



FONDATION POUR LES ETUDES ET RECHERCHES SUR LE DEVELOPPEMENT INTERNATIONAL

**What is hidden behind
the indicators of ethno-linguistic fragmentation?**

**by
Kelly Labar**

Document de travail
Série « Indicateurs de développement innovants » / Idi10
Juin 2010

La FERDI est une fondation reconnue d'utilité publique.
Elle met en oeuvre avec la Fondri l'Initiative pour le Développement et la Gouvernance Mondiale (IDGM).

Abstract :

Ethno-linguistic fragmentation has become an unavoidable factor to consider when studying the determinants of growth. In their article of 1997, Easterly and Levine inaugurated the argument that a fragmented countries' ethnic structure, measured by the ethno-linguistic fragmentation index ELF, negatively influences countries' growth. Since this research, further analyses have either validated or challenged the method used by Easterly and Levine to describe ethnic diversity as well as the results they emphasized. The present research provides consequently an overview of the indicators used by the literature to describe and measure ethno-linguistic fragmentation and the argument put forward by the various authors to support the one or the other indicator. It also challenges the exogenous character of the main ethno-linguistic fragmentation indexes usually assumed in the studies analyzing the link between ethnic fragmentation and economic performance. Having a look at the correlations between exogenous characteristics such as the country's surface, the population density and these main indicators, the article provides potential instrumental variables to be used to control for endogeneity of ethno-linguistic fragmentation index when estimating the impact of ethnic fragmentation on economic performance.

Keywords: ethno-linguistic fragmentation, endogeneity.

1. Introduction

For more than ten years, ethnic fragmentation has become an important factor to consider when studying the determinants of economic performances. The central work of Easterly and Levine (1997) argues that ethnic fragmentation negatively influences economic growth through rent-seeking behavior and lack of political consensus impeding the efficient and equal distribution of public goods. These authors specifically associate these conclusions to African countries characterized by a high level of ethnic fragmentation. Since this controversial study, a huge bench of research has analyzed what is behind Easterly and Levine's findings. The literature studying the link between ethnic fragmentation and economic performances broadly distinguishes between direct and indirect effect of the former on the latter (see Alesina and La Ferrara (2005) for a survey of the literature). The direct link is characterized by the negative effect of ethnic fragmentation on macro-economic elements like economic growth or the provision of public goods (Easterly and Levine (1997), Alesina, Baqir and Easterly (1999)). The indirect link can operate for instance through inter-ethnic conflicts or the lack of democracy (Collier (2001), Collier and Hoeffler (2002), Alesina and La Ferrara (2005)). Linked to linguistic barriers or to various levels of population density, problems of inter-community communication have been emphasized as a factor underpinning the negative relationship between population fragmentation and economic performances. Linguistic more than ethnic fragmentation stresses the communication barriers between different groups. On the other hand, Arcand, Guillaumont and Guillaumont-Jeanneney (2000) emphasize that the lack of communication among a country's population, proxied by illiteracy or low population

density may hamper discussions which would otherwise lead to more efficient policies and provision of public goods.

The more recent literature has cast doubt on fundamental elements like ethnic groups' definition, data or fragmentation indicators used to characterize and quantify tensions between communities (Fearon (2003), Bossert, La Ferrara and D'Ambrosio (2008)). For instance, should we consider more ethnic fragmentation or ethnic polarization to describe socio-economic tensions between ethnic groups? Is the ethnic structure of a country endogenous or exogenous to elements like for instance geography, and/or politics? The limits to the characterization of ethnic groups and to the measurement of ethnic fragmentation stressed above motivate the present research. First the analysis investigates what is behind the construction of the ethno-linguistic fragmentation indexes, their definition and measurement. Then, going deeper into the questioning of the exogenous character of ethno-linguistic fragmentation, the study investigates whether some exogenous geographical and structural characteristics like population density and country's surface are significantly linked to ethno-linguistic fragmentation.

The rest of the paper is consequently organized as follows. In section 2, the literature review sums up studies focusing on the definition of ethnicity and on the appropriateness of the data and indicators used to quantify ethnic fragmentation. Section 3 first develops the idea of endogeneity of ethnic fragmentation highlighted in the literature. Second, it analyzes correlations between different geographical and structural country's characteristics and indicators of ethnic fragmentation to shed light on the potential use of the former to instrument the latter in growth equations. Finally, section 4 concludes.

2. Ethnic fragmentation definition and characteristics

The notion of ethno-linguistic fragmentation (ELF) characterizes a country's ethnic structure going further than the number of ethnic groups living in the country. The number of ethnic groups does not include any information on each group's population, impeding to consider the demographic weight of each ethnic group in the total population. To tackle this limit, Taylor and Huston (1972) calculated the ELF index based on the application of the Herfindahl concentration index to the case of ethnic diversity. It measures the probability that two randomly drawn individuals belong to two different ethnic groups. The higher the ELF index, the more fragmented the country (Cf. Appendix 1 for more details on ELF index mathematical formulae). The application of this measure to countries ethnic structure faces three major issues developed below: how ethnic groups should be defined? To what database do and should we have access to empirically implement ELF index? Is the ELF index the right measure to use when one wants to characterize the tensions existing between ethnic groups?

The first concern about ethnic fragmentation is related to the existence of objective criteria allowing to differentiate between ethnic groups. Fearon (2003) extensively develops this issue

emphasizing the “borderline-arbitrary decisions” required to be able to build ethnic groups or categories. As Fearon stresses: “If there are multiple ways of listing a country’s “ethnic group”, we must be careful that we do not, in effect, choose the coding that best supports our theory, after the fact” (Fearon (2003), p198). This important consideration requires clear definitions of the classification criteria used. An important clarification of ethnic groups characteristics is carried out by Alesina *et al.* (2003) who have built a new ethnic groups’ dataset based on three main criteria: language, ethnicity and religion. The Atlas Narodov Mira used by much of the previous literature on ethnic fragmentation since the work of Taylor and Huston (1972) based its registration of ethnic groups on unclear criteria. Language appears to be the major differentiation criteria but other non specified ethnic characteristics are taken into account. Clarifying the characterization of ethnicity, Alesina *et al.* (2003) make a distinction between linguistic fragmentation and ethnic fragmentation, the second being based on physical characteristics like skin color for instance. The correlation between their fragmentation indexes and ELF index is relatively high but is higher with the linguistic fragmentation index. Reproducing Easterly and Levine (1997) estimations, the use of this new classification does not challenge the results on the negative link between ethnic fragmentation and economic performance (Alesina *et al.* (2003)). Two alternative characterizations of ethnic groups have stressed the limit of the previous grouping methodologies introducing the notions of cultural distance or political involvement. The introduction of cultural distance in the definition of ethnic groups has been considered by Fearon. Defining a resemblance factor based on language distance between ethnic groups, he strengthens the role played by communication underlining the fact that two different ethnic groups can easily share common interests thanks to their close cultures or similar languages (Cf. Appendix 2 for more details on the new formulae developed by Fearon (2003)). To my knowledge, Fearon’s new ethnic groups’ dataset has not yet been used to check for the robustness of Easterly and Levine’s (1997) or Alesina *et al.* (2003) findings. On the contrary, Posner (2004) has used a new concept of politically involved ethnic groups which challenged the negative link between ethno-linguistic fragmentation and economic performance. This author stresses the mismatch between the groups entering in the ethnic fragmentation measurement and the channels through which this fragmentation is meant to influence economic growth. Particularly, Posner reminds that ethnic fragmentation is said to negatively influence growth in negatively influencing policies and decision making. The author considers consequently that only ethnic groups participating in the political game are able to have a significant influence on the decision making process, leading the simple ethnic fragmentation index to be inappropriate. Based on this observation, he identifies the politically relevant ethnic groups and constructs the Herfindahl index from the new ethnic groups’ selection. The created index of ethnic group fragmentation is named PREG index (Politically Relevant Ethnic Groups index). Posner (2004) emphasizes then that his new selection of ethnic groups appears more relevant to quantify the impact of fragmentation on economic performance introducing the important dimension of political involvement in the debate.

The need for a clearer characterization of ethnic groups goes together with further developed databases regarding the dimensions defining ethnicity, the level of geographic aggregation of data and the timeframe considered. As mentioned above, the first limit of the Atlas Narodov Mira is that it suffers from unclear criteria in the identification of ethnic groups. Second, it is important to underline that the Atlas provides data for 1964 what can be seen as out-dated for studies considering recent economic data. Finally, Posner (2004) emphasizes inaccuracies when registering individuals belonging to the various ethnic communities. This issue resides in arbitrary levels of ethnic groups' aggregation. Newly created data sources presented above cope with the problem of ethnic group delineation differentiating between language and ethnic fragmentation (Alesina et.al. (2003), Fearon (2003)). Moreover, they provide more updated data (Alesina et.al. (2003), Fearon (2003), Posner (2004)). However, the level of aggregation of ethnic groups is still questioned. Having access to micro-data for a country allows a higher level of disaggregation thus going deeper into the ethnic structure of the country. Fedderke, Luiz and Kadt (2008) have created such a micro database for South Africa. Case studies on individual countries shed light on the central issue of the link between ethnic fragmentation and economic performances (Cf. section (2.2.)). Nevertheless, despite the wealth brought by microanalysis, huge work would be required to make inter-countries comparisons. Above the question of data updating and given that ethnic fragmentation is meant to influence economic performances which may vary through time, it is worth considering data on ethnic groups observed at different point of the time. This is the case of neither the Atlas Narodov Mira nor databases developed by Alesina *et al.* (2003) and Fearon (2003). The time dimension is taken into account by Fedderke, Luiz and Kadt (2008) and Campos and Kuzeyev (2007) who are investigating the temporal dimension of ethnic fragmentation (Cf. section (2.2.)). Posner (2004) provides decennia measures of ethnic fragmentation between 1960 and 1990.

Finally, the construction of the most common measurement of ethnic fragmentation, ELF index, has been criticized and challenged by the last years' literature. The simplicity of the ELF index impedes it to draw a complete picture of the various fragmented ethnic structures. Posner (2004) provides an example to support this caveat: "Take two hypothetical countries, the first with two groups of equal size and the second with three groups containing two-third, one-sixth and one-sixth of the population, respectively. In both countries, the fractionalization index [...] would be 0.5. Yet, the dynamics of the inter-group competition in each country would [...] certainly be different." (Posner (2004), p.851). The literature trying to address this limit of the ELF index has proposed notions which differ significantly from the initial method and measure. One is the ethnic polarization concept, which is meant to better capture the tensions potentially existing between major groups. Montalvo and Reynald-Queyrol (2005) who developed the ethnic polarization index argue for instance that two major ethnic groups can be more subject to conflict than many small ethnic groups. While the level of polarization is higher in the first situation, the fragmentation measured by ELF index would be higher in the second situation. Polarization appears consequently to better describe tensions impeding political consensus and efficient provision of public goods. That is why Montalvo and

Reynald-Queyroid (2005) highlight a negative and significant correlation between the polarization index and economic performance. Another method extending the ELF measurement of population fragmentation uses a broader set of social characteristics to define homogenous socio-economic groups. Bossert, La Ferrara and D’Ambrosio (2008) propose to consider the degree of similarity between individuals with regards to factors like the level of education, of income, the employment status and the ethnic origin. This last element is thus only one characteristic among others. The various factors are aggregated using the Principal Component Analysis method in order to build a variable synthetically defining homogenous groups. This new variable is then the one taken into account to build the GELF index (Generalized Ethno-linguistic Fragmentation index), following the usual ELF index formulae. Due to the construction of this index, when individuals differ only by their ethnic origins, the GELF index they propose is equal to ELF. Despite the extension represented by this new index, it can be perceived as too far from the initial goal of researchers consisting in describing ethno-linguistic fragmentation. It will consequently not be considered in what follows.

The six major ethnic fragmentation measures provided by the literature and used in the analysis of section 3 are consequently: the **ELF** index¹, the **Ethnic** and **Language** indicators provided by Alesina *et al.* (2003), Fearon’s new calculation of the ELF index (**Fearon**), Fearon’s indicator of cultural fragmentation (**Fearon_cult**) and Posner’s (2004) **PREG** indicators available for 1960, 1970, 1980 and 1990.²

Table 1: Correlations between the six major indicators of ethno-linguistic fragmentation.

	ELF	Ethnic	Language	Fearon	Fearon_cult
ELF	1				
Ethnic	0.7729	1			
Language	0.8755	0.7002	1		
Fearon	0.7818	0.8455	0.6937	1	
Fearon_cult	0.8697	0.6972	0.7142	0.7905	1
PREG	0.6557	0.4544	0.6632	0.4183	0.5040

Source: Author.

The correlations between ELF, Ethnic, Language, Fearon and Fearon_cult are equal or higher than 0.70 showing a relatively high level of correlation between these indicators (Table 1). The highest correlations are observed for the three indicators which consider the language as

¹ An extended version of the ELF index calculated by the author is used here. Actually, using the initial Atlas Narodov Mira allows us to consider a broader set of countries. The mathematical formulation is however the same as for the ELF index constructed by Easterly and Levine (1997) (Herfindahl index). The correlation between the re-calculated ELF index and the one used by Easterly and Levine (1997) is thus of about 0.995.

² For the correlations presented below, the PREG for the decades 1960, 1970, 1980 and 1990 is used. Given that this indicator evolves through time while the other four do not, a lower correlation between PREG on the one hand and ELF, Ethnic, Language, Fearon and Fearon_cult on the other hand is consequently expected.

the major ethnic characteristic differentiating ethnic groups (ELF, Language, Fearon_cult). The correlations with PREG are nonetheless lower as the PREG indicators are based on the very specific criteria of politically relevant ethnic groups. It may consequently be that the links between PREG and our geographical and structural variables of interest demonstrate other patterns than the links between these latter variables and the other indexes of ethnic-fragmentation. Appendix 3 sums up the various datasets on ethnic and ethno-linguistic fragmentation indexes available to this day. We note that most of the indexes use, despite its limit, the Herfindahl index as a measurement of fragmentation. The indexes differ consequently mainly through the datasets they are based on or the subcategories considered. Appendix 4 presents the countries' rank with regards to the various indexes. Even if some countries' ranks differ strongly between indexes, no clear pattern can be identified regarding which quantification of countries' ethnic fragmentation seems to be over or under estimated by one or the other indexes. The variation in the ranks shows however how important the ethnic groups' identification factors are. The case of Brazil presented in the Box 1 mirrors how the different identification criteria considered by the various ethno-linguistic indexes may affect the country's rank and stresses why one indicator can be more appropriate than another to describe Brazil's socio-ethnic problems.

Box 1: An example of contradictions between measures. The case of Brazil.

The values for Brazil ethnic fragmentation indexes and the respective world ranks are the following:

Table 2: Brazilian's ethnic fragmentation

Brazil	ELF	Ethnic	Language	Fearon	Fearon_cult
Values of fragmentation	0.0705	0.541	0.0468	0.549	0.02
Ranks / Countries number	31/149	111/184	21/191	85/146	6/129

We consequently note a clear difference between the values of the indexes but also of Brazil's world rank in terms of fragmentation. From the different elements presented above, these values can be easily interpreted for Brazil. Brazil presents itself as the rainbow country, mirroring the skin color diversity of its population. The country claims to be a harmonious melting pot bringing many ethnic groups together. This high diversity is at best visible in the ethnic fragmentation index "*Ethnic*" which takes mostly into consideration the ethnic character of individuals. We consequently see how high the value of this index is. However, other indexes more based on the language element show significantly lower values (ELF, Language, Fearon). Fearon's cultural diversity index is for instance particularly low, leading the country to be at the 6th place in the world ranking. One can then ask if the ethnic diversity is problematic in Brazil and which ethnic character underpins potential discriminations. The country's social stability and recent high economic growth can be presented as a proof of the secondary role played by ethnic fragmentation. However, it is worth noting that this country

faces very high level of inequalities, which have been shown to be closely related to individual's skin color (Lovell (1999)). The ethnic fragmentation as measured by Ethnic would then be more appropriate to describe which character of ethnic diversity may influence internal ethnic issues in Brazil. However, other conclusions may be emphasized in other countries, leading the language barrier to be the most important factor for instance. The variety of indexes in terms of measurement and ethnic groups' delineation mirrors the variety of countries' ethnic background. Can we then consider only one of these indexes when studying the link between ethnic fragmentation and economic performance? If yes, which one?

The multiplicity of indexes, data sources and of countries' rankings mirror the concern of many authors regarding the perfect exogeneity of ethnic groups delineation and consequently of ethnic fragmentation. While analyzing in more details how to define ethnic groups, considerations linked to the countries' borders, geography, population cannot be ignored. The exogeneity hypothesis has been consequently put into doubt. If ethnic fragmentation is endogenous to such factors, studies which have not taken this endogeneity into account while estimating the link between fragmentation and economic performance may lead to inconsistent results. The following section is dedicated to this specific second concern.

3. Why considering ethnic fragmentation as endogenous?

In case of ethnic fragmentation endogeneity, results provided by econometric estimations emphasizing a negative relationship between ethnic fragmentation and economic performance can be biased. The identification of a causal relationship going from ethnic fragmentation to economic performance cannot for instance be established without instrumenting the variable proxying for ethnic fragmentation. The endogeneity of ELF, if not taken into account would thus cast doubt on the negative relationship previously established between ethnic fragmentation and economic performance (Easterly and Levine (1997), Alesina *et al.* (2003)). Studies which mentioned and/or took into consideration this issue are summed up below and the link between potential exogenous variables like countries' surface or population density and ELF indexes is presented in a second step. This latter link has as a goal to emphasize potential instruments to be used more extensively in the future to address the issue of ELF endogeneity in econometric studies analyzing the relationship between ELF and economic performance.

3.1. The endogeneity of ethnic fragmentation in the literature

The potential endogeneity of linguistic fragmentation has been mentioned as soon as 1956 by linguists. Nettle (2000) quotes for instance Joseph Greenberg who mentioned that "our general expectation [...] is that areas of high linguistic diversity will be those in which communication is poor, and that the increase of communication that goes with greater economic productivity and more extensive political organization will typically lead to [...] the ultimate disappearance of all except a single language." (Nettle (2001), p.335). Nettle has also

put into question the exogeneity of linguistic fragmentation stressing a simultaneity bias between economic growth and ethnic fragmentation. His results emphasize a relationship going from economic performance to linguistic fragmentation, contrarily to the commonly highlighted causal link going from ethnic fragmentation to economic performance. Nettle justifies this matter of fact by the existence of possible common factors influencing both linguistic fragmentation and economic growth.

The economic literature investigating the endogeneity of ethnic and linguistic fragmentation has focused on its underpinning mechanisms. Studies often refer to history or sociological mechanisms able to influence the ethnic structure of a country. Leeson (2005) stresses the role played by colonialism in the perturbation of the pre-existing social structures which went beyond ethnic origins. For this author, the new institutions implemented in colonized countries destroyed the inter-actors' trust based on signals linked to social status (e.g. relationship to the authority, property usage or religious practices). They established instead a hierarchy often based on ethnicity creating tensions between ethnic groups. Apart from this historical theory, the evolution of ethnic structure linked to social contexts is presented in Alesina and La Ferrara (2005). They provide the example of ethnic mimicking which can occur when certain ethnic groups want to be assimilated to the majority one. These changes in ethnic identification can be transmitted to the data collected through changes in self-reported ethnic origin. Alesina and La Ferrara (2005) also stress the role played by migrations between or within countries due among other to conflicts or to better labor or well-being perspective in the foster country. Taking the example of the United States, Alesina and La Ferrara (2005) underline that "changes over time in the economic growth of different metropolitan areas have induced massive flows of migration that have sensibly altered some cities' ethnic composition" (Alesina and La Ferrara (2005), p 791). Empirical studies have also attempted to take endogeneity of ethnicity into consideration in the estimation of the relationship between ethnic fragmentation and economic performance. Fedderke, Luiz and Kadt (2008) introduce the time series microdata for South Africa to investigate this relationship. Focusing on one of the major channel through which ethnic fragmentation influences growth, i.e. political instability, they emphasize a causal link going from ethnic fragmentation to political instability. This result consequently goes in the same direction as findings resulting from country comparisons and highlighting that a higher level of ethnic fragmentation is associated with higher conflict occurrence (Collier (2001)). However, given the temporal dimension and the micro level of their dataset, the authors stress that aggregated measures can bring misleading conclusions missing the evolution of linguistic assimilation for instance. They suggest consequently that more studies should be carried out at each country's level and through time. Campos and Kuzeyev (2007) consider clearly ethnic fragmentation as endogenous. They study the changes through time of ethnic fragmentation in former soviet countries before and after the dissolution of the USSR. They re-estimate the specification of Easterly and Levine (1997) for the different sub-periods and emphasize the same negative

relationship between ethnic fragmentation and growth.³ However, it is worth noting that the instrumental variables they use to control for the endogeneity of ethnic fragmentation are dubious. They use the infant mortality rate, the bank sector reform, the level of infrastructure and the price liberalization as instruments while these are also strongly correlated with economic growth.⁴

In view of the recent literature stressing the potential endogeneity of ethnic fragmentation, the present study investigates whether the ethnic fragmentation is linked to geographical or structural characteristics of countries. The link between countries' geographic characteristics and their ethnic structure is more supported by the historical or intuitive arguments than by empirical research. Introductory results based on correlations are provided in order to underline the need for going behind the ethnic fragmentation measure.

3. 2. Endogenous ethnic fragmentation: geographical and structural characteristics

Geographic characteristics are often mentioned in the literature as a source of potential endogeneity of countries' ethnic structure (Alesina and La Ferrara (2005), Campos and Kuzeyev (2007), Cederman, Rod and Weidmann (2007)). The country's latitude is among the instrumental variables used by Campos and Kutzeyev (2007) to instrument ethno-linguistic fragmentation in the growth equation.⁵ Cederman, Rod and Weidman (2007), focusing on the role played by ethnic fragmentation in conflicts occurrence underline the necessity to consider ethnic geographical repartition in analyses. They stress the role played by rough terrain in the guerilla longevity. It may consequently be that the ethnic structure of countries put forward as influencing economic performance is endogenous to geographical or structural characteristics which themselves are more or less correlated with growth. In that case, the causal relationship going from ethnic fragmentation to economic performance is to be challenged. A more rigorous estimation of this link would then require the use of instrumental variable estimations, as done in Campos and Kuzeyev (2007) but using valid instruments. We argue here that geographical and structural characteristics like country's surface or population density could be good candidates as instruments. We investigate this possibility in the following study of correlations between these two factors and ethno-linguistic fragmentation indexes.

The geographic characteristic considered in the present study is the country's surface. As emphasized by Alesina and La Ferrara (2005), the country's surface is linked to geographical but also historical factors which have contributed to determine the borders. The presence of desert or forests may favor the definition of a bigger territory. If such geographical elements

³ Moreover, they cope with one of the criticisms against Easterly and Levine (1997). The inclusion of variables controlling for the channels through which ethnic fragmentation influences growth decreases the significance of the ethnic fragmentation variable.

⁴ The identification power of these instruments is consequently weak, casting doubt on their final results.

⁵ The hypothesis behind the use of latitude as an instrument is that it mirrors geographical conditions like average temperatures for instance. These geographical conditions may then influence the settlement of populations from different ethnic groups.

exist in a country, different ethnic groups with different culture can be established around or within the areas but belonging to the same country. The hypothesis behind would then be that bigger countries encompass a higher number of ethnic groups, and potentially a higher level of ethnic fragmentation. With regards to the politico-historical determination of countries border, Alesina and La Ferrara (2005) provide an example which sheds light on the link between country's surface determination and ethnic fragmentation: "*after the First World War the superpowers of Britain, France and the United States [...] redrew the world borders in a way that only partially reflected the goal of ethnic homogeneity; they were much more interested in grabbing for their allies as much territory as possible*" (Alesina and La Ferrara (2005), p.791). One can consequently expect that bigger territories cover more diversified ethnic groups.

A structural factor linked to geography and socio-economic behaviors is the population density. Correlated with what is presented above, one can expect that a more dense population results originally in lower ethnic fragmentation. When defining a country's main borders, the population or decision-makers may have looked for a homogenous culture and ethnic structure, particularly for countries with geographically delimited territories.⁶ If we consider European countries like France, Spain, Italy or Great Britain, their territories are delimited by mountains or seas. A national language was established and a homogenous culture emerged contributing to the ethnic homogeneity observed by the beginning of the 1960's (when ethnic groups were listed).

Data on geographical and structural characteristics are taken from the World Development Indicators 2008 for countries for which we have information on ethnic fragmentation. Due to issue of comparability between the different datasets, the French Departments d'Outre-Mer, the former German Democratic Republic and German Federal Republic as well as Czechoslovakia, Yugoslavia, Serbia and Montenegro are excluded from the sample. Regarding the measure of ethnic fragmentation, ELF, Ethnic, Language, Fearon, Fearon_cult and PREG presented previously are considered. The final samples are described in Table 3. Depending on the correlations measured, the sample size may vary. We remind that PREG is measured only for African countries what justifies the lower number of countries observed. The surface of countries considered is invariant through time. Surface and density will be measured in logarithm. Regarding the periods of observation, given that most ethnic fragmentation indexes are invariant through time, the average population density over 2000-2005 is considered. Summary statistics for countries' population is provided for information. In the special case of PREG, for which we have four different values of fragmentation for the years 1960, 1970, 1980 and 1990, the average population density over each decennium is calculated.

⁶ The notion of territory delimitation can be perceived as relative if we consider the colonization process. However, given that we begin our period of observation of ethnic fragmentation by the beginning of the 1960's, most colonization processes have ended and the countries' borders are consequently already determined and invariant through time.

Table 3: Descriptive statistics.

Indicator	Mean	Standard deviation	Number of countries
ELF	0.3901	0.2895	144
Ethnic	0.4361	0.2570	184
Language	0.3901	0.2785	190
Fearon	0.4835	0.2601	152
Fearon_cult	0.3122	0.2108	149
PREG	0.3616	0.2489	42
Country total surface (km ²)	640,722	1,857,228	209
Population density over 2000-2005 (people per km ²)	361.20	1,713.50	207
Population over 2000-2005 (billion people)	3.08	1.21	207

Source: Author.

A first global and visual picture of the link between ethnic fragmentation as measured by ELF index and geographic characteristics is presented on the world map provided below. Figure 1 includes information on the population density (countries' colored surface) and on the ethnic fragmentation represented by circles of varying size increasing with the level of ethnic fragmentation. From this map, we cannot really draw conclusions on the link between country's surface and the level of ethnic fragmentation. Countries like France and the United States of America (USA) show similar levels of ethnic fragmentation while China has a lower level. On the other hand, we see that high ethnic fragmentation is more often observed for countries with a low population density. This is the case for instance in Africa, in Latin America or in countries like Iran, Afghanistan.

Figure 1: World Map. Density and ethno-linguistic fragmentation measured by ELF index.

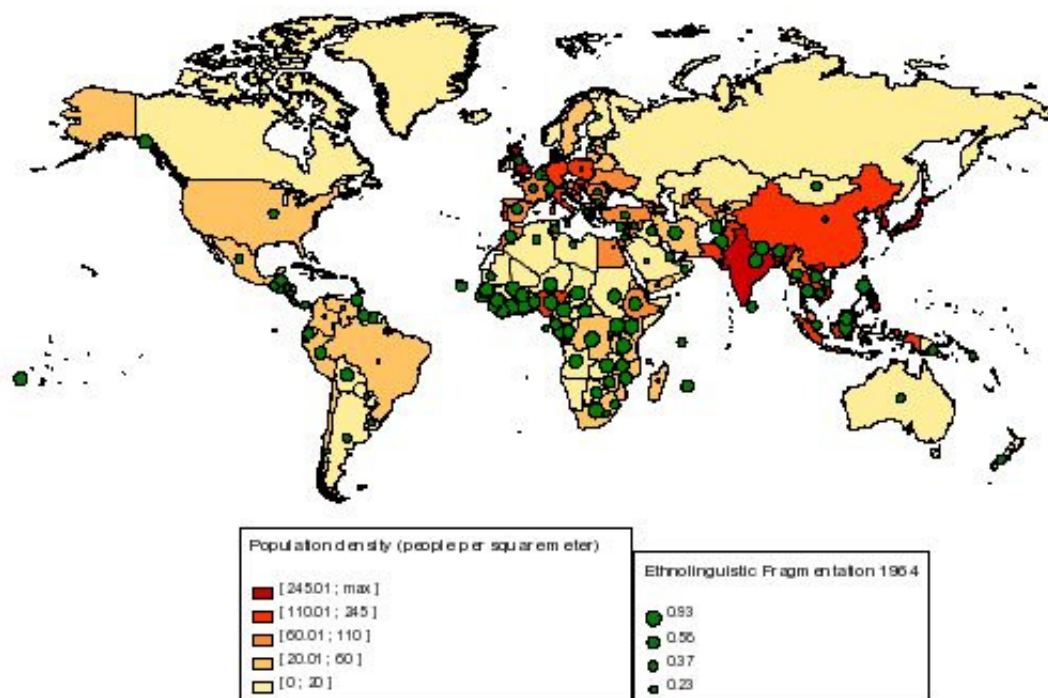


Table 4: Correlations between ethno-linguistic fragmentation indicators and geographic characteristics.

	ELF	Ethnic	Language	Fearon	Fearon_cult	PREG
Log Surface	0.2988***	0.2537***	0.1127	0.1748**	0.1401*	0.2793***
Log Population	0.1592*	0.1116	0.0595	-0.0343	-0.0262	0.4846***
Log Population density	-0.2504***	-0.2916***	-0.1085	-0.2588***	-0.2013**	0.1330**

Note: n.a. means Non Available. *, **, *** represent correlations significant respectively at 10, 5 and 1 percent.

Positive and significant correlations emphasized between the log of countries' surface and most ethnic fragmentation indexes support the idea that the bigger the country, the more ethnically fragmented it is (Table 4). This positive link decreases with the countries' surface as this last is measured in logarithm. This result highlights the non exogenous character of ethnic fragmentation. The country's surface being considered as perfectly exogenous, the significant correlation between this variable and ELF indexes stresses that country's surface may be used as a good instrument to control for the endogeneity of fragmentation variables. The fragmentation indexes for which the correlations are less or not significant are those mostly based on language characteristics to delineate ethnic groups. It seems consequently that language fragmentation is not as much linked to the country's size as is ethnic

fragmentation. In the case of Alesina *et al.* (2003) language fragmentation index and Fearon (2003) cultural fragmentation index, the country's surface would consequently be a weak instrument. We also notice that once one restricts the attention to the politically relevant ethnic groups, a positive and significant correlation between country's surface and the PREG index exists. The country's surface would thus be a relevant instrument for this index as well.

Switching to the pattern observed for population density, we note here a negative and significant correlation between ethno-linguistic fragmentation and the logarithm of population density. Countries which are denser appear consequently to have a lower ethnic fragmentation, thus more homogenous ethnic structures. As mentioned for the country's surface, this result once again casts doubt on the exogenous character of ethnic fragmentation but helps to provide another instrument for fragmentation in estimations of the link between ethnic fragmentation and economic performance. Similarly to the country's surface, the correlation is not significant for Alesina *et al.* (2003) measurement of language fragmentation. Population density would consequently be a weak instrumental variable for this index of language fragmentation. Finally, PREG index is significantly correlated with population density but positively. This result shows that the fragmentation of politically relevant ethnic groups increases when the population is denser. This would support the argument that population density would be a relevant proxy for communication between groups, allowing the formation of more political factions. Population density would thus be a good instrument for the PREG index as well but in the opposite way to the other indexes of ethno-linguistic fragmentation.

Box 2: Population growth and ethnic fragmentation

To strengthen the argument of ethnic fragmentation's endogeneity, which may also be non constant through time, one can have a closer look at the effect on ethnic fragmentation of different population growth rates between two ethnic groups among a country's population. Let us consider a fictive country with two ethnic groups. The ethnic group A has a population of 90,000 people, while the ethnic group B has a population of 60,000 people. The ELF value for this country is then 0.48. Both ethnic groups have an initial population growth rate of 4%. We then make the assumption that people from ethnic group A change their fertility behavior and benefit from a decrease in the death rate leading to a decrease in the population growth which reaches 1% after 18 years and remains unchanged after that.⁷ While people from ethnic group B do not change their behavior. The consequences of such evolution is an increase in the ELF index from 0.48 to 0.50 in less than 20 years and ethnic group B population to be larger than population of ethnic group A in 25 years, inverting the initial population distribution between groups.

⁷ This scenario is realistic if we consider that the ethnic group A is finishing its demographic transition. This can be the case if this group benefits from better socio-economic conditions or/and a higher rate of urbanization for instance.

This short example leads to challenge facts accepted in the literature: first, we see how the fragmentation may change in 20 years, casting doubt on the constant character of fragmentation and consequently of its measurement; second, it shows that besides the evolution of the ethnic fragmentation index, the ethnic groups' weights in terms of population can also change within a country's population. This may influence the forces involved in a political or ethnic conflict for instance. This example shows that a country's ethnic structure may evolve and thus covers more complicated phenomena than one constant value measuring ethnic fragmentation.

4. Conclusion

The negative role played by ethnic fragmentation in the level of economic performance has been highlighted in the early literature. Easterly and Levine (1997) emphasized a negative relationship between ethno-linguistic fragmentation and growth and Alesina *et al.* (2003), using alternative measures of ethnic fragmentation confirmed also this negative link. However, the use of other measures or concepts to delineate ethnic groups or characterize ethnic fragmentation has put this negative relationship into question. Socio-economists have underlined how difficult the definition of ethnic groups can be (Fearon (2003), Posner (2004)); have challenged the data sources used to measure ethnic fragmentation as well as the mathematical tools leading to the quantification of such population fragmentation (Posner (2004)). Another huge bench of research on ethnic fragmentation has focused on the fact that ethnic fragmentation cannot be considered as exogenous. Major factors put forward to highlight the endogeneity of ethnic fragmentation are linked to historical elements or social behaviors for instance. Some empirical studies stress that the causal relationship between ethnic fragmentation and economic performance may go from economic performance to ethnic fragmentation and not in the other direction. Other analyses have considered time series micro data to take further into consideration the endogenous and evolving countries' ethnic structure.

The current study summed up the different concepts existing until now to describe ethno-linguistic fragmentation and shed light on the potential endogeneity of ethnic fragmentation. To investigate further than what the literature has already highlighted about the endogenous character of fragmentation, the link between countries' surface, population density and ethnic fragmentation has been analyzed using correlations. Countries' surface and population density being exogenous, they would be good candidates to be used as instruments in estimations of the link between ethnic fragmentation and economic performances. Through the analyses of correlations, results emphasize how interrelated are ethnic structures, countries' surface and population density, leading to cast doubt on exogeneity of ethnic fragmentation. The positive correlation between most fragmentation indexes and the log of countries' surface first emphasize that the bigger the country, the more fragmented is its ethnic structure. On the contrary, the significant negative link between fragmentation indexes and the log of countries' population density stresses that denser countries' ethnic groups are more homogenous. In both

cases, the significant correlations tend to support the use of countries' surface and population density to instrument ethno-linguistic fragmentation indexes (including the index of fragmentation of politically relevant ethnic groups from Posner (2004)) when estimating their link with economic performance. This would allow establishing a causal relationship, contrarily to what has been done until now in the literature. This central conclusion is however invalid for the measure of linguistic fragmentation of Alesina *et al.* (2003) for which no significant correlation has been found.

The study's findings highlight consequently the need for a deeper analysis of the mechanisms underpinning the formation and evolution of countries' ethnic structures and of a higher availability of updated data on ethnic groups. Conclusions join the suggestions previously stressed in the literature considering ethnic fragmentation as endogenous. First, in showing the interrelationship between ethnic fragmentation and geographical and structural elements, conclusions urge the researchers to consider ethnic fragmentation as endogenous. Second, addressing the issue of endogeneity would require the use of adequate instruments based on exogenous geographical and structural characteristics. Finally and most importantly, it is worth pointing out that more caution must be taken when asserting that such a complex concept as ethnic fragmentation can negatively influence economic performance in a causal manner.

References

- Alesina, A., Baqir, R. and Easterly, W. (1999), "Public Goods and Ethnic Divisions", *The Quarterly Journal of Economics*, 114(4), 1243-1284.
- Alesina, A., A., Devleeschauwer, W., Easterly, S., Kurlat and R., Wacziarg (2003), "Fractionalization", *Journal of Economic Growth*, 8, 155-194.
- Alesina, A. and E., La Ferrara (2005), "Ethnic Diversity and Economic Performance", *Journal of Economic Literature*, 43(3), 762-800.
- Arcand, J.-L., P. Guillaumont and S. Guillaumont –Jeanneney (2000), "Ethnicity, Communication and Growth", *CSAE Working Paper 2000-20*, Oxford: University of Oxford.
- Arcand, J.-L., P. Guillaumont and S. Guillaumont –Jeanneney (2000), "How to make a tragedy: on the Alleged Effect of Ethnicity on Growth", *Journal of International Development*, 12(7), 925-238.
- Bossert, W., C. D'Ambrosio and E., La Ferrara (2008), "A Generalized Index of Fractionalization", *Working Paper 2008-01*, Département de Sciences Economiques, University of Montreal, Canada.
- Campos, N. F. and V. S., Nuzeyev (2007), "On the Dynamics of Ethnic Fractionalization", *American Journal of Political Science*, 51(3), 620-639.
- Collier, P. (2001), "Ethnic Diversity: An Economic Analysis", *Economic Policy*, 32, 127-166.
- Collier, P. and A. Hoeffler (2002), "Greed and Greavance in Civil War", *CSAE Working Paper 2002-01*, Oxford: Oxford University.
- Easterly, W. and R., Levin (1997), "Africa's Growth Strategy: Policies and Ethnic Divisions", *Quarterly Journal of Economics*, 111, 1203-1250.
- Fearon, J. D. (2003), "Ethnic and Cultural Diversity by Country", *Journal of Economic Growth*, 8, 195-222.
- Fedderke, J., J., Luiz and R. de Kadt (2008), "Using Fractionalization Indexes: Deriving Methodological Principles for Growth Studies from Time Series Evidence", *Social Indicator Resources*, 85, 257-278.
- Grimes, J. and B. Grimes (1996), *Ethnologue: Languages of the World*, 13 Ed. Dallas, TX: Summer Insitute of Linguistic.
- Leeson, P. T. (2005), "Endogenizing Fractionalization", *Journal of Institutional Economics*, 1(1), 75-98.
- Levinson, D. (1998), "Ethnic Groups Worldwide. A Ready Reference Handbook", Phoenix, Oryx Press.
- Lovell, P.A. (1999), "Development and Persistence of Racial Inequality in Brazil: 1950-1991", *The Journal of Developing Areas*, 33(3), 395-418.

- Mauro, P. (1995), "Corruption and Growth", *The Quarterly Journal of Economics*, 110(3), 691-712.
- Montalvo, J. G. and M. Reynal-Queyrol (2005), "Ethnic Polarization, Potential Conflict and Civil Wars", *American Economic Review*, 95, 796-816.
- Morrison, D., R. Mitchell and J. Paden (1989), *Black Africa: A Comparative Handbook*, 2nd. Ed. New York: Paragon House.
- Nettle, D. (2000), "Linguistic Fragmentation and the Wealth of Nations: The Fischman-Pool Hypothesis Reexamined", *Economic Development and Cultural Change*, 48(2), 335-348.
- Posner, R.A. (2004), "Measuring Ethnic Fractionalization in Africa", *American Journal of Political Science*, 48(4), 849-863.
- Scarrit, J. and S. Mozaffar (1999), "The Specification of Ethnic Cleavages and Ethnopolitical Groups for the Analysis of Democratic Competition in Africa", *Nationalism and Ethnic Politics*, 5, 82-117.
- Taylor, C. L. and M. C. Hudson (1972), *World Handbook of Political and Social Indicators*, Ed. Ann Arbor.

Appendix:

Appendix 1: Herfindahl formulae of the ELF index

$$ELF = 1 - \sum_{i=1}^I \left(\frac{n_i}{N} \right)^2$$

Where n_i is the number of people in the i th group, N is the total population and I is the total number of ethnic groups in the country.

Appendix 2: Extended Herfindahl formulae of Fearon's cultural fragmentation index.

$$Fearon_cult = 1 - \sum_{i=1}^I \sum_{j=1}^I \left(\frac{n_i}{N} \right) \left(\frac{n_j}{N} \right) r_{ij}$$

With r_{ij} being the resemblance factor of group i compared to group j .

Appendix 3: Characteristics of major ethno-linguistic fragmentation indexes.

	ELF	Ethnic	Language	Fearon	Fearon_cult	PREG
Used by	Mauro (1995), Easterly and Levine (1997), Taylor and Hudson (1972)	Alesina <i>et al.</i> (2003)		Fearon (2003)	Fearon (2003)	Posner (2004)
Formulae	Herfindahl	Herfindahl		Herfindahl	Extended Herfindahl	Herfindahl
Major criteria of ethnic group identification	Unclear: language and others	Ethnic characteristics	Groups' languages	Specific definition of ethnic groups (Fearon, 2003, pp201) Groups larger than 1 percent of country population	Fearon + language distance between groups	Those of the Atlas restricted to politically relevant groups
Data sources	Atlas Narodov Mira	<ul style="list-style-type: none"> - Encyclopedia Britannica (2001) - CIA World Factbook (2000) - Levinson (1998) - Minority Rights Group International (1997) 		<ul style="list-style-type: none"> - Encyclopedia Britannica (2001) - CIA World Factbook (2000) - Library Congress Country Study - Minority at Risk - Scarritt and Mozaffar (1999) - Levinson (1998) - Morrison <i>et al.</i> (1989) 	<ul style="list-style-type: none"> - Similar sources as Fearon - Linguistic classification: Grimes and Grimes (1996) 	<ul style="list-style-type: none"> - Atlas Narodov Mira - 5 to 20 sources per country to identify the politically relevant groups.
Base year(s) for calculation	1964	1997, 1998, 2000, 2001		1989, 1998, 2000, 2001	1989, 1996, 1998, 2000, 2001	1964
Number of observations	144 countries, 1 obs. per country	184 countries, 1 obs. per country	190 countries, 1 obs. per country	152 countries, 1 obs. per country	149 countries, 1 obs. per countries	42 African countries, 3 obs. per country (1960, 1970, 1980, 1990)

Appendix 4 : Countries' rank with regards to each ethno-linguistic fragmentation index.

	ELF	Ethnic	Language	Fearon	Fearon_cult
Country (N^{er} of countries)	(149)	(184)	(191)	(146)	(129)
Afghanistan	111	164	141	119	126
Albania	36	52	20	15	18
Algeria	66	72	109	47	49
American Samoa			59		
Andorra		152	153		
Angola	136	167	172	121	50
Antigua and Barbuda		38	40		
Argentina	72	60	27	41	1
Armenia		30	46	20	24
Aruba			101		
Australia	74	19	90	21	31
Austria	43	24	53	17	22
Azerbaijan		50	64	32	41
Bahamas		90	61		
Bahrain		101	108	87	89
Bangladesh		8	36	37	28
Barbados	30	34	37		
Belarus		69	114	56	47
Belgium	96	117	126	90	90
Belize	97	148	142		
Benin	104	168	173	2	
Bermuda					
Bhutan	113	124	139	95	98
Bolivia	114	160	70	118	124
Bosnia and Herzegovina		129	152	109	30
Botswana	92	83	105	53	35
Brazil	31	111	21	85	6
Brunei	132	112	92		
Bulgaria	57	82	81	45	52
Burkina Faso	115	158	161	111	67
Burma				79	82

Burundi	14	65	80	50	10
Cambodia	68	51	65	30	32
Cameroon	146	179	187	142	129
Canada	130	151	132	94	95
Cape Verde	90	88			
Central African Republic	116			129	97
Chad	139	178	182	126	128
Chile	46	45	62	73	36
China	42	36	49	23	33
Colombia	28	123	9	104	6
Comoros	40	1	4		
Congo	110	180	154	139	108
Congo, Dem. Rep.	148	181	184	145	119
Costa Rica	32	54	23	40	17
Cote d'Ivoire	142	175	171	128	106
Croatia		77	32	57	40
Cuba	17	121		35	6
Cyprus	77	20	102	54	69
Czech Republic		69	87	48	15
Denmark	25	17	39	18	26
Djibouti		170	148	96	79
Dominica		48			
Dominican Republic	15	91	18	58	1
East Timor			125		
Ecuador	93	135	47	103	91
Egypt	21	43	12	25	1
El Salvador	51	47		34	38
Equatorial Guinea	86	74	86	1	
Eritrea		134	147	102	76
Estonia		103	120	77	93
Ethiopia	118	154	175	122	107
Fiji	121	114	129	89	105
Finland	49	32	50	19	27
France	65	22	44	43	53

French Guiana			43		
French Polynesia			140		
Gabon	117	163	170	137	72
Gambia, The	124	166	176	123	104
Georgia		99	117	70	79
Germany (post 1989)		39	55		
Ghana	120	140	151	135	74
Greece	38	37	14	11	13
Greenland			67		
Grenada		62			
Guam			163		
Guatemala	106	106	112	72	94
Guinea	128	159	168	110	92
Guinea-Bissau	137	171	178	133	110
Guyana	100	127	30	97	89
Haiti	4	21		14	1
Honduras	50	46	25	29	36
Hong Kong	8	15	66		
Hungary	37	35	13	31	40
Iceland	26	16	34		
India	145	89	174	131	125
Indonesia	133	156	167	125	99
Iran	131	139	164	106	103
Iraq	79	76	96	86	68
Ireland	22	28	15	27	34
Israel	55	73	130	81	51
Italy	18	26	42	9	10
Jamaica	23	85	41	26	7
Japan	5	3	8	7	5
Jordan	24	122	19	75	12
Kazakhstan		126	149	105	117
Kenya	141	177	186	136	115
Kiribati		10	12		
Korea, Dem. Rep.(nord=	2	4	2	4	2

Korea, Rep.(sud)	2	2	1	5	3
Kuwait		137	93	112	102
Kyrgyzstan		142	135	108	118
Lao People's Dem Rep	102	107	144	67	6
Latvia		119	134	93	86
Lebanon	44	31	48	127	44
Lesotho	58	61	78	42	14
Liberia	140	183	190	143	120
Libya	60	169	31	22	25
Liechtenstein		118	71		
Lithuania		70	85	52	55
Luxembourg	48	108	145		
Macau			77		
Macedonia (Former Yug. Rep)		102	121	83	84
Madagascar	29	182	11	138	43
Malawi	105	141	138	134	65
Malaysia	108	120	136	94	109
Maldives	1				
Mali	135	144	180	120	112
Malta	34	7	35		
Marshall Islands		14	26		
martinique	9				
Mauritania	76	125	88	98	59
Mauritius	99	94	111	99	88
Mayotte			160		
Mexico	71	113	52	84	85
Micronesia		147	165		
Moldova		116	131	76	77
Monaco		143	162		
Mongolia	81	75	97	43	46
Morocco	94	96	115	66	70
Mozambique	109	145	177	124	61
Myanmar (Burma)	88	104	122	80	82
Namibia		130	156	114	111

Nepal	119	138	157	107	103
Netherlands	39	23	123	12	16
Netherlands Antilles			75		
New Caledonia			150		
New Zealand	80	81	56		
Nicaragua	52	97	22	61	20
Niger	125	133	146	100	114
Nigeria	143	176	181	130	123
Northern Mariana Islands			169		
Norway	19	12	28	16	21
Oman	56	93	95	65	79
Pakistan	107	149	159	82	62
Palau		92	83		
Panama	64	115	100		
Papua New Guinea	84	63	94	3	
Paraguay	47	40	137	19	9
Peru	101	136	91	101	96
Philippines	127	56	179	24	23
Poland	12	27	21	10	11
Portugal	3	9	10	9	10
Puerto Rico	10		17		
Qatar	73	161	118		
rda	6			6	4
rfa	11			14	19
Romania	62	67	58	46	57
Russian Federation		57	74	113	113
Rwanda	45	71		28	1
Samoa	7	33	5		
San Marino		64			
Sao Tome and Principe	53		72		
Saudi Arabia	27	42	38	88	80
Senegal	123	146	155	116	78
Seychelles	70	49	54		
Sierra Leone	134	174	166	123	101

Singapore	83	78	99	59	74
Slovak Republic		59	79	51	64
Slovenia		53	68	38	37
Solomon Islands	61	25	124		
Somalia	33	173	16	132	63
South Africa	144	162	183	140	100
Spain	85	87	107	74	56
Sri Lanka	87	86	113	63	73
St Kitts & Nevis		44			
St. Lucia		41	84		
St. Vincent and the Grenadines		66	7		
Sudan	126	153	158	112	127
Suriname	103	155	89		
Swaziland	82	11	57	44	29
Sweden	35	13	63	33	42
Switzerland	91	109	127	91	81
Syria	59	110	60	92	48
Tajikistan		105	128	78	93
Tanzania	149	157	189	146	109
tchecoslovaquie	89			48	15
Thailand	112	131	143	64	83
Togo	122	150	188	141	116
Tonga	13	18	98		
Trinidad and Tobago	98	132	45	102	71
Tunisia	69	5	6	8	8
Turkey	63	68	69	45	66
Turkmenistan		80	103	60	60
Uganda	147	184	191	144	121
Ukraine		95	116	62	54
United Arab Emirates		128	119	117	122
United Kingdom	75	29	24	49	39
United States	78	98	76	71	58
Uruguay	54	58	33	36	1
Uzbekistan		84	106	69	87

Vanuatu		6	133		
Venezuela	41	100	29	68	6
Vietnam	67	55	73	39	45
Virgin Islands (U.S.)			82		
West Bank			51		
Yemen	16		3	13	17
Yugoslavia (pre 1991)	129	172	104	91	75
Zambia	138	165	185	115	42
Zimbabwe	95	79	110	55	28

Source: Author's calculation following Atlas Narodov Mira (1964), Alesina *et al.* (2003) and Fearon (2003).

Note: In grey are stressed examples of rank values which strongly deviate from the average rank observed for the country.