Journal of Open Humanities Data

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: 6. DOI: https://doi.org/10.5334/johd.16

RESEARCH PAPER

Introducing the Historical Index of Ethnic Fractionalization (HIEF) Dataset: Accounting for Longitudinal Changes in Ethnic Diversity

Lenka Drazanova

Migration Policy Centre, European University Institute, IT

lenka.drazanova@eui.eu

Dramatic changes in the ethnic composition of countries in the last decades have sparked new interest among social scientists in studying and uncovering the role of ethnic diversity on social, political and economic outcomes. Yet, most ethnic fractionalization indices used by scholars to study these effects treat ethnic heterogeneity as time-invariant, thus concealing its long-term effects. However, failing to take into account historical developments in ethnic composition might seriously hinder our understanding of their effects on social, economic, and political outcomes. This paper introduces a new dataset containing an annual ethnic fractionalization index for 162 countries across all continents in the period of 1945–2013. The *Historical Index of Ethnic Fractionalization* (HIEF) dataset is a natural extension of previous ethnic fractionalization indices. It offers the opportunity to study the effects of ethnic fractionalization across countries and over time. The article concludes by offering some preliminary descriptive analysis of patterns of change in ethnic fractionalization over time.

Keywords: ethnic fractionalization; ethnic diversity; ethnic groups; intergroup relations

1. Introduction

The demographic composition of most countries has changed dramatically in the last decades [19]. The question of how different people with different backgrounds can peacefully interact in the modern globalized world, and how their societies may prosper, are among the most important challenges in recent years. There has been a growing interest among social scientists and policymakers in studying and uncovering the role ethnic diversity plays in shaping social, political and economic outcomes. Exploring this issue is extremely important and relevant to a range of different public policies including those relating to immigration and integration.

While social scientists can rely on a variety of indices measuring ethnic diversity, to the best of my knowledge, most of these indices treat ethnic fractionalization as a time-invariant phenomenon. Thus, they do not provide country ethnic fractionalization estimates for different years. However, in treating ethnic diversity as time-invariant, we severely limit our understanding of the general long-term effects as well as the effect of slow or radical changes in ethnic fractionalization. The goal of this paper is to characterize the *Historical Index of Ethnic Fractionalization* (HIEF) dataset that provides annual ethnic fractionalization estimates for 162 countries for the years 1945–2013. By introducing a time perspective, the dataset expands previous ethnic fractionalization indices. HIEF estimates are based on the Cline Center for Democracy Composition of Religious and Ethnic Groups (CREG) Project's original data [8] regarding the percentage of principal ethnic groups in each country.

The paper proceeds in three parts. Firstly, I provide general background information on measuring ethnic fractionalization and the problems researchers face when attempting to create ethnic diversity indices. Secondly, I outline how the HIEF dataset was created and describe the data creation in a stepwise manner. Thirdly, I provide a preliminary descriptive analysis of patterns of changes in countries' ethnic fractionalization over time.

2. Measuring ethnic fractionalization

The HIEF dataset builds on conventional measures of heterogeneity used in the literature such as fractionalization or polarization indices. Ethnic fractionalization indices usually measure diversity as a steadily increasing function of the number of groups in a country. They are based on the probability that two randomly drawn individuals from a country belong to two different groups. Theoretically, fractionalization indices range from 0 (when all individuals are members of the same group) to 1 (when each individual belongs to his or her own group). In contrast, polarization indices measure the probability of a potential

conflict that can take place between two groups of equal size. The idea behind polarization indices is that ethnic conflicts will take place in countries where a large ethnic minority faces an ethnic majority. The mere existence of a large ethnic group, and/or ethnic dominance by this group, is not a sufficient condition for an ethnic conflict to develop. There also needs to be an ethnic minority that is large and not divided into many different groups. Theoretically, having a large ethnic minority is the worst possible situation as measures of polarization reach their maximum when two equally sized groups face each other. The two measures represent two different approaches to diversity because ultimate fractionalization occurs when each individual belongs to a different group, whereas ultimate polarization occurs when there are only two types of groups. Thus, the two measures behave quite differently [15]. The HIEF dataset quantifies fractionalization rather than polarization as a first step in providing longitudinal measures of diversity. Nevertheless, since the original data also allow computing a polarization index, this might become a possible future endeavour.

In economics, the majority of studies employ a measure of ethnic fractionalization called Ethno-Linguistic Fractionalization (ELF). The ELF measure was first used in an influential article by Easterly and Levine [14] which argues that given Africa's high ethnic diversity and the strong link between ethnic heterogeneity and slow economic growth, these two factors played a rather important part in the explanation for the region's "growth tragedy". Easterly and Levine's ELF measure is based on the work carried out by a team of Soviet ethnographers in the early 1960s and published as *Atlas Narodov Mira* [6]. Despite ELF's popularity and usage by several generations of political scientists, sociologists and economists, the measure also received criticism and other fractionalization indices have been developed. Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg [1] propose a classification that distinguishes between ethnic, linguistic and religious diversity and creates separate indices for each. Their reasoning is based on the fact that relying largely on linguistic distinctions (as the ELF does) may obscure other aspects of ethnicity like racial origin, skin colour and so forth. For instance, in many countries in South America groups are largely monolingual, yet ethnically divided. Other researchers argued that a distinction must be made between ethnically and culturally diverse groups [17] or between politically relevant ethnic groups [25].

There have also been efforts to overcome simple fractionalization measures by focusing on conjunctures with other heterogeneities such as the index of ethnic inequality [3] that puts forward the inter-section of ethnic diversity and economic inequality or an index that combines five cleavages, namely race, language, religion, region, and income [27]. Other indices make an effort to account for the distance between groups [16], the historical depth of ethnic cleavages [10] or consider heterogeneity between individuals rather than groups [4].

As explained above, heterogeneity may be defined ethnically, religiously, linguistically, culturally, but also economically as income inequality. It is worth underlining that indices regarding ethnic composition are particularly vulnerable to criticism in their attempts to measure ethnicity. To begin with, empirical efforts to create an ethnic index require that we collect data on ethnic groups in different countries. However, there is no uniform criterion on how to define ethnicity. Group identities are complex and mostly socially constructed which means that quantifying and measuring them is inherently problematic. There can be multiple ways to specify ethnic groups in a country all of which may be equally valid concepts of "ethnic groups". Moreover, even within one country, definitions of ethnicities can change over time. Questions related to the definition of diversity become even harder in comparative research that involves multiple countries each of which has its own concept of ethnicity. These facts notwithstanding and being aware of the possible shortcomings in constructing ethnic classifications, the HIEF dataset is largely based on an ethnic, rather than linguistic, distinction between groups.

2.1 Why change over time matters

Definitional issues aside, I argue that a major problem with a large part of the existing social science research on the effects of ethnic diversity is that diversity is often treated as time-invariant. This limits our knowledge about diversity's long-term effects. An increase or decline in ethnic fractionalization over time might have different consequences. For instance, countries with steadily increasing ethnic diversity might be more willing to introduce institutions that effectively manage problems connected to more heterogeneity than countries with shorter histories of ethnically diverse societies or with lower average rates of change in diversity. These institutions may then mediate the relationship between ethnic diversity and social, economic, and political outcomes. Moreover, in instances such as in the case of the dissolution of multi-ethnic states ethnic fractionalization may decrease rapidly which poses completely different challenges to the newly homogeneous societies. Failing to consider these historical developments might seriously hinder our understanding of the effects of ethnic diversity. With HIEF, it is now possible to depict longitudinal relationships that might improve our understanding of the causal relationships between ethnic diversity and relevant outcomes. A number of studies consider changes in ethnic diversity longitudinally in several countries. However, these studies either rely on immigration estimates [20], consider only one country at a time [11], or focus on subnational units [29]. Recently, some scholars published articles that use time-varying measures of ethnic fractionalization [5, 7], but all of the indices used are much more limited than HIEF, either with regard to time-variation or countries covered. Moreover, these studies do not make their original dataset publicly available to be used by other researchers.

3. Creating the Historical Index of Ethnic Fractionalization Dataset

The original data on ethnic groups were gathered from CREG initiated by the Cline Center. The project provided information regarding the percentage of principal ethnic groups present in 162 countries annually for the period 1945–2013 [8]. The main sources for the CREG

data were the Britannica Book of the Year, the CIA World Factbook, and the World Almanac Book of Facts [24]. In the original dataset, data were recorded from the main sources by a group of data collectors and later assessed by a group of data integrators who performed a number of checks. These checks accounted for consistency of group names and data outliers such as if there is "a group that is reported as constituting 25% of the population in one year and 35% in the next" [24: 4], and data inconsistencies when "different editions of the same source reports (*sic*) a group as constituting 18% of the population and 26% of the population in 1968" [24: 4]. Nevertheless, as the original dataset still contained some inconsistencies such as repeated information regarding certain ethnic groups in a single year, the original dataset had to be carefully checked and corrected.

In the HIEF dataset, the degree of ethnic fractionalization has been calculated based on the annual percentage of ethnic groups in each country using the most universally applied formula in the empirical literature which is a decreasing transformation of the Herfindahl concentration index measured by:

$$EF_{ct} = 1 - \sum_{i=1}^{n} S_i^2$$

where EF_{ct} is the level of ethnic fractionalization in country *c* at time *t*, *i* indexes ethnic groups, and S_i is the proportion of the population in unit *c* belonging to ethnic group *i* (*i* = 1, ..., n) at time *t*.

As described above, the ethnic fractionalization index for each country at any given year ranges from 0, where there is no ethnic fractionalization in the country and all individuals are members of the same ethnic group, to 1, where each individual in the country belongs to his or her own ethnic group.

It should be noted that, historically, who was considered as belonging to a certain ethnic group could change, reflecting the politics and science of the times. The relative meaning of being in a certain category may not be the same from one time-point to another [28] both from the societal or individual point of view. The challenge arises especially with the introduction of categories such as "mixed race", mestizo, mulatto and similar categories in data collection. Thus, the measures may only have "nominal equivalence" and lack "functional equivalence" [9] which makes collecting ethnicity data, and measuring changes over time, challenging.

4. Descriptive illustration of the new dataset

The HIEF dataset contains three variables, namely *Country*, *Year* and *EFindex*. The variable *Country* contains the names of countries included in the dataset. Countries that have changed their name and status are included under the official name of the country for the year in question. For example, Bosnia-Herzegovina, Croatia, Slovenia, Macedonia, and Serbia have been part of the Socialist Federal Republic of Yugoslavia from 1945 until 1992, while the other Yugoslav successor states of Kosovo and Montenegro are not included in the HIEF dataset. Thus, the variable *Country* includes the entry "Yugoslavia" for the years 1945–1992 and five separate entries "Bosnia-Herzegovina", "Croatia", "Slovenia", "Macedonia" and "Serbia" for 1993–2013. It follows that countries founded after the year 1945 are included beginning from the year they have been officially established.

The variable *Year* contains the corresponding year of observation for each country, usually ranging from 1945 to 2013. As described above, a shorter time span may be included for certain countries which were yet to be founded or ceased to exist.

The variable *EFindex* contains the actual value of the ethnic fractionalization index in each country for all available years. Every value of the ethnic fractionalization index can be, as described above, interpreted as one minus a weighted sum of population shares p_i where the weights are these shares themselves. **Table 1** summarizes the countries and years for which the ethnic fractionalization index is available.

The HIEF dataset is made available as a .csv file and can be found along with a document briefly introducing the dataset on the Harvard Dataverse repository [12].

4.1 Comparing changes in diversity among European countries

Explorations of the new dataset illustrate the reasons why it is important to take account of historical changes in ethnic diversity within countries. **Figure 1** shows the change of ethnic fractionalization over time in a sample of European countries. We can observe, for example, that Great Britain and the Netherlands had a similar level of ethnic fractionalization in 2013, but since 1949, diversity in the Netherlands has grown at a much faster pace than in Great Britain. In other words, Dutch society had to adapt to diversity more rapidly than the British. In contrast, Finland's ethnic fractionalization has stayed quite stable over the last 50 years and is generally low.

On the other hand, many Central and East European countries are much more ethnically homogenous than they used to be. Moreover, they became homogeneous in a short period. For instance, while former Czechoslovakia used to be an ethnically highly heterogeneous country, its successor states, Czechia and Slovakia, are much more homogeneous. Apart from separations of what used to be ethnically heterogeneous countries such as Czechoslovakia, Yugoslavia, or the Soviet Union, there are a number of reasons why one can observe changes towards more homogeneity in Central and Eastern Europe. Firstly, after the collapse of communism, many workers left their respective countries in search of new economic opportunities. Secondly, in many post-Soviet countries, Russian minorities began to feel unwelcome resulting in return migration [18].

As one can observe in **Figure 2**, many African countries are highly ethnically heterogeneous with relative fractionalization stability. For instance, highly heterogeneous countries such as South Africa or Uganda have not experienced dramatic changes in fractionalization over the years. On the other hand, although its overall fractionalization is quite low, Swaziland has experienced a steady increase in heterogeneity while diversity in Tanzania and the Democratic Republic of Congo has actually declined. Thus, there might be profound political and societal differences

Table 1: Overview of countries and years covered by the Historical Ethnic Fractionalization Index.

Country	Years	Country	Years	Country	Years
Afghanistan	1945–2013	Colombia	1945–2013	German Federal Rep.	1949–2013
Albania	1945–2013	Comoros	1975–2013	Ghana	1957–2013
Algeria	1962–2013	Congo	1960–2013	Greece	1945–2013
Angola	1975–2013	Costa Rica	1945–2013	Guatemala	1945–2013
Argentina	1945–2013	Cote d'Ivoire	1960–2013	Guinea	1958–2013
Armenia	1991 –2013	Croatia	1991–2013	Guinea-Bissau	1974–2013
Australia	1945–2013	Cuba	1945–2013	Guyana	1966–2013
Austria	1945–2013	Cyprus	1960–2013	Haiti	1945–2013
Azerbaijan	1991–2013	Czech Republic	1993–2013	Honduras	1945–2013
Bahrain	1971–2013	Czechoslovakia	1945–1992	Hungary	1945–2013
Bangladesh	1971–2013	Dem. People's Republic of Korea	1948–2013	Indonesia	1945–2013
Belarus	1991–2013	Dem. Republic of Congo	1960–2013	Iran	1945–2013
Belgium	1945–2013	Dem. Republic of Vietnam	1945–2013	Iraq	1945–2013
Benin	1960–2013	Denmark	1945–2013	Ireland	1945–2013
Bhutan	1949–2013	Djibouti	1977–2013	Israel	1948–2013
Bolivia	1945–2013	Dominican Republic	1945–2013	Italy	1945–2013
Bosnia-Herzegovina	1992–2013	East Timor	2002–2013	Jamaica	1962–2013
Botswana	1966–2013	Ecuador	1945–2013	Japan	1945–2013
Brazil	1945–2013	Egypt	1945–2013	Jordan	1945–2013
Bulgaria	1945–2013	El Salvador	1945–2013	Kazakhstan	1991–2013
Burkina Faso	1960–2013	Eritrea	1993–2013	Kenya	1963–2013
Burundi	1962–2013	Estonia	1991–2013	Kuwait	1961–2013
Cambodia	1953–2013	Ethiopia	1945–2013	Kyrgyz Rep.	1991–2013
Canada	1945–2013	Fiji	1970–2013	Laos	1954–2013
Cape Verde	1975–2013	Finland	1945–2013	Latvia	1991–2013
Central African Rep.	1960–2013	Gabon	1960–2013	Lebanon	1945–2013
Chad	1960–2013	Gambia	1965–2013	Lesotho	1966–2013
Chile	1945–2013	Georgia	1991–2013	Liberia	1945–2013
China	1945–2013	German Democratic Rep.	1949–1990	Libya	1951–2013
Lithuania	1991–2013	Portugal	1945–2013	Togo	1960–2013
Macedonia	1991–2013	Qatar	1971–2013	Trinidad and Tobago	1945–2013
Madagascar	1960–2013	Republic of Korea	1948–2013	Tunisia	1956–2013
Malawi	1964–2013	Republic of Vietnam	1954–1975	Turkey	1945–2013
Malaysia	1957–2013	Romania	1945–2013	Turkmenistan	1991–2013
Mali	1960–2013	Russia	1991–2013	Uganda	1962–2013
Mauritania	1960–2013	Rwanda	1960–2013	Ukraine	1991–2013
Mauritius	1968–2013	Saudi Arabia	1945–2013	United Arab Emirates	1971–2013
Mexico	1945–2013	Senegal	1960–2013	United Kingdom	1945–2013
Moldova	1991–2013	Serbia	1991–2013	United States of America	1945–2013
Mongolia	1945–2013	Sierra Leone	1961–2013	Uruguay	1945–2013
Morocco	1956–2013	Singapore	1960–2013	USSR	1945–1991
Myanmar	1948–2013	Slovakia	1993–2013	Uzbekistan	1991–2013

(Contd.)

Country	Years	Country	Years	Country	Years
Namibia	1990–2013	Slovenia	1991–2013	Venezuela	1945–2013
Nepal	1945–2013	Solomon Islands	1978–2013	Yemen Arab Rep.	1945–2013
Netherlands	1945–2013	Somalia	1960–2013	Yemen PDR	1967–1990
New Zealand	1945–2013	South Africa	1945–2013	Yugoslavia	1945–1990
Nicaragua	1945–2013	Spain	1945–2013	Zambia	1964–2013
Niger	1960–2013	Sri Lanka	1948–2013	Zimbabwe	1965–2013
Nigeria	1960–2013	Sudan	1956–2013		
Norway	1945–2013	Swaziland	1968–2013		
Oman	1945–2013	Sweden	1945–2013		
Pakistan	1947–2013	Switzerland	1945–2013		
Panama	1945–2013	Syria	1945–2013		
Paraguay	1945–2013	Taiwan	1949–2013		
Peru	1945–2013	Tajikistan	1991–2013		
Philippines	1946–2013	Tanzania	1961–2013		
Poland	1945–2013	Thailand	1945–2013		

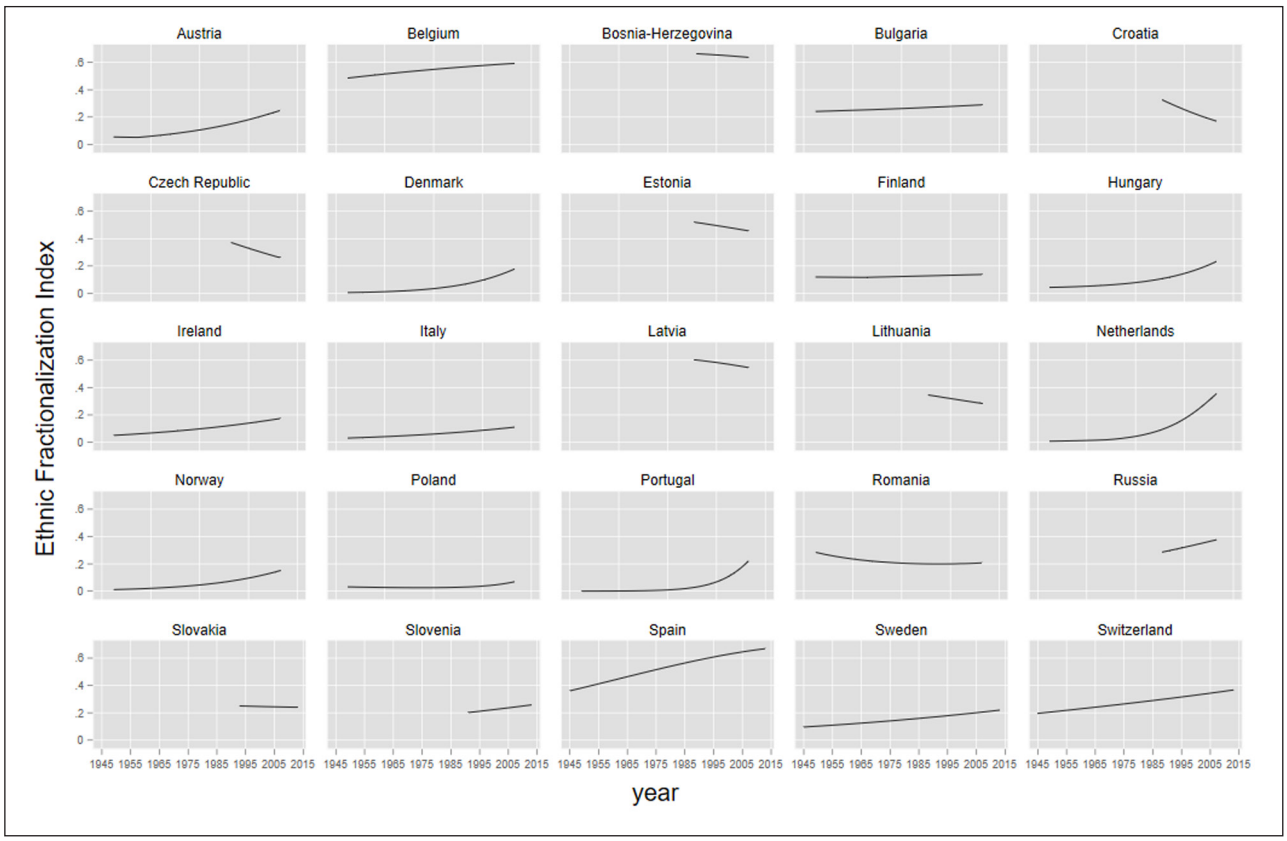

Figure 1: Ethnic fractionalization in a sample of European countries in the years 1945–2013.

between these countries not only concerning current ethnic fractionalization, but also how fast the levels of their current ethnic diversity were achieved in recent years.

4.2 Dataset robustness check

To test the robustness of the HIEF dataset, three new datasets are created that add some noise to the original data. This procedure is adapted from Kolo [21]. The four datasets should not differ in a significant way. As described in detail in Kolo [21], the noise data is created by employing normal randomization, namely by replacing the original group size with a new size produced by a normal distributed random variable. This way, two alternative datasets have been created. Dataset *sigma_1* uses the standard deviation of the group distribution over all observations which is thus equal for all countries, while

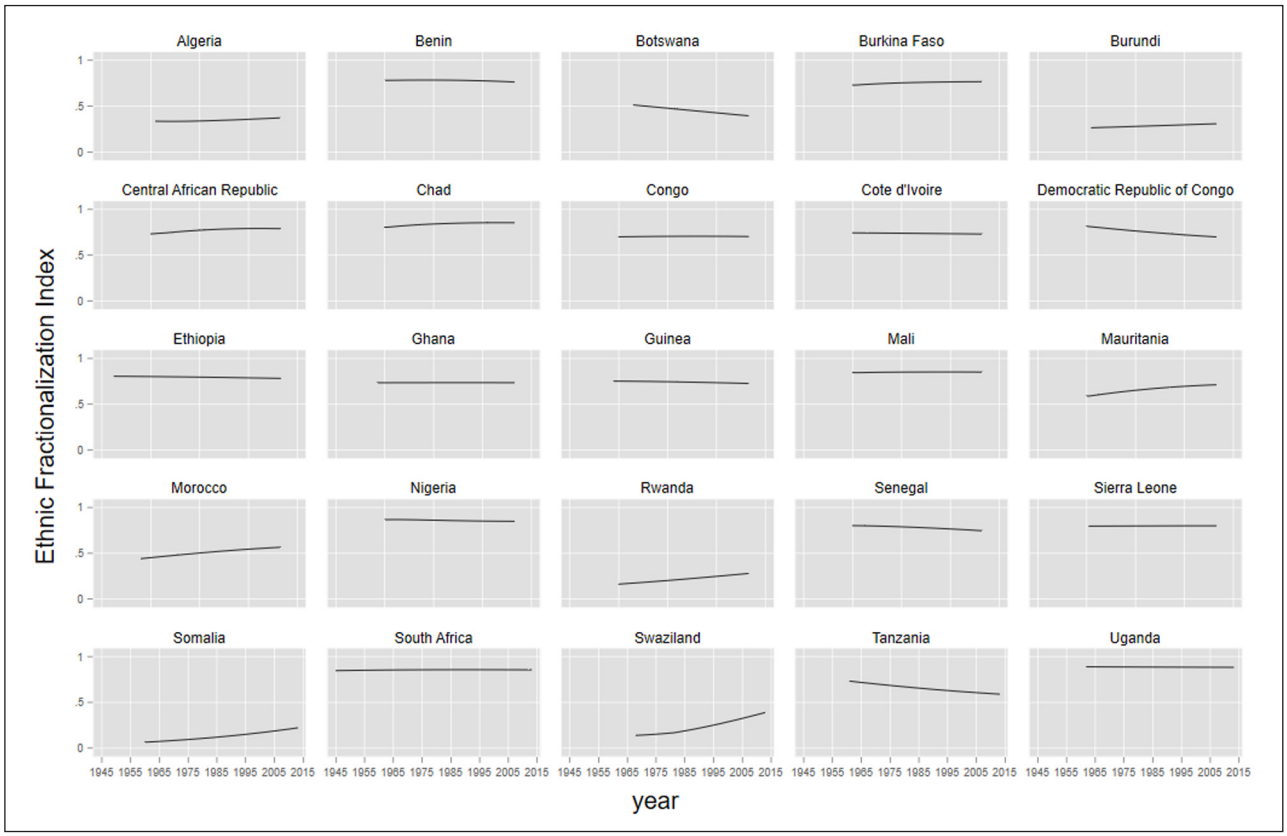

Figure 2: Ethnic fractionalization in a sample of African countries in the years 1945–2013.

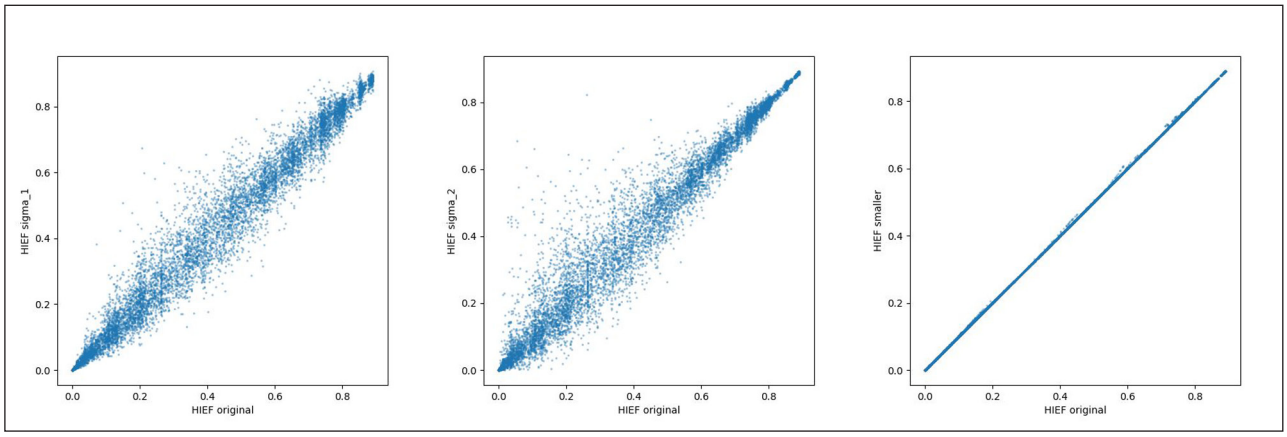

Figure 3: Original HIEF values against newly created random datasets sigma_1 and sigma_2 and against reduced dataset smaller.

dataset *sigma_2* uses a country-specific standard deviation. Finally, as a final robustness test a third *smaller* dataset is created in which the smallest group for each country and for each year is removed. It should be, however, noted that the group is only removed if the number of groups in a country in a given year is greater than one, and the group size of the smallest group is smaller than 1 percent.

Pearson correlations between the original HIEF dataset and the three noisy datasets are all very high (*sigma_1* $r = 0.982$; *sigma_2* $r = 0.974$; *smaller* $r = 1.000$) confirming high congruency. Moreover, there were no statistically significant differences between the original HIEF dataset and the three noisy ones (*sigma_1* $t(17568) = -0.186$, $p = 0.852$; *sigma_2* $t(17568) = -1.411$, $p = 0.158$; *smaller* $t(17568) = -0.062$, $p = 0.949$). **Figure 3** shows the values of the three noisy datasets plotted against the HIEF original data.

5. Conclusion

The aim of this article has been to describe the new *Historical Index of Ethnic Fractionalization* (HIEF) dataset, the procedures used for its calculation and, finally, to illustrate the importance of considering historical developments in ethnic fractionalization. Focusing on country-year estimates for the period 1945–2013, the HIEF dataset complements already existing ethnic fractionalization indices which do not take into consideration the variation of ethnic fractionalization over time. This is an

important advancement as the variation in ethnic heterogeneity over time might be relevant for the effects of ethnic fractionalization on diverse social, economic, and political outcomes. Many studies have concluded that ethnic diversity has a negative impact on economic development [14], macroeconomic stability [2], social trust [26], quality of governance [22], democracy [23] among others. However, I argue that there may be value in rethinking the assumption of ethnic diversity being relatively time-invariant. Changes in heterogeneity might play a role in affecting the relationship between ethnic diversity and social, economic, and political outcomes. Looking at long-term effects and (rapid or slow) time-variant changes in ethnic diversity can help us to advance knowledge about the peaceful co-existence in ethnically diverse societies. For instance, it might help social scientists evaluate the under-explored hypothesis that while people usually react negatively to more diversity, in the long run, ethnic diversity can prove beneficial [26].

Supplementary information and material

Supplementary material is provided in the form of data and open source Python scripts. The data used in this study are archived on Harvard Dataverse [12].

A GitHub repository [13] contains the Python script that has been used to generate the HIEF dataset and its noisy versions for data robustness check. Instructions on how to run the script are available in the Readme file contained in the repository.

Competing Interests

The author has no competing interests to declare.

References

1. **Alesina A, Devleeschauwer A, Easterly W, Kurlat S, Wacziarg R.** Fractionalization. *Journal of Economic Growth*. 2003; 8(2): 155–194. DOI: https://doi.org/10.1023/A:1024471506938
2. **Alesina A, Drazen A.** Why Are Stabilizations Delayed? *The American Economic Review*. 1991; 81(5): 1170–1188. DOI: https://doi.org/10.3386/w3053
3. **Alesina A, Michalopoulos S, Papaioannou E.** Ethnic Inequality. *Journal of Political Economy*. 2016; 124(2): 428–488. DOI: https://doi.org/10.1086/685300
4. **Arbatli CE, Ashraf QH, Galor O, Klemp M.** The Nature of Civil Conflict. *Econometrica*. 2015; 88(2): 727–797. DOI: https://doi.org/10.3982/ECTA13734
5. **Bove V, Elia L.** Migration, Diversity, and Economic Growth. *World Development*. 2017; 89: 227–239. DOI: https://doi.org/10.1016/j.worlddev.2016.08.012
6. **Bruk SI, Alenčenko VS, Telberg VG.** Atlas Narodov Mira. 1964. Moskva: N. N. Miklucho-Maklaja, Institut Etnografii Imeni.
7. **Campos NF, Saleh A, Kuzeyev V.** Dynamic Ethnic Fractionalization and Economic Growth. *The Journal of International Trade & Economic Development*. 2011; 20(2): 129–152. DOI: https://doi.org/10.1080/09638199.2011.538218
8. **Cline Center for Advanced Social Research.** *Composition of Religious and Ethnic Groups* (*CREG*) *Project*. 2014. Available at https://uofi.app.box.com/s/ien7v6e2y7tzzy6t7nljkqb1ge2uww6n [Last accessed 15 August 2020]
9. **Connelly R, Gayle V, Lambert PS.** *Ethnicity and Ethnic Group Measures in Social Survey Research. Methodological Innovations*. 2016; 9: 1–10. DOI: https://doi.org/10.1177/2059799116642885
10. **Desmet K, Ortuño-Ortín I, Wacziarg R.** The Political Economy of Linguistic Cleavages. *Journal of Development Economics*. 2012; 97(2): 322–338. DOI: https://doi.org/10.1016/j.jdeveco.2011.02.003
11. **Dinesen PT, Sønderskov KM.** Trust in a Time of Increasing Diversity: On the Relationship between Ethnic Heterogeneity and Social Trust in Denmark from 1979 until Today. *Scandinavian Political Studies*. 2012; 35(4): 273–294. DOI: https://doi.org/10.1111/j.1467-9477.2012.00289.x
12. **Dražanová L.** Historical Index of Ethnic Fractionalization Dataset (HIEF). *Harvard Dataverse (V1)*. 2019a. DOI: https://doi.org/10.7910/DVN/4JQRCL
13. **Dražanová L.** GitHub Repository of the HIEF Dataset. 2019b. Available at https://github.com/lenkadrazanova/HIEF_dataset [Last accessed 23 August 2020]
14. **Easterly W, Levine R.** Africa's Growth Tragedy: Policies and Ethnic Divisions. *The Quarterly Journal of Economics*. 1997; 112(4): 1203–1250. DOI: https://doi.org/10.1162/003355300555466
15. **Esteban J, Debraj R.** Polarization, Fractionalization and Conflict. *Journal of Peace Research*. 2008; 45(2): 163–182. DOI: https://doi.org/10.1177/0022343307087175
16. **Esteban J, Mayoral L, Debraj R.** Ethnicity and Conflict: An Empirical Study. *American Economic Review*. 2012; 102(4): 1310–1342. DOI: https://doi.org/10.1257/aer.102.4.1310
17. **Fearon JD.** Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*. 2003; 8(2): 195–222. DOI: https://doi.org/10.1023/A:1024419522867
18. **Heleniak T.** Migration of the Russian Diaspora After the Breakup of the Soviet Union. *Journal of International Affairs*. 2004; 57(2): 99–117. URL: https://www.jstor.org/stable/24357868
19. **International Organization for Migration.** *World Migration Report 2020*. 2019. Available at https://publications.iom.int/books/world-migration-report-2020 [Last accessed 17 August 2020]
20. **Kesler C, Bloemraad I.** Does Immigration Erode Social Capital? The Conditional Effects of Immigration-Generated Diversity on Trust, Membership, and Participation across 19 Countries, 1981–2000. *Canadian Journal of Political Science/Revue Canadienne de Science Politique*. 2010; 43(2): 319–347. DOI: https://doi.org/10.1017/S0008423910000077
21. **Kolo P.** Measuring a new aspect of ethnicity: The appropriate diversity index. *IAI Discussion Papers*, 2012; No. 221. Ibero America Institute for Economic Research (IAI), Georg-August-Universität Göttingen. URL: http://hdl.handle.net/10419/92997
22. **Lieberman ES, McClendon GH.** The Ethnicity-Policy Preference Link in Sub-Saharan Africa. *Comparative Political Studies*. 2013; 46(5): 574–602. DOI: https://doi.org/10.1177/0010414012463876

23. **Lijphart A.** *Democracy in Plural Societies*: *A Comparative Exploration*. 1977. New Haven, CT: Yale University Press. URL: https://www.jstor.org/stable/j.ctt1dszvhq
24. **Nardulli PF, Wong CJ, Singh A, Peyton B, Bajjalieh J.** *The Composition of Religious and Ethnic Groups (CREG) Project*. Cline Center for Democracy, University of Illinois at Urbana-Champaign. 2012. URL: http://www.nber.org/ens/feldstein/ENSA_Sources/Cline%20Center/Ethnic%20and%20Religious%20Groups/CREG-White_Paper.pdf
25. **Posner DN.** Measuring Ethnic Fractionalization in Africa. *American Journal of Political Science*. 2004; 48(4): 849–863. DOI: https://doi.org/10.1111/j.0092-5853.2004.00105.x
26. **Putnam RD.** E Pluribus Unum: Diversity and Community in the Twenty-First Century. The 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies*. 2007; 30(2): 137–174. DOI: https://doi.org/10.1111/j.1467-9477.2007.00176.x
27. **Selway JS.** The Measurement of Cross-Cutting Cleavages and Other Multidimensional Cleavage Structures. *Political Analysis*. 2011; 19(1): 48–65. DOI: https://doi.org/10.1093/pan/mpq036
28. **Van Deth JW.** Using Published Survey Data. In: Harkness JA, van de Vijver FJR, Mohler PP. *Cross-Cultural Survey Methods*. 2003; 329–346. New York, NY: Wiley.
29. **Ziller C.** Ethnic Diversity, Economic and Cultural Contexts, and Social Trust: Cross-Sectional and Longitudinal Evidence from European Regions, 2002–2010. *Social Forces*. 2014; 93(3): 1211–1240. DOI: https://doi.org/10.1093/sf/sou088

How to cite this article: 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: 6. DOI: https://doi.org/10.5334/johd.16

Published: 11 September 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 Unported License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

]u[*Journal of Open Humanities Data* is a peer-reviewed open access journal published by Ubiquity Press

OPEN ACCESS