Measuring Ethnic Inequality: An Assessment of Extant Cross-National Indices

This article offers an evaluation of cross-national measures of ethnic socioeconomic inequality. It demonstrates that the measures differ in important ways regarding empirical scope, conceptualization, measurement and aggregation. Despite significant advances in the measurement of ethnic inequality, all measures have shortcomings, such as limited and biased coverage as well as measurement error from the underlying data sources. Moreover, the empirical convergence between conceptually similar measures is strikingly low; some of the measures show no or even negative covariation. Four replication studies also indicate that extant measures of ethnic inequality are generally not interchangeable. Scholars should therefore take the various features highlighted in this evaluation into account before employing any of them. Based on this conclusion, the article offers multiple suggestions for improving existing measures and developing new ones.

Keywords: ethnic inequality; horizontal inequality; conceptualization; measurement validity; data set assessment

Introduction

Economic and social inequalities between ethnic groups – also known as horizontal inequalities (Stewart 2008) – have received increased attention in academia and policy circles in recent years. This growing interest is clear from the dramatic increase in the number of related academic publications.[1] A considerable body of political science research suggests that within-country inequalities between ethnic groups have major negative implications for peace, economic and political development, public goods provision and individual well-being (e.g. Alesina et al. 2016; Baldwin and Huber 2010; Canelas and Gisselquist 2019; Cederman et al. 2015; Houle 2015; Houle and Bodea 2017; Stewart 2008; Wang and Kolev 2019; Ye and Han 2019). Furthermore, the reduction of group-level inequalities is included in Sustainable Development Goal 10 (UN 2020), and the issue was emphasized in a recent OECD report (Deere et al. 2018).

Much of the comparative research that has flourished in the past decade is premised on a series of relatively new datasets. These are valuable tools that can help monitor variation across space and time as well as analyze causes and consequences. A few methodological studies have addressed measurement challenges related to survey and census data (Canelas and Gisselquist 2019), suggested good measurement practices (Stewart et al. 2010) and discussed data sources (Baghat et al. 2017; Tetteh-Baah 2019). However, problems of causal inference have largely overshadowed important problems of conceptualization and measurement, and there are currently no systematic comparative evaluations of how extant cross-national measures relate to each other conceptually and empirically. This also means that we have limited knowledge of the strengths

[1] A keyword search on dimensions.ai for 'ethnic inequality' in social sciences subjects returns 87 publications in 2000 against 419 in 2020. The corresponding numbers for 'horizontal inequality' are 9 in 2000 and 263 in 2020.

and weaknesses of the various measures, including whether they can be considered interchangeable.

Against this background, this article contributes to the emerging literature on ethnic inequalities by discussing and comparing six different cross-national measures offered by Alesina et al. (2016), Cederman et al. (2013), Houle (2015), Baldwin and Huber (2010), Omoeva et al. (2018) and V-Dem (Coppedge et al. 2021a).[2] Scholars have used the evaluated measures in empirical studies to operationalize social or economic ethnic inequality cross-nationally, and they cover the majority of contemporary countries – or at least include countries from several world regions. Even though not all of these measures were created for broad purposes, they are increasingly being used for different empirical research (e.g. Fleming et al. 2020; Ye and Han 2019), which underlines the need for systematic comparison.

The examination of the six indices is inspired by the steps in the integrated assessment framework suggested by Munck and Verkuilen (2002), which provides a comprehensive checklist to evaluate data. However, my examination also goes beyond their framework by providing a series of new data visualizations as well as four replication studies. In the assessment, I find clear differences in conceptualization, measurement, aggregation, and empirical scope. Dramatic differences in coverage influence their relevance for research questions about the causes or consequences of ethnic inequality, which rely on cross-national and, especially, cross-temporal variation. The majority of measures are also afflicted by important biases, such as mainly covering developing countries or focusing exclusively on democracies. Moreover, a comparison of the data sources – including mass surveys, expert surveys, administrative data, and satellite data on

[2] Houle as well as Omoeva, Moussa and Hatch generously shared their data. The remaining datasets were downloaded from online databases.

nightlights – reveals likely sources of measurement error. A number of correlation analyses show that the empirical convergence between the measures is surprisingly low, even when taking into account the differences in conceptualization and aggregation procedures. Notably, two measures based on similar definitions exhibit no significant correlation at all. Moreover, the replication studies suggest that the results of a number of prominent studies are sensitive to measurement choice. The article thus aims to raise awareness about extant measures of ethnic inequality so that their respective strengths and weaknesses are taken into account in the assessment of previous studies and the design of new ones. Based on these findings, I discuss potential avenues forward, including more disaggregated analyses and combining various data-sources.

Conceptualization

At the most general level, inequality is about 'the ability of households to maintain economically a certain standard of living and lifestyle' (Jensen and van Kersbergen 2016: 36). If individuals or families have very different options in terms of how to live their lives, we intuitively consider them as living in an unequal society. Conceptually, we may distinguish between inequality on the individual and group levels. Interpersonal (or 'vertical') inequality is about differences between individuals or households, typically referring to disparities in post-tax-transfer disposable household income in a given year (Jensen and van Kersbergen 2016: 36). The empirics are typically summarized into comparable measures using Gini coefficients, ratios between income percentiles, or income shares going to the top percentiles (Jensen and van Kersbergen 2016: 36-47; Piketty 2014).

Intergroup (or 'horizontal') inequality concerns between-group differences, which are defined according to the type of group identification one is interested in studying, such as ethnicity

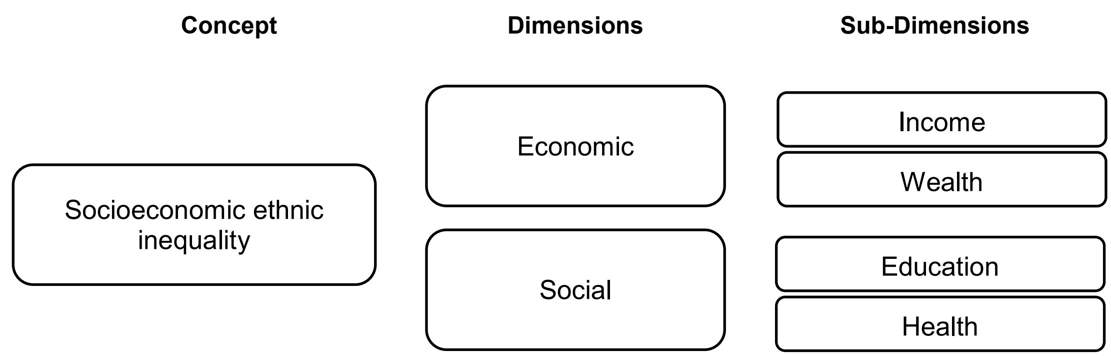
(Stewart 2002: 13). Ethnic inequalities can be measured both at the aggregate, country level (providing a single figure, which represents the entire distribution in a country) and at the group level (providing figures for each group relative to the country mean or another group). This article focuses exclusively on aggregate, cross-national measures, which have been employed by most comparative studies so far (Baghat et al. 2017: 67). They use the average differences in outcomes, such as income or education, between ethnic groups in a society, aggregating them for comparisons across countries and over time.[3] In the surveyed works, ethnicity is generally understood in an encompassing manner consistent with the recent literature on ethnic politics (Canelas and Gisselquist 2018: 306; Chandra 2006: 398; Horowitz 2000). Following the tradition of Max Weber, ethnicity may defined as a subjectively experienced sense of commonality based on a belief in common ancestry and shared culture (Weber 1976 [1922]: 389). Ethnic identity markers indicating a shared ancestry and culture, include language (e.g., Belgium), religion (e.g., Bosnia and Herzegovina), tribe (e.g., Kenya), caste (e.g., India), phenotypical features (e.g., the US), or some combination thereof. In other words, ethnic categories are social constructs linked to descent-based attributes.

Are the surveyed measures of ethnic inequality based on similar conceptual foundations? The examined datasets variously refer to 'economic horizontal inequality' (Cederman et al. 2013:

[3] Economic and social ethnic inequality is closely related to and may be subsumed under the broader concept of *horizontal inequalities*, which deserves brief clarification due to its prominence in the literature. Horizontal inequalities refer to 'inequalities in economic, social, or political dimensions or cultural status between culturally defined groups' (Stewart 2008: 3). Stewart coined this term to distinguish it from interpersonal inequality – also referred to as vertical inequality. Horizontal inequalities are multidimensional (incl. political and cultural inequalities) and potentially refers to any relevant group inequality, such as gender (Deere et al. 2018).

93), 'differences in the economic well-being of groups' (Baldwin and Huber 2010: 645), 'between-ethnic-group inequality (bgi)' (Houle 2015: 470), 'within country differences in well-being across ethnic groups' (Alesina et al. 2016: 429), 'inequalities in education […] between ethnic groups' (Omoeva et al. 2018: 3) and 'inequalities in access to public services […] between particular social groups' (Coppedge et al. 2021a: 218). To invoke a useful distinction by Adcock and Collier (2001), they are not 'systematized concepts' but seem to agree on the 'background concept'; that is, despite different terminologies, all of the surveyed measures share a common conceptual core, as they all reflect asymmetries in socioeconomic conditions between ethnic groups. Importantly, all datasets are explicit about which dimension of ethnic inequality they are capturing (see Stewart 2002). The economic dimension concerns the distribution of income and wealth between ethnic groups (i.e., Alesina et al., Baldwin and Huber, Cederman et al., and Houle); the social dimension concerns the uneven access of groups to public services, such as healthcare and education (i.e., Coppedge et al. and Omoeva et al.). These dimensions not only reflect a common core, socioeconomic ethnic inequality, but are also likely to be highly correlated due to common determinants and reciprocal relationships: Inequality in access to public services may translate into and be highly associated with economic ethnic inequality and vice versa (see Stewart et al. 2010). The conceptual structure is illustrated in Figure 1.

Figure 1: Illustration of conceptual structure



Measurement

The various dimensions and sub-dimensions of socioeconomic ethnic inequality can be operationalized in various ways using a range of indicators. Since the data providers implicitly agree on a background concept (i.e., ethnic inequality concerns differences in standards of living between ethnic groups), a comparison of these measures seems meaningful.

Overview of Extant Measures

Alesina, Michalopoulos, and Papaioannou's (2016) ethnic Gini indices are based on satellite images of nighttime luminosity combined with the homelands of ethnolinguistic groups. This measure reflects differences in "mean income" – as reflected by luminosity per capita across ethnic homelands – between groups within 173 countries (in 1992, 2000, and 2012). Ethnic groups are located using two datasets/maps. First, the Geo-Referencing of Ethnic Groups (GREG), which is the digitized version of the Soviet Atlas Narodov Mira from the 1960s (Weidmann et al. 2010). The second source is the fifteenth edition of the Ethnologue (Gordon 2005), which maps 7,581 language-country groups worldwide in the mid-to-late 1990s. GREG attempts to map major immigrant groups, whereas Ethnologue generally does not. Hence, the two ethnolinguistic mappings capture different ethnic groups, which is particularly important for countries in the Americas (Alesina et al. 2016: 433).

Cederman, Gleditsch, and Buhaug (2013) geographically match subnational economic data (G-Econ data by Nordhaus et al. 2006) with data on geographical boundaries of ethnic settlements from the GeoEPR dataset (Wucherpfennig et al. 2011). While their analytical focus is on investigating group-level data and civil war onset, they also conduct cross-national analyses (see also Buhaug et al. 2014). Strictly speaking, the temporal scope is limited to a single year, because the G-Econ data only reflects 1990 values, whereas only the Geo-EPR is dynamic, taking into account major changes in ethnic settlement patterns over time (Cederman et al. 2013: 101, 106).

Table 1: Scope, Sources and Operationalization of Extant Measures

Data provider and Index	Countries	Years	Country-year observations	Sources	Operationalization
Alesina et al. (2016): - Ethnic Gini	173	1992, 2000, 2012	519	Nightlights/ ethnic homelands	Group Gini (0–1)
Cederman et al. (2013): - G-Econ/Ethnic homeland	163	1990	163	Local econ. data/ethnic homelands	Ratio (poorest/richest group relative to country mean)
Houle (2015): - Between-group inequality (BGI)	75	1960–2007 (unbalanced)	1,641	Mass Survey	BGI indicator (0–6)
Baldwin & Huber (2010) - Between-group inequality (BGI)	46	1996–2006 (unbalanced)	46	Mass Survey	BGI indicator (0–1; stat. standardized scores available)
V-Dem Coppedge et al. (2021) - Access to public services by group	179	1900–2020	18,157	Expert survey	Point estimate and confidence bounds based on IRT-model (original scale: 0–4)
Omoeva et al. (2018): - Educational Group Gini	86	1946–2013 (unbalanced)	4,254	Mass Survey	Group Gini (+ Theil, Cofficient of Variation and Parity Index)

Note: The measures below the line measure the social dimensions; those above it measure the economic dimension (see conceptualization). Note that nightlights may also proxy for access to public services, meaning the distinction between economic and social dimensions is not clear-cut.

Houle (2015) uses information from a range of surveys, including the Demographic and Health Survey (DHS), World Values Survey (WVS) and various regional barometers, to construct an asset-based wealth indicator for within-group, between-group and cross-national ethnic inequality. Since the data were originally gathered to study democratic breakdowns, the measure covers 89 countries from 1960 to 2007, which have been democratic for at least one year and are ethnically heterogeneous. The panel is unbalanced and exhibits limited variation over time (Houle 2015: 500).

Baldwin and Huber (2010) construct a between-group inequality (BGI) measure similar to a group Gini coefficient for 46 democracies based on income variables from a series of surveys. The sample includes democracies from all regions of the world, although Asia and especially Latin America are underrepresented in so far as these regions have a higher proportion of democracies than the data set suggests (Baldwin and Huber 2010: 648). Each country is measured in one year between 1996 and 2006, effectively making the data cross-sectional. The data only includes democracies, since they were originally collected for the purpose of studying public goods provision in heterogeneous democracies. As pioneers in the field, Baldwin and Huber are careful to validate their measure empirically, including comparison of their measure to a handful of countries, where the nature of inequality between groups is widely acknowledged. Moreover, they turn to a number of more fine-grained household surveys that identify income by ethnic group (Baldwin and Huber 2010: 649-50).[4]

Finally, two measures capture unequal access to public services rather than economic outcomes. In the newest data release (v11.1), Varieties of Democracy (V-Dem) provides an expert-coded indicator of inequality in access to basic public services (e.g., primary education, clean water, and healthcare) distributed by 'social group'. The group definition corresponds to a broad conception of ethnicity covering, among other things, language, race, and religion (Coppedge et al. 2021a: 209). The dataset covers all sovereign states in the world since 1900 with the exception of a number of micro-states.

As part of the Education Inequality and Conflict (EIC 2015) project, Omoeva, Moussa, and Hatch (2018: 16) have created measures of inequality in educational attainment between

[4] Wang and Kolev's (2019) dataset explicitly builds on Baldwin and Huber and has thus not been included in the main discussion.

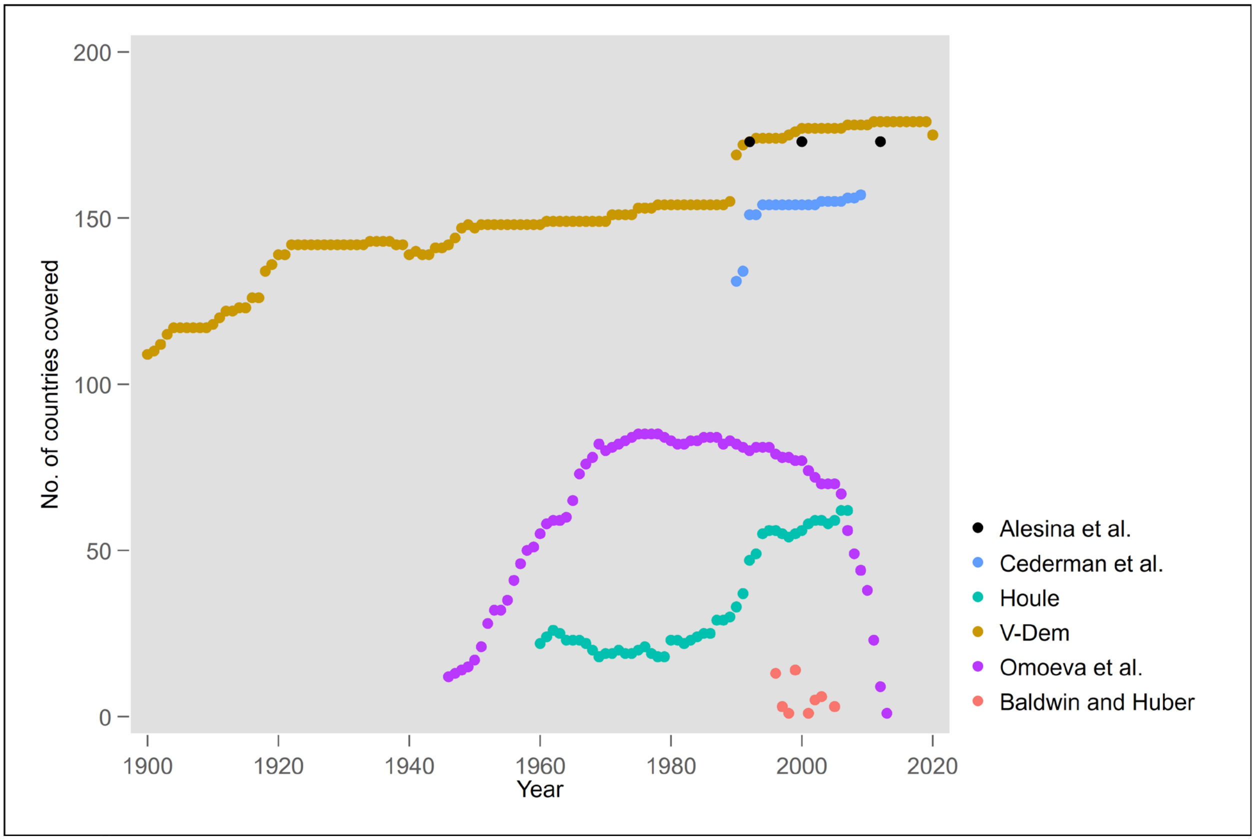
ethnic/religious groups by constructing a group Gini coefficient (as well as Theil Index, Coefficient of Variation and Parity Ratio). They draw on educational attainment data from three public household survey datasets (Omoeva et al. 2018: 15) and fill in missing country-year observations using a logical backward projection technique. The unbalanced dataset covers a set of 86 predominantly developing countries between 1946 and 2013 (Omoeva et al. 2018: 50).

Despite the measures all sharing a common focus on ethnic inequality, they are marked by dramatic differences in scope, ranging from a cross-section of 46 countries to a measure covering most polities since 1900 (see Table 1 and Figure 2).[5] There are also large differences in terms of how time-varying the data are. This is illustrated in Figure 3, which plots the values of the different measures over time for Bolivia, where a high level of ethnic inequality is widely acknowledged (Houle 2015: 485). The V-Dem, Omoeva et al., and Alesina et al. measures exhibit significant variation over time. In contrast, save for a minor change in the Houle measure, the Cederman et al. and Houle measure are time-invariant.[6]

[5] Figure A1-A6 presents maps showing the countries covered by each dataset.

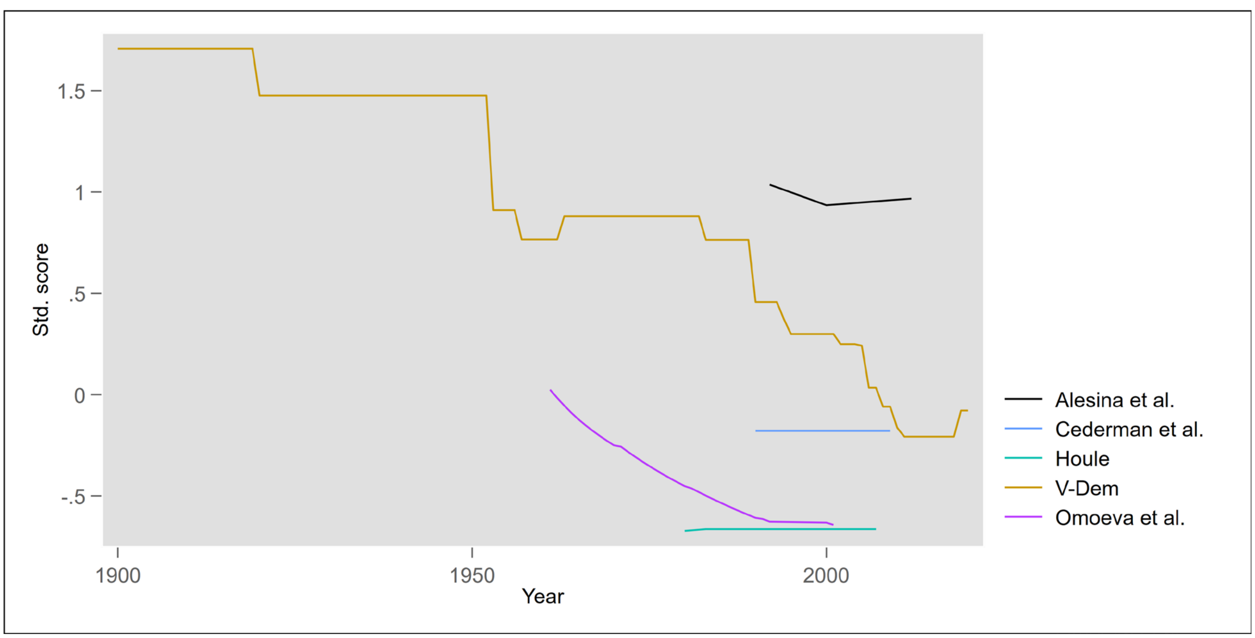
[6] Table A1 also reveals important differences in the temporal granularity of the different measures.

Figure 2: Country-coverage over time by measure



Notes: Each point indicates how many countries are covered in a given year by the dataset in question. V-Dem covers many colonies, which explains its high country-coverage prior to decolonization.

Figure 3: Temporal variation of measures – Bolivia as example



Notes: The cross-sectional Baldwin and Huber measure is not included. To ensure comparability, the variables have been standardized (mean of 0, standard deviation of 1).

The creators of the first cross-national measures, such as Baldwin and Huber (2010), deserve much credit for paving the way with their work to conceptualize and create the first ethnic inequality measures. For the purpose of future empirical studies, however, the restricted empirical scope of most measures limits their value for particular research questions. In particular, the ability to track the developments over time with respect to ethnic inequality is severely restricted.

Finally, there are strong non-random patterns in the data. Most clearly, not all measures support direct comparisons between poor countries versus the experiences of rich, long-enduring democracies. In particular, the Omoeva et al. (2018) data only include a limited number of high-income countries, whereas the Baldwin and Huber (2010) and Houle (2015) datasets only include democracies. I further explore this issue in the Online Appendix (Table A1) with a simple test of non-random missingness (see Rios-Figueroa and Staton 2012: 125). These findings show that most measures provide samples that are not representative regarding GDP/cap, democracy, and state capacity. Looking across the tests, the V-Dem and Alesina et al. measures appear to be least afflicted by non-random missingness. These non-random patterns in the data reduce the ability to infer from the sample to the general population of all countries, and they mean we should avoid being overly confident about any robustness analysis using alternative measures.

Data sources

In addition to well-known measurement constraints for interpersonal (or vertical) inequality, the measurement of ethnic inequality depends on comparable group classifications. This represents a significant challenge, since ethnic identities are not static, people hold multiple identities, and data is often unavailable or incomplete (Bochsler et al. 2021; Canelas and Gisselquist 2019: 161;

Stewart et al. 2010: 10). Dataset creators have creatively addressed these challenges and collected data in three general ways: (1) surveys, which include information on both socioeconomic wellbeing and ethnic group affiliations; (2) spatial datasets, which geographically match economic data with data on geographical boundaries of ethnic settlements; or (3) expert coding.

More specifically, the challenge of identifying comparable ethnic categories has been addressed in three main ways. One strand, which includes Baldwin and Huber (2010); Alesina et al. (2016), Cederman et al. (2013), and Houle (2015), adheres to the ethnic group classification as coded by either Fearon (2003) or Ethnologue (Gordon 2005), or by the Ethnic Power Relations (EPR) dataset or its geocoded extension, GeoEPR (Vogt et al. 2015; Weidmann et al. 2010). Another strand, represented by Omoeva et al. (2018), uses the ethnic categories that have been pre-defined by the teams that develop surveys. Finally, V-Dem (Coppedge et al. 2021a) uses experts' local knowledge to assess ethnic groups based on a prior group definition.[7]

In terms of socioeconomic data sources, Baldwin and Huber (2010), Houle (2015), and Omoeva et al. (2018) all take national household surveys, which include information on both socioeconomic wellbeing and ethnic group affiliations. On the one hand, biased information is unlikely when data are generated from surveys like the Demographic and Health Surveys (DHS),

[7] This represents an alternative approach to measuring identity-based socioeconomic inequality; instead of first identifying the ethnic groups in a state and subsequently determining their mean socioeconomic status (as a basis for calculating e.g., a Gini measure), a scholar may determine the overall degree to which ethnic identities are associated with socioeconomic inequalities. The choice of comparable ethnic categories clearly matter, exemplified by the fact that V-Dem categorizes Qatar as highly unequal, whereas it receives a score of 0 in the Alesina et al. data, which suggests that V-Dem coders also take the large non-citizen populations into account, whereas the Ethnologue does not.

as the original intention was not to assess socio-economic inequalities between ethnic groups.[8] On the other hand, survey and census data on ethnic issues may entail (intentionally or not) incomplete and biased responses: Minority groups may not be accurately represented in national surveys, answers could be significantly affected by the sometimes politically sensitive nature of ethnic identities (Canelas and Gisselquist 2019: 165) and more politically stable countries are more often surveyed in the DHS program. In the African context, for instance, Libya, Eritrea, Somalia, Sudan, and the Central African Republic are not included (Tetteh-Baah 2019: 31).[9]

In light of the gaps and weaknesses in survey- and census-based data on ethnic inequality, Cederman et al. (2013) combine data on ethnic groups' settlement areas with the Nordhaus et al. (2006) G-Econ dataset on local economic activity to measure economic ethnic inequalities. Similarly, Alesina et al. (2016) have worked with various proxy measures to combine geocoded nightlight data with historical maps of ethnic territories or homelands. While these spatial measures provide higher coverage, they also suffer from numerous drawbacks. Measures of local economic activity hinge on the quality of the underlying sources, and data quality is particularly poor for countries with unreliable official statistics and substantial informal economies (Baghat et al. 2017: 82; Chen and Nordhaus 2011). Nightlight emissions from satellite data are an alternative that is independent of governmental bias or the limited quality of official statistical sources. However, like the other measures, this data source is also afflicted by weaknesses, such as

---

[8] Surveys include the DHS, Multiple Indicator Cluster Surveys (MISC) and more opinion-focused surveys, such as the regional barometers and World Values Survey (WVS); see Baghat et al. (2017: 75) for an overview.

[9] See Canelas and Gisselquist (2019: 164-168) and Cederman et al. (2015: 808) for further discussion of survey and census-data challenges. See also Bochsler et al. (2021) for general considerations of measuring ethnic identities quantitatively, including census and survey-based measures.

constituting a relatively indirect proxy for economic development (Chen and Nordhaus 2011), and official data sources are likely to be more accurate in developed countries (Mellander et al. 2015). Moreover, both spatial methods may lead to measurement error in cases where the ethnic group settlement areas largely overlap. Consequently, spatial approaches cannot accurately estimate the economic inequalities between, for example, the Tutsi and Hutu in Rwanda and Burundi (Alesina et al. 2016: 449; Cederman et al. 2015: 807). Returning to the issue of scope and temporal variation, it is worth noting that surveys may or may not be available in a regular time series format, whereas satellite-based measures – as well as updated ethnic homelands data (e.g., GeoEPR) – are available in time series from the 1990's and onward. Consequently, satellite-based measure may help track trend across and within countries with improved temporal granularity in the future.

The V-Dem measure is based on coding by multiple country experts of the question whether 'basic public services, such as order and security, primary education, clean water and healthcare, [are] distributed equally across social groups?' (Coppedge et al. 2021a: 218). The advantage of this approach is the ability to capture latent phenomena based on experts' country-specific knowledge. Given the difficulty of obtaining comparable observable data for public service provision by ethnic group, the assessments made by country experts can become useful when measuring social ethnic inequalities (see Munck et al. 2020: 341). As with any judgement-based data, however, this approach also has its challenges, including the risk of personal biases, limited or biased background information, and reliability issues stemming from inconsistently applied coding criteria (Skaaning 2018: 111-113). As elaborated below, the V-Dem approach increases comparability and reduces the biases inherent in expert codings, which alleviates some of these concerns. However, compared to the other measures, it is much more difficult for us to revisit the data sources, which is relatively easy with the public surveys, G-Econ, or nightlights data. As such, it is impossible to verify, for instance, which ethnic groups form the basis of the expert coding or how much relevant information the expert actually has about ethnic inequality regarding a particular year. Both regarding concept and empirics, we simply cannot know exactly what the coders had in mind when arriving at their assessments, since coders are not required to justify their decisions.

Table 2: Strengths and Weaknesses of Different Data Types

	Surveys	Spatial data	Expert coding
Strengths	● Most direct measure of relative well-being	● Country coverage ● Absence of political biases (for nightlight)	● Country coverage ● Expertise to capture latent phenomena
Weaknesses	● Unrepresentative of ethnic composition ● Answers affected by politically sensitive nature ● Unstable countries/regions undersampled	● Indirect measure ● Data quality of official sources (G-Econ) ● Inability to account for overlapping settlement patterns	● Indirect measure ● Risk of personal biases ● Limited access to relevant information ● Inconsistent application of coding criteria ● Inability to revisit data-sources

The discussed strengths and weaknesses of the data sources are reported in Table 2. As should be apparent, there are no fundamentally superior data sources with the current data availability. In this sense, data choices should be governed by the research question at hand: When studying a specific region, survey measures may prove superior to spatial or expert-coded data, whereas spatial or expert-coded data is more likely to be relevant for global patterns. In that sense, there is a certain tradeoff between the geographical and temporal coverage of the data versus its quality (see also Baghat et al. 2017: 82). I discuss the option of combining various data sources at the end of the paper.

Aggregation

All of the measures are based on different items of information that must be combined to develop the overall measure. Stewart et al. (2010) consider principles of good measures and make the case for three ways to measure aggregate group inequality: the GGini, GTheil, and GCOV, which correspond to the classical Gini coefficient, the Theil index, and the coefficient of variation. Instead of calculating inequality based on each individual's income, it assigns each group's mean income to every member of that group (Baldwin and Huber 2010: 646-648; Stewart et al. 2010: 15). The most established measure, the group Gini index, captures the normalized mean difference between all group incomes in a country, weighted by the population size of each group. Like the Gini coefficient, it ranges from 0 to 1 and offers an interpretation related to the Lorenz curve, as described in detail by Baldwin and Huber (2010: 646). The measure takes on its minimum value when the average incomes of all groups in society are the same, and it takes on 1 when one infinitely small group controls all income (Baldwin and Huber 2010: 646).

The group Gini index is adequate in terms of capturing the general level of inequality across countries over time. Alesina et al. (2016) follow this procedure and construct two 'ethnic Ginis'. Similarly, Omoeva et al. (2018) calculate a group Gini coefficient (as well as a group Theil and a coefficient of variation) for educational attainment across ethnic groups. Although differing in terminology, the Baldwin and Huber (2010) BGI measure is calculated in the same way as the group Gini described above (Baldwin and Huber 2010: 646). Although similar, the aggregate measure by Houle (2015) departs slightly from the Baldwin and Huber GGini or 'BGI' formula.[10]

[10] Houle calculates a population-weighted average of a given country's group-level inequalities. Group-level inequalities, in turn, are calculated following Cederman et al. (2013) as the logarithmized ratio between the average

Another group of measures is employed by scholars who empirically investigate theoretical arguments that only require one group to mobilize. Cederman et al.'s (2013: 143-67) cross-national measure captures the difference between the national average per capita income level and the per capita income of the most (dis)advantaged ethnic group in the country. The authors are explicit that such a 'weakest link logic' is more theoretically relevant when studying civil war onset (Cederman et al. 2013: 145), because measures based on averages or summed features would discount small, atypical groups, especially in large countries. However, such groups might also be the most conflict-prone. While diverging from Stewart et al.'s (2010) suggested approach, the data providers have based their aggregations on explicit theory. Overall, this aggregation procedure means that the measure should differ substantially from the others, since it is not intended to measure overall inequality.

Finally, The V-Dem measure is aggregated using the standard V-Dem methodology. Expert assigned scores are aggregated through a Bayesian item response theory (IRT) measurement model, which also uses information about coder agreement, self-assigned uncertainty estimates, personal coder characteristics, links between countries based on experts assessing more than one country, and responses to vignettes related to the survey questions in order to align the experts' thresholds and calculate uncertainty estimates (Coppedge et al. 2021b: 16-25; Pemstein et al. 2019). This procedure supposedly reduces potential biases, but it cannot eliminate them altogether. For purpose of comparison, the measure has been recoded to go from 0 to 1, with higher values indicating greater inequality.

---

income of members of an ethnic group and the average per capita income of the entire country (Houle, 2015: 482). The justification for this aggregation is unclear.

In sum, the datasets aggregate their data in three different ways. First, measures that reflect the entire distribution of resources or access to public services in a society through measures such as the GGini (Alesina et al.; Baldwin and Huber; Houle; Omoeva et al.). Second, ratio measures focusing explicitly on the poorest (or wealthiest) groups in society relative to the country mean (Cederman et al. 2013). Third, indices summarizing different experts' coding providing an easy-to-interpret number (Coppedge et al. 2021a). As discussed, most measures are aggregated based on existing best-practice or explicit theory. In this sense, each measure is appropriate for different research questions. Most clearly, researchers interested in cross-national differences that take into account the entire group distribution should opt for the first or third category, whereas the second category may be relevant when studying particular ethnic mobilization patterns.

Empirical Comparison

The many differences and similarities in the conceptualizations and measurements of ethnic inequality render it relevant to explore the statistical association between the indices. Comparisons of competitive measures linked to similar background concepts are often assessed by simple correlation tests to clarify whether they tend to tap into the same phenomenon. Following this tradition, Table 3 presents the bivariate correlations between the ethnic inequality indicators. For the purposes of this exercise, I only include the Alesina et al. Ethnologue-based measure, which draws on more recent spatial data (the two measures are correlated at 0.73). Moreover, from Cederman et al. (2013), I only include the ratio of the poorest group relative to the mean (see Table A8 in the Online Appendix for a full correlation analysis covering all measures).

Since all of the measures were argued to reflect the same background concept, share causal determinants and affect each other, we would expect them all to be at least moderately correlated.

Moreover, measures supposed to capture the same dimension (i.e., public services or income/wealth) should show a high level of co-variation. In Table 3, the topmost measures (Alesina et al., Baldwin and Huber, Cederman et al., and Houle) reflect the economic dimensions, whereas the lower two (Coppedge et al., Omoeva et al.) reflect the social dimension. In addition, since Cederman et al. use a distinct aggregation procedure, we expect this measure to exhibit lower correlations with the other measures.

Table 3: Correlations between Measures

	Alesina et al.	Cederman et al.	Houle	Baldwin & Huber	Coppedge et al.	Omeova et al.
Cederman et al. (2013)	0.16 (295)					
Houle (2015)	0.01 (102)	0.12 (970)				
Baldwin & Huber (2010)	n/a	0.04 (46)	0.01 (30)			
V-Dem (2021)	**0.55** (484)	0.05 (3042)	0.31 (1641)	**0.64** (46)		
Omoeva et al. (2018)	**0.40** (160)	-0.07 (1380)	0.30 (750)	0.05 (21)	0.17 (3983)	
Factor loadings (factor 1)	**0.82**	**0.53**	**0.67**	n/a	**0.58**	**0.54**
Factor loadings (factor 2)	-0.33	**0.68**	**0.50**	n/a	-0.63	-0.12

Note: results refer to bivariate Pearson's r correlations (n in parentheses), values over 0.4 in bold. 'n/a' indicates no country-year overlap. The topmost measures reflect the economic dimension, whereas the lower two (in grey) reflect the social dimension. Principal component factor analysis (unrotated).

The most striking observations from Table 3 are the many weak correlations. Only 3 out of 14 are higher than 0.4. To provide a point of comparison, measures of democracy – which also vary substantially in terms of their exact conceptualization and measurement – tend to be highly correlated, typically at 0.8 or higher (Marquez 2016: 11-16). In the same vein, despite varying definitions and data sources, conventional measures of socioeconomic inequality also tend to be highly correlated (in the range between 0.44-0.90).[11] The Alesina et al. measure shows a moderate correlation with the two measures of equal access to social services by V-Dem and Omoeva et al., yet it is virtually uncorrelated with the other measures.[12] Moreover, the V-Dem measure is relatively highly correlated with the Baldwin and Huber measure (0.62). Most surprisingly, the Houle measure shows virtually no correlation with the Alesina et al. measure. Moreover, it is only weakly correlated with the Cederman et al. measure, while showing a slightly stronger co-variation with the two measures capturing equal access to public services (about 0.3). Perhaps equally surprising, the Cederman et al. G-Econ measure is negatively correlated with the Omoeva et al. measure. Contrary to the expectations, the two measures of social ethnic inequality (V-Dem and Omoeva et al.) are only weakly correlated with each other (0.17).[13] To ensure that these results are not simply an artifact of differences in samples, I conduct a series of additional correlation analyses in the Online Appendix, including overlapping time periods and a core set of countries (Table A3-A6). This exercise corroborates the overall pattern of surprisingly low correlations between most measures.

Figure 4 maps the standardized values (mean of 0, standard deviation of 1) for the Alesina et al. and the Cederman et al. data to provide a better sense of the empirical patterns in each dataset and show how individual countries are scored relative to each other. This also provides country-

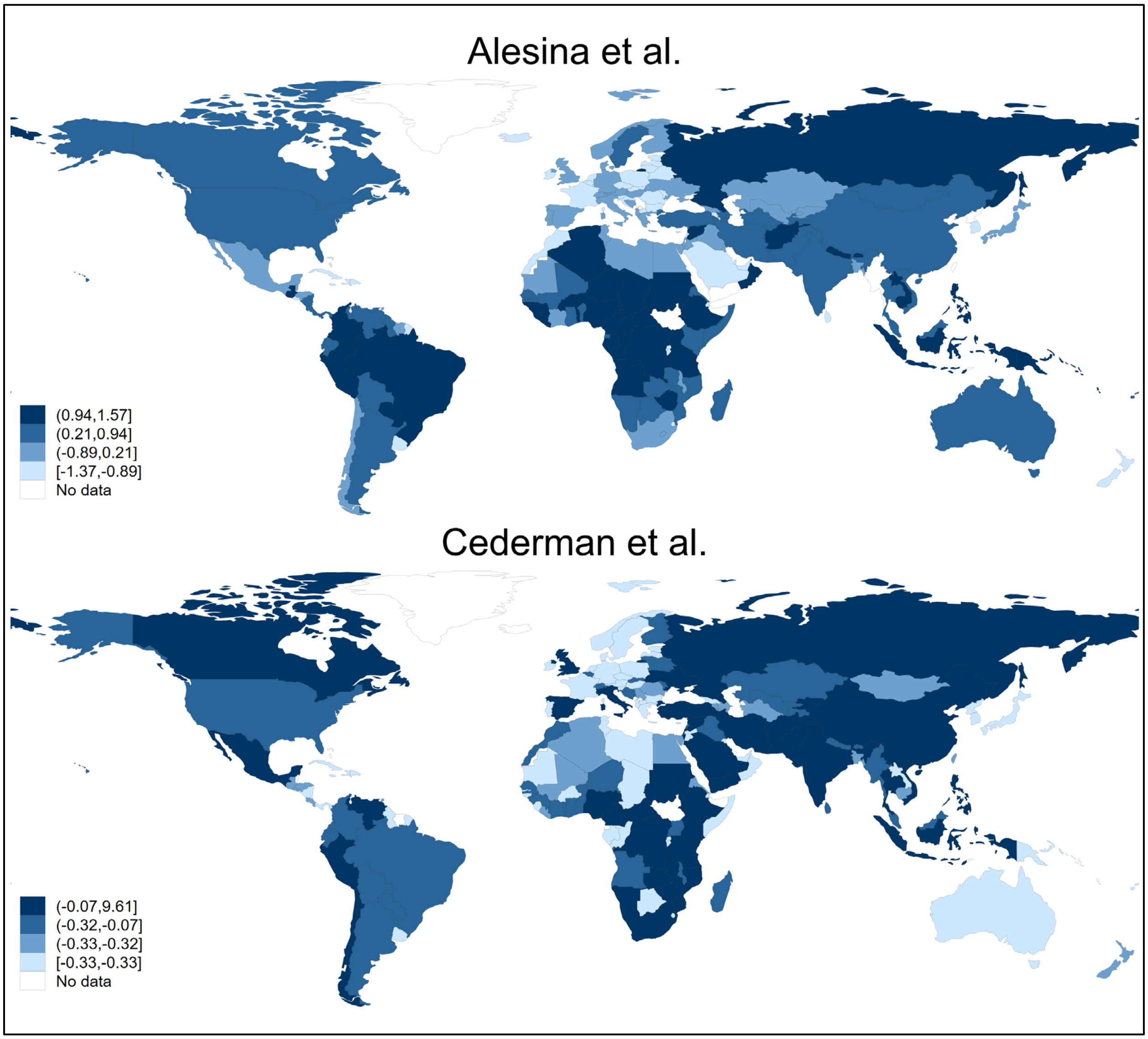
[11] Referring to disposable income Gini by the SWIID, World Development Indicators, a wage share measure as well as V-Dem's measure of inequality in "access to public services by socio-economic status" (see Table A8).

[12] This may partly be a result of the fact that nightlights may also capture access to public services.

[13] Moreover, there is no strong covariation between measures based on overlapping data sources: the Alesina et al. and Cederman et al. spatial measures versus the Houle, Baldwin and Huber, and Omoeva et al. survey-based measures.

or regional experts with an opportunity to assess the face validity of these scores (maps for the other measures are provided in the Online Appendix).

Figure 4: Standardized values for Alesina et al. and Cederman et al. (2000)



Notes: The figure displays the standardized country scores for the data by Alesina et al. (top panel) and Cederman et al. (lower panel) in the year 2000. The values are grouped by quartiles

While there is rough agreement on a number of cases, such as Peru, Congo DRC, and Ethiopia, several important exceptions also stand out. For instance, there is large disagreement with regard to South Africa: The Alesina et al. measure scores it as surprisingly equal (close to the global mean), whereas it is considered as highly unequal in the Cederman et al. data. In this case, the Cederman et al. data are probably closer to the widespread perception that socioeconomic group differences remain high in post-Apartheid South Africa. To take another example, Saudi Arabia emerges as highly equal in the Alesina et al. data, whereas it scores as highly unequal in the Cederman et al. data. Finally, Sweden is scored as relatively unequal in the Alesina et al. data, whereas the Cederman et al. measure score it as highly equal.[14] Overall, such large disagreements between country scores help to explain the low correlation between these measures (0.16).

In the Online Appendix, I graphically explore the non-correlated measures of Alesina et al. and Houle (Figure A13). In addition, in Table A7, I conduct a systematic comparison of the measures for of a number of countries where the nature of ethnic inequality is well-established (South Africa, Guatemala, Peru, Brazil, Nigeria, and Switzerland). The take-away from this exercise is that most measures agree only very roughly on the relative order of a country with significant variation and hard-to-explain exceptions.

Returning to the question of possible clustering, a principal component factor analysis[15] reveals two principal factors with eigenvalues above 1 (see Table 3). The first factor shows moderate-to-high loadings by all measures, suggesting that they tap into a common, latent phenomenon. This corresponds to the previously discussed conceptual logic, in which all dimensions reflect socioeconomic ethnic inequality. The second factor exhibits moderate loadings

[14] According to the EPR data underlying Cederman et al., there is only one politically relevant group in Sweden.

[15] I exclude Baldwin and Huber due to the low N = 46, which would exclude most observations. Moreover, I use an interpolated measure of Alesina et al. that fills the years between 1992, 2000 and 2012 to increase the number of observations from 49 to 415, thereby ensuring sufficient overlap to conduct the factor analysis.

by the Cederman et al. and Houle measures, to which there is no straightforward interpretation.[16] In line with the bivariate correlation analysis, the factor analysis reveals no clustering around an economic and social dimension, respectively.

To further probe my interpretations, I follow Adcock and Collier's (2001: 540) recommendation to assess correlations between the measures and measures of neighboring concepts (discriminant validation). This allows me to check whether the measures diverge from established measures of different yet related concepts. I have thus correlated the various measures with the interpersonal income Gini, interpersonal educational Gini, as well as two measures of ethnic fractionalization. The full analysis is provided in the Online Appendix (Table A3). Most measures behave largely as we would expect, being moderately correlated with the different neighboring concepts.

Meanwhile, the Houle measure demonstrates relatively low correlations with the neighboring concepts (mostly around 0.15–0.20). Strikingly, the Cederman et al. measure has very low and even negative correlations with the neighboring concepts. The low correlations of this measure with neighboring concepts could partly be explained by the ratio aggregation approach, which reflects the poorest (or richest) group in society relative to the mean, whereas the selected neighboring concepts capture aggregate distributions. Because the status of the poorest (or richest) groups in society does not necessarily correspond to the level of ethnic inequality based on the entire distribution of groups, we may see low correlations. In short, these findings further

[16] A partial explanation for the clustering is that both Cederman et al. and Houle use EPR data to classify ethnic groups and calculate their underlying group-level measures in the same fashion. However, they aggregate them differently to the country level.

underscore how the choice between ratio-based and aggregate measures (that represent the entire distribution) has important consequences.[17]

Do the differences matter?

To see whether the reported dissimilarities in conceptualization and measurement affect the findings of empirical analyses, I conduct replication analyses of four prominent studies published in highly recognized journals or book series (Alesina et al. 2016; Cederman et al. 2013; Houle 2015; Baldwin and Huber 2010). In each replication analysis, I have used the original datasets and Stata code, only substituting the measures of ethnic inequality, which have been standardized to ensure comparability.[18] To save space, I only report the main coefficients below, whereas the full regression tables, including controls, are available in the Online Appendix.[19]

Alesina et al. (2016: 454) find a negative and statistically significant cross-country association between ethnic inequality and economic development – measured as the log of per capita GDP in 2000. In Figure 5, I report the ordinary least squares (OLS) regressions, relating logged GDP per capita and the different measures of ethnic inequality. In these analyses, only the coefficients for the Alesina et al. and V-Dem measures are negative and statistically significant, whereas Omoeva et al. have the expected sign yet fail to reach statistical significance. Contrary to

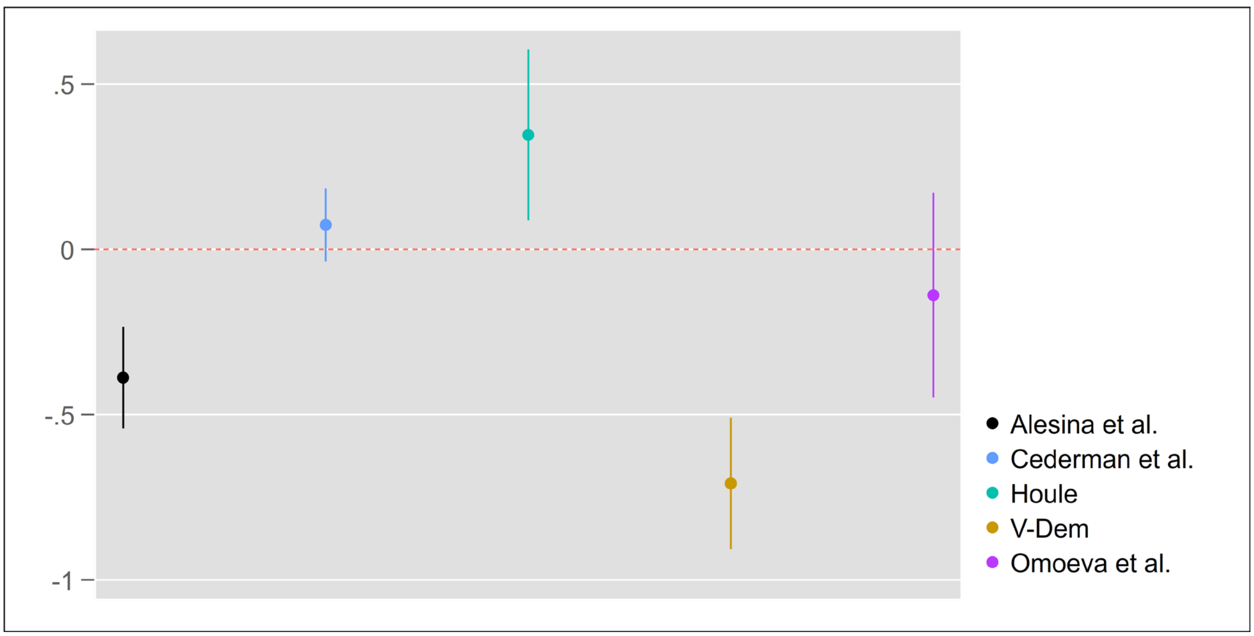
[17] In the Online Appendix, I further discuss how the observed empirical divergences is the product of choices at each of the following levels: 1) ethnic categories, (2) socioeconomic data, and (3) aggregation procedures.

[18] The samples are thus bounded by the original analysis' empirical scope.

[19] I only include the measure by Baldwin in Huber in Figure 8, because it is only available for one year per country between 1996–2005, yielding to few observations for the other replication analyses. Regressions underlying Figures 6-8 are based on an interpolated version of the Alesina et al. measure to provide sufficient observations.

expectations, the coefficients for the Houle and Cederman et al. measures are positive and the coefficient for Houle's measure is statistically significant. However, the results could partly be a product of sample differences in country and temporal coverage. I have thus run regressions based on the exact same sample of countries and years, reported in the Online Appendix Table A11. Overall, these results show that the differences in Figure 5 are not only a product of the different samples; they also reflect measurement differences. While the number of observation drops dramatically, all coefficients remain signed in the same direction, with the exception of the coefficient for Omoeva et al., which turns positive.

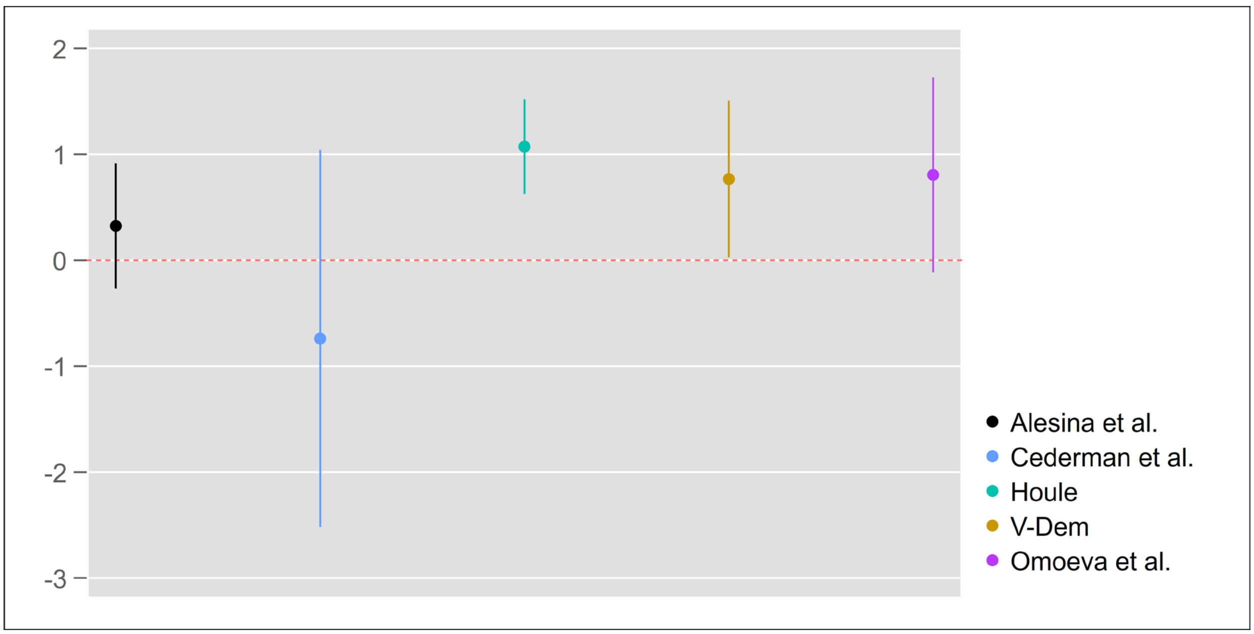
Figure 5: Replication of Alesina et al. 2016: Ethnic Inequality and GDP per capita



Notes: Coefficients for ethnic inequality measures from OLS regressions relating the different measures of ethnic inequality and logged GDP per capita in 2000 (See Table A10 for the underlying full regression table, including controls). 95% confidence interval. Based on Alesina et al. (2016), Table 2, Model 1.

Houle (2015) finds that ethnic inequality (BGI) is associated at the country-level with an increased risk of democratic breakdown, but only when levels of within-group inequality (WGI) are low. In Figure 6, I report the results from the probit estimations of ethnic inequality's association with democratic breakdown. Houle's hypothesis is supported if the coefficient of ethnic inequality is positive. This means that ethnic inequality increases the likelihood of democratic reversals when within-group inequality is zero (Houle, 2015: 491). The results from Figure 6 suggest that – in addition to Houle's own measure – the measures by V-Dem, Omoeva et al., and Alesina et al. show positive associations as expected, though the latter two are not statistically significant. In contrast, the measure by Cederman et al. is signed negatively and is very imprecisely estimated. Again, the result may be influence by differences in country and temporal coverage. Rerunning the analysis with a perfectly overlapping but smaller sample in Table A13, yields similar results, with all variables being signed in the same direction as before.

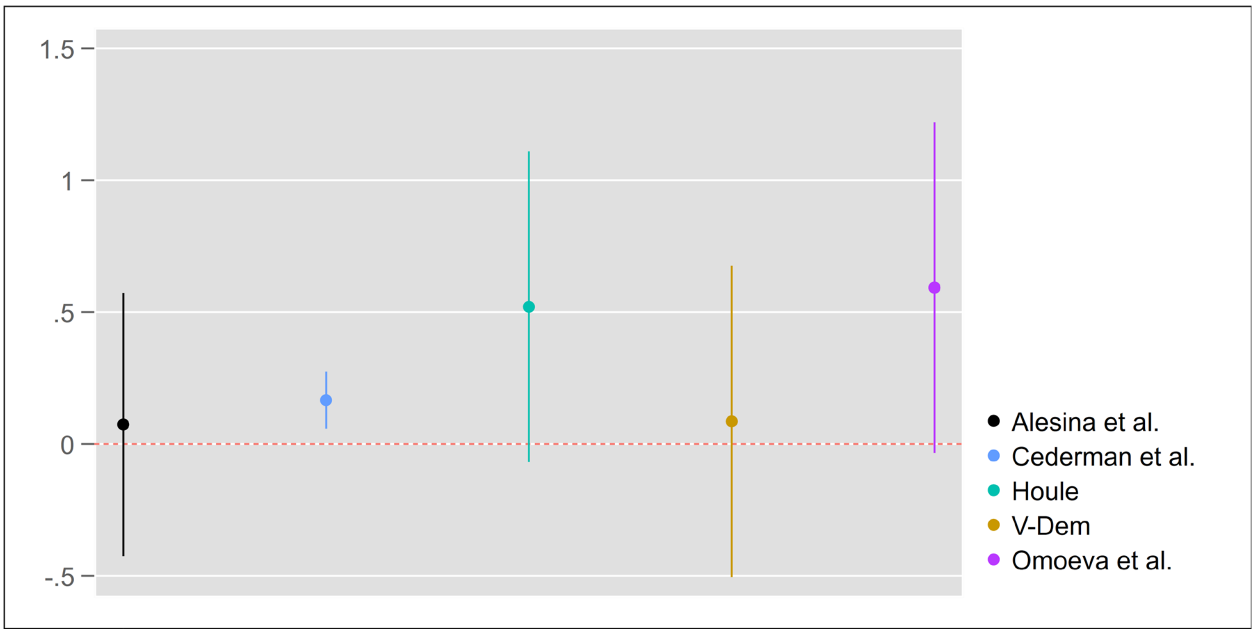
Figure 6: Replication of Houle 2015: Ethnic Inequality and Democratic Breakdown



Notes: Coefficients for ethnic inequality measures from probit regressions of relationship between various ethnic inequality measures and democratic breakdown. (See Table A12 for the underlying full regression table, including controls). 95% confidence interval. Based on Houle (2015), Table 2, Model 1.

Cederman et al. (2013) present country-level evidence that ethnic economic inequality is associated with the risk of civil war onset. Figure 7 shows a replication of the association between the examined ethnic inequality measures and civil conflict. Though all measures are signed in the expected direction, there are important differences. The Cederman et al. measure is estimated precisely, whereas the others are either very close to zero (Alesina et al.; V-Dem) or have very large confidence intervals (Houle, Omoeva et al.). Restricting the analysis to a smaller sample for which all measures have coverage yields somewhat similar results, with all coefficients keeping their original signs (Table A15).

Figure 7: Replication of Cederman et al. (2013): Ethnic Inequality and Civil War

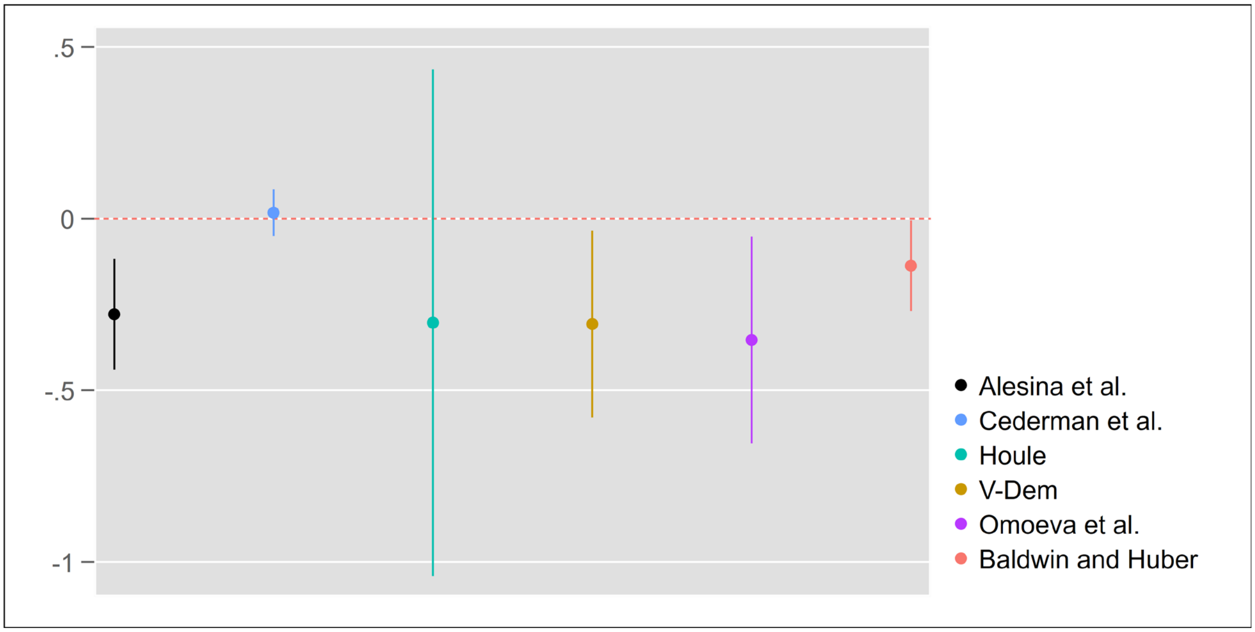


Notes: Coefficients for ethnic inequality measures from logit regressions of relationship between various ethnic inequality measures and civil war onset. (See Table A14 for the underlying full regression table, including controls). 95% confidence interval. Based on Cederman et al. (2013), chapter 7, Model 7.1.

Finally, Baldwin and Huber (2010) find that economic differences between groups are negatively associated with public goods provision. In Figure 8, I show that with the exception of Cederman et al., all measures are negatively associated with public goods provision, although Houle's measure has very large confidence intervals. Since there are only 13 observations for which all measures overlap, checking this replication analysis for sample influence is more difficult.

The findings suggest that the choice of measure has important implications for empirical analysis. The results were generally sensitive to the employed measure, indicating that the examined measures are not interchangeable.

Figure 8: Replication of Baldwin and Huber (2010): Ethnic Inequality and Public Goods Provision



Notes: Coefficients for ethnic inequality measures from logit regressions of relationship between various ethnic inequality measures and public goods provision. (See Table A16 for the underlying full regression table, including controls). 95% confidence interval. Based on Baldwin and Huber (2010), Table 5, Model 4.

Discussion

An overview of the most important strengths and weaknesses in the different datasets indicates that no measure offers a fully satisfactory response to all of the challenges of coverage, conceptualization, measurement, and aggregation (see Table 4). The array of options confronts researchers with a dilemma: Which measure is the most valid and reliable measure of ethnic inequality? First, the answer to this question should rest on theoretical foundations regarding a particular research question. If one is interested in mobilization patterns among severely deprived groups, the theoretical arguments would point toward measures such as that of Cederman et al. (2013), which capture this type of socioeconomic disparity. If one is interested in the causes and consequences of the entire distribution of resources between ethnic groups, the other examined measures are likely to be more appropriate. Second, considering the examined strengths and weaknesses in Table 4, the Alesina et al. (2016) and Coppedge et al. (2021a) measures appear superior in terms of capturing the overall distribution of resources while providing high empirical coverage. That said, researchers considering using one of the measures should still closely study the precise concept and measurement techniques in order to be conscious of biases and errors.

Table 4: Summary of Strengths and Weaknesses of Extant Datasets

Data provider and Index	Strengths	Weaknesses
Alesina et al.: - Ethnic Gini	- Comprehensive spatial scope - Clear, detailed description of measurement and aggregation	- Somewhat restricted temporal scope - Builds on indirect economic proxy
Cederman et al.: - G-Econ/Ethnic homeland	- Comprehensive spatial scope - Detailed conceptual discussion - Clear, detailed description of measurement and aggregation	- Restricted temporal scope (time-invariant) - Builds on crude economic measure
Houle: - Between-group income inequality	- Clear, detailed description of measurement - Face validation of measure	- Restricted + biased empirical scope: only covers democracies - Restricted temporal variation - Aggregation procedure is not justified - Potential survey biases
Baldwin & Huber: - Between-group income inequality	- Thorough validation procedures; Face validity of scores - Clear, detailed measurement discussion	- Severely restricted + biased spatial and temporal scope: only covers 46 democracies - Potential survey biases
Omoeva et al.: - Educational Group Gini	- Relatively comprehensive empirical scope - Multiple, plausible aggregation techniques	- Underrepresentation of developed countries - Exclusive focus on education - Limited conceptual discussion
V-Dem Coppedge et al.: - Access to public services by social group	- Comprehensive empirical scope - Sophisticated aggregation procedure, incl. reliability test - Uncertainty estimates	- Difficult to assess basis of coding decisions - Potential biases in expert coding - Limited conceptual discussion

Although this article has focused on highly aggregated country-level measures, more disaggregated research designs are possible, and have indeed been applied to some of the examined datasets. Cederman et al. (2013) and Houle (2015) present their country-level analyses together with group-level analyses, finding that groups with wealth levels far from the country mean groups are more likely to experience civil war or initiate democratic breakdown, respectively. In the same vein, group-level measures may also help track country-level developments, as illustrated by Bormann et al. (2021) who use nighttime luminosity data from 1992 to 2012 and a global sample of ethnic groups to show how the gap between politically marginalized groups and their included counterparts has narrowed over time.[20]

To the extent that researchers are only interested in two groups – or clusters of groups (e.g. politically included/excluded) – ratios of average achievement of relevant groups constitute a straightforward and intuitive measure of inequality. That said, more aggregate measures are clearly needed if there are larger number of groups and we are interested in a single figure representing the entire distribution (Stewart et al. 2010: 16). Beyond the benefits of including an additional level of analysis, more fine-grained group-level data also hold the promise of more transparency, as it becomes possible to validate the scores for individual groups (see, e.g., Houle 2015, 488-89). Even when presenting highly aggregate country-level measures, data providers should ideally also make public the underlying group-level values that were used to calculate the aggregate measures. This was found to be a clear limitation with the V-Dem data. Since questions involving ethnic inequality usually have clear group-level implications, it is often advisable to supplement country-

[20] Other examples of group-level measures of socioeconomic ethnic inequality include the All Minorities at Risk (AMAR) dataset, which includes the socioeconomic position of a group (Birnir et al. 2017). Moreover, in addition to Houle (2015), Huber and Mayoral (2019) and Kuhn and Weidmann (2015) offer *within-group* inequality measures.

level with group-level analyses. Not least given the discussed measurement and aggregation challenges, additional disaggregated analyses constitute one way to increase our confidence in any findings involving the examined ethnic inequality data.

Another encouraging development led by Cederman et al. (2015) is the introduction of a group-level composite indicator combining the strengths of three different sources of data on local wealth: the G-Econ data, survey data, and nightlight emissions combined with geographical data on the settlement of ethnic groups. They weigh economic data more heavily in countries where official statistics are more trustworthy, and nightlight data more heavily where government statistics are poor or lacking. This triangulated measure has not been included in the main discussion and analysis, as it is not publicly available at the country level.[21] It nevertheless deserves mentioning, since such efforts to overcome the respective weaknesses in the different data sources provide a promising avenue toward more valid and reliable measures of ethnic inequality. This avenue is particularly promising if such measures could be made available for longer time periods. Although this is likely to entail further data collection and to be resource-intensive, it would allow researchers to investigate a range of new and important questions. Finally, providing triangulated measures with different aggregation procedures is crucial if such measures are to be used for broader research purposes.

An additional way forward is to combine various existing cross-national measures into a composite index. This approach relies on the reasonable assumption that socioeconomic ethnic

[21] To see how this measure empirically relates to others, I have transformed it into a cross-national indicator following Cederman et al.'s suggested approach (2013: 150). While the measure exhibits slightly stronger correlations with all of the other evaluated measures than their G-Econ measure, the correlations remain relatively low, i.e. 0.11–0.32 (Table A8).

inequality is imperfectly but more or less accurately observed by the compilers of various existing datasets and that each of them taps into a common dimension. This allows researchers to leverage the enormous effort that scholars have invested in creating an ethnic inequality measure, and it provides a way of dealing with considerable measurement error. Combining measures – based on explicit conceptual foundations – should thus help improve measurement accuracy and minimize the impact of idiosyncratic error associated with particular estimates (see Munck et al. 2020: 345).

In the Online Appendix, I demonstrate this approach with an illustrative example. The resulting index provides plausible values for most countries, and when running the replication analyses with the index, it yields results in line with the original studies in three out of four cases. Since this index is only a relatively crude illustration, the approach should be further exploited using more sophisticated methods, such as latent variable models and item response theory (IRT), which have been employed for other concepts that are impossible or difficult to observe directly (e.g. Fariss 2014; Melton et al. 2015; Solis and Waggoner 2020).

Conclusion

The literature on ethnic (or horizontal) inequalities has made a series of important contributions to political science. This research has relied on new datasets compiled by scholars creatively exploiting a range of different data sources. This paper has compared extant measures, which have been used to operationalize economic and social inequality between ethnic groups at the country level. The assessment has found that measures differ in important ways. Differences in conceptualization and measurement are clearly reflected in the fact that several of the indicators do not correlate highly with each other. Indeed, many of the correlations were surprisingly weak (or even negative). Four replication analyses suggested that the choice of indicator seriously affects our empirical analyses and that the results may depend strongly on the employed indicator. As such, extant measures of ethnic inequality are generally not interchangeable.

Future research can benefit in three ways from the clarifications and critical points put forward in this assessment, which offers helpful information to data users. First, systematic information about the different strengths and weaknesses of various measures of ethnic inequality can help future data users to make conscious choices regarding what measures to use and how. Second, the results suggest that it might be worthwhile to reexamine many of the previous studies using the evaluated measures. Third, the findings can inform the development of new measures that either rely on novel data collection or combine existing indicators in new ways.

Supplementary Material. The Online Appendix is available at: [doi.org link to Cambridge / Journal site once ready]

Data availability Statement. Replication data for this article can be found in Harvard Dataverse at: https://doi.org/10.7910/DVN/6LJEGI

Acknowledgements. I am grateful to Svend-Erik Skaaning, Kees van Kersbergen, Gerardo L. Munck, Kristian Skrede Gleditsch, Christian Houle, Kyle L. Marquardt, Jonathan Doucette, Nicholas Haas, Jacob Nyrup, and Lars Johannsen as well as the three anonymous reviewers and the editor for highly constructive comments and suggestions.

Financial Support. None.

Competing Interests. None.

References

Adcock R. and Collier D. (2001) Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *The American political science review,* 95**,** 529-546.

Alesina A., Michalopoulos S. and Papaioannou E. (2016) Ethnic Inequality. *Journal of Political Economy,* 124**,** 428-488.

Baghat K., Barrett G., Dupuy K., Gates S., Hillesund S., Nygård H.M., Rustad S.A., Strand H., Urdal H. and Østby G. (2017) Inequality and Armed Conflict: Evidence and Data. Background Report for the Un and World Bank Flagship Study on Development and Conflict Prevention. PRIO: Peace Research Institute Oslo.

Baldwin K. and Huber J.D. (2010) Economic Versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision. *American Political Science Review,* 104**,** 644-662.

Birnir J.K., Laitin D.D., Wilkenfeld J., Waguespack D.M., Hultquist A.S. and Gurr T.R. (2017) Introducing the Amar (All Minorities at Risk) Data. *Journal of Conflict Resolution,* 62**,** 203-226.

Bochsler D., Green E., Jenne E., Mylonas H. and Wimmer A. (2021) Exchange on the Quantitative Measurement of Ethnic and National Identity. *Nations and Nationalism,* 27**,** 22–40.

Bormann N.-C., Pengl Y.I., Cederman L.-E. and Weidmann N.B. (2021) Globalization, Institutions, and Ethnic Inequality. *International Organization,* 75**,** 665-697.

Buhaug H., Cederman L.E. and Gleditsch K.S. (2014) Square Pegs in Round Holes: Inequalities, Grievances, and Civil War. *International Studies Quarterly,* 58**,** 418-431.

Canelas C. and Gisselquist R.M. (2018) Horizontal Inequality as an Outcome. *Oxford Development Studies,* 46**,** 305-324.

Canelas C. and Gisselquist R.M. (2019) Horizontal Inequality and Data Challenges. *Social Indicators Research,* 143**,** 157-172.

Cederman L.-E., Gleditsch K.S. and Buhaug H. (2013) *Inequality, Grievances, and Civil War,* Cambridge, Cambridge University Press.

Cederman L.-E., Weidmann N.B. and Bormann N.-C. (2015) Triangulating Horizontal Inequality: Toward Improved Conflict Analysis. *Journal of Peace Research,* 52**,** 806-821.

Chandra K. (2006) What Is Ethnic Identity and Does It Matter? *Annual Review of Political Science,* 9**,** 397-424.

Chen X. and Nordhaus W.D. (2011) Using Luminosity Data as a Proxy for Economic Statistics. *Proceedings of the National Academy of Sciences - PNAS,* 108**,** 8589-8594.

Coppedge M., Gerring J., Knutsen C.H., Lindberg S.I., Teorell J., Altman D., Bernhard M., Cornell A., Fish M.S., Gastaldi L., Gjerløw H., Glynn A., Hicken A., Lührmann A., Maerz S.F., Marquardt K.L., Mcmann K., Mechkova V., Paxton P., Pemstein D., Römer J.V., Seim B., Sigman R., Skaaning S.-E., Staton J., Sundtröm A., Tzelgov E., Uberti L., Wang Y.-T., Wig T. and Ziblatt D. (2021a) "V-Dem Codebook V11.1" Varieties of Democracy (V-Dem) Project.

Coppedge M., Gerring J., Knutsen C.H., Lindberg S.I., Teorell J., Marquardt K.L., Medzihorsky J., Pemstein D., Alizada N., Gastaldi L., Hindle G., Pernes J., Römer J.V., Tzelgov E., Wang Y.-T. and Wilson. S. (2021b) ”V-Dem Methodology V11.1” Varieties of Democracy (V-Dem) Project.

Deere C.D., Kanbur R. and Stewart F. (2018) Chapter 4. Horizontal Inequalities. *In:* Stiglitz Joseph E., Fitoussi Jean-Paul and Durand Martine (eds.) *For Good Measure: Advancing Research on Well-Being Metrics Beyond Gdp.* Paris: OECD.

Eic (2015) Education Inequalities and Conflict Database. Washington, Dc: Education Policy and Data Center.

Fariss C.J. (2014) Respect for Human Rights Has Improved over Time: Modeling the Changing Standard of Accountability. *The American political science review,* 108**,** 297-318.

Fearon J.D. (2003) Ethnic and Cultural Diversity by Country. *Journal of economic growth (Boston, Mass.),* 8**,** 195-222.

Fleming C.M., Manning M., Pham H.-T. and Vorsina M. (2020) Ethnic Economic Inequality and Fatalities from Terrorism. *Journal of interpersonal violence***,** Online ahead of print.

Gordon R.G. (2005) *Ethnologue: Languages of the World. 15th Ed.*, Dallas: SIL Internat.

Horowitz D.L. (2000) *Ethnic Groups in Conflict,* Berkeley, CA, University of California Press.

Houle C. (2015) Ethnic Inequality and the Dismantling of Democracy: A Global Analysis. *World Politics,* 67**,** 469-505.

Houle C. and Bodea C. (2017) Ethnic Inequality and Coups in Sub-Saharan Africa. *Journal of Peace Research,* 54**,** 382-396.

Huber J.D. and Mayoral L. (2019) Group Inequality and the Severity of Civil Conflict. *Journal of Economic Growth,* 24**,** 1-41.

Jensen C. and Van Kersbergen K. (2016) *The Politics of Inequality,* London, Palgrave Macmillan Education.

Kuhn P.M. and Weidmann N.B. (2015) Unequal We Fight: Between- and within-Group Inequality and Ethnic Civil War. *Political Science Research and Methods,* 3**,** 543-568.

Leipziger L. (2022) Replication Data For: Measuring Ethnic Inequality: An Assessment of Extant Cross-National Indices, doi:10.7910/DVN/6LJEGI, Harvard Dataverse, V1, UNF:6:IwftBbyc0ulyg3MhZ40mdg== [fileUNF]

Marquez X. (2016) A Quick Method for Extending the Unified Democracy Scores. *Social Science Research Network Working Paper Series*.

Mellander C., Lobo J., Stolarick K. and Matheson Z. (2015) Night-Time Light Data: A Good Proxy Measure for Economic Activity? *PloS one,* 10**,** e0139779-e0139779.

Melton J., Meserve S. and Pemstein D. (2015) Unifid Democracy Scores (Uds), hdl:10622/Y9G2ZV, IISH Data Collection, V1

Munck G.L., Møller J. and Skaaning S.-E. (2020) Conceptualization and Measurement: Basic Distinctions and Guidelines. *In:* Curini Luigi and Franzese Robert (eds.) *The Sage Handbook of Research Methods in Political Science and Ir.* London: Sage Publications.

Munck G.L. and Verkuilen J. (2002) Conceptualizing and Measuring Democracy: Evaluating Alternative Indices. *Comparative political studies,* 35**,** 5-34.

Nordhaus W., Azam Q., Corderi D., Hood K., Victor N.M., Mohammed M., Miltner A. and Weiss J. (2006) The G-Econ Database on Gridded Output: Methods and Data. *Unpublished*.

Omoeva C., Moussa W. and Hatch R. (2018) The Effects of Armed Conflict on Educational Attainment and Inequality. *EPDC Research Paper No. 18-03*.

Pemstein D., Marquardt K.L., Tzelgov E., Wang Y., Medzihorsky J., Krusell J., Miri F. and Von Römer J. (2019) The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data. *V-Dem Working Paper Series 2019(21).*

Piketty T. (2014) *Capital in the Twenty-First Century,* Cambridge, MA, Harvard University Press.

Rios-Figueroa J. and Staton J.K. (2012) An Evaluation of Cross-National Measures of Judicial Independence. *Journal of Law, Economics, & Organization,* 30**,** 104-137.

Skaaning S.-E. (2018) Different Types of Data and the Validity of Democracy Measures. *Politics and Governance,* 6**,** 105-116.

Solis J.A. and Waggoner P.D. (2020) Measuring Media Freedom: An Item Response Theory Analysis of Existing Indicators. *British journal of political science,* **,** 1-20.

Stewart F. (2002) Horizontal Inequalities: A Neglected Dimension of Development. *Queen Elizabeth House Working Paper Series,* Working Paper Number 81.

Stewart F. (2008) *Horizontal Inequalities and Conflict : Understanding Group Violence in Multiethnic Societies,* Basingstoke, Palgrave Macmillan.

Stewart F., Brown G.K. and Mancini L. (2010) Monitoring and Measuring Horizontal Inequalities. *Centre for Research on Inequality, Human Security and Ethnicity,* 4.

Tetteh-Baah S.K. (2019) Measurement and Impact of Horizontal Inequality. *Doctoral Thesis, Swiss Federal Institute of Technology*.

Un. (2020) *Un Sustainable Development Goal 10: Reduce Inequality within and among Countries. Https://Sdgs.Un.Org/Goals/Goal10* [Online]. [Accessed 15-10-2020 2020].

Vogt M., Bormann N.-C., Rüegger S., Cederman L.-E., Hunziker P. and Girardin L. (2015) Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family. *Journal of Conflict Resolution,* 59**,** 1327-1342.

Wang Y.-T. and Kolev K. (2019) Ethnic Group Inequality, Partisan Networks, and Political Clientelism. *Political research quarterly,* 72**,** 329-341.

Weber M. (1976 [1922]) *Wirtschaft Und Gesellschaft : Grundriss Der Verstehenden Soziologie,* Tübingen, J. B. C. Mohr.

Weidmann N.B., Rød J.K. and Cederman L.-E. (2010) Representing Ethnic Groups in Space: A New Dataset. *Journal of Peace Research,* 47**,** 491-499.

Wucherpfennig J., Weidmann N.B., Girardin L., Cederman L.-E. and Wimmer A. (2011) Politically Relevant Ethnic Groups across Space and Time: Introducing the Geoepr Dataset1. *Conflict management and peace science,* 28**,** 423-437.

Ye F. and Han S.M. (2019) Does Ethnic Inequality Increase State Repression? *Canadian Journal of Political Science,* 52**,** 883-901.