UNITED NATIONS
UNIVERSITY
UNU-WIDER

Project overview note

Group-based inequalities: patterns and trends within and across countries

A UNU-WIDER research initiative

Rachel M. Gisselquist[1], 4 December 2015

1 Project overview

Addressing inequality, both vertical and horizontal, is a core topic for the post-2015 development agenda. Yet, further work is needed to understand horizontal or group-based inequalities in particular, including how they compare across countries, within countries, and over time. This research initiative, supported under UNU-WIDER's 2014-2018 research programme, addresses the measurement of horizontal inequalities in developing countries[2]. It focuses on inequalities among ethnic, racial, religious, and communal groups[3]. This work is an integral part of a larger research effort on the politics of group-based inequalities, which considers causes, correlates, and possibilities for change.

The project aims to build a comprehensive picture of variation in such group-based inequalities across countries at the national level, as well sub-nationally and diachronically for a selected set of developing countries. A core set of 12 countries of focus has been selected on the basis of population size, a minimal degree of ethnic diversity, developing country status, and geographic region: this includes the ten largest developing countries in the world with ethnic fractionalization values greater than 0.20 (India, Indonesia, Brazil, Pakistan, Nigeria, Mexico, the Philippines, Ethiopia, Vietnam, and Iran), plus the next two largest sub-Saharan African (SSA) countries that meet the same criteria (Democratic Republic of the Congo, and South Africa)[4]. The additional SSA countries are included given the particular relevance of SSA to research on ethnic diversity and development (e.g., Easterly & Levine, 1997; Posner, 2004) and the related observation in the literature that ethnic diversity and social structure vary notably across world geographic regions (see, e.g., Alesina et al., 2003; Gisselquist, 2013). This gives a core sample of four countries from SSA, five from Asia, two from Latin America, and one from the Middle East and North Africa. Several additional country case studies may be added subsequently to complement the larger research effort.

In order to support emerging research into group-based inequalities in developing countries, we are aiming to commission most case studies from early-career scholars from each country or region. By facilitating communication and exchange among project contributors – including through an authors' workshop in 2016 – the project also aims to facilitate opportunities for collaboration among early career scholars working on related topics in diverse countries.

The focal points for this research initiative are Dr. Carla Canelas and Dr. Rachel Gisselquist, both UNU-WIDER Research Fellows.

Case studies are commissioned to provide comparable information across all countries included in the project. Project participants should thus familiarize themselves with the detailed guidelines and technical notes provided below in Sections 2-4. The project focal points will be using in particular the "plan of action" memo and the first paper draft, as outlined below, to assess and ensure comparability across the case studies[5]. It is thus important that project contributors follow the guidelines carefully and submit work on schedule.

Assuming that each commissioned case study paper is of satisfactory quality, each paper will be published in the WIDER Working Paper Series, and also possibly included in a UNU-WIDER edited book or special issue on the topic. We will aim for the highest quality outlet for

1 Research Fellow, UNU-WIDER

2 This initiative is formally supported under the UNU-WIDER project component on 'The Political Implications of Group-based Inequalities: Measurement, Implications, and Possibilities for Change' in the 2014-2018 research programme: https://www.wider.unu.edu/project/disadvantaged-groups-and-social-mobility.

3 'Ethnic' as understood in this project refers to categories based on ascriptive attributes that are generally inherited at birth, including skin color, language, tribe, caste, religion, region, kinship, and other communal identity markers (Chandra, 2004; Horowitz, 1985; Htun, 2004). In other words, we adopt one standard definition in the literature on ethnic politics. Strictly speaking, racial, religious, and communal groups may be subsumed under "ethnic" according to this approach. They are mentioned separately here because the broader literature also highlights some ways in which these types of groups may be distinct. Contributors should be aware that the approach adopted here differs in several key ways from the way in which ethnicity is often treated in the economics literature, as well as in more popular writing, with implications for the project design. For instance, while more popular writing often assumes that individuals have a single, unambiguous, and stable ethnic identity, research in ethnic politics shows that this is not necessarily the case, underscoring the value of considering multiple types of ethnic division.

4 Population and developing country status are based on World Bank data. Fractionalization refers to the now standard index of ethno-linguistic fractionalization measure (see Taylor & Hudson, 1972) based on Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003)'s data on 'ethnic' and 'linguistic' fractionalization. Countries with fractionalization values greater than 0.20 of either type are included.

5 For instance, details regarding common coding of specific variables, templates for tables, or common Stata code may be shared.

the collection, taking into account the topic and quality of the papers.[6]

Payment for case studies will be made in two installments, as outlined below. **While payments are thus made prior to book/journal publication, we ask that contributors make a professional commitment to a peer-reviewed book or journal publication, which may require additional revisions after working paper release to meet submission guidelines and in response to a normal journal or book review process, and knowledge sharing activities**. A contributors' workshop will be organized in 2016 to facilitate discussions and paper revisions in preparation for submission.[7]

UNU-WIDER aims through its projects to support and contribute to better policymaking and research on development and thus will undertake various additional activities in support of the dissemination and discussion of findings with relevant policy and scholarly communities. Case study contributors may be invited to participate in some of these activities.

The next section of this document discusses the measurement of group-based inequality and other key measures. It also reviews key data sources and variables to be used in constructing these measures. Drawing on this discussion, Section 3 provides guidelines for the preparation of each case study. Section 4 summarizes key deliverables and timelines for submission.

2 Key measures

Although the literature provides an incomplete empirical picture of group-based inequalities within and across countries, discussion of measurement issues and techniques is relatively well-developed. In political science and sociology, the relationship between ethnicity and class, as well as that between different ethnic and social cleavages (e.g., whether "cross-cutting" or "reinforcing"), have long been core areas of inquiry (e.g., Lijphart, 1979; Lipset & Rokkan, 1967; McAll, 1990; Rae & Taylor, 1970). Likewise in economics, a considerable literature addresses horizontal inequality and its measurement, including comparison to vertical inequality (Kanbur, 2006).

This project thus focuses on empirics and the mapping of patterns within and across countries, drawing on existing measurement techniques. A fuller empirical picture of group-based inequalities in turn is expected to provide a basis for stronger analysis, theory-building, and theory-testing on the causes, correlates, and implications of group-based inequality.

Various approaches to the measurement of group-based inequality are developed in the literature, including innovative efforts to address gaps and weaknesses in the available socioeconomic data. Alesina, Michalopoulos, and Papaioannou (2012), for instance, combine data on nighttime luminosity along with historical ethnolinguistic group homelands to construct measures, while Cederman, Weidmann, and Gleditsch (2011) combine geocoded data on ethnic group settlement areas with spatial wealth estimates. In this research initiative, however, our focus is on using available socioeconomic data as reported by individuals and households in the census, supplemented by information from selected surveys.

Group-based inequality measures will be considered alongside four additional standard measures of ethnic diversity and vertical inequality as defined below.

2.1 Group-based inequality measures

Five measures of group-based inequality will be considered in the case studies. The first three are drawn from Frances Stewart's project on horizontal inequalities and conflict.

6 For information on WIDER special issue and book publications, see: http://www.wider.unu.edu/publications/books-and-journals/en_GB/books-and-journals/. The final decision on which outlets to approach rests with the project focal points and UNU-WIDER, but discussion is welcome. In connection with previous projects, Gisselquist has edited special issues of The ANNALS of the American Academy of Political and Social Science (November 2014), International Peacekeeping (August 2015), Public Administration and Development (August 2014, with Danielle Resnick), The Journal of Globalization and Development (June 2015, with Miguel Niño-Zarazúa), and Third World Quarterly (July 2015).

7 UNU-WIDER will cover basic costs according to UN travelrules, including economy airfare and a standard allowance for accomodation and meals. For papers with multiple authors, expenses for one author generally are covered.

Stewart, Brown, and Mancini (2010) explores measurement considerations and presents the case for using these measures over others:

1 **Group-weighted coefficient of variation (GCOV):**

$$GCOV = \frac{1}{\overline{y}}\left(\sum_{r}^{R} p_r\left(\left(\overline{y}_r - \overline{y}\right)^2\right)\right)^{\frac{1}{2}}$$, where $\overline{y}_r = \frac{1}{n_r}\sum_{i}^{n_r} y_{ir}$ is group *r*'s mean value; *R* is the number of groups; p_r is group *r*'s population share; y_{ir} is the quantity of the variable of interest (e.g., years of education) of the *i*th member of group *r.*

The coefficient of variation is a common measure of regional disparities. GCOV is weighted by the population size of each group, so that changes in the position of small groups get less weight than those of larger groups (Mancini, 2005).

2 **Group-weighted Gini coefficient (GGini):**

$$GGini = \frac{1}{2\overline{y}}\sum_{r}^{R}\sum_{s}^{S} p_r p_s \left|\overline{y}_r - \overline{y}_s\right|$$, where $\overline{y}_r = \frac{1}{n_r}\sum_{i}^{n_r} y_{ir}$ is group r's mean value; *R* is group *r*'s population size; p_r is group *r*'s population share; y_{ir} is the quantity of the variable of interest (e.g., years of education) of the *i*th member of group *r.*

GGINI compares every group with every other group (as opposed to calculating the difference from the mean).[8]

3 **Group-weighted Theil (GTheil):**

$$GTheil = \sum_{r}^{R} p_r \frac{\overline{y}_r}{\overline{y}} \log \frac{\overline{y}_r}{\overline{y}}$$, where $\overline{y}_r = \frac{1}{n_r}\sum_{i}^{n_r} y_{ir}$ is group *r*'s mean value; *R* is group *r*'s population size; p_r is group r's population share; y_{ir} is the quantity of the variable of interest (e.g., years of education) of the *i*th member of group *r.*

GTheil compares each group with the mean. It is especially sensitive to the lower end of the distribution. It can be used to divide vertical inequality into "within group" and "between group" components.

Generally, Stewart et al.'s analysis suggests that these three measures of horizontal inequality are well correlated with each other, but not with vertical inequality.

The second two measures are based on Selway (2011)'s refinement of the concept of "crosscuttingness" used in the literature in political science and political sociology:

4 **Crosscuttingness (CC)** is identified when "group i on cleavage x is identically distributed among groups on cleavage y with all other groups on cleavages x" (Selway, 2011, p. 51).

$$CC = 1 - \sqrt{\Sigma \frac{(O-E)^2}{E} \Big/ nm}$$, where O is the observed frequency in the subgroup cell, E is the expected frequency = (column %)(row %)(total sample size), *n* is the sample size, and *m* is the smaller of either the number of columns minus 1 or the number of rows minus 1.

8 Baldwin and Huber (2010) similarly calculate "between-group inequality" as $BGI = \frac{1}{2\overline{y}}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \left|\overline{y}_j - \overline{y}_i\right|$, where is $\overline{y}$ the mean income in the country and $\overline{y}_i$ is the mean income of group *i.*

CC is based on the normalization of the chi-square statistic given by Cramer (Agresti, 2002) and subtracted from 1, so that higher values imply higher crosscuttingness.

5 **Cross-Fractionalization (CF)** is "the extent to which individuals who are in the same group on one cleavage are in different groups on the other cleavage" (Selway, 2011, p. 52) and based on Rae and Taylor (1970)'s measure of crosscuttingness.[9]

$CF = \sum_{x=1}^{n} p_x^2 + \sum_{y=1}^{n} p_y^2 - 2\sum_{x,y}^{n} p_{xy}^2$, where p_x is the proportion of population at cleavage x, p_y is the proportion at cleavage y, and p_{xy} is the proportion of population at both x and y.

It is closely related to the Herfindahl index, which gives the fractionalization score for the groups on cleavage x as $1 - \sum_{x=1}^{n} p_x^2$.

For two groups, CF is defined as the sum of the number of pairs that share the same group on the first cleavage but not on the second and the number of pairs that share the same group on the second but not the first, divided by N(N-1), where N is the total number of pairs.

2.2 Consideration of gender

Each of the above measures should be calculated for all salient types of "ethnic" divisions (as discussed further below), as well as separately for each type for men and women. Each of the above measures should also be calculated for gender-group-based inequalities, so that we can compare, e.g., inequalities between men and women with those between ethnic groups.

2.3 Additional measures

At least four additional measures of ethnic diversity and vertical inequality should also be calculated. The two ethnic measures are:

1 **Fractionalization (Taylor & Hudson, 1972)**

$Frac = 1 - \sum_{i=1}^{n} p_i^2$ where p_i is the proportion of individuals who belong to group i and n is the number of groups

2 **Polarization (Montalvo & Reynal-Querol, 2005)**

$Pol = 1 - \sum_{i=1}^{n} \left(\frac{0.5 - p_i}{0.5} \right)^2 p_i$ where p_i is the proportion of group i and n is the number of groups.

The two inequality measures are:

3 **Gini coefficient:** See above.

4 **Theil:** See above.

2.4 Key data sources and variables

The key source for analysis in the case studies is the national census. In the ideal situation, a final case study would be based on census data from 1960 (or the closest year before or after)

9 In Selway's terms, it captures both crosscuttingness and sub-group fractionalization.

until 2015 (or the closest year) - e.g., 1960, 1970, 1980, 1990, 2000, and 2010. These data would then be used to describe and analyze patterns and trends in the above indicators for multiple groups across major sub-national regions and over time at the national and sub-national levels.

However, given data constraints and challenges, we recognize that this ideal is unlikely for many – if not all – country case studies included in this project. For instance, census data may be unavailable for multiple years or may require substantial work to put into a form that can be analyzed. Sub-national boundaries may also change over time, complicating comparisons. This may mean that it is only feasible to analyse information from a more limited number of censuses. In addition, many censuses do not collect or report comprehensive data on ethnic and other groups, and many also provide only limited information that can be used to assess socio-economic status. This may mean that in order to provide a decent discussion of empirical patterns, it is necessary to supplement census data with information from standardized surveys such as the Demographic and Health Survey (DHS) or Living Standards Measurement Survey (LSMS).

Plan of action memo: Decisions regarding precisely which data sources will be used in each case study will thus vary across countries and should be made in consultation with the project focal points early on in the research process. For this reason, applicants under the call for papers are required to submit a brief "plan of action memo." The plan of action memo should summarize the data available, the key variables included in each data source, and the author's proposed plan of action for the analysis. In formulating and justifying the proposed plan of action, applicants should keep in mind that the purpose of the case studies in the broader research project is to provide as comprehensive a picture as possible of subnational and diachronic patterns, noting the technical issues summarized in this section. The memo thus should provide:

- detailed notes on the relevant variables available in the census (and other sources as appropriate), including precise questions asked and possible responses/values
- discussion of any challenges to comparisons over time, e.g., given changes in subnational boundaries or in census/survey questions
- an overview of the data available; e.g., if datasets, questionnaires, or codebooks cannot be obtained for some years, this should be noted and explained.

Ethnic variables: We are primarily interested in groups commonly labeled 'ethnic', 'racial', 'linguistic', 'tribal', 'religious', or 'communal' in census and survey questionnaires. (See footnote 2 for more details.) The objective is to capture all major divisions which are politically and/or socially 'salient' (see Posner, 2004). It is well established in the literature that national censuses do not necessarily include all major salient divisions, and that even when relevant ethnic variables are included, data may need to be recoded to reflect salient divisions. For instance, the Philippine census question on 'ethnicity by blood' records 142 possible responses in 2000 and 182 in 2010, but a strong case can be made in terms of ethnic salience for regrouping these responses into a much smaller set of categories (McDoom & Gisselquist, 2015). Analyzing all recorded 'ethnicities by blood' may be interesting from an anthropological perspective, but in terms of group-based inequalities will obscure insight, e.g., into discrimination faced by indigenous peoples as a group. It is thus important that the plan of action memo provide details on relevant ethnic variables and a proposed recoding and rationale as appropriate.

Socio-economic variables: Group-based inequality can be assessed in terms of a number of economic, political, and social markers. We are primarily interested in indicators describing income and wealth, education, occupation, and poverty and living standards. Available data will vary significantly across countries, thus we expect that each case study will consider some different indicators. The plan of action memo should outline relevant variables.

In addition, as the objective is also to use the case studies to improve cross-country comparisons, we aim to identify collectively at least one economic indicator that is common across all or most case study countries and that can be coded in a comparable manner to provide at least a blunt measure of group-based inequality for cross-country comparison. As specific income and wealth estimates are unavailable in many countries, we expect that the two most likely possibilities will be (1) an educational scale and perhaps (2) a (blunt) relative poverty scale. The plan of action memo should explicitly address the availability of data relevant to these two indicators:

- For the education scale: Is sufficient information available to code a 6-point scale (0=no formal education, 1=some primary schooling; 2=completion of primary school and some secondary school; 3=completion of secondary school; 4=some tertiary education; 5=university degree or above). Is sufficient information given to code years of education in a more disaggregated manner?
- For the poverty scale: Is sufficient information available to assess whether respondents are above or below the national poverty line?

Unit of analysis: In general, case study analysis should be done for individuals 18 years and older. Household analysis may also be considered. However, individual analysis will facilitate consideration of gender and of ethnicity in ethnically-mixed households.

3 Case study guidelines

Country case studies are intended to provide an empirical picture of patterns and trends in group-based inequality, with particular attention to sub-national and diachronic variation. They should draw first on national census data, with possible supplementation from nationally-representative surveys. The final version of each case study paper should be approximately 8,000 words, including tables in the main text and references. It should include the following:

1. An introductory section that provides a) a summary of key findings, b) a brief introduction to the country, highlighting population size, poverty rates, levels of inequality, and salient ethnic divisions and groups, and c) a brief review of published work relevant to patterns of group-based inequality in the country and related analysis.

2. Description of data sources used in the analysis, discussion of key ethnic and socio-economic variables used to construct the measures listed above, and basic descriptive statistics. Along the lines discussed above, this section should also review any major data weaknesses, measurement challenges, and recoding of variables.

3. Core analysis: What are the patterns in group-based inequality within the country? What type of group-based inequality is 'worst' (e.g., that based on language, race, religion, etc.)? How well do different measures correlate with each other or provide additional information? How does group-based inequality compare with other key inequalities – in particular vertical inequality, inequalities between men and women, and regional inequalities? How does it compare with measures of ethnic division or diversity? Discussion should be based on analysis of subnational and diachronic patterns in group-based inequality based on the measures described above. At a minimum, this should include consideration of all five measures of group-based inequality for all major ethnic groups in the available data using several different socioeconomic indicators of inequality, including the 'standard' one which will be identified collectively as above.

4. **As long as the core analysis discussed above is provided, papers may also be framed by contributors to advance a broader theoretical or methodological argument. The strongest papers will do so and this will collectively improve publication prospects. For instance, a paper could draw on theory to consider correlates of group-based inequality, develop additional measures of group-based political inequality and discuss how they relate to socioeconomic inequality measures, analyze an alternative method of measuring group-based inequality, or consider how and why sub-national variation at the municipal or village level compares to that at higher administrative levels.**

5. Conclusion.

6. Reference list and appendix of additional tables as appropriate.

7. In addition, final papers should be accompanied by replication datasets and Stata do-files. Contributors should also provide to UNU-WIDER unedited copies of all datasets used in the analysis, along with related material such as codebooks and questionnaires.

4 Schedule of Deliverables

1. Plan of action memo, as summarized in Section 2 above. Deadline under the call for papers: 18 January 2016. The focal points will provide feedback on the memo as needed and use the memos to coordinate standardization across the case studies as described above.

2. First draft. Deadline: 1 May 2016. The first payment will be made upon submission of a satisfactory first draft.

3. Revised final draft. Deadline: 1 August 2016. The second payment will be made upon submission and review of a satisfactory revised draft.

Works cited

Agresti, A. (2002). Categorical Data Analysis. New York: Wiley-Interscience.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. Journal of Economic Growth, 8(2), 155-194. doi: 10.1023/a:1024471506938

Alesina, A., Michalopoulos, S., & Papaioannou, E. (2012). Ethnic Inequality. NBER Working Paper No. 18512. NBER. Cambridge, MA.

Baldwin, K., & Huber, J. D. (2010). Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision. American Political Science Review, 104(04), 644-662. doi: doi:10.1017/S0003055410000419

Cederman, L.-E., Weidmann, N. B., & Gleditsch, K. S. (2011). Horizontal Inequalities and Ethnonationalist Civil War: A Global Comparison. American Political Science Review, 105(3), 478-495.

Chandra, K. (2004). Why Ethnic Parties Succeed: Patronage and Ethnic Head Counts in India. Cambridge: Cambridge University Press.

Easterly, W., & Levine, R. (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. The Quarterly Journal of Economics, 112(4), 1203-1250. doi: 10.1162/003355300555466

Gisselquist, R. M. (2013). Ethnic Politics in Ranked and Unranked Systems: An Exploratory Analysis. Nationalism and Ethnic Politics, 19(4), 381-402. doi: 10.1080/13537113.2013.847598

Horowitz, D. L. (1985). Ethnic Groups in Conflict. Berkeley: University of California Press.

Htun, M. (2004). 'Is Gender like Ethnicity? The Political Representation of Identity Groups'. Perspectives on Politics, 2(3), 439-458.

Kanbur, R. (2006). The Policy Significance of Inequality Decompositions. Journal of Economic Inequality, 4(3), 367-374.

Lijphart, A. (1979). Religious vs. Linguistic vs. Class Voting: The 'Crucial Experiment' of Comparing Beligium, Canada, South Africa, and Switzerland. The American Political Science Review, 73(2), 442-458.

Lipset, S. M., & Rokkan, S. (1967). Party Systems and Voter Alignments: Cross-National Perspectives. New York: Free Press.

Mancini, L. (2005). Horizontal Inequality and Communal Violence: Evidence from Indonesian Districts CRISE Working Paper No. 22. Oxford: University of Oxford.

McAll, C. (1990). Class, Ethnicity, and Social Inequality. Montreal: McGill Queen's University Press.

McDoom, O. S., & Gisselquist, R. M. (2015). The Measurement of Ethnic and Religious Divisions: Spatial, Temporal, and Categorical Dimensions with Evidence from Mindanao, the Philippines. Social Indicators Research, 1-29. doi: 10.1007/s11205-015-1145-9

Montalvo, J., & Reynal-Querol, M. (2005). Ethnic Polarization, Potential Conflict, and Civil Wars. American Economic Review, 95(3), 796-816.

Posner, D. N. (2004). Measuring Ethnic Fractionalization in Africa. American Journal of Political Science, 48(4), 849-863. doi: 10.1111/j.0092-5853.2004.00105.x

Rae, D. W., & Taylor, M. (1970). The Analysis of Political Cleavages. New Haven: Yale University Press.

Selway, J. S. (2011). The Measurement of Cross-cutting Cleavages and Other Multidimensional Cleavage Structures. Political Analysis, 19(1), 48-65. doi: 10.1093/pan/mpq036

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

Taylor, C., & Hudson, M. (1972). World Handbook of Political and Social Indicators (2nd Ed.). New Haven: Yale University Press.