Online Appendix for

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

Contents

Maps of country coverage for each measure

The following figures (A1-A6) show, which countries are covered by the respective dataset with at least one observation. Black shades indicate that the country is covered the dataset, whereas light grey indicates missing values.

Figure A1: Alesina et al.

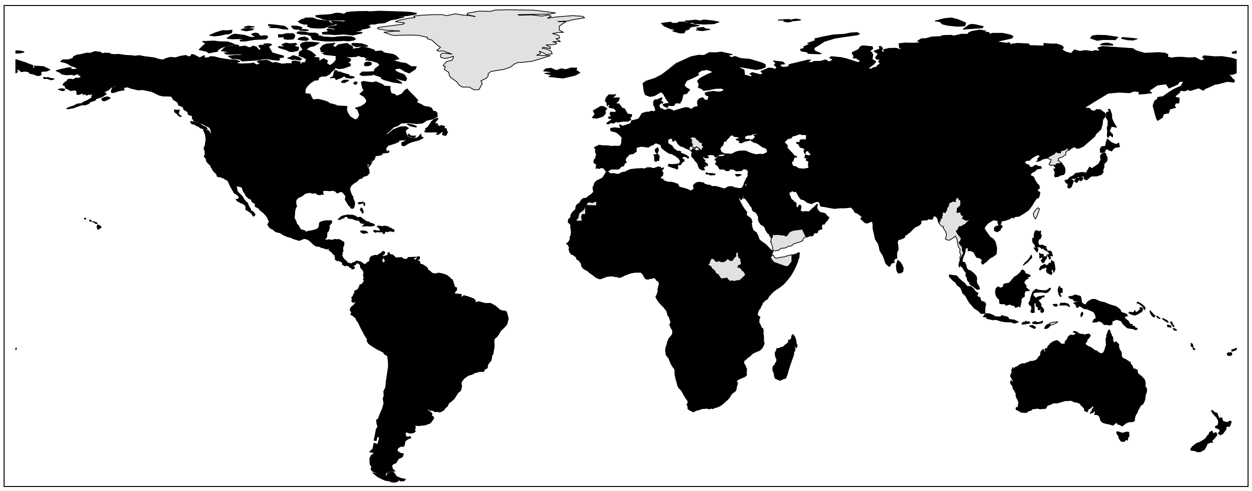


Figure A2: Cederman et al.

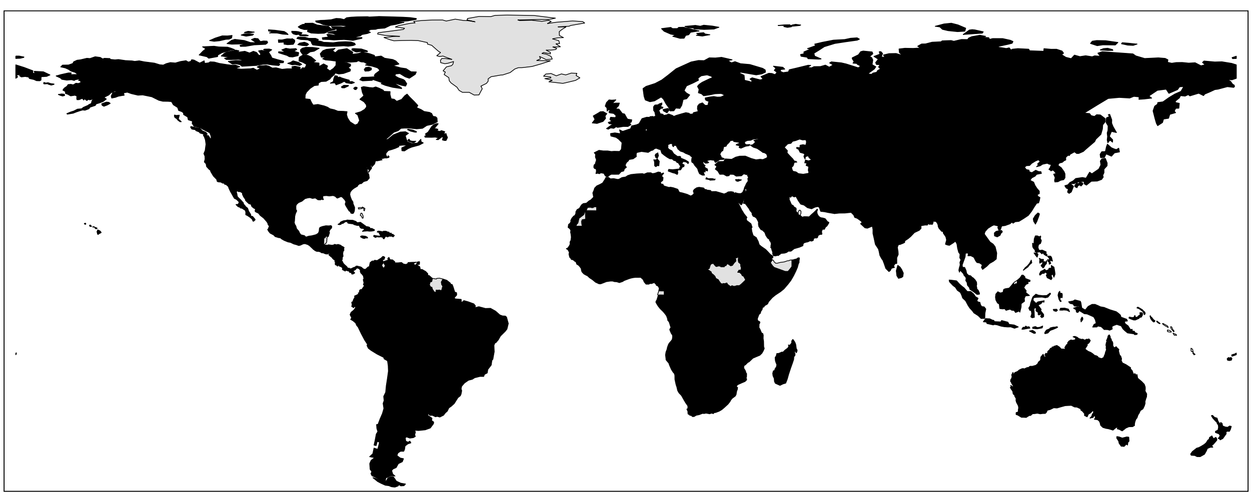


Figure A3: Baldwin and Huber

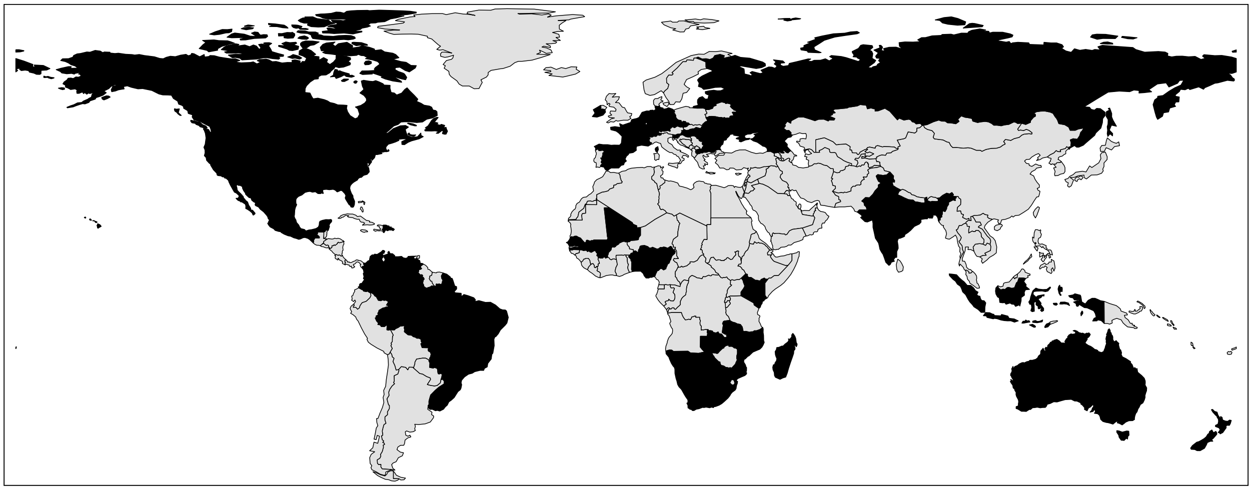


Figure A4: Houle

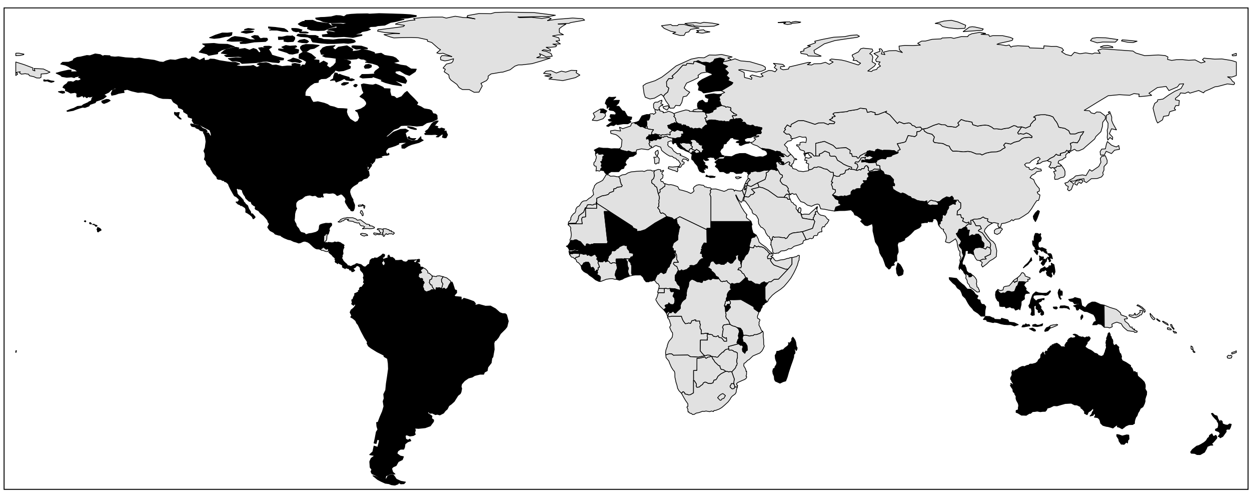


Figure A5: Coppedge et al. (V-Dem)

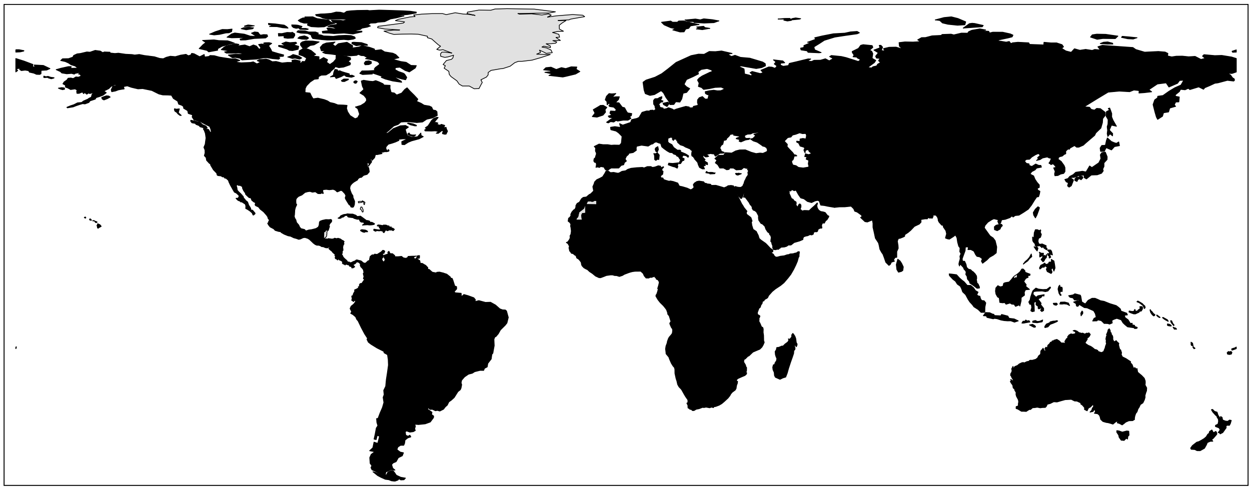
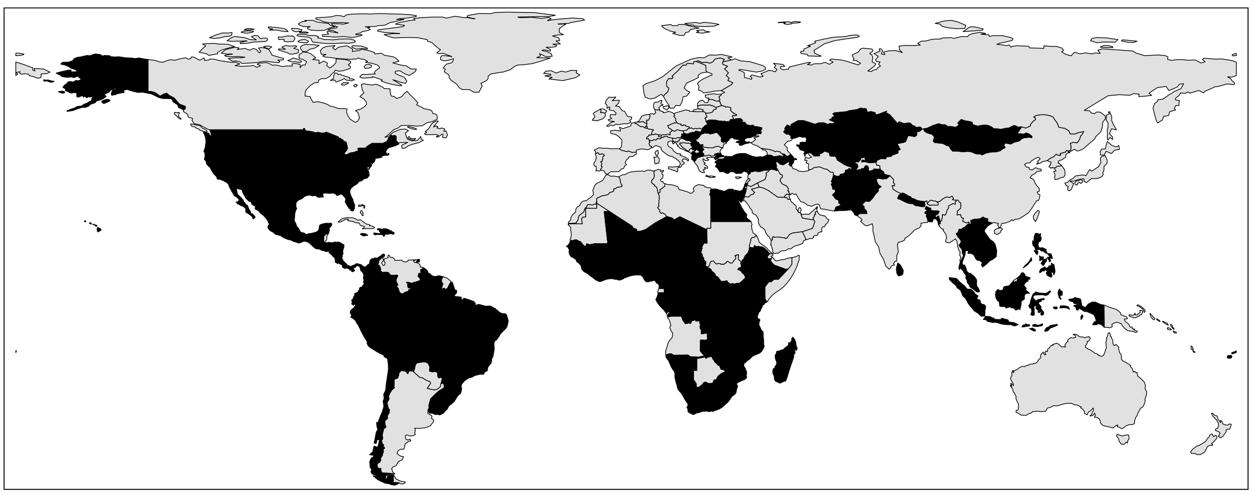


Figure A6: Omoeva et al.



Temporal variation

Table A1 reports the between- and within country standard deviation of the examined measures. While all measures vary significantly more between countries that within, there are considerable differences in within-country variation. The Omoeva and V-Dem measure exhibit relatively high within-country-variation, whereas Houle, Cederman et al. and Alesina et al.'s measures exhibit lower within-country variation.

Table A1: Between and within-country variation

Data provider	Std. Dev. (between)	Std. Dev. (within)
Alesina et al. (2016)	.99	.15
Cederman et al. (2013)	.97	.22
Houle (2015)	.95	.31
Baldwin & Huber (2010)	1	0
V-Dem / Coppedge et al. (2021)	.86	.53
Omoeva et al. (2018)	.87	.55

Note: All variables have been z-transformed (mean of 0, overall standard deviation of 1) to ease interpretation and comparison. The Cederman et al. measures is not presented as time-varying in nature by the authors. However, there is temporal variation in some countries, e.g. South Africa in 1994. For reference, a z-transformed version of V-Dems polyarchy measures has a significantly higher within-country variation (0.67).

Non-random patterns of missing data

To further explore systematic patterns of missing data, I conduct a simple test of non-random missingness (see Rios-Figueroa and Staton 2012: 125). I consider whether missingness can be explained by a state's level of economic development, its level of democracy, state capacity or land area. The test checks whether there is a statistically significant difference between the mean scores across the subsamples for which we have data, and the subsample for which we have missing values. The included variables are GDP/cap. (Inklaar et al. 2018), a dichotomous measure of electoral democracy (Boix et al. 2013), V-Dem's measure of rigorous and impartial public administration as a proxy for state capacity (Coppedge et al. 2021: 175) and land area (Haber and Menaldo 2011; Weidmann et al. 2010).

Table A2: Simple non-random patterns in missing values in ethnic inequality measures (2000)

	Alesina et al.	Cederman et al.	Houle	Baldwin & Huber	Omoeva et al.
GDP/cap (ln)	-0.05 (8.70-8.75)	0.73* (9.41-8.68)	-0.35* (8.63-8.98)	-0.23 (8.79-9.01)	1.45** (9.40-7.95)
Democracy (dichotomous)	-0.03 (0.50-0.53)	0.04 (0.56-.53)	-0.57** (0.34-.91)	-0.34** (0.53-.87)	0.08 (0.56-0.49)
State capacity: rigorous and impartial adm. (0-1)	-0.06 (0.49-0.55)	0.03 (0.57-0.54)	-0.08** (0.52-0.61)	-0.10** (0.55-0.64)	0.09** (0.59-0.49)
Land area (ln)	-2.41** (9.61- 12.02)	-4.28** (8.03- 12.31)	-0.64* (11.70- 12.34)	-0.87** (11.88-12.75)	-0.61* (11.64- 12.24)

Note: Entries are average differences (subtracted means in parentheses) between countries that are covered and missing, respectively by the ethnic inequality measures. *p<0.1, **p<0.01 (two-tailed test). The measure by Baldwin and Huber is included for the period 1996-2006 due to unbalanced year-coverage. V-Dem is not included in the analysis as it covers all countries, except for a range of micro states. Numbers rounded up separately from Stata output.

The findings in Table A2 show that the measures provide samples that are not representative regarding GDP/cap, democracy, state capacity and country size. Notice also that the direction of the bias as expressed by the coefficients differs. For instance, countries covered by Cederman et al. and Omoeva et al. are relatively poorer and have lower state capacity, whereas countries covered by Houle and Baldwin & Huber are significantly more democratic and demonstrate higher state capacity. All datasets are skewed toward countries with relatively large land areas. Looking across the tests, Alesina et al.'s measure appears to be least afflicted by non-random missingness, which is expectable given their high country-coverage. However, to some extent, the missingness bias affects all the datasets, even the ones which approximate universal coverage (i.e., V-Dem which does not cover all micro states).

Overall, such non-random missingness obviously reduces the ability to infer from the sample to the general population of all countries. Scholars should either abstain from general inferences or, alternatively, justify why it is possible to infer to missing countries. Moreover, the missing data problem and the consequent lack of interchangeability means we should avoid being overly confident about any robustness analysis using alternative measures (see Rios-Figueroa and Staton 2012: 125-26).

Additional Correlation Analyses

Because there are only 13 country-years for which all measures have observations, I report the results from a series of alternative approaches. First, I report the correlations if the sample is restricted to the year 2000 (Table A3) or the years 1995 to 2005 (Table A4), in which the overlap is stronger. Second, I report correlations from a sample for which at least four of the six measures offer observations (Table A5). Finally, Table A6 reports the results from a group of countries, which have a high numbers of observations across measures. While there are a number of changes in individual correlations, these go in both directions, i.e. producing both higher and lower correlations. (Non-trivial changes compared to the main results are marked with bold). In short, this exercise confirms the overall pattern of surprisingly low correlations between most measures.

Table A3: Correlations with sample restricted to the year 2000

	Alesina et al.	Cederman et al.	Houle	Baldwin & Huber	Coppedge et al.	Omeova et al.
Cederman et al. (2013)	0.16 (149)					
Houle (2015)	-0.02 (55)	**0.03 (56)**				
Baldwin & Huber (2010)	n/a	n/a	n/a	n/a		
Coppedge et al. (2021): V-Dem	0.56 (162)	0.07 (154)	**0.07 (56)**	n/a		
Omoeva et al. (2018)	0.46 (74)	-0.05 (74)	**0.61 (27)**	n/a	0.18 (77)	

Note: results refer to bivariate Pearson's r correlations (n in parentheses)

Table A4: Correlations with sample restricted to the years 1995-2005

	Alesina et al.	Cederman et al.	Houle	Baldwin & Huber	Coppedge et al.	Omeova et al.
Cederman et al. (2013)	0.16 (149)					
Houle (2015)	-0.02 (55)	0.13 (625)				
Baldwin & Huber (2010)	N/A	0.04 (46)	0.01 (30)			
Coppedge et al. (2021): V-Dem	0.56 (162)	0.06 (1697)	**0.10 (625)**	0.64 (46)		
Omoeva et al. (2018)	0.46 (74)	-0.05 (796)	0.27 (294)	0.05 (21)	0.19 (826)	

Table A5: Correlations based on sample where at least four measures have coverage (up to 625 obs)

	Alesina et al.	Cederman et al.	Houle	Baldwin & Huber	Coppedge et al.	Omeova et al.
Cederman et al. (2013)	0.10 (202)					
Houle (2015)	0.01 (102)	**0.35 (517)**				
Baldwin & Huber (2010)	n/a	**0.31 (38)**	0.01 (30)			
Coppedge et al. (2021): V-Dem	0.55 (202)	0.07 (625)	**0.12 (517)**	0.61 (38)		
Omoeva et al. (2018)	0.41 (149)	-0.01 (555)	**0.18 (447)**	0.05 (21)	0.14 (555)	

Note: There are only 13 country-years for which all measures are available. Correlations based on sample where at least four measures have coverage thus constitutes a pragmatic attempt to approach an overlapping sample without losing too many observations.

Table A6: Correlations based on sample with group of countries with high coverage across measures

	Alesina et al.	Cederman et al.	Houle	Baldwin & Huber	Coppedge et al.	Omeova et al.
Cederman et al. (2013)	0.10 (48)					
Houle (2015)	**-0.41 (35)**	**0.27 (342)**				
Baldwin & Huber (2010)	n/a	0.01 (9)	**0.16 (7)**			
Coppedge et al. (2021): V-Dem	0.46 (72)	0.04 (480)	0.33 (601)	0.45 (9)		
Omoeva et al. (2018)	**0.24 (45)**	**-0.23 (383)**	**0.15 (520)**	**-0.38 (9)**	0.20 (1265)	

Note: countries included are the United States, Mexico, Ghana, Colombia, Brazil, Bolivia, Honduras, Mali, Peru, Senegal, Kenya, Philippines, Thailand, Benin, Nepal, Nicaragua, Malawi, Panama, Turkey, Madagascar, Sri Lanka and Trinidad and Tobago. The countries where in the top 10 percentiles in terms of cumulated country-year observations across measures.

Maps of standardized values for each index

The following maps (Figure A8-A13) show the standardized values for the different measures in the year 2000 to provide a sense of the scores assigned to various countries. It also gives a sense of the differences in empirical scope for a year, in which coverage is relatively high across measures.

Figure A7: Alesina et al. in 2000 (also appears in main text)

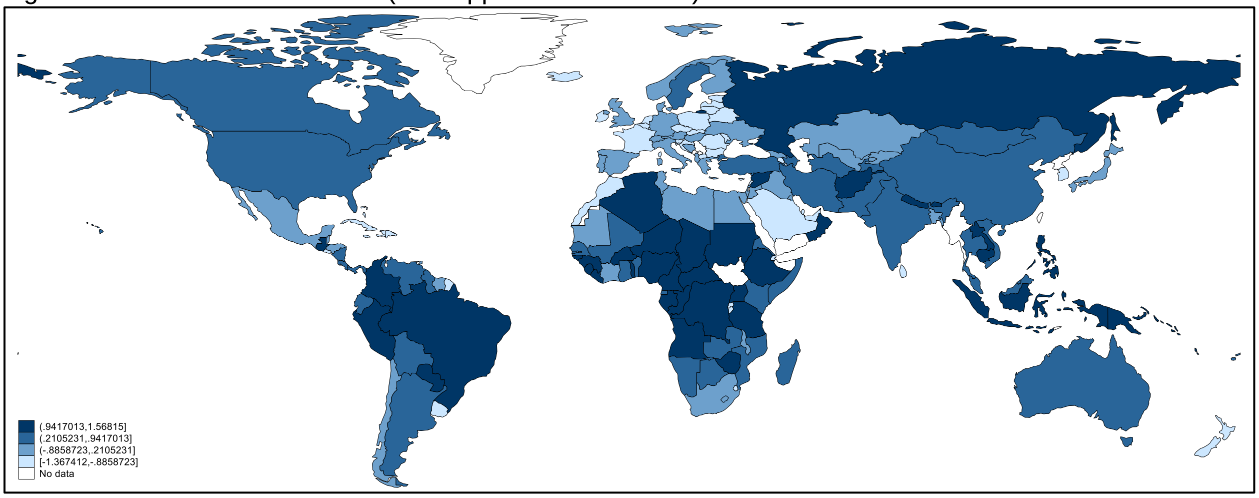


Figure A8: Cederman et al. in 2000 (also appears in main text)

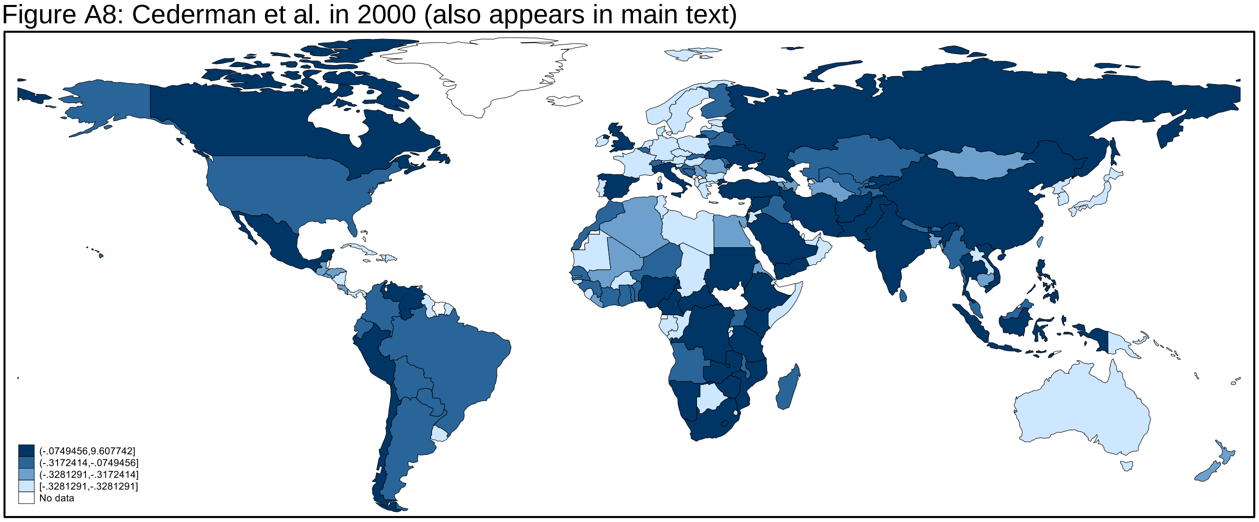


Figure A9: Houle in 2000

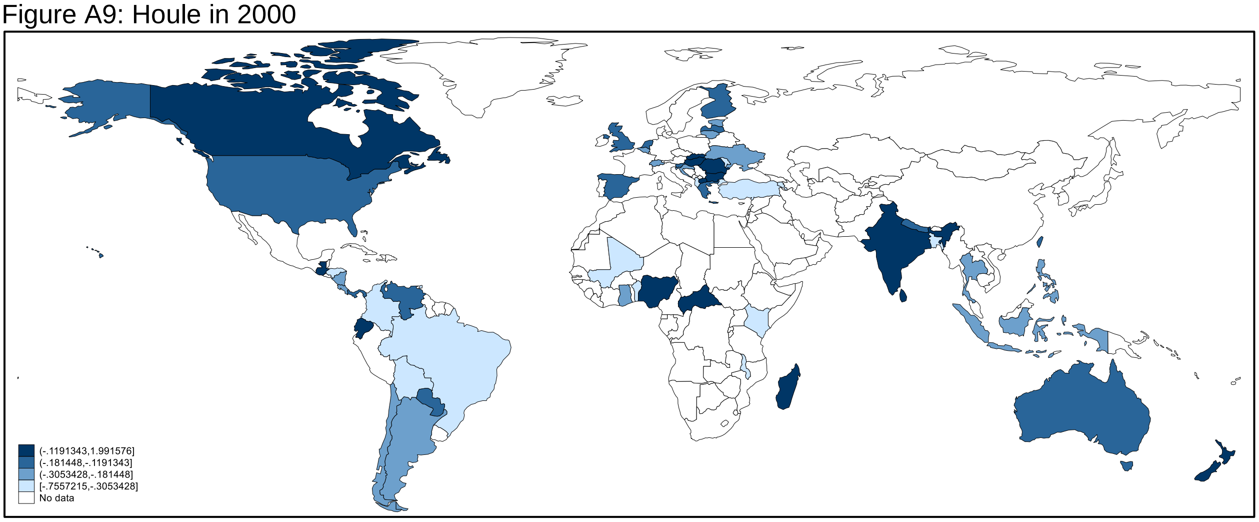


Figure A10: Baldwin and Huber in 2000 (values for available year 1996-2005)

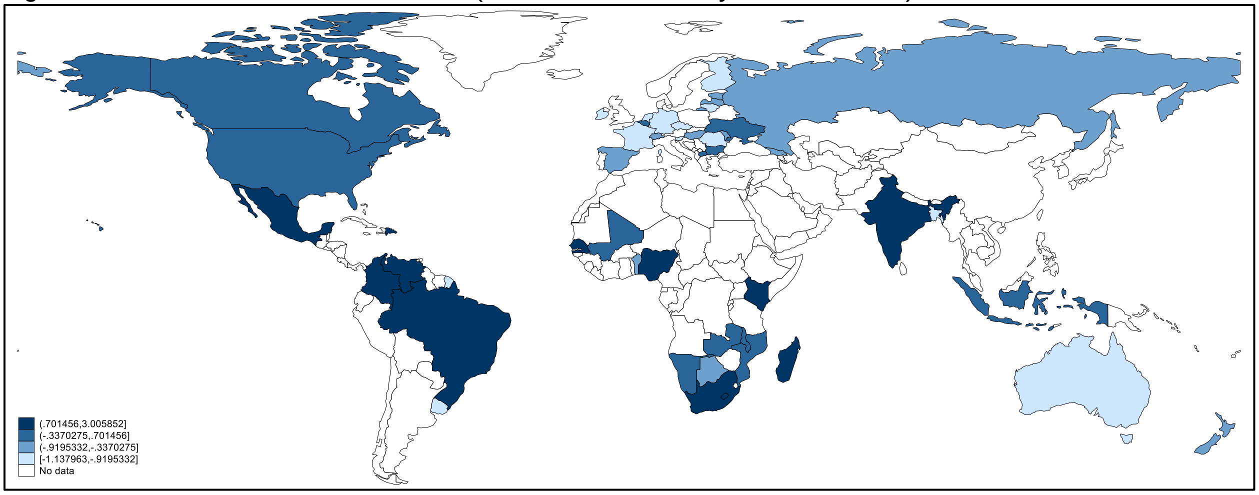


Figure A11: V-Dem in 2000

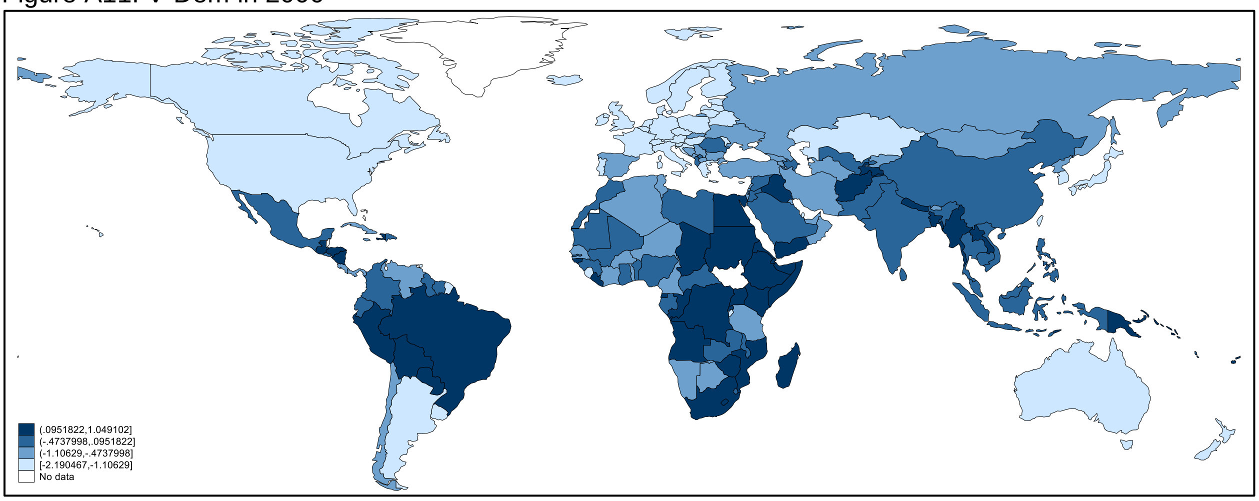
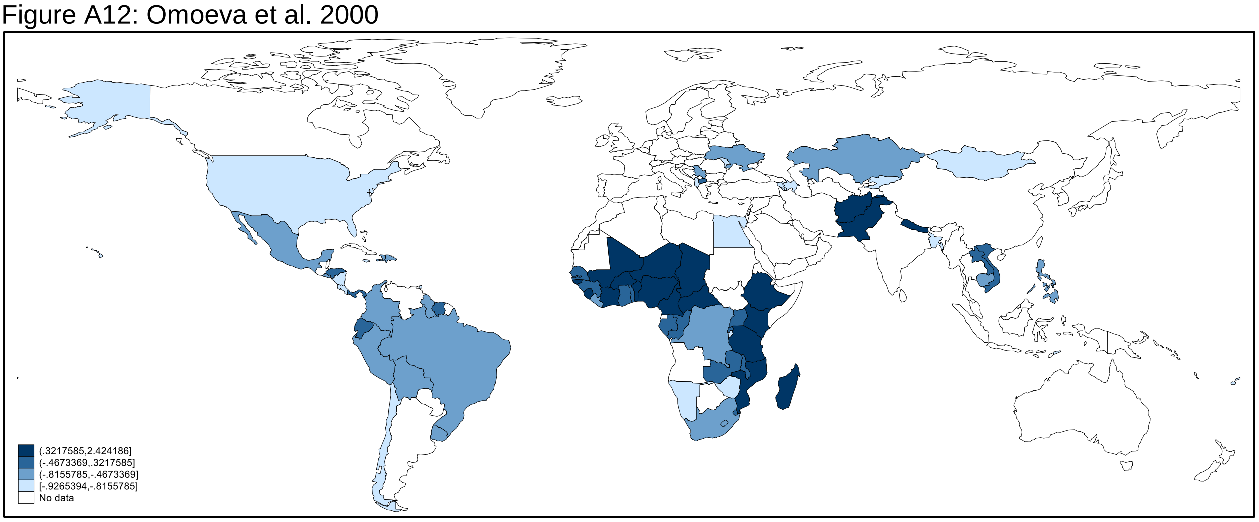


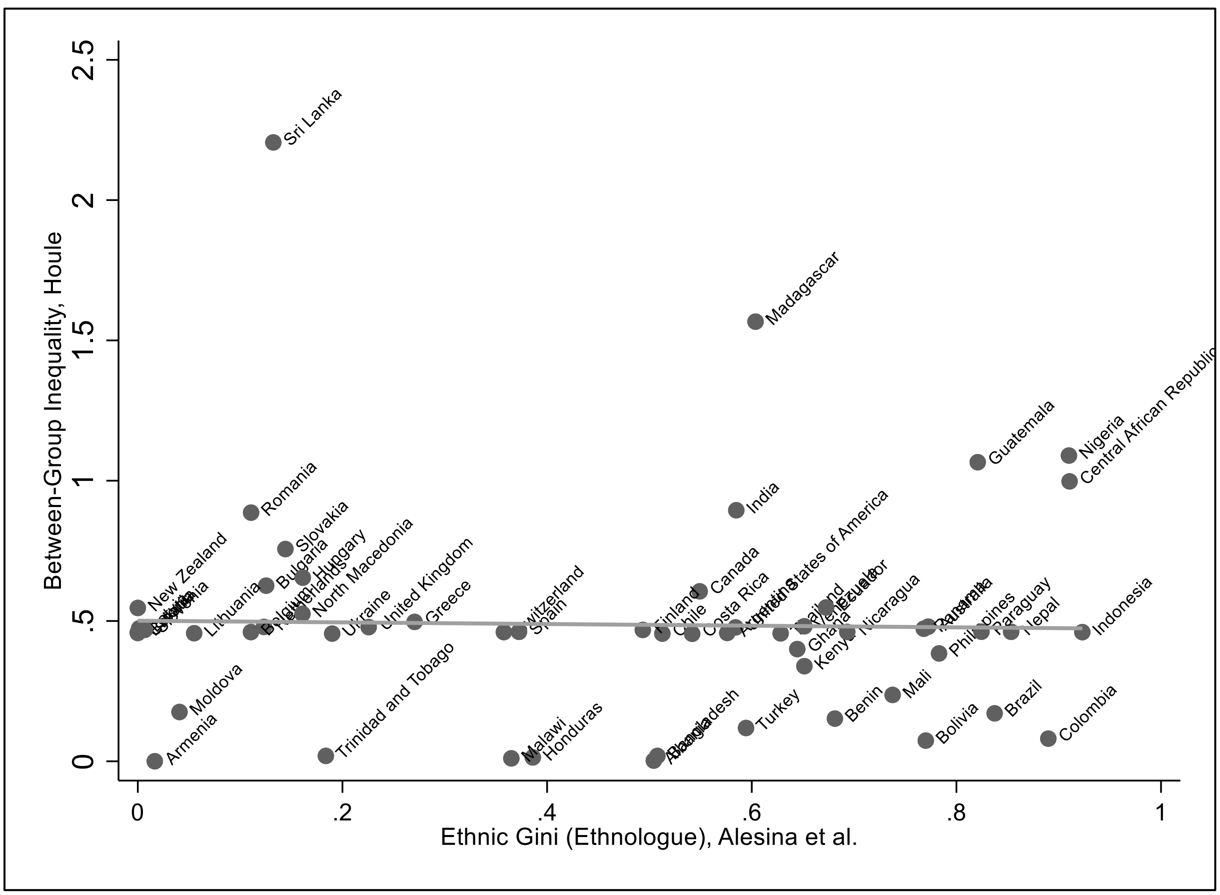
Figure A12: Omoeva et al. 2000



Graphical inspection of non-correlated measures

Figure A7 provides a scatterplot of the non-correlated Alesina et al. and Houle measures in the year 2000. Most of the countries in the Houle dataset ultimately receive a score of around 0.5. At the same time, a series of countries in the Alesina et al. dataset have a value of 0 while receiving a 0.5 score in the Houle dataset. Whereas the scores in the Alesina et al. data are more evenly distributed, we find more extreme values in the Houle data, such as Sri-Lanka and Madagascar. The datasets are also in disagreement on countries such as Bolivia, Brazil and Colombia, which score very high in the Alesina et al. dataset but low in Houle's data. The same can be said at the 'low end' of the spectrum, where, for instance, Central and Eastern European countries such as Romania, Slovakia, Bulgaria and Hungary are assigned relatively higher scores in the Houle data compared to the Alesina et al. data.

Figure A13:Correlation between Houle and Alesina et al. in 2000



Despite the absence of benchmark data, it is possible to gain some sense of the validity of the measures in Figure A7 by exploring whether they reflect economic disparities in a range of cases where the nature of ethnic inequality is widely recognized. It is reassuring that both measures assign relatively high scores to Guatemala and Nigeria, where ethnic inequalities are pronounced (Archibong 2018; Canelas and Gisselquist 2018). Meanwhile, the very low scores assigned to Bolivia and Brazil in Houle's data seem implausible in light of the widely recognized ethnic inequalities in these countries (Leivas and dos Santos 2018; Molina 2007).

Standardized scores for selected countries

To explore the face validity issue further, I consider six well-known cases ethnic inequality is across all six measures: Switzerland, South Africa, Guatemala, Peru, Nigeria and Brazil. To simplify this comparison, I have standardized the measures (mean of 0, standard deviation of 1; higher values indicate higher inequality) and compared the scores for these countries in the year 2000 across the examined measures. Letters A-E indicate the relatively country rank for each measure, with A indicating the most equal country (see Table A7). Switzerland has minor inequalities between language groups (Baghat et al. 2017: 71), whereas South Africa, Guatemala, Peru, Nigeria and Brazil are known to have extensive socioeconomic differences between ethnic groups (Archibong 2018; Canelas and Gisselquist 2018; Figueroa and Barrón 2005; Leivas and dos Santos 2018).

Table A7: Standardized scores for selected countries in the year 2000

	Alesina	Cederman	Houle	B&H	V-Dem	Omoeva
Switzerland	-0.30 (B)	-0.30 (B)	-0.18 (B)	-0.71 (A)	-1.54 (A)	Missing
South Africa	-0.52 (A)	5.33 (F)	Missing	1.56 (D)	0.44 (E)	-0.71 (A/B)
Guatemala	1.09 (C)	-0.33 (A)	0.57 (C)	Missing	0.88 (F)	Missing
Peru	1.57 (F)	1.5 (E)	Missing	Missing	0.30 (C)	-0.48 (C)
Brazil	1.14 (D)	-0.26 (C)	-0.54 (A)	1.91 (C)	0.35 (D)	-0.71 (A/B)
Nigeria	1.35 (E)	0.09 (D)	0.60 (D)	0.98 (B)	0.01 (B)	2.42 (D)

Note: Standardizes scores (z-transformed) for year 2000. Baldwin and Huber for available year between 1995-2005.

Reassuringly, all of the available measures for Switzerland have below-average ethnic inequality scores, albeit with important variation ranging from –.18 (Houle) to –1.54 (Coppedge et al.). The picture for South Africa is less clear-cut; while some measures take high values (Baldwin and Huber; Cederman et al.), others are much lower than we should expect (Alesina et al.; Omoeva et al.). In particular, the Alesina et al. measure indicates slightly *lower* ethnic inequality in South Africa than Switzerland in 2000. One interpretation could be that overlapping ethnic settlement patterns in relatively urbanized South Africa introduce measurement error. However, this should also be the case for the Cederman et al. measure, which also relies on geo-spatial data for ethnic groups; nevertheless, the scores for ethnic inequality for South Africa in 2000 are very high. While the exact reason is unclear, this finding indicates the possibility of measurement error in the nightlight data.

Most measures agree that Guatemala is relatively unequal and assign it an above-the-mean score. However, the data by Cederman et al. provide a relatively low score for Guatemala, which suggests the country is slightly more equal than Switzerland. Similarly, all measures assign relatively high levels of ethnic inequality to Peru – though with big variation and with the exception of Omoeva et al. who assign a below-the-mean score (-0.48). Brazil is variously relatively equal (Omoeva et al. with -0.71) or relatively unequal (Baldwin and Huber with 1.91). Finally, all measures assign high scores to Nigeria, though with rather large variation, i.e. between 0.01 and 2.42. Overall, this patterns indicates that most measures agree only very roughly on the relative order of a country, and with significant variation and hard-to-explain exceptions.

Correlation Analysis: Vertical Inequality

Despite varying definitions and data sources, conventional measures of socioeconomic inequality tend to be more highly correlated. In Table A8, I have compared (1) disposable income Gini by the SWIID, (2) the Gini from the World Development Indicators, (3) a wage share measure by Knutsen (2015), as well as (4) V-Dem's measure of inequality in "access to public services by socio-economic group". The correlations range between 0.44-0.90, with the Gini measures correlating the most. Note that I have rescaled V-Dem's measure (Coppedge et al. 2021) and Knutsen's (2015) wage share measure to ensure that higher levels indicate more inequality, in line with the Gini measures.

Table A8: Correlations between different measures of vertical inequality

	Gini, disp (SWIID)	Gini (WDI)	Public service access (V-Dem)	Wage share (Knutsen 2015)
WDI	0.90			
V-Dem	0.65	0.62		
Wage share	0.44	0.44	0.47	

Test of discriminant validation

To further probe my interpretations, I follow Adcock and Colliers' (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 Gini net coefficient (Solt 2019), interpersonal educational Gini (Clio-Infra 2020) well as two measures of ethnic fractionalization (Alesina et al. 2003; Fearon 2003). While interpersonal income and educational inequality are distinct concepts from ethnic inequality, they should theoretically exhibit a moderate level of correlation with all surveyed measures. This is because greater discrepancies between groups will also partly translate into greater interpersonal differences. We should have similar expectations with regard to ethnic heterogeneity, because a minimum level of heterogeneity is precondition for ethnic inequality, i.e. there should be no fully homogenous countries with high levels of ethnic inequality. The results are presented in Table A9.

The measure by Alesina et al. (2016) is moderately correlated with the different neighboring concepts, Similarly, the measure by Baldwin and Huber (2010) has relatively high correlations with most neighboring concepts (0.59-0.63), yet somewhat lower with regard to educational Gini (0.31). The V-Dem (Coppedge et al. 2021) measure also shows low to moderate correlations with ethnic fractionalization (0.32-0.36) to relatively high correlations with the interpersonal income Gini's (0.62). The relatively high correlations between V-Dem and both income Gini and educational Gini may indicate that at least some country experts had interpersonal rather than ethnic inequality in mind when coding values. This worry is supported by the observation that some homogenous countries – like South Korea in 1950 – show higher ethnic inequality than we would expect. Omoeva et al.'s educational Gini data is moderately correlated with all neighboring concept measures (0.45-0.56) except for the income Gini, which is strikingly low (0.01). Meanwhile, the measure by Houle (2015) demonstrates relatively low correlations with the neighboring concepts (mostly around 0.15-0.20). Strikingly, the G-econ measure by Cederman et al. (2013) has very low and even negative correlations with the neighboring concepts.

Table A9: Correlations between measures and neighboring concepts

	Alesina et al.	Cederman et al.	Houle	Baldwin & Huber	Coppedge et al.	Omeova et al.
Gini, income, net (Solt, 2019)	0.37 (400)	0.13 (2552)	0.13 (1323)	**0.59** (46)	**0.61** (5431)	0.04 (1942)
Educational Gini (Clio Infra, 2020)	0.39 (264)	-0.11 (2552)	0.17 (1534)	0.31 (42)	**0.61** (10867)	**0.56** (3513)
Ethnic fractionalization (Alesina et al. 2003)	**0.46** (519)	-0.00 (2932)	0.14 (1630)	**0.60** (46)	0.34 (16637)	**0.48** (4127)
Ethnic fractionalization (Fearon, 2003)	**0.46** (450)	0.06 (2916)	0.15 (1616)	**0.63** (46)	0.38 (15530)	**0.45** (4086)

Note: results refer to bivariate Pearson's r correlations (n in parentheses), with values over 0.4 made bold.

To use this exercise in a more constructive way, we may also use these correlations to disentangle particular features that each measure seems to capture. In one interpretation, Cederman et al.'s low correlations with neighboring concepts could party be explained by Cederman et al.'s 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 underscores that choosing between ratio-based and aggregate measures – that represent the entire distribution – has important consequences.

Where do the differences between measures emerge?

The source of observed divergence in the empirical analysis stems from a combination of choices regarding (1) ethnic categories, (2) socioeconomic data (survey, nightlights and official sources on local economic activity) and (3) aggregation procedures. The examination suggested that all three stages help explain differences: The choice of comparable ethnic categories clearly matter, as suggested by the .73 correlation between the two Alesina et al. measures, which are identical, save the use of different maps of ethnic homelands (Table A18). In the same vein, 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 incorporate the large non-citizen populations, whereas the Ethnologue does not.

Socioeconomic data sources also matter to a significant degree. Some measures relying on similar conceptualizations and aggregation procedures (e.g., Houle and Alesina et al. data), yet employing different data sources, exhibited no correlation.1 This is also supported by the Cederman et al. (2015: 811) finding that group-level measures from different data sources differ significantly. In the same vein, when I empirically compare the three sources (G-Econ, nightlights and surveys) that go into their composite group-level measure, there are significant differences despite having aggregated them in the same fashion (correlations range between 0.18 and 0.39). Finally, the distinct empirical patterns associated with the Cederman et al. data also suggested that the aggregation procedure has a significant effect on the divergences.

Regression Tables for Replication Analysis

In each replication analysis, I have used the original datasets and Stata code, only substituting the measures of ethnic inequality. The samples are thus bounded by the original analysis' empirical scope. Nevertheless, I also present additional analysis with smaller samples that only cover overlapping country-years for which all measures are available. All ethnic inequality measures are standardized (mean of 0, standard deviation of 1) to ensure comparability. Regressions underlying Tables A12-A17 (replications of Cederman et al., Houle, and Baldwin and Huber) are based on the interpolated values for the Alesina et al. measures to provide a sufficient number of observations.

Replication of Alesina et al. (2016)

Table A10: Replication of Alesina et al. 2016

	GDP per capita (logged) in 2000				
	(1) Alesina et al. (Ethnologue)	(2) Cederman et al.	(3) Houle	(4) Coppedge et al.	(5) Omoeva et al.
Ethnic Ineq.	-0.388***	0.074	0.346**	-0.708***	-0.139
	(0.078)	(0.056)	(0.128)	(0.101)	(0.155)
Constant	8.448***	8.337***	8.689***	8.010***	7.533***
	(0.061)	(0.066)	(0.078)	(0.078)	(0.074)
N	173	149	55	162	74
R2	0.670	0.671	0.799	0.728	0.622
Region F-E	✓	✓	✓	✓	✓

OLS. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A11: Replication of Alesina et al. 2016: Overlapping sample

	GDP per capita (logged) in 2000				
	(1) Alesina et al. (Ethnologue)	(2) Cederman et al.	(3) Houle	(4) Coppedge et al.	(5) Omoeva et al.
Ethnic Ineq.	-0.058	0.163	0.096	-0.501*	0.147
	(0.225)	(0.292)	(0.184)	(0.229)	(0.118)
Constant	7.917***	7.916***	7.925***	7.757***	7.922***
	(0.145)	(0.096)	(0.106)	(0.107)	(0.090)
N	27	27	27	27	27
R2	0.839	0.840	0.839	0.875	0.843
Region F-E	✓	✓	✓	✓	✓

OLS. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Replication of Houle (2015)

Table A12: Replication of Houle

	Democratic Breakdown				
	(1) Alesina et al. (Ethnologue)	(2) Cederman et al.	(3) Houle	(4) Coppedge et al.	(5) Omoeva et al.
Ethnic Inequality	0.324	-0.740	1.072^{***}	0.767^{*}	0.805
	(0.301)	(0.908)	(0.228)	(0.377)	(0.470)
WGI_1	3.090	4.298^{**}	2.276^{*}	2.096^{*}	1.710
	(1.650)	(1.512)	(0.895)	(0.858)	(1.045)
Ethnic Inequality*WGI_1	-0.479	3.393	-2.931^{***}	-1.619	-1.925
	(0.862)	(2.612)	(0.641)	(0.884)	(1.116)
Oil	0.401	0.123	-0.397	-0.146	-0.256
	(0.421)	(0.470)	(0.289)	(0.201)	(0.259)
Ethnic frac.	-0.013	-0.009	-0.007	-0.013	-0.011
	(0.011)	(0.010)	(0.006)	(0.007)	(0.007)
GDP pc	0.035	0.147	-0.100	-0.054	-0.011
	(0.272)	(0.189)	(0.175)	(0.179)	(0.210)
Growth	-0.023	0.004	0.020	0.022	-0.005
	(0.015)	(0.030)	(0.020)	(0.021)	(0.019)
Muslim	0.005	0.005	0.002	0.002	0.001
	(0.005)	(0.006)	(0.003)	(0.003)	(0.003)
Percent world dem.	6.615	-3.580	-4.680^{***}	-3.586^{***}	-3.205^{*}
	(6.688)	(2.764)	(0.798)	(0.855)	(1.270)
Age	0.012	0.019	-0.008	-0.009	0.024^{*}
	(0.012)	(0.011)	(0.006)	(0.007)	(0.012)
Size dom.	-2.623^{**}	-2.621^{**}	-1.761^{***}	-1.546^{**}	-2.550^{***}
	(0.803)	(0.935)	(0.431)	(0.502)	(0.604)
Western	0.000	0.000	-0.770	-0.386	0.000
	(.)	(.)	(0.490)	(0.517)	(.)
Power sharing	0.376	0.201	-0.084	0.200	0.462
	(0.422)	(0.410)	(0.229)	(0.241)	(0.378)
Constant	-5.501	-1.116	1.974	1.042	0.896
	(5.032)	(1.739)	(1.651)	(1.724)	(2.061)
N	710	768	1607	1607	694

Probit. Robust standard errors clustered by country in parentheses; all explanatory variables are lagged.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A13: Replication of Houle – overlapping sample

	Democratic Breakdown				
	(1) Alesina et al. (Ethnologue)	(2) Cederman et al.	(3) Houle	(4) Coppedge et al.	(5) Omoeva et al.
Ethnic Inequality	1.112	-0.462	0.886	1.512	1.761*
	(0.604)	(1.094)	(0.539)	(0.882)	(0.811)
WGI_1	3.273	4.557**	2.713	1.536	5.361*
	(2.350)	(1.650)	(1.787)	(2.349)	(2.132)
Ethnic Inequality*WGI_1	-1.721	2.445	-1.906	-2.077	-2.966
	(1.317)	(2.872)	(1.253)	(1.516)	(1.515)
Oil	1.775	0.879	1.002	1.717	1.498*
	(0.913)	(0.768)	(0.700)	(0.957)	(0.718)
Ethnic frac.	-0.030	-0.024	-0.018	-0.021	-0.023
	(0.019)	(0.014)	(0.012)	(0.012)	(0.013)
GDP pc	-0.261	0.143	-0.110	-0.248	0.194
	(0.618)	(0.296)	(0.381)	(0.609)	(0.520)
Growth	-0.065**	-0.049**	-0.058**	-0.071***	-0.087**
	(0.025)	(0.017)	(0.020)	(0.018)	(0.027)
Muslim	0.010	0.009	0.009	0.012	0.005
	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)
Percent world dem.	6.762	5.760	2.915	5.636	7.232
	(6.237)	(6.507)	(6.617)	(5.686)	(7.072)
Age	0.042*	0.034*	0.046*	0.051*	0.054*
	(0.020)	(0.015)	(0.021)	(0.022)	(0.022)
Size dom.	-3.805**	-3.357**	-3.332**	-4.108**	-3.523**
	(1.394)	(1.185)	(1.130)	(1.462)	(1.203)
Western	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)
Power sharing	0.629	0.589	0.651	0.751	0.162
	(0.583)	(0.606)	(0.610)	(0.647)	(0.687)
Constant	-2.924	-5.793	-2.081	-1.739	-7.294
	(6.034)	(4.732)	(5.021)	(6.379)	(5.758)
N	398	398	398	398	398

Probit. Robust standard errors clustered by country in parentheses; all explanatory variables are lagged.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Replication of Cederman et al. (2013)

Table A14: Replication of Cederman et al. (2013)

	Civil War Onset				
	(1) Alesina et al. (Ethnologue)	(2) Cederman et al.	(3) Houle	(4) Coppedge et al.	(5) Omoeva et al.
Ethnic Inequality	0.074	0.166**	0.520	0.086	0.593
	(0.255)	(0.055)	(0.300)	(0.301)	(0.320)
Democracy	0.224	-0.230	1.335	-0.252	-0.225
	(0.565)	(0.549)	(2.609)	(0.509)	(0.874)
Gini	-0.012	-0.021	-0.000	-0.021	-0.009
	(0.018)	(0.015)	(0.025)	(0.016)	(0.021)
Max exclusion	-0.628	-0.111	-1.967	-0.218	0.864
	(0.981)	(0.726)	(1.317)	(0.805)	(0.832)
Ethnic diversity	1.159	0.787	3.869**	0.802	-0.431
	(0.651)	(0.559)	(1.391)	(0.550)	(0.802)
Population size	0.270*	0.136	0.406	0.195	0.131
	(0.110)	(0.122)	(0.208)	(0.111)	(0.109)
GDP/capita	-0.202	-0.258	-0.173	-0.145	-0.110
	(0.192)	(0.173)	(0.339)	(0.222)	(0.262)
Previous conflicts	0.266	0.471	-0.858	0.422	-0.149
	(0.274)	(0.338)	(0.566)	(0.303)	(0.563)
Peace years	0.436***	0.353**	0.322	0.328**	0.544*
	(0.131)	(0.127)	(0.271)	(0.120)	(0.226)
Constant	-4.868**	-2.479	-8.034**	-3.722	-3.525
	(1.651)	(1.973)	(2.996)	(2.051)	(2.243)
N	2296	2467	913	2467	1177

Logit. Robust standard errors in parentheses. Estimates for three natural cubic splines not shown
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A15: Replication of Cederman et al. (2013) – overlapping samples

	Civil War Onset				
	(1) Alesina et al. (Ethnologue)	(2) Cederman et al.	(3) Houle	(4) Coppedge et al.	(5) Omoeva et al.
Ethnic Inequality	0.872	0.747	1.292*	1.081	1.274
	(0.595)	(0.585)	(0.551)	(0.785)	(0.813)
Democracy	4.434	4.382	3.695	5.100	5.698
	(5.556)	(4.902)	(3.421)	(5.517)	(4.651)
Gini	-0.003	-0.007	-0.045	0.002	0.037
	(0.043)	(0.048)	(0.041)	(0.039)	(0.032)
Max exclusion	-0.612	-0.043	-0.842	-0.206	1.892
	(1.734)	(1.503)	(1.702)	(1.556)	(1.993)
Ethnic diversity	9.316**	11.588**	11.146*	10.956**	7.255*
	(3.193)	(4.309)	(4.694)	(4.059)	(3.064)
Population size	0.196	0.108	0.012	0.147	0.191
	(0.246)	(0.304)	(0.338)	(0.291)	(0.234)
GDP/capita	0.159	-0.033	-0.107	0.025	-0.046
	(0.956)	(0.890)	(0.643)	(1.034)	(0.558)
Previous conflicts	-2.916**	-2.464*	-2.711	-2.937*	-2.717*
	(1.085)	(1.244)	(1.649)	(1.184)	(1.099)
Peace years	0.561	0.517	0.170	0.626	0.645
	(0.445)	(0.474)	(0.542)	(0.505)	(0.489)
Constant	-13.736*	-12.593	-8.120	-13.207	-13.881**
	(6.911)	(8.231)	(6.710)	(8.857)	(5.079)
N	399	399	399	399	399

Logit. Robust standard errors in parentheses. Estimates for three natural cubic splines not shown.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Replication of Baldwin and Huber (2010)

Table A16: Replication of Baldwin and Huber (2010)

	Public Goods Provision					
	(1) Alesina et al. (Ethnologue)	(2) Cederman et al.	(3) Houle	(4) Coppedge et al.	(5) Omoeva et al.	(6) Baldwin & Huber
Ethnic Inequality	-0.279**	0.017	-0.303	-0.307*	-0.353*	-0.137*
	(0.080)	(0.034)	(0.356)	(0.135)	(0.139)	(0.065)
GDP(ln)	0.583***	0.502*	0.299	0.392*	0.459	0.492**
	(0.160)	(0.203)	(0.262)	(0.189)	(0.228)	(0.172)
Population	-0.142*	-0.241**	-0.258*	-0.172**	-0.427	-0.200***
	(0.056)	(0.075)	(0.098)	(0.053)	(0.352)	(0.056)
Polity 2	0.036	0.196	0.332	0.120	0.113	0.172
	(0.126)	(0.153)	(0.170)	(0.128)	(0.132)	(0.116)
Afrobarometer	-0.554	-0.656	-1.010*	-0.677*	-0.664	-0.620*
	(0.297)	(0.332)	(0.481)	(0.313)	(0.455)	(0.289)
WVS	-0.185	0.006	0.065	-0.123	-0.242	-0.043
	(0.158)	(0.149)	(0.230)	(0.161)	(0.391)	(0.133)
CSES	0.085	0.213	0.211	0.098	0.690	0.141
	(0.171)	(0.183)	(0.217)	(0.167)	(0.372)	(0.183)
Constant	0.160	0.123	0.100	-0.057	0.046	0.149
	(0.136)	(0.149)	(0.219)	(0.107)	(0.343)	(0.137)
N	46	46	30	46	21	46
R2	0.875	0.845	0.866	0.865	0.826	0.858

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A17: Replication of Baldwin and Huber (2010) – overlapping samples (NOTE: very small sample)

	Public Goods Provision					
	(1) Alesina et al. (Ethnologue)	(2) Cederman et al.	(3) Houle	(4) Coppedge et al.	(5) Omoeva et al.	(6) Baldwin & Huber
Ethnic Inequality	-0.545	-0.192	0.057	-0.660	-0.330	-0.287
	(0.216)	(1.663)	(0.816)	(0.270)	(0.527)	(0.244)
GDP(ln)	0.424	0.284	0.244	0.024	0.474	0.434
	(0.220)	(0.392)	(0.496)	(0.188)	(0.570)	(0.341)
Population	-0.464	-0.707	-0.772	-0.482	-0.624	-0.884
	(0.252)	(0.736)	(0.662)	(0.257)	(0.566)	(0.479)
Polity 2	0.343	0.428	0.428	0.277	0.260	0.377
	(0.198)	(0.338)	(0.398)	(0.334)	(0.400)	(0.278)
Afrobarometer	-1.112	-1.471	-1.648	-1.884	-0.761	-1.667
	(0.496)	(2.097)	(1.804)	(0.946)	(2.038)	(1.127)
WVS	-0.521	-0.404	-0.509	-0.680	-0.193	-0.869
	(0.595)	(1.603)	(1.479)	(0.779)	(1.332)	(1.086)
CSES	-0.138	0.473	0.265	-0.186	0.550	0.149
	(0.588)	(2.038)	(1.481)	(0.747)	(1.171)	(0.978)
Constant	0.786	0.632	0.787	0.475	0.277	1.143
	(0.367)	(1.473)	(1.334)	(0.490)	(1.196)	(0.829)
N	13	13	13	13	13	13
R2	0.950	0.838	0.838	0.918	0.855	0.864

Robust tandard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Correlations with additional measures

Table A18 provides a correlation analysis that also includes Alesina et al.'s alternative measure (based on the Geo-referencing Ethnic Groups, GREG) data as well as the triangulated measure by Cederman et al. (2015). I have transformed the triangulated measure by Cederman et al. (2015) to a cross-national indicator following Cederman et al.'s suggested approach (2013: 150). The measure exhibits slightly stronger correlations with all the other evaluated measure than their G-Econ measure, yet the correlations are still relatively low.

Table A18: Correlation analysis including additional measures

	Alesina et al. (ethn)	Alesina et al. (greg)	Cederman et al.	Cederman et al. (comp.)	Houle	Baldwin & Huber	Coppedge et al.	Omeova et al.
Alesina et al. (greg)	0.75							
Cederman et al.	0.16	0.16						
Cederman et al. (comp.)	0.34	0.32	0.53					
Houle	0.01	-0.19	0.12	0.20				
Baldwin & Huber	/	/	0.04	0.16	0.01			
V-Dem	0.55	0.47	0.05	0.18	0.31	0.64		
Omoeva et al. (2018)	0.40	0.31	-0.07	0.11	0.30	0.05	0.17	

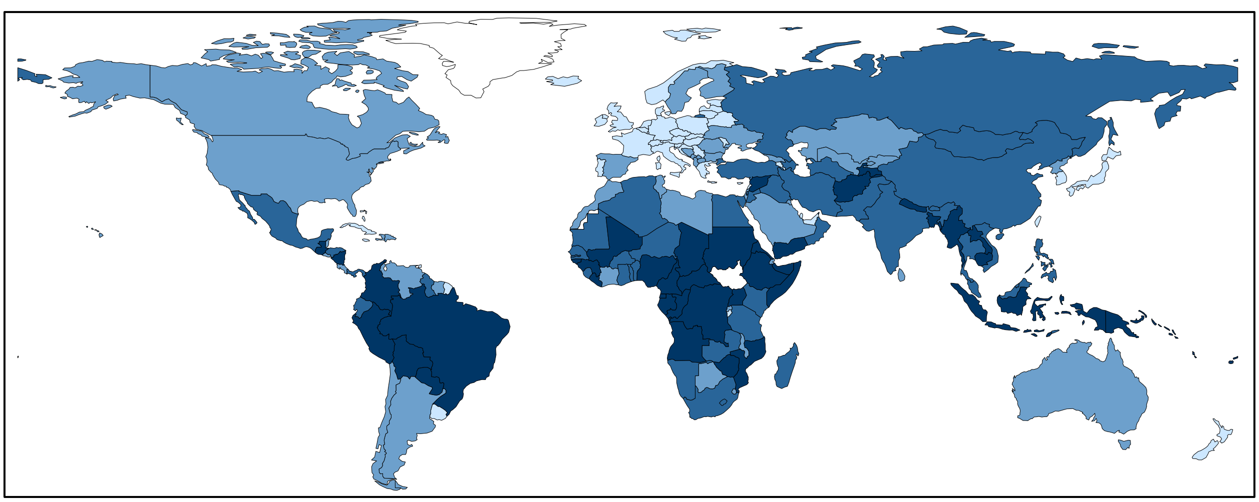
Note: results refer to bivariate Pearson's r correlations (n in parentheses). The topmost measures reflect the economic dimension, whereas the lower two (in grey) reflect the social dimension

Illustrative Example of Index

In the following, I will provide an illustrative example by combining the Alesina et al. and Coppedge et al. (V-Dem) measures to create a measure of overall socioeconomic ethnic inequality. As a first step, this type of measurement should rest on a theoretically valid conceptualization of how the latent variable is expected to manifest (Linzer and Staton 2015: 224). I have previously argued that socioeconomic inequality between ethnic groups entails an economic dimension (income, wealth) and a social dimension (education, health etc.), which together reveal differences in standards of living between the average members of different ethnic groups. The Alesina et al. and Coppedge et al. measures both reflect one of these dimensions, while also providing broad empirical coverage. Combining the two measures is also consistent with the theoretical perspective that economic and social ethnic inequality, although not identical conceptually, are interrelated in practice (Stewart 2002). The empirical analysis supported this logic as the two measures were correlated at 0.55 and loaded on the same factor (Table 3).

Second, we should carefully consider how combining the selected indicators may improve measurement. As the examination made clear, all surveyed measures are subject to nontrivial measurement error, and each of the indicators unto themselves might be a less valid measure of the underlying concept. More specifically, I discussed biases stemming from the inability of the nightlight data to account for overlapping settlement patterns. Local V-Dem experts' intimate knowledge of a country may correct this systematic problem. Meanwhile, concerns with cross-national comparability may occur because different experts code different countries, despite V-Dem's extensive attempts in this regard. For instance, experts may perceive scales differently and use regional standards as a reference point. The nightlight data may help to correct for such cross-national differences in scale perceptions. The values for the year 2000 are visualized in Figure A14. The index was constructed using Stata's alpha command with standardized items. Missing values were replaced with the value of the other, non-missing item.

Figure A14: Index based on Alesina et al. and Coppedge et al. for the year 2000

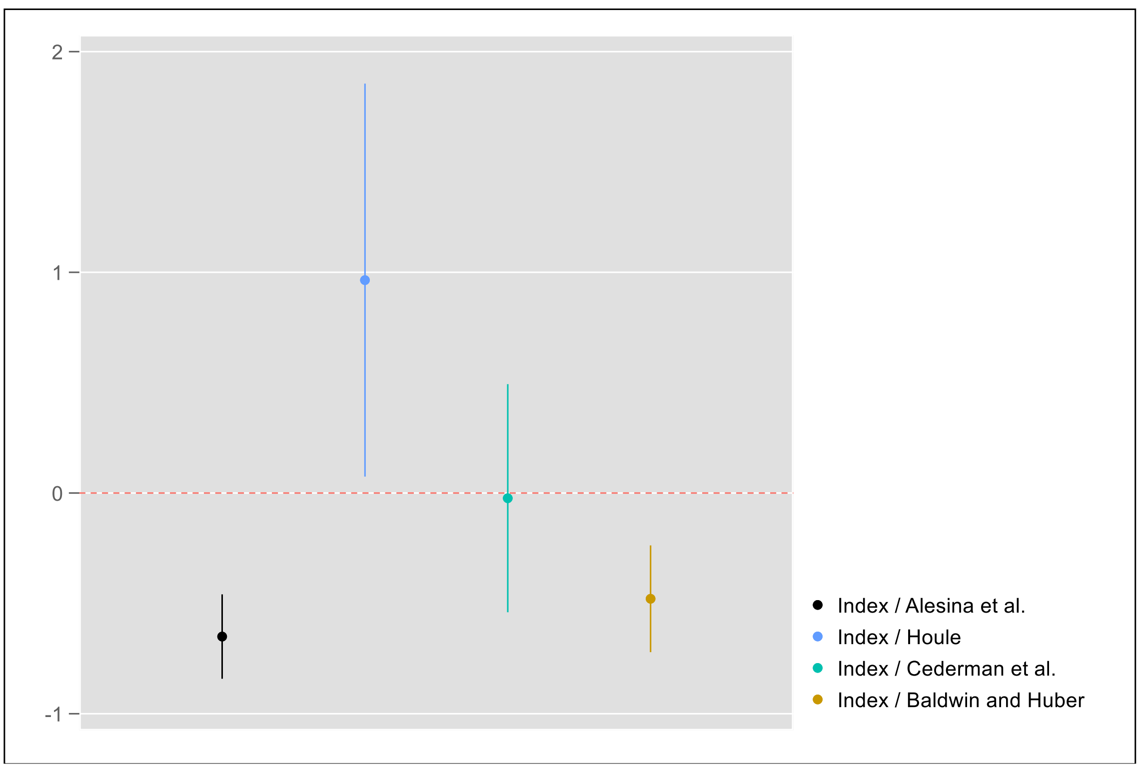


Note: Color shades according to quartiles, with darker shades indicating higher ethnic inequality.

Combining the strengths and weaknesses of these data sources provides plausible estimates for most countries. For instance, the values for the previously mentioned cases of Guatemala, Peru, Brazil and Nigeria, where high levels of ethnic inequality are widely acknowledged, confirm our expectations. Without offering a panacea or substituting for continued data collection, this approach capitalizes on the tendency of each existing measure to capture similar aspects of what makes a country more or less ethnically unequal.

When running the replication analyses with the index, it yields results in line with the original studies in three out of four studies, i.e. it predicts *lower* economic development (Alesina et al), *higher* likelihood of democratic breakdown (Houle), as well as *reduced* public goods provision (Baldwin and Huber). However, it does not predict a higher likelihood of civil war (see Figure A15). Overall, this could indicate that the index contains less measurement error than the measures separately.

Figure A15: Replication studies using index based on Alesina et al. and Coppedge et al. for the year 2000



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., Devleeschauwer A., Easterly W., Kurlat S. and Wacziarg R. (2003) Fractionalization. *Journal of economic growth (Boston, Mass.),* 8**,** 155-194.

Archibong B. (2018) Historical origins of persistent inequality in nigeria. *Oxford Development Studies,* 46**,** 325-347.

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.

Boix C., Miller M. and Rosato S. (2013) A complete data set of political regimes, 1800–2007. *Comparative Political Studies,* 46**,** 1523-1554.

Canelas C. and Gisselquist R.M. (2018) Human capital, labour market outcomes, and horizontal inequality in guatemala. *Oxford Development Studies,* 46**,** 378-397.

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.

Clio-Infra. (2020) *Educational inequality gini coefficient: Https://clio-infra.Eu/. From v-dem dataset v10.* [Online]. [Accessed].

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. (2021) "V-dem codebook v11.1" varieties of democracy (v-dem) project.

Fearon J.D. (2003) Ethnic and cultural diversity by country. *Journal of economic growth (Boston, Mass.),* 8**,** 195-222.

Figueroa A. and Barrón M. (2005) Inequality, ethnicity and social disorder in peru. *Centre for Research on Inequality, Human Security and Ethnicity: Working Paper No. 8,* 8.

Haber S. and Menaldo V. (2011) Do natural resources fuel authoritarianism? A reappraisal of the resource curse. *The American Political Science Review,* 105**,** 1-26.

Inklaar R., De Jong H., Bolt J. and Van Zanden J.L. (2018) *Rebasing 'maddison': New income comparisons and the shape of long-run economic development* [Online]. Groningen

Growth and Development Center. Available: https://ideas.repec.org/p/gro/rugggd/gd-174.html [Accessed].

Knutsen C.H. (2015) Reinvestigating the reciprocal relationship between democracy and income inequality. *Review of Economics and Institutions,* 6.

Leivas P.H.S. and Dos Santos A.M.A. (2018) Horizontal inequality and ethnic diversity in brazil: Patterns, trends, and their impacts on institutions. *Oxford Development Studies,* 46**,** 348-362.

Linzer D., A. and Staton J.K. (2015) A global measure of judicial independence, 1948–2012. *Journal of law and courts,* 3**,** 223-256.

Molina G.G. (2007) Ethnic politics in bolivia: 'Harmony of inequalities' 1900-2000. *Centre for Research on Inequality, Human Security and Ethnicity, CRISE,* working paper, no. 15.

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.

Solt F. (2019) The standardized world income inequality database, versions 8-9. V5 ed.: Harvard Dataverse.

Stewart F. (2002) Horizontal inequalities: A neglected dimension of development. *Queen Elizabeth House Working Paper Series,* Working Paper Number 81.

Weidmann N.B., Kuse D. and Gleditsch K.S. (2010) The geography of the international system: The cshapes dataset. *International interactions,* 36**,** 86-106.