

MEASURING OF ETHNIC HOMOGENEITY OF THE POPULATION – ONE NEW APPROACH

Dimitar ARKADIEV
Faculty of Economics
"St. Cyril and St. Methodius" University
Veliko Tarnovo, Bulgaria
arkadiev@abv.bg

Abstract

The necessity of constructing a measure of ethnic homogeneity of the population of the different countries is founded. The advantages and shortcomings of different indices used for this purpose are discussed. The indicators according to which the different ethnic communities and groups are converging (respectively diverging) are also discussed. It is established that the language and territorial proximity are most important. Their definition does not cause special difficulties. For other types of proximity - religious, racial, economic, etc., considerable difficulties arise for their measurement. A coefficient of population ethnic homogeneity is proposed. It accounts for the number of the respective ethnic group and the degree of proximity to the basic (the most numerous) ethnic group in the country. For the definition of this proximity a range scale is applied. A classification of the coefficient of ethnic homogeneity is done. Examples for calculating the coefficient for different countries are given.

Keywords: economic comparison, indexes of ethnic heterogeneity, ethnic indicators, ethnic data, ethnic classification, coefficient of ethnic homogeneity.

JEL classification: A13, B49, C19, C46, D72, J15, N30, O15, O57, R15, Z13.

1. INTRODUCTION

Ethnically based conflicts are quite common in contemporary world. They have got deep historical roots and have been fuelled by political and economic interests.

National development of ethnic communities is different. Some of them have managed to create national states. Others have failed to do so for various reasons.

The development of statehood in the world is also different. On the European continent most countries are established on the basis of one ethnic community - Albania, Bulgaria, Germany, Denmark, Portugal, Poland, etc. Many of these countries covers entirely or big part of the ethnic territory. Few European countries are on the basis of two or more ethnic communities: Belgium (Flemish and Walloons), Spain (Spanish, Catalan, Galician, and Basque), Switzerland and others.

Totally different situation exists in Africa on the south of Sahara desert. Here most of the state borders are formed randomly during the colonial period. They reflect the interests and arrangements among colonial empires. Many communities are split between several

countries. Therefore national development of these countries is difficult and meets with larger or smaller contradictions between ethnic communities and groups. There are similar cases on Asian continent. National development of the countries on other continents - America, Australia and Oceania also has got a significant whimsicality.

Usually after the establishment of most national states in the past on the basis of ethnic majority governments of these countries have followed a policy of ethnic homogeneity formation. This policy is implemented by more or less violent means. The idea to get an ethnic homogeneity by this way is not uninteresting to any governments even on European continent. One such example is the development and disintegration of former Yugoslavia.

Most of the economists who have conducted research about the relationship between income inequality and growth, tried to identify its determinants. One of the most effective hypotheses was written by Simon Kuznets [Kuznets, 1955, 1-28].

Simon Kuznets' hypothesis is an important thesis about economic growth and income inequality. Kuznets thought that economic development shifts economic inequality from lower to higher stages, and that after a period of time, it will lower again.

The pattern of income inequality which first rises then falls as economic development occurs is called the "Kuznets curve" or "the inverted-U hypothesis." Kuznets' idea made a big impression among other economists who worked on the relationship between economic growth and income inequality.

Kuznets' hypothesis was based on cross-section regression of Gross Domestic Product (GDP) per capita and income distribution across a large number of countries. Kuznets and other economists who use cross-section analysis for assessing the relationship between economic growth and income inequality looked across different countries at nearly the same time and examined how the income inequality fluctuates in moving from lower-income, less-developed countries, to higher income, developed countries.

Now, many economists want to measure directly the impact of ethnic, linguistic, religious, racial and other diversity of separate countries on economic growth and income inequality. There are numerous scientific publications (books, articles, etc.) where that problem is discussed [Alesina, 2005, 766-767; Das, 2010, 91-113; Emigh, 2001; Fearon, 2003, 75-90; Fielding, 2005, 279-301; Genovese, 2006; Hegre, 2004; Kim, 2008; Knack, 2002, 91-107; Leigh, 2006, 121-125].

Naturally, the question how to measure an ethnic diversity is very important.

2. INDICATORS OF ETHNIC HOMOGENEITY OR HETEROGENEITY

The objective situation of ethnic homogeneity or heterogeneity needs a summary measure. That need arises when there are researches and comparisons in various fields - economic, political, cultural, historical, and others, and particularly when statistics and econometrics method and models have applied to determine the objective laws governing the development of society. Some of the variables in these models reflect different economic and social phenomena, which directly can be measure by indicators such as gross domestic product, population, income, unemployment and others. Some phenomena as corruption, social security, political freedoms and others can not be measure directly. Special constructions popular so-called indexes have been used for these phenomena.

Such indexes are built and used for characterization of ethnic, linguistic and religious homogeneity or heterogeneity. Some of the most frequently proposed and used indices in this regard are the following.

Index of ethno-linguistic fractiolization - *ELF*. It is also known under the abbreviation *ETHFRAC*.

It is the following:

$$ELF = 1 - \sum_{i=1}^N p_i^2 \quad (i = 1, 2, \dots, N), \quad (1)$$

where p_i - the proportion of the ethno-linguistic group in the country's population,

N – number of ethno-linguistic groups.

The index measures the likelihood that two individuals taken not to belong to the same ethno-linguistic group.

ELF is moved in the range from 0 to 1 when is presented as a factor, and from 0 to 100 when is expressed as a percentage (for this purpose it must be multiplied by 100). Low rates show little fragmentation and high values - big fragmentation (heterogeneity).

Similarly, the indexes of ethnic (EF), language (LF) and religious (RF) fragmentation can be built separately. Everywhere they consider a sum of squares of the corresponding proportion of ethnic linguistic and religious groups in the country. Therefore the abbreviation **FRAC** is used to reflect totally these indexes [Alesina, 2003, 155-194].

Index of ethnic heterogeneity - *EHET*. This index is built on the combination of three sub-indexes: for racial, for linguistic and for religious divisions [Vanhanen, 1999, 55-79]. Each of them is defined the relationship between 100 percent and the proportion of the largest racial group (p_R), the largest language group (p_L) and the largest religious group (p_{RE}) as a percentage of the population in the country.

Formally, the index is the following:

$$EHET = \frac{100}{p_R} + \frac{100}{p_L} + \frac{100}{p_{RE}} \quad (2)$$

The index shows how many times the most relevant group is contains in the total population (retroactive magnitude of the share).

The index should move in the range from 0 to 144. The interpretation is similar to *ELF*.

A careful examination shows that the lower limit may not be 0, and at least 3. This value would realize if the entire population of the country consists only of one race, one language and one religious group (complete homogeneity).

Index of ethnic homogeneity. This index is calculated as a sum of the squares of the seven largest ethnic group proportions in the country population. It is ranged from 0 to 1. Low point corresponds to the complete homogeneity and up point - complete heterogeneity [Portable ..., 2002].

The index is similar to *ELF*. What is the difference between them? That index only is relative to the first seven ethnic groups, while *ELF* is relative to all ethnic groups. Therefore its gradation is in reverse order.

Other indexes. Some other indexes are applied besides the discussed above.

One of them measures the degree of *ethnic dominance*. It shows a ratio between sizes of first largest ethnic group to the second largest ethnic group.

Furthermore, different combinations of these indexes are put into practice of econometric analysis [Reynal-Querol, 2002, 29-54].

3. SOURCES OF DATA ON ETHNIC BACKGROUNDS AND THEIR FEATURES

The construction of indexes which measure ethnic, linguistic and religious population diversity in different countries needs by adequate data.

The critical comments prevail over data regarding indexes which are discussed above. Usually, there are not so any comments about construction (content) of the indexes [Laitin, 2001, 13-17; Hegre, 2004, etc.].

Perhaps the reasons for this attitude are various. One of important reason is the inclusion of these indexes as independent variables in econometric models for international comparisons. Sometimes one problem appears. The parameters of fractiolization variables are statistically significant in some cases but in other cases – not. That may occur even when similar cases are investigated. Of course, when preliminary expectations (theoretical, hypothetical, etc.) are not confirmed, the researchers are so zealous in this regard.

We want too much of our data (as well as one author is said in another case). We want to check with their help our theories to obtain estimates of the most important parameters and then - to clarify the exact type of relationship between different variables [Griliches, 1967, 16-49].

Some of discussed above indexes are built by many analysts in other studies. One of best construction is made by Alberto Alesina and scientists from Harvard University (USA) and other institutions [Alesina, 2003, 155-194]. It may applied by different researches for analytical purposes. The authors indicate that earlier index which is used ELF is built on data available in Atlas narodov mira (the text is in Russian language). It is issued in 1964 [GUGK, 1964]. The data on ethno-linguistic structure of the population of each country is translated to mid 1961. The main drawback here indicates that they do not reflect contemporary political-territorial structure of the world. In general they already are outdated as quantitative indicators.

To overcome these shortcomings authors seek to build the index in the same way but with updated data. For greater opportunity use three indexes are calculated: for ethnic, language, and religious fragmentation.

The sources of data using for this purpose are: 1) Encyclopaedia Britannica – for the ethnic composition in 124 countries and territories around the world (some of which no longer exists); for language composition - in 200 countries and territories; for religious composition - in 214 countries and territories [Encyclopaedia Britannica, 2001]; 2) Levinson's reference book – for the ethnic composition in 23 countries [Levinson, 1998]; 3) reference book of Central Intelligence Agency to the U.S. government (CIA) – for the ethnic composition in 25 countries [CIA, 2000]; 4) International Group of Minority Rights in London - for the ethnic composition in 13 countries [MRGI, 1997]; 5) Official censuses data – for the ethnic composition in 4 countries. Additionally, to the population of Somalia are used data which are published in a scientific article.

A summary it can be said that of these 215 countries and territories, the index of ethnic fragmentations is set for 190 of them, the index of linguistic fragmentation – for 200, and the index of religious fragmentation – for 214.

The analysis of the results of this development shows that the following notices can be made:

1. Used data of the population compositions is not obtained by the same or nearly the same methodology.

2. There are large differences in time for the data on ethnic composition among countries.

3. Available data of official sources (censuses) are not used sufficiently in general, and for separate countries, too.

It is known that most of the data on ethnic, linguistic and religious composition of the population in separate countries are mined by official censuses. They generally apply the same or similar methodology.

Only a part of countries in the world put a direct question about ethnicity in the census questionnaire. There are some differences among countries. Relatively comparable data exists for the following countries:

1. In Europe: Russia, Estonia, Latvia, Lithuania, Belarus, Ukraine, Moldova, Romania, Bulgaria, Macedonia, Serbia, Montenegro, Bosnia and Herzegovina, Croatia, Slovenia, Hungary, Slovakia, Poland, the Czech Republic, etc. United Kingdom data is likely to become comparable in the next census because the current prevailing racial classification criterion.

2. In Asia - Kazakhstan, Kyrgyzstan, Tajikistan, Afghanistan, Uzbekistan, Turkmenistan, Mongolia, China, Vietnam, Philippines, Malaysia, Myanmar, Nepal, Sri Lanka, Iraq, Cyprus, Armenia, Georgia, Azerbaijan, etc. In countries like India, Bangladesh and Taiwan ethnicity is defined only to the part of population – usually for aboriginal (indigenous) people.

3. In Africa - Senegal, Gambia, Guinea, Mali, Sierra Leone, Liberia, Ivory Coast, Burkina-Faso, Ghana, Togo, Benin, Nigeria, Chad, Equatorial Guinea, Namibia, Swaziland, Lesotho, Mauritius, Malawi, Tanzania, Rwanda, Burundi, Uganda, Kenya, Ethiopia, Eritrea, Djibouti, Sudan, etc. Already any countries have been put an end to collect ethnicity data.

4. In Australia and Oceania - New Zealand, Fiji and some other countries and territories. In Australia the last census is counted the ancestry, which does not correspond fully to ethnicity.

Slightly more different approach is applied in Central and South America. The first group of countries determines the ethnic composition only to a part of the population. For example, some of them thereby determine the number of indigenous peoples and other groups (mainly race). That approach is used in Honduras, Costa Rica, Panama, Colombia, Venezuela, Bolivia, Chile, Argentina, Paraguay, and others. The second group of countries is determined mixed ethnicity and race. The second group of countries determines mixed ethnicity and race. That group includes Belize, Ecuador, Jamaica, Guyana, Surinam, Trinidad and Tobago and other countries and territories on the Caribbean islands. The third group of countries determines race - Brazil, Cuba and others.

Countries in North America as United States (U.S.A.) and Canada apply different approaches, too. U.S.A. specifies both race and ancestry. Canada counts ethnicity (ethnic origin) and “visible minorities”, which generally corresponds to the definition of racial background.

In some countries the issue of ethnicity is replaced by questions about the country of birth, nationality, race, origin, ancestry and others.

Situation to determine the language composition of the population is similar to that of ethnicity.

The question of religious affiliation is more complicated. At this point there are big differences, mainly on the details of the classifications of religious trends. This makes the data (where collected) incomparable.

Total uniformity of the obtained data can not expect soon and probably quite forward in the future. These data are obtained from various sources and in many cases are estimates without sufficient justification. Thus, the computed indexes are approximate and ultimately are not comparable, but the comparability is a fundamental requirement in the analysis. It can be achieved only for limited number of countries and territories.

The situation is such with the comparability of data over time. The used data have been collected during a different time. In some cases the difference among countries reaches more than 15 years. If these data are applying now, the difference may be more than 20 years. This is more or less important for separate countries. It depends on the development of ethnic and linguistic processes. The ethnic, linguistic, and religious composition of the population is changed more quickly or more slowly over time. It can not be separated from general social development of the country from changes in the economy, political processes, etc. The demonstration of relevant ethnic affiliation is the result of the interaction and integration of these factors. The speed of change is different and the change is very diverse in each country. For example, ethno-consolidation process is running in many African countries. Its result is a reduced fragmentation. Conversely, in some countries, which are an attractive center of immigrants, the degree of ethnic diversity has increased. The comparative analysis has to "catch" it. Also, the development of science should take into account and particularly - the classification of ethnic groups and languages, and the degree of kinship between them.

Regarding the sources of data on ethnic, linguistic and religious composition should give priority to official sources, such as censuses. This resource is not used sufficiently.

Usually, these sources of census data about ethnic, linguistic, and religious composition are:

1. United Nations Demographic yearbook. There are special tables which contain census data for many countries and census years [Demographic Yearbook, 1956, 1963, 1971, 1973, 1979, 1983, 1988, 1993, 2005-2006].
2. Official publications of the national statistical services and institutions. They may be in the form of books, websites, CD, etc.

Should also be noted that encyclopaedic and other reference publications in general can not allocate space to the respective theme and the style here is telegraphic. That sometimes hides any important details.

The main conclusion is following. Completely comparable index of fragmentation, heterogeneity, homogeneity, etc. is not possible to obtain for the present.

From a formal point of view on the use of indices as variables in statistics, mathematical models for analysis of various laws also has some peculiarities

The using of indexes as variables in econometric models also has some peculiarities. The indexes represent aggregate values, which have got a less variation. The dimension of index as numbers is important, when any disturbing events as heteroscedasticity are appeared.

4. INDICATIONS REFLECTING THE PROXIMITY (REMOTENESS) OF ETHNIC COMMUNITIES/GROUPS

In connection with the above features and clarify any misunderstandings, it is necessary to briefly consider the most important classifications of different ethnic communities and groups.

The most common classifications of people are [Osnovai ..., 1968, 18-34]:

1. *Language (linguistic) classification*. This is the most widespread classification. It is fined expression in affiliations of research language to language family, language group, and language subgroup. For example, the Bulgarians belong to Indo-European language family, Slavic language group, South- Slavic language subgroup. Most often (but not always) closest linguistic affiliation means common origin of the respective ethnic communities, similar culture, historical development and destiny, contacts, etc. Usually, the ethnic groups with greater linguistic propinquity to the main ethnic community in the country will be integrated more easily to it than those with more distant languages. This is assigned usually to states with significant influx of immigrants and proves by the practical development of that process. One available source of such classification is a guide "Naselenie mira" (in Russian language) [Bruk, 1981]

2. *Geographical classification*. It is fined expression in co-habitation of many ethnic communities on geographical and historical area which may include several countries. This implies the existence of similar traits of behavior, character, culture, etc. These ethnic communities may speak similar or completely different languages, and even confess different religions. For example, the ethnic communities inhabiting on the Balkans have a like mentality, although speak different languages (Slavic, Romance, Turkic, and others), and have a different religion.

3. *Economic and cultural classification*. That classification is built on the base of adherence to a traditional type of economy - agriculture, stock-breeding, fishing, etc. At the present time it almost is lost its topicality, especially in Europe, but on other continents and countries is still in force.

4. *Racial classification*. It is expressed in belonging of the people to big races (Caucasoid, Congoid, Capoid, Mongoloid, Australoid) and their divisions. Integration processes are easier where external physical differences between people do not provoke any negative reactions. Conversely, major differences can lead to isolation of the groups and hinder integration processes.

5. *Religious classification*. It is characterized in belonging of the people to particular religion and its divisions. The religion has very great significance in some countries and regions in the world. The uniformity and similarity of religion usually stimulates integration processes. Otherwise its differences sometimes may develop any drastic conflicts

On the basis of these classifications, which reflect proximity or remoteness among ethnic communities, should seek the appropriate measurement of ethnic homogeneity or heterogeneity (fragmentation).

5. CONSTRUCTION OF THE COEFFICIENT OF ETHNIC HOMOGENEITY (KEE)

The ethnic homogeneity of population is determined by many factors. One of the main is the quantitative factor. The integration processes are slowed or accelerated. They are dependent on number of persons belonging to the ethnic group or community - whether it is numerous or relatively small. Understandably, the more numerous groups more difficult and more slowly are included in the main ethnicity of the country, if any extraordinary conditions have not existed. Accordingly, small ethnic communities more easily lose its detachment. That does not depend on the number of ethnic communities/groups but on their quantify participation in the country's population. For example, the processes in Bulgaria are

in the interaction among three main ethnic communities - Bulgarian, Turkish and Roma (Gypsy). All other ethnic groups only give variety to ethnic processes.

All indices of measuring the ethnic, linguistic or religious diversity which are mentioned at the beginning report only quantitative factor.

One new indicator of ethnic homogeneity in the rough could be built by following procedure:

The procedure needs a census data. The number of persons who have not self-identification, and have not responded to the question of ethnicity, and mixed group "Other" must subtract from total population.

Thereafter, the registered ethnic groups and communities must be ranked in descending order of number of persons belonging to each of them. The most numerous group/community receives rank 1, the second numerous group – rank 2, etc.

Farther each rank is multiplied by the number of persons belonging to group/community. The products must add. The sum of these products is divided by the reduced total population.

The resulting figure represents the arithmetic mean of ranks of all recorded ethnic groups/communities. Here ranks of more numerous ethnic groups/communities will "weight down".

Formally, the ratio may be present as:

$$KEE1 = \sum R_i N_i / \sum N_i, \quad (3)$$

where R_i – the rank of ethnic group/community,

N_i – number of persons belonging to the ethnic group/community.

The number of persons belonging to the ethnic community/group may be replaced as a percentage. Results will be the same.

The lower limit of that ratio is 1. Theoretically upper limit can be quite large but values greater than 10 are very rare in practice. The ratio is closed to 1 as ethnic homogeneity is bigger and vice versa. The same coefficient can be used to measure linguistic and religious homogeneity.

This ratio also indicated reflect only quantitative "burden" of ethnic community/group. It does not reflect the degree of proximity between ethnic communities and groups and especially - the basic (most numerous) ethnicity in the country. What are the main factors determining this close?

6. CONSTRUCTION OF INDICATORS SCALE

The situation requires a construction of more precise coefficient of ethnic homogeneity. It has to reflect the degree of proximity among ethnic (languages and religions) groups. The following considerations could be take into note.

The degree of proximity may determine to the most numerous ethnic community/group in the country. Rank scale of measurement relatively has biggest opportunities. That does not preclude using other scales.

Thus, the arrangement of ethnic communities/groups reflects only their numbers. It will be changed by the degree of proximity on different signs. Method of "additive values" is used for this purpose [Torgerson, 1952, 401-419].

Regarding *the language* all other ethnic communities/groups could be arranged by rank scale as follows: a) the same language - Rank (-2), b) the same linguistic subgroup - Rank (-

1); c) another language subgroup, the same language group - Rank (0) d) other linguistic group, the same linguistic family - Rank (1) d) other language families - Rank (2). The sum of ranks gives zero. Arrangement can be traced in Table no. 1.

The logic of this scale is the following. From a formal point of view negative ranks will reduce the sum of the values in the coefficient numerator. That will lead to the reduction of ratio. The result is bigger ethnic homogeneity. Otherwise, positive ranks will increase the coefficient numerator and thereafter - to increase its value. This will be an expression of less ethnic homogeneity.

Table no. 1 - Language proximity of ethnic communities/groups

Proximity	Degree	Scale
1.The same language	Concurrent	-2
2. The same linguistic subgroup	Very close	-1
3. Another linguistic subgroup, the same language group	Near	0
4. Another language group, the same linguistic family	Not very close	1
5. Other linguistic family	Distant	2

Identification of ethnic (linguistic) groups in the population census needs adequate scientific classification of ethnic groups, their languages, and the degree of kinship between them. Sometimes these classifications are defective or absent. Therefore the census organizers are forced to isolate group "unknown" in publications of results.

For *geographical proximity* the classification of ethnic community/group could be the following: 1) inhabited mainly in the same country (state), 2) inhabited mainly in neighbouring countries, 3) inhabited mainly to another country, but on the same continent, 4) inhabited mainly on another continent. An intermediate proximity is "inhabited mainly in no neighbouring country, but from the same geographical and historical area". There are any difficulties at determination whether a country belongs to the concrete geographical and historical area or not.

The same difficulty arises when ethnic group do not have own state and do not come from other countries where they mainly inhabit. For example, these groups include Roma (Gypsies), Wallachian, Karakachan, and others in Bulgaria. Such is the case with many ethnic groups in Africa, Asia and other parts of the world. A settlement by compromise is that the group inhabits in the same country. Any substantial difficulties may arise if it is assumed that the group should be classified according to the country where its biggest part inhabits.

The logic of construction and scale is similar to the language proximity.

Table no. 2 - Geographical and historical proximity of ethnic communities/groups

Proximity	Degree	Scale
1.The same country	Concurrent	-2
2. The neighbouring country	Very close	-1
3. An other country, the same geographical and historical region	Near	0
4. An other country, the same continent	Not very close	1
5. An other continent	Distant	2

Economic and cultural proximity in the present time has got sense only for some countries in Asia, America, Africa and Oceania. There are ethnic communities with relatively primitive forms of economy. At the same time, communities at a higher level of development also inhabit here.

That economic and cultural proximity is important only for communities in same or neighbouring countries. Many migrants from distant countries (other regions and continents) who have been practiced similar forms of economy could not consider so close with native ethnic groups. In practical terms the inclusion of this proximity is difficult.

Racial proximity or remoteness furthers preservation or more rapid integration of separate ethnic group. Roma (Gypsies) are a typical example in this respect. Perhaps their racial type is contributed to keep them significantly to our days. They belong to Caucasoid human race but to one of its small races – North-Indian. Some distinguishing features (notably a darker skin colour and hair) are different between Roma and majority of European people. Over time, these differences have become fuzzy but most of Roma apparently differ from local population. The differences among representatives of main big races are more distinct.

In this regard it should be noted that clean racial ethnic communities/groups practically do not exist, but certain racial characteristics prevalent in each of them.

Here the proximity among major races could be set with a great deal of conditionality (Table no. 3).

The classification of small races, race types and variations is very difficult because generally it usually requires special erudition. The determination of racial structure usually requires special anthropological studies.

It should be given to the following. In most countries around the world race population differences are not essential and not hold the attention of society. In others, but very few countries where the differences are obvious, this creates certain problems. These countries record the race in the population censuses. Such countries are U.S.A., Brazil, South Africa, part of the countries belonging to the Caribbean and others. Now it should be noted that this registration is not made on purely scientific criteria, but most often in appearance, origin, adherence to a certain lifestyle, etc.

Therefore the inclusion of the proximity or remoteness is not important for all countries.

Table no. 3 - Racial proximity of ethnic communities/groups

Proximity	Degree	Scale
1. The same main race	Close	-1
2. Mix between main races and contact races	Not very close	0
3. An other main race	Distant	1

Religious proximity or remoteness also is important. It may use simple classification: 1) the same religion, 2) other (Table no. 4).

Table no. 4 - Religious proximity of ethnic communities/groups

Proximity	Degree	Scale
1. Same religion or same religion predominantly	Close	-1
2. Another predominant religion or another religion	Distant	1

The determination is made on the principle of prevalence, but that simple classification does not reflect the full complexity of the matter. Here certain difficulties arise when separate ethnic group profess many religious denominations and not one of them prevails. For this purpose, there must be evidence of the proportions among the different denominations.

There are many difficulties by statistical nature.

Some of the most important are: a) census question of religion is not always requires self-identification (an answer is not obligatory) and possibly a large proportion of the population will not answer, b) the classification of religious denominations may be differ in separate countries, c) furthermore, many religious denominations have got various directions; sometimes the differences among them may be quite strong.

In connection with the above it should be noted that most often mixed population usually inhabits in countries which are located next to borders among main religious denominations - Christianity, Islam, Buddhism, etc. In this case the influence of the religious factor is not correct to be isolated from the overall socio-economic and historical development of the country and region. Conflicts are generated in religious heterogeneous countries (for example, Bosnia and Herzegovina) and in religious homogeneous countries (for example, Tajikistan, Somalia, etc.).

7. CONSTRUCTION OF THE EXPANDED (IMPROVED) COEFFICIENT OF ETHNIC HOMOGENEITY (KEE)

From a theoretical point of view all indications of proximity to the basic (predominant) ethnic community/group have to be included. It is desirable but not always possible. In practical terms the inclusion of language and territorial proximities only is most reasonable. The information shall not be altered (at least sufficiently long interval of time) and not to contain implicit indications - common origin, common cultural traits, etc.

Furthermore, the inclusion of more indications may reduce excessively the coefficient. It may fall below the theoretical limit - 1 or becomes negative, which openly is unacceptable. In this perspective other possible solutions have to seek.

On this base the expanded coefficient of ethnic homogeneity with the inclusion of language proximity is the following:

$$KEE2 = \left\{ R_1 N_1 + \sum_{i=2}^N (R_i + LG_i) N_i \right\} / \left(N_1 + \sum_{i=2}^N N_i \right), \quad (4)$$

where R_1 – the rank of main (the most numerous) ethnic group ($R_1 = 1$),

N_1 – number of persons belonging to the main ethnic group,

R_i – rank of ethnic group,

LG_i – rank of language proximity,

N_i – number of persons belonging to the ethnic group.

The second coefficient is more prudent estimate of proximity and the differences among ethnic groups. It change the ranking of the respective ethnic groups depending on the proximity of a given ethnic group to the main (most numerous) ethnic group of the country. Then other groups which are more close to the main ethnic group will receive greater weight.

When the ranks of geographical proximity are included the coefficient is transformed. Its form is:

$$KEE3 = \left\{ R_1 N_1 + \sum_{i=2}^N [R_i + (LG_i + GP_i)] N_i \right\} / \left(N_1 + \sum_{i=2}^N N_i \right), \quad (5)$$

where GP_i - rank of geographical proximity. The other symbols are known.

For comparative purposes among countries and territories the coefficient always must be determine in the same way, despite what items are included or excluded from it.

The ethnic homogeneity should classify for more precise definition. One possible classification is contained in Table no. 5.

Table no. 5 – Classification of ethnic homogeneity coefficient (KEE)

Group	Coefficient	Homogeneity
1	To 1.1	Very large
2	1.2 – 1.4	Large
3	1.5 – 2.0	Significant
4	2.1 – 3.0	Moderate
5	3.1 – 5.0	Small
6	5.1 and more	Very small

8. PRACTICAL TESTING OF KEE

Practical testing of that approach is applied on data of ethnic groups in Bulgaria which are recorded by 2001 census. Auxiliary values for determining the coefficients KEE1 and KEE2 are contained in Table no. 6.

The number of persons belonging to the separate ethnic group is put in column 1. They are arranged in descending order (from largest to the smallest). The sum of these numbers (7836345) is located on bottom of the table. The sum in this column is less than the total population (7928901). The difference is due to the exclusion of groups "others", "no self-identify", and "no answer".

Column 2 contains ranks of ethnic groups. The products of ranks and numbers are put in column 3. The sum of these products (9740451) is on the bottom of table.

Then the results according formula (3) are follows:

$$KEE1 = 9740451 / 7836345 = 1.2340.$$

According to the classification in Table no. 5 on this scale Bulgaria's population has got large ethnic homogeneity.

Table no. 6 - Calculation of ethnic homogeneity coefficient of the population in Bulgaria according to 2001 census data

Ethnic group	Number	R_i	$R_i N_i$	LG_i	$R_i + LG_i$	$(R_i + LG_i) N_i$
a	1	2	3=1x2	4	5=2+4	6=1x5
Bulgarian	6655210	1	6655210	x	x	x
Turkish	746664	2	1493328	2	4	2986656
Roma	370908	3	1112724	1	4	1483632
Russian	15595	4	62380	0	4	62380
Armenian	10832	5	54160	1	6	64992

Ethnic group	Number	R_i	$R_i N_i$	LG_i	$R_i + LG_i$	$(R_i + LG_i) N_i$
a	1	2	$3=1 \times 2$	4	$5=2+4$	$6=1 \times 5$
Macedonian	5071	7	35497	-2	5	25355
Karakachan	4107	8	32856	1	9	36963
Greek	3408	9	30672	1	10	34080
Ukrainian	2489	10	24890	0	10	24890
Arab	2328	11	25608	2	13	30264
Tatar	1803	12	21636	2	14	25242
Jewish	1363	13	17719	2	15	20445
Romanian	1088	14	15232	1	15	16320
Polish	825	15	12375	0	15	12375
Vietnamese	635	16	10160	2	18	11430
Gagauz	540	17	9180	2	19	10260
German	436	18	7848	1	19	8284
Serbian	422	19	8018	-1	18	7596
Circassian	367	20	7340	2	22	8074
Czech	316	21	6636	0	21	6636
American	293	22	6446	1	23	6739
Albanian	278	23	6394	1	24	6672
French	195	24	4680	1	25	4875
Hungarian	169	25	4225	2	27	4563
Slovak	161	26	4186	0	26	4186
Kurdish	147	27	3969	1	28	4116
African	78	28	2184	2	30	2340
Slovene	28	29	812	-1	28	784
Bosnian	23	30	690	-1	29	667
Total	7836345	x	9740451	x	x	4984778
KEE	x	x	x	1.2430	x	1.4854

Source: Census of population, housing and farms in 2001. Vol. 1. Population, Paper 1, Demographic and social characteristics of the population. Tabl.8.1.14. Population by ethnic group, place of residence and sex, p.181, National Statistical Institute, Statprint, Sofia, 2004.

In column 4 ranks are assigned to language proximity to the Bulgarian ethnic group. For example, the Turkish language belongs to another language family (Altaic) - Rank 2, Roma (Gypsy) language belongs to another group (Indo-Aryan), but the same language family (Indo-European) - Rank 1, the Russian language belongs to the same group (Slavic) - Rank 0, etc.

New ranks of ethnics groups are obtained in column 5. For example, the rank of Turkish ethnic groups is $(2 + 2 = 4)$, Roma (Gypsy) - $(3 + 1 = 4)$, Russian - $(4 + 0 = 4)$, etc. Column 6 contains the products of these ranks by the number of the ethnic group. The total sum of these products is 4,984,778. It must be added the number of the most numerous ethnic group - Bulgarian (6655210).

The next result is obtained according formula (4):

$$KEE2 = \{6655210 + 4984778\} / (7836345) = 11639988 / 7836345 = 1.4854.$$

According to the classification in Table no. 5 on this scale Bulgaria's population is between moderate and large ethnic homogeneity with regard to the linguistic proximity of other ethnic groups to Bulgarian. This result is due to the fact that the language of largest

ethnic group after the Bulgarian - Turkish belongs to another family, and the next - Roma (Gypsy) - to the same linguistic family, but another language group.

Necessary parameters are defined in Table no. 6A to determine the coefficient only taking into account both linguistic and geographical/historical proximity (KEE3).

Table no. 6A - Calculation of ethnic homogeneity coefficient of the population in Bulgaria according to 2001 census data

Ethnic group	Number	R_i	LG_i	GP_i	$R_i + LG_i + GP_i$	Product
a	1	2	3	4	5=2+3+4	6=1x5
Bulgarian	6655210	1	x	x	x	x
Turkish	746664	2	2	-1	3	2239992
Roma	370908	3	1	-2	2	741816
Russian	15595	4	0	1	5	77975
Armenian	10832	5	1	2	8	86656
Wallachian	10566	6	1	-2	5	52830
Macedonian	5071	7	-2	-1	4	20284
Karakachan	4107	8	1	-2	7	28749
Greek	3408	9	1	-1	9	30672
Ukrainian	2489	10	0	1	11	27379
Arab	2328	11	2	2	15	34920
Tatar	1803	12	2	1	15	27045
Jewish	1363	13	2	2	17	23171
Romanian	1088	14	1	-1	14	15232
Polish	825	15	0	1	16	13200
Vietnamese	635	16	2	2	20	12700
Gagauz	540	17	2	-2	17	9180
German	436	18	1	1	20	8720
Serbian	422	19	-1	-1	17	7174
Circassian	367	20	2	1	23	8441
Czech	316	21	0	1	22	6952
American	293	22	1	2	25	7325
Albanian	278	23	1	1	25	6950
French	195	24	1	1	26	5070
Hungarian	169	25	2	1	28	4732
Slovak	161	26	0	1	27	4347
Kurdish	147	27	1	2	30	4410
African	78	28	2	2	32	2496
Slovene	28	29	-1	1	29	812
Bosnian	23	30	-1	1	30	690
Total	7836345	x	x	x	x	3509920
KEE	x	x	x	x	x	1.2972

In column 4 ranks are assigned to geographic/ historical proximity. For example, the Turkish ethnic group dwell mainly in neighbouring countries - rank (-1), Roma (Gypsy) - in the same country - rank (-2), Russian - in another country, but on the same continent - rank 1, etc.

The sum of all ranks - for ranking in number, of language and geographic proximity is in column 5. For example, the Turkish ethnic group - (2 + 2 - 1) = 3, Roma (Gypsy) - (3 + 1 - 2) = 2, Russian - (4 + 0 + 1) = 5, etc.

The products of the sum of ranks and the number of the ethnic group are in column 6. The sum of these products is 3,509,920. This sum is added to the number of the most numerous ethnic group - Bulgarian (6,655,210).

Then the third result according formula 5 is:

$$KEE3 = \{6655210 + 3509920\} / (7836345) = 10165130 / 7836345 = 1.2972.$$

From the resulting figure the addition of geographical proximity in this case is increasing ethnic homogeneity, because the two largest ethnic communities (excepting Bulgarian) in Bulgaria inhabit mainly the same and neighbouring countries.

Applying of discussions approach to different countries gives the results shown in Table no. 7.

Table no. 7 - Ethnic homogeneity coefficient of some countries

Country	Year	KEE1	KEE2	KEE3
Bulgaria	2001	1.243	1.485	1.297
Romania	2002	1.203	1.377	1.267
Serbia	2002	1.730	1.727	1.570
Poland	2002	1.039	1.040	1.024
Czech Republic	2001	1.189	1.098	1.018
Myanmar	1983	2.148	2.475	1.982
Kenya	1989	5.326	5.505	3.983

The practical application of the three versions of the coefficient shows some interesting conclusions.

In countries where the bulk of the population belongs to one ethnic community three options are generally close in value. This case is typical for countries in Europe, some countries in Asia, and on other continents.

In countries where one of the most numerous ethnic group, but alongside it there are other relatively numerous ethnic groups, differences between the three versions of the coefficient can be quite large. Usually in these countries have several centers of ethnic integration. This case is typical for most countries in Africa, some areas in Asia and other continents. For example Kenya magnitudes are (1989 years): 5.326, 5.505 and 3.983. According to the classification in Table no. 5 in the first and second variant population is very small, according to a third - with little ethnic homogeneity. The difference is due to the fact that many of ethnic communities/groups are cognate of the largest and also inhabited wholly or mainly in that country.

9. CONCLUSIONS

1. Due to various reasons, most countries in the world have a diverse ethnic composition of the population.
2. Summary indicators of ethnic homogeneity or heterogeneity have to build for different purposes – research, management, etc.
3. Most frequently used summary indicators of ethnic diversity are ELF (index of ethno-linguistic fragmentation) and ENET (index of ethnic disparity).
4. All indicators reflect only quantitative ratio (proportions) of different ethnic, linguistic or religious groups in the population.

5. The sources of data to construct indicators of ethnic diversity are various - censuses, estimations, registers and others. The data are derived by different methodology and often have great distance in time. Therefore, they are incomparable.

6. The computed indices are based on various data sources for the various population groups. They have got an approximate and tentative character.

7. Under existing conditions it is not possible to determine the indexes for each country in the world in terms of comparability.

8. A new indicator of ethnic homogeneity is proposed. It is built on Ranks scale of measurement. It has got three variants: a) taking into account only the quantitative ratios between ethnic and other groups (as well as all indexes discussed above); b) taking into account linguistic proximity of all other groups to the basic ethnic group in the country; c) when reporting both linguistic and geographical/historical proximity. The last two variants report quality differences among ethnic (and other) groups. Six degrees of homogeneity are differentiated.

9. Practical application of proposed indicator shows the following: a) in countries with one predominant ethnic group, the three variants are close in value; b) in countries with more numerous ethnic groups, the three variants may be quite different.

References

- [1] Alesina A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R., "Fractionalization", *Journal of Economic Growth*, 8, 2003.
- [2] Alesina, A., La Ferrara, E., "Ethnic Diversity and Economic Performance", *Journal of Economic Literature*, 3/43, 2005.
- [3] *Atlas narodov mira*, GUGK, Moskva, 1964.
- [4] Bruk, S.I., *Naselenie mira, Etho-demograficheskii spravochnik*, Nauka, Moskva, 1981.
- [5] Das, J., DiRienzo, C., "Is Ethnic Diversity Good for the Environment? A Cross-Country Analysis", *The Journal of Environment & Development*, 1/19, 2010.
- [6] *Demographic Yearbook 1956, 1963, 1971, 1973, 1979, 1983, 1988, 1993*, United Nation, N.Y.
- [7] *Demographic Yearbook, Special Census Topics 2000 Round*, United Nation, N.Y., 2006-2007, at <http://unstats.un.org/unsd/demographic/products/dyb/dyb2.htm>, accessed on April 1, 2010.
- [8] Emigh, R. J., Fodor, E., Szelényi, I., "The Racialization and Feminization of Poverty?", *On-Line Working Paper Series*, 2001, California Center for Population Research, UC Los Angeles, at http://escholarship.org/uc/ccpr_olwp, accessed on April 1, 2010.
- [9] *Encyclopedia Britannica 2001*, Encyclopaedia Britannica, Inc., Chicago, 2001.
- [10] Fearon, J.D., Laitin, D.D., "Ethnicity, Insurgency, and Civil War", *American Political Science Review*, 1/97, 2003.
- [11] Fielding, D., Torres, S., "A Simultaneous Equation Model of Economic Development and Income Inequality", *Journal of Economic Inequality*, 4, 2005.
- [12] Genovese, N., La Spada, M.G., "Diversity and Pluralism: An Economist's View", *FEEM Working Paper*, 62, 2006, at <http://papers.ssrn.com/sol3/>, accessed on April 1, 2010.
- [13] Griliches, Z., "Distributed lags: A survey", *Econometrica*, 1/35, 1967.
- [14] Hegre H., Sambanis N., *Sensitivity Analysis of the Empirical Literature on Civil War Onset*, Centre for the Study of Civil War, PRIO & Yale University, Yale, 2004.
- [15] Kim, H. S., *The Determinants of a Difference Between Conflict and Non-Conflict Countries in the Developing World*, Georgia Political Science Association • Conference Proceedings • 2008, at www.ciaonet.org/olj/shjdir/v7n2/v7n2_09.pdf, accessed on April 1, 2010.
- [16] Knack, S., Zak, P. J., "Building Trust: Public Policy, Interpersonal Trust, and Economic Development", *Supreme Court Economic Review*, 10, 2002.

- [17] Kuznets, S., "Economic Growth and Income inequality", *American Economic Review*, 4, 1955.
- [18] Laitin, D., Posner, D., "The implications of constructivism for constructing ethnic fractionalization indices", *APSA-CP*, 1/12, 2001.
- [19] Leigh, A., "Does equality lead to fraternity?", *Economics Letters*, Elsevier, 1/93, 2006.
- [20] Levinson, D., *Ethnic Groups Worldwide, A Ready Reference Handbook*, Oryx Press, Phoenix, 1998.
- [21] *Osnovii etnofrafi*, Visshaia shkola, Moskva, 1968.
- [22] *Portable Data Dictionary*, State Failure Task Force Phase III Model Variables, Data Sources, 2002, at <http://globalpolicy.gmu.edu/pitf/Phase%20III%20Data%20Dictionary.pdf>, accessed on April 1, 2010.
- [23] Reynal-Querol, M. "Political Systems, Stability and Civil Wars", *Journal of Conflict Resolution*, 1/46, 2002.
- [24] *The World Factbook*, Central Intelligence Agency (CIA), 2000, at <https://www.cia.gov/library/publications/the-world-factbook/>, accessed on April 1, 2010.
- [25] Torgerson, W. S., "Multidimensional scaling: theory and method", *Psychometrika*, 4/17, 1952.
- [26] Vanhanen, T., "Domestic Ethnic Conflict and Ethnic Nepotism", *Journal of Peace Research*, 1/36, 1999.
- [27] *World Directory of Minorities and Indigenous People*, Minority Rights Group International (MRGI), London, 1997, at <http://www.minorityrights.org/directory>, accessed on April 1, 2010.