Hoffler (2004) and Fearon and Laitin (2003) use Deininger and Squire (1996, 1998) data in their analyses. But, as discussed above, the observations in this dataset are rarely comparable across countries.

[21] See `http://iresearch.worldbank.org/PovcalNet/index.htm?0,2` for methodological details on this dataset.

[22] To make these regressions comparable to those in our main analysis below, we consider data from 1994 onwards, since the bulk of our inequality data –223 out of the 233 surveys– have been gathered after 1994.

Table 3: Country-level results. Baseline specification with PRIO25

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{\text{GINI}}$	3.234							
	(2.951)							
$\overline{\text{GINI-BGI}}$		1.597						
		(3.986)						
$\overline{\text{GINI-WGI}}$			-3.032					
			(4.760)					
$\overline{\text{BGI}}$		5.294		-0.301	0.894	-0.0144		
		(4.329)		(5.119)	(5.131)	(5.309)		
BGI_t							0.060	0.203
							(0.433)	(0.285)
$\overline{\text{WGI}}$			13.761**	13.752**	11.982**	13.771**		
			(6.256)	(6.422)	(6.074)	(6.410)		
WGI_t							0.822**	0.559*
							(0.397)	(0.303)
$\overline{\text{OV}}$				-8.011	-8.069	-8.032		
				(7.221)	(5.668)	(7.214)		
OV_t							0.468	-0.022
							(0.446)	(0.444)
GDP	-0.281	-0.350	-0.219	-0.339	-0.382	-0.339	-0.363	0.033
	(0.254)	(0.270)	(0.261)	(0.274)	(0.283)	(0.272)	(0.229)	(0.025)
POP	0.400***	0.390***	0.361***	0.319**	0.394**	0.382**	-0.541	0.034**
	(0.132)	(0.135)	(0.136)	(0.142)	(0.163)	(0.155)	(0.451)	(0.017)
F	2.676**	2.134*	9.067***	9.932***	8.504**	10.077***		
	(1.219)	(1.219)	(3.493)	(3.789)	(3.471)	(3.759)		
P	1.517	1.802*	1.560	2.091**	2.144**	2.089**		
	(1.002)	(0.985)	(1.007)	(0.992)	(0.946)	(0.987)		
NCONT	1.098	1.370*	1.107	1.705**	1.755**	1.697**		
	(0.671)	(0.739)	(0.673)	(0.758)	(0.719)	(0.765)		
MOUNT	0.011	0.011	0.012	0.011	0.008	0.010		
	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)		
XPOL	0.031	0.037	0.022	0.030		0.038	-0.009	0.006
	(0.041)	(0.043)	(0.042)	(0.044)		(0.055)	(0.019)	(0.007)
ANOC					1.366***			
					(0.501)			
DEMOC					1.129**			
					(0.508)			
(XPOL^2)						-0.005		
						(0.016)		
OIL/DIAM	-0.294	-0.264	-0.272	-0.224	-0.390	-0.338		
	(0.337)	(0.337)	(0.360)	(0.374)	(0.382)	(0.381)		
LAG	4.655***	4.617***	4.565***	4.465***	4.501***	4.494***	0.334**	0.682***
	(0.624)	(0.622)	(0.611)	(0.601)	(0.590)	(0.607)	(0.143)	(0.085)
CONST	-9.334***	-8.142***	-13.200***	-11.936***	-12.281***	-12.403***	8.574	-0.301
	(2.440)	(2.786)	(3.289)	(3.402)	(3.541)	(3.667)	(5.333)	(0.419)
(Pseudo) R^2	0.626	0.627	0.629	0.631	0.635	0.630	0.435	–
Obs	1044	1044	1044	1044	1054	1044	210	214
Reg. E/FE	Reg. E	Reg. E	Reg. E	Reg. E	Reg. E	Reg. E	FE	FE

Notes: Dependent variable is PRIO25. There are 88 countries in columns (1)–(6) and 53 in columns (7)–(8). Robust standard errors (clustered by country) are in parentheses. All columns contain year indicator variables. Reg. E (FE) denotes models containing regional effects (country fixed effects). Columns (1)–(6) have been estimated by maximum likelihood in a logit regression, column (7) using OLS in a linear model with fixed effects (since the algorithm in the conditional logit regression didn't converge) and column (8) using system GMM.

* p<.10, ** p<.05, *** p<.01

the effect of all inequality unrelated to BGI on conflict, and the coefficient on $\overline{\text{BGI}}$ estimates the effect only of inequality across groups (and not of BGI through Gini). We find that the effect of inequality unrelated to $\overline{\text{BGI}}$ is positive but insignificant, and that the coefficient on $\overline{\text{BGI}}$ is positive but also measured with substantial error. Model 3 includes $\overline{\text{WGI}}$ on the right-hand side (and includes $\overline{\text{GINI-WGI}}$ as the control for non-WGI inequality). We find that $\overline{\text{WGI}}$ has a positive and significant coefficient. The coefficient on $\overline{\text{GINI-WGI}}$ has a negative sign but it is not precisely estimated. Column 4 includes all three components of the Gini separately. The coefficient of $\overline{\text{WGI}}$ remains very similar to that in column 3. Overlap has a negative coefficient whereas $\overline{\text{BGI}}$ has a positive one, but neither coefficient is precisely estimated. Models 5 and 6 are similar to model 4 but employ alternative ways of controlling for the nature of the political system. Column 5 uses dummy variables for anocracies and democracies (computed from Polity IV) to capture potential nonlinearities in democracy score measures (Fearon and Laitin, 2003).[23] Column 6 introduces xPolity and its square value for a similar purpose. The results for our key variables, WGI, BGI and Overlap remain unchanged.

Finally, models 7 and 8 exploit within-country variation in the inequality measures. As described in Section 3, for many countries, there exist measures of inequality for more than one year. Thus, it is possible to construct a (very unbalanced) panel with time-varying inequality measures. The obvious advantage of doing this is that we can introduce country fixed effects in the regressions, reducing the risk of omitted variable bias. However, the results should be taken with caution since the size of the sample in these regressions shrinks dramatically. Column 7 examines the Gini decomposition variables with all time-varying controls in column 3. We also estimated this model without the lagged conflict variable, producing nearly identical results. Both columns are estimated in a linear model since conditional logit does not converge. Column 8 uses system GMM to control for the endogeneity of lagged conflict. The results are qualitatively identical to those obtained in the previous columns: WGI_t has a positive coefficient and is significant at the 5% level in column 7 (as well as in the model without lagged conflict) and at the 10% level in column 8 (p-value is 0.065), while the coefficients for BGI_t and OV_t are imprecisely measured in both models.

The estimated effect of within group inequality in not only precisely estimated, it is sub-

[23]Vreeland (2008) argues that the connection between anocracies and conflict stems from the fact that certain components of the Polity index are linked to political violence so, not surprinsingly, they exhibit a strong relationship with civil war. This is why in our baseline specifications we use the xPolity variable, which has these components removed.

stantively large. Using the results from column 4, moving from the median country's $\overline{\text{WGI}}$ (Cyprus, .182) to the country in the 90th percentile (Vietnam, .283) while holding holding other variables at their means, the predicted probability of experiencing conflict (i.e, the probability of observing strictly positive values of PRIO25) rises from .052 to .174, which implies an increase of more than 300%.[24]

The country-level results provide preliminary support for the arguments in Esteban and Ray about within-group inequality and conflict, and they cast doubt on arguments that inter-group grievances spur conflict. But each of these arguments is more appropriately tested at the group-level, something we undertake in the next section. Here we wish to underline that the results in Table 3 can illuminate why it has been so difficult to find empirical relationships between overall inequality and conflict.

When we consider the results for the Gini and its three components that are based on individual ethnic or religious identities, we find that only WGI has a robust positive relationship with conflict. As described in Section 4, although WGI is the largest component of the Gini, it explains little of the Gini's variation – it has a small coefficient of variation and unlike BGI and OV, it has a low correlation with Gini. But the two components which are most closely associated with variation in Gini have no relationship with conflict. Their coefficients are always measured with considerable error and typically have opposite signs. Thus, in spite of the strong connection between WGI and conflict, it cannot be captured through the Gini coefficient. This underscores the importance of examining all three components of the Gini decomposition in order to uncover the relationship between inequality and conflict. More generally, for any other dependent variable that is typically associated with inequality (eg., development, public good provision, corruption, etc.), in ethnically heterogeneous societies, it seems crucial to consider the Gini decomposition in order to provide a meaningful interpretation of the relation between those variables and inequality.

[24]This interpretation of the magnitude of the effect of WGI draws on the cross-country distribution of the inequality components. Although for a particular country it is difficult to change one of the Gini components while leaving the others 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 data set 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.

5.3 Robustness checks

This section describes some additional robustness checks that confirm the strong link between WGI and the incidence and intensity of conflict at the country level, and the absence of a robust relationship between BGI or OV and conflict. Tables in this section can be found in Appendix A.

Using similar controls as column 4 in Table 3, Table A.2 shows that the results in the former table are robust to using other conflict variables, more specifically, PRIOINT (first 4 columns) and PRIOCW (last 4 columns). Models containing regional as well as fixed effects are considered. The results are very similar to those in Table 3: the Gini is not significantly related to conflict but WGI is positively and significantly associated with it (at least at the 10% level, with the exception of the fixed effects regression with PRIOCW estimated by OLS). The other two components, BGI and Overlap, are not significant or, when they are so, have the wrong sign.

We have also considered alternative ways of controlling for the heterogeneity in our surveys. The results are presented in Table A.3, where the first 4 columns have PRIO25 as dependent variable and the last 4, PRIOINT. Columns 1, 2, 4 and 5 use a conventional correction to adjust the observations from the different surveys. Each of the (unadjusted) Gini components have been regressed on country, time and survey dummies. Next, the coefficients of the survey dummies are employed to correct for systematic (average) differences across surveys. The reference category in columns 1 and 4 (that is, the survey dummy that is omitted) is HES. Since the number of HES observations is small and the analysis in Section 3 revealed that the WVS produced estimates of the Gini and its subcomponents that are reasonably close to those of SWIID and HES, the inequality variables in columns 2 and 5 have been adjusted using HES and WVS as reference survey types (i.e., both the HES and WVS dummies have been omitted in the adjusting regressions). Finally, columns 3 and 6 use the original inequality variables without any adjustment. The conclusions are robust to all these variations.

6 Group-level analysis of civil conflict

We now turn to the group-level analysis, examining whether group-specific attributes related to inequality are associated with the propensity of groups to engage in conflict. In addition to the group-level inequality indices described in Section 3.2.2, we use the following data:

Conflict. Our data on which specific groups engage in conflict are taken from the Ethnic Power Relations data set (EPR, Cederman et al. 2009), and are accessed through the ETH Zurich's GROWup data portal (`http://growup.ethz.ch`).[25] We use two measures of conflict incidence and intensity. PRIO25g is a binary measure taking a value of 1 for those years where an ethnic group becomes involved in armed conflict against the state resulting in more than 25 battle-related deaths.[26] PRIOINTg aims to capture conflict intensity and takes 3 values: 0 if no conflict, 1 if PRIO25g is 1 and the total number of battle deaths that year is smaller than 1000 and 2 if PRIO25g is 1 and the total number of battle deaths exceeds 1000 for that given year.

Controls. In addition to the country-level controls used in the previous section, the group-level regressions also control for GROUPSIZE, which is the relative size of the group, measured using Fearon (2003) group shares.

Table 4 reports the main results of our baseline specification at the group level and provides details about the estimation procedures employed. Column 1 includes the same country-level control variables used in column 4 of Table 3 in the country-level analysis. In addition, it has the three group-level variables, GROUPGINI, HORINEQ and GROUPSIZE. The coefficient for GROUPGINI is positive and significant (p-value is 0.051), indicating that more inequality within a group is associated with more conflict for the group. GROUPGINI is negatively and significantly associated with conflict, and there is no significant relationship between Horizontal Inequality and conflict. The remaining models in Table 4 explore the robustness of these results. Column 2 drops GROUPGINI from the model. When we do this, HORINEQ becomes significant at the 5% level, a result similar to that found in Cederman et al. (2011), which does not control for within-group inequality.[27] Column 3 drops the variables that are not significant in models 1 or 2, including HORINEQ. The results for GROUPGINI remain robust. Column 4 replaces the regional dummy variables with country dummy variables. The magnitude of the GROUPGINI coefficient almost doubles (while that of HORINEQ remains very similar). The remaining models use PRIOINTg as

[25]Although the EPR utilizes a slightly different definition of groups than Fearon, we found it very straightforward to map from the EPR group definitions to the Fearon definitions used here.

[26]Ethnic groups are coded as engaged in conflict if a rebel organization involved in the conflict expresses its political aims in the name of the group and a significant number of members of the group participate in the conflict, see Wucherpfennig et al. (2012) for details.

[27]Cederman et al. (2011) use conflict onset rather than conflict incidence as measure of civil strife. We have also run regressions using onset as dependent variable obtaining similar results as those presented in Table 4: the coefficient of GROUPGINI is positive and significant while that of HORINEQ is not.

Table 4: Group level results. Baseline Specification with PRIO25 and PRIOINT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GROUPGINI	6.574*		5.594**	11.993**	6.272*		5.288*	0.250**
	(3.372)		(2.812)	(5.836)	(3.301)		(2.711)	(0.124)
HORINEQ	0.926	0.891**		1.162	0.819	0.815*		0.031
	(0.630)	(0.437)		(1.967)	(0.624)	(0.452)		(0.024)
GROUPSIZE	-3.178*	-2.986*	-3.679**	-4.134*	-3.534*	-3.281*	-4.066**	-0.061*
	(1.802)	(1.657)	(1.689)	(2.427)	(2.126)	(1.838)	(2.047)	(0.032)
GDP	-0.222	-0.120		0.784	-0.278	-0.168		0.012
	(0.452)	(0.361)		(0.761)	(0.443)	(0.394)		(0.017)
POP	0.280	0.366		-0.131	0.285	0.327		0.001
	(0.305)	(0.327)		(0.517)	(0.355)	(0.355)		(0.041)
F	3.945	3.016		-2.683	4.055	3.080		0.005
	(2.778)	(2.682)		(5.810)	(2.911)	(2.888)		(0.156)
P	-1.636	-0.832		-7.564	-1.338	-0.475		-0.070
	(2.289)	(1.807)		(5.627)	(2.325)	(1.891)		(0.140)
OIL/DIAM	-2.974***	-2.747***	-1.930***	0.067	-2.866***	-2.658***	-1.848***	0.000
	(0.952)	(0.862)	(0.515)	(2.906)	(0.903)	(0.858)	(0.475)	(0.005)
XPOL	0.158**	0.146*	0.100	-0.069*	0.162**	0.146*	0.104	-0.001
	(0.074)	(0.080)	(0.065)	(0.041)	(0.080)	(0.084)	(0.067)	(0.002)
NCONT	-0.450	-0.080		-2.350	-0.543	-0.054		0.040
	(1.343)	(1.575)		(3.694)	(1.464)	(1.716)		(0.143)
MOUNT	0.027**	0.021*	0.031*	0.018	0.028**	0.022*	0.031*	0.001
	(0.014)	(0.012)	(0.017)	(0.031)	(0.013)	(0.012)	(0.017)	(0.002)
LAG	3.099***	2.955***	3.298***	1.964**	2.362***	2.317***	2.554***	0.178
	(0.808)	(0.779)	(0.860)	(0.790)	(0.675)	(0.730)	(0.694)	(0.141)
CONSTANT	-9.067**	-8.277**	-7.103***	-6.462		-7.734*		-0.170
	(3.708)	(4.103)	(1.296)	(11.510)		(4.412)		(0.358)
(Pseudo) R^2	0.328	0.297	0.303	0.316	0.287	0.280	0.265	
Obs	5502	5502	5502	1515	5502	5502	5502	5502
Reg. E./FE	Reg. E.	Reg. E.	Reg. E.	FE.	Reg. E.	Reg. E.	Reg. E.	FE
N. of groups	446	446	446	114	446	446	446	114

Note: Dependent variable is PRIO25g in columns (1)–(4) and PRIOINTg in columns (5)–(8). Robust standard errors (clustered at the group level) are in parentheses. Year indicator variables are included in all models. Reg. E (FE) denotes models containing regional dummies (country dummies). Columns (1)–(4) are estimated in a logit regression while model (5), in an ordered logit (ologit) regression. Models (6) and (8) are estimated in a linear model since the algorithm in the ordered logit regression didn't converge. * p<.10, ** p<.05, *** p<.01

dependent variable. Model 5 uses a specification similar to that in column 1. The magnitude for the coefficient for GROUPGINI is very close to that in column 1 and is still significant at the 10% level (p-value is 0.057). Column 6 drops GROUPGINI and a similar result as in column 2 is obtained, i.e., HORINEQ is significant (at the 10% level) once GROUPGINI is dropped. Column 7 drops the control variables that are not significant in the previous regressions while column 8 replaces the regional fixed effects by country fixed effects. In both cases the coefficient of GROUPGINI is significant at the 5% level. Summarizing, this table shows that heterogeneity in incomes within one group is positively associated with conflict but inequality between groups is not.

6.1 Robustness checks

As a robustness check, we have considered alternative ways of adjusting the group-level inequality measures. Table A.4 in Appendix A shows that the results presented in Table 4 are robust to other ways of adjusting the data (including non-adjustment). The dependent variables are PRIO25g (first four columns) and PRIOINTg (last 4). Columns 1 and 4 use the same specification as column 1 in Table 4), but in this case GROUPGINI is left unadjusted. In columns 2 and 3 both GROUPGINI and HORINEQ are adjusted following a procedure similar to that described in Section 5.3. First, both GROUPGINI and HORINEQ are regressed on country, time and survey dummies; second, the coefficients of the survey dummies are used to correct for systematic differences among survey types. The omitted survey dummy in column 2 is HES while in they are HES and WVS in column 3. Column 4 uses GROUPGINI adjusted as in the baseline specification and HORINEQ adjusted as in (3). The last four columns in Table A.4 have identical structure as the first four but use PRIOINTg as dependent variable. The results are robust to these variations, with the coefficient of GROUPGINI being significant at the 10% level and that of HORINEQ being insignificant.

7 Conclusion

Using individual-level surveys to calculate the components of the Gini coefficient across a wide range of countries, we find that within-group economic differences have a strong association with civil conflict, whereas between-group economic differences do not. These results, which are consistent with Esteban and Ray's argument about the importance of labor and capital for waging conflict, hold at both the country and group level, and are robust to a wide range of models, including the inclusion of country fixed effects, different measures of civil conflict, and different approaches to adjusting the heterogeneity that exists in the measures of income across surveys. The group-level inequality data also help us to understand why previous research has not found a robust relationship between overall inequality and civil conflict. On average, most inequality within countries occurs within ethnic groups, whereas inequality across ethnic groups typically accounts for a relatively small proportion of overall inequality. However, variation in the Gini coefficient itself is strongly correlated with inequality between groups but uncorrelated with inequality within groups. Since inequality between groups has no relationship with civil conflict (when we control for inequality within groups) but inequality within groups has a strong association with conflict, it should be expected that overall inequality has no association with conflict. More generally, the analysis underlines the difficulties in cross-national research that associated with interpreting results from measures of overall inequality because such measures mask quite different types of inequality that exist when group affiliations are taken into consideration. An important challenge is to develop theoretical models that link these different types of group-based inequality to outcomes of importance, such as levels of economic growth, public goods provision, levels of corruption and democratic performance.

8 References

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

Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou. 2013. "Ethnic Inequality." *NBER Working Papers*, 18512, revised on July 1, 2013.

Angel, S. 2012. *Planet of Cities*. Lincoln Institute of Land Policy, MA: Cambridge

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, Brian Min, and Andreas Wimmer. 2009. Ethnic Power Relations dataset, hdl:1902.1/11796.

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.

Cramer, Christopher. 2003. "Does Inequality Cause Conict?" *Journal of International Development*, 15: 397–412.

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

Dahrendorf, R. 1959. *Class and class conflict in industrial society.* Stanford, CA: Stanford University Press.

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

Deininger, Klaus, and Lyn Squire. 1998. "New ways of looking at old issues: inequality and growth." *Journal of Development Economics, 57(2): 259–287.*

Doyle, Michael and Nicholas Sambanis. *2006. Making War and Building Peace. Princeton: Princeton University Press.*

Esteban, Joan, and Debraj Ray. *2008. "On the Salience of Ethnic Conflict." American Economic Review 98: 2185–2202.*

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. *2000. "Violence and the Social Construction of Ethnic Identity." International Organization, Cambridge University Press, vol. 54(04): 845–877.*

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.*

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

Gurr, Ted R. *1970. Why Men Rebel. Princeton: Princeton University Press.*

Gurr, Ted R. *1980. Why Men Rebel. Handbook of Political Conflict: Theory and Research, New York Free Press.*

Kuhn, Patrick M., and Nils B. Weidmann. *2013. "Unequal We Fight: The Impact of Economic Inequality Within Ethnic Groups on Conflict Initiation." Working Paper, Princeton University.*

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.*

Letu, Husi, Masanao Hara, Gegen Tana, Fumihiko Nishio. *2012. "A saturated light correction method for DMSP/OLS nighttime satellite imagery." IEEE Transactions on Geoscience and Remote Sensing, 50: 389-396.*

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.*

Melvern, Linda. *2000. A people betrayed: the role of the West in Rwanda's genocide. London; New*

York, N.Y.: Zed Books.

Mitra, Anirban and Debraj Ray. *2012. "Implications of an economic theory of conflict: Hindu-Muslim Violence in India." Mimeo.*

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

Morelli, Massimo and Dominic Rohner. *2013. "Resource Concentration and Civil Wars." Typescript, Columbia University.*

Nordhaus, William D. *2006. "Geography and macroeconomics: New data and new findings." Proceedings of the National Academy of Sciences of the USA, 103(10): 3510-3517.*

Penn World Table. *2011. Dataset,*

https://pwt.sas.upenn.edu/php_site/pwt_index.php.

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. 29?54.*

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. *2009. Understanding Civil War: Evidence and Analysis: Africa. The World Bank, Vol.1. .*

Sambanis, Nicholas, and Branko Milanovic. *2011. "Explaining the demand for sovereignty." Policy Research Working Paper Series, 5888, The World Bank.*

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

Stewart, Frances. *2000. "Dynamic Interactions between the Macro-Environment, Development Thinking and Group Behaviour." Development Working Papers, 143, University of Milano.*

Stewart, Frances. *2002. "Horizontal Inequalities: A Neglected Dimension of Development." Annual*

Lecture No. 5, UNU World Institute for Development Economics Research.

Tilly, Charles. *1978. From mobilization to revolution. Reading, MA: Addison-Wesley.*

Verwimp, Philip. *2002. "An economic profile of peasant perpetrators of genocide. Micro-level evidence from Rwanda." Journal of Development Economics, 77: 297-323.*

Vreeland, James R. *2008. "The Effect of Political Regime on Civil War". In Journal of Conflict Resolution, 52(3): 401–425.*

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.*

World Bank. *2013. Dataset,*

http://iresearch.worldbank.org/PovcalNet/index.htm?0,2.

Wucherpfennig, Julian, Nils W. Metternich, Lars-Erik Cederman, and Kristian S. Gleditsch. *1991. "Ethnicity, the state, and the duration of civil war". World Politics, 64(1): 79–115.*

Yanagizawa-Drott, David. *2012. "Propaganda and conflict: Theory and Evidence from the Rwandan Genocide." Harvard University.*

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.*

A Appendix (For Online Publication)

Table A.1: Country-level results. Inequality (Gini) and conflict (PRIO25)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GINI	3.110	-8.358	0.024	-0.044	10.390	-0.615	3.234
	(1.949)	(7.572)	(0.020)	(0.063)	(6.469)	(0.572)	(2.951)
F	1.671*		2.009**		6.047***		2.676**
	(0.857)		(0.871)		(1.867)		(1.219)
P	1.291		1.416		0.623		1.517
	(0.886)		(0.894)		(1.392)		(1.002)
NCONT	0.746		0.667		1.187		1.098
	(0.466)		(0.470)		(0.894)		(0.671)
MOUNT	0.013**		0.010*		0.002		0.011
	(0.006)		(0.006)		(0.010)		(0.009)
GDP	-0.143	1.635	-0.163	2.024	-0.665	-0.151	-0.281
	(0.181)	(1.827)	(0.189)	(2.191)	(0.612)	(0.136)	(0.254)
POP	0.376***	-7.174	0.409***	-7.924	1.042***	-0.626*	0.400***
	(0.095)	(4.643)	(0.101)	(6.145)	(0.345)	(0.350)	(0.132)
XPOL	-0.014	-0.122**	-0.023	-0.129**	-0.018	-0.028**	0.031
	(0.037)	(0.057)	(0.039)	(0.060)	(0.070)	(0.014)	(0.041)
OIL/DIAM	-0.373		-0.450		-2.034*		-0.294
	(0.325)		(0.357)		(1.062)		(0.337)
CONST	4.563***		4.573***		4.875***		4.655***
	(0.456)		(0.468)		(0.994)		(0.624)
CONST	-9.142***		-9.510***		-17.345***	7.726*	-9.334***
	(1.800)		(1.950)		(6.048)	(4.152)	(2.440)
(Pseudo) R^2	0.625	0.098	0.623	0.097	0.646	0.099	0.626
Obs	1652	451	1590	419	392	411	1044
Countries	138	37	136	38	101	106	88
Reg. E./ FE	Reg. E.	FE	Reg. E.	FE	Reg E.	FE	Reg.E

Note: Dependent variable is PRIO25. Robust standard errors (clustered by country) are in parentheses. Gini in columns (1) to (4) comes from SWIID and is computed using net income in columns (1) and (2) and gross income in columns (3) and (4). The source of Gini in columns (5) and (6) is the World Bank (Povcalnet) while in column (7) is the Gini based on surveys introduced in this paper. Year indicator variables are included in all models. Reg. E (FE) denotes models containing regional effects (country fixed effects). Columns (1), (3), (5) and (7) have been estimated in a logit regression, columns (2), (4) by conditional logit and column (6) in a linear model containing fixed effects, since conditional logit didn't converge.
* p<.10, ** p<.05, *** p<.01

Table A.2: Country-level results. Other measures of conflict: PRIOINT and PRIOCW

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\overline{\text{GINI}}$	3.505				1.626			
	(2.594)				(3.272)			
$\overline{\text{BGI}}$		1.240				-14.639***		
		(4.971)				(5.568)		
BGI_t			-0.214	0.001			0.230	0.117
		(4.971)	(0.458)				(0.332)	(0.267)
$\overline{\text{WGI}}$		9.992*				33.502***		
		(5.841)				(9.675)		
WGI_t			0.762*	0.762***			0.473	0.904***
			(0.392)	(0.281)			(0.355)	(0.331)
$\overline{\text{OV}}$		-3.118				-38.680***		
		(6.288)				(9.337)		
OV_t			0.345	0.215			0.556	0.426
			(0.395)	(0.433)			(0.399)	(0.324)
F	2.021*	6.279*			3.658**	28.950***		
	(1.075)	(3.647)			(1.453)	(6.164)		
P	1.440	1.692*			1.251	3.096**		
	(1.027)	(0.975)			(1.161)	(1.225)		
NCONT	0.765	1.073*			1.041	3.124***		
	(0.614)	(0.616)			(0.669)	(0.820)		
MOUNT	0.011	0.010			0.009	0.007		
	(0.008)	(0.008)			(0.010)	(0.011)		
GDP	-0.228	-0.216	-0.602**	0.030	0.198	0.199	-0.282*	0.024
	(0.217)	(0.234)	(0.237)	(0.025)	(0.272)	(0.252)	(0.163)	(0.022)
POP	0.356***	0.299**	-0.424	0.032	0.639***	0.471***	-0.538	0.045***
	(0.133)	(0.139)	(0.533)	(0.020)	(0.178)	(0.177)	(0.354)	(0.017)
XPOL	0.017	0.010	-0.012	0.014	-0.021	-0.025	-0.005	0.006
	(0.037)	(0.040)	(0.021)	(0.009)	(0.061)	(0.069)	(0.012)	(0.011)
OIL/DIAM	-0.420	-0.422			-0.779**	-1.069**		
	(0.321)	(0.353)			(0.391)	(0.436)		
LAG	3.943***	3.843***	0.361***	0.695***	6.470***	5.945***	0.444***	0.707***
	(0.465)	(0.455)	(0.105)	(0.064)	(0.745)	(0.736)	(0.163)	(0.108)
CONSTANT			9.483	-0.271	-15.447***	-26.242***	7.464*	-1.028***
			(5.868)	(0.475)	(2.795)	(4.893)	(4.107)	(0.302)
(Pseudo) R^2	0.560	0.563	0.457	–	0.780	0.796	0.387	–
Obs.	1044	1044	210	214	1042	1042	210	214
Reg. E/FE	Reg. E	Reg. E	FE	FE	Reg. E	Reg. E	FE	FE

Note: Dependent variable is PRIOINT in columns (1)–(4) and PRIOCW in columns (5)–(8). Robust standard errors (clustered by country) are in parentheses. Year indicator variables are included in all models. Reg. E (FE) denotes models containing regional effects (country fixed effects). Columns (1)–(2) and (5)–(6) have been estimated in an ordered logit and a logit regression, respectively; Columns (3) and (7) by OLS in a model containing fixed effects, and models (4) and (8) by system GMM.
* $p<.10$, ** $p<.05$, *** $p<.01$

Table A.3: Country-level results. Alternative ways of addressing survey heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)
$\overline{\text{BGI}}$	5.468	5.975	5.368	3.279	3.818	2.675
	(6.042)	(6.204)	(7.624)	(5.866)	(5.965)	(7.314)
$\overline{\text{WGI}}$	23.265***	17.790**	16.883**	16.443**	12.726**	12.327**
	(8.073)	(7.226)	(6.997)	(7.172)	(6.142)	(5.991)
$\overline{\text{OV}}$	2.302	-1.307	-4.148	0.888	-1.125	-2.912
	(7.094)	(6.250)	(5.827)	(6.984)	(6.153)	(5.705)
GDP	-0.201	-0.105	-0.204	-0.105	-0.018	-0.065
	(0.275)	(0.318)	(0.309)	(0.233)	(0.269)	(0.255)
POP	0.252*	0.268*	0.300**	0.247*	0.254*	0.278*
	(0.139)	(0.150)	(0.152)	(0.142)	(0.146)	(0.147)
F	8.992***	7.832**	8.126**	6.195**	5.351*	5.706**
	(3.337)	(3.331)	(3.317)	(2.818)	(2.732)	(2.777)
P	2.211*	2.620**	2.384*	1.638	1.944	1.696
	(1.252)	(1.331)	(1.371)	(1.226)	(1.299)	(1.327)
NCONT	2.496***	2.877***	2.799***	1.624***	1.869***	1.793***
	(0.644)	(0.694)	(0.691)	(0.598)	(0.646)	(0.649)
MOUNT	0.023**	0.022**	0.021**	0.017***	0.017**	0.017**
	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
XPOL	0.027	0.045	0.049	0.006	0.019	0.020
	(0.046)	(0.048)	(0.048)	(0.040)	(0.042)	(0.041)
OIL/DIAM	-0.212	-0.373	-0.374	-0.414	-0.521	-0.523
	(0.406)	(0.436)	(0.446)	(0.387)	(0.417)	(0.427)
LAG	4.429***	4.358***	4.343***	3.781***	3.740***	3.733***
	(0.582)	(0.580)	(0.573)	(0.428)	(0.427)	(0.424)
CONSTANT	-16.927***	-14.675***	-13.412***	–	–	–
	(4.565)	(4.491)	(4.452)			
(Pseudo) R^2	0.638	0.638	0.639	0.567	0.568	0.568
Obs	1044	1044	1044	1044	1044	1044

Note: Dependent variable is PRIO25 in columns (1)–(3) and PRIOINT in columns (4)–(6). There are 88 countries. Robust standard errors (clustered by country) are in parentheses. The inequality variables in columns (1), (2), (4) and (5) have been adjusted by subtracting average survey differences (reference survey type is HES in columns (1) and (4) and WVS and HES in columns (2) and (5). In columns (3) and (6) they are the original (unadjusted) inequality measures. Year and regional indicator variables are included in all models. Estimation has been carried out by max. lik. in a logit regression in columns (1)–(3) and in an ordered logit regression in columns (4)–(6).
* p<.10, ** p<.05, *** p<.01

Table A.4: Group-level results. Alternative ways of of addressing survey heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GROUPGINI	8.844*	9.160*	10.110*	6.338*	8.798*	9.045*	10.103*	6.060*
	(4.941)	(5.450)	(5.563)	(3.345)	(4.917)	(5.497)	(5.558)	(3.283)
HORINEQ	0.480	0.709	0.503	0.932	0.399	0.611	0.407	0.824
	(0.649)	(0.701)	(0.673)	(0.607)	(0.630)	()0.699)	(0.657)	(0.604)
GROUPSIZE	-3.036*	-3.011*	-3.040*	-3.147*	-3.359*	-3.391*	-3.394*	-3.515*
	(1.677)	(1.713)	(1.701)	(1.795)	(1.928)	(2.027)	(1.999)	(2.128)
GDP	0.040	0.017	0.218	-0.202	-0.019	-0.035	0.161	-0.259
	(0.440)	(0.429)	(0.449)	(0.448)	(0.459)	(0.445)	(0.461)	(0.440)
POP	0.222	0.133	0.069	0.266	0.218	0.138	0.071	0.273
	(0.321)	(0.269)	(0.289)	(0.306)	(0.381)	(0.311)	(0.324)	(0.357)
F	4.586	3.739	4.599	3.974	4.789	3.862	4.777	4.080
	(3.410)	(2.826)	(3.096)	(2.766)	(3.512)	(2.953)	(3.173)	(2.899)
P	-1.117	-1.139	-0.924	-1.648	-0.887	-0.908	-0.681	-1.342
	(2.184)	(2.110)	(2.088)	(2.285)	(2.230)	(2.185)	(2.133)	(2.321)
XPOL	0.176*	0.155*	0.166**	0.160**	0.182*	0.158*	0.170**	0.163**
	(0.092)	(0.083)	(0.080)	(0.074)	(0.098)	(0.088)	(0.086)	(0.080)
OIL/DIAM	-3.711***	-2.845***	-3.214***	-2.994***	-3.643***	-2.778***	-3.148***	-2.883***
	(1.421)	(0.929)	(1.061)	(0.967)	(1.412)	(0.889)	(1.038)	(0.916)
NCONT	0.123	-0.081	0.226	-0.408	0.007	-0.189	0.088	-0.508
	(1.340)	(1.377)	(1.235)	(1.338)	(1.435)	(1.484)	(1.294)	(1.460)
MOUNT	0.025*	0.025*	0.028**	0.027**	0.025**	0.026*	0.029**	0.027**
	(0.013)	(0.014)	(0.014)	(0.013)	(0.012)	(0.014)	(0.014)	(0.013)
LAG	2.933***	3.003***	2.958***	3.110***	2.250***	2.288***	2.270***	2.370***
	(0.786)	(0.821)	(0.795)	(0.807)	(0.654)	(0.696)	(0.659)	(0.659)
CONSTANT	-11.555**	-11.669**	-12.612**	-9.024**	–	–	–	–
	(5.074)	(4.700)	(5.002)	(3.703)				
(Pseudo) R^2	0.329	0.328	0.334	0.329	0.290	0.288	0.294	0.287
Obs	5502	5502	5502	5502	5502	5502	5502	5502

Note: Dependent variable is PRIO25g in columns (1)–(4) and PRIOINTg in columns (5)–(8). Robust standard errors (clustered at the group level) are in parentheses. Inequality measures are left unadjusted in columns (1) and (5). GROUPGINI has been adjusted in columns (2), (3), (6) and (7) by subtracting average survey differences (reference survey type is HES in columns (2) and (6) and WVS and HES in columns (3) and (7)). In columns (4) and (8) GROUPGINI has been adjusted as in the baseline specification and HORINEQ, as in column (2).
Year and regional indicator variables are included in all models. Estimation has been carried out in a logit regression in columns (1)–(3) and in an ordered logit regression in columns (4)–(6).
* p<.10, ** p<.05, *** p<.01

Table A.5: SUMMARY STATISTICS

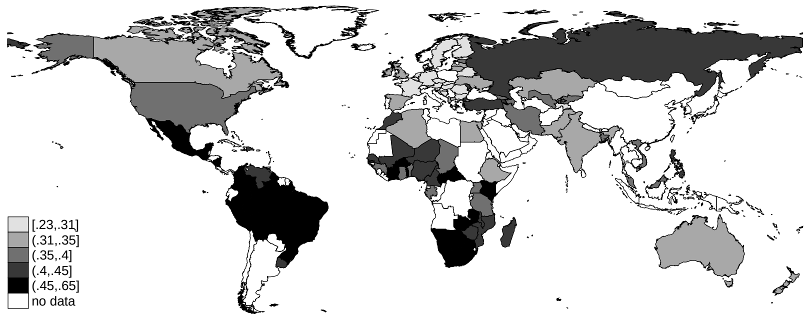
	Obs.	Mean	Std. Dev.	Min.	Max.
Inequality variables					
$\overline{\text{GINI}}$	1044	0.388	0.092	0.231	.652
$\overline{\text{BGI}}$	1044	0.061	0.055	0.001	0.307
$\overline{\text{WGI}}$	1044	0.179	0.081	0.011	0.378
$\overline{\text{OV}}$	1044	0.148	0.086	0.011	0.399
BGI_t	210	0.067	0.072	0.000	0.564
WGI_t	210	0.183	0.085	0.011	0.433
Ov_t	210	0.150	0.094	0.010	0.454
GROUPGINI	5154	0.380	0.124	0	0.940
HORINEQ	5154	0.105	0.249	0.004	2.28
Conflict variables					
PRIO25	1044	0.183	0.387	0	1
PRIOINT	1044	0.212	0.476	0	2
PRIOCW	1042	0.147	0.358	0	1
PRIO25g	5502	0.028	0.164	0	1
PRIOINTg	5502	0.031	0.191	0	2
Controls					
F	1044	0.507	0.240	0.077	0 .953
P	1044	0.580	0.199	0.154	0.986
NCONT	1044	0.155	0.362	0	1
MOUNT	1044	15.950	19.610	0	81
GDP	1044	8.400	1.350	5.620	10.830
POP	1044	9.628	1.361	6.631	13.950
XPOL	1044	3.503	4.030	-5	7
ANOC	1044	0.294	0.456	0	1
DEMOC	1044	0.615	0.487	0	1
OIL/DIAM	1046	0.266	0.442	0	1
SIZE	5502	0.190	0.266	0.000	0.994

Table A.6: Inequality Surveys

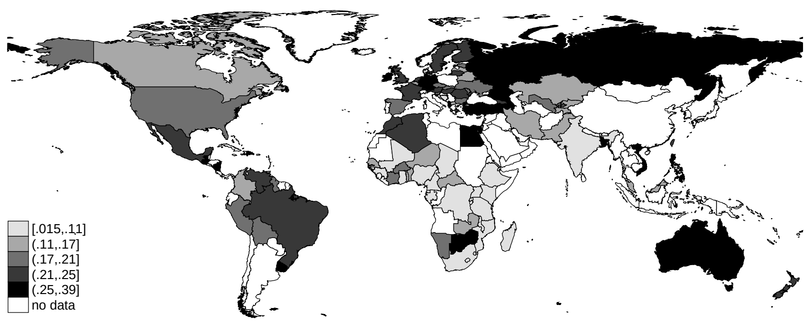
Albania	2002(WVS) 2005(HES-LSMS)	Kyrgyz Rep	1997(DHS) 2003(WVS)
Algeria	2002(WVS)	Latvia	1996(WVS) 1999(WVS)
Armenia	1997(WVS) 2000(DHS)	Lithuania	1997(CSES, WVS)
Australia	1995(WVS) 1996(CSES) 2004(CSES) 2005(WVS)	Macedonia	1998(WVS) 2001(WVS)
Austria	2000(LIS)	Madagascar	2005(AFRO)
Azerbaijan	1995(HES-ASLC) 1997(WVS) 2006(DHS)	Malawi	2000(DHS) 2003(AFRO) 2004(DHS) 2005(AFRO)
Bangladesh	1996(WVS) 1997(DHS) 2000(DHS) 2002(WVS) 2004(DHS) 2007(DHS)	Malaysia	2006(WVS)
Belarus	1996(WVS) 2001(CSES)	Mali	1995(DHS) 2001(DHS) 2002(AFRO) 2005(AFRO) 2006(DHS)
Belgium	1999(CSES, WVS)	Mexico	1997(CSES, WVS) 2000(WVS) 2003(CSES)
Benin	1996(DHS) 2001(DHS) 2005(AFRO) 2006(DHS)	Moldova	1996(WVS) 1999(WVS) 2005(DHS) 2006(WVS)
Bolivia	2002(HES-MECOVI) 2003(DHS)	Morocco	2001(WVS) 2007(WVS)
Bosnia	1998(WVS) 2001(WVS) 2004(HES-LIBP)	Mozambique	2002(AFRO) 2005(AFRO)
Botswana	2003(AFRO) 2005(AFRO)	Namibia	2000(DHS) 2003(AFRO) 2006(AFRO)
Brazil	1996(DHS) 1997(WVS) 2002(CSES, HES-IPUMS) 2006(WVS, HES-PNAD)	Netherlands	1999(WVS)
Bulgaria	1995(HES-IHS) 1997(WVS) 2001(CSES) 2006(WVS)	New Zealand	1996(CSES) 1998(WVS) 2002(CSES)
Burkina Faso	1992(DHS) 1998(DHS, HES-EP2) 2003(DHS)	Nicaragua	2001(HES-EMNV)
Cameroon	1998(DHS) 2004(DHS)	Niger	1992(DHS) 1998(DHS) 2006(DHS)
Canada	1997(CSES, HES) 2000(WVS) 2001(HES-IPUMS) 2006(WVS)	Nigeria	2000(WVS) 2005(AFRO)
Central African Rep	1994(DHS)	Pakistan	2001(WVS)
Chad	1997(DHS) 2004(DHS)	Peru	2000(DHS) 2004(DHS, HES) 2008(WVS)
Colombia	1998(WVS)	Philippines	1993(DHS) 1998(DHS) 2003(DHS) 2008(DHS)
Cote d'Ivoire	1998(DHS)	Romania	1996(WVS, CSES) 1997(HES) 2005(WVS)
Cyprus	2006(WVS)	Russia	1995(WVS) 1999(CSES) 2000(CSES, HES) 2006(WVS)
Czech Rep	1996(CSES)	Senegal	1992(DHS) 2002(AFRO) 2005(AFRO, DHS)
Dominican Rep	1998(WVS)	Singapore	2002(WVS)
DRC	2007(DHS)	Slovakia	1998(WVS)
Egypt	1995(DHS) 2000(WVS) 2005(DHS) 2008(DHS)	Slovenia	1996(CSES)
Estonia	1996(WVS) 1999(WVS) 2000(HES)	Spain	1995(WVS) 1996(CSES) 2000(CSES, WVS) 2004(CSES) 2007(WVS)
Ethiopia	2000(DHS) 2005(DHS)	South Africa	1996(WVS) 1998(DHS) 2001(HES-IPUMS) 2002(AFRO) 2006(AFRO) 2007(WVS)
Finland	2003(CSES) 2004(HES) 2005(WVS)	Sweden	2005(HES) 2006(WVS)
France	1999(WVS) 2002(CSES) 2006(WVS)	Taiwan	1995(WVS) 1996(CSES) 2004(CSES)
Gabon	2000(DHS)	Tajikistan	1996(HES-LSS)
Georgia	1996(WVS)	Tanzania	1993(HES-HRDS)
Germany	1999(WVS) 2004(HES) 2006(WVS)	Togo	1998(DHS)
Ghana	1993(DHS) 1998(DHS) 2003(DHS) 2008(DHS)	Turkey	1993(DHS) 2007(WVS)
Guatemala	1995(DHS) 1998(DHS) 2000(HES-ENCOVI) 2005(WVS) 2006(HES)	Uganda	1995(DHS) 2005(AFRO)
Guinea	1999(DHS) 2005(DHS)	UK	2004(HES)
Guyana	2005(DHS)	Ukraine	1996(WVS) 1998(CSES) 2006(WVS)
Hungary	2002(CSES)	United States	1996(CSES) 1997(HES) 2000(WVS) 2004(CSES) 2005(HES-IPUMS) 2006(WVS)
India	1995(WVS) 2001(WVS) 2006(WVS)	Uruguay	1996(WVS) 2006(WVS)
Iran	2007(WVS)	Uzbekistan	1996(DHS)
Ireland	1999(WVS)	Venezuela	1996(WVS) 2000(WVS)
Israel	1995(HES-IPUMS) 2005(HES)	Vietnam	1997(DHS) 2002(DHS) 2005(DHS)
Kazakhstan	1995(DHS) 1999(DHS)	Zambia	1996(DHS) 2001(DHS) 2003(AFRO) 2005(AFRO) 2007(WVS, DHS)
Kenya	1993(DHS) 1998(DHS) 2003(DHS, AFRO) 2005(AFRO) 2008(DHS)	Zimbabwe	2001(WVS) 2004(AFRO) 2005(AFRO)

Figure A.2: Inequality and its subcomponents

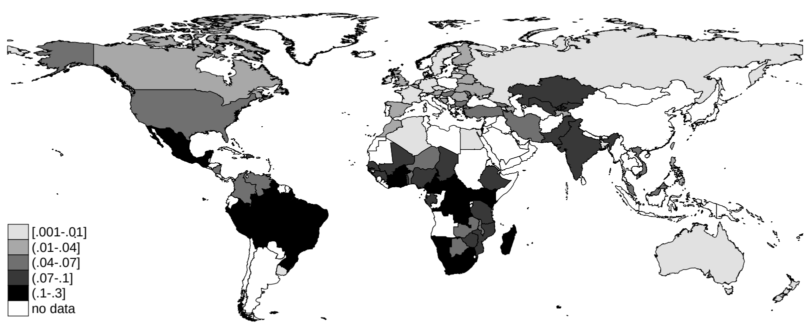
(a) Gini



(b) Within-group inequality



(c) Between-group inequality



(d) Overlap