

Accepted Manuscript

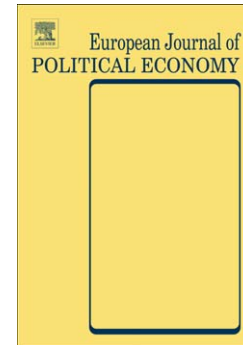
State history, historical legitimacy and modern ethnic diversity

Michael Bleaney, Arcangelo Dimico

PII: S0176-2680(16)30001-5
DOI: doi: [10.1016/j.ejpoleco.2016.03.001](https://doi.org/10.1016/j.ejpoleco.2016.03.001)
Reference: POLECO 1550

To appear in: *European Journal of Political Economy*

Received date: 21 October 2015
Revised date: 19 February 2016
Accepted date: 17 March 2016



Please cite this article as: Bleaney, Michael, Dimico, Arcangelo, State history, historical legitimacy and modern ethnic diversity, *European Journal of Political Economy* (2016), doi: [10.1016/j.ejpoleco.2016.03.001](https://doi.org/10.1016/j.ejpoleco.2016.03.001)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

State History, Historical Legitimacy and Modern Ethnic Diversity

Word Count: 7,513

Michael Bleaney^a and Arcangelo Dimico^b

^a University of Nottingham

^b Queen's University Belfast

Abstract

How much does the antiquity of states, and the sometimes arbitrary nature of colonial boundaries, explain the modern degree of ethnic diversity? It is shown that states with greater historical legitimacy (more continuity between the pre-colonial and post-colonial state) have less ethnic diversity. Historical legitimacy is more strongly correlated with ethnic diversity than are the antiquity of states, genetic diversity or the duration of human settlement. Although historical legitimacy is particularly pertinent to Africa, the correlation also holds outside Africa.

Keywords: History, Legitimacy, Ethnic Fractionalization, Development

JEL Classification: O.10

a. Michael Bleaney, University of Nottingham, School of Economics, University Park, ng72rd, email: michael.bleaney@nottingham.ac.uk

b. Arcangelo Dimico, Queen's University Belfast, 185 Stranmillis Rd, Belfast, bt7 1ht. tel. 02890974176. email:a.dimico@qub.ac.uk

1. Introduction

The effects of ethnic diversity have been widely studied both in terms of economic growth (Easterly and Levine, 1997; Alesina *et al.*, 2003; Alesina and La Ferrara, 2005; Posner, 2004; Collier and Hoeffler, 1998, Busse and Hefeker, 2007; Aisen and Veiga, 2013; Menkyna, 2014) and civil war (Fearon and Laitin, 2003; Collier and Hoeffler, 2004; Hegre and Sambanis, 2006; Wimmer *et al.* 2009; Montalvo and Reynal-Querol, 2005; Esteban and Ray, 2008). Over a period of decades (the typical time-span of such research), measures of the ethnic diversity of a state are effectively fixed, in the absence of unusual circumstances such as boundary changes, mass migrations or genocides. On a longer view, however, ethnic diversity as perceived today is not just a matter of genetic composition (Ashraf and Galor, 2013), but also of the history of statehood. The majority of the population of England defines itself today as English, but in 400 A.D., the ethnic structure of the country was very different. Exactly the same applies to Turkey. Moreover, in each of these cases the genetic composition of the population is dominated by the inhabitants who pre-dated the arrival of the invaders that gave the country its name and language; these original inhabitants are scarcely represented in the modern cultural ethnic map of these countries (Oppenheimer, 2006; Yardumian and Yardumian, 2011).

In organized states, the ruling language tends to expand at the expense of others, because of its official status and its association with power and wealth. This is most evident in the case of long-lasting empires, as can be seen from the history of China. The legacy of the Roman Empire is reflected in the predominance of Latin languages in Portugal, Spain, France and Romania, even though Latin was not spoken at all in these regions before their conquest by Rome. To cite another example, most of the population of North African countries defines itself as Arab, but none would have done so before the Arab invasions. In short, in perceived ethnic identity, history trumps genetics.

In this paper we explore further the relationship between history and modern cultural ethnic diversity. We capture the history of statehood using two measures, both of which are discussed in greater detail below: the state antiquity index of Bockstette *et al.* (2002), and an index of the state's historical legitimacy developed by Englebert (2000). We find that historical legitimacy in particular is significantly correlated with modern ethnic diversity, and these results hold even when we use an instrumental variables approach in order to overcome

possible issues in previous research.

Two papers that are closely related to ours are Ashraf and Galor (2013) and Ahlerup and Olsson (2012). Ashraf and Galor (2013) hypothesize that “genetic diversity, determined predominantly during the migration of humans out of Africa tens of thousands of years ago, is a fundamental determinant of observed ethnic and cultural heterogeneity, as reflected by the number of ethnic groups and the levels of ethnolinguistic fractionalization and polarization within national boundaries,” and they provide robust empirical evidence to support this hypothesis. Ahlerup and Olsson (2012) present a simple model of hunter-gatherer societies that predicts an increase in ethnolinguistic diversity with the length of time since original human settlement. They supplement this with a second hypothesis that state formation changes this dynamic, which they test using the Bockstette *et al.* (2002) data on state antiquity. Their empirical evidence supports these hypotheses.

Neither of these papers investigates the role of the structure of colonial empires as captured by Englebert’s measure of historical legitimacy. This is the contribution of our paper. Moreover the empirical evidence underlying the Ahlerup-Olsson model seems a good deal less impressive once it is recognized that the time since original human settlement has a correlation of 0.91 with a dummy for sub-Saharan Africa, since it took mankind a very long time to find a way across or round the surrounding seas and deserts. An alternative interpretation of their results would be not that time creates ethnic diversity, but simply that sub-Saharan Africa is different. In fact, as we show, if we omit sub-Saharan Africa from the sample, variables like the duration of human settlement and genetic diversity lose their explanatory power completely, whereas historical legitimacy does not.

The paper is structured as follows. Previous research is reviewed in Section Two. Hypotheses are presented in Section Three, and data are discussed in Section Four. The main results are presented in Section Five, with further discussion in Section Six. Conclusions are presented in Section Seven.

2. Literature Review

A significant strand of literature focuses on ethnic politics and the impossibility of enforcing a stable democratic regime when ethnic cleavages are present in the country (Horowitz, 1985;

Huntington, 1996). When voters and politicians find it efficient to build their electoral support along ethnic lines, citizens lack confidence in political institutions (Norris and Mattes, 2003). As a result, ethnic divisions lead to political systems which “*benefit few citizens at the expense of many*” (Keefer and Khemani, 2005). Empirical research shows that ethnic diversity is particularly high in Africa, and ethnic politics seems to be more crucial in African countries (Mattes and Gouws, 1999; Mattes and Piombo, 2001; Norris and Mattes, 2003). This is because of the prevalence of poor institutions across Africa, which tends to amplify the problems typically associated with fractionalization (Easterly, 2001).

It is possible, however, that there are deeper characteristics of Africa that explain both its current ethnic diversity and its weak politics and institutions. Ashraf and Galor (2013) show that genetic diversity strongly predicts contemporaneous ethnic diversity across countries. Both genetic diversity and ethnic diversity are particularly high in Africa. Michalopoulos (2012) argues that geographical factors, as reflected by variations in land quality and elevation, contributed significantly to ethnolinguistic fractionalization in Africa. Ahlerup and Olsson (2012) find that the timing of initial settlement by modern humans can explain a large fraction of existing differences in ethnic fractionalization, though they also report a strong effect of modern states (i.e. state formation after 1800) on ethnicity.

In a separate strand of research, many authors have stressed the arbitrariness of African state boundaries that were established in the colonial period (e.g. Ajala, 1983; Asiwaju, 1985; Barbour, 1961; Bello, 1995; Brownlie, 1979; Davidson, 1992; Kum, 1993; Nugent and Asiwaju, 1996; Touval, 1966). The effect of the colonial “Scramble for Africa” on current national African states has also been emphasized by Englebert (2000). Englebert argues that African states may be perceived as illegitimate because of the lack of historical continuity between the pre- and the post-colonial period. “*The variations in the extent to which post-colonial state institutions clash with pre-existing ones largely account for what differentiates state capacity and economic growth across the region. The greater the incongruence between pre- and post-colonial institutions, the greater the relative power payoffs to domestic elites of adopting neo-patrimonial policies over developmental ones*” (Englebert, 2000: 7). He uses this idea to construct an index of state legitimacy based on historical continuity (HL) which, he argues, significantly explains the poor performance of African states in terms of development. Essentially, states that have been colonized are regarded as historically legitimate only if the post-colonial state conforms closely with pre-colonial political entities.

Englebort and Tull (2008) suggest that lack of historical legitimacy also accounts for the number of failed states in Africa.

A somewhat different idea is captured by the index of state antiquity (SA) developed by Bockstette *et al.* (2002). This is a measure of how long national (non-tribal) states have been in existence (whether colonial or not). Chanda and Putterman (2007) and Putterman and Weil (2010) show that this variable is correlated with current economic development, whilst Ang (2013a, 2013b) demonstrates that it is also robustly related to institutions and financial development, which could serve as alternative mechanisms linking the effects of statehood to economic well-being. State history may matter because of cultural and institutional development over time, so that states with a longer history have better institutions, and so experience less civil conflict. In fact Historical Legitimacy and State Antiquity are less highly correlated across countries than one might expect (the correlation coefficient is 0.26), because the former focuses on the continuity between tribal and national states, while the latter only considers the period during which current national states have been in place. Of course, European states, which Englebort (2000) considers as legitimate, are also among the most antique national states in Bockstette *et al.* (2002), but the variation across former colonies between the two sources is quite different, so the two concepts are by no means identical in practice.

The importance of state formation and state continuity has long been debated in political science, mainly with regard to the process of national identity, which is considered essential for the legitimacy and sovereignty of the state. Gellner (1983) considers the coincidence of ethnic and political boundaries to be central to the development of a national sentiment and identity. On the other hand Eriksen (1993, 2002) views nationalism as a necessary process to establish the ideological justification of the state. The idea of historical continuity within the context of nationalism is discussed by Hastings (1997), who focuses on the development of an English national sentiment in Mediaeval Europe and the importance of shared languages, customs, and origins. Using a similar approach, Green (2010) argues that the pre-colonial African kingdom of Buganda constituted an authentic pre-colonial nation-state. However, while in England the coincidence of “ethnic” and political boundaries was retained through history, Buganda was arbitrarily reshuffled under colonial rule, leading to the formation of a larger state (Uganda) with split ethnic groups, which jeopardized the development of a national identity. The insistence of the Buganda on maintaining a separate political identity is

considered a major factor in the destabilization of Uganda after the country obtained independence in 1962 (Barkan, 2011).

There is also a recent large literature which exploits this division of ethnic groups to analyze the persistence of the effect of tribal institutions on development. Gennaioli and Rainer (2006) and Michalopoulos and Papaioannou (2013) document a close association between pre-colonial ethnic political centralization and measures of national and regional development. Dimico (2014), Rainer and Trebbi (2014) and Francois *et al.* (2012) look at the effect of ethnic divisions on political representation, policy, and development.

This strand of political thought stresses the importance of history in establishing a sense of community and shared identity within a state. How people define themselves in ethnic terms today is potentially an index of this, and may therefore be a by-product as much as a cause of the legitimacy and quality of its institutions and the degree of social peace that prevails. In other words the empirical correlation between ethnic diversity and the propensity to civil conflict may be less causative than is currently thought. Both may be reflections of deeper causative historical processes that shape the strength and legitimacy of political and social institutions. This is the issue under investigation in this paper.

3. Hypotheses

We first of all investigate whether ethnic diversity (ED) in the modern era reflects the antiquity (SA) and the historical legitimacy (HL) of states. Our hypothesis is that older states with greater continuity between pre-colonial and post-colonial times will have less ethnic diversity, because of the tendency for minorities to be absorbed by the ruling group.

Therefore we can summarize our hypothesis as follows:

- 1) Countries which have a long history are more likely to have gone through a state-formation process, forging a unique identity, which then reduces the degree of fractionalization of groups because people consider themselves as citizens of that state first and members of ethnic or religious groups second.
- 2) The stronger is continuity between pre- and post-colonial state history and boundaries, the more likely it is that citizens will consider the state legitimate. Countries which

have been completely reshuffled during the colonial period (as is the case with most African and some Latin American and Asian countries) are less likely to develop a national identity, and because of that they will develop kinship relationships, with people considering themselves as members of a group rather than as citizens of the state.

From an econometric point of view we can specify these hypotheses as follows. The relationship between ethnic diversity and state history can be written as:

$$ED_i = a + bSA_i + cHL_i + u_i \quad (1)$$

where ED_i denotes a measure of ethnic diversity - either ethnic fractionalization from Fearon (2003) or a measure of ethnic polarization constructed from group shares in Fearon (2003) using the formula of Montalvo and Reynal-Querol (2005)¹. The variable SA_i represents an index of State Antiquity which we collect from Bockstette *et al.* (2002), while HL_i represents a measure of Historical Legitimacy of the state according to Englebort (2000). This regression is performed for a cross-country sample using data for the year 2000.

4. Data

We use a large cross-country data set with up to 152 countries, depending on data availability, and one observation per country. In this section we describe the data; data sources are summarized in Appendix Table A1, and summary statistics are presented in Appendix Table 2.

Ethnic diversity is commonly measured by a fractionalization index, which captures the probability that two randomly chosen individuals come from different groups. Ethnic fractionalization (F) is calculated as

¹ We prefer to construct a measure of polarization using group shares in Fearon (2003) rather than using existing data on polarization in order to maintain a closer comparability between the measure of ethnic fractionalization and ethnic polarization. In fact differences between the two measures may capture differences between the distribution of groups if indices are obtained from different data on group shares.

$$F = 1 - \sum_{i=1}^n \pi_i^2 \quad (2)$$

where π_i represents the population share of group i and n is the total number of groups. This measure has a minimum of zero, when there is just one group, and a maximum of one. In practice it is strongly negatively correlated with the share of the largest group. The fractionalization data are from Fearon (2003); as a robustness check we have also used data from Alesina et al. (2003).

As an alternative we also investigate an index of polarization, which reflects the division of society into sizeable groups, since authors such as Esteban and Ray (2008) and Montalvo and Reynal-Querol (2005) have argued that ethnic polarization is particularly conducive to conflict. Montalvo and Reynal-Querol's (2005) index of ethnic polarization (P) is:

$$P = 1 - \sum_{i=1}^n 4\pi_i(0.5 - \pi_i)^2 = 4 \sum_{i=1}^n \pi_i^2(1 - \pi_i) \quad (3)$$

The polarization index also has a minimum of zero and a maximum of one. Fearon's data on group shares are used to construct a measure of ethnic polarization using this formula.

We turn now to our measures of state history. Bockstette *et al.*'s (2002) measure of State Antiquity (SA) is constructed by dividing the period from the year 0 to 1950 A.D. into 39 half-centuries, assigning a score to each half-century, and then summing these scores to a present value by applying a discount rate. The score for each half-century is obtained by multiplying together three sub-scores. The first sub-score is the answer to the question: "is there a government above the tribal level?" (score 1 if yes, 0 if no). The second sub-score depends on whether the government is foreign- or locally based (score 1 if local, 0.5 if foreign (i.e. the country is a colony), 0.75 for intermediate cases). The third sub-score reflects how much of the modern country's territory was covered by this government (score 1 if over 50%, 0.75 if between 25% and 50%, 0.5 if between 10% and 25%, and 0.3 if less than 10%). After applying the chosen discount rate, the score is divided by the maximum score across all countries to give an index with a range from zero to one. The discount rate potentially makes a big difference: with a high discount rate, the recent past will be weighted much more heavily relative to the distant past. Appendix B of Bockstette *et al.* (2002) gives

the data for various discount rates. We use a discount rate of 5% per half-century that the authors themselves favour (labelled by them *statehist05*), but as a robustness check we have also used discount rates of 1% and 10%. Note that this measure of state antiquity does not just reflect the length of time for which a state has existed, but also (through the second sub-score) a country's independence, so to that extent it might be interpreted as including elements of legitimacy. Nevertheless the most important element is the issue of whether a state exists at all, since the score is zero for any half-century for which the answer to this question is zero.

Englebert's measure of Historical Legitimacy is a pure binary variable "*that differentiates endogenous from imported statehood with the idea that political institutions which evolve endogenously to a society, as a result of domestic social relations or in opposition to neighboring societies, are presumed historically legitimate from a societal point of view*" (Englebert, 2000, p. 1827). The emphasis is on the historical basis of the present state, rather than on how recently it evolved. A state is deemed historically legitimate if it passes *any one* of the following five tests: (1) it was not colonized in modern times; (2) on reaching independence, it recovered its previous sovereignty, identity or previous existence; (3) the country was uninhabited before colonization; (4) the colonizers reduced the pre-existing societies to numerical insignificance and became citizens of the new country; and (5) the post-colonial state did not do severe violence to pre-existing political institutions. In short, countries that have not experienced colonization are classified as historically legitimate, and those that have experienced colonization can still be classified as historically legitimate, but only if there is a good deal of continuity between the pre-colonial and post-colonial state, or if the country is numerically dominated by the descendants of colonists rather than by the original inhabitants. In our sample, 64% of countries are classified as historically legitimate.

5. Ethnic Diversity and History

We begin by investigating the relationship between ethnic fractionalization and the antiquity and legitimacy of the state, and then we evaluate the same relationship for ethnic polarization. The aim of these regressions is to understand how history has influenced the ethnic diversity of states.

In Table 1 we look at the relationship between ethnic fractionalization and state history. In Model 1 we only control for the measure of historical legitimacy (HL) of Engelbert (2000), which is a binary variable. The R-squared is relatively high for a bivariate regression (0.35), and the coefficient implies that historical legitimacy is associated with a lower level of fractionalization by 35 percentage points. The average level of fractionalization for legitimate countries is 0.31, whereas the average for illegitimate countries is 0.65. In Model 2 we replace HL with the measure of State Antiquity (SA), and we also control for the duration of human settlement (HSD), so that the model is similar to that of Ahlerup and Olsson (2012)². Both variables have a significant correlation with fractionalization, although the *t*-statistic is considerably higher for HSD, and the R-squared is only slightly less than for Model 1. As Ahlerup and Olsson (2012) find, a longer duration of human settlement and later state development are associated with greater ethnic fractionalization. Since the standard deviation of SA is 0.26, a one-SD increase in SA is associated with a fall in fractionalization of $0.26 \times 0.18 = 0.05$. On the other hand the standard deviation for the duration of human settlement is 5.95, and the change in ethnic fractionalization associated with a one-SD in the duration of human settlement is almost 0.15.

In Model 3 we add HL to Model 2. The R-squared rises to 0.48, but SA loses its statistical significance, whilst HL and HSD are both significant at a one percent level. In Model 4 we add to Model 3 the proxy for genetic diversity from Ashraf and Galor (2013), but this variable does not improve the model and is not significant. In Model 5 we add a full set of geographical variables suggested by Michalopolous (2012) (crop suitability, elevation, distance from a waterway and precipitation)³ and a dummy for the Old World as additional controls. The Old World dummy is included to capture the potential effect of mass migrations from the Old World to the New World over the last 500 years (Ashraf and Galor, 2013). Inclusion of these extra controls improves the R-squared further to 0.57. The coefficient on HL in Model 5 is 0.25 (i.e. historical legitimacy is associated with 25 percentage points' less ethnic fractionalization) and remains highly significant, with a *t*-statistic of -5.49. Settlement duration and genetic diversity are both significant at a five

² Ahlerup and Olsson (2012) argue that new peripheral groups could have emerged over time, reflecting the inefficient provision of public goods from core groups. Because of that the duration of human settlement in a given location since prehistoric times could have contributed to ethnic diversity.

³ Geographical variables are obtained from Ashraf and Galor (2013) which are computed using spatial data from the FAO-GAEZ. In order to obtain a measure of crop suitability the FAO-GAEZ considers soil moisture conditions together with other climate characteristics (radiation and temperature) in order to model the timing when sufficient water is available to sustain crop growth.

percent level, but SA is not statistically significant.⁴

Table 1: Ethnic Fractionalization and State History

<i>Dependent Variable: Ethnic Fractionalization</i>					
Estimation Method: OLS	Model 1	Model 2	Model 3	Model 4	Model 5
Historical Legitimacy (HL)	-0.312*** (-8.67)		-0.251*** (-5.20)	-0.248*** (-5.00)	-0.253*** (-5.49)
State Antiquity (SA)		-0.179** (-2.28)	-0.103 (-1.33)	-0.101 (-1.29)	0.0593 (0.62)
Human Settlement Duration		0.0257*** (6.98)	0.0136*** (3.03)	0.0147*** (2.78)	0.0131** (2.43)
Genetic Diversity				-0.306 (-0.42)	1.796** (1.98)
Average Crop Suitability					-0.256*** (-3.27)
Average Elevation					0.0435 (1.16)
Distance from Waterway					0.0237 (0.64)
Average Precipitation					0.0966*** (2.64)
Old World Dummy					-0.163** (-2.41)
Constant	0.663*** (23.51)	0.396*** (8.15)	0.582*** (8.67)	0.795 (1.58)	-0.685 (-1.04)
Observations	145	128	123	123	123
R-squared	0.347	0.325	0.483	0.484	0.570

*** p<0.01, ** p<0.05, * p<0.1

Robust t statistics in parentheses

In Table 2 we estimate similar models for a measure of ethnic polarization. The fit is much poorer, with R-squareds of only 0.11 in Model 4 and 0.22 in Model 5. In these models HL is still statistically significant at a one percent level, but no other variable reaches a five percent significance level. The average level of polarization for illegitimate states exceeds that for legitimate states by only 0.10, which is only one-third of the effect for fractionalization, even though the two indices are measured on the same zero-one scale. The duration of human

⁴ A referee has raised the point that, because the dependent variable is constrained to lie between zero and one, OLS errors may not be normal. Using the method of Baum (2008) as an alternative yields similar results.

settlement and the proxy for genetic diversity even have negative coefficients. These results reflect the fact that a high degree of polarization requires at least two sizeable ethnic groups, and it is not clear why any of these explanatory variables should bring this about.

Table 2: Ethnic Polarization and State History

<i>Dependent Variable: Ethnic Polarization</i>					
Estimation Method: OLS	Model 1	Model 2	Model 3	Model 4	Model 5
Historical Legitimacy (HL)	-0.0982*** (-2.69)		-0.142*** (-3.28)	-0.133*** (-3.05)	-0.139*** (-3.11)
State Antiquity (SA)		-0.148* (-1.77)	-0.115 (-1.35)	-0.109 (-1.24)	-0.0258 (-0.25)
Human Settlement Duration		0.00110 (0.32)	-0.00556 (-1.36)	-0.00219 (-0.44)	-0.00155 (-0.30)
Genetic Diversity				-1.005 (-1.18)	-0.281 (-0.24)
Average Crop Suitability					-0.191* (-1.90)
Average Elevation					0.0553* (1.79)
Distance from Waterway					0.0241 (0.58)
Average Precipitation					-0.00890 (-0.20)
Old World Dummy					-0.148* (-1.87)
Constant	0.644*** (26.53)	0.644*** (12.61)	0.749*** (11.48)	1.449** (2.45)	1.042 (1.24)
Observations	145	128	123	123	123
R-squared	0.043	0.030	0.100	0.112	0.216

*** p<0.01, ** p<0.05, * p<0.1

Robust t statistics in parentheses

One of the concerns with previous results is that many of these variables may be a proxy for sub-Saharan Africa, because observation of African experience has played a significant role in the development of these theories. To investigate this, Table 3 shows the correlation with a dummy for sub-Saharan African countries, and with each other, for four variables: HL, SA, duration of human settlement, and genetic diversity. The proxy for the duration of human

settlement, in particular, has a very high correlation (almost 0.91) with the sub-Saharan Africa dummy. As a result it is not clear if the estimated effect of the variable in previous tables is the result of longer human settlement, as argued by Ahlerup and Olsson (2012), or some entirely different feature of sub-Saharan Africa. The other three variables also have a significant correlation with the sub-Saharan Africa dummy, but this is much lower than the proxy for the duration of human settlement (i.e. the correlation is not above 0.56).

Table 3: Pairwise Correlation

	SSAfrica	HL	SA	Settlement Duration	Genetic Div.
Sub-Saharan Africa	1.0000				
Historical Legitimacy	-0.5640	1.0000			
State Antiquity	-0.4000	0.2336	1.0000		
Settlement Duration	0.9071	-0.5392	-0.2685	1.0000	
Genetic Diversity	0.4684	-0.1485	-0.0474	0.4890	1.0000

In order to understand whether the estimated effects in previous tables are specific to African countries or not, we re-estimate Tables 1 and 2 excluding sub-Saharan African countries. Table 4 replicates results in Table 1 for this smaller sample. The coefficient of HL in Model 1 drops from 0.31 to 0.14, but the effect is still significant at a one percent level, and remains so in Models 3 to 5 with the inclusion of different controls. By contrast the three variables State Antiquity, the duration of human settlement and genetic diversity are never significant once sub-Saharan Africa is omitted from the sample.

Table 5 reports the results for ethnic polarization omitting sub-Saharan Africa, which overall resemble results for ethnic fractionalization in Table 4. HL is the only variable which retains a five percent significance level across models. The only notable difference compared to Table 4 is a significant effect of SA in Model 2, Model 3, and Model 4. However the variable turns to be insignificant once a full set of geographical controls is entered in the model (Model 5). We have also estimated these models using the ethnic composition data of Alesina et al. (2003), with similar results (not shown).

In the light of these results the idea that it is simply the passage of time since human settlement that creates ethnic diversity does not seem particularly robust. If variables largely

capture the fact that Africa represents the origins of humanity, it is not possible to disentangle the effects of these variables from the simple fact that Africa is different.

Table 4: Ethnic Fractionalization and State History Excluding Sub-Saharan Africa

<i>Dependent Variable: Ethnic Fractionalization</i>					
Estimation Method: OLS	Model 1	Model 2	Model 3	Model 4	Model 5
Historical Legitimacy (HL)	-0.144*** (-2.85)		-0.167*** (-3.07)	-0.153*** (-2.91)	-0.149*** (-2.67)
State Antiquity (SA)		-0.181 (-1.63)	-0.139 (-1.29)	-0.125 (-1.12)	0.0169 (0.11)
Human Settlement Duration		0.0155 (1.19)	0.00718 (0.60)	0.00899 (0.77)	-0.00257 (-0.19)
Genetic Diversity				-1.013 (-1.36)	0.754 (0.62)
Average Crop Suitability					-0.200** (-2.25)
Average Elevation					0.0945** (2.21)
Distance from Waterway					0.0455 (0.89)
Average Precipitation					0.0813 (1.59)
Old World Dummy					-0.0974 (-1.05)
Constant	0.491*** (10.94)	0.419*** (8.24)	0.549*** (8.00)	1.251** (2.40)	-0.0479 (-0.05)
Observations	103	91	86	86	86
R-squared	0.076	0.036	0.141	0.158	0.322

*** p<0.01, ** p<0.05, * p<0.1

Robust t statistics in parentheses

Table 5: Ethnic Polarization and State History Excluding Sub-Saharan Africa

<i>Dependent Variable: Ethnic Polarization</i>					
Estimation Method: OLS	Model 1	Model 2	Model 3	Model 4	Model 5
Historical Legitimacy (HL)	-0.129** (-2.29)		-0.168*** (-2.91)	-0.159*** (-2.69)	-0.158** (-2.60)
State Antiquity (SA)		-0.277** (-2.03)	-0.237* (-1.79)	-0.228* (-1.67)	-0.0737 (-0.36)
Human Settlement Duration		0.00812 (0.56)	-0.00115 (-0.08)	0.00002 (0.00)	-0.00951 (-0.56)
Genetic Diversity				-0.657 (-0.70)	1.108 (0.73)
Average Crop Suitability					-0.232* (-1.92)
Average Elevation					0.0955** (2.43)
Distance from Waterway					0.0322 (0.57)
Average Precipitation					0.0671 (1.05)
Old World Dummy					-0.125 (-1.08)
Constant	0.672*** (13.97)	0.692*** (11.22)	0.823*** (10.10)	1.278* (1.96)	0.0201 (0.02)
Observations	103	91	86	86	86
R-squared	0.043	0.054	0.132	0.137	0.265

*** p<0.01, ** p<0.05, * p<0.1

Robust t statistics in parentheses

The relationship between ethnic polarization and fractionalization is complex. In relatively homogeneous societies with levels of fractionalization below 0.3, which covers approximately one-third of countries, indices of fractionalization and polarization are both very low and, in fact, virtually proportional to one another. At the other extreme, when fractionalization is high, they are negatively correlated. This raises the issue whether our models truly explain polarization at all; if they do not, the results should change radically if we omit countries with a fractionalization index below 0.3, in which range polarization is essentially a proxy for fractionalization. If we re-estimate Model 1 of Tables 1 and 2 for this reduced sample, the results are radically different, as shown in Table 6. For countries with a

fractionalization index above 0.3, the correlation between fractionalization and HL is still significantly negative, but the correlation between ethnic polarization and HL is now significantly *positive*. This suggests that HL is not, in general, associated with a low degree of ethnic polarization, except in cases where polarization is a proxy for fractionalization, as is the case in highly homogeneous countries. This may be because large minorities, whose existence the polarization measure is designed to capture, are more resistant to absorption into the ruling group.

Table 6: Ethnic Diversity and Historical Legitimacy if Fractionalization > 0.3

<i>Estimation Method: OLS</i>	<i>Ethnic Polarization</i>	<i>Ethnic Fractionalization</i>
Historical Legitimacy (HL)	0.076*** (2.92)	-0.184*** (-6.45)
Constant	0.673*** (33.05)	0.703*** (32.04)
Observations	100	100
R-squared	0.08	0.29

*** p<0.01, ** p<0.05, * p<0.1

t statistics in parentheses

6. Instrumental Variables and Causality

Of course one of the potential problems with OLS results is the fact that there may be omitted variables which can bias effects. For this reason in this section we re-estimate models similar to those above using instrumental variables (IV). As an instrument for Historical Legitimacy we use the number of years for which the country was colonized, because only colonized countries can experience a lack of historical continuity between pre- and post-colonization, and this lack of continuity is likely to be more severe for countries which have suffered a longer colonial history.

We use the elapsed time since the transition to agriculture from Chanda and Putterman (2007) as an instrument for State Antiquity, given that the transition to agriculture fostered property

rights and state formation (Bowles and Choi, 2013). Diamond (1997) argues that a prerequisite of a sustained process of institutional development was biogeographical potential, in the form of the availability of plants and animals suitable for domestication, which stimulated the transition to agriculture. Olsson (2005) documents a strong relationship between a measure of biogeography diversity and institutional development, which alone explains almost 45 percent of the variation in current institutions. After reviewing a series of possible determinants of institutions, he argues that the most plausible candidates for explaining the broadest cross-continental variance in institutional quality are those focusing on historical differences in biogeographical potential for early agriculture and on the importance of disease geography for European colonization strategy.

Table 7 reports IV estimates for ethnic fractionalization. Model 1 includes a full set of controls except the duration of human settlement. Model 2 also includes the latter variable. The first-stage statistics confirm a significant effect of the proxy for colonial history on HL and an insignificant effect on SA. The first-stage F-statistics for SA and HL are 11.55 and 14.79 respectively. The number of years under colonization has a significant and positive effect on HL and an insignificant effect on SA. On the other hand the transition to agriculture has opposite effects (significant for SA but not for HL). Therefore the two instruments adequately isolate the variation in each of the two endogenous regressors without causing problems of non-orthogonality between the explained variation and each of the endogenous variables in the model. Weak identification is also rejected, given that the partial first-stage F-statistics are either much larger than 10 (in Model 1) or quite close to 10 (in Model 2). Results from the second-stage regression confirm a significant effect of HL on ethnic diversity and an insignificant effect of SA. The estimated coefficient from the IV is almost twice the one estimated using an OLS estimator, indicating an increase in the level of ethnic heterogeneity by almost 50 percent for countries lacking historical continuity between the pre- and post-colonial period. The proxy for genetic diversity is significant at a 1 percent level as well while SA is not statistically significant.

Table 8 reports IV estimates for ethnic polarization, and the results in Table 2 are largely confirmed. Apart from HL no other variables are statistically significant at even a 10 percent level. However the estimated effect of HL on polarization is much smaller than on fractionalization.

Table 7: Ethnic Fractionalization and State History – IV

<i>Ethnic Fractionalization</i>		
Estimation Method: IV	Model 1	Model 2
Historical Legitimacy (HL)	-0.507*** (-4.40)	-0.495*** (-2.66)
State Antiquity (SA)	0.178 (0.75)	0.205 (0.73)
Genetic Diversity	3.264** (2.51)	3.259** (2.53)
Average Crop Suitability	-0.184** (-1.99)	-0.187* (-1.88)
Average Elevation	0.0678* (1.65)	0.0638 (1.25)
Distance from Waterway	0.0151 (0.32)	0.0184 (0.33)
Average Precipitation	0.0972** (1.98)	0.0989* (1.81)
Old World Dummy	-0.237** (-2.10)	-0.249* (-1.93)
Human Settlement Duration		0.00177 (0.13)
Constant	-1.555 (-1.60)	-1.572 (-1.61)
Anderson Canon. LR Stat.	13.828	5.968
Observations	123	123
R-squared	0.432	0.444
<i>First Stage Stats. Model 1</i>		
Dependent Variable	SA	HL
Years Since the Transition to Agric.	0.239*** (5.33)	0.142 (1.24)
Years as a Colony	-0.008 (-0.23)	-0.346*** (-4.23)
First Stage Partial F-statistics	11.55	14.79
<i>First Stage Stats. Model 2</i>		
Dependent Variable	SA	HL
Years Since the Transition to Agric.	0.212*** (3.93)	-0.0853 (-0.63)
Years as a Colony	0.006 (0.13)	-0.230*** (-2.91)
First Stage Partial F-Statistics	4.44	8.03

*** p<0.01, ** p<0.05, * p<0.1, Robust t statistics in parentheses

Table 8: Ethnic Polarization and State History - IV

<i>Ethnic Polarization</i>		
Estimation Method: IV	Model 1	Model 2
Historical Legitimacy (HL)	-0.315** (-2.40)	-0.384* (-1.83)
State Antiquity (SA)	0.430 (1.46)	0.266 (0.85)
Genetic Diversity	1.712 (0.91)	1.743 (0.95)
Average Crop Suitability	-0.145 (-1.44)	-0.127 (-1.12)
Average Elevation	0.0405 (1.04)	0.0651 (1.26)
Distance from Waterway	0.0571 (1.02)	0.0371 (0.56)
Average Precipitation	0.0231 (0.39)	0.0125 (0.19)
Old World Dummy	-0.370** (-2.35)	-0.296* (-1.93)
Human Settlement Duration		-0.0107 (-0.76)
Constant	-0.397 (-0.29)	-0.299 (-0.22)
Anderson Canon. LR Stat.	13.828	5.968
First Stage F-Stat.	14.79	8.03
Observations	123	123
R-squared	0.035	0.028
<i>First Stage Stats. Model 1</i>		
Dependent Variable	SA	HL
Years Since the Transition to Agric.	0.239*** -5.33	0.142 (1.24)
Years as a Colony	-0.008 (-0.23)	-0.346*** (-4.23)
First Stage Partial F-statistics	11.55	14.79
<i>First Stage Stats. Model 2</i>		
Dependent Variable	SA	HL
Years Since the Transition to Agric.	0.212*** (3.93)	-0.0853 (-0.63)
Years as a Colony	0.006 (0.13)	-0.230*** (-2.91)
First Stage Partial F-Statistics	4.44	8.03

*** p<0.01, ** p<0.05, * p<0.1, Robust t statistics in parentheses

7. Conclusions

National identity has always been considered an important element of the process of consolidation of the state, because it provides the common ground necessary for positive inter-group relations and interaction between citizens from different backgrounds. This process normally takes time, and because of that it is expected that states which have a longer history have had more chance to develop those shared values, which are normally considered important in terms of legitimacy. In some cases, the state formation process has been shattered by the imposition of arbitrary institutions or boundaries, which have caused a discontinuity between the pre-existing society and current institutional arrangements. As previous authors have identified, this problem is particularly severe for African countries.

We have shown that state history and historical legitimacy are both important determinants of ethnic diversity, although legitimacy is considerably more significant than the antiquity of the state. Ethnic fractionalization is noticeably higher in less historically legitimate and, to a lesser extent, younger states. Ethnic polarization is much more weakly associated with these same factors. This may reflect the fact that large minorities are much more resistant to absorption into the ruling group than small ones, and so their numbers have less of a tendency to be eroded over time.

Historical legitimacy does not only help to explain why ethnic diversity is high in sub-Saharan Africa; it also has significant explanatory power outside the region, which shows that it is more than just a substitute for an unidentified SSA effect, as may be the case for some other variables that have been used, such as the duration of human settlement.

References

- Ahlerup, Pelle, and Ola Olsson (2012). "The Roots of Ethnic Diversity", *Journal of Economic Growth*, 17(2): 71-102.
- Aisen, A. and F. J. Veiga (2013). "How Does Political Instability Affect Economic Growth", *European Journal of Political Economy*, Volume 29, Pages 151-167
- Ajala, Adekunle (1983). "The Nature of African Boundaries;" *Afrika Spectrum*, 18, 177-188

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003). "Fractionalization," *Journal of Economic Growth*, 8, 155-194

Alesina, Alberto, and Eliana La Ferrara (2005) "Preferences for Redistribution in the Land of Opportunities", *Journal of Public Economics*, 89 (5-6), 897-931

Ang, James. B., (2013) "Institutions and the Long-Run Impact of Early Development" *Journal of Development Economics*, 105, 1-18

Ang, James B., (2013) "Are Modern Financial Systems Shaped by State Antiquity?," *Journal of Banking and Finance*, vol. 37(11), pages 4038-4058.

Arbatli, C.E, Ashraf, Q., and Galor, O., (2013). "The Nature of Civil Conflict," Working Papers 2013-15, Brown University

Ashraf, Q. and O. Galor. (2013). "Genetic Diversity and the Origins of Ethnic Fragmentation," *American Economic Review P&P*, 103, 528-533.

Asiwaju, A. I. (1985). "The Conceptual Framework," in A. I. Asiwaju (Ed.), *Partitioned Africans* (pp. 1-18). New York: St. Martin's.

Barbour, K. M. (1961). *A Geographical Analysis of Boundaries in Inter-Tropical Africa*. In K. M. Barbour & R. M. Prothero (Eds.), *Essays on African population* (pp. 303-323). London: Routledge Kegan Paul

Barkan, Joel D. (2011). "Uganda: Assessing Risks to Stability", *Center for Strategic & International Studies*

Baum, C.F. (2008). "Modelling Proportions", *The Stata Journal* 8, 299-303.

Bello, Adebayo (1995). "The Boundaries Must Change," *West Africa*, p. 546

Bockstette, Valerie, Areendam Chanda, and Louis Putterman (2002). "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, 7(4): 347-69

Bowles, S., and J.K. Choi (2013). "Coevolution of farming and private property during the early Holocene", *Proceedings of the National Academy of Science of the United States of America (PNAS)*, vol. 110(22): 8830-8835

Brownlie, Ian (1979). "African Boundaries: A Legal and Diplomatic Encyclopedia," London: C. Hurst

Busse, M and C. Hefeker (2007), "Political Risk, Institutions and Foreign Direct Investment", *European Journal of Political Economy*, Volume 23, Issue 2, Pages 397-415

Chanda, Areendam, and Louis Putterman (2007). "Early Starts, Reversals and Catch-up in the Process of Economic Development," *Scandinavian Journal of Economics* 109, 387-413

Collier, Paul, and Anke Hoeffler (1998). "On Economic Causes of Civil War," *Oxford*

Economic Papers, 50(4), pages 563-73

Collier, Paul, and Anke Hoeffler (2004). "Greed and grievance in civil war," *Oxford Economic Papers*, 56(4), pages 563-595

Collier, P., A. Hoeffler and D. Rohner, 2009. "Beyond Greed and Grievance: Feasibility and Civil War," *Oxford Economic Papers* 61(1):1-27.

Davidson, Basil (1992). *The Black Man's Burden: Africa and the Curse of the Nation-State*, New York: Times Book

Dimico, A. (2014). "Size Matters: The Effect of the Size of Ethnic Groups on Development," *QUCEH WP*, 14-02

Easterly, William, and Levine, Ross (1997). "Africa's Growth Tragedy: Policies and Ethnic Divisions," *The Quarterly Journal of Economics*, vol. 112(4): 1203-50

Easterly, W. (2001), 'Can institutions resolve ethnic conflict?', *Economic Development and Cultural Change*, **49**: 687-706

Englebert, P. (2000b). Solving the Mystery of the AFRICA dummy, *World Development*, 28(10): 1821-1835

Englebert, Pierre and Denis M. Tull (2008). "Post-Conflict Reconstruction in Africa: Flawed Ideas About Failed States," *International Security* 32: 106-39

Eriksen T. H. (1993) *Ethnicity and Nationalism: Anthropological Perspective*, London: Pluto Press

Eriksen, T.H. (2001) *Small Places, Large Issues: An Introduction to Social and Cultural Anthropology*, London: Pluto Press

Esteban, Joan, and Debraj Ray (2008). "Polarization, Fractionalization and Conflict," *Journal of Peace Research*, 45,163-182.

Fearon, James D (2003): "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, Springer, vol. 8(2), pages 195-222

Fearon, James, and David D. Laitin (2003). "Ethnicity, Insurgency, and Civil War". *American Political Science Review* 97(1): pp.75-90

Francois, Patrick, Ilia Rainer and Francesco Trebbi (2012): How Is Power Shared In Africa? *NBER Working Paper* No. 18425

Gellner, Ernest (1983). *Nations and Nationalism*. Oxford: Blackwell

Gennaioli, N. and I. Rainer (2006): *Precolonial Centralization and Institutional Quality in Africa, in Institutions and Norms in Economic Development*, by M. Gradstein and K. Conrad, (Eds.) MIT Press

- Green, E.D. (2010). Ethnicity and Nationhood in Precolonial Africa: the Case of Buganda, *Nationalism and Ethnic Politics*, 16, 1-21
- Gurr, Ted (1970). *Why Men Rebel*, Princeton University Press
- Hastings, Adrian (1997). *The Construction of Nationhood*. Cambridge: Cambridge University Press.
- Hegre, Håvard, and Nicholas Sambanis (2006). "Sensitivity Analysis of Empirical Results on Civil War Onset," *Journal of Conflict Resolution*, 50:508-535
- Horowitz, Donald L. (1985) *Ethnic Groups in Conflict*, University of California Press
- Huntington, S. P. (1991) *The Third Wave: Democratization in the Late Twentieth Century*, University of Oklahoma Press, 1991
- Keefer, P., and Khemani, S. (2005). "Democracy, Public Expenditures, and the Poor: Understanding Political Incentives for Providing Public Services," *World Bank Research Observer*, 20 (1), 1-27
- Kum, Joseph M. (1993). "The Central African Subregion," in *Disarmament: Workshop on the Role of Border Problems in African Peace and Security* (pp. 49-71). New York: United Nations
- Mattes, Robert B., and Amanda Gouws (1999). "Race, Ethnicity and Voting Behavior: Lessons from South Africa," In *Electoral Systems and Conflict in Divided Societies*, Ed. Andrew Reynolds and Ben Reilly. Washington, DC: National Academy Press.
- Mattes, Robert B., and Jessica Piombo (2001). "Opposition parties and the voters in South Africa's General Election of 1999," *Democratization* 8(3): 101-12
- Menkyna, F., T. (2014). "A Theory of Ethnic Diversity and Income Distribution: A Legislative Bargaining Approach", *European Journal of Political Economy*, Volume 34, Pages 52–67
- Michalopoulos, S. (2012). "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 102, 1508-1539.
- Michalopoulos, S. and E. Papaioannou (2013): Pre-colonial Ethnic Institutions and Contemporary African Development, *Econometrica*, 2013, 81(1): 113152.
- Montalvo, José G., and Marta Reynal-Querol (2005). "Ethnic Polarization, Potential Conflict, and Civil Wars." *American Economic Review*, 95(3): 796–816
- Montalvo, José G., and Marta Reynal-Querol (2010). "Ethnic Polarization and the Duration of Civil Wars," *Economics of Governance* 11: 123-43
- Norris, Pippa, and Robert Mattes (2003). "Does Ethnicity Determine Support for the Governing Party?," *Afrobarometer Paper* no. 26

Nugent, Paul (1996). "Arbitrary Lines and the People's Minds: A Dissenting View on Colonial Boundaries in West Africa," in Paul Nugent & A. I. Asiwaju (Eds.), *African Boundaries* (pp. 35-67). London: Pinter

Olsson, Ola (2005). Geography and Institutions: Plausible and Implausible Linkages *Journal of Economics* Volume 10, 2005, pp 167-194

Oppenheimer, Stephen (2006) *The Origins of the British: a Genetic Detective Story*, London: Constable and Robinson.

Posner, Daniel (2004), "Measuring Ethnic Fractionalization in Africa," *American Journal of Political Science*, 48: 849–863

Posner, Daniel (2005) *Institutions and Ethnic Politics in Africa*, New York: Cambridge University Press

Putterman, Louis, and David Weil (2008). "Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality," *NBER Working Paper* no. 14448

Rainer, Ilia and Francesco Trebbi (2014): *New Tools for the Analysis of Political Power in Africa, in African Successes: Government and Institutions*, edited by Sebastian Edwards, Simon Johnson, and David N. Weil; University of Chicago Press

Singer, David J & Melvin Small (1982) *Resort to Arms: International and Civil Wars, 1816–1980*. Beverly Hills, CA: Sage.

Singer, David J & Melvin Small (1994) *Correlates of War Project: International and Civil War Data, 1816–1992: A Statistical Handbook*. New York: John Wiley.

Touval, Saadia (1969). "The Sources of Status Quo and Irredentist Policies," in Carl G. Widstrand (Ed.), *African Boundary Problems* (pp. 101-118). Uppsala, Sweden: Scandinavian Institute of African Studies

Wimmer, A., L.-E. Cederman and B. Min (2009). Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set, *American Sociological Review*, 74, 316-337

Yardumian, A. and T.K. Yardumian (2011) "Who Are the Anatolian Turks?" *Anthropology and Archaeology of Eurasian* 50: 6-42

*Appendix Table A1:**Variables and Data Sources*

Variable	Source
Ethnic Fractionalization	Fearon (2003)
Ethnic Polarization	Computed using group shares from Fearon (2003) to maintain a comparability between group distributions for EF and EP
Historical Legitimacy	Englebert (2000)
State Antiquity (SA)	Bockstette et al. (2002)
Settlement Duration (ORIGTIME)	Ashraf and Galor (2013)
Genetic Diversity	Ashraf and Galor (2013)
Average Crop Suitability	Ashraf and Galor (2013)
Average Elevation	Ashraf and Galor (2013)
Distance from Waterway	Ashraf and Galor (2013)
Average Precipitation	Ashraf and Galor (2013)
Old World Dummy	Ashraf and Galor (2013)

Appendix Table A2: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Ethnic Fractionalization	152	.4756996	.2566556	.001999	.952575
Ethnic Polarization	154	.5894079	.2298821	.0049919	.9856
Historical Legitimacy (HL)	157	.611465	.488977	0	1
State Antiquity (SA)	139	.4597747	.2503054	.0217776	1
Human Settlement Duration	139	6.257194	4.946311	.12	16
Genetic Diversity	139	.7270694	.0272777	.6278865	.7743011
Average Crop Suitability	139	.3869487	.243879	.0033928	.9512565
Average Elevation	139	.6113424	.5610907	.0005218	2.836526
Distance from Waterway	139	.3682802	.4662995	.0204568	2.38558
Average Precipitation	139	.8604824	.5872799	.0291064	2.599521
Old World Dummy	139	.8201439	.3854566	0	1
Religious Fractionalization	152	.3805033	.2164094	0	.7828

Highlights

- States with greater historical legitimacy have less ethnic diversity
- Historical legitimacy is more strongly correlated with ethnic diversity than State Antiquity
- The correlation also holds outside sub-Saharan Africa.
- Other factors lose their explanatory power if we omit sub-Saharan Africa