

**LAND, HISTORY OR MODERNIZATION?: EXPLAINING
ETHNIC FRACTIONALIZATION¹**

ERIC KAUFMANN

Department of Politics, Birkbeck College, University of London

Malet Street, London, United Kingdom WC1E 7HX

e.kaufmann@bbk.ac.uk

¹ The author acknowledges the support of British Academy grant # SG091163 which made this research possible. I would like to thank Anthony D. Smith, Stelios Michalopoulos and Michael Hechter for their feedback and assistance. I am also grateful for feedback on an earlier version of the paper presented at the Intrastate Conflict Seminar at the Kennedy School of Government, Harvard, 2008-9 and American Political Science Association conference, Seattle, 2011.

Jared Diamond argues that geoclimactic factors endowed Europe, the Middle East and much of Asia with a head start in the transition to agricultural civilization, the effects of which continue to be felt today (Diamond 1997). Related geographic arguments are used to explain the divergent economic performance of individual states. These explanations focus on states' distance to the sea and transport costs to markets; disease environments and their effect on agricultural productivity; and the stock of readily accessible industrial resources such as coal and oil. Sub-Saharan Africa tends to fare worst, and Western Europe best, on most of these measures (Gallup, Sachs and Mellinger 1999; Braudel 1972).

Technology has levelled many geographic barriers to growth, but geography's social and political traces may be harder to dislodge. For instance, a second-order variant of the geoclimactic thesis holds that ecological diversity within a state produces elevated ethnic diversity (Michalopoulos, forthcoming). This in turn contributes to lower levels of economic development (Easterly and Levine 1997; Alesina et al. 1999) and, in the opinion of some, an enhanced risk of violent conflict (Montalvo and Reynal-Querol 2005; Urdal 2008). Clearly much is at stake when it comes to the relatively neglected question of what explains international variation in ethnic fractionalization.

Geoclimactic diversity offers an important but contested account of why some countries are more ethnically mixed than others. A prominent critique of Diamond's work (Acemoglu and Robinson 2012) claims that the wealth and poverty of nations hinges on the openness of their political institutions. This is unrelated to geography: the product of long-term institutional drift interacting with rare critical junctures. On this view, the performance of today's economies is rooted in path-dependent historical processes reaching back into premodernity. Historical accident, not geography, holds the key to explaining world economic rankings. The economic underperformance of Sub-Saharan Africa, for instance, is traceable to the destabilizing effects of the Slave Trade. Beginning in the sixteenth century in

West Africa, and even earlier in the East, slavery established extractive economic practices and exclusive political institutions. Colonialism continued the pattern while post-colonial rulers, despite their rhetoric, tuned in to the incentives laid down by existing structures. Critical junctures, often premodern, set processes of institutional drift in train that shape economic performance, but also set the ethnic morphology that in turn affects economic outcomes (Acemoglu and Robinson 2012: 327).

Set against geographic and historical institutionalist interpretations are those of modernization theorists. These writers consider the modern period to represent a sharp break with the past (Giddens 1991). The shift to an industrial mode of production, bureaucratization, conscription, mass political participation and a literate public sphere are held to radically alter the rhythms of daily life. In the process, local identities are transcended and citizens inducted into the mass culture of the nation. Geographic and premodern historical endowments pale beside indicators of the intensity of modernization in explaining outcomes. Sub-state identities emerge as responses to modernity: peripheral political entrepreneurs mobilise countermovements to those of the state (Tönnies and Loomis 1940; Gellner 1983; Breuilly 1993). Ethnic fractionalization is therefore related less to longstanding institutions or primordial geography than to variations in the intensity of modernity and the prevalence of political opportunities. This paper tests the geoclimactic, historical institutionalist and modernist accounts of the sources of ethnic fractionalization.

Ethnic Fractionalization

All contemporary questions of nationalism and ethnic conflict begin with the imperfect overlap between ethnic/national communities and political units. Ethnic fractionalization indices provide a quantitative measure of one manifestation of this: the

degree of ethnic pluralism contained under the political roof of each of the world's states. The ethno-linguistic fractionalization index, or ELF, measures the likelihood that any two random individuals in a state's population are members of the same ethnic group. The greater the number of ethnic groups and the more even their relative size, the more fractionalized the population¹. More recent formulations attempt to modify the original 1964 ELF measure by considering the size of the largest and second largest ethnic groups (Fearon and Laitin 2003: 84). Critically, cultural diversity in the form of language, as measured by Ethnologue, for instance, must be distinguished from ethnic diversity, which is based on self-identity measures. Ethnic identity in turn is not coterminous with politically-relevant ethnicity: in some societies, notably in sub-Saharan Africa, ethnicity has an 'onion'-like character, with several different levels, only the highest of which may be politically important (Posner 2005). Similarly, in North America, 'white' is now politically relevant in a way Croatian or German is not. Jewish and Mormon, however, remain politically relevant groups despite their size. This paper considers a wide range of ethnic fractionalization measures, seeking to evaluate which factors are most closely associated with different forms of ethnic fractionalization across the world's states.

Incumbent upon the pathbreaking work of Easterly and Levine (1997), Alesina et al. (1999) and Fearon and Laitin (2003), an extensive literature now exists on the relationship between ethnic fractionalization and political and economic outcomes such as economic growth (Easterly et al. 2006), public goods provision (Banerjee and Somanathan 2007) and violent conflict. In terms of violent conflict, the literature is divided. Studies which take conflict *onset* as the independent variable tend to find no relationship with ethnic fractionalization (i.e. Sambanis 2001; Fearon and Laitin 2003; Collier and Hoeffler 2004;

¹ Initial research on ELF used data from a 1960s Soviet ethnographic atlas (Bruk, S. I. and V. S. Apenchenko, eds. 1964. Atlas narodov mira. Moscow: Glavnoe upravlenie geodezii i kartografii gosudarstvennogo geologicheskogo komiteta SSSR and Institut etnografii im. H. H. Miklukho-Maklaia, Akademiia nauk SSSR.)

Schneider and Wiesehomeier 2009). Those that focus on the *incidence* of civil war, by contrast, typically report a significant association (i.e. Ellingsen 2000; Montalvo and Reynal-Querol 2005; Urdal 2008).

Few have turned the question around to ask why some states are more fractionalized than others. Such questions are of more than intrinsic interest. They are also important because it is vital to understand the upstream determinants of the fractionalization which may be producing malign political and economic effects. Moreover, identifying fixed or slow-changing correlates of ethnic fractionalization enables scholars to deploy these measures as bedrock independent or instrumental variables in their analyses. On the other hand, if ethnic fractionalization is at the mercy of faster-moving political and economic modernization factors, then it should properly be viewed as endogenous to modern economic and political change - and therefore less important.

Theories of Ethnicity and Nationalism

The geoclimactic-historical-modernist triptych has its analogue in the three main families of nationalism theory: primordialism, ethnosymbolism and modernism (Özirimili 2010). Primordialism locates ethnicity in universal aspects of human psychology. For primordialists, our evolutionary psychology represents the successful adaptation of humanity to conditions obtaining during prehistoric time. Primary among these is geography. Geoclimactic isolation of people produces cultural and genetic drift. Genetic distance creates divergent genetic interests which are the basis for group boundaries and conflict. Genetics are invisible because neighbouring groups rarely look different, thus culture is activated as a marker of genetic boundaries when groups leave their ecologies and come into contact with others (Salter 2001; Van den Berghe 2002; Pinker 2011: 353-55).

In effect, ethnic boundaries based on genetic and cultural difference preserve the variations incubated in ecological niches. In this manner, geoclimactic explanations for variations in ethnic fractionalization are most closely related to primordialism. An alternative 'geoculturalist' interpretation would claim that geography shapes ethnogenesis via cultural diversity and its usefulness as a group marker, but that this diversity is subsequently amenable to being eroded and reshaped by social processes (Cavalli-Sforza 2001). This second formulation would predict that while traces of geography's effects may remain in ethnic fractionalization patterns, the two will be more loosely connected than if primordialist assumptions hold.

Critics of primordialism suggest that kin-selection impulses are deflected toward constructed forms of community like teams, religious groups or political nations. They also reject the primordialist position that ethnicity can exist in small groups, averring that by definition, ethnic communities must involve a larger scale of human community. Bonds therefore need to be culturally imagined rather than merely experienced in the form of face-to-face *gemeinschaft* relationships. The ethnosymbolist school, for example, concurs with primordialists that ethnic groups predate the modern era, but emphasizes the importance of political and cultural institutions over geography. Ethnosymbolists claim that ethnic groups do not arise until the late neolithic period when writing, religion, recorded history and extra-local mobilization allowed for the formation of communities knit together by 'imagined' bonds of territory, memory and ancestry (Smith 1986: 44-5). Often ethnic consciousness remained the preserve of a small elite, as with the Anglo-Saxon English consciousness of the Venerable Bede and King Alfred (Hastings 1997: 35-9). Some ethnic groups (i.e. Jews, Amhara, Armenians, Persians) had ancient origins, while many more emerged in the medieval and early modern periods through tribal confederation (i.e. Arabs, Kurds), conquest agglomeration (i.e. Gothic founders of Spain) or dynastic competition (i.e. Scots, Catalans).

In all cases, territorial identities extending beyond the locale came to be established (Smith 1986; Armstrong 1982).

There are two major forms of ethnicity, according to Francis' (1976: 6) schema: *primary* ethnicity, in which members of the group occupy their ancestral 'homeland' territory; and *secondary* ethnicity, whereby groups acknowledge that they are diaspora and not native, and thus their homeland lies elsewhere. Since immigrants form just 2.7 percent of the world's population, it is primary ethnic fractionalization which is generally captured by ELF and is by far the most important form of fractionalization when it comes to economic development and conflict (Demeny and McNicoll 2006, ch.1). Most of the premodern entities studied by ethnosymbolists are primary ethnic groups, even if they sometimes spawn secondary offshoots like the Jewish, Parsee and Armenian diasporas.

The appearance of translations of religious texts such as the bible into vernacular languages, and the numerous recorded premodern references to *natio*, *gens* and *ethnos* is cited in favour of the theory (Hastings 1997). This explains why historians of the medieval period tend to be ethnosymbolists (Zimmer and Scales 2005). This argument predicts that ethnic identities, once formed, are highly path-dependent and durable. Ethnic sentiments are reproduced by both state and vernacular institutions. The vernacular rootedness of ethnicity means that it is capable of inspiring collective action and resisting 'official' political and identity constructs imposed by the state. Though more culturalist than materialist in orientation, ethnosymbolist theory nests most comfortably within historical institutionalism, evincing greater skepticism of explanations tied to geoclimactic or modernizing factors.

The *modernist* account contests the ethnosymbolist and primordialist version of events. It argues that premodern identities were strictly local – for the peasant masses, or imperial-religious for military and religious elites (Anderson [1983] 1991; Gellner 1983;

Giddens [1985] 1996). Modernity fractures the horizontal ties between cosmopolitan elites, as with Latin Christendom, which fragmented into nation-states with their own vernacular languages. Beneath them, locals were 'invited into history', and came to be connected to wider, self-conscious territorial communities (Nairn 1977). Print capitalism, mass conscription, mass education, secularization and more intensive transport networks combine to orient local identities toward a common, this-worldly community. Pre-existing cultures are orthogonal to this process, which is driven by political and economic imperatives. 'Nationalism, which sometimes takes preexisting cultures and turns them into nations, sometimes invents them, and often obliterates preexisting cultures: that is a reality,' writes Eric Hobsbawm. 'In short, for the purposes of analysis nationalism comes before nations. Nations do not make states and nationalisms but the other way round' (Hobsbawm 1990: 10).

Beyond the majority ethnic identity - created by the new nation-state - shared ethnic identities are forged in modern times through states' internal administrative boundary marking activity. This reflexive demarcation institutionalizes ethnic diversity, as with Soviet Republics or colonial administrative departments. Divide-and-rule policies by imperial rulers and missionaries contribute to the process (Trevor-Roper 1983; Brass 1991; Brubaker 1996; Wimmer 2002). Anti-state mobilization by political entrepreneurs, often driven by the imperative to control important industrial resources such as oil, is another vector of ethnogenesis. Ethnic entrepreneurs may have experienced blocked upward mobility within central state structures (Gellner 1983), or may use ethnic and national movements as a vehicle to acquire more power or wealth than they might through conventional political channels (Breuilly 1993). Modern processes, not geography and premodern history, are responsible for spawning diversity.

Existing Studies of Ethnic Fractionalization

Work on ethnic diversity as a dependent variable is in its early stages. The best developed line of inquiry concerns geoclimactic predictors. Broadly speaking, these writers find that the wider the diversity in land quality and topography in a territory, the greater the ethnic fractionalization (Michalopoulos, forthcoming; Sutherland 2003; Collard and Foley 2002; Moore et al. 2002; Nettle 1996; Nichols 1992). Ahlerup and Olsson (2011) add that an early incidence of initial prehistoric human settlement, together with geo-climactic factors, predicts fractionalization. States far from mankind's East African origins, such as Sweden, were settled later than equatorial regions, and hence possess less ethnic diversity. Laitin and Robinson (2011) also advance a geoclimactic argument, applying Jared Diamond's continental axis theory to individual states. They uncover compelling evidence that linguistic diversity is greater in states characterized by a North-South cartographic skew as compared with those which spread in a more East-West direction. This is conceived as stemming from the greater difficulty which premodern agricultural states experienced when attempting to expand through what Diamond posits to be the steeper ecological isolines of the North-South axis. Techniques adapted to one latitude work less well in other latitudes. Trans-longitudinal adaptation is much less of a problem.

Historical institutional factors feature in work with the State Antiquity dataset (Bockstette et al. 2002). This uncovers a significant (at $p < .05$ level) negative association between ELF and the date of initial state formation coupled with the degree of indigenous control of the state in the ensuing period. The logic is that older states, and those where the indigenous population had greater political control, could spread their culture and identity and are therefore less diverse than newer states - or those ruled by foreigners. However the connection between state history and ethnic fractionalization is a bivariate finding since this relationship was not the primary focus of the authors. Fletcher and Iyigun (Fletcher and

Iyigun 2009) formulate an alternate historical argument, claiming that a higher incidence of Muslim-Christian conflict between 1400 and 1900 predicts lower ethnic fractionalization in today's European and Middle Eastern states while those which experienced Protestant-Catholic conflict or anti-Jewish pogroms are more fractionalized. Nunn (2008: 164), also working from a historical institutional perspective, considers the role of the slave trade in producing ethnolinguistic diversity in Africa. The internal tribe-on-tribe raiding that characterized African slave economies is linked to weaker precolonial states and, by extension, more ethnic fractionalization (Acemoglu and Robinson 2012: 87).

Modernist approaches to this subject are in their infancy. Green (2011), using Philip Roeder's 1961 and 1985 datasets (Roeder 2001), contends that urbanization in postcolonial states in Africa during 1961-85 is associated with declining levels of ethnic diversity. This echoes qualitative work which notes the presence of ethnic fusion in modernizing locations such as the colonial Zambian Copperbelt settlements of the 1930s where miners from formerly distinct groups amalgamated into larger ethnic entities based on cultural relatedness. They even invented new rituals to mark themselves off from other ethnic groups. These constructs were transmitted back to ethnic hinterlands and led to a second-order enlargement of ethnic consciousness in the countryside (Eriksen 1993: 20-21). The case has also been made for regionally-fractured European groups like the Ukrainians and Italians, who often developed an awareness of themselves as larger entities only after settling together in ethnic neighbourhoods of American cities in the late nineteenth and early twentieth centuries (i.e. Thomas and Znaniecki 1918). This paper breaks new ground by introducing new data on historical institutional and international factors. It integrates this within a comprehensive approach that incorporates geoclimactic diversity, historical indicators, modernization and international determinants. Finally, expanding beyond previous work, it encompasses linguistic, identity-based and politically-relevant ethnic fractionalization measures.

Geoclimactic Variation and Ethnic Fractionalization

The three major theories make different predictions regarding the relationship between geography and ethnic fractionalization. Primordialists view extreme ecological diversity - as in the New Guinea case - as diversity-enhancing. Yet for ethnosymbolists, extreme isolation produces sub-ethnic localism, impeding imagined community. This localism produces disorganization, reducing resistance to the modern state when it eventually penetrates the periphery, and renders the task of nation-building easier than might be the case in a situation where larger ethnic groups have mobilized. However, above a certain threshold - perhaps several thousand in population - an ethnosymbolist would grant that difficult terrain acts to increase the number of competing premodern polities. Rival ethnic identities can take root so long as there are literate intellectuals and institutions that help spread myths, symbols and memories beyond the local. Terrain that permits this mobilization while preventing wider integration is optimal in producing ethnic fragmentation. Thus more challenging terrain would be expected to lead to greater ethnic heterogeneity, but less so than in the primordialist case.

For modernists, varied terrain acts as a barrier to the state in its quest to homogenize populations and facilitates secessionists' strategy of escaping to peripheral redoubts from which they can invent their interest-reinforcing ethnic projects. On the other hand, variegated terrain hampers the networks of coordination (Laitin 2007) which incentivize participants to join an imagined community. Though ecological variation is associated with greater ethnic fractionalization in all three theories, this relationship would be expected to be stronger under assumptions of primordialism than for competing theories. We can test for this by examining

the relationship between a country's geoclimactic diversity and its ethnic diversity. Thus our first hypothesis:

H₁: Geoclimactic variation is associated with ethnic fractionalization. The stronger the relationship, the greater the support for primordialist theories

Historical Institutions and Ethnic Fractionalization

According to ethnosymbolist theory, modern nations typically form around premodern dominant ethnic groups (Smith 1986). In most cases, the dominant ethnic group is also largest because popular sovereignty and democratization spread their influence down the social scale and render exclusive dominant minorities like Syria's Alawis rare (Kaufmann and Haklai 2008). Therefore we may approach the question of ethnic absorption through measures of ethnic and state antiquity. The State Antiquity dataset asks when a polity above the tribal level was founded on the territory of an existing state; whether this was indigenously-controlled or foreign; and further, what proportion of the territory of the present-day state was under native rule. This is determined for every 50-year period since 1 A.D. Different rates of discounting past periods are applied by the authors, with the most common measure being a 5 percent discount every 50 years (Putterman 2007).

An alternative approach is to attempt to code the founding date of the largest ethnic group in a state. This serves as a measure of ethnic absorption because older dominant ethnic groups will have had more time in which to assimilate neighbouring or subaltern groups than newer groups. Dominant ethnic groups frequently emerge as assimilationist actors - fractionalization-reducing nuclei - within multi-ethnic states. Connor (1994a: 96) notes that

homogeneous nation-states occur in less than 10 percent of the world, but that a substantial majority of states contain an ethnic majority. All but five of 156 countries in Vanhanen's (1999) dataset feature a plurality group of a third or more of the population. In other words, some form of ethnic dominance appears to be nearly universal (Kaufmann 2004).

Modernists would be somewhat more circumspect. They would explain ethnic homogeneity as a result of nation-building and the ethnic exclusions practiced by modern states (Wimmer 2002). States established earlier in the modern period would be expected to contain less ethnic diversity than more recent states, but the age of premodern *ethnic* groups should bear no relation to contemporary fractionalization after controlling for the age of the modern state.

How to measure the founding date of the largest ethnic group? This is nowhere near as straightforward as the founding date of states. In this paper, ethnic founding dates are operationalized as the *first imagining of the group by a putative member of the group*. The Ethnic Plurality Group Founding Dates dataset has been developed through a survey of historians and social scientists with expertise on particular countries. Accepting the 'reality' of these ethnic founding dates does not entail embracing the ethnosymbolist perspective. Premodern ethnic imaginings are explained by modernists as the musings of individuals - with no consequences for mass social and political behaviour. Primordialists, too, would consider these visions to be subsidiary to spontaneous collective nepotism in the process of ethnic fusion.

Therefore, an ethnosymbolist would expect that:

H₂: States with plurality ethnic groups that are comparatively old will have lower degrees of ethnic fractionalization; and

H₃: States that have an older tradition of indigenous control will have lower degrees of ethnic fractionalization

Modernists would qualify this as follows:

H₄: Older modern states (defined as post-1789 phenomena) will have lower degrees of ethnic fractionalization, but pre-1789 ethnic plurality founding dates or state antiquity scores should not affect ethnic fractionalization net of the age of the modern state

Primordialism grounds its claims in the ethnic substratum. It is skeptical of states' ability to fuse ethnic groups together, even in the long run. It treats ethnicity as being of prehistoric provenance. Therefore neither ethnic nor state age should affect fractionalization measures.

Modernists locate the source of ethnic fragmentation in modern political and economic factors. States with an abundance of exportable primary commodities such as oil are more likely to experience rent-seeking and ethnic entrepreneurialism and consequently will be more ethnically fractionalized than others. States experiencing political instability in the form of a transition between autocratic and democratic governance are more likely to offer opportunities to political entrepreneurs. Democracy, however, once attained, should

permit dissent to be expressed through the state rather than via extra-statal insurgent movements, lowering fractionalization. Finally, urbanization and income per capita are indicators of the intensity of modernity in a state. Thus modernists would predict that:

H₅: States with higher levels of urbanization, democracy and income per capita should be less ethnically fractionalized; and

H₆: States whose economy is based on exportable primary commodities and/or those undergoing political instability should be more ethnically fractionalized

Data

Secondary data sources are listed in Appendix 1. Data on the founding date of the largest ethnic group in each state has been collected through our British Academy-funded survey of experts, supplemented with textual sources. Methodology, questionnaire and detailed response data for the Ethnic Plurality Group Founding Dates dataset may be found at: <http://www.sneps.net/ethnic/ethnicdates.htm>. We were able to compile a set of dates for 129 of the 156 countries listed in the Vanhanen dataset. Note that this data does not provide an iron-clad answer to the question of how effectively an ethnic identity has been diffused down the social scale and out to neighbouring populations. This is due to the fact that there is a considerable distance to travel between the first imagining of a group by a putative insider and the realization of mass ethnic consciousness. Walker Connor (1994b), for instance, argues that a majority of ethnic group members must be conscious of a common identity in order for a group to be deemed in existence.

In some cases, the elapsed time between first elite imagining and collective action is brief. In others, it is protracted. For example, kingdoms which formed the ultimate origin of the largest ethnic groups in the Democratic Republic of Congo, Nigeria, Sierra Leone and several other West African states date to the 14th and 15th centuries. Yet these were hierarchical entities destabilized by the slave trade. It is unclear whether these foundations endow their successor ethnic groups with much of an assimilatory advantage over their twentieth century equivalents in neighbouring Liberia or Cameroon.

Another possibility is that even if their early kingdoms achieved little mass consciousness, those with a time-hallowed pedigree (such as the Wolof of Senegal or Kikuyu of Kenya, both successful assimilators) can point to these antecedents as a form of cultural capital which endows them with a patina of prestige (Johnson 2004). Prestige attracts power and adherents from other groups, increasing the group's share of the population (Deutsch 1966 (1953): 120). In conjunction with ethnic date and indigenous control of territory, measures of continuity of consciousness and speed of social diffusion would be required to more accurately predict the size of the largest ethnic group. These are avenues for future research.

Finally, we also code for states which emerged from successful ethnic national self-determination movements as well as states with longstanding or current dominant minorities (see Appendix 4) and refined the data through discussion with Anthony Smith, a leading expert on the history and sociology of ethnonationalist movements.

Results

We are restricted to cross-sectional models due to the absence of time-series data on ethnic fractionalization. Many independent variables are also unavailable over time. Most datasets ignore smaller island states, reducing the universe of cases. Finally, gaps in the data result in a small degree of listwise deletion. The number of cases (109 to 146 depending on the specification) limits our degrees of freedom, thus the need for a multi-model approach which examines a wide array of variables and successively tests the predictive power of competing theories before bringing the most promising parameters together in a combined analysis.

Geoclimactic Model

Results of a series of OLS regressions of geoclimactic predictors on leading measures of ethnic fractionalization appear in table 1.

[Table 1 here]

We find strong confirmation of hypothesis H_1 , which is associated with a neo-Diamondian interpretation of ethnic difference. Rougher terrain, in the form of standard deviation of mean elevation, is significant (at $p < .05$) in all but one model. For every measure of ethnic fractionalization at least one geoclimactic variable reaches significance. In most cases, several do. The power of geoclimactic measures is clearly strongest when it comes to predicting linguistic diversity and weakens as dependent variables begin to reflect political considerations. This is evident in the discrepancy between the R^2 of .630 attained in the

linguistic diversity (NMBRLANG) model, and the .450 to .570 achieved in other models. Geoclimactic coefficients are weaker when regressed on coefficients for politically-relevant ethnic diversity (ELFPREG) and the size of the largest politically-relevant ethnic group (MAXPOP) than for other measures.

Politically-relevant ethnic diversity measures automatically exclude states where ethnicity is deemed irrelevant to politics. The 22 states excluded in this way tend to be highly homogeneous, such as Denmark, or, in a few instances such as Tanzania, extremely heterogeneous. Since the average excluded state has an 81 percent ethnic majority as against 67 percent for included states, this weakens the predictive power of parameters which distinguish highly homogeneous states from moderately diverse ones. Furthermore, politically-relevant ethnic group measures represent an aggregation of ethnic groups into politically-relevant entities on the national stage which in some cases may be considered supra-ethnic rather than ethnic. Thus it taps processes of supra-ethnic amalgamation as well as ethnic diversity. Even so, one or both of ruggedness of terrain and mean agricultural suitability of land reached significance in models of politically-relevant ethnic fractionalization. Other geoclimactic correlates were often in the expected direction even if, in the relatively small universe of world states, they failed to attain significance. Our control variables for state land area and three world regions reached significance in at least some models. Sub-Saharan Africa is the most consistently powerful parameter in the dataset: it is positively and significantly associated with all measures of ethnic fractionalization in virtually every specification though is noticeably less effective in predicting linguistic diversity. Part of this, as we shall find, relates to the legacy of slavery in the region.

Historical Institutional Model

Our second and third hypotheses, H₂ and H₃, predict that an older tradition of indigenous control of the state and an older plurality ethnic group are significantly associated with less ethnic fractionalization. Both relationships are confirmed in bivariate tests, where the two measures (state antiquity and ethnic origin date) attain significance across most ethnic fractionalization measures.² Importantly, these two historical institutional predictors are not associated with linguistic diversity, and associations with politically-relevant ethnic fractionalization are weaker than for ELF. Moreover, state antiquity and origin date effects wash out when other parameters are added to the model in table 2. Two other historicist predictors, the number of millennia since the transition to agriculture (Diamond's dependent variable), and population density in 1500, reach significance ($p < .05$) in some models. Transition to agriculture is the more important of the two, and even where insignificant at the $p < .05$ level, its sign is in the expected direction. More striking is the predictive power of slave exports as a proportion of historic African states' populations. It is significant in all but the linguistic and politically-relevant fractionalization models.

[Table 2 here]

However, the historical institutional story is not yet complete. Figure 1 plots the relationship between Fearon and Laitin's ethnic fractionalization measure and the date when the largest ethnic group in each of the world's states was first established (see <http://www.sneps.net/ethnic/ethnicdates.htm> for details).

² State antiquity predicts 11 percent and ethnic founding date 5 percent of the variation in EF. Both are significant at $p < .001$ level in bivariate analysis.

[Figure 1 here]

The six ancient data points from before the birth of Christ skew the results: if removed, ethnic group founding date becomes significant in five of the seven multivariate analyses in table 2 (see Appendix 3). This relationship is robust to different time subsets within the post-0 A.D. data. This makes sense insofar as ethnic groups formed from city-states or land empires in the ancient period have experienced more discontinuity than the dynastic kingdoms and tribal confederations which arose in the post-Roman era. It strengthens somewhat when the 11 cases which contain dominant minorities are dropped (specification not shown in this paper). This follows because politically-dominant minorities tend to impede the assimilation efforts of the largest ethnic groups in their state.

The post-Classical period turns out to be especially important for another reason: the relationship between the age of the largest ethnic group and its share of a state's population is not linear but categorical.³ It turns out that the 25 plurality ethnic groups founded between 450 and 1099 A.D. are exceptionally homogeneous and explain much of the power of the ethnic origin date variable (see figure 2).

[Figure 2 here]

³ Tests for a curvilinear relationship showed no improvement in the predictive power of ethnic group founding date.

The most likely explanation for this pattern is that the period from 450-1099 A.D. was one in which some of the earliest continuous 'ethnic states' (Smith 1986), with elite myths of descent and cultural codes, were formed. It encompasses many states which occupy lands captured during the Sunni Arab conquests of the 7th-11th centuries. In addition, a number of durable East Asian kingdoms arose at this time and many West European states emerged out of the Germanic barbarian successor dynasties which replaced the western Roman Empire.

Compared to ethnic groups founded during this period, those established later inhabit states which are much less homogeneous. This helps explain the strong relationship between date and ethnic fractionalization noted for the post-Classical period. When we apply a dummy variable for these early Medieval ethnic groups (Origin 450-1099 A.D.), this turns out to be highly significant in all models in table 2, including those for linguistic diversity and politically-relevant ethnic fractionalization.

[Table 2 here]

Modernist Models

Models focusing on modern political and economic variables are reported in table 3. Standard errors in this model are biased downward, and predictive power upward, because of the highly plausible reverse causation effect by which ethnic fractionalization impedes modernization. Given the relatively modest correlations in this model and the theoretical aim of this paper, the variation is not partitioned through the use of instrumental variables, though this remains a fertile area for future work.

[Table 3 here]

At first glance, the modernization hypothesis H_4 appears to be refuted, because the age of the modern state is not significant (at $p < .05$) in any specifications. Later we shall see that this variable, in combination with geoclimactic and historical predictors, is in fact associated with fractionalization and only in this context can we properly assess H_4 , i.e. its predictive power as compared with ethnic group founding date.

Parameters that test H_5 - urbanization, income per capita and democracy - are modestly associated with certain measures of ethnic fractionalization. The lack of a strong urbanization effect is particularly surprising. When it comes to H_6 , the picture is more mixed: oil exporting countries are significantly more fractionalized in four models, supporting rational-choice and neo-Marxist variants of modernist arguments. The proportion of primary commodity exports as a share of GDP (not included in table 1) led to considerable listwise deletion and this term fell out of larger models but was significant in the expected direction in several specifications (see Appendix 3). Political instability was insignificant in all models.

Though there is a case to be made that more fractionalized societies rely more on oil exports, this is less convincing than the reverse causation argument for H_5 - whereby low ethnic fractionalization is as much a cause as an effect of modernization. All told, modernist models explain a similar proportion of the variation in ethnic fractionalization as historical institutional models, and less than geoclimactic models. In addition, model fit is likely to be biased upward owing to the effects of reverse causation.

International Determinants of Ethno-Political Fit

Ethnic fractionalization may owe as much to the way politics carves up the globe as it does to the diversity beneath it that is produced by geography, history and modernity. Fractionalization is produced by the overlay of political fragmentation atop ethnolinguistic variation. There are currently close to 200 states superimposed on as many as 6000 language groups producing a variegated ethno-linguistic fractionalization grid as we move from state to state. Fractionalization tends to decrease as ethnicity and politics move toward congruence. This can occur for three major reasons: a) ethnic nuclei absorb smaller groups or attract members away from them; b) the number of political units increases; or c) ethnicity and politics come into closer alignment through ethnic secession and partition.

We have focused mainly on a) above, because this has the most to contribute to our understanding of theory. The absence of data on changes in fractionalization over time prevents us from assessing proposition b). Yet completeness demands a consideration of c), namely international determinants of ethno-political congruence. The degree of ethnic self-determination of state boundaries varies from state to state. It issues from either 'top-down' variation in the ethno-cartographic versimilitude of Great Power boundary demarcation, or a 'bottom-up' variable: the nature and extent of ethno-national self-determination. Green alludes to the 'top-down' aspect, citing variation in colonial practices in Africa, suggesting that the large size of states and the arbitrariness of political boundaries produced high levels of fractionalization:

In particular, Englebert et al. (2002, p. 1096) have noted that some 44% of colonial borders were straight lines, leading to as many as 177 ethnic groups split across two and sometimes three colonial borders; when added together these partitioned ethnic groups represented 43% of the average African state's population (Green 2011).

What of bottom-up processes? The more a movement for national self-determination defines itself in ethnic terms (i.e. 'ethnic' Irish nationalism vs. 'civic' Eritrean nationalism) and the greater a movement's degree of efficacy in securing its ethnic territory, the lower will be its ELF.

However, even ethnonational movements and their Great Power sponsors are greedy for territory and cannot resist ruling others. Ulster Protestants and Lebanese Christians, in consultation with their Great Power guarantors, aimed at about a two-thirds majority in the new states (or provinces) created for them in 1921 and 1926. This satisfied these dominant groups' desire to be in a comfortable majority but also slaked their ambition to control significant swathes of minority-dominated territory (Mansergh 1978: 38; Zisser 2000: 6-7). In other cases, 'winning' nationalizing states, like Romania after WWI, blatantly sought out the most territory they could hold (Brubaker 1996; Zimmer 2003b).

On the other hand, ethnonational movements that fail to achieve independence inflate their state's ethnic fractionalization index. Post-World War I Romania was too successful, but the Kurds too weak, to lower ELF. An ethnic group must be strong *vis á vis* the powers-that-be, but not too strong, to achieve low ELF. Most irridentisms, for example, would increase ethnic diversity (Saideman and Ayres 2008). Conversely, ex-empires emerge from collapse, and multiethnic states from secession, as more homogeneous. The partitioning of Hungary

and Ireland, for instance, or the dismemberment of the Turkish, Soviet, Austro-Hungarian and Japanese empires greatly increased rump states' homogeneity (Brubaker 1998).

The relative balance between ethnonationalists' degree of self-determination, their success at territorial aggrandizement and the ethno-political accuracy of others' determination of boundaries governs the congruence of borders. To tap ethnonational 'fit', this paper offers several indicators. State formation tends to occur in distinct bursts which punctuate long periods of relative calm. Waves of state formation are coded around key historical junctures as follows: 1) pre-1815; 2) 1816-1918; 3) 1919-45; 4) 1946-90; 5) 1991-present. Roughly speaking, these correspond to 1) state nations of the western core; 2) secessions from European land empires; 3) self-governing white dominions; 4) ex-colonies; and 5) post-Cold War secessions.

This posits a connection between the period of state formation and the degree of ethno-national self-determination. This is not a continuous or ordinal variable. In particular, states forming through decolonization (virtually all new states formed during 1946-90) should produce the highest fractionalization scores since ethnic self-determination is least evident. This produces:

H₄: The wave in which a state was created should be associated with ethnic fractionalization, with former European overseas colonies being most fractionalized;

and

H₅: States formed on the basis of nationalist movements which define themselves in explicitly ethnic terms should have lower levels of ethnic fractionalization than other states

The results of table 4 broadly confirm both hypotheses. Ex-colonies are more fractionalized than others (pre-1815 wave states are the reference category) in four of seven models and states formed from successful ethnonational secessions (see Appendix 4) are less fractionalized in five of seven models. In both cases, these parameters, where not significant at the $p < .05$ level, were signed in the expected direction.

In addition we test for the impact of secessions and dominant minorities. States which have experienced secession are expected to have lower fractionalization than others while states with dominant minorities should have higher fractionalization. Interestingly, neither of these hypotheses are borne out in the analysis in table 4.

Combined Model

Drawing on the most promising regressors from all four models while including GDP per capita as a control yields the results in the combined model in table 5. This model predicts over 60 percent of the variation in the four identity-based ethnic fractionalization measures. It performs less well when the dependent variable is pure linguistic diversity or politically-relevant ethnic fractionalization, but still captures half or more of the variation. Four geoclimactic predictors attain significance in at least three models. Higher and drier countries have less diversity. Those with greater variation in elevation and soil productivity are more fractionalized. The geoclimactic model in table 1 outperforms the combined model when it comes to predicting linguistic diversity. However, as one progressively moves from

language to identity to politically-organized groups, the power of a strictly geoclimactic model weakens. In our two models of politically-relevant ethnic fractionalization (ELFPREG and MAXPOP), no geoclimactic predictors remain significant.

One should not overplay the importance of this finding. Of non-geographic predictors, only oil output per capita and the sub-Saharan Africa dummy are significantly associated with politically-relevant ethnic fractionalization. Moreover, variation in elevation and soil type approach significance and relationships are signed in the expected direction in these models. One must bear in mind, however, that homogeneous countries are deemed not 'relevant' in ethnic terms on this measure and supra-ethnic mobilizing dynamics (rather than mere ethnic diversity) play a key part in the politically-relevant ethnicity measures.

The combined model allows a test of H₄. Namely, is the founding date of the largest ethnic group or that of the state more central in predicting the four identity-based fractionalization measures (ELF, EF, PLURALITY, PCTMAJ)? In table 4, states whose largest ethnic group was founded between 450 and 1099 A.D. are significantly less fractionalized in all four identity-based models while state founding date is significant in three. On the other hand, state date is a more powerful predictor than ethnic date in two of three models. Replacing the 450-1099 A.D. dummy with a continuous variable for date weakens the predictive power of ethnic group founding date unless we exclude the six ancient cases from before the Birth of Christ (see Appendix 3 for alternative specification with ethnic group founding date instead of 450-1099 A.D. origin).

Against this, it could be argued that reverse causation is possible for state founding date insofar as more ethnically fractionalized territories develop into modern states later. This was confirmed in table 4, where we found that states formed in post-1815 waves, notably ex-colonies, are more fractionalized than others. That said, including an ex-colony dummy did

not alter the power of state date in the combined model. Overall, the results are inconclusive when it comes to arbitrating between ethnosymbolist and modernist theory.

Oil output per capita is an important predictor in four of seven models, supporting 'greed'- based modernist arguments based on ethnic entrepreneurialism (Collier and Hoeffler 1994). Modernization variables offer a mixed picture: democracy enters just one of seven models and GDP per capita none. Infant mortality rate (not shown) did not approach significance in any specifications. However the robustness of state founding date across the main ethnic fractionalization models offers modest support for modernist approaches. Qualified support is also provided for historical institutionalist approaches. The sub-Saharan Africa dummy, the only variable to perform well across all models, partitions fairly evenly into effects related to the volume of historic slave exports per capita and regional effects unrelated to the legacy of slavery.⁴ Along with ethnic group founding date effects, this also offers modest support to historical institutional approaches.

Other variables, not readily assignable to the three major theories, figure prominently in the combined model. Population density in 1995, which springs from geoclimactic, historical institutional and modernist sources, is significant in five of seven models, though its sign changes when the dependent variable moves from linguistic to identity-based measures of fractionalization. In general, denser populations are associated with less ethnic fractionalization. Table 2 shows that this effect persists with a control for population density in 1500 indicating that more recent variation in population growth may underpin this correlation. Larger territories are expected to contain more groups, and this is borne out in the data: land area is significant in three models and correctly signed in all seven.

⁴ Rerunning the four identity-based models (EF, ELF, PLURAL, PCTMAJ) with an interaction term for sub-Saharan African slave exports results in the sub-Saharan Africa dummy and sub-Saharan African slave exports each falling out of two models. Both remain signed in the expected direction across all models. Sub-Saharan African slave exports was not included in the combined model due to issues of degrees of freedom, but this specification is shown in appendix 3.

Importantly, a term capturing whether a country is an ex-European colony did not reach significance in the combined model. This questions the general wisdom that colonization, by running roughshod over ethnic boundaries, is primarily responsible for sub-Saharan Africa's high ethnic fractionalization. It seems the institutional inheritance of slavery plays a more important role in this development, as Nunn (2008) and Acemoglu and Robinson (2012) surmise.

Finally, states which emerged on the back of nationalist movements defined in ethnic terms are, as expected, more homogeneous: ethnic nationalism reaches significance in two of seven models, though it is signed in the expected direction in all. This intimates that successful ethnic nationalists tend not to achieve their homogeneous utopias, probably because most bear few qualms about controlling territory populated by other groups.

Conclusion

This paper attempts to explain variation in ethnic fractionalization between countries at the end of the twentieth century. This is important because ethnic fractionalization has been connected to a series of malign political and economic outcomes. In addition, measures of ethnic fractionalization are frequently deployed in econometric and political analysis. This article brings existing sources together with newly-collected data on plurality ethnic group founding dates and international factors to provide the first comprehensive account of worldwide variation in ethnic fractionalization. It offers a quantitative test of competing theories of ethnicity and nationalism.

The article likewise assesses geoclimactic, historical institutional and modernist theories of socio-political development. Established work on the importance of geographic

variation receives substantial support. States with a greater difference between their highest and lowest points are significantly more diverse than others. The same holds for those with a wider range of soil types. Higher and drier countries are less fractionalized than low, wet ones. Overall, geoclimactic variables are more strongly associated with ethnic fractionalization than historical institutionalist and modernist predictors. This speaks to the importance of interpretations which hold that ecological diversity lays the basis for linguistic and ethnic diversity, as exemplified by the case of Papua New Guinea.

This neo-Diamondian thesis must be qualified, however, by noting that geoclimactic variables are less strongly associated with politically-relevant ethnic diversity. This is an important corrective to the existing literature on geographic determinants of ethnic fractionalization and suggests that there are limits to the efficacy of purely geoclimactic paradigms in explaining politico-economic variation among states.

Historical institutional legacies have an important bearing on ethnic fractionalization, though somewhat less so than geography. Ethnic group founding date is a significant inverse predictor of ethnic diversity because older groups have had longer to assimilate neighbours. Moreover, their patina of age confers prestige. This variable is not linear, however: plurality ethnic groups formed in the period between 450 and 1099 A.D. are associated with highly homogeneous states while ancient groups and those formed after 1099 are located in more diverse states. It is postulated that ethnic groups of the immediate post-Classical era were better able to maintain a continuity of consciousness into the modern era than those dating from the ancient period. Meanwhile groups emerging after 1099 have had less time to assimilate neighbours into their culture and myth of ancestry than those formed in the seven centuries before. In terms of historical factors, this study also corroborates the claims of Nunn (2008) and Acemoglu and Robinson (2012) that African states with a history of slave exports

are more fractionalized than other countries. There is also modest evidence that the date of transition to agriculture helps explain variation in ethnic fractionalization.

Most modern economic and political variables are not associated with ethnic fractionalization. Cross-national differences in urbanization, health and income had little or no predictive power. Of all modernization variables, only the founding date of the modern state proved robustly associated with ethnic fractionalization: its effect is similar to that of ethnic group founding date which suggests that there is merit in both ethnosymbolist and modernist approaches to nationalism. Modernist explanations based on the competition for lootable resources receive some support from this study. Oil exports per capita are strongly associated with ethnic fractionalization, and one of the few consistently significant predictors of politically-relevant ethnic diversity. Per capita commodity exports is significantly associated with some fractionalization measures in a few models. On the other hand, political instability fails to predict fractionalization in any model.

There is a less theoretically-classifiable input into fractionalization from international factors. Ethnic diversity tends to decrease when state and ethnic boundaries converge, and vice-versa. This can occur as a) the number of political units increases and/or b) ethnicity and politics come into closer alignment through secession and partition. Tests for the effects of new state creation are limited by the lack of a time dimension in global ethnic census and fractionalization data. Yet states which experienced secessions are no more homogeneous than others in the data. This may be because more fractionalized states such as Russia/USSR and Serbia/Yugoslavia are more apt to experience secession such that the two effects negate each other. Furthermore, there is strong evidence that states formed through ethnically-defined national self-determination movements are more homogeneous than more 'civic' states originating on the basis of ideology or realpolitik.

Finally, the historic era in which states were created is important. Those which formed prior to the Congress of Vienna in 1815 are most homogeneous, while those emerging during the period of decolonization are most diverse. Having said this, ex-colonies are not significantly more fractionalized than other countries once geographic, historical institutional and modernist factors are introduced into the model. This questions the conventional wisdom that colonial borders condemned African states to ethnic fractionalization. Instead, these results suggest that variegated and low-lying terrain, a history of slavery and the presence of lootable resources better explains the ethnic diversity of sub-Saharan Africa. Indirect effects of colonization, such as the size and recent vintage of Africa's states, are a further contributing factor.

The most striking finding of this paper is that ethnic diversity is rooted in the geography, climate and history of a country. This renders it largely exogenous to modern political and economic change, and therefore a useful parameter in political and economic models. These results question the strong variant of constructionist theory which claims that ethnolinguistic diversity may be created *ex nihilo*. What this research suggests is that ethnic entrepreneurs can politicize pre-existing linguistic divisions or activate previously latent ethnic identities, but will have difficulty creating language and ethnicity anew. The Ijaw movement in Nigeria in the late 1990s, for example, represents the emergence of a new politically-relevant ethnic group in response to the glaring inequalities and pollution generated by local oil resources (Osaghae 2008). However, the Ijaw have been constructed on the basis of subgroups which Ethnologue classifies as speaking a related language.

Linguistic invention for political reasons is not impossible: Bosnian, Croat and Serb variants of Serbo-Croat have only recently been developed. Yet, the principal source of ethnolinguistic difference is geographical and historical. Ethnogenesis requires a plausibility structure and a degree of popular resonance which limits the scope for invention (Zimmer

2003a: 174). Hence nakedly political attempts at ethnogenesis such as the Padanian movement in Northern Italy, the Cruithin-Pictish interpretation of Northern Ireland Protestant origins or the Arab myth among Trinidadian Muslim Indians have proven conspicuous failures (Kaufmann 2008).

Modernization counts for more when it comes to reducing ethnic fractionalization. The founding date of the state is significant in several specifications, suggesting that national integration is an important solvent of ethnic bonds. It is also vital to appreciate that this is an analysis of variation between countries rather than time points. The static nature of the dependent variable biases the data against faster-moving predictors, hence these results do not negate the importance of modernizing processes in reducing diversity over time. Predictors of variation between countries at one point in time are often different from those which predict variation within countries over time (Smith 1995; Kittel 2001: 233).

'Every 14 days a language dies,' claims the National Geographic's *Enduring Voices Project* team. 'By 2100, more than half of the more than 7,000 languages spoken on Earth—many of them not yet recorded—may disappear.'⁵ Qualitative evidence that modernization is driving this decline, or, similarly, that secession reduced ethnic fractionalization in Russia/USSR after 1989, can only be uncovered using time-series data. The global coverage of the large-scale Demographic and Health Survey (DHS), while imperfect, will eventually furnish a global database to conduct time-series ethnic fractionalization research. Innovative historical approaches, such as those utilizing sub-state census data for one or more countries (Urdal 2008) or providing ELF data at two time points (Roeder 2011; Green, 2011) offer new pathways toward understanding the temporal aspects of this phenomenon.

⁵ See the National Geographic's Enduring Voices project website at: <http://travel.nationalgeographic.com/travel/enduring-voices/>.

References

- Acemoglu, Daron, and James Robinson. 2012. *Why Nations Fail: the Origins of Power, Prosperity and Poverty*. London: Profile Books.
- Ahlerup, Pelle, and Ola Olsson. 2011. "The Roots of Ethnic Diversity." Stockholm Institute of Transition Economics.
- Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114 (November):1243–84.
- Anderson, Benedict. [1983] 1991. *Imagined Communities*. London: Verso.
- Armstrong, John. 1982. *Nations Before Nationalism*. Chapel Hill: University of North Carolina Press 201-3.
- Banerjee, A., and R. Somanathan. 2007. "The political economy of public goods: Some evidence from India." *Journal of Development Economics* 82 (2):287-314.
- Bockstette, V., A. Chanda, and L. Putterman. 2002. "States and markets: The advantage of an early start." *Journal of Economic Growth* 7 (4):347-69.
- Brass, P. 1991. "Ethnic Groups and Ethnic Identity Formation." In *Ethnicity and Nationalism*, ed. Brass. New Delhi; Newbury Park, CA: Sage Publications.
- Breuilly, J. 1993. *Nationalism and the State*. Manchester: Manchester University Press.
- Brubaker, R. 1996. *Nationalism Reframed: Nationhood and the National Question in the New Europe*. Cambridge: Cambridge University Press.
- Brubaker, R. 1998. "Migrations of ethnic unmixing in the "New Europe"." *International Migration Review* 32 (4):1047-65.
- Cavalli-Sforza, L. L. 2001. *Genes, Peoples and Languages*. Stanford, CA: University of California Press.
- CIESIN 2012. Center for International Earth Science Information Network (CIESIN), Columbia University. Low Elevation Coastal Zone (LECZ) Urban-Rural Estimates, Global Rural-Urban Mapping Project (GRUMP), Alpha Version. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at <http://sedac.ciesin.columbia.edu/gpw/lec2>. (downloaded January 18, 2012).
- Collard, I., and R. Foley. 2002. "Latitudinal Patterns and Environmental Determinants of Recent Human Cultural Diversity: Do Humans Follow Biogeographic Rules?" *Evolutionary Ecology Research* 4 (3):371-83.
- Collier, P., and A. Hoeffler. 2004. "Greed and grievance in civil war." *Oxford Economic Papers-New Series* 56 (4):563-95.
- Connor, W. 1994a. *Ethnonationalism: The Quest for Understanding*. Princeton, N.J.: Princeton University Press.
- . 1994b. "A nation is a nation, is a state...." In *Ethno-nationalism: The Quest for Understanding*, ed. Connor. Princeton, N.J.: Princeton University Press.
- Demeny, P., and G. McNicoll, eds. 2006. *The Political Economy of Global Population Change, 1950-2050*. New York: Population Council.
- Deutsch, Karl. 1966 (1953). *Nationalism and Social Communication: An Inquiry into the Foundations of Nationality*, 2nd Ed. Cambridge, MA: M.I.T. Press.
- Diamond, J. 1997. *Guns, Germs, and Steel: the Fates of Human Societies*. New York; London: W.W. Norton & Co.
- Easterly, W., and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 111 (4):1203-50.
- Easterly, William, Jozef Ritzen, and M. Woolcock. 2006. "Social Cohesion, Institutions, And Growth." *Economics and Politics* 18 (2):103-20.

- Ellingsen, T. 2000. "Colorful Community or Ethnic Witches' Brew: multiethnicity and domestic conflict during and after the Cold War." *The Journal of Conflict Resolution* 44 (2):228-49.
- Englebert, P., S. Tarango, and M. Carter. 2002. "Dismemberment and suffocation - A contribution to the debate on African boundaries." *Comparative Political Studies* 35 (10):1093-118.
- Eriksen, Thomas H. 1993. *Ethnicity and Nationalism: Anthropological Perspectives*. London: Pluto Press.
- Fearon, James D, and D Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97 (1):75-90.
- Fearon, J. D. 2005. "Primary commodity exports and civil war." *Journal of Conflict Resolution* 49 (4):483-507.
- Fletcher, E., and M. Iyigun. 2009. "Culture, Clashes and Peace." In *IZA Discussion Paper no. 4116*.
- Francis, E.K. 1976. *Interethnic Relations: An Essay in Sociological Theory*. New York, NY: Elsevier Scientific.
- Gellner, Ernest. 1983. *Nations and Nationalism*. Oxford: Blackwell.
- Giddens, Anthony. 1991. *Modernity and Self-Identity*. Cambridge: Polity Press.
- . [1985] 1996. *The Nation-State and Violence*. Cambridge: Polity Press.
- Green, Elliott. 2011. "Explaining African Ethnic Diversity'." In *American Political Science Association*. Seattle, WA.
- Hastings, Adrian. 1997. *The Construction of Nationhood: Ethnicity, Religion and Nationalism*. Cambridge & New York: Cambridge University Press.
- Hobsbawm, E. J. 1990. *Nations and Nationalism since 1780. Programme, Myth, Reality*. Cambridge: Cambridge University Press.
- Johnson, Nancy Kwang. 2004. "Senegalese into Frenchman or Peasants into Senegalese?: The Politics of Language, Culture and Assimilation: A Colonial and Post-Independence Critique (Senegal)," unpublished paper, *International Studies Association* conference, Montreal.
- Kaufmann, E. P. 2004. "Dominant ethnicity: from background to foreground." in *Rethinking Ethnicity: Majority Groups and Dominant Minorities*, ed. E. Kaufmann. London: Routledge.
- Kaufmann, Eric. 2008. "The Lenses of Nationhood: An Optical Model of Identity." *Nations & Nationalism* 14 (3):449-77.
- Kaufmann, Eric., and Oded Haklai. 2008. "Dominant Ethnicity: From Minority to Majority." *Nations and Nationalism* 14 (4):743-67.
- Kittel, Bernhard. 2001. "Sense and Sensitivity in Pooled Analysis of Political Data." *European Journal of Political Research* (35):225-53.
- Laitin, David D. 2007. *Nations, states, and violence*. Oxford ; New York: Oxford University Press.
- Laitin, David., and Amanda Robinson. 2011. "The Continental Axis Theory Revisited," unpublished paper, *American Political Science Association*. Seattle, WA.
- Mansergh, Nicholas. 1978. *The prelude to partition : concepts and aims in Ireland and India*. Cambridge [Eng.] ; New York: Cambridge University Press.
- Michalopoulos, Stelios. forthcoming. "The Origins of Ethnolinguistic Diversity." *American Economic Review*.
- Montalvo, J. G., and M. Reynal-Querol. 2005. "Ethnic polarization, potential conflict, and civil wars." *American Economic Review* 95 (3):796-816.

- Moore, J. L., L. Manne, T. Brooks, N. D. Burgess, R. Davies, and C. Rahbek. 2002. "The Distribution of Cultural and Biological Diversity in Africa." *Proceedings of the Royal Society of London* 296 (1501):1645-53.
- Nairn, Tom. 1977. *The Break-Up of Britain*. London: NLB.
- Nettle, D. . 1996. "Language Diversity in West Africa." *Journal of Anthropological Archaeology* 15 (4):403-38.
- Nichols, J. . 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nunn, Nathan. 2008. "The Long Term Effects of Africa's Slave Trade." *Quarterly Journal of Economics* 123 (1):139-76.
- Osaghae, E. E. 2008. "Constructions of belonging: Igbo communities and the Nigerian state in the twentieth century." *African Affairs* 107 (429):666-7.
- Özirimli, Umut. 2010. *Theories of nationalism : a critical introduction*. 2nd ed. ed. Basingstoke: Palgrave Macmillan.
- Pinker, Steven. 2011. *The better angels of our nature : why violence has declined*. New York: Viking.
- Posner, Daniel. 2005. "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science* 48 (4):849-63.
- Putterman, L. 2007. *The State Antiquity Dataset*. Brown University [cited February 2, 2012.] Available from http://www.econ.brown.edu/fac/Louis_Putterman/antiquity%20index.htm.
- Roeder, P. 2010. *Ethnolinguistic Fractionalization (ELF) Indices, 1961 and 1985* 2001 [cited 2010]. Available from <http://weber.ucsd.edu/~proeder/elf.htm>.
- Saideman, Steve, and R. W Ayres. 2008. *For Kin or Country: Xenophobia, Nationalism, and War* New York, N.Y.: Columbia University Press.
- Salter , Frank K. 2001. "A Defense and an Extension of Pierre van den Berghe's Theory of Ethnic Nepotism," in *Evolutionary Theory and Ethnic Conflict*, ed. P. James and D. Goetze. Westport, CT and London: Praeger.
- Sambanis, N. 2001. "Do ethnic and nonethnic civil wars have the same causes? A theoretical and empirical inquiry (part 1)." *Journal of Conflict Resolution* 45 (3):259-82.
- Schneider, Robert A, and N. Wiesehomeier. 2009. "Ethnic Polarization, Potential Conflict, and Civil Wars: Comment." *working paper, University of Konstanz*.
- Smith, Anthony. 1986. *The Ethnic Origins of Nations*. Oxford: Blackwell.
- Smith, M.A. 1995. "Time serial designs and cross-sectional designs: Uncovering the structural logic of pooled analyses." In *Paper presented at the 1995 American Political Science Association Meetings*.
- Sutherland, W.J. 2003. "Parallel Extinction Risk and Global Distribution of Languages and Species." *Nature* 423 (6937):276-9.
- Thomas, William Isaac, and Florian Znaniecki. 1918. *The Polish peasant in Europe and America; monograph of an immigrant group*. Chicago, Ill.: The University of Chicago press.
- Tönnies, Ferdinand, and Charles Price Loomis. 1940. *Fundamental concepts of sociology (Gemeinschaft und gesellschaft)*. New York, Cincinnati [etc.]: American Book Company.
- Trevor-Roper, Hugh. 1983. "The Invention of Tradition: The Highland Tradition of Scotland," in *The Invention of Tradition*, ed. E. Hobsbawm and T. Ranger. Cambridge: Cambridge University Press.
- Urdal, H. 2008. "Population, resources, and political violence - A subnational study of India, 1956-2002." *Journal of Conflict Resolution* 52 (4):590-617.

- Van den Berghe, P.L. 2002. "Multicultural democracy: can it work?" *Nations and Nationalism* 8 (4):433-49.
- Vanhanen, T. 1999. *Ethnic Conflicts Explained by Ethnic Nepotism*. Stamford, CT: Jai Press.
- Wimmer, Andreas. 2002. *Nationalist Exclusion and Ethnic Conflict. Shadows of Modernity*. Cambridge: Cambridge University Press.
- 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 (2):316-37.
- Zimmer, O. 2003a. "Boundary mechanisms and symbolic resources: towards a process-oriented approach to national identity." *Nations and Nationalism* 9 (2):173-93.
- _____. 2003b. *Nationalism in Europe: 1890-1940*. London: Palgrave/Macmillan.
- Zimmer, O., and Len Scales, eds. 2005. *Power and the Nation in History*. Cambridge: CUP.
- Zisser, Eyal. 2000. *Lebanon: The Challenge of Independence*. London & New York: I.B.Tauris.

Table 1. Geoclimactic Model

	NMBRLANG	ELF	EF	ELFPREG	PLURALITY	PCTMAJ	MAXPOP
Mean	0.071***	0.007*	0.010***	0.007	-0.008**	-0.701**	-0.006
Temperature	(0.015)	(0.003)	(0.003)	(0.004)	(0.003)	(0.242)	(0.003)
Mean	0.008***	0.001**	0.001**	0.000	-0.001**	-0.105***	0.000
Precipitation	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.028)	(0.000)
Distance to	-0.379	0.049	0.001	-0.027	0.002	-4.011	0.026
the Sea	(0.317)	(0.066)	(0.060)	(0.072)	(0.055)	(5.239)	(0.064)
Mean	-0.263	-0.084	-0.073	-0.017	0.082	3.438	0.039
Elevation	(0.323)	(0.067)	(0.061)	(0.078)	(0.056)	(5.336)	(0.070)
S.d. of Mean	1.193**	0.263**	0.248**	0.129	-0.248**	-14.235*	-0.192*
Elevation	(0.431)	(0.090)	(0.082)	(0.106)	(0.074)	(7.120)	(0.095)
Mean Agric.	-0.360	-0.112	-0.244**	-0.288**	0.195*	15.220*	0.231*
Suitability	(0.432)	(0.090)	(0.082)	(0.107)	(0.075)	(7.140)	(0.095)
S.d. of Mean	2.756*	0.305	0.312	0.479	-0.253	-29.660	-0.326
Agric. Suitability	(1.078)	(0.225)	(0.204)	(0.264)	(0.186)	(17.819)	(0.236)
Land Area	0.562***	0.034*	0.021	0.021	-0.017	-1.263	-0.010
(Square km)	(0.068)	(0.014)	(0.013)	(0.018)	(0.012)	(1.124)	(0.016)
Sub-Saharan	0.474	0.354***	0.327***	0.238***	-0.327***	-24.354***	-0.280***
Africa	(0.250)	(0.052)	(0.047)	(0.064)	(0.043)	(4.135)	(0.057)
Eastern Europe	0.527	0.227***	0.221***	0.048	-0.171**	-15.98**	-0.012
	(0.297)	(0.062)	(0.056)	(0.072)	(0.051)	(4.911)	(0.065)
Western	-0.633*	-0.091	-0.031	-0.117	0.032	-1.455	0.090
Hemisphere	(0.264)	(0.055)	(0.050)	(0.067)	(0.046)	(4.369)	(0.060)
constant	-0.469	-0.029	0.056	0.278	1.015***	105.478***	0.833***
	(0.380)	(0.079)	(0.072)	(0.103)	(0.066)	(6.288)	(0.092)
N	146	146	146	118	146	146	117
R²	0.627	0.525	0.557	0.449	0.570	0.488	0.508

* $p \leq 0.05$, ** $p \leq 0.01$; *** $p \leq 0.001$; robust standard errors in parentheses. Dependent variable at top of each column. NMBRLANG, ELF, EF and ELFPREG are measures of ethnic fractionalization, PLURAL, PCTMAJ and MAXPOP of the size of the largest ethnic group. See Appendix 1 for further details.

Table 2. Historical Institutional Model

	NMBRLANG	ELF	EF	ELFPREG	PLURALITY	PCTMAJ	MAXPOP
State	-0.690	0.149	-0.007	0.192	0.014	1.673	-0.167
Antiquity	(0.599)	(0.110)	(0.099)	(0.118)	(0.090)	(8.164)	(0.107)
Ethnic	0.000	0.000	0.000	0.000	0.000	-0.003	0.000
Origin Date	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)
Ethnic Origin	-0.939**	-0.184**	-0.183**	-0.197**	0.137*	14.851**	0.138*
0-1100 A.D.	(0.351)	(0.065)	(0.058)	(0.073)	(0.053)	(4.789)	(0.066)
Transition to	0.011	0.013	0.021*	0.013	-0.021*	-1.686	-0.022*
Agriculture	(0.064)	(0.012)	(0.011)	(0.012)	(0.010)	(0.871)	(0.011)
log Population	0.267*	-0.020	0.002	0.003	-0.005	-1.253	0.011
Density 1500	(0.110)	(0.020)	(0.018)	(0.022)	(0.017)	(1.505)	(0.020)
Land Area	0.555***	0.029	0.007	0.041	-0.008	-0.980	-0.021
(Square km)	(0.104)	(0.019)	(0.017)	(0.023)	(0.016)	(1.424)	(0.020)
Slave Exports	0.049	0.018**	0.013*	0.007	-0.011*	-1.210*	-0.009
Per Capita	(0.037)	(0.007)	(0.006)	(0.008)	(0.006)	(0.503)	(0.007)
Population	-2.095	-0.316	-0.780**	-0.439	0.716**	68.395**	0.442
Density (1995)	(1.596)	(0.294)	(0.263)	(0.302)	(0.240)	(21.768)	(0.273)
Population	0.138	0.019	0.031	-0.008	-0.024	-0.966	0.002
(log)	(0.110)	(0.020)	(0.018)	(0.023)	(0.017)	(1.500)	(0.021)
Sub-Saharan	-0.137	0.182*	0.217**	0.240*	-0.219**	-11.096	-0.258**
Africa	(0.486)	(0.090)	(0.080)	(0.099)	(0.073)	(6.631)	(0.089)
Western	0.362	-0.013	0.079	-0.015	-0.086	-11.254*	-0.007
Hemisphere	(0.365)	(0.067)	(0.060)	(0.074)	(0.055)	(4.972)	(0.067)
Eastern	-0.726*	0.077	0.015	-0.078	0.010	0.671	0.109
Europe	(0.314)	(0.058)	(0.052)	(0.061)	(0.047)	(4.289)	(0.055)
constant	0.378	-0.151	-0.225	0.470	1.239***	100.609***	0.793*
	(1.868)	(0.345)	(0.308)	(0.396)	(0.281)	(25.476)	(0.353)
N	130	130	130	109	130	130	109
R²	0.492	0.469	0.518	0.447	0.529	0.472	0.500

* $p \leq 0.05$, ** $p \leq 0.01$; *** $p \leq 0.001$; robust standard errors in parentheses. Dependent variable at top of each column. NMBRLANG, ELF, EF and ELFPREG are measures of ethnic fractionalization, PLURAL, PCTMAJ and MAXPOP of the size of the largest ethnic group. See Appendix 1 for further details.

Table 3. Modernist Model

	NMBRLANG	ELF	EF	ELFPREG	PLURALITY	PCTMAJ	MAXPOP
State	0.001	0.001	0.001	0.001	0.000	-0.033	0.000
Found. Date	(0.003)	(0.000)	(0.000)	(0.001)	(0.000)	(0.037)	(0.001)
Percent	-1.285	-0.033	0.183	0.097	-0.168	-19.000	-0.072
Urbanized	(0.816)	(0.142)	(0.142)	(0.177)	(0.128)	(11.270)	(0.161)
Instability	0.217	0.029	-0.001	-0.010	-0.006	-2.011	-0.010
	(0.307)	(0.053)	(0.053)	(0.060)	(0.048)	(4.242)	(0.053)
Democracy	-0.006	-0.001	-0.001	0.000	0.001	0.166**	0.000
(Polity IV)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.058)	(0.001)
GDP per	0.075	-0.014	-0.070	-0.044	0.074*	5.484	0.052
capita	(0.206)	(0.036)	(0.036)	(0.042)	(0.032)	(2.846)	(0.038)
Oil Output	-0.015	-0.003	0.006*	0.010*	-0.006*	0.162	-0.011**
Per Capita	(0.017)	(0.003)	(0.003)	(0.004)	(0.003)	(0.239)	(0.004)
Population	0.067	0.013	0.009	-0.013	-0.007	0.377	0.012
(log)	(0.104)	(0.018)	(0.018)	(0.021)	(0.016)	(1.442)	(0.019)
Land Area	0.491***	0.044**	0.028	0.049**	-0.022	-2.159	-0.036*
(Square km)	(0.085)	(0.015)	(0.015)	(0.018)	(0.013)	(1.175)	(0.016)
Sub-Saharan	0.041	0.252***	0.243***	0.166*	-0.233***	-15.350**	-0.206**
Africa	(0.359)	(0.062)	(0.062)	(0.072)	(0.056)	(4.954)	(0.064)
East Europe	0.217	0.030	0.091	-0.047	-0.074	-7.438	0.027
	(0.313)	(0.055)	(0.055)	(0.065)	(0.049)	(4.331)	(0.058)
Western	-0.785*	0.034	0.002	-0.127	0.010	1.421	0.121*
Hemisphere	(0.345)	(0.060)	(0.060)	(0.066)	(0.054)	(4.761)	(0.059)
constant	-0.322	-1.338	-0.626	-0.645	1.092	91.332	0.695
	(6.109)	(1.063)	(1.063)	(1.239)	(0.958)	(84.402)	(1.112)
N	136	136	136	112	136	136	113
R²	0.453	0.519	0.455	0.439	0.478	0.466	0.496

* $p \leq 0.05$, ** $p \leq 0.01$; *** $p \leq 0.001$; robust standard errors in parentheses. Dependent variable at top of each column. NMBRLANG, ELF, EF and ELFPREG are measures of ethnic fractionalization, PLURAL, PCTMAJ and MAXPOP of the size of the largest ethnic group. See Appendix 1 for further details.

Table 4. International Determinants Model

	NMBRLANG	ELF	EF	ELFPREG	PLURALITY	PCTMAJ	MAXPOP
Dominant	-0.106	-0.040	-0.011	0.037	0.010	1.722	-0.005
Minority	(0.239)	(0.041)	(0.040)	(0.047)	(0.037)	(3.467)	(0.043)
Secessions	0.102	-0.021	-0.007	0.000	0.008	0.735	0.002
	(0.098)	(0.017)	(0.017)	(0.018)	(0.015)	(1.424)	(0.016)
Ethnic	-0.394	-0.208**	-0.188**	-0.133	0.155**	15.596**	0.072
Nationalism	(0.359)	(0.062)	(0.061)	(0.075)	(0.056)	(5.213)	(0.069)
State Found.	0.668	0.025	0.069	0.099	-0.075	-6.401	-0.086
1816-1918	(0.456)	(0.079)	(0.077)	(0.100)	(0.071)	(6.626)	(0.092)
State Found.	0.455	0.134	0.160	0.264*	-0.133	-9.558	-0.234*
1919-1945	(0.493)	(0.085)	(0.083)	(0.103)*	(0.076)	(7.161)	(0.095)
State Found.	0.908*	0.112	0.162*	0.208*	-0.140*	-7.162	-0.162
1946-1990	(0.424)	(0.073)	(0.072)	(0.094)	(0.066)	(6.158)	(0.087)
State Found.	0.937	0.061	0.097	0.135	-0.079	-8.506	-0.095
1991-1999	(0.665)	(0.115)	(0.112)	(0.128)	(0.103)	(9.665)	(0.118)
Land Area	0.499***	0.046**	0.029*	0.029	-0.024*	-1.702	-0.023
(Square km)	(0.076)	(0.013)	(0.013)	(0.016)	(0.012)	(1.111)	(0.014)
Sub-Saharan	0.356	0.317***	0.274***	0.180**	-0.272***	-24.066***	-0.238***
Africa	(0.299)	(0.052)	(0.051)	(0.060)	(0.046)	(4.342)	(0.056)
Eastern Europe	0.002	0.014	0.048	-0.082	-0.039	-6.225	0.039
	(0.350)	(0.061)	(0.059)	(0.072)	(0.054)	(5.081)	(0.066)
Western	-0.662	0.236*	0.139	-0.009	-0.101	-9.667	0.040
Hemisphere	(0.534)	(0.092)	(0.090)	(0.101)	(0.083)	(7.757)	(0.093)
constant	1.910***	0.439***	0.456***	0.442***	0.679***	67.932***	0.732***
	(0.557)	(0.096)	(0.094)	(0.114)	(0.086)	(8.094)	(0.105)
N	140	140	140	113	140	140	114
R²	0.366	0.494	0.456	0.443	0.463	0.384	0.473

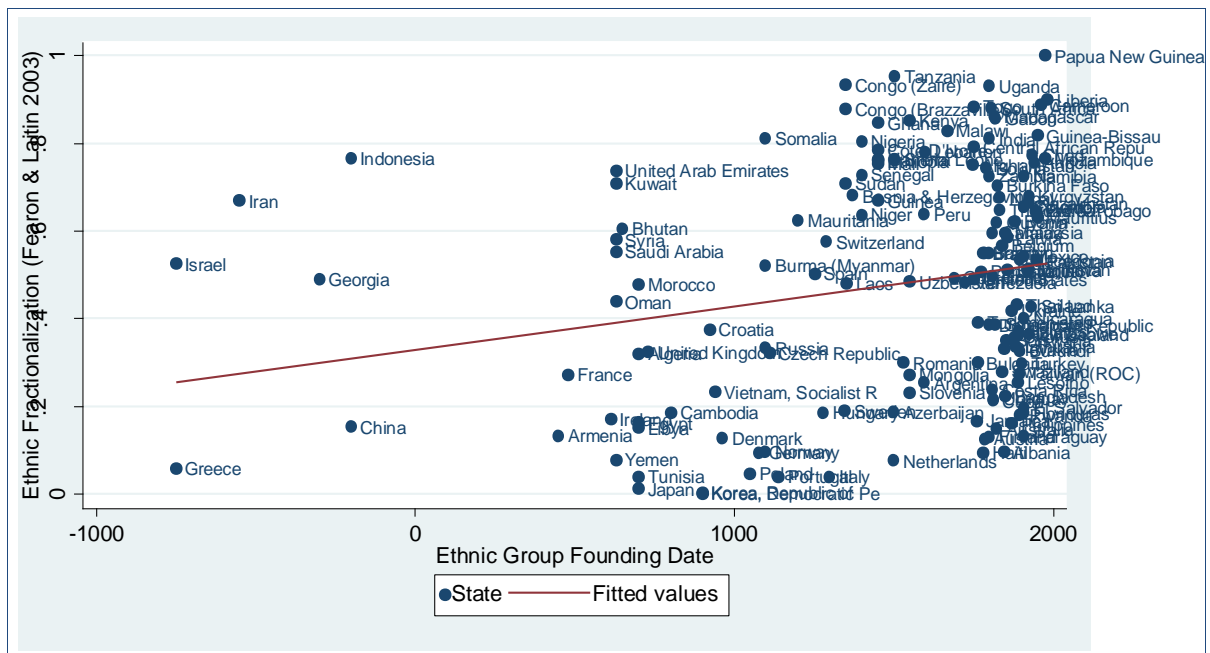
* $p \leq 0.05$, ** $p \leq 0.01$; *** $p \leq 0.001$; robust standard errors in parentheses. Reference category for state founding wave is states established before 1816. Dependent variable at top of each column. NMBRLANG, ELF, EF and ELFPREG are measures of ethnic fractionalization, PLURAL, PCTMAJ and MAXPOP of the size of the largest ethnic group. See Appendix 1 for further details.

Table 5. Combined Model of Ethnic Fractionalization

	NMBRLANG	ELF	EF	ELFPREG	PLURALITY	PCTMAJ	MAXPOP
State	0.002	0.001***	0.001**	0.001	-0.001*	-0.046	0.000
Found. Date	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.027)	(0.000)
Land Area	0.555***	0.038**	0.017	0.038*	-0.014	-1.054	-0.024
(Square km)	(0.070)	(0.012)	(0.011)	(0.016)	(0.010)	(0.912)	(0.015)
Democracy	-0.006	-0.001	-0.001	0.000	0.001	0.148**	0.000
(Polity IV)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.049)	(0.001)
GDP per	-0.047	0.006	-0.017	-0.016	0.026	0.893	0.026
capita	(0.156)	(0.026)	(0.025)	(0.033)	(0.023)	(2.046)	(0.030)
Oil Output	0.009	-0.001	0.009**	0.015**	-0.009**	-0.184	-0.015***
Per Capita	(0.016)	(0.003)	(0.003)	(0.004)	(0.002)	(0.215)	(0.004)
Ethnic Origin	-0.287	-0.141**	-0.126**	-0.077	0.085*	12.921**	0.030
0-1100 A.D.	(0.288)	(0.049)	(0.047)	(0.068)	(0.043)	(3.772)	(0.062)
Mean	0.008***	0.001	0.001*	0.000	-0.001**	-0.061**	0.000
Precipitation	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.023)	(0.000)
S.d. of Mean	2.043	0.518*	0.582**	0.506	-0.507**	-41.721*	-0.344
Agric. Suitability	(1.214)	(0.206)	(0.198)	(0.257)	(0.182)	(15.928)	(0.232)
S.d. of Mean	1.208*	0.171*	0.190*	0.048	-0.186*	-9.561	-0.127
Elevation	(0.491)	(0.083)	(0.080)	(0.102)	(0.074)	(6.437)	(0.093)
Mean	-0.751*	-0.105	-0.116*	0.003	0.114*	7.097	0.019
Elevation	(0.325)	(0.055)	(0.053)	(0.068)	(0.049)	(4.267)	(0.062)
Sub-Saharan	0.695*	0.239***	0.222***	0.242**	-0.234***	-12.801**	-0.315***
Africa	(0.315)	(0.054)	(0.052)	(0.068)	(0.047)	(4.134)	(0.061)
Ethnic	-0.134	-0.091	-0.110*	-0.100	0.076	7.960*	0.031
Nationalism	(0.276)	(0.047)	(0.045)	(0.061)	(0.041)	(3.626)	(0.056)
Population	0.063*	-0.267**	-0.587**	-0.223	0.486*	51.546**	0.182
Density (1995)	(1.235)	(0.210)	(0.202)	(0.253)	(0.185)	(16.204)	(0.230)
constant	-1.384	-2.428**	-1.439	-1.052	1.858**	144.659*	1.067
	(4.546)	(0.772)	(0.743)	(0.980)	(0.681)	(59.625)	(0.893)
N	136	136	136	112	136	136	113
R²	0.565	0.636	0.618	0.497	0.621	0.617	0.537

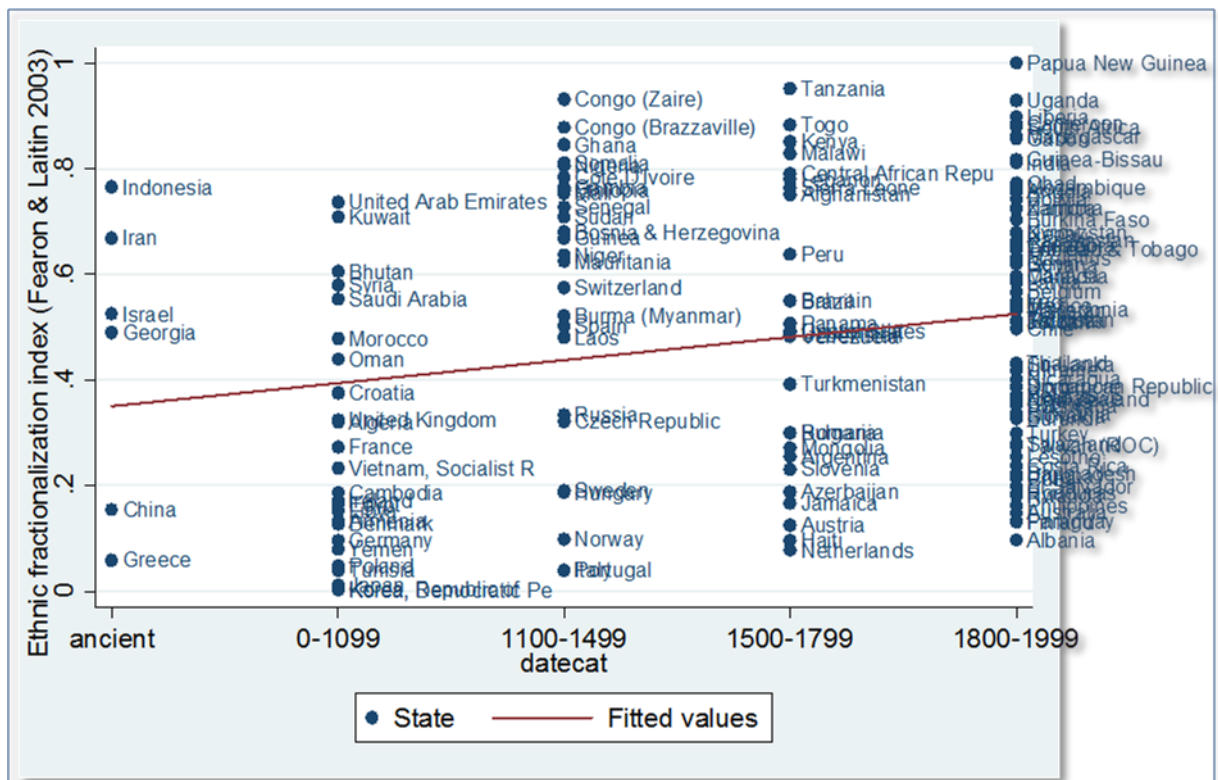
* $p \leq 0.05$, ** $p \leq 0.01$; *** $p \leq 0.001$; robust standard errors in parentheses. Dependent variable at top of each column. NMBRLANG, ELF, EF and ELFPREG are measures of ethnic fractionalization, PLURAL, PCTMAJ and MAXPOP of the size of the largest ethnic group. See Appendix 1 for further details.

Figure 1. Plurality Ethnic Group Founding Dates and Ethnic Fractionalization



Source: Own data.

Figure 2. Plurality Ethnic Group Founding Date Category and Ethnic Fractionalization



Source: Own data.

Appendix 1. Data Sources

Variable	Source
Log of number of languages spoken in country (NMBRLANG)	Michalopoulos, forthcoming
Ethnolinguistic fractionalization index (ELF)	Laitin and Fearon 2003
Ethnic fractionalization (EF)	Laitin and Fearon 2003
Ethnic fractionalization of politically-relevant groups (ELFPREG)	Wimmer et. al. 2009
Largest ethnic group (PLURAL)	Laitin and Fearon 2003
Largest ethnic group (PCTMAJ)	Vanhanen 1999
Largest politically-relevant ethnic group (MAXPOP)	Wimmer et. al. 2009
Mean Temperature	Michalopoulos, forthcoming
Mean Elevation	Michalopoulos, forthcoming
Mean Precipitation	Michalopoulos, forthcoming
Distance to the sea	Michalopoulos, forthcoming
Standard deviation of mean elevation	Michalopoulos, forthcoming
Mean agricultural suitability	Michalopoulos, forthcoming
Standard deviation of mean agricultural suitability	Michalopoulos, forthcoming
State antiquity	Putterman 2007
Ethnic origin date	Own data
Origin 450-1099 A.D.	Own data
Transition to agriculture	Michalopoulos, forthcoming
log Population density 1500	Michalopoulos, forthcoming
State founding date	Correlates of War (COW) Interstate Data
Political instability	Laitin and Fearon 2003
Percent urbanized	CIESIN, accessed 2012
Democracy (Polity IV) score	CIESIN, accessed 2012
GDP per capita	Laitin and Fearon 2003
Oil output per capita	Laitin and Fearon 2003
Log population	Wimmer et. al. 2009
Land area (sq. km.)	Michalopoulos, forthcoming
Sub-Saharan Africa	Laitin and Fearon 2003
Western hemisphere	Laitin and Fearon 2003
East Europe	Laitin and Fearon 2003
Commodity exports per capita	Fearon 2005
Dominant minority	Own data
Ethnic nationalism	Own data
Secessions in state history	Correlates of War (COW) Interstate Data
State founded pre-1815	Correlates of War (COW) Interstate Data
State founded 1816-1918	Correlates of War (COW) Interstate Data
State founded 1919-1945	Correlates of War (COW) Interstate Data
State founded 1946-90	Correlates of War (COW) Interstate Data
State founded 1991-1999	Correlates of War (COW) Interstate Data
Population density (1995)	Michalopoulos, forthcoming
Infant mortality rate	Wimmer et. al. 2009
Slave exports per historic population	Nunn 2008
Ex-colony (British or French)	Laitin and Fearon 2003

Appendix 2. Summary Statistics

	N	Mean	S.d.	Min	Max
Log of number of languages spoken in country (NMBRLANG)	149	2.46	1.51	0.00	6.14
Ethnolinguistic fractionalization index (ELF)	153	0.41	0.28	0.00	0.93
Ethnic fractionalization (EF)	153	0.47	0.26	0	1
Ethnic fractionalization of politically-relevant groups (ELFPREG)	122	0.50	0.27	0.02	1.00
Largest ethnic group (PLURAL)	153	0.65	0.24	0	1
Largest ethnic group (PCTMAJ)	156	69.28	21.37	16	100
Largest politically-relevant ethnic group (MAXPOP)	122	0.63	0.25	0.02	0.99
Mean Temperature	150	17.98	8.26	-4.79	28.74
Mean Elevation	150	0.59	0.50	0.03	2.52
Mean Precipitation	150	89.53	62.06	4.00	278.16
Distance to the sea	150	0.35	0.38	0.01	1.98
Standard deviation of mean elevation	150	0.37	0.37	0.01	1.95
Mean agricultural suitability	150	0.45	0.25	0.00	0.96
Standard deviation of mean agricultural suitability	150	0.19	0.09	0.00	0.41
State antiquity (5% discounting of each 50-year period)	142	0.45	0.24	0.02	0.96
Ethnic origin date	156	1467.55	574.28	-750	1981
Origin 450-1099 A.D.	156	0.16	0.37	0	1
Transition to agriculture	145	4.87	2.36	0.40	10.50
log Population density 1500	145	0.92	1.50	-3.82	3.84
State founding date	148	1924.86	56.39	1816	1993
Political instability	153	0.16	0.36	0	1
Percent urbanized	156	0.51	0.24	0.07	0.98
Democracy (Polity IV) score	152	49.64	35.57	0	100
GDP per capita	152	8.40	1.06	6.20	10.35
Oil output per capita	148	1.81	6.74	0.00	50.51
Log population	156	16.31	1.48	11.11	21.02
Land area (sq. km.)	149	0.61	1.48	-2.83	4.73
Sub-Saharan Africa	153	0.27	0.45	0	1
Western hemisphere	156	0.17	0.37	0	1
East Europe	153	0.17	0.38	0	1
Commodity exports per capita	131	0.14	0.14	0.00	0.71
Dominant minority	156	0.10	0.30	0	1
Ethnic nationalism	156	1.16	0.37	1	2
Secessions in state history	156	0.37	2.03	0	23
State founded pre-1815	148	0.04	0.19	0	1
State founded 1816-1918	148	0.26	0.44	0	1
State founded 1919-1945	148	0.09	0.28	0	1
State founded 1946-1990	148	0.44	0.50	0	1
State founded 1991-1999	148	0.12	0.33	0	1

Infant mortality rate	151	47.36	39.02	4	170
log population density (1995)	149	-3.18	1.31	-9.20	-0.25
Former colony (British or French)	153	0.45	0.50	0	1
African slave exports per historic population	156	2.86	4.86	0	14

Appendix 3. Alternative Specifications

	EF1	EF2	PCTMAJ1	PCTMAJ2	EF3	PCTMAJ3
State founding date	0.083* (0.035)	0.085* (0.034)	-0.029 (0.029)	-0.030 (0.028)	0.083 (0.043)	-0.041 (0.039)
Land area (sq. km)	0.021 (0.012)	0.019 (0.012)	-1.264 (0.992)	-1.127 (0.992)		
Democracy (Polity IV) score	0.021 (0.012)	0.019 (0.012)	-1.264** (0.992)	-1.127 (0.992)		
log GDP per capita	-0.025 (0.026)	-0.016 (0.026)	1.367 (2.104)	0.836 (2.126)	-0.034 (0.024)	4.107 (2.148)
Oil output per capita	0.009** (0.003)	0.008** (0.003)	-0.237 (0.228)	-0.180 (0.230)	0.008* (0.003)	-0.256 (0.264)
Ethnic group founding date	0.115** (0.043)	0.118** (0.042)	-0.013*** (0.003)	-0.013*** (0.003)		
Ethnic origin 450-1099 A.D.					-0.092 (0.050)	10.007* (4.484)
Mean precipitation	0.001* (0.000)	0.001* (0.000)	-0.067* (0.026)	-0.062* (0.026)	0.001* (0.000)	-0.073** (0.027)
Mean agricultural suitability	-0.120 (0.076)	-0.122 (0.075)	11.901 (6.256)	11.977 (6.227)		
S.d. of agricultural suitability	0.542* (0.216)	0.404 (0.223)	-46.098* (17.753)	-38.335* (18.467)	0.558* (0.228)	-51.678* (20.513)
S.d. of mean elevation	0.229** (0.082)	0.228** (0.081)	-11.370 (6.753)	-11.271 (6.722)	0.143* (0.058)	-5.056 (5.235)
Mean elevation	-0.124* (0.054)	-0.095 (0.055)	7.594 (4.437)	6.006 (4.551)		
Sub-Saharan Africa	0.257** (0.051)	0.111 (0.085)	-15.867*** (4.211)	-7.620 (7.076)	-0.056 (0.098)	3.280 (8.787)
Sub-Saharan Africa x Slave exports		0.016* (0.007)		-0.896 (0.620)	0.025** (0.009)	-1.602* (0.769)
Ethnic nationalism	-0.086 (0.047)	-0.093* (0.047)	4.862 (3.866)	5.271 (3.858)	-0.128* (0.054)	11.170* (4.828)
Commodity exports as share of GDP					0.340* (0.172)	-18.216 (15.450)
Ex-British or French colony					0.056 (0.040)	-0.623 (3.556)
_cons	-1.188 (0.767)	-1.281 (0.757)	134.914* (62.913)	140.116* (62.722)	-1.022 (0.898)	124.699 (80.820)
R²	130	130	130	130	114	114
N	0.621	0.636	0.606	0.613	0.666	0.581

* $p \leq 0.05$, ** $p \leq 0.01$; *** $p \leq 0.001$; robust standard errors in parentheses. Models based on ethnic fractionalization (EF) and majority ethnic share (PCTMAJ) as dependent variables. Model 1s use ethnic group founding date after year 0 instead of a dummy for ethnic founding dates 450-1099. Model 2s as Model 1s but with addition of interaction term for sub-Saharan African slave exports. Model 1s and 2s exclude cases prior to year 0. Model 3s test effect of commodity exports as a share of GDP and term for ex-British or French colonies.

Appendix 4. List of States for Ethnic Nationalism and Dominant Minority Variables

<i>Ethnic Basis for National Self-Determination</i>	<i>Dominant Minority</i>
Albania	Bahrain
Armenia	Burundi
Azerbaijan	Ethiopia
Bangladesh	Iraq
Bulgaria	Kenya
Croatia	Liberia
Estonia	Lebanon
Finland	Niger
Georgia	Rwanda
Germany	Syria
Greece	United Arab Emirates
Hungary	
Ireland	
Israel	
Italy	
Latvia	
Lithuania	
Mongolia	
Norway	
Poland	
Romania	
Serbia	
Singapore	
Slovakia	
Slovenia	