Historical Index of Ethnic Fractionalization Dataset (HIEF)

Abstract: Historical Index of Ethnic Fractionalization (HIEF) dataset contains an ethnic fractionalization index for 165 countries across all continents. The dataset covers annually the period 1945-2013. The ethnic fractionalization index corresponds to the probability that two randomly drawn individuals within a country are not from the same ethnic group. The applications of HIEF pertain to the pattern of ethnic diversity across countries and over time.

Source of data

The original data was gathered from the Composition of Religious and Ethnic Groups (CREG) project initiated by the Cline Center for Democracy, University of Illinois at Urbana-Champaign. It provided information regarding the percentage of principal ethnic groups present in 165 countries annually for the time period 1945 - 2013 and is available at https://clinecenter.illinois.edu/project/Religious-Ethnic-Identity/composition-religious-and-ethnic-groups-creg-project. As the original dataset contained some inconsistencies (for instance it contained repeated information regarding certain ethnic groups) the original dataset had to be carefully checked and corrected.

Historical Index of Ethnic Fractionalization Dataset

Based on the annual percentage of ethnic groups in each country The Historical Index of Ethnic Fractionalization (hereafter, HIEF) calculates the degree of ethnic fractionalization (EF) using the most universally applied formula in the empirical literature, which is a decreasing transformation of the Herfindahl concentration index measured by:

$$EF_c = 1 - \sum_{i=1}^{n} Si^2$$

where EF_c is the level of ethnic fractionalization in country c, i indexes ethnic groups and S_i is the proportion of the population in unit c belonging to ethnic group i (i = 1, …, n).

Theoretically, the ethnic fractionalization index reflects the likelihood that two people chosen at random within a given country will be from different ethnic groups. The ethnic fractionalization index ranges from 0, when there is no ethnic fractionalization and all individuals are members of the same ethnic group to 1, where each individual belongs to his or her own ethnic group.

The unit of observation is country-year.

List of variables

HIEF contains three variables;
Country - name of country
Year – corresponding year between 1945 – 2013. For some countries, years are available only for the time period when country became independent
EFindex – corresponding ethnic fractionalization index for each country-year.