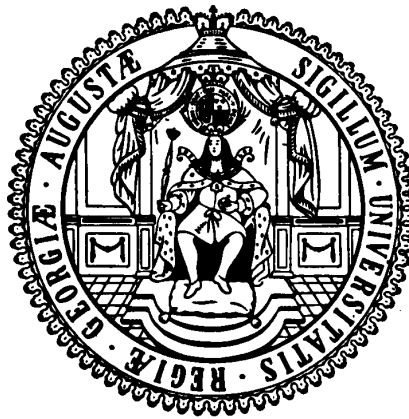# Ibero-Amerika Institut für Wirtschaftsforschung
Instituto Ibero-Americano de Investigaciones Económicas
Ibero-America Institute for Economic Research
(IAI)

Georg-August-Universität Göttingen
(founded in 1737)

Diskussionsbeiträge · Documentos de Trabajo · Discussion Papers

---

Nr. 210

Questioning Ethnic Fragmentation's Exogeneity -
Drivers of Changing Ethnic Boundaries

Philipp Kolo

May 2011

Questioning Ethnic Fragmentation's Exogeneity – Drivers of Changing Ethnic Boundaries

Philipp Kolo[*]

Georg-August University Goettingen, Germany

May 2011

Abstract

Ethnic fragmentation is a variable increasingly used in the economic literature to explain differences in economic development level, growth or the incidence of conflicts. Nearly all articles have in common that they treat ethnic fragmentation as a static, exogenous fact. Only recently some contributions outlined first ideas, why different levels of ethnic fragmentation evolved based on biodiversity and evolutionary theories.

This article has two main goals. In connecting with these recent findings, the article boldly confirms their results that a 'base-level' of fragmentation evolved due to geographical and evolutionary factors. Additionally, it draws the attention to the impact of colonization on fragmentation, especially on how a country was colonized. The main goal, however, is to show that ethnic fragmentation is not only evolving over centuries, but changes over a short period of time. As static factors, e.g. geographical ones, can't be responsible for changes in the short run, the article offers a structured assessment of factors that may influence diversity levels in the short term. Although migration is the most obvious factor, urbanization and especially education play an even more important role in influencing a country's ethnic boundaries.

Key words: Colonization, Endogeneity, Ethnic fractionalization (ELF), Heterogeneity.
JEL classification: C23, F54, I29, O10, Z10

1 Introduction

"Every valley is still a little world that differs from neighboring world as Mercury does from Uranus" (Weber, 1976, p. 47). In this quote Weber is not referring to a developing country in the heart of the African continent where ethnic heterogeneity is claimed to be at the roots of its growth tragedies[1]. Instead, it is a citation of an economist describing France in the second half of the 19th century. Only 36 out of 89 *départments* were fully French-speaking, and Weber (1976) concludes that "French was a foreign language for a substantial number of Frenchmen, including almost half the children who would reach adulthood in the last quarter of the century" (Weber, 1976, p. 67). Despite the language heterogeneity, Weber describes in great detail how diversity was persistent in every part of life, from cultural traits over measurement systems, currencies and various beliefs in contrast to the officially proclaimed Christianity. Some decades later, in the middle of the 20th century, demographic estimates already showed the more common picture of France as the homogenous *grande nation*[2].

This paves the way to investigate the dynamics of a country's ethnic diversity[3] and to question the static nature on which most economic literature bases their analyses of the role of ethnic diversity[4]. Although most authors admit that there is some endogeneity involved, they do not pursue this fact further and proclaim that fragmentation is at least not changing over a short period of time[5]. But, in a time where conflicts, migration and globalized trade are shaping countries and their populations, shouldn't one be able to observe rather huge shifts in a country's ethnic set-up?

In contrast to this literature, some recent publications try to shed some more light on what roots diversity might have and why it developed so differently over the globe. Ahlerup and Olsson (2007) analyze the influence of human settlement. The duration of uninterrupted settlement leaves more time to diverge

---

[1]See the influential paper of Easterly and Levine (1997) about 'Africa's growth tragedy'.

[2]Héran *et al.* (2002) assess that less than 10% of parents did not speak French with their children in 1950.

[3]Ethnic fragmentation and diversity is used in this article interchangeable, which is quite common in most of the literature. However, diversity is a much more elaborated aspect of ethnicity because it requires to take (dis)similarities between groups into account. For a methodological-technical discussion of the prerequisites to measure diversity, see Bossert et al. (2003) and Nehring and Puppe (2002). For the assessment of a new index to capture ethnic diversity, see Kolo (2011a).

[4]For a more detailed overview of pathways through which diversity is affecting the economic outcome of a country via its influence on institutional and policy drivers of growth, see for example Alesina and La Ferrara (2005). For a good overview of diversity's influence on conflict incidence, type and duration, see additionally Garcia-Montalvo and Reynal-Querol (2003).

[5]A rare exception is Fedderke *et al.* (2008) with a case study on South Africa. They employ changing values of racial fragmentation for each decade in their analysis on its role on economic growth.

into different groups, leading to an increased diversity. The existence of modern sates and its institutions lowered a country's fractionalization[6]. Additionally, policies might directly or indirectly promote 'assimilation'.

Michalopoulos (2008) bases his article on Darwin's evolutionary theory. He argues that various geographical conditions are "the ultimate cause of the emergence and persistence of ethnic diversity" (Michalopoulos, 2008, p. 2). These different settings in turn lead to the emergence of different species adapted to their specific niche, which is also true for the modern human.

Whereas both Michalopoulos (2008) and Ahlerup and Olsson (2007) explore rather long-term historical and geographical determinants of diversity, Campos and Kuzeyev (2007) analyse changes in heterogeneity in the former Soviet republics after the fall of the Iron Curtain. Their approach thus comes closest to the intention of this paper. They show that over the decade that followed 1989 ethnic fractionalization decreased in most countries, language diversity did not change significantly and religious diversity demonstrated a slight increase[7]. Unfortunately, Campos and Kuzeyev (2007) remain with these findings and do not try to analyze empirically the reasons for the different changes.

Using new data, this article supports the above findings that a 'base-level' of diversity evolved due to a set of geographical and historical variables. It additionally offers a new interpretation of colonization's impact on shaping a country's diversity. The approach the colonial powers followed in their pursuit plays the most important role. The main finding of this article is that diversity did already change over the rather short period of twenty years. Static factors can't be responsible for these dynamics. Migration is the most obvious factor in a more integrated and globalized world, which is confirmed by this study. However, it shows that urbanization and especially education play a significant and even more important role.

The rest of the paper is organized as follows. In section 2 ethnicity is shortly framed and the major views on its dynamics are introduced. Section 3 structures and discusses the various drivers that might be responsible for changes in a country's ethnic set-up. Section 4 outlines the empirical strategy and discusses the data sources used, their coverage, limitations and first insights into descriptive statistics. Section 5 then controls the empirical significance of the drivers for a wide range of countries. Finally, section 6 summarizes the key findings, concludes and gives an outlook for further research.

---

[6]See also Ranis (2009), who argues that kinship relationships are a mere compensation for not existing official social security networks.

[7]For a discussion of 'association webs' between various forms of fractionalization and other social, political and institutional dimensions in a case study for South Africa, see Fedderke and Luiz (2007).

2 Framing ethnicity

In line with most of the economic literature, this paper relies on the ethno-linguistic fractionalization index (ELF) as a measure for a country's ethnic set-up. Mostly, a combination of ethnic, language and religious characteristics are used to assess the ethno-linguistic groups of a country[8]. It was first calculated in this context by Taylor and Hudson (1972) and was then continuously used in the economics literature. The ELF is calculated based on a Herfindahl concentration index:

$$ELF = 1 - \sum_{i=1}^{N}(p_i^2), \qquad i = 1, ...N \tag{1}$$

where p_i are the relative group sizes. The measure ranges between zero (only one group and thus complete homogeneity) and one (complete heterogeneity). It reflects the probability that two randomly selected individuals from a population come from different groups and generally increases with the number of groups[9]. However, to define ethnic heterogeneity and its measurement does not yet explain why ethnic groups formed and why these groups should be subject to change. Two main schools of thought try to explain this fact: an evolutionary and a constructivist one.

van den Berghe (1981) sees ethnic groups as nothing but an extension of the concept of kinship. This purely nepotistic behavior is observed in all mammal species and is the result of an evolutionary survival strategy. Living in an environment with only limited resources, sticking with your kin, led to "greater reproductive success and tend[s] to dominate all populations" (Ahlerup and Olsson, 2007, p. 6). As these kinship groups grew, they developed common (cultural) traits or markers to sustain the structure also for a more extended group. Horowitz (1985) sees no "bright line to be drawn between kinship and ethnicity, especially in societies where the range of recognized family relationships is wide and the importance of kinship ties is great" (Horowitz, 1985, p. 60). All these dynamics thus proved as efficient evolutionary concepts serving

---

[8]See for example Alesina *et al.* (2003), and Fearon (2003), who build their measures on this combined taxonomy. For more details on language groups and the mutual differences, see Lewis (2009) and Fearon (2003). For some specific analysis on the role of religion, see for example Guiso *et al.* (2004) or Barro and McCleary (2003) and Garcia-Montalvo and Reynal-Querol (2003) for the role of religious polarization.

[9]For details on other measures, see Garcia-Montalvo and Reynal-Querol (2003, 2005, 2008) for an index of polarization, Posner (2004) on his restricted index of politically relevant ethnic groups, and Fearon (2003) for the idea of ethnic distance that is further explored by Kolo (2011a).

to enforce rules, avoiding free riding and sustain loyalty within the group[10].

On the other hand, more recent factors and the emergence of nations also left their traces on the development or construction of ethnic groups. According to this constructivist view, major changes from early human development to modern national states did have a huge effect. According to Olsson (2007), this process started as people became sedentary farmers. This in turn led to an unprecedented population growth and for the first time to the emergence of specialists' class within the population. Subsequently, the formation of nations and modern states shaped and changed the group construction and identification drastically.

This article does not want to reconcile these two groups. For the main argument of this article that ethnic boundaries are subject to change, both offer a comprehensible argumentation. In analysing drivers of changes in a country's ethnic set-up, both approaches subsequently deliver potential explanations and influential factors.

Some theoretical frameworks and mathematical models offer additional motivation for the dynamics of changing ethnic boundaries. Constant and Zimmermann (2007) discuss in a simple framework the main strategies of immigrants with respect to their ethnic heritage. According to them, immigrants follow either an assimilation, integration, marginalization or a separation strategy. Depending on the strategy chosen different effects on the ethnic composition in the destination country would emerge. Darity *et al.* (2006) use an evolutionary game theory model to show different 'acculturation' outcomes linked with the potential wealth accumulation in the overall society. Ahlerup and Olsson (2007) build their model on kinship-based social organization providing public goods. Caselli and Coleman (2008) discuss a model of coalition formation based on the excludability of others. They root the emergence of conflicts in on the possibility to exclude the defeated group from the seized assets. Lazear (1999) model assimilation processes of language groups to sustain or ameliorate trade. Kolo (2011b) extends this approach and covers the main dynamics this article tries to prove empirically. It balances the gains of increased trade possibilities due to learning a new language with the costs of doing so. The costs are strongly influenced by the proximity of two languages and the infrastructure both for learning as well as trading. Trading gains in turn are defined by the size of trade partners, i.e. the size of the respective language groups. The extended model shows that with a rising development, a continuous process of assimilation into the majority group is expected. Increasing education lowers the

---

[10]Finally, Ranis (2009) points out that the kinship relationships are a good substitute for social security networks and that they can be efficient in providing public goods.

costs of learning and more individuals would decide in favor of an assimilation. Higher transportation costs (or less integration or infrastructure) decrease the value of the trade option and would thus make an assimilation less probable. Migration is not specifically covered in the model. However, with an increasing exchange and trade, migration also gets more important. A higher development level would increasingly attract immigrants and would render it - at least in the short-term - more heterogeneous. Thus, the model of Kolo (2011b) gives some first points of reference for the further discussion of potential drivers for a changing ethnic set-up.

3 Drivers of diversity change

Ethnic boundaries that are based on tradition, ancestry and conveyed habits, are certainly nothing that is subject to instant fluctuation. However, the environment, in which generations are raised, be it economically, socially or educationally, should leave their marks and thus lead to a changing ethnic identification; especially in an increasing globalizing world. A key difference between the prospective drivers for change might be their time dimension. The geographic outline of a country is fixed. The access to remote areas can be alleviated, but this is rather a policy decision regarding infrastructure than a per se change in geographical conditions. Others, can change rather quickly and are susceptible to political influence. Depending on the ease of change, the variables can be categorized within two groups: evolutionary and historical factors, as well as socioeconomic and policy factors.

3.1 Evolutionary and historical factors

Location and geographical conditions One of the most basic location characteristics of a country is its latitude. Michalopoulos (2008) points to the fact that biodiversity is decreasing with the distance from the equator. Cashdan (2001) roots the high biodiversity around the equatorial region to its tropical climate, the associated habitat diversity, and its higher pathogen load. The lack of climate variability in tropical areas leads to specialization regarding a very specific environment or niche. Areas with high climatic variability (e.g. hot summer — cold winter) lead to a more generalized approach and lower variation. Additionally, a country that is covered in large parts by mountains offers more niches and at the same time makes an exchange between valleys much more difficult. For both reasons, one would expect more mountainous countries to be more diverse. Large countries that cover a huge area should

6

encompass more bio-geographic niches and should thus demonstrate a greater diversity.

Human development The historical duration of uninterrupted human settlement since centuries basically left more time for humans to diverge into different groups. Ahlerup and Olsson (2007) rebuild the way in which the modern human migrated from its birthplace in East Africa to all other parts of the world. In doing so, the development follows a constant process of genetic fractionalization. Michalopoulos (2008) underlines the importance of geographical conditions in catalyzing the emergence of different human groups. Whereas more time since the emergence of the modern human already leads to a diversification just based on genetic mutations, geographical conditions help to shape and maintain diversity in the various locations. Ahlerup and Olsson (2007) direct attention to Papua New Guinea for an example as to how both drivers jointly affect ethnic diversity. Its special geography spans a wide array of bio-geographic niches, and its long population time of some 65,000 years led to many isolated and distinct ethnic and language groups. Some 860 indigenous languages spoken within a total population of only around 4 million inhabitants are still reported today[11].

Modern state history and colonization Modern states with their institutions can play a decisive role in homogenizing countries. Well functioning institutions that include codified laws, security and military protection rendered ethnic and cultural forms of interaction less important[12] and should have led to an assimilation process into the major group. Olsson and Hibbs (2005) point to the transformation from a hunter-gatherer economy to sedentary agricultural production as one of the most important events in shaping societies. This transition led to a very basic set of institutions, which led to an increase in productivity. This in turn was the basis for the development of a non-producing class. Freeing this class from production obligations left room for the development and organization of knowledge leading to the expansion of science, technology, and state formation. The time since this agricultural transition is thus supposed to be a factor influencing civilizations and their respective diversity[13].

---

[11]The 860 languages represent over one tenth of the world's total (Lewis, 2009).

[12]See for example Greif (1993) on an example of ancient trade relationships in the Maghreb region.

[13]Ahlerup and Olsson (2007) explore how experiences with a modern state over the last hundred years significantly reduced diversity. Yet they admit that causality in this aspect is not clear, and more homogeneous countries might have developed a modern state more easily and thus earlier in history.

In many developing countries the arrival of colonizers meddled lastingly with existing structures and it is a significant factor that created and shaped countries and societies. Colonizers tried to introduce their legal and political systems. Often they additionally forced their own language on the occupied countries. From a language point of view, Latin America displays a strong homogeneity as Spanish was widely adopted. The same is true of many French-speaking countries in Africa. The identity of the colonizer and the time span of colonization might be crucial factors for changes in ethnic boundaries. This leads to an argument by Acemoglu *et al.* (2001). Depending on the interest of the colonial power, they either pursued the 'divide-and-rule' approach and just exploited the country without any long term interest (mainly in Africa) or did indeed establish institutions to sustain a long term development and settlements (e.g. Canada, Australia or Singapore). Acemoglu *et al.* (2001) attribute these two approaches to the differences in living conditions the colonizer came upon at that time. They measure these conditions as the mortality rate among the Europeans arriving in their respective colonies. In countries with higher mortality rates, the colonizers did not want to create lasting structures and institutions targeted on long term settlements. A more extractionary approach specifically exploited differences between groups, deepened them and spurred the groups on each other. This was pursued by the Belgians in Rwanda with the Hutu-Tutsi split still in the twentieth century[14]. In countries with higher mortality rates that were subsequently exploited and experienced lower levels of institutional developments one might find a higher degree of ethnic diversity.

3.2 Socioeconomic and policy factors

Demographic factors The global international migrant stock rose between 1960 and 2005 from 77 million to 195 million people (World Bank, 2009). Coleman (2009) sees immigration as the primary reason responsible for an increasing diversity with respect to ancestry, ethnic origins, and religions that will have long-term changing effects on the population make-up.

Schüler and Weisbrod (2010) analyze whether the effect of ethnic heterogeneity on economic performance changes when migration as a cause of higher diversity is taken into account. They conclude that migrants increase trade as they import information about their home country, thereby reducing transaction costs and simultaneously increasing trade due to their preferences for home country products. However, they do not analyze what impact immigration has

---

[14]For a broader discussion of the 'divide-and-rule' strategy as a principle of mere exploitation, see Ahlerup and Olsson (2007). For the Rwandan case, see also Caselli and Coleman (2008), who discuss their theoretical model with this conflict.

on the diversity level of a country that may be significant[15].

Fertility rates and population growth are affected by a wide range of factors. Ultimately, not only a woman's personal experience but also her heritage plays a decisive role[16]. Different preferences of fertility rates between a country's historic population and immigrant groups might be important. Most host countries (mainly developed countries) experienced their fertility transition, significantly lowering the birth per women, earlier than most less developed countries, where many immigrants originate. This might have a significant impact on the destination countries[17]. For Coleman (2009), these differences are the second most important driver of a shift in diversity.

A rising population density will mainly affect very small countries. The growth of metropolitan regions might be more susceptible for changes in a broader set of countries. The population density in urban areas might even increase when the country density remains constant due to high rural-urban migration flows. Cashdan (2001) showed in her work on biodiversity that an increased density of species leads to a higher degree of specialization on a smaller area and thus finally to a higher diversity. Urban areas are an agglomeration of people all struggling over limited resources. Thus an coordination along ethnic ties to better sustain economic or social development could be expected. However, as the newly arriving population needs to interact with the existing masses, an integration into this mainstream is also expectable. In Bates (2006) one finds the argument that urbanization erodes cultural foundations and replaces ethnic ties with rather interest-based liasons. This could have an effect on the ethnic differences between groups. Ethnic borders become less pronounced leading to a more homogeneous civilization. Finally the demographic set-up in terms of the age structure ought to have implications for the transmission of ethnic traits, language or religion between generations. In countries with an exploding population, the relation between youth and parents or older members of the communities is decreasing very fast. The continuously transmission from cultural traits might be less intense and the possibility of an assimilation to a new or adapted 'youth culture' or a 'main stream culture' might be higher[18].

---

[15]Especially their diversity measure does not change even for high immigration countries.

[16]Fernandez and Fogli (2009) find that the heritage-induced fertility is a significant and persistent factor within second generation immigrant mothers in the United States.

[17]Hispanic and Asian 'minority' groups in the United States are projected to account for around 36% of the total population by 2050 (Coleman, 2009). For a detailed analysis of the emergence of mixed ethnic groups in the United States, see Perez and Hirschman (2009).

[18]For a discussion on the differences between vertical and horizontal transmission of culture, see Bisin and Verdier (2001), Saez-Marti and Sjögren (2007) or Spolaore and Wacziarg (2009).

Conflicts A wide set of literature tries to link an increased conflict incidence with higher ethnic diversity[19]. The reverse causal chain is not addressed so far in empirical papers but some theoretical models capture this dynamic[20]. What remains unquestioned is that the various forms of conflict has a significant impact on a country's population. This might be a direct one as for instance death from prosecution or combat, or through refugee-induced migration. Not only on the country where the conflict is rooted, but also on the neighboring countries. The violent construction of ethnic identities, ethnic cleansing and genocides are the most brutal form in this regard. In line with the constructivist view, additionally, the question arises, whether ethnic identities arise or are shaped in the onset of ethnic conflicts. Elites might agitate their peers and strategically use potentially salient ethnic divisions for their ambitions. Fearon and Laitin (2000) analyze a wide range of case studies concluding that elites, but in some cases also ordinary folk, systematically construct ethnic identities in order to strengthen or seek their hold on power[21].

Economic factors There is a growing literature on factors benefiting the economic growth of a country, including various measures of institutions, financial indicators, trade, education or infrastructure[22]. Thus, it would be obvious to include GDP figures in the regressions. However, it is hard to see why the economic development level per se should have altering effects on the ethnic diversity of a country if not through various variables highly linked to it. To better elaborate which of these variables affect heterogeneity, a set of variables highly linked to the GDP per capita measures is included.

As Olsson and Hibbs (2005) discussed, there are structural changes within an economy over its development path. A different economic structure could be more susceptible to different values of diversity. Gellner (1983) reasons that the industrial revolution and the accompanying higher division of production steps led to a need for higher homogenization. To face the new division of labor and efficiently work together, there was a need for a certain level of assimilation or homogeneity.

---

[19]The first to analyze the effect of ethnic diversity on conflicts were Collier and Hoeffler (1998). Subsequently, Fearon and Laitin (1999) analyzed the question with a focus on minority groups, Collier (1998) with a focus on democratic institutions and Fearon (2003) with a more general approach.

[20]See for example Caselli and Coleman (2008) or Darity *et al.* (2006) and more generally Ahlerup and Olsson (2007).

[21]Fearon and Laitin (2000) also give a good general overview of the theory on social construction of ethnic identities.

[22]An exemplary selection of papers analyzing economic growth factors that also deal with ethnic diversity are Alesina and La Ferrara (2005); Bellini *et al.* (2009); Collier (2000); Easterly and Levine (1997); Mauro (1995) and Sachs (2001).

Assimilation might not only take place within one economy, but can also have the effect of a mutual rapprochement between two different countries. For Janeba (2004), imported Western products are responsible to crowd out locally manufactured goods and might even marginalize local culture. In general, trade makes a higher variety of (foreign) products available and normally also reduces the price of these goods. The access and a low relative price decrease the overall cost of non-conformity with the own culture and paves the way to a more globalized or generalized culture. If the increased choice of products wash out the ethnic identity of a country, the group of people still sticking to the old habits can be worse off than in autarky when everybody was conform to their preferences[23]. In some constellations of his model, this might even outweigh the gain of trade.

For international trade, language might play a special role, because a common language facilitates trade (Lazear, 1999). The knowledge of a main international language should give an advantage to a country's group that speaks this language natively. But all others also have an incentive to learn the (economically) dominant language and to understand or familiarize oneself with this language[24].

Institutions and policy factors   Institutions in general and their underlying ideology might play an important role. The development of state structures, codified law, governing institutions and a common military protection changed the mode of living together. Ethnic identity might always be in some field of tension with a nation state promoting cultural similarity and integration. The relationship between ethnic diversity, the emergence of institutions and vice versa is not completely clear. Institutions can grant equality, human rights and freedom to pursue cultural expressions. They can also be used as an excessive form of nationalism, excluding culturally deviant citizens with various forms of pressure or even brutality[25]. This kind of uniforming policy can be present in all forms of state activities, always with the intention of considerably altering the ethnic composition of the national state. In forming a French identity, as outlined in the introduction, the mode was rather peaceful. In the last century, some examples showed an unimaginable brutality.

---

[23]Dreher (2006) for example proxy social globalization inter alia with the number of McDonald's restaurants.

[24]For Leeson (2005), learning a language is a form of signaling that the distant individual has a strong desire to trade as he bears the high costs of assimilation (learning a foreign language, adapting to new customs or joining a new religion).

[25]For a discussion of the blurred transition between ethnicity and nationalism, see Eriksen (1991).

Linked with institutions, the question of the role of democracy is inevitable. Both Alesina and La Ferrara (2005) and Collier (1998) show that more democratic regimes moderate the potential detrimental effect of ethnic fractionalization on the economic development. This could indicate a more tolerant environment in democratic countries in which more diverse views are accepted. Campos and Kuzeyev (2007) hold the more tolerant environment of democratization in the former Soviet republics after the fall of the Iron Curtain responsible for an increased religious heterogeneity. However, this might have been a special case, as religious activity was especially disregarded under the communist regime. More autocratic or dictatorial regimes that are mostly built around a very nationalistic ideology might display a significantly lower diversity. Again, the role of democratic regimes and the direction of causality is not clear[26]. However, there is some indication that this kind of political regime at least leaves more room for cultural activity, which might be represented in a more diverse religious or ethnic identification[27].

Education plays a key role for the development of a country (Barro, 1999; Knack and Keefer, 1997) and for its democratization (Akdede, 2010; Barro, 1999). Bolt and Bezemer (2009) well describe the different effects education might have. In a general interpretation education increases one's human capital. Being endowed with higher human capital one's social and economic vulnerability declines. Less vulnerable groups need less to rely on ethnic differentiation or identification to pursue their (economic) activities. It also increases tolerance and leads to more rational decisions. Both effects back the argument, that ethnic identification becomes less important with an increasing education.

Transporting a common history and culture can lead to a better mutual understanding but can also be used as a form of exerting an influence over the young citizens. Education is in the context of this paper also interpreted as a strong expression of state power. Bolt and Bezemer (2009) subsume this under the educations' "purpose of cultural repression" (Bolt and Bezemer, 2009, p. 28). For minorities education often includes language education, as they might have been raised in their native language[28]. It seems that the early education

---

[26]Collier (1998), for example, discusses, how more democratic regimes might emerge only or more easily in countries where ethnic differences are less problematic.

[27]An interesting other aspect is discussed in Bezemer and Jong-A-Pin (2008). They try to find support for the thesis that the combination of democracy and globalization makes conflict in countries with an ethnic minority dominating the market more probable. However, they only find limited support for this thesis in Sub-Saharan Africa.

[28]Turkey, for example, still partly prohibits the native Kurdish language and promotes an education system exclusively in Turkish. Aimed at marginalizing this culture and to in repress its minorities, it still uses discriminatory language in school books (European Commission, 2006).

has the most significant effects, as it is the first time when in many countries a young citizen is confronted with the influence of state institutions. The shift from no schooling to primary schooling is thus probably the most important one. A country with a higher primary enrollment rate or educational coverage thus might be more homogeneous. The impact and role of higher (secondary or tertiary) education is, however, less obvious[29].

Despite geographical hurdles, modern forms of infrastructure and communication make an exchange between remote areas possible. Roads, on which goods and services may travel, are crucial to start business with the so far 'outside world'. Infrastructure can counterbalance geographical disadvantages to enable participation in national or international trade[30]. Accordingly, Cashdan (2001) shows that ethnic diversity is indeed lower where land and water transportation are more efficient. One would expect the same to result from the analysis of this article.

4   Empirical strategy and data

To connect with the existing literature, some of the key results of Ahlerup and Olsson (2007) and Michalopoulos (2008) of a 'base-level' of diversity are reproduced. This analysis takes up the effects of the evolutionary and historical factors discussed in section 3 that will stay constant or will not change also over long periods of time. The corresponding ordinary least square (OLS) regressions are for:

$$ELF_i = \beta_0 + \beta_i \cdot X_i + \epsilon_i \tag{2}$$

where ELF_i are the ELF levels in country i. X_i is a vector of the various independent variables, and ϵ_i is a random error term. The model uses heteroskedasticity robust estimators.

Having analyzed the static variables influencing ELF levels, some new insights as to how diversity is changing over a rather short period is the focus of the second step. An adjusted growth model taking into account level data that does not change over the period and the relevant variables that should be responsible for the change of ELF levels is used here. The linear regression

---

[29]Barro (1999) also finds differences in terms of explanatory power of the various education levels on democratization. Whereas average years of attainment and the gender gap at the primary level have high explanatory power, secondary and higher levels of education do not.

[30]For a detailed survey of infrastructure and their impact on trade flows, see Limao and Venables (2001).

model is specified as follows:

$$\Delta ELF_i = \beta_0 + \beta_i \cdot Z_i + \gamma_i \cdot \Delta X_i + \epsilon_i \tag{3}$$

where ΔELF_i is the change of the ELF value of country i between the two observation points. Vector Z_i contains level data that is static (e.g. country size) and was analyzed in the previous section. These factors are controlled, as the timing or magnitude of changes could be influenced by their presence. In a very mountainous country, diversity might be much more stable than in a small country that does not have any geographical barriers. ΔX_i instead contains the relevant changes of the socioeconomic and policy variables over the period covered. ϵ_i is a random error term, and again, the model uses heteroskedasticity robust estimators.

The key question for the empirical operationalization is which source for the ELF values should be applied. Defining ethnic groups is very much liable to the subjective decision of its authors. Combining two sources over different points in time is highly difficult. A distinction between differences in definitions and real changes in a country's ethnic set-up is all but impossible. The only data source that offers ethnic diversity data on two points in time is the *Atlas Narodov Mira* (*ANM*) compiled by Russian ethnographers (Bruk, 1964; Bruk and Pučkov, 1986). Although only the first edition of the Atlas Narodov Mira (Bruk, 1964) is widely used in the literature, there is a second edition from the mid-1980s (Bruk and Pučkov, 1986)[31]. Some later critique centered around the *ANM*'s bias to a higher linguistic than ethnic split of groups. This underscores the fractionalization in regions like Latin America, where Spanish is widely spoken by minority populations. More important for this article is that the definition of the groups follows the same lines in both points in time and less whether the chosen group characterization is the correct one. Despite the critique on the *ANM* data, comparing them with the two main alternatives, Alesina *et al.* (2003) and Fearon (2003) yields high correlations as displayed in *Table 1*[32]:

Additionally, one might argue that the data has been assembled under the auspices of the Soviet Union with a significant bias between Eastern and West-

---

[31] As both are published in Russian only, this article relies on Roeder (2001), who calculated and published ELF values based on these two editions. Roeder (2001), additionally, calculates ELF values in three different ways, depending on the aggregation levels of sub-groups reported in the original data. Following the approach of Alesina *et al.* (2003), this analysis is based on the most disaggregated values that use all sub-groups reported.

[32] For their ELF indices both combine different sources, mainly the *CIA Factbook* (CIA, 2009) and the Encyclopædia Britannica (2009). Whereas Alesina *et al.* (2003) pursue to always select the most granular source, Fearon (2003) limits the data on groups that at least constitute 1% of a country's population. Due to the very subjective decisions between the data sources selected, both measures also have some severe limitations.

	ANM '61	ANM '85	Alesina	Fearon
ANM '61	1			
ANM '85	0.9517	1		
Alesina	0.8473	0.7850	1	
Fearon	0.8147	0.8430	0.8628	1

Table 1: Pairwise correlation of main ELF indices

ern countries. Taylor and Hudson (1972) tested for this point right from the beginning but did not find any hint for this theory. Finally, Weidmann *et al.* (2010) conclude that the *ANM* data "is complete and carefully researched, it relies on a uniform group list that is valid across state borders." The last point is probably the most important for my analysis.

Based on the sources used to calculate the ELF values, Roeder (2001) reports the data to be for the years 1961 and 1985. As yearly data on most of the covariates used to explain ethnic heterogeneity and its trends is scarcely available, average values for 1960—65 for the first point in time and for 1975—80 for the second are used[33]. An important reason to take the average of several years instead of single ones is to avoid or at least reduce the impact of cyclical deviations. Additionally, the focus of this paper is on a longer time frame and not on annual variations. For the later time span one could alternatively use 1980—85 instead of 1975—80. The period from 1975—80 is preferable for two reasons. First, if diversity adjusts in reaction to policy changes, as is argued in this paper, it needs time to adapt and will not change immediately. Taking a lag of five years, gives some room for these adjustments[34]. Second, time having elapsed between changes in policy variables and the ELF adaptations, limits the suspicions of reverse causality that ELF changes are responsible for policy adjustments.

Roeder (2001) reports data for 138 countries at both points in time based on the respective edition of the Atlas Narodov Mira[35]. *Table 2* displays the distribution of ELF values across regions for both years. The highest median level is found, as expected, in Sub-Saharan Africa (SSA) and the lowest in Western countries[36]. This picture is consistent in both years. The same is

---

[33]In the early 1960s, data is often only available in five-year spans. Taking six-year averages increases the data availability for many countries for the first point in time.

[34]Analyzing the adjustment times between policy changes and ELF value changes, which might differ considerably between variables, exhibits an interesting area for future research.

[35]In total data is reported for 151 countries for the two points in time. However, some (former) countries where no additional data was available and countries that changed considerably over the time due to secession (e.g. Pakistan/Bangladesh) or union (e.g. Vietnam) were excluded.

[36]This includes besides the European Countries also developed nations like Australia, Canada, Japan, New Zealand and the United States. Categorization is taken from Fearon (2003).

Region	ANM 1961			ANM 1985			Delta ('85-'61)		
	Obs.	Mean	Std.D.	Obs.	Mean	Std.D.	Obs.	Mean	Std.D.
World	138	0.463	0.278	168	0.461	0.272	138	0.006	0.086
Asia	22	0.483	0.295	27	0.467	0.306	22	-0.035	0.053
E. Europe	5	0.138	0.094	26	0.371	0.207	5	-0.029	0.038
L. America	25	0.446	0.194	26	0.443	0.213	25	0.012	0.061
MENA	19	0.318	0.165	20	0.342	0.222	19	0.040	0.177
SSA	45	0.674	0.226	46	0.663	0.235	45	-0.011	0.037
W. Countries	22	0.231	0.210	23	0.273	0.227	22	0.050	0.076

Table 2: Summary statistics of Atlas Narodov Mira data for 1961, 1985 and its change between 1961 and 1985

true for intermediate ELF values for Asia, Latin America and the Middle East and North Africa (MENA). The huge change of ELF values in Eastern Europe between the 1961 values and the 1985 values comes entirely from an increase in observation from five countries in 1961 to 26 in 1985.

Regions that became more homogeneous (a decreasing ELF value) show negative values, whereas regions that became more heterogeneous (an increasing ELF value) show positive values. Although the median country per region did not change much, all except for 19 countries report a change in their respective ELF value. The biggest changes were experienced in the MENA region, where countries moved significantly in both directions. Nevertheless, some tendencies of regional drift can be noted. Whereas Asia experienced rather a homogenization, Latin America and the Western countries showed some heterogenization. Sub-Saharan Africa did not experience much variation over the 20 years in question.

5 Results

5.1 Influential factors on a 'base-level' of diversity

The regressions of *Table 4* are based on equation (2) and includes the major geographical variables already discussed. *Latitude* reflects the distance from the equator, *Altitude* measures the altitude variation that is found within a country, and *Area* is its surface area. The more distant a country is located from the equator, we would expect a decreasing biodiversity and in turn also a lower ethnic diversity. *Latitude* has the expected negative sign and is highly significant. Larger and more mountainous countries have a higher probability to encompass different habitats. This opens more solitary areas that facilitate the development of different species and ethnic groups and additionally acts

as barrier to sustain them. Both *Altitude* and *Area* show a significant positive impact on higher diversity levels.

The fourth variable included in the first regression is *Agritime*. It captures the time elapsed since the transition from a hunter-gatherer economy to agricultural production. The earliest countries transitioned around 8500 B.C. and the latest only around 1600 A.D.[37]. Countries that made the transition earlier in time should then show a lower level of diversity as they had more time to develop into more advanced civilizations. As expected, *Agritime* displays a negative sign that is significant at the 5% level. The different transition times between the first and the last countries (approx. 10.000 years) lead to 0.16 lower ELF values.

In regression (2), another variable used by Ahlerup and Olsson (2007) is included. The experience of a modern state captures how many years in the time between 1800 and 1950 a country had power over its territory. It has a comparable interpretation as *Agritime* but in a way captures the final result or how well an early civilizations developed into modern civilizations. Therefore, it comes as no surprise that both variables point in the same direction. As soon as *Modern* is included it remains significant whereas *Agritime* loses its significance.

Regression (3) controls for more specific geographical characteristics, including a *Tropics* variable and regional dummies. The *Tropics* variable measures the percentage of a country's total areas classified as being exposed to tropical climate. As expected, one finds a positive and significant correlation between tropical climate and diversity. Except for the Sub-Saharan Africa dummy, which is marginally significant, none of the other regional dummies is significant. *Latitude*, which was highly significant in all previous regressions, loses its significant explanatory power when the regional dummies are included. This is not too surprising as the regional division partly reflects the distance from the equator. Additionally, *Tropics* seem to better capture the idea of a different habitat around the equator. Nevertheless, the major geographical variables *Altitude* and *Area* rather maintain their significance. Thus, latitude per se is not the driver of a different diversity structure but the different geographical and climatic conditions found along the latitudinal stretch.

More democratic regimes are considered to give their citizens more freedom of personal expression and might thus also show a higher level of diversity. Democratic tradition is measured by the average Polity score after World War

---

[37]The first were Israel, Jordan, Lebanon and the Syrian Arab Republic, whereas Mauritius and Australia were the last.

II (1945—1960)[38] developed by Marshall and Jaggers (2008). *Democratic Tradition* displays the expected positive sign at least at the 10% level[39].

Regressions (5)-(8) of *Table 4* replicate the first four regressions for the second *ANM* data set of the 1980s. As the covariates did not change between the two points in time, there should not be a different result in taking the later one. As expected, there is no qualitative difference between the two data sets and the results remain very much comparable[40]. All results so far are in line with the results of Ahlerup and Olsson (2007) and Michalopoulos (2008). As these authors test their hypotheses not on the *ANM* data but on the ELF indices from Alesina *et al.* (2003) and Fearon (2003), regression (1) and (2) from *Table 4* are replicated for both alternative sources. The results are reported in *Table 5* and generally support all findings, discussed so far. These results give additional credibility to the *ANM* data.

Exceeding the scope of Ahlerup and Olsson (2007) and Michalopoulos (2008), this article investigates further on the grounds of colonization. *Table 6* shows the main results[41]. Regression (1) is the already known set-up for the full set of countries. In regression (2), a *Colony* dummy is included to control if former colonies generally show differences in their diversity set-up from non colonial countries. Former colonies seem to display significantly lower levels of diversity. Not only the statistic significance but also the economic one is big. Former colonies are attributed with an around 21% lower level of diversity. This result could be driven by the linguistic bias of the diversity data. Especially in Latin America, the colonial regime rather left a common language. Regression (3) tries to prove this by entering interaction terms of the colony and the regional dummies for Latin America and Sub Saharan Africa. Although they have the expected sign, Latin America negative and Sub-Saharan Africa positive, both

---

[38]Alesina and Zhuravskaya (2011) use a comparable time frame to assess a democratic tradition variable. Additionally, Campos and Kuzeyev (2007) also found significant changes in the ethnic set-up over a short time frame of 13 years after the fall of the Iron curtain. However, if the time frame for the *Democratic Tradition* variable is extended to 1900–1960, the results do not change, but the observations are further reduced.

[39]A caveat is that the inclusion of the *Democratic Tradition* variable nearly halves the number of observations. That is also why these variables in coming regressions are not included unless explicitly controlled for the role of democracy.

[40]In an additional regression GDP/capita levels in 1960 were included in the regressions of *Table 4*. Whereas the GDP/capita levels in 1960 had only a significant influence in less than halve of the regressions, *Latitude*, *Area*, *Modern* and the *MENA* and the *SSA* dummy retained their significance or improved. *Altitude*, *Agritime* and *Democratic tradition* lost significance. As the number of observation is again significantly reduced, these results are not explicitly reported here.

[41]In the following regressions the 1985 *ANM* data is used as it contains more observations than the earlier version. However, the results do not change compared to the earlier data.

are not significant. The result of the *Colony* variable is not altered very much[42]. The longer the colonial powers stayed the more settlers might have domiciliated in the new countries permanently. Aligned with earlier findings of Ahlerup and Olsson (2007), the colonial duration (*Duration*) has a positive, but barely significant, impact on diversity displayed in regression (4). Controlling in regression (5) for the colonizer's homeland, one finds no significant correlation with the French, Spanish or British colonizers.

In regression (6) and (7), the idea of Acemoglu *et al.* (2001) was picked up, exploring the implication how, rather than by whom countries were colonized. In countries with high mortality rates, colonial powers did not show any interest in establishing permanent settlements. This 'divide-and-rule' approach should have left its mark in the form of higher diversity levels. Ethnic differences were often leveraged to play groups off against each other. Building on prevalent tensions, lead to the emergence and a segmenting of groups along group lines. Indeed, *Mortality* shows a significant positive correlation with the level of diversity[43]. Including the *Mortality* estimate also affects the colonizer homeland dummies, rendering the British dummy significant. This is very much in line how Bolt and Bezemer (2009) describe the educational system in the British colonies. Whereas the French colonizers pursued an approach of assimilation and introduced a centrally controlled system, the British colonizers pursued a more indirect influence. They exercised a rule, "where traditional structures and institutions were left intact" and acted with a "relative tolerance towards local customs" (Bolt and Bezemer, 2009, p. 30)

Analyzing the influence of evolutionary and historical factors, two important insights turned out. Earlier findings with different data sets, showing that geographical attributes (especially *Altitude, Area* and *Latitude*) are highly responsible for the 'base level' of diversity, are confirmed. Second, attention is drawn to the role of colonization. This article argues that the homeland of the colonizers is less important for a former colony's diversity than how colonial powers pursued their endeavors.

5.2 Drivers of diversity level changes over a short period

Table 7 reports the first results of the regressions based on equation (3). It contains all variables that display a change over the period covered, i.e., variables

---

[42] An additional caveat is that it is hard to distinguish whether the effect does not reflect a reverse causality and colonial powers just chose more homogeneous countries for their colonization efforts.

[43] Including the mortality variable, increases the explanatory power of the model, increasing the adjusted R^2 from 0.39 to 0.52 between regression (1) and (5). However, the number of observations decreased again significantly.

of the vector ΔX_i. *Latitude, Altitude, Area, Agritime* and the *ANM* values in 1961 are included as static variables of vector Z_i. Although the variables of vector Z_i do not show any changes over time, they might have a mediating role for diversity adaptations. This is why they are controlled for in all regressions in this table. However, nearly none of the variables are significant, and the values are not explicitly reported here.

As discussed earlier, data availability poses a major limitation to the regressions. This article tries to make the best possible trade-off between including additional variables and thereby reducing the risk of omitted variables by not too much downsizing the number of observations available.

Regression (1) controls for the most important changes in developing countries regarding their settlement and population pattern. Metropolitan areas attract people from the countryside with the prospect of a better economic future. Many old traditions are left behind, and one tries to merge into a more mainstream culture of major cities. Indeed, a significant negative impact of the change in *Urbanization*, the percentage of the population living in urban areas, on the diversity level is found. As expected, the most obvious effect of *Immigration* on heterogeneity is positive. Both are significant at the 5% level. Comparing both effects, immigration plays a bigger role. An increase in one standard deviation of immigration change increases the diversity change by 0.45 standard deviations, whereas the same change in urbanization leads to a decrease of -0.19 standard deviations. Population density (*Density*) shows no significant impact in this first regression.

In regression (2), primary schooling rates (*Primary Schooling*) are included. This variable does not only cover the educational attainment and in a way the level of education in a country but can be understood as a proxy for state influence on an increasing part of the population. *Primary Schooling* seems to be a very important variable. It shows a significant negative impact and lowers the size and significance level of both, *Urbanization* and *Immigration*. *Primary Schooling* and *Immigration* display the highest impact with beta-coefficients of -0.26 and 0.41, respectively. Controlling for various other variables in regressions (3)-(6), the significant influence of *Urbanization*, *Immigration* and *Primary Schooling* persists at least at the 5% level. Neither a change in the level of democracy (*Polity IV*), the number of conflicts (*Conflicts*) nor *Trade* and Infrastructure (*Telephones*) show any significant impact. Including changes in GDP per capita levels (*GDP/capita*) in regression (7), turns all variables except for *Primary Schooling* and *Urbanization* insignificant. Although the variable carries the expected negative sign, it is per se insignificant. Most of the socioe-

conomic policy variables are very strongly associated with higher wealth levels of a country, reflected in growing GDP/capita levels. That *Primary Schooling* and *Urbanization* remain significant, although the GDP/capita increase is included, confirm their robustness. Regression (7) is also the only one where *Immigration* loses its significance. As immigrants are attracted by prosperous countries, i.e., countries with high GDP/capita growth rates, a high correlation with immigration is inevitable. Controlling for regions (8)), does not add any new insights as all of these variables are not significant. They also do not lead to any major changes in the variables focussed here.

As was already pointed out in the discussion of the economic and policy factors, it is hard to see why GDP/capita levels should have a direct impact on diversity. The regressions in *Table 7* already showed some influential factors that all are highly linked to the GDP/capita levels and the progress, the GDP/capita growth rates. However, as the overall economic development of a country plays a crucial role, it is also controlled for it here. This is done less in order to generate new insights than as an additional robustness check. Taking selected regressions of *Table 7* in *Table 8* various measures of (economic) development are included. Regressions (1) and (2) are the already known ones. In regression (3) and (4), additionally the GDP/capita level in 1960 based on the Penn World Tables (Heston *et al.*, 2009) is included in the otherwise unchanged set-up[44]. *Urbanization* and *Immigration* that show the highest correlation with the GDP/capita level turns insignificant. Instead the GDP/capita level at the beginning of the period is positive in all regressions at least at the 5% level. *Primary Schooling* shows lower, but still significant values if GDP/capita levels are included. If the GDP growth (change in GDP/capita levels) are included the significance fades. This has two important interpretations. First, the results for *Primary Schooling* are robust. Although the GDP/capita level variable absorbs some of its influencing value, its significance does not change considerably. Countries that are richer already have much higher primary schooling figures, so changes would be expected to be smaller. Still the influence persists. Second, countries that already have a higher development level seem already to move in the other direction, thus getting more heterogeneous. Most of the highly developed countries are classic immigration countries, like the US, Australia and Canada. That hints to the curvilinear relationship of ethnic identification and development or 'modernization' discussed in Bannon *et al.* (2004). Ethnic fragmentation is not necessarily a sign of backwardness.

---

[44]The results displayed are based on the Laspeyres index of the Penn World Tables. The regressions with the Chain index yield the same results.

Regressions (5) and (6) use HDI numbers (UNDP - United Nations Development Programme, 1994). The Human Development Indicator (HDI) is a broader indicator of development, not only taking into account the GDP/capita levels but also health and education figures. In general, the results are very much comparable to the results discussed above. The broader construction of the HDI, especially including schooling variables, explains why the HDI variable is the only one where also the change variable has a significant and negative impact, taking up the influence of the *Primary Schooling* variable.

As additional robustness checks, the key regressions of *Table 7* are run again with different model specifications. Both fixed-effect (FE) and the random-effect (RE) models are tested. Using the FE model, a correlation between the entity specific error term and the explanatory variables is allowed. Furthermore, all level variables that are time-invariant are removed from the regressions to assess only the regressors' net effect. The influence of the time-invariant characteristics that were also controlled in the earlier regression would thus be covered in the error terms. The RE model, in contrast, assumes the independence between the entity error term and the explanatory variables. From the discussion above, the better suitability of the FE model is clear. However, the Hausman-test only supports in half of the regression pairs clearly the FE model[45]. *Table 9* summarizes the results of both models. Although the values of the coefficients vary, the significant positive or negative effects of the main variables *Urbanization, Immigration* and *Primary Schooling* are clearly confirmed.

Because *Primary Schooling* seems to play a crucial role, *Table 10* depicts the influence of different measures of education as well as various education levels to test the robustness of the finding. Regression (1) corresponds to the second regression in *Table 7*. *Immigration* and *Primary Schooling* are both significant[46]. In regression (2), additionally *Secondary Schooling* and *Tertiary Schooling* are included. The coefficient of *Primary Schooling* remains significant and increases in size. Looking at the role of higher education, reveals an additional interesting insight. *Secondary Schooling* enters the regression with a significant positive sign. Higher education apparently has a different effect on diversity than primary education. While the effect of primary education is uniformly negative, secondary education enters mostly positively into the

---

[45]This is the case for the regression pairs (2/6) and (3/7). Results of the Hausman test are not reported here.

[46]As in the previous regression, controls for *Latitude, Altitude, Area, Agritime* and the ELF 'base-level' in 1961 (*ANM '61*) are included in the regressions; details on the results are not explicitly reported here.

equations[47].

Regressions (3)-(7) try to fundament the findings with different measures of education offered by Barro and Lee (2010). The total sum of all years of schooling (*Schooling total*) does not show any significant impact. This is not surprising. As primary and higher educational levels enter the regression with opposite signs, they seem to cancel out if summed up together. All other regressions confirm the homogenizing impact of primary education. In most cases, the positive impact of higher education is also confirmed. However, the coefficients are no more significant. These robustness checks confirm the apparent importance of primary schooling for a country's homogenization and does not depend on the definition or measure of primary education.

In section 4, the time frame chosen was already discussed. For the reason of endogeneity and time needed for potential adjustments of diversity, the time frame 1960/65—1975/80 was chosen. Nevertheless, the results should not entirely depend on the choice of the time frame. As an additional robustness check, the time frame for all policy variables was changed from 1960/65—1975/80 to 1960/65—1980/85. The results are reported in *Table 11*. Although the coefficient sizes vary slightly, the significance levels of all variables discussed only change marginally.

6   Conclusion

In line with the recent publications of Ahlerup and Olsson (2007) and Michalopoulos (2008) on the roots of ethnically diverse countries, the major results are confirmed. Although different data and data sources were partly used, the results remain robust. Geographical characteristics, like a country's surface and altitude variation, and evolutionary factors, like the transition from sedentary farming, are major drivers of a 'base-level' of ethnic fragmentation. To the analysis of geographical and historical factors a more detailed view on colonization is added. Whereas the homeland of the colonizer seem to play no important role, the way how a country was colonized does show a significant impact. Countries, where colonial powers did not have any incentive to settle and build good institutions but rather exploited the country's resources, show a significant higher level of ethnic fragmentation. Ethnic boundaries seemed to

---

[47]Knack and Keefer (1997) find in their analysis of education's role on trust a comparable differentiated result for primary and secondary education. Additionally, Bannon *et al.* (2004) comparably find in their analysis of ethnic identification for a small set of African states that students identify themselves more along ethnic lines than farmers.

be an easy line for playing one group off against the other. Mistrust and rifts between ethnic groups seem to persist also after independence — mirrored in higher fragmentation levels.

What this paper mainly wants to add to the recent discussion, is that ethnic fragmentation can not be treated as being exogenous or being only rooted in geographic and historical factors. Especially since the beginning of the 20th century, various policy and economic factors significantly changed the dynamics between ethnic groups, their interchange and assimilation, as well as migration patterns. Migration proves to be the most important factor in changing a country's diversity. Gulf countries, relying heavily on immigrants, show this trend most clearly. Doubtless, migration plays an even bigger role in the globalized world subsequently the time frame analyzed in this paper. Its impact might thus be even more pronounced today. The same is true for the other variables shown to have a significant impact on a country's ethnic fragmentation. Other more policy-induced variables, like urbanization and especially primary education, leave their marks on a country's diversity. Urbanization and the growth of metropolitan areas, attracting huge parts of the population, lead to an erosion of old habits and to an assimilation into or the emergence of a 'mainstream' culture. Education is, according to the findings of this paper, not only a measure of a higher educational level attained. Because primary education is in general the first contact point with the state authorities, it is also a good proxy for the government's influence. By expanding the government's reach for more remote areas and in turn by increasing the primary education numbers, more and more people are exposed to its influence. In line with recent findings of other authors, education is not influencing diversity uniformly. The empirical results support the theoretical findings that higher educational levels lead to a more heterogeneous society.

Nevertheless, the paper also faces some limitations. Due to data limitation in the early 1960s, not all of the variables discussed can be tested. Only data on ethno-linguistic fragmentation and not on other concepts regarding language or religion was available. In line with Campos and Kuzeyev (2007), the distinction between ethnic, linguistic and religious diversity could be an interesting field for future research. Not only that these different characteristics might be driven by different factors, the time span in which changes occur and the direction might also be different. Additionally, the changes may be less visible due to the highly linguistic definition of the *ANM* data. Both Campos and Kuzeyev (2007) and Fedderke and Luiz (2007) find more significant changes in ethnic and racial set-up than for the linguistic and religious characteristics.

Despite these limitations, the set of variables and data used for this article show clear and very robust results. They are a very good basis to refute the assumption of diversity's exogeneity. Thus, this paper is not only a caveat for the interpretation of most of the current studies to be aware of these changes, but it also marks a first assessment of what might be the drivers of these changes.

Admitting that a country's ethnic set-up changes and that it can be influenced, turns back to the growing literature on the effects of ethnic diversity. Having seen that the ethnic composition is changing with variables that are highly linked to the development level of a country, using a fixed measure of ethnicity for economic growth analysis seems rather unreasonable. This would attach greater importance to older measures of the ex ante ethno-linguistic composition of a country in the analysis of economic outcomes, because the ethnic set-up may have been endogenously determined by the factors under investigation. This is exactly what Campos and Kuzeyev (2007) find for their data set on former Soviet Republics. Whereas the effect of an exogenous diversity measure on growth is limited, the dynamic, endogenous measure illustrates a significantly negative effect.

More than a caveat, this article offers a first attempt to venture into the dynamics of diversity and gives some better understanding as to how policy, intentionally or not, can shape a country's ethnic set-up.

References

Acemoglu, D., Johnson, S. and Robinson, J. A. (2001) The Colonial Origins of Comparative Development: An Empirical Investigation, *American Economic Review*, **91**, 1369–1401.

Ahlerup, P. and Olsson, O. (2007) The Roots of Ethnic Diversity, Working Papers in Economics 281, Department of Economics, Göteborg University.

Akdede, S. H. (2010) Do more Ethnically and Religiously Diverse Countries have Lower Democratization?, *Economics Letters*, **106**, 101–104.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003) Fractionalization, *Journal of Economic Growth*, **8**, 155–194.

Alesina, A. and La Ferrara, E. (2005) Ethnic Diversity and Economic Performance, *Journal of Economic Literature*, **43**, 762–800.

Alesina, A. and Zhuravskaya, E. (2011) Segregation and the Quality of Government in a Cross-Section of Countries, *American Economic Review*, p. forthcoming.

Bannon, A., Miguel, E. and Posner, D. N. (2004) Sources of Ethnic Identification in Africa, Afrobarometer Working Papers 44, Afrobarometer.

Barro, R. J. (1999) Determinants of Democracy, *The Journal of Political Economy*, **107**, 158–183.

Barro, R. J. and Lee, J.-W. (2010) A New Data Set of Educational Attainment in the World, 1950-2010, NBER Working Papers 15902, National Bureau of Economic Research.

Barro, R. J. and McCleary, R. (2003) Religion and Economic Growth, NBER Working Papers 9682, National Bureau of Economic Research.

Bates, R. H. (2006) Ethnicity, in *The Elgar Companion to Development Studies* (Ed.) D. A. Clark, Edward Elgar Publishing, Cheltenham, pp. 167–173.

Bellini, E., Ottaviano, G. I. P., Pinelli, D. and Prarolo, G. (2009) Cultural Diversity and Economic Performance: Evidence from European Regions, Working Papers 2009.63, Fondazione Eni Enrico Mattei.

Bezemer, D. J. and Jong-A-Pin, R. (2008) World on Fire? Democracy, Globalization and Ethnic Violence, MPRA Paper 7027, University Library of Munich.

Bisin, A. and Verdier, T. (2001) The Economics of Cultural Transmission and the Dynamics of Preferences, *Journal of Economic Theory*, **97**, 298–319.

Bolt, J. and Bezemer, D. (2009) Understanding Long-Run African Growth: Colonial Institutions or Colonial Education?, *The Journal of Development Studies*, **45**, 24–54.

Bruk, S. I. (1964) *Atlas Narodov Mira*, N. N. Miklucho-Maklaja, Moskva, Institut Etnografii Imeni.

Bruk, S. I. and Pučkov, P. I. (1986) *Naselenie Mira*, Nauka, Moskva, 2 edn., Etnodemografičeskij Spravočnik.

Campos, N. F. and Kuzeyev, V. S. (2007) On the Dynamics of Ethnic Fractionalization, IZA Discussion Papers 2822, Institute for the Study of Labor.

Caselli, F. and Coleman, W. J. (2008) On the Theory of Ethnic Conflict, CEDI Discussion Paper Series 08-08, Centre for Economic Development and Institutions, Brunel University West London.

Cashdan, E. (2001) Ethnic Diversity and its Environmental Determinants: Effects of Climate, Pathogens, and Habitat Diversity, *American Anthropologist*, **103**, 968–991.

CIA (2009) *The World Factbook*, `https://www.cia.gov/library/publications/the-world-factbook/`.

Coleman, D. (2009) Divergent Patterns in the Ethnic Transformation of Societies, *Population and Development Review*, **35**, 449–478.

Collier, P. (1998) The Political Economy of Ethnicity, Working Paper Series 98-8, Centre for the Study of African Economies, University of Oxford.

Collier, P. (2000) Ethnicity, Politics and Economic Performance, *Economics and Politics*, **12**, 225–245.

Collier, P. and Hoeffler, A. (1998) On Economic Causes of Civil War, *Oxford Economic Papers*, **50**, 563–573.

Constant, A. and Zimmermann, K. F. (2007) Measuring Ethnic Identity and Its Impact on Economic Behavior, IZA Discussion Papers 3063, Institute for the Study of Labor.

Darity, W. A., JR., Mason, P. L. and Stewart, J. B. (2006) The Economics of Identity: The Origin and Persistence of Racial Identity Norms, *Journal of Economic Behavior and Organization*, **60**, 283–305.

Dreher, A. (2006) Does Globalization Affect Growth? Evidence from a New Index of Globalization, *Applied Economics*, **38**, 1091–1110.

Easterly, W. and Levine, R. (1997) Africa's Growth Tragedy: Policies and Ethnic Divisions, *The Quarterly Journal of Economics*, **112**, 1203–1250.

Encyclopædia Britannica (Ed.) (2009) *The New Encyclopædia Britannica*, Encyclopædia Britannica Inc., Chicago, Ill., 15. ed. edn.

Eriksen, T. H. (1991) Ethnicity versus Nationalism, *Journal of Peace Research*, **28**, 263–278.

European Commission (2006) *Turkey 2006 Progress Report*, Commission staff working document – SEC(2006) 1390, Brussels.

Fearon, J. D. (2003) Ethnic and Cultural Diversity by Country, *Journal of Economic Growth*, **8**, 195–222.

Fearon, J. D. and Laitin, D. D. (1999) Weak States, Rough Terrain, and Large-scale Ethnic Violence since 1945, Atlanta, Ga., Presented at the Annual Meeting of the American Political Science Association.

Fearon, J. D. and Laitin, D. D. (2000) Violence and the Social Construction of Ethnic Identity, *International Organization*, **54**, 845–877.

Fedderke, J., Luiz, J. and de Kadt, R. (2008) Using Fractionalization Indexes: Deriving Methodological Principles for Growth Studies from Time Series Evidence, *Social Indicators Research*, **85**, 257–278.

Fedderke, J. W. and Luiz, J. M. (2007) Fractionalization and Long-run Economic Growth: Webs and Direction of Association Between the Economic and the Social: South Africa as a Time Series Case Study, *Applied Economics*, **39**, 1037–1052.

Fernandez, R. and Fogli, A. (2009) Culture: An Empirical Investigation of Beliefs, Work, and Fertility, *American Economic Journal: Macroeconomics*, **1**, 146–177.

G-Econ (2006) *Geographically Based Economic Data 1.3*, `http://gecon.yale.edu/`.

Garcia-Montalvo, J. and Reynal-Querol, M. (2003) Religious Polarization and Economic Development, *Economics Letters*, **80**, 201–210.

Garcia-Montalvo, J. and Reynal-Querol, M. (2005) Ethnic Diversity and Economic Development, *Journal of Development Economics*, **76**, 293–323.

Garcia-Montalvo, J. and Reynal-Querol, M. (2008) Discrete Polarisation with an Application to the Determinants of Genocides, *The Economic Journal*, **118**, 1835–1865.

Gellner, E. (1983) *Nations and Nationalism*, Cornell University Press, Ithaca, 1 edn.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M. and Strand, H. (2002) Armed Conflict 1946-2001: A New Dataset, *Journal of Peace Research*, **39**, 615–637.

Greif, A. (1993) Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition, *The American Economic Review*, **83**, 525–548.

Guiso, L., Sapienza, P. and Zingales, L. (2004) Cultural Biases in Economic Exchange, NBER Working Papers 11005, National Bureau of Economic Research.

Héran, F., Filhon, A. and Deprez, C. (2002) La dynamique des langues en France au fil du XXe siècle, Bulletin mensuel d'information 376, L'institut national d'études démographiques.

Heston, A., Summers, R. and Aten, B. (2009) Penn World Table Version 6.3, Data Set, Center for International Comparisons of Production, Income and Prices, University of Pennsylvania.

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

Janeba, E. (2004) International Trade and Cultural Identity, NBER Working Papers 10426, National Bureau of Economic Research.

Knack, S. and Keefer, P. (1997) Does Social Capital Have an Economic Payoff? A Cross-Country Investigation, *The Quarterly Journal of Economics*, **112**, 1251–1288.

Kolo, P. (2011a) Measuring a New Aspect of Ethnicity – Construction of an Appropriate Diversity Index, mimeo, Georg-August University Goettingen.

Kolo, P. (2011b) On the Dynamics of Ethnicity, mimeo, Georg-August University Goettingen.

Lazear, E. P. (1999) Culture and Language, *The Journal of Political Economy*, **107**, 95–126.

Leeson, P. T. (2005) Endogenizing Fractionalization, *Journal of Institutional Economics*, **1**, 75–98.

Lewis, M. P. (2009) *Ethnologue; Languages of the World*, Summer Institute of Linguistics (SIL), Dallas, Tex., 16 edn.

Limao, N. and Venables, A. J. (2001) Infrastructure, Geographical Disadvantage, Transport Costs, and Trade, *World Bank Economic Review*, **15**, 451–479.

Marshall, M. G. and Jaggers, K. (2008) *Political Regime Characteristics and Transitions, 1800-2008: Polity IV Project*, `http://www.systemicpeace.org/polity/polity4.htm`.

Mauro, P. (1995) Corruption and Growth, *The Quarterly Journal of Economics*, **110**, 681–712.

Michalopoulos, S. (2008) The Origins of Ethnolinguistic Diversity: Theory and Evidence, MPRA Paper 11531, University Library of Munich.

Olsson, O. (2007) On the Institutional Legacy of Mercantilist and Imperialist Colonialism, Working Papers in Economics 247, Department of Economics, Göteborg University.

Olsson, O. and Hibbs, D. A. (2005) Biogeography and Long-Run Economic Development, *European Economic Review*, **49**, 909–938.

Perez, A. D. and Hirschman, C. (2009) The Changing Racial and Ethnic Composition of the US Population: Emerging American Identities, *Population and Development Review*, **35**, 1–51.

Posner, D. N. (2004) Measuring Ethnic Fractionalization in Africa, *American Journal of Political Science*, **48**, 849–863.

Putterman, L. (2008) *State Antiquity Index*, `http://www.econ.brown.edu/fac/Louis_Puttermann`.

Putterman, L. and Weil, D. N. (2010) Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality, *The Quarterly Journal of Economics*, **125**, 1627–1682.

Ranis, G. (2009) Diversity of Communities and Economic Development: An Overview, Center Discussion Paper 977, Economic Growth Center, Yale University.

Roeder, P. G. (2001) *Ethnolinguistic Fractionalization (ELF) Indices, 1961 and 1985*, `http://weber.ucsd.edu/~proeder/elf.htm`.

Sachs, J. D. (2001) Tropical Underdevelopment, NBER Working Papers 8119, National Bureau of Economic Research.

Saez-Marti, M. and Sjögren, A. (2007) Peers and Culture, Working Paper Series 349, Institute for Empirical Research Economics, University of Zurich.

Schüler, D. and Weisbrod, J. (2010) Ethnic Fractionalisation, Migration and Growth, *Empirical Economics*, **39**, 457–486.

Spolaore, E. and Wacziarg, R. (2009) The Diffusion of Development, *The Quarterly Journal of Economics*, **124**, 469–529.

Taylor, C. L. and Hudson, M. C. (1972) *World Handbook of Political and Social Indicators*, Yale University Press, New Haven.

UNDP - United Nations Development Programme (1994) *Human Development Report 1994*, Oxford University Press, New York.

van den Berghe, P. L. (1981) *The Ethnic Phenomenon*, Elsevier, New York.

Weber, E. (1976) *Peasants into Frenchmen; the Modernization of Rural France, 1870-1914*, Stanford University Press, Stanford, Calif.

Weidmann, N. B., Rød, J. K. and Cederman, L.-E. (2010) Representing Ethnic Groups in Space: A New Dataset, *Journal of Peace Research*, **47**, 491–499.

World Bank (2009) *World Development Indicators 2009: CD-ROM and Book*, The World Bank Group, Washington, DC.

A   Appendix

Table 3: Overview of variables, definitions and sources

Variable name	Description	Source
ANM	Atlas Narodov Mira (ANM) Ethno-linguistic fractionalization index (ELF)	Roeder (2001)
Alesina	Ethno-linguistic fractionalization index (ELF) of Alesina	Alesina *et al.* (2003)
Fearon	Ethno-linguistic fractionalization index (ELF) of Fearon	Fearon (2003)
Latitude	Absolute value of the latitude of a country's capital, scaled to take values between 0 and 1, where 0 is the equator	Cepii (2011)
Altitude	Average absolute deviation between grid and country mean altitude	Based on G-Econ (2006)
Area	Log of country area in km2	World Bank (2009)
Agritime	Years since transition to agriculture (in '000 years) in relation to the base year 2000	Putterman (2008)
Modern	State power over territory between 1800 and 1950 in years*	Putterman and Weil (2010)
Democratic tradition	Average polity2 score (ranging from -10 to 10), with lower values indicating a less democratic or autocratic regimes (negative values) for the years after WWII 1945—1960. Only countries with observation for at least half of the years included	Marshall and Jaggers (2008)
Tropics	% land area in Koeppen-Geiger tropics and subtropics (Af+Am+Aw+Cw)	Sachs (2001)
Regional dummies	Dummy for Eastern Europe, Latin America, North Africa and Middle East, Sub-Saharan Africa, Western countries and Asia	Based on Fearon (2003)
Colony	Dummy variable that takes value one if country was colonized and 0 if not	Based on data in Olsson (2007)
Duration	Total number of years under colonial rule*	Olsson (2007)
Spanish, French, British, Portuguese Colony	Dummy variable for Spanish, French, British or Portuguese colonization of the country	Cepii (2011)
Mortality	Log of potential settler mortality, measured in terms of deaths per annum per 1.000 "mean strength" of settlers	Acemoglu *et al.* (2001)
Urbanization	% of population living in urban areas	World Bank (2009)
Immigration	International migrant stock (% of population)	World Bank (2009)
Density	Population density (people per km2)	World Bank (2009)
Polity IV	Average polity score (ranging from -10 to 10), with lower values indicating a less democratic or autocratic regimes (negative values)	Marshall and Jaggers (2008)
Conflicts	Years with summed magnitudes of all societal major events of political violence (MEPV) higher than 1	Gleditsch *et al.* (2002)
Trade	Log of trade (% of GDP)	World Bank (2009)
Telephones	Mobile and fixed-line telephone subscribers (per 100 people)	World Bank (2009)
GDP/capita - Penn	Log of real GDP per capita in constant international dollars (Laspeyres index)	Heston *et al.* (2009)
HDI	Human Development Indicator, measures development along three dimensions: healthy life, GDP/capita and education	UNDP - United Nations Development Programme (1994)
Prim., Sec., Tert. Enrollment	% of population aged 15 and over that attained respective school	Barro and Lee (2010)
Prim., Sec., Tert. Completion	% of population aged 15 and over that completed respective school	Barro and Lee (2010)
Prim., Sec., Tert. Schooling	Average years of respective school attainment of population aged 15 and over	Barro and Lee (2010)

* For better readability in regression tables, rescaled to decades.

Table 4: Influence of geographic and historical variables on Atlas Narodov Mira
ELF scores in 1961 and 1985

	(1) ANM '61	(2) ANM '61	(3) ANM '61	(4) ANM '61	(5) ANM '85	(6) ANM '85	(7) ANM '85	(8) ANM '85
Latitude	-0.880***	-0.703***	-0.377	-0.718***	-0.610***	-0.461***	-0.376	-0.702***
	(-8.51)	(-5.93)	(-1.44)	(-5.32)	(-5.66)	(-4.14)	(-1.44)	(-5.04)
Altitude	0.101*	0.147**	0.149***	0.203**	0.0664	0.143**	0.137***	0.191**
	(1.81)	(1.99)	(2.71)	(2.54)	(1.30)	(2.34)	(2.68)	(2.40)
Area	0.0274***	0.0411***	0.0304**	0.0365*	0.0241***	0.0409***	0.0230*	0.0334*
	(3.12)	(4.05)	(2.56)	(2.00)	(2.65)	(4.16)	(1.91)	(1.82)
Agritime	-0.0161**	-0.0135	-0.00418	-0.00717	-0.0147*	-0.0214**	-0.00557	-0.00871
	(-2.10)	(-1.66)	(-0.37)	(-0.60)	(-1.77)	(-2.56)	(-0.49)	(-0.72)
Modern		-0.0265***		-0.0227**		-0.0274***		-0.0167*
		(-4.44)		(-2.42)		(-4.93)		(-1.69)
Tropics			0.211**				0.175*	
			(2.40)				(1.88)	
Asia			0.0148				-0.0508	
			(0.14)				(-0.51)	
Eastern Europe			-0.0747				0.0991	
			(-1.34)				(1.59)	
Latin America			-0.0975				-0.113	
			(-0.98)				(-1.15)	
N. Africa, Middle East			0.0323				-0.000686	
			(0.42)				(-0.01)	
SSA			0.171*				0.134	
			(1.71)				(1.31)	
Democratic Trad.				0.00598*				0.00890**
				(1.70)				(2.40)
_cons	0.609***	0.736***	0.260	0.653***	0.576***	0.716***	0.355*	0.608***
	(9.42)	(10.78)	(1.29)	(3.47)	(8.95)	(10.64)	(1.80)	(3.14)
N	130	114	124	66	158	142	151	66
adj. R^2	0.462	0.516	0.544	0.427	0.279	0.385	0.350	0.359
F	38.62	25.30	32.31	13.07	17.79	21.05	12.24	10.46

Heteroscedasticity robust standard errors used; t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Comparison between various ELF measures - influence of geographic and historical variables

	(1) ANM '61	(2) ANM '61	(3) ANM '85	(4) ANM '85	(5) Alesina	(6) Alesina	(7) Fearon	(8) Fearon
Latitude	-0.880***	-0.703***	-0.610***	-0.461***	-0.897***	-0.778***	-0.738***	-0.549***
	(-8.51)	(-5.93)	(-5.66)	(-4.14)	(-8.46)	(-7.80)	(-7.08)	(-5.39)
Altitude	0.101*	0.147**	0.0664	0.143**	0.0538	0.0978	0.0692	0.163***
	(1.81)	(1.99)	(1.30)	(2.34)	(0.88)	(1.25)	(1.47)	(2.83)
Area	0.0274***	0.0411***	0.0241***	0.0409***	0.0347***	0.0440***	0.0178*	0.0318***
	(3.12)	(4.05)	(2.65)	(4.16)	(3.81)	(4.66)	(1.78)	(3.11)
Agritime	-0.0161**	-0.0135	-0.0147*	-0.0214**	-0.00476	-0.00364	0.00303	-0.00901
	(-2.10)	(-1.66)	(-1.77)	(-2.56)	(-0.55)	(-0.43)	(0.35)	(-1.00)
Modern		-0.0265***		-0.0274***		-0.0212***		-0.0291***
		(-4.44)		(-4.93)		(-3.17)		(-5.62)
_cons	0.609***	0.736***	0.576***	0.716***	0.526***	0.636***	0.572***	0.747***
	(9.42)	(10.78)	(8.95)	(10.64)	(8.96)	(9.29)	(7.68)	(9.89)
N	130	114	158	142	127	117	150	139
adj. R^2	0.462	0.516	0.279	0.385	0.441	0.501	0.281	0.410
F	38.62	25.30	17.79	21.05	30.67	27.05	18.65	24.37

Heteroscedasticity robust standard errors used; t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Influence of various colonization characteristics on Atlas Narodov Mira ELF scores

	(1) Full sample	(2) Full sample	(3) Full sample	(4) Colonies	(5) Colonies	(6) Colonies	(7) Colonies
Latitude	-0.461*** (-4.14)	-0.766*** (-5.12)	-0.660*** (-4.38)	-0.843*** (-4.13)	-0.488*** (-3.46)	-0.808*** (-4.13)	-0.921*** (-4.46)
Altitude	0.143** (2.34)	0.142** (2.44)	0.159*** (2.69)	0.146* (1.68)	0.140** (2.16)	0.200** (2.19)	0.315*** (2.88)
Area	0.0409*** (4.16)	0.0499*** (5.16)	0.0455*** (4.77)	0.0686*** (6.13)	0.0461*** (4.45)	0.0639*** (4.73)	0.0493*** (3.67)
Modern	-0.0274*** (-4.93)	-0.0289*** (-5.38)	-0.0254*** (-4.00)	-0.0368*** (-5.34)	-0.0393*** (-5.51)	-0.0367*** (-4.21)	-0.00709 (-0.72)
Agritime	-0.0214** (-2.56)	-0.0394*** (-3.41)	-0.0327*** (-2.86)	-0.0458*** (-2.65)	-0.0295*** (-3.25)	-0.0416** (-2.04)	-0.0527** (-2.43)
Colony		-0.208** (-2.59)	-0.185** (-2.30)				
Colony*LA			-0.00606 (-0.10)				
Colony*SSA			0.0923 (1.60)				
Duration				0.00259* (1.67)			0.000343 (0.18)
Spanish colony					0.000350 (0.00)		-0.122 (-1.51)
French colony					-0.00475 (-0.08)		0.0675 (0.89)
British colony						-0.0533 (-1.05)	0.151** (2.14)
Mortality						0.0307* (1.84)	0.0522*** (3.35)
_cons	0.716*** (10.64)	0.984*** (8.24)	0.867*** (6.99)	0.734*** (9.16)	0.850*** (9.61)	0.666*** (4.52)	0.329* (1.96)
N	142	142	142	86	117	60	58
adj. R^2	0.385	0.414	0.420	0.417	0.396	0.514	0.545
F	21.05	21.01	16.02	15.83	18.13	11.79	8.668

Heteroscedasticity robust standard errors used; t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Influence of socioeconomic and policy factors - dependent variable, change in Atlas Narodov Mira ELF scores

	(1) ANM change	(2) ANM change	(3) ANM change	(4) ANM change	(5) ANM change	(6) ANM change	(7) ANM change	(8) ANM change
Urbanization	-0.0592**	-0.0476*	-0.118***	-0.0797***	-0.0990***	-0.0798***	-0.0610**	-0.0597*
	(-2.50)	(-1.96)	(-3.33)	(-3.19)	(-4.19)	(-3.24)	(-2.39)	(-1.81)
Immigration	0.00589**	0.00543*	0.0116***	0.00985***	0.0112***	0.0118***	0.00537	0.00542*
	(2.05)	(1.91)	(3.52)	(3.48)	(3.34)	(3.86)	(1.25)	(1.85)
Density	-0.0350	-0.0111	0.122*	0.0979	0.190***	0.0758	-0.0224	-0.0154
	(-0.49)	(-0.16)	(1.83)	(1.53)	(2.78)	(1.21)	(-0.37)	(-0.22)
Primary Schooling		-0.0559**	-0.0442**	-0.0434**	-0.0558***	-0.0409**	-0.0312*	-0.0527**
		(-2.35)	(-2.18)	(-2.26)	(-2.76)	(-2.16)	(-1.76)	(-2.29)
Polity IV			-0.000507					
			(-0.53)					
Conflict				0.000572				
				(0.48)				
Trade					-0.00399			
					(-0.15)			
Telephones						0.0124		
						(0.92)		
GDP/cap - Penn							-0.0204	
							(-0.97)	
SSA								0.0174
								(0.45)
Latin America								0.0148
								(0.48)
Asia								-0.0339
								(-1.22)
_cons	0.0149	0.0703**	0.0557	0.0556	0.0446	0.0315	0.0670**	0.0540
	(0.51)	(2.14)	(1.37)	(1.54)	(1.00)	(0.77)	(2.22)	(1.13)
Level var. included	yes	yes	yes	yes	yes	yes	yes	yes
N	130	116	99	94	86	89	91	116
adj. R^2	0.208	0.255	0.470	0.550	0.558	0.589	0.161	0.263
F	3.024	2.799	3.090	2.380	9.435	4.418	3.198	2.397

Included level variables (Z_i): *Latitude, Altitude, Area, Agritime* and the *ANM* values in 1961
Heteroscedasticity robust standard errors used; t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Influence of various economic and human development levels at the beginning of the period (average 1960-65) - dependent variable, change in Atlas Narodov Mira ELF scores

	(1) ANM ch.	(2) ANM ch.	(3) ANM ch.	(4) ANM ch.	(5) ANM ch.	(6) ANM ch.
Urbanization	-0.0476*	-0.0610**	-0.0296	-0.0217	0.0228	0.0273
	(-1.96)	(-2.39)	(-1.01)	(-0.72)	(0.64)	(0.76)
Immigration	0.00543*	0.00537	0.00439	0.00442	0.00166	0.00148
	(1.91)	(1.25)	(1.15)	(1.15)	(1.21)	(1.14)
Density	-0.0111	-0.0224	-0.00124	-0.0196	-0.0264	-0.0249
	(-0.16)	(-0.37)	(-0.02)	(-0.32)	(-0.68)	(-0.67)
Primary Schooling	-0.0559**	-0.0312*	-0.0315*	-0.0254	-0.0363**	-0.0248
	(-2.35)	(-1.76)	(-1.93)	(-1.56)	(-2.13)	(-1.54)
GDP/cap. level ('60-'65) - Penn			0.0200*	0.0217**		
			(1.86)	(1.99)		
GDP/cap. - Penn		-0.0204		-0.0285		
		(-0.97)		(-1.39)		
HDI level ('60-'65)					0.142***	0.155***
					(3.40)	(3.66)
HDI						-0.365**
						(-2.43)
_cons	0.0703**	0.0670**	-0.112	-0.114	-0.0194	-0.00313
	(2.14)	(2.22)	(-1.23)	(-1.26)	(-0.52)	(-0.08)
Level var. included	yes	yes	yes	yes	yes	yes
N	116	91	91	91	98	98
adj. R^2	0.255	0.161	0.194	0.197	0.182	0.222
F	2.799	3.198	2.924	3.053	3.290	3.219

Included level variables (Z_i): *Latitude, Altitude, Area, Agritime* and the *ANM* values in 1961
t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Test of various other model specifications (fixed effects, random effects) - dependent variable, change in Atlas Narodov Mira ELF scores

	(1) Fixed effects	(2) Fixed effects	(3) Fixed effects	(4) Fixed effects	(5) Random effects	(6) Random effects	(7) Random effects	(8) Random effects
Urbanization	-0.0605**	-0.108***	-0.0743***	-0.0601***	-0.0604***	-0.0732***	-0.0678***	-0.0492***
	(-2.53)	(-4.11)	(-3.59)	(-2.66)	(-3.29)	(-3.32)	(-3.58)	(-3.06)
Immigration	0.00582*	0.0128***	0.0125***	0.00477	0.00594**	0.00960**	0.0127***	0.00424*
	(1.95)	(3.55)	(4.23)	(1.14)	(2.03)	(2.01)	(5.11)	(1.80)
Density	-0.00963	0.142**	0.0590	-0.0390	-0.0431***	-0.0354	-0.0359**	-0.0497***
	(-0.15)	(2.27)	(1.18)	(-0.88)	(-3.08)	(-1.63)	(-2.28)	(-3.55)
1975	0.0602***	0.0314	0.0327*	0.0682***	0.0643***	0.0724***	0.0682***	0.0693***
	(3.21)	(1.54)	(1.72)	(3.84)	(6.01)	(5.85)	(6.58)	(6.52)
Primary Schooling	-0.0496**	-0.0518***	-0.0397**	-0.0337**	-0.0403***	-0.0430***	-0.0289***	-0.0351***
	(-2.27)	(-2.89)	(-2.42)	(-2.03)	(-4.12)	(-4.66)	(-2.81)	(-3.36)
Trade		0.00504				-0.00686		
		(0.17)				(-0.29)		
Telephones			0.00753				-0.0122	
			(0.58)				(-1.05)	
GDP/capita - Penn				-0.0196				-0.0220
				(-1.00)				(-1.14)
_cons	0.779***	0.388*	0.539***	1.008***	0.858***	0.892***	0.799***	1.024***
	(4.77)	(1.87)	(3.58)	(5.13)	(13.97)	(9.16)	(12.01)	(7.46)
N	243	198	206	210	243	198	206	210
adj. R^2	0.278	0.554	0.618	0.218				
F	4.253	10.84	6.484	4.765				
r2_o					0.246	0.218	0.252	0.262
chi2					81.00	95.19	106.1	93.69

Cluster robust standard errors used; t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Differences in various education measures - dependent variable, change in Atlas Narodov Mira ELF scores

	(1) ANM change	(2) ANM change	(3) ANM change	(4) ANM change	(5) ANM change	(6) ANM change	(7) ANM change
Urbanization	-0.0476* (-1.96)	-0.0227 (-0.89)	-0.0559** (-2.17)	-0.0376 (-1.54)	-0.0359 (-1.43)	-0.0256 (-0.97)	-0.0252 (-0.96)
Immigration	0.00543* (1.91)	0.00494* (1.85)	0.00597* (1.96)	0.00603** (2.10)	0.00598** (2.02)	0.00514* (1.82)	0.00479* (1.79)
Density	-0.0111 (-0.16)	-0.000276 (-0.00)	-0.0297 (-0.40)	-0.0381 (-0.55)	-0.0354 (-0.50)	-0.00142 (-0.02)	0.0148 (0.20)
Prim. Schooling	-0.0559** (-2.35)	-0.0697*** (-2.65)					
Sec. Schooling		0.0537** (2.22)					
Tert. Schooling		0.0494 (0.62)					
Schooling total			-0.00604 (-0.49)				
Prim. Completion				-0.00285*** (-2.95)	-0.00276** (-2.58)		
Sec. Completion					0.000000303 (0.00)		
Tert. Completion					0.00246 (0.49)		
Prim. Enrollment						-0.00267** (-2.22)	-0.00380** (-2.03)
Sec. Enrollment							-0.00239 (-1.34)
Tert. Enrollment							0.000698 (0.19)
_cons	0.0703** (2.14)	0.0493 (1.54)	0.0295 (0.83)	0.0428 (1.53)	0.0391 (1.24)	0.0148 (0.54)	0.0320 (1.06)
Level var. included	yes	yes	yes	yes	yes	yes	yes
N	116	116	116	116	116	116	116
adj. R^2	0.255	0.297	0.196	0.253	0.240	0.267	0.278
F	2.799	4.275	2.216	3.001	2.690	2.907	4.731

Included level variables (Z_i): *Latitude, Altitude, Area, Agritime* and the *ANM* values in 1961
Heteroscedasticity robust standard errors used; t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Alternative time frame 1960/65-1980/85 - dependent variable, change in ANM ELF scores

	(1) ANM change	(2) ANM change	(3) ANM change	(4) ANM change	(5) ANM change	(6) ANM change	(7) ANM change	(8) ANM change
Urbanization	-0.0499**	-0.0372*	-0.0945***	-0.0666***	-0.0832***	-0.0586***	-0.0446**	-0.0446*
	(-2.52)	(-1.93)	(-3.15)	(-3.42)	(-4.11)	(-2.71)	(-2.20)	(-1.70)
Immigration	0.00532**	0.00487**	0.00951***	0.00830***	0.00970***	0.0107***	0.00492	0.00491**
	(2.26)	(2.06)	(3.37)	(3.64)	(3.34)	(4.44)	(1.38)	(2.01)
Density	-0.0327	-0.00617	0.0997*	0.0757	0.161***	0.0360	-0.0152	-0.0119
	(-0.60)	(-0.11)	(1.77)	(1.49)	(2.97)	(0.76)	(-0.33)	(-0.22)
Primary Schooling		-0.0448**	-0.0398**	-0.0403**	-0.0504***	-0.0420***	-0.0287**	-0.0438**
		(-2.58)	(-2.41)	(-2.62)	(-3.26)	(-2.66)	(-2.15)	(-2.58)
Polity IV				-0.000591				
				(-0.61)				
Conflict					0.000661			
					(0.55)			
Trade					-0.000565			
					(-0.03)			
Telephones						-0.00585		
						(-0.47)		
GDP/cap - Penn							-0.0200	
							(-0.91)	
SSA								0.0187
								(0.51)
Latin America								0.0165
								(0.55)
Asia								-0.0332
								(-1.21)
_cons	0.0162	0.0736**	0.0614	0.0646*	0.0516	0.0857*	0.0710**	0.0578
	(0.52)	(2.36)	(1.53)	(1.88)	(1.19)	(1.80)	(2.44)	(1.25)
Level var. included	yes	yes	yes	yes	yes	yes	yes	yes
N	130	116	99	94	85	90	91	116
adj. R^2	0.219	0.275	0.465	0.560	0.576	0.583	0.197	0.286
F	3.456	3.489	2.592	2.562	8.009	4.974	3.892	2.717

Included level variables (Z_i): *Latitude, Altitude, Area, Agritime* and the *ANM* values in 1961
Heteroscedasticity robust standard errors used; t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$