UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2022/25

Ethnic dominance and exclusion

Unpacking cross-national data

Andrea Vaccaro*

March 2022

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: It is widely accepted that the distribution of power between ethnic groups within a country plays a key role in major social, political, and economic outcomes. Researchers working on the topic have various measures of ethnic dominance and exclusion, and other closely related aspects such as ethnic fractionalization, polarization, and heterogeneity, available to them. These measures are, however, often used without sufficient critical reflection and their empirical differences are not entirely understood. The literature does not provide too much guidance either. In fact, efforts to comparatively evaluate data on how power is distributed between ethnic groups are in short supply. To address this gap in the literature, the study at hand provides a comprehensive descriptive analysis of publicly available common cross-national country-level measures on ethnic dominance and exclusion. The findings of this study suggest that Varieties of Democracy's *Exclusion by Social Group* and Ethnic Power Relations' *Ethnic Exclusion* are more alike than the other measures under scrutiny, yet in general the empirical differences among common measures of ethnic dominance and exclusion are substantial. Scholars should thus reject an 'any measure will do' approach when choosing among competing measures.

Key words: ethnic dominance, ethnic exclusion, ethnicity, data, indicators, measurement

JEL classification: D70, J15, N30

Acknowledgements: The author is grateful to Rachel M. Gisselquist for very helpful comments and suggestions.

* UNU-WIDER, Helsinki, Finland, and University of Insubria, Varese, Italy, andrea.vaccaro@wider.unu.edu

This study has been prepared within the UNU-WIDER project Addressing group-based inequalities.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-156-3

https://doi.org/10.35188/UNU-WIDER/2022/156-3

Typescript prepared by Joseph Laredo.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Ethnicity is today a major social science discipline of its own and an important research topic in many social science subfields, including comparative politics, international relations, and political economy. Empirical evidence suggests that ethnicity in its different forms and meanings is related to a variety of significant social, political, and economic factors such as civil war (Cederman et al. 2010; Collier and Hoeffler 2004), regime type (Gerring et al. 2018; Jensen and Skaaning 2012), quality of government (Alesina et al. 2003; Charron 2009), economic growth (Alesina and La Ferrara 2005; Easterly and Levine 1997), public goods provision (Alesina et al. 1999; Gisselquist et al. 2016), and social trust (Dinesen and Sønderskov 2015; Putnam 2007). Even if most experts agree that ethnicity matters for many social scientific outcomes, empirical evidence on ethnicity and its correlates remains far from conclusive, in terms of both signs and causal directions.

A classic background assumption in the ethnicity literature has long been that ethnic complexity within countries impedes progress and thus is not an asset but a curse. More recent studies on the topic have, however, started to move on from such relatively simplistic assessments of the relationship between ethnicity and various social, political, and economic phenomena to sort out and focus more precisely on the mechanisms that drive these associations. For instance, it is now widely believed that in many circumstances certain ethnopolitical power structures matter more than diversity per se (Bates 2000; Cederman and Girardin 2007; Cederman et al. 2010; Collier 2001; Franck and Rainer 2012; Rørbæk and Knudsen 2017). Ethnic diversity might lead to adverse outcomes, but only when it entails an unequal and unfair distribution of political power along ethnic lines—in other words, when a society suffers from pervasive *ethnic dominance and exclusion*.

This increased understanding of the specific mechanisms that link ethnicity to relevant outcomes has been accompanied by the emergence of a new generation of measures of ethnicity. Scholars have developed and started to use more nuanced measures of ethnicity, which capture aspects closely related to ethnic dominance and exclusion. Nevertheless, the scholarship that evaluates measures of ethnicity has not entirely kept up with these developments. There is an important body of research that scrutinizes different cross-national measures of ethnicity (Alesina et al. 2003; Baldwin and Huber 2010; Canelas and Gisselquist 2018, 2019; Chandra and Wilkinson 2008; Desmet et al. 2009; Fearon 2003; McDoom and Gisselquist 2016; Montalvo and Reynal-Querol 2005; Posner 2004) and provides crucial guidance to scholars working on ethnicity. However, these studies do not focus on measures of ethnic dominance and exclusion but on measures of other aspects of ethnic divisions. Marquardt and Herrera (2015) discuss some of the advantages and disadvantages of common measures of ethnic dominance and exclusion, but their useful review of ethnicity data does not involve any quantitative comparison of measures of ethnic dominance and exclusion.

Simply put, no comprehensive comparative study focusing on measures of ethnic dominance and exclusion has been published to date. The study at hand aims to fill this gap in the literature by providing an in-depth descriptive comparative analysis of currently relevant measures of ethnic dominance and exclusion. Even if the main focus of my study is on ethnic dominance and exclusion, I will also discuss some important theoretical and empirical distinctions between different types of ethnicity data. Generally speaking, gaining knowledge on the quantitative information used in social science research is helpful, because data users tend to be uncritical when choosing one measure over another (Mudde and Schedler 2010). Unpacking measures of ethnicity is, however, crucial because measuring ethnicity is problematic and scholars using ethnicity data are not always aware of the conceptual and empirical differences between the measures available

to them. This is particularly true of recent data on ethnic dominance and exclusion, which have not yet been systematically evaluated.

Ultimately, then, this study aims to provide guidance to users of measures of ethnicity and, especially, to users of measures of ethnic dominance and exclusion. On the whole, the findings of my study prove that frequently used measures of ethnic dominance and exclusion share some conceptual underpinnings but are at best weakly convergent in terms of their numerical content. In other words, different measures of ethnic dominance and exclusion portray different pictures of the world. This means that users of these measures should be careful when choosing among different datasets, because their conclusions are more than likely to be conditional on the chosen measure. Varieties of Democracy's *Exclusion by Social Group* and Ethnic Power Relations' *Ethnic Exclusion* seem to be the most similar pair of currently relevant measures of ethnic dominance and exclusion.

My study is structured as follows. In Section 2, I select the measures that are examined in this study. I present their characteristics, consider their differences and similarities in terms of content, and discuss briefly the expected linkages among the measures. In Section 3, I delineate my research strategy. In Section 4, I run the empirical analysis and discuss my results. The scope of this section is primarily global, although I also focus on different regions of the world and in particular on sub-Saharan Africa (SSA). Section 5 concludes.

Before moving on to the next section, I would like to stress that acquiring a more thorough understanding of measures of ethnic dominance and exclusion is important not only to push forward the social scientific scholarship on ethnicity by providing guidance to researchers working on the topic, but also to support policy-makers and global leaders in their decisions. Ensuring political and social inclusion, irrespective of identity, is one of the explicit targets of the United Nations' Sustainable Development Goals. Understanding whether we are making any progress on this global objective and gaining knowledge on how to reach our policy targets requires greater command of cross-national ethnicity data.

2 Measures of ethnicity: selection, characteristics, and expected linkages

Experts working on ethnicity have more and more measures of ethnic dominance and exclusion to choose from. It goes without saying that each and every measure of ethnic dominance and exclusion cannot be analysed and compared comprehensively in a single paper. Therefore, only the most relevant measures are selected for further analysis. Since my study focuses on frequently used measures of ethnic dominance and exclusion, the main selection criterion is popularity in previous social science research on ethnic dominance and exclusion. This presents me with four primary measures for comparison. Nonetheless, I collect a number of additional measures of ethnicity, which are analysed in relation to common measures of ethnic dominance and exclusion. This set of supplementary measures contains not only less frequently used measures of ethnic dominance and exclusion, but also frequently used measures of the broader concept of ethnicity, including measures of ethnic fractionalization, ethnic polarization, and ethnic heterogeneity. Widely used measures of ethnic dominance and exclusion are thus not only compared with each another but also with measures of other aspects of ethnicity. This allows me to extend of the scope of the study at hand to an investigation of the differences and similarities between measures of ethnic dominance and exclusion and measures of the broader concept of ethnicity.

An extensive review of the ethnicity literature suggests that currently the most frequently used measures of ethnic dominance and exclusion are Ethnic Power Relations' *Ethnic Exclusion* (Vogt

et al. 2015), Varieties of Democracy Institute's *Exclusion by Social Group* (Coppedge et al. 2021), and All Minorities at Risk's *Largest Ethnic Group* and *Political Discrimination* indicators (Birnir et al. 2017). These four measures are also the most exhaustively examined measures of ethnicity in my study. As already said, however, my comparative analysis is supplemented with some additional measures of ethnicity. These supplementary measures are (in alphabetical order) O'Brochta's (2020) *Cabinet Diversity Index*, Afrobarometer's (2020) *Ethnic Discrimination*, Alesina et al.'s (2003) *Ethnic Fractionalization*, Vanhanen's (1999) *Ethnic Heterogeneity*, Reynal-Querol's (2002) *Ethnic Polarization*, Ruedin's (2009) *Ethnic Representation Score*, Cederman and Girardin's (2007) *Ethnonationalist Exclusion*, Fund for Peace's (2017) *Group Grievance*, Drazanova's (2020) *Historical Index of Ethnic Fractionalization*, and Minorities at Risk Project's (2009) *Largest Ethnic Group*. Information on the sources of the selected data is presented in Table 1. An overview of the main features of the selected measures is presented in Table 2.

Table 1: Data sources

Measure	**Source**
Principal	
Ethnic Exclusion (EPR)	Ethnic Power Relations (Vogt et al. 2015)
Exclusion by Social Group (VDEM)	Varieties of Democracy (Coppedge et al. 2021)
Largest Ethnic Group (AMAR GP)	All Minorities at Risk (Birnir et al. 2017)
Political Discrimination (AMAR PDIS)	All Minorities at Risk (Birnir et al. 2017)
Supplementary	
Cabinet Diversity Index (CDI)	O'Brochta (2020)
Ethnic Discrimination (ABR)	Afrobarometer (2020)
Ethnic Fractionalization (EF)	Alesina et al. (2003)
Ethnic Heterogeneity (EH)	Vanhanen (1999)
Ethnic Polarization (POL)	Reynal-Querol (2002)
Ethnic Representation Score (ERS)	Ruedin (2009)
Ethnonationalist Exclusion (N*)	Cederman and Girardin (2007); Vogt et al. (2015)
Group Grievance (FSI)	Fund for Peace (2017)
Historical Index of Ethnic Fractionalization (HIEF)	Drazanova (2020)
Largest Ethnic Group (MAR GP)	Minorities at Risk Project (2009)

Source: author's construction.

Table 2: Selected measures of ethnicity

Measure	Year(s)	Countries	Range	Description
Principal				
Ethnic Exclusion (EPR)	1946–2020	176	0–1, low to high	MEG (marginalized ethnic group) population as a fraction of ethnically relevant population
Exclusion by Social Group (VDEM)	1900–2020	179	0–1, most excluded to least excluded	'Exclusion is when individuals are denied access to services or participation in governed spaces … based on their identity or belonging to a particular group' (Coppedge et al. 2021: 296)
Largest Ethnic Group (AMAR GP)	2001–07 (CS)	161	0–1, low to high	Proportion of largest socially relevant ethnic group
Political Discrimination (AMAR PDIS)	1980–2006	129	0–4, low to high discrimination on ordinal scale	0 = No discrimination; 1 = Neglect/remedial policies; 2 = Neglect/no remedial policies; 3 = Social exclusion/neutral policy; 4 = Exclusion/repressive policy
Supplementary				
Cabinet Diversity (CDI)	1967–2017	149	0–1, least diverse to most diverse	Based on the diversity of the names of cabinet ministers
Ethnic Discrimination (ABR)	2016/2018 (CS)	34	0–100, where higher values indicate more discrimination	How often the respondent has been ethnically discriminated against
Ethnic Fractionalization (EF)	1979–2001 (CS)	190	0–1, most homogeneous to least homogeneous	The likelihood that two randomly chosen people in a country come from different ethnic groups
Ethnic Group Representation (ERS)	2006 (CS)	95	0.72–1.00, where 1 represents perfect representation	'Difference between the proportions of citizens and parliamentarians falling into certain ethnic groups' (Ruedin 2009: 339)
Ethnic Heterogeneity (EH)	1990–96 (CS)	183	0–177, low to high	Relative degree of ethnic heterogeneity
Ethnic Polarization (POL)	1970–80 (CS)	137	0–1, where 1 represents perfect polarization	How similar in size two ethnic groups are
Ethnonationalist Exclusion (N*)	1946–2020	88	0–1, where higher values indicate more exclusion	Relative size of ethnic group in power
Group Grievance (FSI)	2005–20	178	1–10, where higher values represent more grievances	'Divisions and schisms between different groups in society' (Fund for Peace 2017)
Historical Ethnic Fractionalization (HIEF)	1945–2013	162	0–1, most homogeneous to least homogeneous	The likelihood that two randomly chosen people in a country come from different ethnic groups
Largest Ethnic Group (MAR GP)	1950–2006	117	0–1, low to high	Proportion of largest minority ethnic group

Note: CS = cross-section (time constant) data.

Source: author's construction.

2.1 Principal measures of ethnic dominance and exclusion

The Ethnic Power Relations dataset provides expert coded annual data both at group level and at country level. The dataset includes data for all politically relevant groups in countries with at least 500,000 inhabitants in 1990 from 1946 on. Political relevance, according to the producers of the dataset, means that the interests of a given group must be represented in the national political arena and, to be included in the dataset, members of a group must be 'systematically and intentionally discriminated against in the domain of public politics' (Wimmer et al. 2009: 325). In total, the dataset includes 817 ethnic groups. Its *Ethnic Exclusion* (EPR) indicator has become a standard measure of ethnic exclusion used in many studies in the field (e.g. Rørbæk 2019; Rudolfsen 2017). EPR quantifies the marginalized ethnic group population as a fraction of the ethnically relevant population within a country. The measure ranges from 0 to 1, where a higher score indicates a more marginalized population and thus a higher level of ethnic exclusion.

The Varieties of Democracy Institute is mainly known for its data on democracy and political institutions. Nevertheless, since 2019, it has also published expert-survey-based data on social and political exclusion. In particular, its *Exclusion by Social Group* (VDEM) index has been used in multiple recent studies to quantify ethnic exclusion (e.g. Piazza 2021; Uzonyi et al. 2021). The index captures the extent to which 'individuals are denied access to services or participation in governed spaces […] based on their identity or belonging to a particular group' (Coppedge et al. 2021: 296). It is based on five sub-indicators, which capture specific aspects of exclusion based on social groups: (1) unequal civil liberties, (2) unequal distribution of political power, (3) unequal access to public services, (4) unequal access to public sector jobs, and (5) unequal access to state business opportunities. VDEM ranges from 0 to 1, where a higher score reflects more ethnic exclusion. The index provides yearly data for virtually all countries in the world from 1900 on.

The All Minorities at Risk (AMAR) dataset has its roots in the pioneering Minorities at Risk (MAR) dataset and can be seen as an upgraded version of MAR. Both datasets provide expert coded information on ethnic groups in countries around the world and contain more or less the same set of indicators. AMAR, however, addresses MAR's problems of sample selection bias (Birnir et al. 2017). While the MAR sample includes only minorities that are judged to be 'at risk', the AMAR sample includes all socially relevant groups within a county. As a consequence, AMAR codes as many as 1,202 ethnic groups, whereas MAR codes only 282. In both datasets the basic unit of analysis is at group level, not country level. To facilitate the use of the AMAR/MAR data in studies where the unit of analysis is country(-year), scholars have transformed AMAR's/MAR's group-level measures to the country level. This transformation has been typically done by taking the highest annual value of a given measure within countries (Caprioli and Trumbore 2003; Piazza 2012). Following the above procedure, I compute country-level scores of AMAR's *Political Discrimination* (AMAR PDIS), which quantifies the level of political discrimination faced by a given ethnic group and has been used in several studies to capture the level of ethnic dominance and exclusion (e.g. Heger 2015; Piazza 2012). The indicator runs on a five-point ordinal scale, where 0 represents no discrimination at all and 4 represents the highest level of exclusion from political participation in relation to the other groups. The yearly coverage of AMAR PDIS currently extends from 1980 to 2006.

AMAR PDIS is not the only measure that I collect from the AMAR dataset. Since ethnic groups that constitute a majority of the population can reasonably aspire to dominate politically (Bates 2008), the size of the largest ethnic group has been used to capture various aspects of ethnicity, including ethnic concentration (Bates 2008), ethnic homogeneity (Fearon and Laitin 2003), multiethnicity (Ellingsen 2000), and ethnic dominance (Collier and Hoeffler 2004; Jackman 1978). I derive from AMAR's group-level data a *Largest Ethnic Group* (AMAR GP) indicator to capture the level of ethnic dominance within a country. The indicator quantifies the proportion of the

largest socially relevant group as a share of the total population and is thus a purely demographic indicator of ethnic dominance and exclusion. AMAR GP is time-invariant, based on 2001/2007 population data, and ranges from 0 (low ethnic dominance) to 1 (high ethnic dominance).

The four measures of ethnic dominance and exclusion discussed above—EPR, VDEM, AMAR PDIS, and AMAR GP—are the main measures analysed in this paper. As already said, however, I complement my descriptive investigation with some additional measures of ethnicity. These ten supplementary measures are presented below.

2.2 Supplementary measures of ethnicity

The *Cabinet Diversity Index* (CDI) is one of the most recently published measures of ethnic diversity. It takes advantage of the NamePrism classification algorithm to capture the ethnic diversity of cabinets by quantifying the predicted probability, based on their names, that cabinet ministers belong to different ethnic groups (O'Brochta 2020). These predicted probabilities are ultimately transformed into the final CDI scores by using the well-known Herfindahl–Hirschman concentration index. The information on the names of cabinet members is collected from the CIA's *Chiefs of State and Cabinet Members of Foreign Governments* database. CDI thus does not focus on ethnic diversity in the general population, but only on the ethnic diversity of cabinet members. The index ranges from 0 (low diversity) to 1 (high diversity) and its country-year data are available from 1967 to 2017 for 149 countries in the world.

Afrobarometer provides data on citizens' perceptions concerning social, political, and economic issues in Africa. Its seventh survey round (2016/2018) provides data for 34 African countries. While most of the survey questions are related to topics other than ethnicity, question Q86C (ABR) asks how often respondents have been personally discriminated against on account of their ethnicity in the past year (Afrobarometer 2020). The respondents are evaluated on a four-point Likert scale: 0 = never, 1 = once or twice, 2 = several times, 3 = many times. Based on the responses to this question, for the study at hand, I have created a measure of ethnic discrimination that additively aggregates the percentages of participants who responded either 'several times' or 'many times'.

Ethnic fractionalization indices are arguably the most widely used category of ethnicity measures. These indices use the Herfindahl–Hirschman index to estimate the probability that two randomly picked individuals in a country do not belong to the same ethnic group. Multiple measures of ethnic fractionalization have been developed over the years. Here, I select two of them for further analysis: (1) Alesina et al.'s (2003) *Ethnic Fractionalization* index (EF), which is the most widely cited measure of ethnicity to date[1] and provides cross-sectional data on 190 countries in the world, and (2) Drazanova's (2020) *Historical Index of Ethnic Fractionalization* (HIEF), which is the most complete longitudinal measure of ethnic fractionalization to date and covers 162 countries in the world from 1945 to 2013. Both indices range from 0 to 1, where a lower score denotes less ethnic diversity.

The *Ethnic Representation Score* (ERS) captures ethnic power structures through 'the difference between the proportions of citizens and parliamentarians falling into certain ethnic groups' (Ruedin 2009: 339). Therefore, ERS is a measure of the imbalance between the share of each ethnic group and its proportional political representation at the country level. The measure is available for 95 democracies in 2006 and is based on representation in lower chambers (or single chambers in unicameral systems). In theory, the measure ranges from 0 to 1, where 1 represents a perfect balance between ethnic groups and elected members of parliament. In practice, however, the

[1] The paper in which the measure is presented had 6,476 Google Scholar citations as of 7 January 2022.

measure ranges from 0.72 to 1. ERS is 'based on newly collected data on the ethnic distribution of representatives' (Ruedin 2009: 342).

Vanhanen's (1999) *Ethnic Heterogeneity* (EH) captures ethnic divisions along three categories of ethnic groups: those based on racial differences, those based on linguistic differences, and those based on religious differences. For each category, ethnic division is first measured by the percentage of the largest ethnic group as a share of the total population. The three inverse percentages are then additively aggregated into the final index of heterogeneity (Vanhanen 1999). A higher score thus denotes a higher level of ethnic heterogeneity. EH ranges from 0 to 177 and provides cross-sectional time-constant information on 183 countries in the world for the period 1990–96. The information on ethnic groups and their composition comes from multiple sources.

Ethnic polarization indices are another well known category of measures of ethnicity used in cross-national quantitative social science research. These measures capture the extent to which the distribution of ethnic groups follows a bimodal distribution (Reynal-Querol 2002). Perfect polarization occurs when a society is composed of two equally sized ethnic groups. One of the most widely used indices of ethnic polarization is Reynal-Querol's (2002) *Ethnic Polarization* (POL). This measure is chosen for further analysis in the study at hand. The primary source for POL's ethnic group classification is the World Christian Encyclopedia. POL ranges from 0 to 1, where a higher score denotes more ethnic polarization. The measure provides scores for a cross-sectional time-constant sample of 137 countries in the world.

Ethnonationalist Exclusion (N*) quantifies 'the relative degree to which ethnic groups are included or excluded from state power' (Cederman and Girardin 2007: 176) by focusing on the country-level demographic balance of ethnic group configuration. The underlying idea of N* is that countries in which the ethnic group in power is a demographic majority are less likely to fall into ethnic conflict. The coding of the measure originates in extant information on ethnicity from Fearon (2003), Heger and Salehyan (2007), Minorities at Risk, and the *CIA World Factbook*. N* ranges from 0 to 1 and 'goes up as long as the dominant group's share of the total population declines' (Cederman and Girardin 2007: 177). In other words, higher scores reflect a higher likelihood of ethnic conflict. Initially coded for 88 countries in Cederman and Girardin (2007), the measure is now available for 146 countries in the latest version of the *Ethnic Power Relations* dataset (Vogt et al. 2015).

Fund for Peace's *Fragile States Index* is commonly employed by both policy-makers and researchers to capture aspects of state fragility. One of its sub-indicators—*Group Grievance* (FSI)—'focuses on divisions and schisms between different groups in society' and answers, for instance, the question 'Are groups oppressed or do they feel oppressed?' (Fund for Peace 2017: 7). Even if the indicator has not been widely used in the ethnicity literature, it seems to be a content-valid measure of grievances and divisions between ethnic groups. Yearly FSI scores are available for up to 178 countries in the world since 2005. Its scores are based on a combination of expert review, content analysis, and quantitative secondary data. The measure ranges from 1 to 10, where a lower score denotes fewer group grievances.

The latest version of the Minorities at Risk (MAR) dataset offers group-level data on ethnopolitical groups that suffer or benefit from discrimination in relation to other groups or engage in collective action to promote or defend their own interests (Minorities at Risk 2009). The MAR dataset thus includes only a fraction of all existing ethnic groups; it includes only minorities at risk, not all socially relevant ethnic groups, as does AMAR. While MAR and AMAR have many similarities, considering their disagreements in sample size, we would expect to find substantive differences in the numerical content of some of their measures. I derive from the MAR dataset a *Largest Ethnic Group* (MAR GP) indicator, which captures the proportion of the largest ethnic group as a share

of total population in a given country. Keeping in mind the disagreements in the samples of the MAR and AMAR datasets, MAR GP in fact captures the proportion of the largest ethnic minority at risk, whereas AMAR GP captures the proportion of the largest socially relevant ethnic group. The latest version of MAR data provides yearly data from 2004 to 2006.[2]

2.3 Expected linkages

Having presented the selected measures of ethnicity, I would like to discuss briefly the theoretical expectations on the relationships between the four principal measures of ethnic dominance and exclusion, as well as, more generally, those between these four measures of ethnic dominance and exclusion and the other selected measures of ethnicity. First, considering the content of the measures of ethnic dominance and exclusion, it seems plausible that the measures explicitly capturing ethnic exclusion (VDEM, AMAR PDIS, and EPR) are positively related among one another. Ultimately, these three measures should be capturing the same concept, albeit with slightly different approaches. Therefore, we would expect at least these three of our four main measures of ethnic dominance and exclusion to be highly convergent among one another.

In contrast, we would expect AMAR GP, which is a purely demographic measure of ethnic dominance quantifying the share of the largest ethnic group out of the total population, to be less clearly related to VDEM, AMAR PDIS, and EPR. On the one hand, it has been argued that if one demographically dominant ethnic group forms a coherent majority in a given country, the rest of the ethnic groups in this country face permanent exclusion from power (Horowitz 1985). The size of the largest group could thus be positively related to the extent of exclusion. On the other hand, if minority groups are small, the extent of exclusion cannot be wide-reaching in demographic terms, although it could still be substantial in terms of intensity. If the largest ethnic group is ruled by a minority group, then we would expect repression and exclusion to be pervasive (Bates 2008).

When it comes to the associations between our four main measures of ethnic dominance and exclusion and our ten additional measures of ethnicity, we have somewhat heterogeneous expectations. We expect MAR GP to be positively related to VDEM, AMAR PDIS, and EPR because, as the demographic size of the largest minority group increases, the number of possibly discriminated-against and excluded people increases. Moreover, since we know that ethnic minorities are often excluded from political power (Ruedin 2009) and that dominant ethnic groups often engage in coethnic favouritism (Cederman et al. 2010; Franck and Rainer 2012), the existence of a large minority group could entail more ethnic exclusion. This exclusion could be even more intense if the largest minority group holds power. The relationship between MAR GP and AMAR GP is likely to be inversely related, at least to a certain extent, simply because if the largest ethnic group is demographically large, the largest minority must be demographically small.

If we assume that ethnic minorities are more likely to be marginalized than ethnic majorities, then ERS should be inversely related to VDEM, AMAR PDIS, and EPR. An unfair composition of parliament that does not reflect the proportions of ethnic groups in a country could be seen as an indication of ethnic dominance and exclusion. The relationship between ERS and AMAR GP is harder to predict. Conversely, because of conceptual similarities, I expect N*—a measure of ethnonationalist exclusion—to be positively related to VDEM, AMAR PDIS, and EPR. Countries that score higher on N* should also have more social exclusion (VDEM), more political discrimination (AMAR PDIS), and more marginalized ethnic groups (EPR). N* and AMAR GP are both likely to be high in countries where the demographically largest ethnic group is excluded from power. Conversely, I expect N* to be inversely related to AMAR GP if the demographically

[2] Previous versions of the MAR dataset provide data from 1950 to 2003.

largest ethnic group holds power. I expect FSI to be positively related to VDEM, AMAR PDIS, and EPR because grievances between ethnic groups are likely to increase if exclusion along ethnic lines increases. The relationship between ABR and VDEM, AMAR PDIS, and EPR is plausibly positive as well, because citizens' perceptions of individual ethnic discrimination are likely to reflect the extent of group-level ethnic exclusion, at least in part.

A priori, the relationship between measures of ethnic diversity (CDI, EF, EH, and HIEF), ethnic polarization (POL), and VDEM, AMAR PDIS, and EPR remains relatively unclear. It is obvious that ethnic dominance and exclusion requires a certain degree of ethnic diversity and polarization, yet ethnic diversity and polarization do not necessarily imply ethnic dominance and exclusion. On the whole, the relationship between measures of ethnic diversity and AMAR GP should be inverse, because a low score on measures of diversity means that a given country is relatively homogeneous and has one demographically large ethnic group, whereas a high score means that a given country has many small ethnic groups.

3 Research strategy

I begin the statistical comparison by exploring the convergence of our measures of ethnic dominance and exclusion, and also their association with the other selected measures. Correlations are a classic tool for evaluating the convergence of a set of indicators. Bivariate correlations are run with both Pearson's and Spearman's methods to ensure the robustness of the results and to take into account possible non-linearity of the relationships. If bivariate correlations among our measures are strong, we can conclude that different cross-national measures of ethnicity are convergent. Conversely, if bivariate correlations among our measures are weak, we can conclude that different cross-national measures of ethnicity are not convergent. Since the correlation analysis indicates that AMAR PDIS is weakly convergent with the other three measures of ethnic dominance and exclusion, I complement my evaluation of convergence with a ridge plot analysis of the distributional features of VDEM, EPR, and AMAR GP across different levels of AMAR PDIS.

I then shift my focus to our four main measures of ethnic dominance and exclusion in different regions of the world. First, with box plots and bar plots, I explore differences in ethnic dominance and exclusion in each region in 2006, which is the most recent year of common observations. Besides providing a summary of some of the most important features of the data, box plots allow us also to identify possible outlier observations region by region and whether these outliers are idiosyncratic to a given measure or consistent across different measures. Second, I take advantage of the time-series information of some of our measures to examine the evolution of ethnic dominance and exclusion (VDEM, EPR, AMAR PDIS) and ethnic diversity (CDI, HIEF) across regions of the world. Differences and similarities between measures are illustrated and analysed with line plots.

Last but not least, I pay particular attention to our measures of ethnicity in SSA. Again I begin by running bivariate correlations among different measures of ethnicity to assess the convergence of ethnicity data, but this time with a sample of SSA countries. This analysis of convergence allows us also to investigate whether there are differences in the results vis-à-vis the previously used global sample of countries. Then I pick four interesting SSA countries (Congo DR, Rwanda, Sierra Leone, Zimbabwe) for further time-series analysis. I explore the differences and similarities among some of our measures (VDEM, EPR, AMAR PDIS, CDI, HIEF) in these four countries.

I give special importance to SSA because a large part of the ethnicity literature has focused on this region. It has been argued that SSA's ethnic diversity impedes development (Easterly and Levine

1997), that 'ethnic fragmentation is arguably the single most important threat to many African countries' (Herbst 2014: 173), and that ethnicity plays a key role in mobilizing groups for collective action, especially in SSA (Rudolfsen 2017). Put simply, experts believe that ethnicity matters especially in SSA. Therefore, in a study on measures of ethnic dominance and exclusion, it seems reasonable to pay special attention to this region of the world. Since ABR provides scores only for African countries, the measure is included only in this part of the empirical analysis. On the contrary, ERS is excluded from this last part of the study, because it codes only 13 countries in SSA.

4 Empirical results and analysis

I begin the empirical analysis with an evaluation of the convergence of our principal measures of ethnic dominance and exclusion, and also their convergence with different measures of the broader concept of ethnicity. The results of the correlation analysis (Table 3) are quite puzzling. While some of the measures are highly correlated, others are hardly associated at all with the other measures. Out of 78 Pearson's correlation coefficients, 19 are non-significant at conventional levels and only 13 are larger than 0.50 or smaller than –0.50. Out of 78 Spearman's correlation coefficients, 20 are non-significant at conventional levels and 22 are larger than 0.50 or smaller than –0.50.

Considering all the measures of ethnicity included in the correlation analysis, with Pearson's method, the strongest bivariate correlations are between N* and EPR (0.86), EF and HIEF (0.84), HIEF and EH (0.71), HIEF and AMAR GP (–0.73), and EF and AMAR GP (–0.75). The weakest bivariate correlations are between N* and CDI (0.0001), AMAR PDIS and CDI (0.004), CDI and EF (0.02), AMAR PDIS and HIEF (–0.01), CDI and POL (–0.02). If we consider only our four main measures of ethnic dominance and exclusion (VDEM, EPR, AMAR GP, AMAR PDIS), the strongest Pearson's correlations are between VDEM and EPR (0.48), AMAR GP and VDEM (–0.38), and AMAR GP and EPR (–0.43), whereas the weakest Pearson's correlations are between AMAR GP and AMAR PDIS (–0.06), AMAR PDIS and VDEM (0.24), and AMAR PDIS and EPR (0.29).

With Spearman's method, if we take into consideration all our measures of ethnicity, the strongest bivariate correlations are between N* and EPR (0.91), EF and HIEF (0.84), AMAR GP and EH (–0.78), AMAR GP and EF (–0.77), and EH and HIEF (0.75). The weakest bivariate correlations are between CDI and N* (–0.02), CDI and POL (–0.01), CDI and EF (0.01), CDI and EPR (0.02), and CDI and AMAR PDIS (0.02). If we consider only our four main measures of ethnic dominance and exclusion, the strongest Spearman's correlations are again between VDEM and EPR (0.46), AMAR GP and VDEM (–0.38), and AMAR GP and EPR (–0.43), whereas the weakest Spearman's correlations are between AMAR GP and AMAR PDIS (–0.09), VDEM and AMAR PDIS (0.24), and EPR and AMAR PDIS (0.24).

Generally speaking, CDI seems to be the least convergent measure of all the examined measures of ethnicity. The correlation coefficients of the bivariate relationships between CDI and the other measures range from –0.11 to 0.18 with Pearson's method and from –0.08 to 0.19 with Spearman's method. CDI is significantly correlated only with VDEM with both correlation methods. The relationship between VDEM and CDI is positive, albeit weak. AMAR PDIS is also at best weakly correlated with the other measures. Bivariate correlations between AMAR PDIS and the other measures range from –0.10 to 0.31 with Pearson's method and from –0.09 to 0.30 with Spearman's method. If we consider both correlation methods, AMAR PDIS is significantly (and positively) related only to VDEM, EPR, N*, and FSI.

Table 3: Correlations between measures of ethnicity in the world

	VDEM	EPR	CDI	ERS	MAR GP	AMAR GP	AMAR PDIS	POL	HIEF	EH	EF	N*	FSI
VDEM	1.00 (175)	0.46*** (170)	0.19* (171)	-0.08 (82)	0.38*** (117)	-0.38*** (161)	0.24** (128)	0.26** (129)	0.37*** (157)	0.49*** (170)	0.47*** (167)	0.46*** (132)	0.63*** (170)
EPR	0.48*** (170)	1.00 (172)	0.02 (169)	-0.49*** (81)	0.56*** (117)	-0.43*** (161)	0.24** (129)	0.44*** (127)	0.47*** (156)	0.51*** (167)	0.45*** (165)	0.91*** (132)	0.51*** (168)
CDI	0.18* (171)	0.03 (169)	1.00 (172)	-0.03 (80)	0.07 (116)	-0.08 (159)	0.02 (128)	-0.01 (128)	-0.05 (156)	0.14 (168)	0.01 (165)	-0.02 (130)	0.08 (168)
ERS	-0.05 (82)	-0.45*** (81)	-0.11 (80)	1.00 (82)	-0.42** (56)	0.33** (76)	0.06 (59)	-0.51*** (62)	-0.34** (75)	-0.40*** (82)	-0.36*** (82)	-0.38** (62)	-0.14 (81)
MAR GP	0.36*** (117)	0.55*** (117)	0.06 (116)	-0.27* (56)	1.00 (117)	-0.51*** (117)	0.11 (117)	0.44*** (87)	0.38*** (113)	0.57*** (117)	0.51*** (117)	0.75*** (105)	0.36*** (116)
AMAR GP	-0.38*** (161)	-0.43** (161)	-0.09 (159)	0.18 (76)	-0.29** (117)	1.00 (161)	-0.09 (128)	-0.32*** (121)	-0.74*** (153)	-0.78*** (159)	-0.77*** (158)	-0.51*** (131)	-0.29*** (160)
AMAR PDIS	0.24** (128)	0.29** (128)	0.004 (128)	0.08 (59)	-0.10 (117)	-0.06 (128)	1.00 (129)	0.12 (97)	-0.04 (124)	0.14 (128)	0.03 (127)	0.22* (111)	0.30*** (127)
POL	0.29*** (129)	0.39*** (127)	-0.02 (128)	-0.43*** (62)	0.36*** (87)	-0.32*** (121)	0.21* (97)	1.00 (129)	0.50*** (117)	0.54*** (129)	0.51*** (128)	0.36*** (96)	0.26** (127)
HIEF	0.36*** (157)	0.43*** (159)	0.04 (156)	-0.31** (75)	0.21* (113)	-0.73*** (153)	-0.01 (124)	0.57*** (117)	1.00 (159)	0.75*** (155)	0.84*** (154)	0.47*** (125)	0.28*** (155)
EH	0.46*** (170)	0.44*** (167)	0.16* (168)	-0.33** (82)	0.39** (117)	-0.70*** (159)	0.14 (128)	0.50*** (129)	0.71*** (155)	1.00 (170)	0.77*** (167)	0.52*** (130)	0.41*** (168)
EF	0.46*** (167)	0.45*** (165)	0.02 (165)	-0.35** (82)	0.30*** (117)	-0.75*** (158)	0.04 (127)	0.56*** (128)	0.84*** (154)	0.70*** (167)	1.00 (167)	0.54*** (129)	0.31*** (165)
N*	0.42*** (132)	0.86*** (132)	0.0001 (130)	-0.17 (62)	0.59*** (105)	-0.43*** (131)	0.25** (111)	0.24* (96)	0.30*** (125)	0.30*** (130)	0.36*** (129)	1.00 (132)	0.39*** (130)
FSI	0.63*** (170)	0.47*** (168)	0.08 (168)	-0.10 (81)	0.32*** (116)	-0.31*** (160)	0.31*** (127)	0.29*** (127)	0.30*** (155)	0.38*** (168)	0.33*** (165)	0.30*** (130)	1.00 (170)

Note: Pearson's correlation coefficients (bottom-left quadrant) and Spearman's correlation coefficients (upper-right quadrant) from 1980 to 2006 (average). Number of observations in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

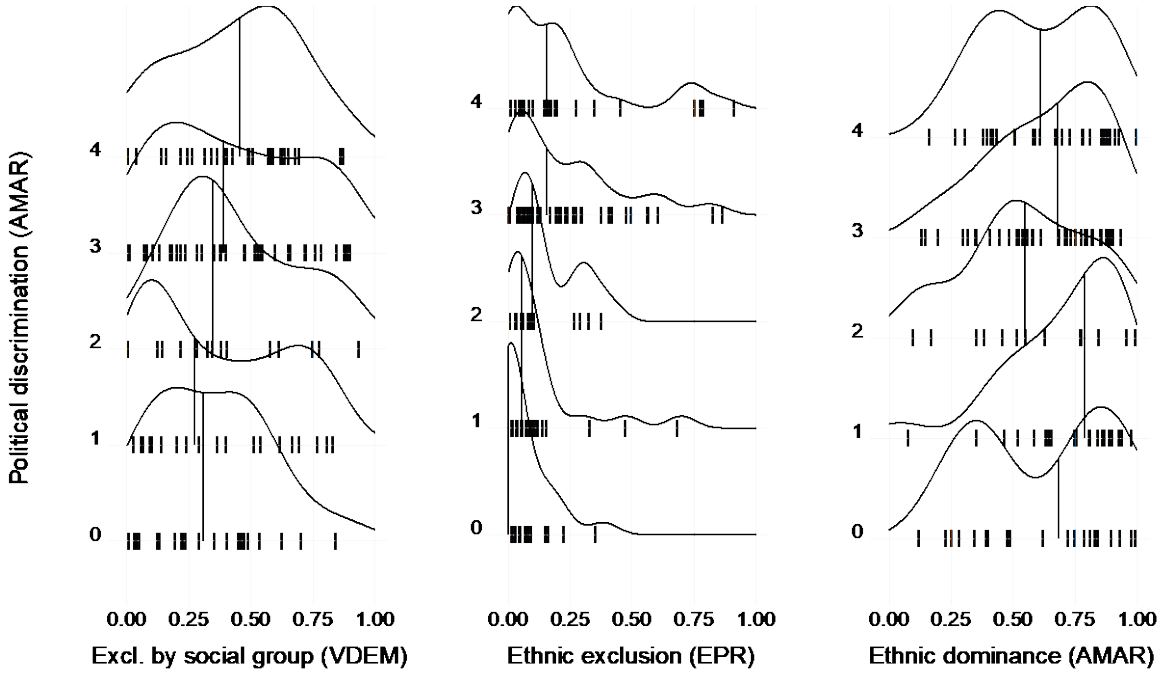
Source: author's construction.

The correlation analysis indicates that in most cases both Pearson's and Spearman's methods lead to similar results. Nevertheless, interestingly, the strength of some of the bivariate associations depend significantly on the chosen correlation method. Virtually without exception, when our two methods of correlation result in large discrepancies, Pearson's method leads to weaker correlations than Spearman's method. For instance, the bivariate correlations between AMAR GP and ERS (Pearson's r: 0.18; Spearman's rho: 0.33) and AMAR GP and MAR GP (Pearson's r: –0.29; Spearman's rho: –0.51) are considerably stronger with Spearman's method than with Pearson's. Since Pearson's correlations capture the magnitude of the linear relationship between two variables and Spearman's correlations capture the magnitude of the monotonic relationship between two variables, these large discrepancies deriving from the chosen correlation method can be taken as an indication of non-linearity in some of the relationships.

Bivariate correlations between our main four measures of ethnic dominance and exclusion are all statistically significant, except for the correlation between AMAR GP and AMAR PDIS, regardless of the employed correlation method. AMAR GP is inversely related to the other measures of ethnic dominance and exclusion, whereas the remaining three measures of ethnic dominance and exclusion are positively related among one another. AMAR PDIS is clearly the least convergent measure of ethnic dominance and exclusion.

To investigate in more detail the weak convergence between AMAR PDIS and the remaining three measures of ethnic dominance and exclusion, since AMAR PDIS is ordinal in nature, we can examine how these three measures are distributed across different levels of political discrimination. The ridge plots in Figure 1 provide such an alternative view of the relationship between AMAR PDIS and the other three measures of ethnic dominance and exclusion. If AMAR PDIS were similar to VDEM, EPR, and AMAR GP, the distributions of these three measures would follow a constant increase or decrease from the bottom level of AMAR PDIS to the top level of AMAR PDIS.

Figure 1: Distribution of measures of ethnic exclusion and dominance according to level of political discrimination (AMAR PDIS) in the world in 2006



Source: author's construction.

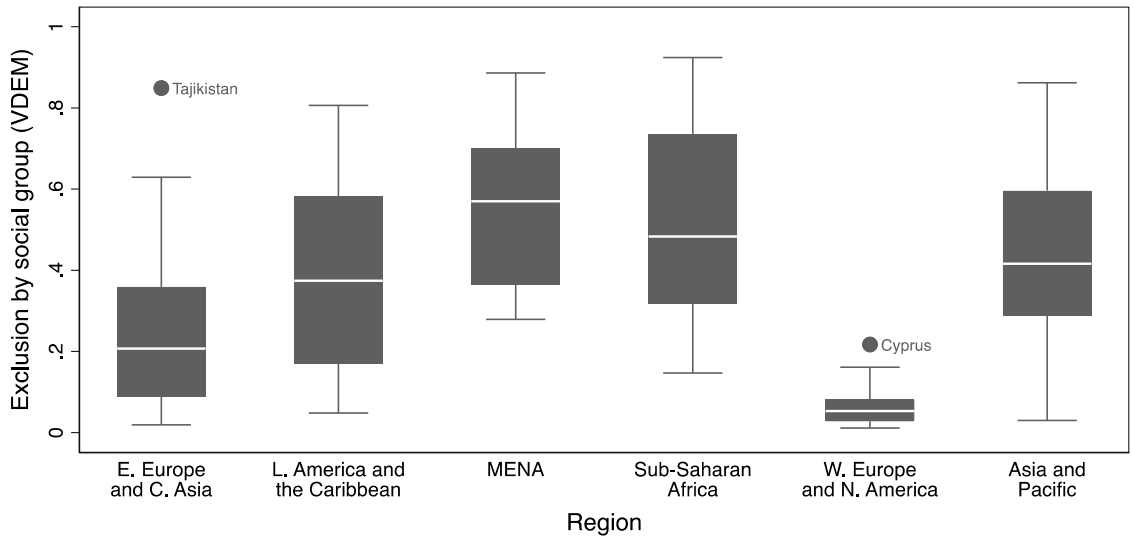
As suggested by the results of the correlation analysis, however, there seems to be no clear pattern between political discrimination (AMAR PDIS) and demographic dominance (AMAR GP). There are both large and small dominant ethnic groups at all levels of discrimination. On average, the highest level of demographic dominance occurs at level 1 of political discrimination, which is the level of discrimination in countries such as Canada, Australia, and Germany. Conversely, the ridge plots confirm that ethnic exclusion increases as the level of political discrimination increases. The median level of EPR increases with every level of political discrimination, except from level 3 to 4. While low levels of ethnic exclusion exist at all levels of political discrimination, the highest levels of ethnic exclusion exist only at the highest levels of discrimination. Likewise, the median level of VDEM increases with every level of political discrimination, except from level 0 to level 1, where we find a small decrease. The main peak of VDEM is at the upper half of the scale only at the highest level of political discrimination. Yet, we can find countries with both high and low levels of exclusion at any level of political discrimination.

For now, the focus of my empirical analysis has been global. However, it is well known that in some regions of the world issues related to ethnic dominance and exclusion play a more important role than in others. We would thus expect to see some similarities in how different measures depict different regions of the world. The box plots in Figures 2–4 provide information on the distribution of observations by regions of the world in the most recent year of common observations. Each box corresponds to the interquartile range, the lines in the middle of the box stand for the medians, and the whiskers extend to the lowest and highest observations in the data, with the exception of possible outliers, which are represented by dots above or below the whiskers. These box plots are especially useful in showing whether there are common outliers within regions among different measures and whether different measures of ethnicity portray similar or different pictures of ethnicity across the world's regions. Countries are classified into regions following the *e_regionpol_6C* classification retrieved from the Varieties of Democracy dataset (Coppedge et al. 2021).

Figure 2 shows the distribution of exclusion by social group (VDEM) in the world's regions. On average, in Western Europe and North America, there is far less ethnic exclusion than in the other regions. The highest levels of exclusion can be found in the Middle East and North Africa (MENA) and in SSA. The latter region, together with Latin America and the Caribbean, also has the broadest interquartile range, which means in practice that the intermediate values in these two regions are more spread out than in the other four regions. Tajikistan is an outlier in Eastern Europe and Central Asia, whereas Cyprus is an outlier in Western Europe and North America. Both countries have a substantially higher level of exclusion than the other countries in their respective regions.

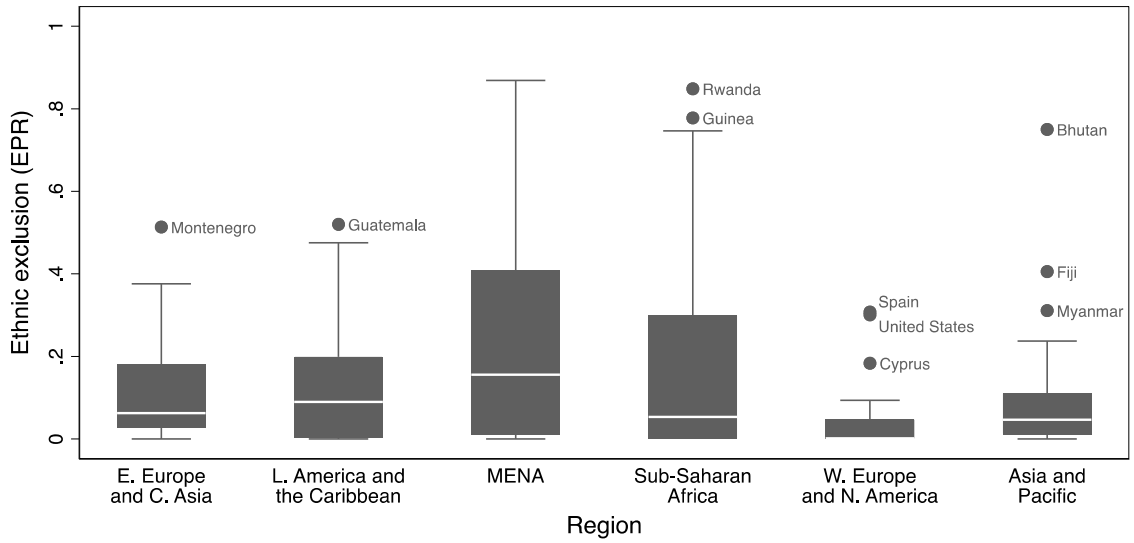
Figure 3 provides similar information on the distribution of ethnic exclusion (EPR) in different regions in the world. As before, MENA and SSA have the uppermost boxes and whiskers, indicating that in general these two regions have much higher exclusion rates than the other regions. Likewise, as before, Western Europe and North America have on average the lowest levels of ethnic exclusion. Nonetheless, with EPR, there are outlier observations in all regions except MENA. Montenegro is an outlier in Eastern Europe and Central Asia, Guatemala is an outlier in Latin America and the Caribbean, Rwanda and Guinea are outliers in SSA, Spain, USA, and Cyprus are outliers in Western Europe and North America, and Bhutan, Fiji, and Myanmar are outliers in Asia and Pacific. All these outlier countries have a considerably higher level of ethnic exclusion than the other countries in their respective regions.

Figure 2: Box plot of exclusion by social group (VDEM) by regions of the world in 2006



Source: author's construction.

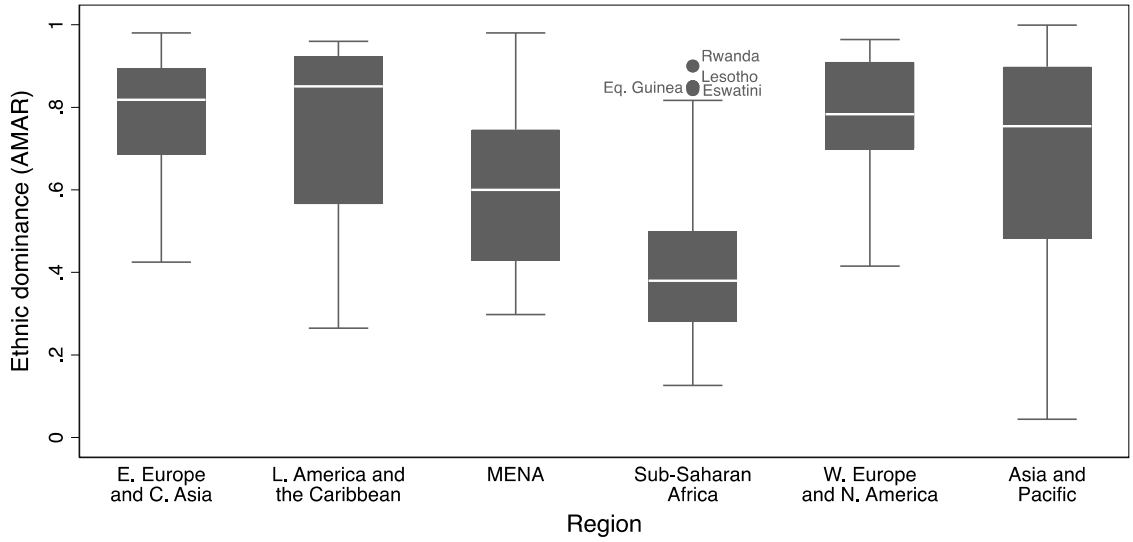
Figure 3: Box plot of ethnic exclusion (EPR) by regions of the world in 2006



Source: author's construction.

Figure 4 shows the distribution of ethnic dominance (AMAR GP) in the world. This time, SSA is clearly different from the other regions. According to AMAR GP, most countries in SSA have demographically small dominant ethnic groups, at least in comparison with the other regions. In SSA, the median size of the largest ethnic groups is less than 40 per cent of the total population within a country. In the other regions of the world, in contrast, the median size of the largest ethnic group is at least 75 per cent of the total population, with the exception of MENA, where the median size of the largest ethnic group is around 60 per cent of the total population. Nevertheless, there are some outlier countries in SSA—Equatorial Guinea, Eswatini, Lesotho, and Rwanda—in which the dominant ethnic group is considerably larger than in the rest of the region. In Asia and Pacific, data is particularly spread throughout the scale, which means that the region has both countries with small and countries with large ethnically dominant groups.

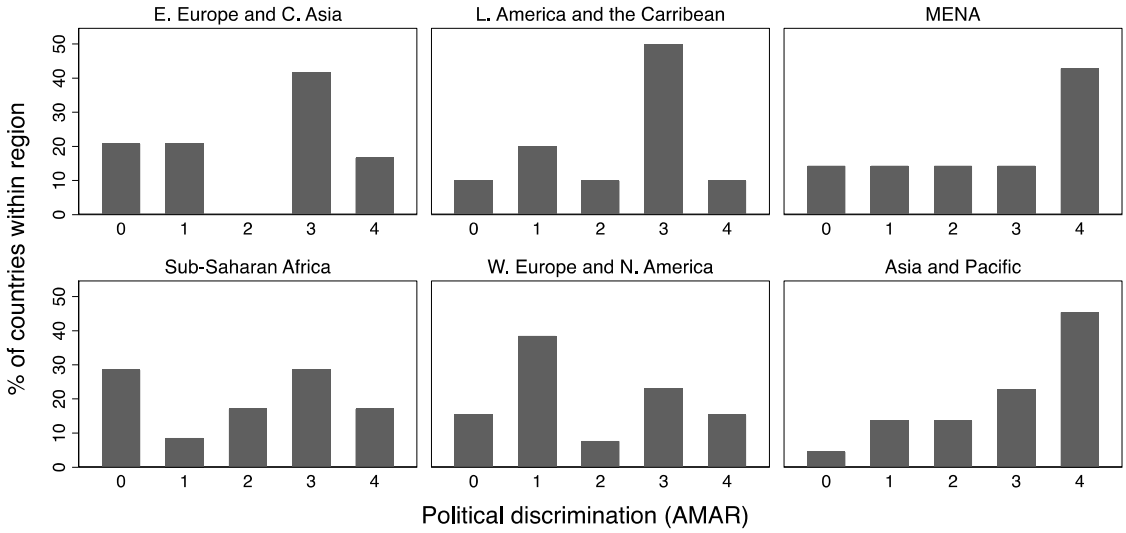
Figure 4: Box plot of ethnic dominance (AMAR GP) by regions of the world in 2006



Source: author's construction.

Figure 5 shows the frequencies of countries (as a percentage of total countries within a region) according to level of political discrimination (AMAR PDIS) by regions of the world. MENA and Asia and Pacific have many countries in which some ethnic groups face high discrimination. To be specific, over 40 per cent of the countries in these two regions have at least one ethnic group with the highest possible level of political discrimination. More than half of the rated countries in all regions except Western Europe and North America and SSA have ethnic groups at least at the second-highest level of discrimination. Interestingly, SSA has comparatively the largest share of countries (29 per cent) with no discriminated ethnic groups.

Figure 5: Bar plot of political discrimination (AMAR PDIS) by regions of the world in 2006



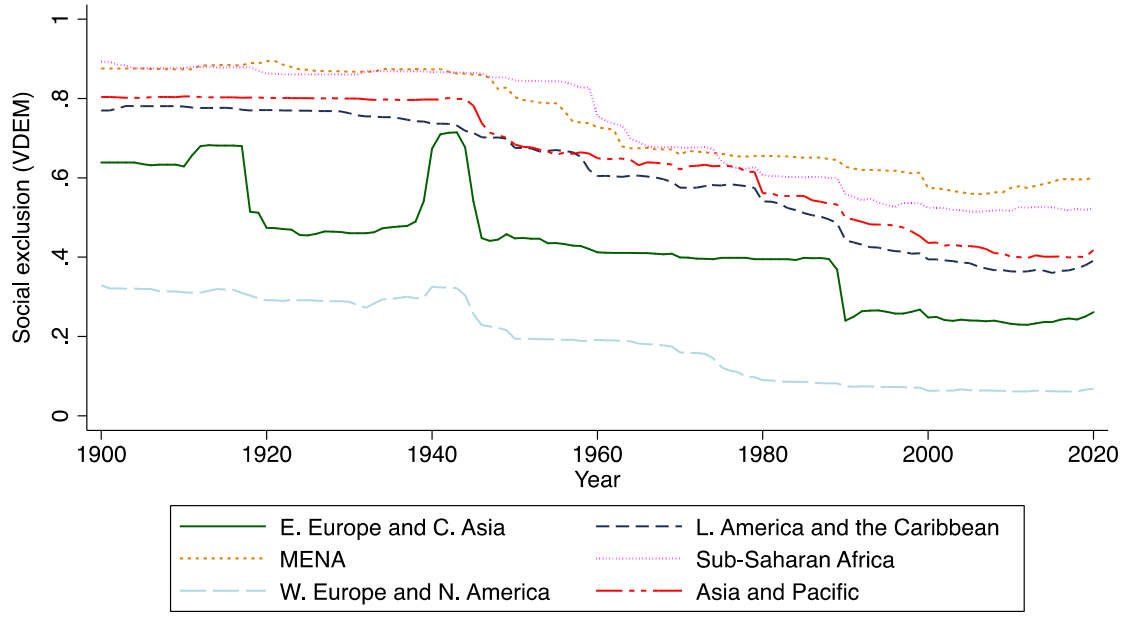
Source: author's construction.

If the above box plots and bar plots show a static picture of ethnicity in the most recent year of common observations, we can take advantage of the time-series dimension of some of the measures to explore the evolution of our different measures of ethnicity in different regions of the

world with line plots. Needless to say, only measures that change over time can provide such information. Hence, the measures of ethnicity that are time-invariant are excluded from the time-series analysis.

Figure 6 shows the evolution of ethnic exclusion in the world's regions as measured by VDEM. The index extends more than a hundred years back in time (to 1900), so it provides a particularly long view of the evolution of ethnic exclusion. In general, levels of ethnic exclusion have decreased in all regions since 1900, yet the relative positions of different regions have remained more or less the same during the observed period. From the first year of available observations until today, SSA and MENA have had the highest levels of ethnic exclusion, whereas Western Europe and North America has had the lowest levels. Eastern Europe and Central Asia, on the other hand, has experienced some sharp increases and decreases in ethnic exclusion: dramatic decreases after World War I and World War II and in the aftermath of the collapse of the Soviet Union, and a dramatic increase at the beginning of World War II.

Figure 6: Evolution of exclusion by social group in the world (VDEM)

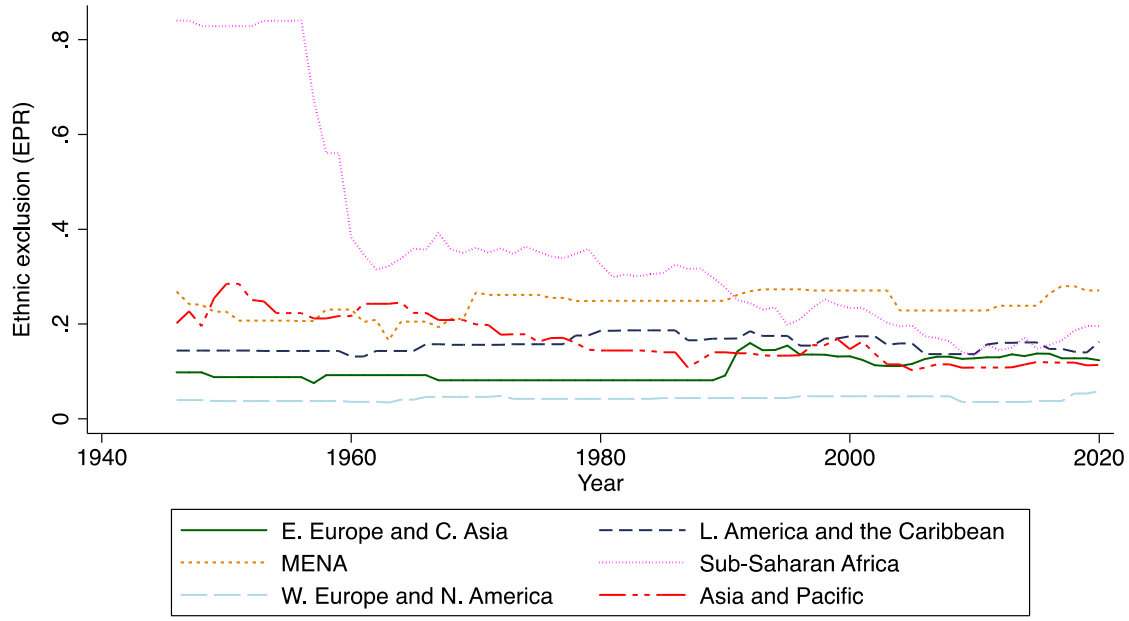


Source: author's construction.

Figure 7 illustrates how the level of ethnic exclusion (EPR) has changed in different regions of the world since 1946. As with VDEM, SSA and MENA had the highest levels of ethnic exclusion in 1946 and continue to be the most ethnically excluded regions in the world today. However, EPR records a massive reduction in the average level of exclusion in SSA around the late 1950s. Western Europe and North America has had the lowest levels of exclusion throughout the period. In contrast to the positive picture painted by VDEM, however, according to EPR, Eastern Europe and Central Asia experienced a sharp increase in ethnic exclusion around the years of the collapse of the Soviet Union. Furthermore, since the first year of available data, according to EPR, ethnic exclusion has also increased in MENA, Latin America and the Caribbean, and Western Europe and North America.

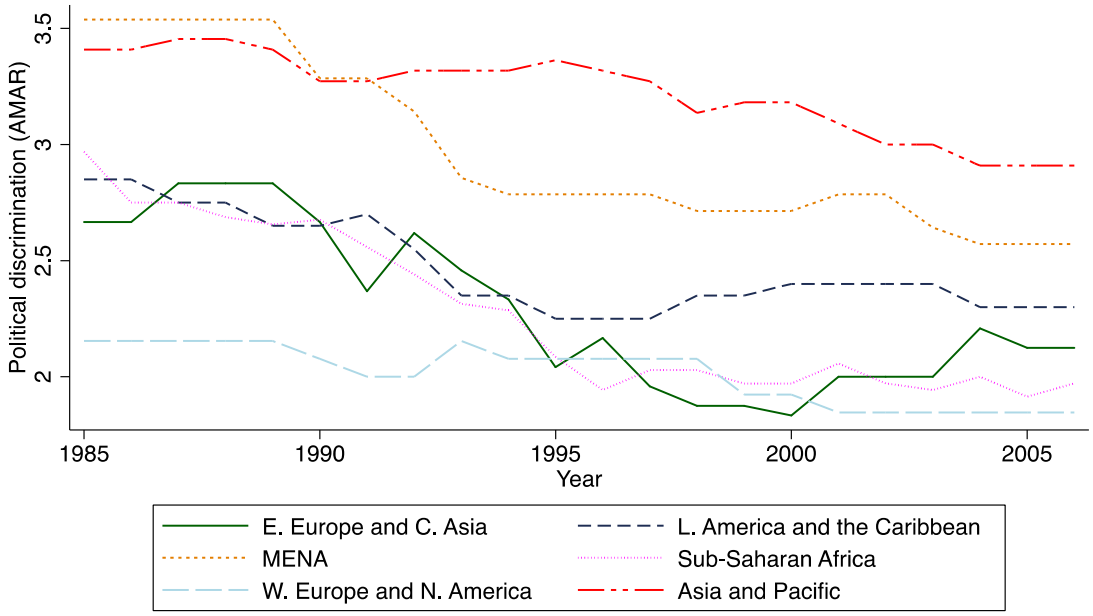
Figure 8 illustrates the evolution of political discrimination (AMAR PDIS) in the world's regions. In this case, data availability allows us to analyse only the period between 1985 and 2006[3].

Figure 7: Evolution of ethnic exclusion in the world (EPR)



Source: author's construction.

Figure 8: Evolution of political discrimination in the world (AMAR PDIS)



Source: author's construction.

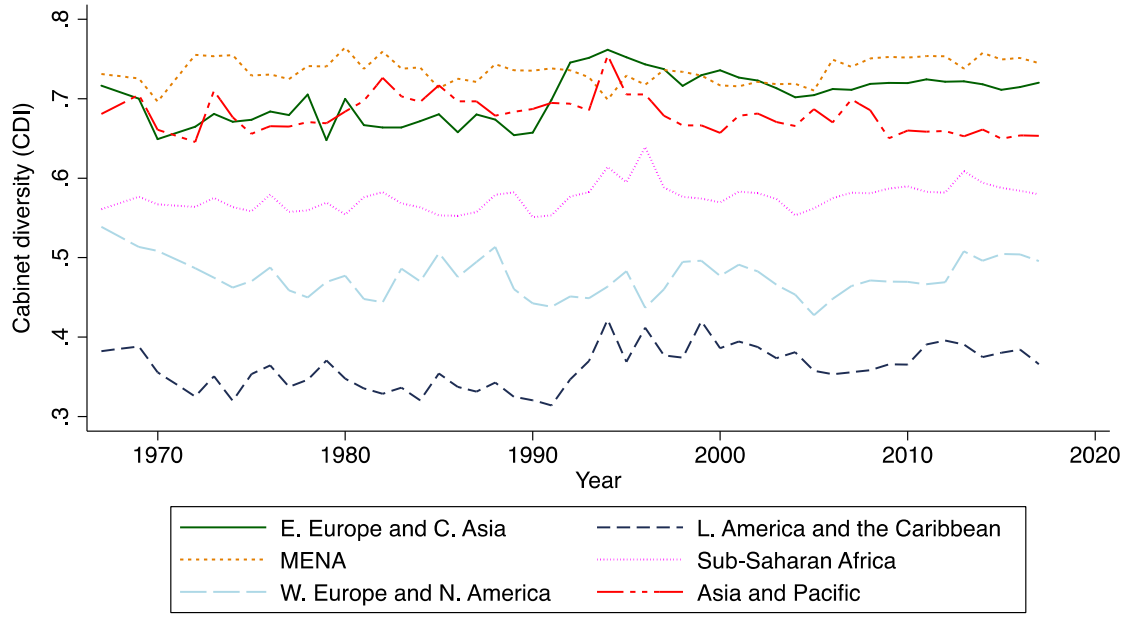
In spite of the relatively short period, the graph shows that overall there has been a clear downward trend and an evident drop in levels of political discrimination along ethnic lines in all regions of

[3] The first available year in AMAR PDIS is 1980, but yearly scores are available only from 1985 onwards.

the world. In contrast to the picture painted by VDEM and EPR, according to AMAR PDIS, Asia and Pacific has been the region with the highest levels of political discrimination from the 1990s onwards. Since the mid-1990s, SSA, Western Europe and North America, and Eastern Europe and Central Asia have in contrast been the regions with the lowest average level of political discrimination. According to AMAR PDIS, Eastern Europe and Central Asia is the only region where political discrimination has increased relatively steadily in since 2000.

Figure 9 shows the evolution of cabinet diversity (CDI) in the world's regions. There seems to be no clear trend or pattern in any region, except that on average cabinet diversity has been relatively volatile over time. CDI has seen some sharp changes, such as the increases of diversity in Latin America and the Caribbean and Eastern Europe and Central Asia in the early 1990s, but it has not seen any persistent increases or decreases over time. Latin America and the Caribbean had the least diverse cabinets in late 1960s and continues to have the least diverse cabinets in 2017, which is the most recent year of available observations. MENA had the most diverse cabinets in the late 1960s and continues to have them in 2017.

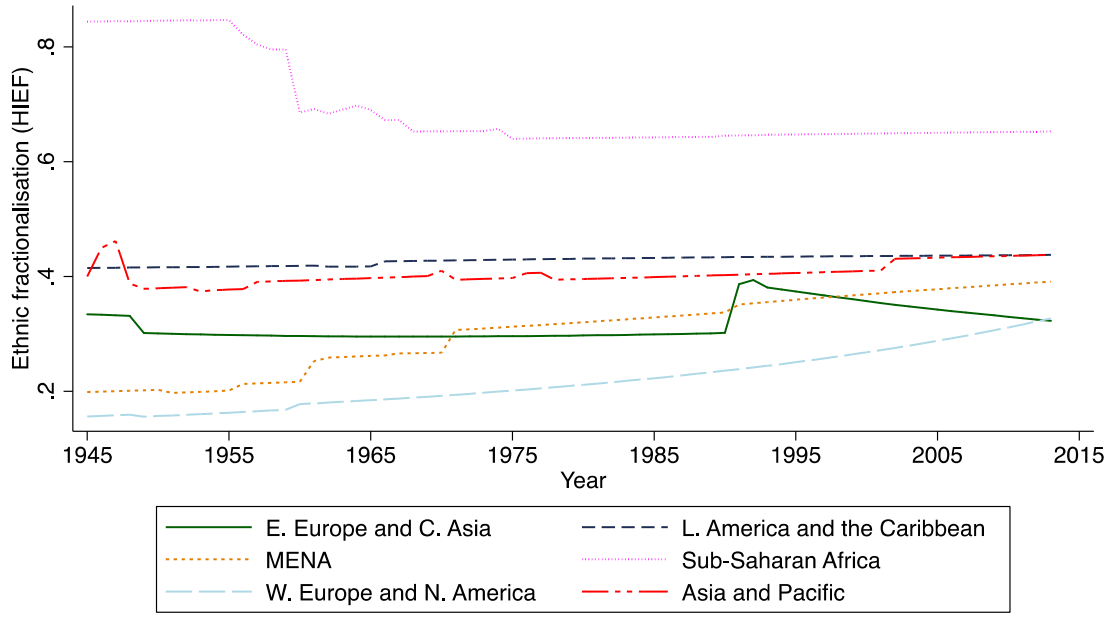
Figure 9: Evolution of ethnic diversity in the world (CDI)



Source: author's construction.

Figure 10 shows the evolution of ethnic fractionalization (HIEF) in the world's regions. According to HIEF, SSA experienced a substantial reduction in ethnic diversity in the 1950s and the 1960s. A similar decrease is recorded by VDEM and EPR, but not by CDI. Conversely, with HIEF we can notice clear and persistent increases in ethnic diversity in Western Europe and North America and MENA. Western Europe and North America, Eastern Europe and Central Asia, and MENA have been the three least heterogeneous regions since 1945, whereas SSA has been by far the most ethnically heterogeneous region throughout the period of available data.

Figure 10: Evolution of ethnic fractionalization in the world (HIEF)



Source: author's construction.

Last, as already discussed, my study focuses briefly on SSA, which is arguably the region of the world where ethnicity-based questions are currently the most salient. First, I explore the convergence of the measures in SSA. Then, I examine the longitudinal evolution of the scores in selected SSA countries. Correlations of measures in SSA (Table 4) show many similarities with the previously run correlations concerning all countries in the world. As with the broader global sample of countries, CDI is not significantly related to any of the measures of ethnicity, N* and EPR are strongly correlated (Pearson's r: 0.85; Spearman's rho: 0.86), and HIEF and EF are relatively strongly correlated (Pearson's r: 0.70; Spearman's rho: 0.56). Yet, there are some major differences too. First, ethnic polarization (POL), which was significantly related to all our main measures of ethnic dominance and exclusion in the global sample, is not related to our main measures of ethnic dominance and exclusion in the SSA sample of countries. Second, AMAR GP and MAR GP, which were inversely related in the global sample, are positively related in the SSA sample, although the coefficient is statistically significant only with Pearson's method.

Interestingly, when it comes to our four main measures of ethnic dominance and exclusion, VDEM and EPR (Pearson's r: 0.56; Spearman's rho: 0.55), VDEM and AMAR PDIS (Pearson's r: 0.36; Spearman's rho: 0.39), and EPR and AMAR PDIS (Pearson's r: 0.45; Spearman's rho: 0.37) are more strongly (and significantly) related to each another in the sample of SSA countries than in the global sample. Conversely, AMAR GP, which was strongly (and inversely) related to both VDEM and EPR in the global sample, is not significantly related to either of the measures in the SSA sample.

Afrobarometer's measure of ethnic discrimination (ABR), which was not analysed in the global sample of countries because it covers only 34 African countries, is positively and significantly related to some our principal measures of ethnic dominance and exclusion. ABR is moderately related to VDEM (Pearson's r: 0.42; Spearman's rho: 0.45) and EPR (Pearson's r: 0.42; Spearman's rho: 0.50), but not significantly related to AMAR GP and AMAR PDIS. These comparatively strong linkages between ABR, VDEM, and EPR indicate that citizens' perceptions of individual ethnic discrimination, expert surveys on country-level ethnic exclusion, and expert coded group-level ethnic exclusion match quite well.

Table 4: Correlations between measures of ethnicity in sub-Saharan Africa

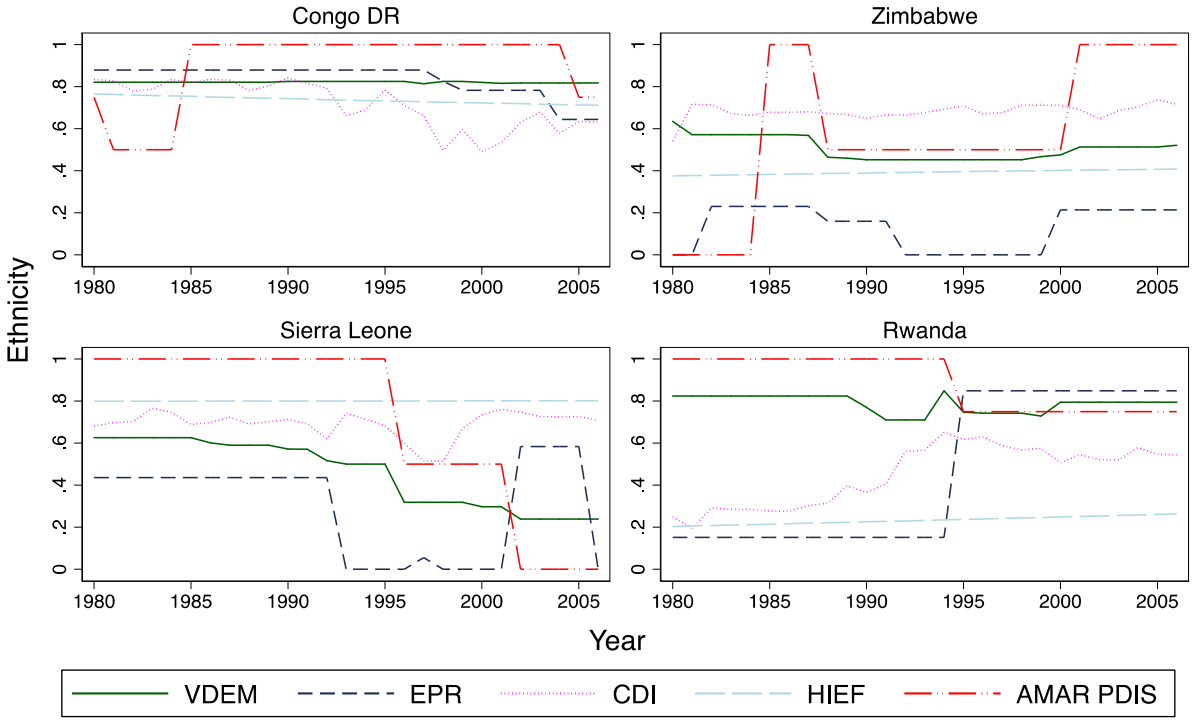
	VDEM	EPR	CDI	ABR	MAR GP	AMAR GP	AMAR PDIS	POL	HIEF	EH	EF	N*	FSI
VDEM	1.00	0.55***	0.18	0.45**	0.38*	-0.03	0.39*	0.06	0.13	0.35*	0.30*	0.33	0.50***
	(49)	(46)	(48)	(32)	(30)	(45)	(35)	(42)	(43)	(48)	(47)	(36)	(48)
EPR	0.56***	1.00	-0.06	0.50**	0.38*	-0.16	0.37*	0.19	0.37*	0.28	0.24	0.86***	0.56***
	(46)	(46)	(46)	(31)	(30)	(45)	(35)	(41)	(43)	(46)	(46)	(36)	(46)
CDI	0.22	-0.005	1.00	0.11	-0.32	0.05	-0.05	0.13	-0.12	0.19	-0.12	-0.25	-0.08
	(48)	(46)	(48)	(32)	(30)	(45)	(35)	(42)	(43)	(48)	(47)	(36)	(48)
ABR	0.42*	0.42*	0.16	1.00	0.55*	-0.28	0.25	0.11	-0.56**	0.43*	0.51**	0.27	0.19
	(32)	(31)	(32)	(32)	(19)	(30)	(24)	(29)	(29)	(32)	(31)	(22)	(32)
MAR GP	0.36	0.37*	-0.21	0.54*	1.00	0.09	0.05	0.05	0.14	0.12	0.10	0.58**	0.28
	(30)	(30)	(30)	(19)	(30)	(30)	(39)	(27)	(29)	(30)	(30)	(24)	(30)
AMAR GP	-0.02	-0.20	0.05	-0.35	0.37*	1.00	0.08	0.10	-0.52***	-0.37*	-0.73***	-0.14	-0.06
	(45)	(45)	(54)	(30)	(30)	(45)	(35)	(40)	(42)	(45)	(45)	(36)	(45)
AMAR PDIS	0.36*	0.45**	-0.06	0.23	0.10	0.10	1.00	0.16	-0.08	0.18	0.05	0.51**	0.52**
	(35)	(35)	(35)	(24)	(30)	(35)	(35)	(31)	(33)	(35)	(35)	(27)	(35)
POL	0.06	0.24	0.02	0.12	0.02	0.05	0.30	1.00	0.05	0.31*	-0.05	0.25	0.06
	(24)	(41)	(42)	(29)	(27)	(40)	(31)	(42)	(39)	(42)	(42)	(32)	(42)
HIEF	0.07	0.35*	-0.10	0.57**	-0.28	-0.64***	-0.04	0.34*	1.00	0.44**	0.56***	0.10	0.13
	(43)	(43)	(43)	(29)	(29)	(42)	(33)	(39)	(43)	(43)	(43)	(34)	(43)
EH	0.32*	0.33*	0.15	0.45*	-0.16	-0.53***	0.10	0.35*	0.53***	1.00	0.56***	0.14	0.38**
	(48)	(46)	(48)	(32)	(30)	(45)	(35)	(42)	(43)	(48)	(47)	(36)	(48)
EF	0.26	0.27	-0.10	0.48**	-0.36*	-0.73***	0.05	0.21	0.70***	0.69***	1.00	0.20	0.29*
	(47)	(46)	(47)	(31)	(30)	(45)	(35)	(42)	(43)	(47)	(47)	(36)	(47)
N*	0.33*	0.85***	-0.21	0.29	0.44*	-0.12	0.49**	0.25	0.003	0.05	0.07	1.00	0.40*
	(36)	(36)	(36)	(22)	(24)	(36)	(27)	(32)	(34)	(36)	(36)	(36)	(36)
FSI	0.46**	0.58***	-0.07	0.19	0.26	-0.04	0.51**	0.08	0.14	0.36*	0.26	0.38*	1.00
	(48)	(46)	(32)	(32)	(30)	(45)	(35)	(42)	(43)	(48)	(47)	(36)	(48)

Note: Pearson's correlation coefficients (bottom-left quadrant) and Spearman's correlation coefficients (upper-right quadrant) from 1980 to 2006 (average). Number of observations in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Source: author's construction.

The evolution of ethnicity is examined with line plots in selected SSA countries (Figure 11). These line plots show without doubt that different measures of ethnicity capture different aspects of dominance and exclusion within a country.

Figure 11. Different measures of ethnicity in selected countries in sub-Saharan Africa from 1980 to 2006



Source: author's construction.

In Congo DR (upper left panel), according to VDEM, the level of ethnic exclusion remained more or less the same from 1980 to 2006. HIEF records only a slight decrease, but no major changes in ethnic fractionalization. The other measures, however, portray a completely different picture. CDI indicates that the ethnic diversity of cabinet ministers has been relatively inconsistent over time, but has decreased overall from its highest values in the 1980s. In contrast, AMAR PDIS shows that political discrimination increased in the 1980s and decreased after the advent of the new millennium. A similar decrease in the level of ethnic exclusion is recorded by EPR, starting from the late 1990s. Overall, whether it is about diversity, exclusion, or discrimination, all measures agree that the 'level of ethnicity' in Congo DR has been relatively high.

In Zimbabwe (upper right panel), there are more disagreements among our measures of ethnicity. According to VDEM, exclusion by social group has somewhat decreased since the early 1980s, whereas according to all the other measures, there has been an increase in ethnicity during the observed period. For HIEF, this increase is negligible. For AMAR PDIS, in contrast, it is substantial. Cabinet diversity (CDI) has increased as well, despite being relatively stable after an important increase in early 1980s. Likewise, EPR records increases in ethnic exclusion in the early 1980s and late 1990s.

In Sierra Leone (lower left panel), according to HIEF, there were no changes in the level of ethnic fractionalization from 1980 to 2006. In contrast, VDEM and AMAR PDIS portray a clear decrease in exclusion and discrimination, respectively, particularly from 1995 onwards. The level of ethnic diversity of cabinet members (CDI), in contrast, saw a minor decrease in the mid-1990s but then increased to its previous levels at the end of the 1990s. Similarly, according to EPR, there were no

(or almost no) ethnically excluded groups from the early 1990s to early 2000s, but then an increase to the highest levels ever from 2001 to 2002.

In Rwanda (lower right panel), we notice some increases and some decreases in ethnicity, depending on the measure. The level of political discrimination (AMAR PDIS) along ethnic lines has decreased since the highest values from 1980 to mid-1990s but remains relatively high (at the second highest level). On the contrary, the amount of politically excluded ethnic population (EPR) increased substantially in the mid-1990s. The ethnic diversity in cabinets (CDI) saw a relatively persistent increase from 1980 to the mid-1990s and ethnic fractionalization (HIEF) slightly increased throughout the period of analysis. According to VDEM, exclusion by social group saw some volatility but no major persistent changes over time from 1980 to 2006.

5 Concluding remarks

The study at hand offers one of the first comparative analyses of measures of ethnic dominance and exclusion. In this study, I have reviewed and compared some of the most relevant measures of ethnicity in terms of both content and empirical substance. I have scrutinized the convergence of the measures with correlations, looked into the similarities and differences between the measures through the descriptive characteristics of the data, and examined the longitudinal evolution of ethnicity in different regions and countries using different measures. First, I have focused on all countries in the world with available data. Second, I have divided the data by regions and focused on the six macro-regions of the world. Third, I have shifted my focus to sub-Saharan Africa and compared the measures in arguably the most topical region in the world for the field of ethnicity.

Generally speaking, we can conclude that measures of ethnic dominance and exclusion are not similar to each other. Even if our four main measures—VDEM, EPR, AMAR PDIS, and AMAR GP—supposedly capture more or less the same concept, the measures are at best moderately convergent. No expert would expect these measures to be completely equivalent, yet my findings suggest that the different approaches of these measures to ethnic dominance and exclusion lead to surprisingly divergent country-level information. According to my results, VDEM and EPR are the most similar pair of measures of ethnic dominance and exclusion. They are more strongly correlated than the other measures and depict many of the world's regions in a relatively similar way. AMAR GP and AMAR PDIS, on the other hand, are the least similar pair of measures, and in fact do not seem to be related at all. Considering these empirical differences, scholars using measures of ethnic dominance and exclusion should not rely on an 'any measure will do' approach. Common measures of ethnic dominance and exclusion might have similar conceptual underpinnings, but they are likely to lead to different conclusions.

The study at hand has also compared these four measures of ethnic dominance and exclusion with other measures of the broader concept of ethnicity. The findings indicate that, overall, VDEM and EPR are positively related to commonly used measures of ethnic fractionalization, polarization, and heterogeneity, whereas AMAR GP is negatively related to common measures of ethnic fractionalization, polarization, and heterogeneity. AMAR PDIS, in contrast, is not significantly related to these common measures of ethnic diversity in a global sample of countries. What is interesting, however, is that AMAR PDIS is more strongly related to several measures of ethnicity in sub-Saharan African countries than globally. Similarly, the relationships between measures of ethnic dominance and exclusion and some of the measures of the broader concept of ethnicity seem to depend on the region under scrutiny. In particular, ethnic polarization (POL) is more weakly related to measures of ethnic dominance and exclusion in SSA countries than globally.

These notable differences in the results between the sample of global countries and the sample of SSA countries call for further investigation. The preliminary findings of this study suggest that the convergence of some of the measures might depend on the context.

References

Afrobarometer (2020). 'Data codebook for a Round 7 Afrobarometer Survey'. East Lansing, MI: Michigan State University.

Alesina, A., and E. La Ferrara (2005). 'Ethnic diversity and economic performance'. *Journal of Economic Literature*, 43(3): 762–800. https://doi.org/10.1257/002205105774431243

Alesina, A., R. Baqir, and W. Easterly (1999). 'Public goods and ethnic divisions'. *Quarterly Journal of Economics*, 114(4): 1243–84. https://doi.org/10.1162/003355399556269

Alesina, A., A. Devleeschauwer, W. Easterly, and S. Kurlat (2003). 'Fractionalization'. *Journal of Economic Growth*, 8(2): 155–94. https://doi.org/10.1023/A:1024471506938

Baldwin, K., and J.D. Huber (2010). 'Economic versus cultural differences: forms of ethnic diversity and public goods provision'. *American Political Science Review*, 104(4): 644–62. https://doi.org/10.1017/S0003055410000419

Bates, R.H. (2000). 'Ethnicity and development in Africa: a reappraisal'. *American Economic Review*, 90(2): 131–34. https://doi.org/10.1257/aer.90.2.131

Bates, R.H. (2008). 'State failure'. *Annual Review of Political Science*, 11: 1–12. https://doi.org/10.1146/annurev.polisci.11.060606.132017

Birnir, J.K., D.D. Laitin, J. Wilkenfeld, D.M. Waguespack, A.S. Hultquist, and T.R. Gurr (2017). 'Introducing the AMAR (All Minorities at Risk) data'. *Journal of Conflict Resolution*, 62(1): 203–26. https://doi.org/10.1177/0022002717719974

Canelas, C., and R.M. Gisselquist (2018). 'Horizontal inequality as an outcome'. *Oxford Development Studies*, 46(3): 305–24. https://doi.org/10.1080/13600818.2018.1508565

Canelas, C., and R.M. Gisselquist (2019). 'Horizontal inequality and data challenges'. *Social Indicators Research*, 143(1): 157–72. https://doi.org/10.1007/s11205–018–1932–1

Caprioli, M., and P.F. Trumbore (2003). 'Ethnic discrimination and interstate violence: testing the international impact of domestic behavior'. *Journal of Peace Research*, 40(1): 5–23. https://doi.org/10.1177/0022343303040001842

Cederman, L.E., and L. Girardin (2007). 'Beyond fractionalization: mapping ethnicity onto nationalist insurgencies'. *American Political Science Review*, 101(1): 173–85. https://doi.org/10.1017/S0003055407070086

Cederman, L.E., A. Wimmer, and B. Min (2010). 'Why do ethnic groups rebel? new data and analysis'. *World Politics*, 62(1): 87–119. https://doi.org/10.1017/S0043887109990219

Chandra, K., and S. Wilkinson (2008). 'Measuring the effect of "ethnicity"'. *Comparative Political Studies*, 41(4–5): 515–63. https://doi.org/10.1177/0010414007313240

Charron, N. (2009). 'Government quality and vertical power–sharing in fractionalized states'. *Publius*, 39(4): 585–605. https://doi.org/10.1093/publius/pjp015

Collier, P. (2001). 'Implications of ethnic diversity'. *Economic Policy*, 16(32): 127–66. https://doi.org/10.1111/1468–0327.00072

Collier, P., and A. Hoeffler (2004). 'Greed and grievance in civil war'. *Oxford Economic Papers*, 56(4): 563–95. https://doi.org/10.1093/oep/gpf064

Coppedge, M., et al. (2021). 'V–Dem Codebook v11.1'. Varieties of Democracy (V–Dem) Project. https://doi.org/10.2139/ssrn.3802627

Desmet, K., I. Ortuno–Ortín, and S. Weber (2009). 'Linguistic diversity and redistribution'. *Journal of the European Economic Association*, 7(6): 1291–318. https://doi.org/10.1162/JEEA.2009.7.6.1291

Dinesen, P.T., and K.M. Sønderskov (2015). 'Ethnic diversity and social trust: evidence from the micro–context'. *American Sociological Review*, 80(3): 550–73. https://doi.org/10.1177/0003122415577989

Drazanova, L. (2020). 'Introducing the Historical Index of Ethnic Fractionalization (HIEF) dataset: accounting for longitudinal changes in ethnic diversity'. *Journal of Open Humanities Data*, 6: 1–8. https://doi.org/10.5334/johd.16

Easterly, W., and R. Levine (1997). 'Africa's growth tragedy: policies and ethnic divisions'. *Quarterly Journal of Economics*, 112(4): 1203–50. https://doi.org/10.1162/003355300555466

Ellingsen, T. (2000). 'Colorful community or ethnic witches' brew? Multiethnicity and domestic conflict during and after the Cold War'. *The Journal of Conflict Resolution*, 44(2): 228–49. https://doi.org/10.1177/0022002700044002004

Fearon, J.D. (2003). 'Ethnic and cultural diversity: keys to power'. *Journal of Economic Growth*, 8(2): 195–222. https://doi.org/10.1023/A:1024419522867

Fearon, J.D., and D.D. Laitin (2003). 'Ethnicity and civil war'. *American Political Science Review*, 97(1): 75–90. https://doi.org/10.1017/S0003055403000534

Franck, R., and I. Rainer (2012). 'Does the leader's ethnicity matter? Ethnic favoritism, education, and health in sub-Saharan Africa'. *American Political Science Review*, 106(2): 294–325. https://doi.org/10.1017/S0003055412000172

Fund for Peace (2017). 'Fragile States Index methodology'. Washington, DC: The Fund for Peace.

Gerring, J., M. Hoffman, and D. Zarecki (2018). 'The diverse effects of diversity on democracy'. *British Journal of Political Science*, 48(2): 283–314. https://doi.org/10.1017/S000712341600003X

Gisselquist, R.M., S. Leiderer, and M. Niño–Zarazúa (2016). 'Ethnic heterogeneity and public goods provision in Zambia: evidence of a subnational "diversity dividend"'. *World Development*, 78: 308–23. https://doi.org/10.1016/j.worlddev.2015.10.018

Heger, L.L. (2015). 'Votes and violence: pursuing terrorism while navigating politics'. *Journal of Peace Research*, 52(1): 32–45. https://doi.org/10.1177/0022343314552984

Heger, L., and I. Salehyan (2007). 'Ruthless rulers: coalition size and the severity of civil conflict'. *International Studies Quarterly*, 51(2): 385–403. https://doi.org/10.1111/j.1468–2478.2007.00456.x

Herbst, J. (2014). *States and power in Africa: comparative lessons in authority and control*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400852321

Horowitz, D.L. (1985). *Ethnic groups in conflict*. Berkeley, CA: University of California Press.

Jackman, R.W. (1978). 'The predictability of coups d'état: a model with African data'. *American Political Science Review*, 72(4): 1262–75. https://doi.org/10.2307/1954538

Jensen, C., and S.E. Skaaning (2012). 'Modernization, ethnic fractionalization, and democracy'. *Democratization*, 19(6): 1117–37. https://doi.org/10.1080/13510347.2011.619777

Marquardt, K.L., and Y.M. Herrera (2015). 'Ethnicity as a variable: an assessment of measures and data sets of ethnicity and related identities'. *Social Science Quarterly*, 96(3): 689–716. https://doi.org/10.1111/ssqu.12187

McDoom, O.S., and R.M. Gisselquist (2016). 'The measurement of ethnic and religious divisions: spatial, temporal, and categorical dimensions with evidence from Mindanao, the Philippines'. *Social Indicators Research*, 129(2): 863–91. https://doi.org/10.1007/s11205–015–1145–9

Minorities at Risk (2009). 'Minorities at Risk dataset'. College Park, MD: Center for International Development and Conflict Management. Available at: www.mar.umd.edu (accessed 15 February 2022).

Montalvo, J.G., and M. Reynal-Querol (2005). 'Ethnic polarization, potential conflict, and civil wars'. *American Economic Review*, 95(3): 796–816. https://doi.org/10.1257/0002828054201468

Mudde, C., and A. Schedler (2010). 'Introduction: rational data choice'. *Political Research Quarterly*, 63(2): 410–16. https://doi.org/10.1177/1065912910371286

O'Brochta, W. (2020). 'Ethnic diversity in central government cabinets'. *Politics, Groups, and Identities*, online. https://doi.org/10.1080/21565503.2020.1756360

Piazza, J.A. (2012). 'Types of minority discrimination and terrorism'. *Conflict Management and Peace Science*, 29(5): 521–46. https://doi.org/10.1177/0738894212456940

Piazza, J.A. (202)1. 'The impact of rural–urban economic disparities on terrorist organizations' survival and attacks'. *Defence and Peace Economics*, online. https://doi.org/10.1080/10242694.2021.1916684

Posner, D.N. (2004). 'Measuring ethnic fractionalization in Africa'. *American Journal of Political Science*, 48(4): 849–63. https://doi.org/10.1111/j.0092–5853.2004.00105.x

Putnam, R.D. (2007). 'E pluribus unum: diversity and community in the twenty–first century' (2006 Johan Skytte Prize Lecture). *Scandinavian Political Studies*, 30(2): 137–74. https://doi.org/10.1111/j.1467–9477.2007.00176.x

Reynal-Querol, M. (2002). 'Political systems, stability and civil wars'. *Defence and Peace Economics*, 13(6): 465–83. https://doi.org/10.1080/10242690214332

Rørbæk, L.L. (2019). 'Ethnic exclusion and civil resistance campaigns: opting for nonviolent or violent tactics?' *Terrorism and Political Violence*, 31(3): 475–93. https://doi.org/10.1080/09546553.2016.1233872

Rørbæk, L.L., and A.T. Knudsen (2017). 'Maintaining ethnic dominance: diversity, power, and violent repression'. *Conflict Management and Peace Science*, 34(6): 640–59. https://doi.org/10.1177/0738894215612996

Rudolfsen, I. (2017). 'State capacity, inequality and inter-group violence in sub-Saharan Africa: 1989–2011'. *Civil Wars*, 19(2): 118–45. https://doi.org/10.1080/13698249.2017.1345541

Ruedin, D. (2009). 'Ethnic group representation in a cross-national comparison'. *Journal of Legislative Studies*, 15(4): 335–54. https://doi.org/10.1080/13572330903302448

Uzonyi, G., N.K. Kim, N. Jahanbani, and V. Asal (2021). 'Genocide, politicide, and the prospects of democratization since 1900'. *Journal of Conflict Resolution*, 65(9): 1521–50. https://doi.org/10.1177/00220027211000445

Vanhanen, T. (1999). 'Domestic ethnic conflict and ethnic nepotism: a comparative analysis'. *Journal of Peace Research*, 36(1): 55–73. https://doi.org/10.1177/0022343399036001004

Vogt, M., N.C. Bormann, S. Rüegger, L.E. Cederman, P. Hunziker, and L. Girardin (2015). 'Integrating data on ethnicity, geography, and conflict: the ethnic power relations data set family'. *Journal of Conflict Resolution*, 59(7): 1327–42. https://doi.org/10.1177/0022002715591215

Wimmer, A., L.-E. Cederman, and B. Min (2009). 'Ethnic politics and armed conflict: a configurational analysis of a new global data set'. *American Sociological Review*, 74(2): 316–37. https://doi.org/10.1177/000312240907400208