Inequality, Ethnicity and Civil Conflict

John D. Huber and Laura Mayoral*

July 7, 2014

Abstract

We explore the connection 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, there is a strong positive association between the level of inequality within a group and the group's propensity to engage in civil conflict. In addition, we find that countries with higher levels of inequality within ethnic groups are most likely to experience civil wars. By contrast, inequality across ethnic groups is not associated with the civil conflict. By breaking down measures of inequality into group-level components, the analysis also reveals why it is difficult to identify a relationship between general inequality and conflict, and it highlights more generally why it will often be difficult to draw substantive conclusions in cross-national research 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 it should be unsurprising if standard measures of overall inequality are not associated with civil conflict because such measures do not 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. First, a central focus in existing studies that examine inequality and the engagement of ethnic groups in conflict have focused on group grievances, and thus on "horizontal inequality" – on how the average level of well-being in a group affects group incentives to engage in conflict (Stewart 2002, Cederman et al., 2011). As we discuss below, however, theoretical expectations about horizontal inequality are not unambiguous. If one group is particularly poor, for example, it may lack the means to wage violence. And re-

[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.

cent empirical research has found that an increase in the income of poorer groups is associated with an intensification of conflict. Although we estimate the effects of horizontal inequality in our analysis below, our empirical focus, inspired by a the theoretical model in Esteban and Ray (2008 and 2011), focuses instead on the *ability* of groups to sustain conflict. To this end, we focus our attention on inequality *within* groups. 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. Using group-level models, we find strong support 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 group participation in conflict.

Second, if groups that have high levels of inequality are more likely to engage in conflict, then we might expect that countries that have high levels of group-based inequality will have a higher incidence of civil conflict. We test this possibility by also estimating models at the country level. 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. In our country-level empirical models, 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, we show that its variability is considerably smaller than that of the other two components, and that its correlation with the Gini coefficient is small. If inequality within groups is central to conflict, 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. Our analysis therefore sheds light on why it should be difficult to find a relationship between measures of overall inequality, such as the Gini coefficient, and conflict.

A by-product of this effort represents our third contribution: a new data set on inequality that uses individual-level surveys to measure the three components of the Gini in 89 countries.[3] We draw on a wide range of surveys, including high quality household expenditure surveys from the

[3]Baldwin and Huber (2010) also use surveys to measure group-based inequality, but they use a far smaller number of countries, do not utilize surveys that include household expenditures, and do not provide group-level data.

Luxembourg Income Study and other similar household expenditure surveys. To obtain measures for a large number of groups and countries, however, we also utilize surveys that gauge economic well-being less precisely. Our analysis therefore invokes two standard approaches for adjusting the inequality measures to account for survey heterogeneity, and the analysis utilizes measures resulting from both approaches to assess robustness. Although this approach is not without its limitations, it also has advantages over existing approaches that utilize the spatial location of groups to measure group-based inequality. We discuss the trade-offs below.

The paper is organized as follows. Section 2 describes the relevant existing theoretical and empirical literature on inequality, group identity and civil conflict, and provides illustrative examples. Sections 3-5 focus on data and measurement. Section 3 describes the inequality measures we use, as well as surveys used to construct these measures. Section 4 describes the two approaches used to address heterogeneity in the survey measures of economic well-being, and section 5 discusses the strengths and weaknesses of the survey approach and the main alternative in the literature, which centers on the spatial location of groups. Our core analysis follows in section 6, where we estimate group-level models of conflict. This is followed in section 7 by country level analysis. Section 8 concludes.

2 Group-based inequality and conflict

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, such as ethnicity or religion (e.g., 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.

Previous research emphasizes the role of both rich and poor in ethnic conflict. Typically, the rich ethnic elites instigate conflict for their own benefit, and they provide funds for combat labor. Fearon and Laitin (2000, p. 846), for example, note that "a dominant or most common narrative...is that large-scale ethnic violence is provoked by elites seeking to gain, maintain or increase their hold on political power." Brass (1997) argues that opportunistic leaders are often responsible for publicly coding existing disputes as "communal violence" and that this coding serves to foster larger scale communal violence. In addition, several writers have noticed that financial support from diaspora communities is one of the most significant factors that fuel ethnic conflict (Anderson 1992, Carment 2007). And there is considerable evidence suggesting that fighters in ethnic conflicts are recruited from the poor. As noted by Brubaker and Laitin (1998) most ethnic leaders are well educated and from middle-class backgrounds while the lower-ranking troops are more often poorly educated and from working-class backgrounds. In their study of Sierra Leone's civil war Humphreys and Weinstein (2008) find that factors such as poverty, a lack of access to education, and political alienation are good predictors of conflict participation and that they may proxy, among other factors, for a greater vulnerability to political manipulation by elites. Justino (2009) also emphasizes that poverty is a leading factor in explaining participation in ethnic conflict.

Esteban and Ray (2008, 2011) (henceforth "ER") develop a theory about ethnic violence that explicitly analyzes the role of rich and poor within a group. Their main argument is highly intuitive: effectiveness in conflict requires various inputs, most notably, financial support and labor (i.e, fighters). 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. It has also been argued that heterogeneity in incomes within a group 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.

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). However, in their study on conflict participation, Humphreys and Weinstein (2008) challenge this interpretation since the factors usually associated with grievance-based accounts (poverty, political alienation, etc.) predict violent action in both rebellion and counterrebellion, whose goal is to defend the status quo.

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 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.

2.1 Existing empirical studies

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 within-group inequality 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, the main dependent variable in KH's study is conflict onset. As mentioned before, ER's theory does 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 and use conflict onset as a robustness check. Finally, KW's methodology for computing within group inequality using nightlight emissions has limitations (see below for a description) that the use of survey-based data can help alleviate.

With respect to horizontal inequality, Stewart (2002) use case studies to document a positive connection between horizontal inequality and conflict, as do many essays in Stewart (2008). 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 examples from Africa, Asia and Europe to illustrate the causal mechanisms in their theory. In their survey of the literature on ethnic conflict, Fearon and Laitin (2000) report several examples where the elites promote ethnic conflict and combatants are recruited from the lower class to carry out the killings. Summazing the accounts in Brass (2007), Fearon and Laitin (2000) conclude,

> [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)

Fearon and Laitin (2000) provide examples of this behavior from 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), Sri Lanka (where the ethnic war on the ground was fought on the Sinhalese side by gang members), and Burundi. A more recent example can be found in Ukraine, where Rinat Akhmetov, its richest man, has sent thousands of his own steelworkers to establish control of the streets in Eastern Ukraine in opposition to the pro-Kremlin militants.

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.

3 Measuring ethnic inequality using surveys

To compute measures of ethnic inequality we need data on the joint distribution of income and ethnicity. We draw on individual level surveys containing such data. A challenge associated with this approach lies in identifying surveys from a large number of countries with information on group identity and economic well-being. Ideally, surveys would have fine-grained income or household expenditure data, but unfortunately the number of surveys with such information is quite small (and as we note below, in some contexts even such fine-grained data masks important levels of inequality among the least well-off). We are therefore left with a trade-off: (1) cast a wide net to include as many countries as possible and face the issue that different surveys will take different approaches to measuring economic well-being, or (2) cast a narrow net, focusing on countries that have comparable, high-quality measures of economic well-being, but face the problem of a small set of countries. Our main approach is to cast the wide net, and then to implement two existing approaches to account for heterogeneity in the measures of individual economic well-being. We will also present results that rely exclusively on the World Values Surveys, and thus that do not have issues associated with survey heterogeneity.

3.1 The surveys

Casting the wide net to include a variety of surveys yields three different categories of surveys. The first category, which we refer to as HES (for "Household Expenditure Survey") includes the best surveys available in the world for calculating inequality. These include the Luxembourg Income Study, the Living Standards Monitoring Surveys, other similar household expenditure surveys, as well as a handful of national censuses. The second type of survey uses 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, which is conducted in relatively poor countries, 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 poor individuals whom do not make substantial cash transactions, and thus where individual income cannot be used to meaningfully distinguish the economic well-being of many individuals from each other. In such cases, social scientists often use 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 (AFRO). We use the household assets to measure individual economic well-being. 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 household "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.

One concern about surveys is that they may fail to represent accurately the ethnic structure of a country. To identify the relevant ethnic groups in a country, we rely on the list of groups from Fearon (2003), who provides a set of clear and reasonable criteria for identifying the socially relevant ethnic, religious, racial and/or linguistic groups across a wide range of countries that is widely used in the literature. We use identity questions from the surveys to code a respondent's "ethnic group." Since the relevant identity categories from Fearon (2003) could be related to ethnic identity, religion, race or language, different variables are used in different surveys to map the respondents to the Fearon groups.[4] We discard surveys that do not adequately map to the Fearon groups. Specifically, if there exist one or more groups on Fearon's list that we cannot identify in the survey, we sum the proportion of the population that these groups represent per Fearon's data. If this sum is greater than .10, we do not utilize the survey.[5]

[4]For example, we have a DHS survey from 1997 in Bangladesh. Fearon lists two ethnic groups in Bangladesh as Bengalis (87.5 percent of the population per Fearon) and Hindus (10.5 percent). The DHS survey has a religion variable where 89.7 percent of respondents are Muslim, 0.26 percent are Buddhist, 0.16 percent are Christian and 9.91 percent are Hindu. We use this variable to code the Hindus, and the Bengalis are coded as the Muslims. As a practical matter, the coding of the Buddhists and Christians is irrelevant because they are a trivial percentage of the population. The replication materials describe for each survey the mapping from survey questions to Fearon categories.

[5]As an example, consider the Afrobarometer survey for Nigeria in 2003, for which it is possible to use a language

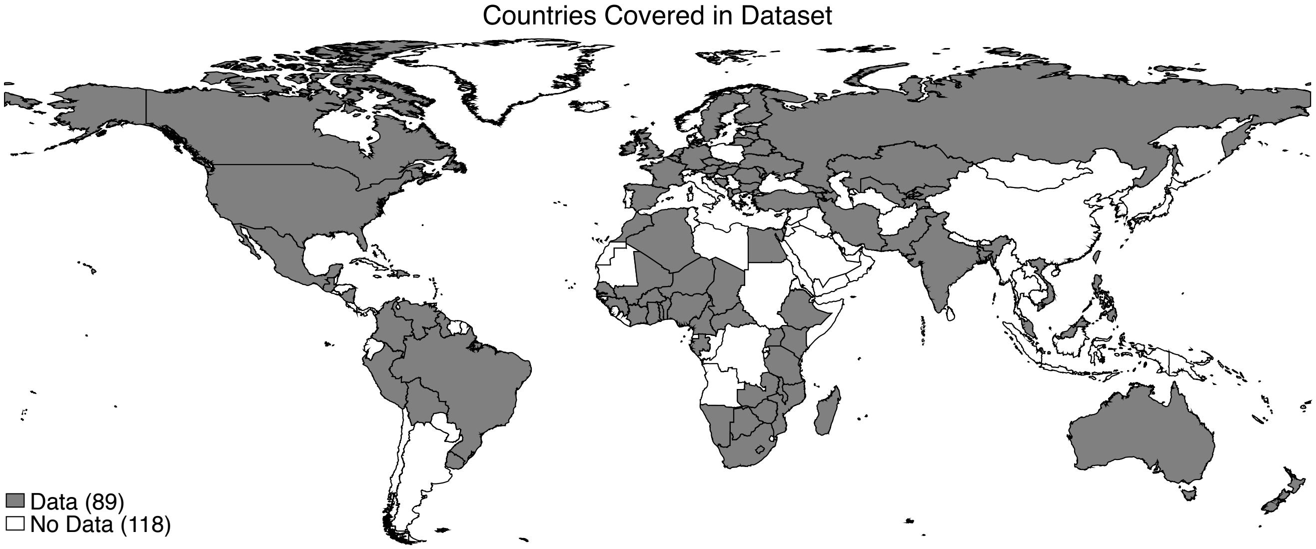

Figure 1: Countries included in data set

This approach yields 232 surveys from 89 countries depicted in the map in Figure 1. Surveys were conducted from 1992 to 2008.[6] The WVS provides the largest number of surveys (79), and the number of surveys in the remaining categories are 70 (DHS), 30 (HES), 29 (CSES) and 24 (AFRO). For 29 countries, we have only one survey, whereas in others we have multiple surveys, at most 7. Fifteen pairs of observations correspond to the same country/year. Therefore, the empirical analysis is based on 217 *distinct* country/year observations.

3.1.1 Group-level measures

The central argument we wish to test concerns whether groups with higher levels of inequality are more likely to engage in civil conflict. To this end, we use the surveys to measure the Gini coefficient of inequality for each group. For a group g it is given by

$$G_g = \frac{\sum_{k=1}^{N_g} \sum_{l=1}^{N_g} \mid y_{g_k} - y_{g_l} \mid}{2N_g^2 \bar{y_g}}, \tag{1}$$

where N_g is the size of group g, y_{g_j} is the income of individual $j = \{k, l\}$ of group g and $\bar{y_g}$ is the average income of group g. In addition, to test arguments about the impact of horizontal

variable to map to many of Fearon's groups. But one of his groups is "Middle Belt," and it is not possible to identify these individuals in the Afrobarometer survey. Since Fearon's data suggest they represent 18 percent of the population (which exceeds our threshold), we exclude this survey.

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

inequalities on conflict, we follow Cederman et al. (2011) and measure

$$HI_g = log(\bar{y_g}/\bar{y})^2, \tag{2}$$

where $\bar{y}$ is the mean income in society. HI_g measures the deviation of a group's average income from the country's average income, and thus takes high values for both high and low income groups.

3.1.2 Country-level measures

To explore whether countries with the highest within-group income disparities are more likely to experience civil conflict than countries with lower levels of such disparities, we estimate within-group inequality (or "WGI"), one of three components of the well-known Gini coefficient. 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 with a small proportion of resources in society will contribute less to WGI than will high inequality in a group with a large proportion of resources). Using discrete data, WGI can be written as

$$WGI = \sum_{g=1}^{m} G_g n_g \pi_g, \tag{3}$$

where m denotes the number of groups and π_g and n_g are the proportion of total income going to group g and its relative size, respectively.

The second component of the Gini is between-group inequality ("BGI"), a measure of the average difference in group mean incomes in a society. BGI calculates the society's Gini based under 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

$$BGI = \frac{1}{2\bar{y}}(\sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j \mid \bar{y}_i - \bar{y}_j \mid). \tag{4}$$

Overlap, the third component, is the residual that remains when BGI and WGI are subtracted from the Gini (G), and it is written as

$$OV = G - WGI - BGI. \tag{5}$$

When the groups' income support do not overlap, OV is zero, so scholars have interpreted this 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. Since the Gini coefficient does not decompose neatly into BGI and WGI components, scholars have at times turned to general entropy measures like the Theil index, which cleanly decomposes 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, or between men and women, across states).

We will therefore use BGI and WGI to test arguments about ethnic inequality and civil conflict at the national level. Although these two components do not capture all inequality in a society, our main focus is not on overall inequality, and BGI and WGI have straightforward and substantively appropriate definitions for the purposes here.

4 Estimates of ethnic inequality

To compute the measures defined above, we use the data on the economic well-being of group members from the surveys and data on group size from Fearon (2003). Since the surveys vary in their measures of economic well-being, we face the problem of comparability in inequality measures across surveys. This is a standard challenge faced by efforts to measure inequality across units that have heterogeneous measures of economic well-being. For instance, the observations in Deininger and Squire's classic (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), perhaps the most comprehensive data set of income inequality, presents identical shortcomings. Thus, if scholars wish to conduct broad, cross-national research on inequality using such measures, they must adopt methodologies to adjust the data to make them comparable. We consider two approaches.

4.1 The "intercept approach" to adjusting the survey measures of inequality

The first approach to adjusting the inequality measures shares the same spirit as the original Deininger and Squire (1996) exercise. The idea is to remove average differences due to different survey methodologies. To implement this approach, we regress the group-level inequality measures (G_g and HI_g) on survey, time and country dummies, with HES as the omitted category. We use the HES as reference since these surveys are probably the best-available estimates of income distribution in the world. The shift coefficients on the survey dummies are then used to adjust the inequality measures so as to remove average differences that could be traced to different survey types.

To adjust the country-level measures of inequality we proceed in a similar fashion. We regress the 3 components of the Gini (WGI, BGI and OV) on region, time and survey dummies. We then subtract the coefficients of the survey dummies from the Gini components in order to get rid of average differences due to survey methodology. The adjusted country-level Gini is obtained by summing the adjusted components.

Since inequality variables vary only slowly over time, in most of our empirical analysis we use time-invariant inequality measures. To compute these measures at the group-level, we take the average of the adjusted inequality measures from all the available surveys for a group and assign these average values to all years, beginning with the first year for which a survey exists for the group. Define $G_g^{ADJ_I}$ as this average group Gini using measures adjusted with the intercept approach. Data are missing in years prior to the first available survey year. For the country-level measures of the Gini, we adopt an identical approach, averaging all available observations for the same country and assigning them to all years starting with the first year for which a survey is available for that country. We label this country-level variable G^{ADJ_I}. A comprehensive list of all variables used in the analysis below is given in the Appendix.

4.2 The "ratio approach" to adjusting the components of the Gini

The second approach draws on external data on the Gini – the Standardized World Income Inequality Dataset (SWIID)– to adjust the group-level measures of the Gini as well as the three components of the Gini decomposition. The SWIID (Solt 2009) provides comparable Gini indices of gross and net income inequality for 173 countries from 1960 to the present and is one of the finest attempts to tackle the comparability challenge (see Solt 2009 for details on the methodology).

The basic idea of our approach is to use the SWIID data and a methodology similar to Solt (2009) to obtain (time-varying) adjustment factors for the overall country Gini from each country and year. We apply these country-level factors to the group-level measures of the Gini as well as to the three components of the (country-level) Gini decomposition. Central to our justification of this approach is our observation that although some of the surveys tend to produce measures that systematically underestimate the overall inequality in society (and, thus, the *level* needs to be adjusted), surveys provide much more reliable estimates of the *proportion* of inequality that is attributable to each of the Gini's three components. Section A.1 in the Appendix provides evidence for these claims.

Let $G^{SWIID}_{c,t}$ be the SWIID Gini for country c in year t and $G^s_{c,t}$ be the Gini from country c and year t using survey s. The ratio approach involves 4 steps:

Step 1: Whenever a survey Gini and the SWIID Gini are available for the same country and year, we compute their ratio, $R^s_{c,t} = \frac{G^s_{c,t}}{G^{SWIID}_{c,t}}$.

Step 2: For the 201 available ratios, we regress $R^s_{c,t}$ on country and year dummy variables. Specifically, we estimate:

$$R^s_{c,t} = \alpha_c + \delta_t + \epsilon^s_{c,t}. \tag{6}$$

Step 3. Following Solt (2009), for each survey we use the parameter estimates from eq. (6) to obtain the predicted values of the ratios, $\hat{R}^s_{c,t}$, for all surveys. For those surveys where ratios exist, the predicted ratios are of course very close to the actual ratios (r=.98), but the predicted ratios also can be derived from Eq. (6) for the 16 surveys where the SWIID Gini is missing. This is justified by the fact that the factors that affect these ratios tend to change only slowly over time within a given country and, hence, the missing ratios can be predicted based on available data on the same ratio in the same country in proximate years.

Step 4. To obtain the adjusted measures using the ratio approach, denoted by the superscript ADJ$_R$, we take the product of the original measures (e.g., $WGI^s_{c,t}$ for WGI in country c, year t using survey s) and the predicted ratios.

$$G^{ADJ_R}_{c,t,s} = \hat{R}^s_{c,t} G^s_{c,t} \tag{7}$$

$$WGI^{ADJ_R}_{c,t,s} = \hat{R}^s_{c,t} WGI^s_{c,t} \tag{8}$$

$$BGI^{ADJ_R}_{c,t,s} = \hat{R}^s_{c,t} * BGI^s_{c,t} \tag{9}$$

$$OV^{ADJ_R}_{c,t,s} = \hat{R}^s_{c,t} * OV^s_{c,t} \tag{10}$$

In this way, the weight of each of the components of the Gini is preserved but their level is adjustment to match the adjusted overall Gini. And we use the predicted ratios to obtain an adjusted group-Gini:

$$G^{ADJ_R}_{g,c,t,s} = \hat{R}^s_{c,t} * G_{g,t,s}. \tag{11}$$

Step 4 yields the measures we use in our empirical analysis using the "ratio" approach. As in the intercept approach, time-invariant measures are computed by averaging all observations available for one group/country and assigning the average values to all years, beginning with the first year for which data is available. Define $G^{ADJ_R}_g$ as the average group-level Gini adjusted using the ratio approach, define WGI^{ADJ_R} as the average country-level measure of WGI, adjusted with the ratio approach, with other components similarly defined.

Both the intercept and ratio approaches are well-established in the literature. The ratio approach has the advantage of utilizing a well-known time-varying external benchmark, the Gini coefficient, to adjust the Gini and its components. But it has the disadvantage of forcing us to assume that each component of the Gini must be adjusted by the same amount. The intercept approach avoids this assumption, allowing us to adjust each component separately based on benchmarking against the HES. But the intercept approach has the disadvantage that the benchmark HES observations, unlike the external measures of the Gini, are available for a relatively small set of countries. As a practical matter, the two approaches yield rather similar results. For example, the correlation of $G_g^{ADJ_I}$ and $G_g^{ADJ_R}$ is .75, although $G_g^{ADJ_I}$ has a somewhat higher mean (.45) than that of $G_g^{ADJ_R}$ (.38).[7] We are agnostic regarding which approach to use and instead wish to understand if the empirical results are robust to the alternative approaches. In addition, estimating models using only the WVS unadjusted allows us to apply the same measure of income to all countries (albeit a relatively small subset of them).

5 Strengths and weaknesses of the survey-based data and alternatives

There are a number of potential limitations associated with using surveys to measure ethnic inequality. One is that the approach can only be implemented in countries with useful surveys, and the set of such countries might be unrepresentative in important ways. In particular, one might worry that the countries where surveys exist might be correlated with ethnic conflict itself, or with variables related to ethnic conflict.

Table 1 examines this issue empirically.[8] The table compares the sample of countries obtained from our surveys to a broader set of countries from the SWIID data set. 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 136 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 SWIID countries

[7] If we consider the country-level data, the two approaches also produce very similar results: the correlations of the two WGI variables is .89, of the two BGI variables is .93, and of the two Overlap variables is .90. More information about these components is provided below.

[8] 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.

Table 1: Sample representativeness

	SWIID sample	Survey sample
Number of countries	136	88
Percentage of countries in:		
Central Europe	19.8	26.4
Latin America	16.2	12.5
Middle East	5.9	3.4
Africa	28.7	30.1
Neo-Europe	16.2	18.2
East Asia	8.8	5.7
South Asia	4.4	3.4
Average Real GDP/capita	$9,836	$10,288
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

Notes. This table compares the sample of countries included in the dataset presented in this paper (88 countries) and the SWIID (137).

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: GDP/capita, ethnic fractionalization (F), ethnic polarization (P), level of democracy (xPolity), level of inequality, and the incidence of civil conflict.[9] For each of these variables, the means for the set of countries in SWIID are quite similar to the means for the set of survey countries. Thus, although there are limits on the number of countries we can analyze using surveys, the sample of countries obtained using surveys seems reasonably unbiased with respect to the variables of central interest in the analysis here.

Another concern may be that the surveys themselves do not accurately represent the groups in society. As noted above, a strategy we employ for addressing this possibility is to use the group size data from Fearon (2003) and and to utilize only surveys that adequately represent the Fearon

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

groups (by discarding surveys for which there exist 10 percent of the population (per Fearon) that cannot be identified using the identity questions in the survey). But it is also important to note that the correlation of ELF from the surveys and ELF from Fearon's data is an impressive .93. Moreover, when we calculate the the components of the Gini decomposition using the surveys' measure of group size, we obtain measures of the Gini components that are extremely similar to those based on the Fearon group sizes: the correlations of WGI using the surveys' measure of group size and the measures of WGI using Fearon group size is .95. For BGI or for Overlap the correlations are both .94.

Although this is reassuring evidence that neither the sample of countries nor the sample of groups from the surveys is particularly biased, the accuracy with which the surveys measure individual "income" of course remains a concern. In particular, we face the challenge described above of incorporating measures based on different metrics. We have described two strategies for addressing this issue, and our analysis below will incorporate the resulting measures in a variety of models to asses robustness. But there may still be concern that survey respondents may not tell the truth about their income on surveys. While this is always a potential concern with surveys, we can take some reassurance from the fact the proportions of the Gini coefficient are very similar across the survey types, and that one of the survey types (HES) uses very careful household expenditure surveys which provide the best information available about economic well-being.

To put these limitations with survey data in perspective, it is worth discussing the main alternative in the literature, which combines geo-referenced data on the geographic location of ethnic groups with geo-referenced estimates of economic development. Examples include Cederman et al. (2011) and Alesina et al. (2013), who focus primarily on inequality between groups, and Kuhn and Weidmann (2013), who examine inequality *within* groups. The data on the geographic location of groups has been taken from a variety of sources. Alesina et al., for example, utilize the GREG data set (the Geo-Referencing of Ethnic Groups data set, published by Weidmann, Rob and Cedarman (2010), and based on the Soviet Atlas Narodov Mira) and the *Ethnologue*, which provides information on the spatial location of linguistic groups in much of the world. Cederman et al (2011) utilize the GeoEPR data set, which is described in Wucherpfennig et al. (2011) and which utilizes an expert survey to determine the identity and location of politically relevant ethnic groups. The spatial data on group locations can has been linked to spatial data on economic output, for example using Nordhauss (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 Kuhn and Weidmann 2013).

The geo-coded inequality data have an advantage vis-a-vis surveys when it comes to country coverage. Depending on the definition of groups used (e.g., the GeoEPR data set covers more countries than the *Ethnologue*), the data sets can cover the vast majority of countries in the world. Like the surveys, however, the spatial approach entails tradeoffs with respect to measuring the representativeness of groups in the population and the measurement of economic well-being. Potential limitations with respect to the representativeness of groups stem principally from two issues. First, these approaches rely on expert estimates of the spatial location of groups, and thus they risk measurement error because the experts themselves often do not have data on which to base their estimates of the group locations. Indeed, the best data on which experts could draw would be some sort of careful survey or census, so any biases with respect to the coverage of groups in the survey data are going to be also present in the spatial data. Indeed, the biases might be worse in the spatial data because experts are asked to state precisely where the groups reside.

Second, the spatial approach is limited in the way it treats urban dwellers. In some countries, groups might be relatively geographically segregated in the country side. But in urban areas, this is unlikely to be true, and it seems very challenging for country experts to accurately determine which ethnic groups are located in specific urban neighborhoods. Thus, providing representative estimates of the spatial location of groups can be particularly challenging in urban areas, which are often excluded from geo-coded analyses.

When we consider the measurement of economic well-being, a clear strength of the spatial approach – and particularly the night-light approach – is that it applies a consistent criterion across countries, potentially reducing problems with cross-national comparisons of economic measures. This is particularly important in countries that have weak infrastructure for collecting economic data, or in countries where the government may have incentives to misrepresent data about the economy. Although there are a number of issues associated with using night-light data to measure economic activity, this approach clearly provides valuable information about economic well being, at least in relatively large geographic areas.[10] But to our knowledge there has as yet been little

[10]See Chen and Nordhaus (2011), Bhandari and Roychowdhury (2011), Chosh et al (2013) and Mellander et al

effort to understand the potential strengths and weaknesses of using nightlight data to measure within- and between-group economic differences, and we feel there are reasons for caution in this regard.

One limitation of the spatial approach is the need 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 area have the same income. Neither assumption is attractive. 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 essentially assume what one is trying to measure.

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. 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. Additionally, the urbanization of a country may be correlated with other factors that are related to civil war, raising concerns that the biases may be correlated with conflict. It is also the case that using spatial data in this way to measure WGI should yield results that are 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. So like surveys, the spatial approach has strengths and weaknesses.

6 Group-level analysis of civil conflict

The survey-based measures make it possible to examine empirically whether group-based inequality is related to the propensity of groups to engage in conflict. Empirical measures of civil war distin-

(2013).

guish between conflict onset (the year a civil conflict begins) and conflict incidence (the presence and intensity of conflict in a given year). Esteban and Ray argue that 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 provides 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 fuel violence. We will therefore focus primarily on measures of conflict incidence (although we also present results for onset).

The data on conflict are taken from the Ethnic Power Relations data set (EPR, Cederman et al. 2009), which describes which groups are engaged in conflict in a given country-year.[11] 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). We begin by focusing on CONFLICT25$_G$, a binary measure taking a value of 1 for those years where an ethnic group is involved in armed conflict against the state resulting in more than 25 battle-related deaths. 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. Our regressions are based on (at most) 88 countries and 449 groups over the period 1992-2009. For each group and for each country, the first observation that enters the regressions is the first one for which survey data is available.

Table 2 presents six models using the two approaches to adjust the survey-based measures: models 1-3 use the ratio approach to adjust the measure of the group Gini and models 4-6 use the intercept approach. The dependent variable in each model is CONFLICT25$_G$. All models include country and year indicator variables as well as a lagged dependent variable, and the models are estimated in a logit specification, with standard errors clustered at the ethnic group level. Model 1 includes three group-level variables: G$_g$, HI$_g$ and POP$_g$, the population of the group.[12] The group Gini variable has a positive and precisely estimated coefficient (p=.023), but the coefficients of the other group-level variables are estimated with substantial error. Model 2 adds two time-varying country level controls: GDP is lagged value of the log of GDP per capita, and XPOLITY is a democracy

[11]The data are accessed through the ETH Zurich's GROWup data portal (`http://growup.ethz.ch`). 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.

[12]A detailed list of all variables is provided in the Appendix.

score based on Polity IV, lagged one year. It combines 3 out of the 5 components of Polity IV but leaves out the two components (PARCOMP and PARREG) that are related to political violence (see Vreeland 2008). Neither of the coefficients for these two variables is measured precisely and the results for the group level variables are virtually the same as those in model 1. Finally, some scholars have argued that poverty can lead to violence (Fearon and Laitin 2003, Miguel et al 1994). To control for this possibility, model 3 includes GDP_g, which is the (lagged) GDP per capita of the group. Although the coefficient is negative (implying poorer groups are more likely to engage in conflict), it is measured with substantial error. The results for the other variables are not affected by the inclusion of this variable.

The estimated effect of inequality within the group is very substantial. Using the results from column 3, moving from the median G_g (Arabs in Turkey, .36) to the group in the 90th percentile (Nama/Damara in Namibia, .54) while holding other variables at their means, the predicted probability of experiencing conflict (i.e, the probability of observing strictly positive values of CONFLICT25_G) rises from .057 to .27, which implies an increase of almost 500%. This effect is very robust across the different specifications considered in Table 2 since the coefficients associated with G_g are very stable across all columns.

Models 4-6 have the same structure as columns 1–3 but use inequality measures adjusted according to the intercept approach. As in the first three models, group inequality has a positive and precisely estimated coefficient, and HI does not.

Table 3 re-estimates the same models as in Table 2, but with two differences. First, in models 1-3, we estimate the model using only data from the WVS. This limits the number of countries to 53, compared with 88 when the full data set is used, but doing so makes it possible to apply the same measure of income in each country, and thus makes it possible to utilize the data with no adjustment. The results show a strong robust relationship between a group's Gini and conflict by the group. In addition, there is a strong positive relationship between HI and group conflict, with group's that are more distant from the mean income more likely to engage in conflict. Group GDP and group size do not have a strong association with conflict. Models 4-6 utilize all the data, but estimate the models without adjusting any of the survey-based measures. The only group-level variable that is statistically significant is the group Gini, which has a positive and precisely estimated coefficient.

Table 2: Ethnic inequality and group-level conflict: Baseline (CONFLICT25$_G$)

	[1]	[2]	[3]	[4]	[5]	[6]
$G_g^{ADJ_R}$	10.080**	10.077**	10.076**			
	(0.024)	(0.026)	(0.024)			
$G_g^{ADJ_I}$				10.754*	10.749*	10.753*
				(0.051)	(0.054)	(0.051)
HI_g^{UNAD}	1.547	1.559	1.550			
	(0.352)	(0.350)	(0.352)			
$HI_g^{ADJ_I}$				1.404	1.416	1.408
				(0.391)	(0.388)	(0.391)
POP$_g$	-3.448*	-3.409	-3.657**	-3.379*	-3.376	-3.635**
	(0.057)	(0.136)	(0.049)	(0.056)	(0.134)	(0.047)
GDP		0.111			0.088	
		(0.899)			(0.917)	
XPOLITY		-0.023			-0.024	
		(0.625)			(0.584)	
GDP$_g$			-0.186			-0.221
			(0.585)			(0.513)
CONFLICT25$_G$(lag)	0.458**	0.459**	0.419*	0.436**	0.437**	0.388*
	(0.031)	(0.027)	(0.054)	(0.035)	(0.031)	(0.067)
CONST	19.032	17.903	22.290	17.033	16.405	20.955
	(0.167)	(0.443)	(0.140)	(0.206)	(0.474)	(0.159)
(Pseudo) R^2	0.212	0.212	0.212	0.202	0.203	0.203
Obs.	1627	1579	1627	1627	1579	1627

*Note.*The dependent variable is CONFLICT25$_G$. Columns 1-3 adjust the inequality data using the ratio approach and columns 4-6 using the intercept approach. All models include country and year fixed effects, and all models are estimated with logit. Robust standard errors clustered at the group level have been computed with p-values in parentheses. The period considered is 1992-2009 and the number of countries is 88.
* p<.10, ** p<.05, *** p<.01.

Using Conflict25$_G$, we have found a strong positive relationship between a group's Gini and its participation in conflict. The results are robust to different model specifications, to different ways of adjusting the inequality data (or not), and to using only countries included in the WVS. Next we consider whether the results are robust to different measures of conflict. First we estimate ordered logit models using CONFLICT-INT$_G$ as the dependent variable. This measure of conflict can take one of 3 values: 0 if a group is not engaged in conflict, 1 if a group is engaged in a conflict that results in more than 25 battle deaths in the country-year, and 2 if the group is engaged in a conflict that results in more than 1,000 deaths in the country-year. Table 4 presents the results, using the same six models as in Table 2. As in Table 2, the coefficients for group inequality are positive and precisely estimated across all six models. No other variable has a coefficient that is consistently estimated with precision.

Finally, although Esteban and Ray model the ability of groups to sustain rather than initiate

Table 3: Ethnic inequality and group-level conflict: unadjusted data (CONFLICT25$_G$)

	[1]	[2]	[3]	[4]	[5]	[6]
G_g^{WVS}	33.550**	33.481**	34.004**			
	(0.021)	(0.021)	(0.018)			
G_g^{UNAD}				12.762**	12.792**	12.762**
				(0.015)	(0.017)	(0.015)
HI_g^{WVS}	17.740***	17.797***	17.756***			
	(0.005)	(0.005)	(0.005)			
HI_g^{UNAD}				1.383	1.394	1.385
				(0.369)	(0.366)	(0.369)
POP$_g$	-4.990	-3.972	-3.634	-3.432*	-3.391	-3.679**
	(0.468)	(0.512)	(0.562)	(0.059)	(0.139)	(0.050)
GDP		3.123			0.111	
		(0.102)			(0.898)	
XPOLITY		-0.046			-0.022	
		(0.765)			(0.632)	
GDP$_g$			0.717			-0.217
			(0.403)			(0.524)
CONFLICT25$_G$	0.132	0.136	0.267	0.446**	0.447**	0.399*
	(0.704)	(0.711)	(0.390)	(0.033)	(0.029)	(0.062)
CONST	21.592	-13.980	4.569	18.744	17.593	22.565
	(0.671)	(0.760)	(0.918)	(0.176)	(0.451)	(0.137)
(Pseudo) R^2	0.430	0.434	0.432	0.214	0.215	0.214
Obs.	504	504	504	1627	1579	1627

*Note.*The dependent variable is CONFLICT25$_G$. Columns 1-3 use only WVS surveys with no adjustment and columns 4-6 use all data with no adjustment. All models include country and year fixed effects, and are estimated with logit. Robust standard errors clustered at the group level have been computed with p-values in parentheses. The period considered is 1992-2009 and the number of countries is 88. * p<.10, ** p<.05, *** p<.01.

conflict, we might expect that if a group has the labor and capital to sustain conflict, it might also be more inclined to get involved in conflict in the first place. Table 5 therefore estimates logit models where ONSET$_G$, a measure of the onset of civil conflict, is the dependent variable. Using this variable obviously results in far fewer observations, so the results should be taken with some caution. But we again find that group inequality is associated with conflict onset across all six models. In addition, GDP$_g$ is negative and significant in both models, suggesting that poorer groups are more likely than richer ones to initiate conflict. And POP$_g$ has a precisely estimated coefficient in half the models. The coefficient for HI is never precisely estimated.

Using models that include country and year dummy variables, we have found a strong positive relationship between group-based inequality and the participation of groups in civil conflict, one that is robust to different model specifications, to two different approaches to adjusting the measures of group inequality, to using only WVS countries or using unadjusted data, and to three different measures of conflict.

Table 4: Ethnic inequality and group-level conflict: CONFLICT-INT$_G$

	[1]	[2]	[3]	[4]	[5]	[6]
$G_g^{ADJ_R}$	9.751**	9.746**	9.739**			
	(0.023)	(0.025)	(0.023)			
$G_g^{ADJ_I}$				10.609*	10.609*	10.598**
				(0.050)	(0.053)	(0.050)
HI_g^{UNAD}	1.533	1.543	1.534			
	(0.337)	(0.335)	(0.337)			
$HI_g^{ADJ_I}$				1.409	1.419	1.411
				(0.374)	(0.372)	(0.374)
POP$_g$	-2.788	-2.520	-3.041	-2.706	-2.456	-2.996
	(0.220)	(0.397)	(0.210)	(0.224)	(0.401)	(0.210)
GDP		0.270			0.257	
		(0.781)			(0.783)	
XPOLITY		-0.016			-0.018	
		(0.725)			(0.695)	
GDP$_g$			-0.234			-0.261
			(0.499)			(0.447)
CONFLICT-INT$_G$ (lag)	0.234	0.238	0.192	0.226	0.230	0.178
	(0.244)	(0.231)	(0.341)	(0.246)	(0.232)	(0.362)
(Pseudo) R^2	0.421	0.424	0.421	0.416	0.419	0.417
Obs	6149	6065	6149	6149	6065	6149

Note. The dependent variable is CONFLICT-INT$_G$. Columns 1-3 adjust the inequality data using the ratio approach and columns 4-6 using the intercept approach. All models include country and year fixed effects, and all models are estimated with ordered logit. Robust standard errors clustered at the group level have been computed with p-values in parentheses. The period considered is 1992-2009 and the number of countries is 88.
* p<.10, ** p<.05, *** p<.01.

7 Country-level analysis

The existing arguments about inequality within-groups and inequality between groups are focused on group-level dynamics. For a number of reasons, however, it is also useful to consider national-level analysis. First, substantively it is important to understand which types of countries are likely to experience civil war, and thus whether the arguments about group-based inequality can contribute in this regard. It is not unreasonable to expect that these existing group-based theories will have purchase in country-level analysis: if group-based inequality leads a group to engage in civil conflict, for example, we might reasonable expect countries with high levels of group-based inequality to be most likely to experience civil war. Second, a national-level analysis makes possible a more nuanced understanding of the relationship between overall inequality and conflict. As noted above, scholars have concluded that overall inequality and ethnic conflict are unrelated (i.e., Fearon and Laitin 2003 and Collier and Hoeffler 2004). The survey data on the Gini decomposition

Table 5: Ethnic inequality and group-level conflict: ONSET_G

	[1]	[2]	[3]	[4]	[5]	[6]
$G_g^{ADJ_R}$	7.730*	7.539*	7.760*			
	(0.085)	(0.070)	(0.082)			
$G_g^{ADJ_I}$				10.619*	10.320**	10.683**
				(0.050)	(0.041)	(0.047)
H_g^{UNAD}	0.401	1.269	0.476			
	(0.817)	(0.459)	(0.791)			
$H_g^{ADJ_I}$				0.138	1.398	0.256
				(0.953)	(0.532)	(0.913)
POP_p	-39.901***	5.212	-42.302***	-39.659***	6.678	-42.144***
	(0.000)	(0.802)	(0.001)	(0.000)	(0.754)	(0.001)
GDP		49.159***			50.904***	
		(0.001)			(0.002)	
XPOLITY		-1.645***			-1.679***	
		(0.006)			(0.006)	
GDP_g			-1.345*			-1.407*
			(0.069)			(0.071)
CONST	479.849***	-443.262	519.008***	475.309***	-475.924	515.945***
	(0.000)	(0.178)	(0.001)	(0.000)	(0.170)	(0.001)
(Pseudo) R^2	0.209	0.305	0.220	0.219	0.317	0.230
Obs.	118	118	118	118	118	118

Note. The dependent variable is ONSET_G. Columns 1-3 adjust the inequality data using the ratio approach and columns 4-6 using the intercept approach. All models include country and year fixed effects, and all models are estimated with logit. Robust standard errors clustered at the group level have been computed with p-values in parentheses. The period considered is 1992-2009 and the number of countries is 88.
* p<.10, ** p<.05, *** p<.01.

can be used to explore why this might be true.

The country-level conflict data is taken from the UCDP/PRIO data set.[13] Our baseline variable is PRIO25_C, an indicator variable that takes the value 1 in a country-year if there is a conflict with 25 or more battle deaths in that year.[14] For robustness, we also consider different conflict definitions. PRIOINT_C takes the value 0 if there is peace in a given year, the value 1 if there are events satisfying PRIO25_C 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. PRIOCW_C is 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. ONSET_C is a dummy that switches on in a particular year if the incidence requirement is met (at the level of PRIO25_C), but not in 2 or more previous years.

[13]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.

[14]See Appendix B for an exact definition.

We use a standard set of controls: population (POP), GDP per capita (GDP), a dummy variable for oil and/or diamond producing countries (OIL/DIAM), the percentage of mountainous terrain (MOUNT), a dummy variable for noncontiguity of country territory (NCONT), fractionalisation (F), polarisation (P) and a variable measuring the extent of democracy (XPOLITY). The justification for each of these controls can be found elsewhere (see Fearon and Laitin 2003, Vreeland 2008 and Esteban et al 2012). See Appendix B for exact definitions and sources. In addition, all our regressions contain year dummies and region (or country) fixed effects.

Table 6 presents the results using PRIO25$_C$ as the dependent variable. The inequality measures employed in these regressions have been computed by averaging all the inequality observations available for each country and assigning this value to the whole period, starting by the first year for which survey data are available. All models contain the control variables discussed above, including the year indicator variables and the regional dummies, but our discussion will focus on the inequality variables. In columns 1–3 the inequality variables have been adjusted using the 'ratio' approach, as described in Section 4.2. Column 1 presents results relating the overall country GINI, G, and conflict. The coefficient of G is positive but not significant. We have also explored the connection between conflict and overall country Ginis using alternative datasets (SWIID and Povcalnet –World Bank–), time period (1960 onwards) and estimation approach (including country fixed effects). In line with previous literature, the overall conclusion is that the lack of connection between country Ginis and conflict is a very robust result.

Column 2 introduces WGI. Since WGI is also a component of the Gini, we compute a new variable by subtracting WGI from the Gini and include this new variable, G-WGI, on the right-hand side (instead of G, itself). The coefficient on G-WGI therefore estimates the effect of all inequality unrelated to WGI on conflict, and the coefficient on WGI estimates the effect only of inequality within groups (and not of WGI through G). We find that the effect of WGI is positive and significant, indicating that countries with more within-group dispersion of incomes are more likely to experience conflict. The coefficient of G-WGI is negative but not significant. Column 3 includes all three components of the Gini separately. The coefficient of WGI remains very similar to that in model 2. Overlap has a negative coefficient whereas BGI has a positive one, but neither coefficient is precisely estimated. Models 4–6 are similar to those in columns 1–3 but with inequality variables adjusted using the 'intercept' approach. The qualitative conclusions are very similar: the only

Table 6: Ethnic inequality and Country-level conflict: Baseline (PRIO25$_C$)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
G^k	2.162			7.771				
	(0.463)			(0.100)				
WGI^k		13.437**	13.144**		21.853**	19.502*	130.101***	16.474**
		(0.031)	(0.042)		(0.028)	(0.073)	(0.001)	(0.049)
BGI^k			-1.552			8.590	-60.490**	7.208
			(0.763)			(0.172)	(0.019)	(0.327)
OV^k			-11.363			-1.119	-52.253	-3.049
			(0.119)			(0.871)	(0.167)	(0.555)
G-WGI^k		-4.687			3.113			
		(0.326)			(0.549)			
F	2.652**	9.532***	10.672***	2.426*	8.403**	7.782**	65.454***	7.683**
	(0.028)	(0.006)	(0.005)	(0.057)	(0.013)	(0.030)	(0.000)	(0.019)
P	1.543	1.588	2.249**	2.390**	2.119*	2.487**	-0.794	2.555**
	(0.117)	(0.110)	(0.021)	(0.038)	(0.078)	(0.030)	(0.906)	(0.037)
NCONT	1.117*	1.133*	1.898**	1.581**	1.699**	2.066***	11.408***	2.438***
	(0.097)	(0.093)	(0.011)	(0.025)	(0.014)	(0.002)	(0.000)	(0.001)
MOUNT	0.011	0.011	0.010	0.012	0.015	0.015	0.095	0.015
	(0.240)	(0.209)	(0.282)	(0.200)	(0.113)	(0.131)	(0.129)	(0.126)
GDP	-0.265	-0.204	-0.345	-0.223	-0.291	-0.457	1.791	-0.290
	(0.292)	(0.435)	(0.211)	(0.376)	(0.252)	(0.101)	(0.104)	(0.368)
POP	0.400***	0.359***	0.307**	0.334**	0.327**	0.304**	0.880***	0.320**
	(0.002)	(0.009)	(0.030)	(0.018)	(0.018)	(0.032)	(0.009)	(0.032)
XPOL	0.035	0.026	0.038	0.035	0.019	0.022	0.190	0.041
	(0.407)	(0.555)	(0.407)	(0.444)	(0.672)	(0.648)	(0.113)	(0.396)
OIL/DIAM	-0.270	-0.262	-0.206	-0.316	-0.231	-0.148	-2.623*	-0.292
	(0.432)	(0.479)	(0.599)	(0.421)	(0.537)	(0.694)	(0.099)	(0.469)
LAG	4.647***	4.556***	4.443***	4.636***	4.580***	4.504***	3.416***	4.402***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
CONST	-8.980***	-13.061***	-11.438***	-11.573***	-15.225***	-13.090***	-71.840***	-12.612***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.000)	(0.002)
(Pseudo) R^2	0.625	0.629	0.632	0.628	0.631	0.633	0.843	0.636
Obs.	1044	1044	1044	1044	1044	1044	586	1044

Notes. Dependent variable is PRIO25$_C$. A logit model has been estimated in all cases. All models include year indicator variables and regional dummies. Robust standard errors clustered at the country level have been computed and p-values are in parentheses. The inequality variables in columns 1–3 (4–6) have been adjusted according to the ratio (k=ADJ$_R$) (intercept, k=ADJ$_I$) approach. Column 7 uses data from the WVS exclusively to compute the Gini decomposition (k=WVS) . Column 8 uses unadjusted inequality measures (k=UNAD). The period considered is 1992-2009. The number of countries is 88, except for column 7 which is 49. The inequality measures have been computed by averaging all the observations available for each country and assigning this value to the whole period. For each country, the first observation that enters the regression is the first one for which survey data is available for that country.
* $p<.10$, ** $p<.05$, *** $p<.01$.

component of the Gini that is significantly associated with conflict is WGI. The effect of within group inequality is not only precisely estimated, it is substantively large. Using the results from column 3, moving from the median country's WGI (Cyprus, .178) to the country in the 90th percentile (Egypt, .284) while 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 .049 to .172, which implies an increase of more than 300%.[15]

Model 7 is the same as model 3 but is based on data from the WVS exclusively. The estimated coefficient of WGI is still positive and significant. The coefficients for BGI and OV have the wrong sign, but only the former is significant. Column 8 employs the original inequality data with no adjustment. The results are in line with those obtained in columns 3 and 6: WGI is the only component of the Gini decomposition positive and significantly associated with civil conflict. In Appendix A, we show that these results are robust to the use of different definitions of civil conflict and to the inclusion of country fixed effects.

These country level results, when linked to information about the components of the Gini coefficient and how they are related to the Gini, help us to understand why general inequality is not related to civil conflict. To this end, consider Table 7, which provides basic information about the Gini decomposition using the two methods for adjusting the data. WGI is, on average, the largest component of the Gini, with the mean of WGI being about .18 from both the ratio and the intercept approach. The smallest component is between-group inequality, with BGI^{ADJ_R} an average of one-third that of WGI^{ADJ_R} and BGI^{ADJ_I} roughly one-half of WGI^{ADJ_I}. Overlap is the second largest component of the Gini, and is only slightly smaller on average than WGI using both approaches.

If WGI is the largest component of the Gini and is also related to civil conflict, why is it the case that the Gini itself is unrelated to civil conflict? Although WGI is the largest component of the Gini, the data reveal that the variability of WGI, as captured by the coefficient of variation in the third column of Table 7, is considerably smaller than that of BGI and OV. And variation in the Gini

[15]A very similar interpretation is obtained if the results from column 6 are used instead: in this case moving from the median's country WGI to the country in the 90'th percentile while holding other variables at their means increases the probability of conflict from .052 to .209. 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.

Table 7: The Gini decomposition across countries

Variable	Mean	Coeff.Var	Corr. with Gini
WGI^{ADJ_R}	0.179	0.447	-.02
BGI^{ADJ_R}	0.060	0.912	.70
OV^{ADJ_R}	0.146	0.585	.63
WGI^{ADJ_I}	0.181	0.381	-.34
BGI^{ADJ_I}	0.094	0.455	.72
OV^{ADJ_I}	0.146	0.491	.65
G^{ADJ_R}	0.385	0.24	–
G^{ADJ_R}	0.421	0.126	–
G (SWIID)	0.087	0.201	–

coefficient is strongly related to variation in BGI and Overlap, but not to variation in WGI. Using the ratio approach, for example, the correlation between G and WGI is non-existent (r=-0.02), but there is a strong correlation between Gini and the other two components (0.70 with BGI and 0.63 with Overlap). Similar results exist for the intercept approach.

This finding about the elements of the Gini decomposition help explain why overall inequality is unrelated to civil conflict. Within-group inequality is the only element of the Gini that is related to civil conflict, but this element is uncorrelated with the Gini. Put differently, country-level inequality measures do not reflect the cross-country variations in inequality within the groups that are are associated with conflict. More generally, Table 7 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.

8 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 while the correlation with inequality within groups is much smaller. Thus, much of the variation in the Gini coefficient is driven by the between group inequality component. Since the latter term has no relationship with civil conflict (when we control for inequality within groups), it should be expected that overall inequality has no association with conflict. More generally, the analysis underlines the difficulties in cross-national research that are 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.

9 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.

Anderson, Benedict . 1992. "Long-Distance Nationalism: World Capitalism and the Rise of Identity Politics." Wertheim Lecture, Center for Asian Studies, Amsterdam.

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.

Bhandari, Laveesh and Koel Roychowdhury. 2011. "Night Lights and Economic Activity in India: A study using DMSP-OLS night time images." *Proceedings of the Asia-Pacific Advanced Network* 32: 218-236.

Brass, Paul R. 1997. *Theft of an Idol: Text and Context in the Representation of Collective Violence.* Princeton, N.J.: Princeton University Press.

Brubaker, Rogers and David D. Laitin. 1998. "Ethnic and Nationalist Violence." *Annual Review of Sociology,* 24, 42–452.

Carment, David. 2007. "Exploiting Ethnicity" *Harvard International Review*, 28.

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.

Chen, Xi and William D. Nordhaus. 2011. "Using luminosity data as a proxy for economic statistics." *Proceedings of the National Academy of Sciences* 108(21): 8589-8594.

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).

Ghosh, Tilottama; Anderson, Sharolyn J.; Elvidge, Christopher D.; Sutton, Paul C. 2013. "Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being." Sustainability 5, no. 12: 4988-5019.

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.

Humphreys, Macarta and Jeremy M. Weinstein (2008). “Who Fights? The Determinants of Participation in Civil War.?? *American Journal of Political Science*, 52, 436?455.

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.

Mellander, Charlotta, Kevin Stolarick, Zara Matheson and José Lobo. 2013. “Night-Time Light Data: A Good Proxy Measure for Economic Activity?” Working paper, Martin Prosperity Research.

Melvern, Linda. 2000. *A people betrayed: the role of the West in Rwanda’s genocide.* London; New York, N.Y.: Zed Books.

Miguel, Edward, Shanker Satyanath and Ernest Sergenti. 2004. *Economic Shocks and Civil Conflict: An Instrumental Variables Approach. Journal of Political Economy*, 112(4): 725–753.

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: Additional Analysis

A.1 Survey heterogeneity and inequality measures

As noted in the main text, an important issue of concern is that surveys use different methods and definitions of "income." This section explores the impact of these alternative definitions on the obtained inequality measures. We focus both on the *level* of the overall country Gini as well as on *proportion* of inequality that is attributable to each of the Gini's three components.

Some surveys measure "income" using some form of income or expenditure (CSES, WVS and HES) while others use a list of asset indicators (DHS and AFRO). We can explore the extent to which this difference influences the inequality measures derived from either type of survey. Table A.1 presents a number of regressions comparing inequality measures from those survey types. "Asset" is an indicator variable that takes the value 1 if the survey is based on a list of asset indicators and 0 if the survey is based on a measure of income. In column 1 the dependent variable is the original overall Gini in society and the right-hand side variables include year and regional dummy variables, as well as a dummy variable corresponding to the surveys based on assets. The coefficient for the "Asset" indicator is very small and not significantly different from zero. Models 2 and 3 present similar regressions using the proportions of WGI and BGI in overall inequality (i.e., WGI/G and BGI/G) as dependent variables, respectively. In all cases, the coefficients for "Asset" are very small in absolute value and estimated with a very large error. Thus, on average there are no systematic differences in inequality measures across surveys that use assets and surveys that use income to measure economic well-being.

Next we explore whether there exist systematic differences across the various surveys, taking advantage of the fact that one of our survey types, the HES, provides very precise income or expenditure data.[16] We run regressions that are similar to those above, introducing dummy variables for all the surveys except HES. The coefficients on the survey variables therefore describe how inequality measures from the surveys other than the HES differ from the HES. In model 4, the dependent variable is the overall Gini in society and the right-hand side variables include year and regional dummy variables in addition to the survey dummy variables. All the coefficients of the survey variables are negative and highly significant, suggesting that the non-HES surveys tend to

[16] The correlation of the survey Ginis from HES with the SWIID Ginis is .92.

underestimate overall inequality.

Although the non-HES surveys underestimate overall inequality, do they also have biases in the proportion of overall inequality due to the various components? Column 5 presents the same analysis as column 4 but the dependent variable is the proportion of inequality due to WGI. The coefficients on all survey indicators are small and measured with considerable error. And Wald tests strongly reject the possibility that any of the survey coefficients differ from each other. Thus, there is no evidence that the proportion of inequality due to WGI varies systematically with survey type. Column 6 use the proportion of inequality due to BGI as the dependent variable. In this case, there are a number of coefficients that are statistically significant and negative. But column 7 estimates the same model without South Africa and Peru, two enormous outliers in the measurement of BGI, and only the the coefficient on WVS is measured with some precision (p-value 0.094). Thus, although there are some concerns about WVS underestimating the true proportion of BGI, in general, the surveys are producing very similar estimates of the proportion of inequality that is due to between-group economic differences.

Finally, the correlations of the proportions of the Gini components using different surveys for the same country are very high. For example there are ten countries for which there are three pairs of survey types: WVS-CSES, DHS-WVS, and HES-WVS. When considering the proportion of the Gini due to BGI, the correlations are .75 for WVS-CESES; .78 for DHS-WVS and .90 for DHS-WVS. For the proportion of inequality attributable to WGI, the correlations are for .99 WVS-CESES; .89 for DHS-WVS and .99 for DHS-WVS.

The analysis therefore provides evidence that the surveys employed in the creation of the inequality dataset tend to underestimate the level of inequality and, thus, some correction needs to be introduced to make them comparable in this regard. However, the proportions attributable to each of the components of the Gini coefficient do not seem to present such biases. These findings justify employing the ratio approach.

A.2 Country-level robustness tests

This section describes some additional tests of the robustness of the results presented in Section 7. For the sake of brevity we focus on inequality variables adjusted using the ratio approach. The

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Asset	-0.011	0.024	0.014				
	(0.517)	(0.492)	(0.527)				
WVS				-0.153***	0.004	-0.054**	-0.042*
				(0.000)	(0.924)	(0.036)	(0.094)
DHS				-0.120***	0.029	-0.023	0.012
				(0.000)	(0.535)	(0.608)	(0.670)
CSES				-0.144***	0.040	-0.063**	-0.036
				(0.000)	(0.445)	(0.046)	(0.128)
AFRO				-0.194***	0.044	-0.081	-0.047
				(0.000)	(0.504)	(0.115)	(0.227)
Constant	0.280***	0.673***	-0.017	0.388***	0.672***	0.011	0.129
	(0.002)	(0.000)	(0.886)	(0.000)	(0.001)	(0.825)	(0.101)
Dep. Var.	G	WGI/G	BGI/G	G	WGI/G	BGI/G	BGI/G
R^2	0.189	0.559	0.334	0.520	0.561	0.364	0.403
Obs.	217	217	217	217	217	217	208
C	89	89	89	89	89	89	87

Table A.1: THE RELATION BETWEEN INEQUALITY AND THE SURVEYS

Notes. The dependent variables in columns 1 and 4 is overall Gini, in columns 2 and 5 is the proportion of the Gini attributable to to each of WGI, and in columns 3, 6 and 7 is BGI/G. All models contain year and region indicators. Robust standard errors clustered at the country level have been computed. p-values are in parentheses.

results are similar when the intercept approach is employed to adjust the data. To simplify the notation, the super index of the inequality measures has been dropped.

Table A.2 replicates columns 1 and 3 in Table 6 using three different dependent variables: PRIO-INT$_C$ (columns 1 and 2), PRIOCW$_C$ (columns 3 and 4) and ONSET$_C$ (columns 5 and 6). For each dependent variable, the first column reports estimates relating the Gini coefficient to civil conflict. In all cases we find no significant association. Columns 2, 4 and 6 present results using the Gini decomposition. WGI is positive and significantly related to conflict (at least at the 10% level) for the two "incidence" conflict variables whereas it is not in the Onset regression. It is useful to recall that Esteban and Ray's theory is about the ability of groups to sustain conflict and that it is silent on the determinants of conflict onset.

We have also explored whether the main result holds when the within-country variation in the inequality measures is used to identify the parameters. Since for many countries there exist measures of inequality for more than one year, 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 sample size in these regressions shrinks dramatically. Table A.3 presents the results. The dependent variable is PRIO25_C in columns 1–3 and PRIOINT_C in columns 4–6. Column 1 examines the Gini decomposition variables with all time-varying controls with the exception of lagged conflict. This regression has been estimated by OLS in a linear model with fixed effects, since the algorithm in the conditional logit regression didn't converge.[17] We obtain very similar qualitative results as before: the only variable significantly associated with conflict incidence is WGI. Column 2 introduces lagged conflict in an otherwise identical regression and the conclusions from column 1 remain unchanged. To avoid the effect of the Nickel bias, we've reestimated column 2 using system GMM. As a result, the magnitude of the coefficient of WGI decreases a bit but it is still significant at the 10% level (p-value is 0.064), while those of BGI and OV remain insignificant. Columns 4–6 replicate the previous 3 columns using PRIOINT_C as dependent variable. The results are very similar to those obtained for PRIO25_C.

[17]Notice that the coefficients reported in Table A.3 are not comparable to those in the previous tables. This is due to the fact that a linear specification has been used in this case, while logit or ordered logit models where employed in Tables 6 and A.2 above).

Table A.2: Ethnic inequality and country-level conflict: Alternative dep. variables

	[1]	[2]	[3]	[4]	[5]	[6]
G	2.781		1.905		-2.323	
	(0.287)		(0.571)		(0.499)	
WGI		9.744*		31.549***		-6.363
		(0.094)		(0.001)		(0.333)
BGI		0.252		-13.217**		3.818
		(0.959)		(0.017)		(0.544)
OV		-5.188		-36.833***		-4.169
		(0.421)		(0.000)		(0.597)
F	2.003*	6.781*	3.656**	27.414***	1.492	-0.139
	(0.061)	(0.062)	(0.012)	(0.000)	(0.235)	(0.974)
P	1.456	1.773*	1.245	3.104**	2.928**	3.363***
	(0.149)	(0.063)	(0.284)	(0.011)	(0.019)	(0.007)
NCONT	0.777	1.169*	1.036	3.063***	1.429*	1.941**
	(0.207)	(0.056)	(0.122)	(0.000)	(0.056)	(0.014)
MOUNT	0.010	0.009	0.010	0.008	-0.000	-0.002
	(0.177)	(0.244)	(0.329)	(0.470)	(0.996)	(0.847)
GDP	-0.214	-0.204	0.193	0.191	-0.863**	-0.919**
	(0.319)	(0.385)	(0.479)	(0.448)	(0.025)	(0.019)
POP	0.356***	0.290**	0.637***	0.480***	0.542**	0.536**
	(0.007)	(0.037)	(0.000)	(0.007)	(0.013)	(0.017)
XPOLITY	0.019	0.013	-0.022	-0.022	-0.023	-0.010
	(0.624)	(0.748)	(0.721)	(0.756)	(0.764)	(0.882)
OIL/DIAM	-0.402	-0.414	-0.778**	-1.027**	0.494	0.471
	(0.215)	(0.254)	(0.047)	(0.018)	(0.335)	(0.358)
LAG	3.938***	3.827***	6.468***	5.960***		
	(0.000)	(0.000)	(0.000)	(0.000)		
CONST			-15.504***	-25.489***	-4.388	-2.656
			(0.000)	(0.000)	(0.103)	(0.418)
(Pseudo) R^2	0.560	0.562	0.781	0.795	0.197	0.200
Obs.	1044	1044	1042	1042	887	887

Note. Dependent variable is PRIO-INT$_C$, in columns 1 and 2, PRIOCW$_C$ in columns 3 and 4 and ONSET$_C$ in columns 5 and 6. An ordered logit (logit) specification has been employed in models 1 and 2 (4–6). All models include year indicator variables and regional dummies. The inequality variable have been adjusted using the Ratio approach. Robust standard errors clustered at the country level have been computed and p-values are in parentheses. The period considered is 1992-2009. The number of countries is 88, except for column 7 which is 49. The inequality measures have been computed by averaging all the observations available for each country and assigning this value to the whole period. For each country, the first observation that enters the regression is the first one for which survey data is available for that country.
* p<.10, ** p<.05, *** p<.01.

Table A.3: Ethnic inequality and country-level conflict: country fixed effects

	[1]	[2]	[3]	[4]	[5]	[6]
WGI	0.808*	0.811**	0.541*	0.911*	0.753*	0.687**
	(0.051)	(0.041)	(0.064)	(0.057)	(0.055)	(0.045)
BGI	-0.021	0.060	0.006	-0.154	-0.209	0.176
	(0.954)	(0.890)	(0.982)	(0.690)	(0.645)	(0.600)
OV	0.407	0.474	0.069	0.511	0.349	0.342
	(0.311)	(0.288)	(0.871)	(0.183)	(0.375)	(0.497)
GDP	-0.151	-0.377*	0.029	-0.352	-0.617***	0.046*
	(0.461)	(0.086)	(0.254)	(0.265)	(0.009)	(0.091)
POP	-0.806*	-0.523	0.033**	-0.457	-0.422	0.039*
	(0.098)	(0.235)	(0.042)	(0.375)	(0.414)	(0.059)
XPOLITY	-0.018	-0.008	0.010	-0.020	-0.011	0.008
	(0.242)	(0.626)	(0.164)	(0.287)	(0.570)	(0.453)
LAG		0.338**	0.683***		0.363***	0.822***
		(0.019)	(0.000)		(0.001)	(0.000)
CONST	9.356*	8.525	-0.254	7.675	9.595*	-0.571
	(0.097)	(0.101)	(0.533)	(0.203)	(0.094)	(0.266)
R^2	0.247	0.356		0.226	0.400	
Obs.	213	213	213	213	213	213

Note. Dependent variable is PRIO25$_C$ in columns 1–3 and PRIOINT$_C$ in columns 4–6. Time-variant inequality measures have been employed in all models. Columns 1, 2, 4 and 5 have been estimated using OLS in a linear model with fixed effects (since the algorithm in the conditional logit regression didn't converge) while columns 3 and 6 have been estimated by system GMM. The inequality variables have been adjusted using the ratio approach. Robust standard errors clustered at the country level have been computed and p-values are in parentheses. The period considered is 1992-2009 and the number of countries is 88.
* p<.10, ** p<.05, *** p<.01.

B Appendix: Variables, summary statistics and surveys used

B.1 Variable definitions and summary statistics

This section provides definitions and sources for all the variables employed in our empirical analysis.

Conflict.

PRIO25$_C$. "Armed conflict" : a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths per year and per incompatibility. We consider only types 3 and 4 from the database; these refer to internal armed conflict. Source: PRIO.

PRIO-INT$_C$. "Conflict intensity" : we assign a value of 0 if there is peace in a given year, a value of 1 if there are events satisfying PRIO25$_C$ and the total number of battle deaths that year does not exceed 1000, and a value of 2 if the number of battle deaths is larger than 1000. Source: PRIO.

PRIOCW$_C$. "Intermediate armed conflict" : includes all PRIO25 conflicts that result in a minimum of 1000 deaths over the course of the conflict. We consider only types 3 and 4 (internal armed conflict). Source: PRIO.

ONSET$_C$: "Conflict Onset": Binary variable that takes a value of 1 in year t if PRIO25 is equal to 1 that year but equal to zero in the two previous years. Source: PRIO.

CONFLICT25$_G$: "Group level Armed conflict". A binary measure taking a value of 1 for those years where an ethnic group is involved in armed conflict against the state resulting in more than 25 battle-related deaths. Ethnic groups are coded as engaged in conflict if a rebel organisation 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. Source: Wucherpfennig et al. (2012).

CONFLICT-INT$_G$: "Group level Conflict intensity". We assign a value of 0 if group G is at peace in a given year, a value of 1 if there are events satisfying CONFLICT25$_G$ and the total number of battle deaths that year does not exceed 1000, and a value of 2 if the number of battle deaths is larger than 1000. Source: Wucherpfennig et al. (2012).

ONSET$_G$: "Group level Conflict Onset". A binary measure reflecting the first year in which a group enters a conflict, as defined in CONFLICT25$_G$ above.

Inequality measures.

G^k, $k \in \{\mathrm{ADJ}_I, \mathrm{ADJ}_R, \mathrm{UNAD}, \mathrm{WVS}\}$: Country-level Gini coefficient, adjusted according to the intercept approach ($k =$ADJ$_I$), ratio approach ($k =$ADJ$_R$), unadjusted ($k =$UNAD), and based on WVS data exclusively ($k =$WVS). For each country, all available observations have been averaged so these variables are time invariant.

WGI^k, $k \in \{\mathrm{ADJ}_I, \mathrm{ADJ}_R, \mathrm{UNAD}, \mathrm{WVS}\}$: Within group inequality, as defined in (3). For each country, all available observations have been averaged so these variables time invariant

BGI^k, $k \in \{\mathrm{ADJ}_I, \mathrm{ADJ}_R, \mathrm{UNAD}, \mathrm{WVS}\}$: Between group inequality, as defined in (4). For each country, all available observations have been averaged so these variables are time invariant.

OV^k, $k \in \{\mathrm{ADJ}_I, \mathrm{ADJ}_R, \mathrm{UNAD}, \mathrm{WVS}\}$: Overlap as defined in (5). For each country, all available observations have been averaged so these variables are time invariant.

G^k_t, BGI^k_t, WGI^k_t, OV^k_t, $k \in \{\mathrm{ADJ}_I, \mathrm{ADJ}_R, \mathrm{UNAD}, \mathrm{WVS}\}$: time-varying inequality measures (i.e., non-averaged over countries), otherwise defined as above.

G^k_g, $k \in \{\mathrm{ADJ}_I, \mathrm{ADJ}_R, \mathrm{UNAD}, \mathrm{WVS}\}$: Group-level Gini coefficient as defined in (1), adjusted according to the intercept approach ($k =$ADJ$_I$), ratio approach ($k =$ADJ$_R$), unadjusted ($k =$UNAD), and based on WVS data exclusively ($k =$WVS). For each group, all available observations have been averaged so these variables are time invariant.

HI^k_g, $k \in \{\mathrm{ADJ}_I, \mathrm{UNAD}, \mathrm{WVS}\}$: Horizontal inequality as defined in (2). For each group, all available observations have been averaged so these variables are time invariant.

Controls.

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

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

XPOLITY: 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.

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

NCONT: an indicator variable taking the value 1 in countries with territory holding at least 10,000 people and separated from the land area containing the capital city either by land or by 100 kilometers of water, as measured in Fearon and Laitin (2003).

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

OIL/DIAM: 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).

POP$_g$: Group population. It is computed by multiplying the group share by total population and taking the log. Source: Fearon (2003) and PWT.

GDP$_g$: Group GDP per capita. The surveys provide information to compute this variable. Since they use heterogeneous income definitions, we compute the share of the group's income in total income. and multiply this share by GDP per capita (from the PWT), taking the log.

Table B.1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	Obs
PRIO25$_C$	0.183	0.387	0	1	1044
PRIO-INT$_C$	0.213	0.476	0	2	1044
PRIOCW$_C$	0.147	0.354	0	1	1044
ONSET$_C$	0.027	0.163	0	1	627
CONFLICT25$_G$	0.114	0.318	0	1	1627
CONFLICT-INT$_G$	0.125	0.363	0	2	1627
ONSET$_G$	0.02	0.141	0	1	1627
G^{ADJ_I}	0.425	0.05	0.265	0.563	1044
G^{ADJ_R}	0.387	0.091	0.231	0.647	1044
G^{UNAD}	0.304	0.063	0.161	0.563	1044
G^{WVS}	0.27	0.046	0.161	0.382	611
WGI^{ADJ_I}	0.181	0.07	0.016	0.315	1044
WGI^{ADJ_R}	0.178	0.081	0.011	0.378	1044
WGI^{UNAD}	0.142	0.067	0.016	0.272	1044
WGI^{WVS}	0.154	0.057	0.035	0.275	611
BGI^{ADJ_I}	0.095	0.042	0.017	0.259	1044
BGI^{ADJ_R}	0.061	0.055	0.001	0.307	1044
BGI^{UNAD}	0.047	0.041	0.001	0.204	1044
BGI^{WVS}	0.031	0.037	0.001	0.215	611
OV^{ADJ_I}	0.149	0.073	0.016	0.451	1044
OV^{ADJ_R}	0.148	0.085	0.011	0.339	1044
OV^{UNAD}	0.115	0.071	0.01	0.451	1044
OV^{WVS}	0.086	0.042	0.01	0.165	611
G$_g^{ADJ_I}$	0.464	0.082	0.173	0.714	1627
G$_g^{ADJ_R}$	0.375	0.102	0	0.688	1627
G$_g^{UNAD}$	0.298	0.075	0	0.51	1627
G$_g^{WVS}$	0.24	0.07	0	0.352	504
H$_g^{ADJ_I}$	0.093	0.193	0	1.333	1627
H$_g^{UNAD}$	0.095	0.198	0	1.214	1627
H$_g^{WVS}$	0.07	0.2	0	1.188	504
GDP	8.398	1.348	5.62	10.829	1044
POP	9.628	1.361	6.631	13.947	1044
XPOLITY	3.503	4.03	-5	7	1044
F	0.507	0.24	0.077	0.953	1044
P	0.58	0.199	0.154	0.986	1044
NCONT	0.155	0.362	0	1	1044
MOUNT	15.949	19.606	0	81	1044
OIL/DIAM	0.266	0.442	0	1	1044
POP$_g$	10.215	1.681	7.593	13.961	1627
GDP$_g$	7.247	0.842	5.434	10.249	1627

Table B.2: 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)