

Arbath, Cemal Eren; Ashraf, Quamrul; Galor, Oded

**Working Paper**

## The nature of civil conflict

Working Paper, No. 2013-15

**Provided in Cooperation with:**

Department of Economics, Brown University

*Suggested Citation:* Arbath, Cemal Eren; Ashraf, Quamrul; Galor, Oded (2013) : The nature of civil conflict, Working Paper, No. 2013-15, Brown University, Department of Economics, Providence, RI

This Version is available at:

<https://hdl.handle.net/10419/102630>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The Nature of Civil Conflict\*

Cemal Eren Arbatli<sup>†</sup>  
HSE Moscow

Quamrul Ashraf<sup>‡</sup>  
Williams College

Oded Galor<sup>§</sup>  
Brown University and NBER

## Abstract

This research empirically establishes that the emergence, prevalence, and recurrence of civil conflict in the modern era reflect the long shadow of prehistory. Exploiting variations across contemporary national populations, it demonstrates that genetic diversity, as determined predominantly tens of thousands of years ago, has contributed significantly to the frequency, incidence, and onset of both overall and ethnic civil conflicts over the last half century, accounting for a large set of geographical and institutional correlates of civil conflict, as well as measures of economic development. These findings arguably reflect the adverse effect of genetic diversity on interpersonal trust and cooperation, the potential impact of genetic diversity on income inequality, the potential association between genetic diversity and divergence in preferences for public goods and redistributive policies, and the contribution of genetic diversity to the degree of fractionalization and polarization across ethnic and linguistic groups in the population.

---

\*We thank Yann Algan, Carl-Johan Dalgaard, David de la Croix, Pauline Grosjean, Stelios Michalopoulos, Nathan Nunn, Enrico Spolaore, Romain Wacziarg, Fabrizio Zilibotti, and participants at the 2013 AEA Annual Meeting, the 2013 NBER Summer Institute Workshop on “Income Distribution and Macroeconomics,” and the conference on “Deep Determinants of International Comparative Development” at Brown University for helpful comments and discussions. Ashraf acknowledges research support from the Hellman Fellows Program at Williams College. Ashraf and Galor both acknowledge research support from the NSF (SES-1338738).

<sup>†</sup>Department of Economics, Higher School of Economics, 26 Shabolovka St., Building 3, 3116A, Moscow, Russia (e-mail: darbatli@hse.ru).

<sup>‡</sup>Department of Economics, Williams College, Schapiro Hall, 24 Hopkins Hall Dr., Williamstown, MA 01267 (email: Quamrul.H.Ashraf@williams.edu).

<sup>§</sup>Department of Economics, Brown University, Robinson Hall, 64 Waterman St., Providence, RI 02912 (email: Oded.Galor@brown.edu).

# 1 Introduction

Civil conflicts in the post World War II era have resulted in more than 16 million casualties across the globe, surpassing the loss of human life associated with interstate wars. Regions plagued by civil conflicts have experienced significant human casualties, substantial loss of productive resources, and considerable reduction in productivity, investment, and trade flows. Further, more than 20% of all nations experienced at least 10 years of civil conflict during the 1960–2006 time period. While the proportion of countries with active conflicts declined from 52 at its peak in the early 1990s, more than 30 countries were still experiencing one or more civil conflicts in 2008.

This research advances the hypothesis that the emergence, prevalence, and recurrence of civil conflicts across the globe reflect the long shadow of prehistory. The analysis establishes that the genetic diversity of contemporary national populations, as determined predominantly in the course of the exodus of humans from Africa tens of thousands of years ago, has contributed to the frequency, incidence, and onset of both civil and ethnic conflicts over the last half century, accounting for the potentially confounding influence of a wide variety of factors, including traditional group-based measures of diversity such as ethnic and linguistic fractionalization and polarization.

Genetic diversity may contribute to the emergence of civil conflicts through three direct channels. First, as argued by Ashraf and Galor (2013a), genetic diversity has had an adverse effect on trust and cooperation among members of society, and it can thereby reduce the socio-cultural threshold for the onset of a conflict. Second, to the extent that diversity in preferences over public goods and political outcomes reflect diversity in genetic traits, social conflicts could be more prevalent in more diverse societies. Third, genetic diversity and its manifestation in heterogeneity in cognitive and physical traits may have contributed to inequality and, thus, to socio-political instability and the propensity for civil conflicts.

Moreover, genetic diversity may influence the prevalence of conflicts indirectly, via its effect on the emergence and evolution of ethnic groups and their distinct identities (Ashraf and Galor, 2013b). Over the course of the human history, following the “out of Africa” migration, the initial level of genetic diversity in indigenous settlements across the globe presumably facilitated the formation of distinct ethnic groups through a process of endogenous group formation, based on the trade-off between the costs and benefits associated with heterogeneity and scale. As the forces of cultural drift augmented intergroup divergences in language, customs, and norms over time, reinforcing the barriers to intergroup assimilation, distinct ethnic identities were formed. While heterogeneity raised the likelihood of disarray and mistrust, reducing cooperation and thus adversely affecting group-specific productivity, complementarities across diverse productive traits and preferences stimulated productivity. In particular, since diminishing marginal returns to diversity and homogeneity would imply an optimal size for each group, higher initial genetic diversity may have positively contributed to the number of groups, and thus to the degree of ethnic fractionalization. Further, to the extent that higher initial diversity did not lead to an excessively

large number of groups, it would have positively contributed to the degree of ethnic polarization as well.

The empirical analysis establishes that genetically diverse countries have had a significantly higher propensity for civil and ethnic conflicts, accounting for the potentially confounding effects of ethnic fractionalization and polarization, institutional characteristics, income, population size, and several geographical factors that have received much attention in the empirical literature on cross-country comparative development and civil conflict. Furthermore, it is shown that the results are qualitatively insensitive to the data sources and the underlying codings for conflicts. Moreover, the analysis establishes that the reduced-form impact of genetic diversity on conflicts may, indeed, partly operate through two mediating channels, namely interpersonal trust and ethnolinguistic fragmentation.

Population geneticists measure the extent of diversity in genetic material across individuals within a given population (e.g., an ethnic group) using an index called expected heterozygosity. Like most other measures of diversity, this index may be interpreted simply as the probability that two individuals, selected at random from the relevant population, are genetically different from one another with respect to a given spectrum of traits. Specifically, the expected heterozygosity measure for a given population is constructed by geneticists using sample data on allelic frequencies. Given allelic frequencies for a particular gene or DNA locus, a gene-specific heterozygosity statistic may be computed (i.e., the probability that two randomly selected individuals differ with respect to the gene in question), which when averaged over multiple genes or DNA loci yields the overall expected heterozygosity for the relevant population.

While existing data on genetic diversity pertain only to ethnic groups, many national populations today are composed of multiple ethnicities, some of which may not be indigenous to their current geographical locations. Hence, the index of genetic diversity for contemporary national populations, as constructed by Ashraf and Galor (2013a), accounts for the expected heterozygosity of each sub-national group as well as the additional component of diversity at the country level that arises from the pairwise genetic distances amongst its precolonial ancestral populations.

The genetic diversity of contemporary national populations could therefore partly reflect historical processes such as interregional migrations that were, in turn, determined by historical patterns of civil conflicts, raising potential concerns about endogeneity issues that are present in the current indexes of ethnic fractionalization and polarization. Given that cross-country migrations in the post-1500 era significantly affected the distribution of genetic and cultural diversity, especially across countries in the New World, the measure of contemporary genetic diversity developed by Ashraf and Galor (2013a) is employed as the main independent variable. Moreover, since this measure may be spuriously correlated with the component of contemporary ethnolinguistic diversity that was shaped by the forces of colonialism and globalization over the past 500 years, the analysis employs the migratory distance of native (i.e., pre-1500) populations from East Africa as an instrument for contemporary genetic diversity in a global sample of countries. Exploiting the

highly significant negative first-stage relationship between genetic diversity and migratory distance from the cradle of humankind, this strategy mitigates concerns regarding the potential endogeneity between genetic diversity and civil conflicts.

In particular, the identification strategy appeals to two well established theories from the field of population genetics: the “Out of Africa” hypothesis and the serial founder effect. According to the well-established “Out of Africa” hypothesis, the human species, having evolved to its modern form in East Africa some 150,000 years ago, embarked on populating the entire globe in a stepwise migration process beginning about 70,000–90,000 BP. Furthermore, the contemporary distribution of diversity across indigenous ethnic groups reflects a serial founder effect originating in East Africa. Accordingly, since the populating of the world occurred in a series of stages where sub-groups left initial colonies to create new colonies further away, carrying with them only a portion of the genetic diversity of their parental colonies, contemporary genetic diversity in human populations decrease with increasing distance along prehistoric migratory paths from East Africa (e.g., Ramachandran et al., 2005; Prugnolle et al., 2005).

The baseline findings of the empirical analysis suggest that, conditional on the full set of controls for geographical and institutional covariates, as well as measures of economic development and variables reflecting the extent of ethnic fractionalization and polarization, a move from the 10<sup>th</sup> to the 90<sup>th</sup> percentile of the cross-country genetic diversity distribution leads, on average, to an increase in the number of new conflict onsets per year by 0.026 – an effect that is comparable to a shift from the 50<sup>th</sup> to the 90<sup>th</sup> percentile of the cross-country conflict distribution. In contrast, the conditional correlations of ethnic fractionalization and polarization with conflict are not significantly different than zero. Further, the partial  $R^2$  statistic associated with genetic diversity suggests that the residual cross-country variation in genetic diversity explains about 6.6 percent of the residual cross-country variation in civil conflicts, whereas the residual variations in ethnic fractionalization and polarization together explain only 0.7 percent of this variation.

## 2 Related Literature

The origin of civil conflicts has been the subject of intensive research in the past decades. The role of ethnic fractionalization and polarization in the onset of ethnic and civil conflicts has been at the center of this exploration, reflecting the prevailing hypothesis that competition for resources and political power across ethnic groups, along with conflicting preferences for public policies, would be harder to reconcile in highly fractionalized societies.

Nevertheless, the evidence that has emerged about the effect of ethnic and religious heterogeneity on the incidence and outbreak of civil conflicts has been largely inconclusive (Fearon and Laitin, 2003; Collier and Hoeffler, 2004). Some of these findings reflect the conceptual limitation of the fractionalization index, and indeed its replacement by the index of ethnic polarization has

generated superior although still somewhat fragile results (Montalvo and Reynal-Querol, 2005; Esteban et al., 2012).

Importantly, due to voluntary and forced migration, triggered by genocides, atrocities, and the destruction of wars, the degree of ethnic fractionalization and polarization within a country cannot be viewed as independent of past and present conflicts. Thus, the existing evidence suffers from endogeneity bias, and in particular, they are likely to capture the effects of reverse causality. In contrast to measures of ethnic linguistic fractionalization and polarization, the use of a measure of diversity that incorporates information on heterogeneity within and across groups, accounting explicitly for distances between ethnic groups, permits the current study to capture the effect of diversity (properly measured) on civil conflict, as well as to shed light on the causes of intra-group conflicts.

The current study also advances the literature on the relationship between inter-ethnic diversity and civil conflict. The existing literature features three distinct approaches. Primordialist explanations (e.g., Horowitz, 1985, 1999; Brewer, 1979, 1991, 1997) hold the view that ethnicity is deeply rooted in “primordial” notions of kinship and group belonging. According to this view, interethnic relations are charged with the potential for conflict, as group behavior is conditioned by old sources of enmity or the desire to dominate other ethnic groups. In contrast to the primordialist view, non-essentialist explanations of ethnic conflict (e.g., Caselli and Coleman, 2006; Esteban and Ray, 2008, 2011) argue that social conflict may manifest itself along ethnic or religious lines for pragmatic reasons, including but not limited to the use of ethnic networks as devices for effective monitoring, enforcement against free-riding, and easier access to financing.<sup>1</sup> Finally, the third view, advocated by modernist theories, stresses that ethnic conflict arises from increased intergroup competition for scarce resources, once groups that are excluded from social and political power experience economic modernization and thus begin to challenge the status quo (Gellner, 1983; Bates, 1986; Wimmer et al., 2009).

While the evidence provided in this paper is not necessarily relevant for the assessment of modernist theories, it is consistent with both primordialist and non-essentialist viewpoints. Genetic diversity may have functioned as a basis for endogenous group formation in the distant past, regardless of whether these groups today are mobilized by ethnic entrepreneurs trying to reinforce ethnic identities for their private interests or whether group identities entirely reflect primordial notions of kinship and belonging.<sup>2</sup>

---

<sup>1</sup>For example, Esteban and Ray (2008) study a model of coalition formation, showing that ethnic cleavages can be more salient than class-based cleavages when ethnic divisions generate a more suitable complementarity of finance and physical inputs into conflict. Eifert et al. (2010) offer some survey evidence from African countries that ethnic identities matter more when they are useful in the competition for political power.

<sup>2</sup>The hypothesis that genetic diversity contributed to ethnic heterogeneity is consistent with an evolutionary view of ethnic origins, mainly rooted in the sociobiological theory of ethnicity (Shils, 1957; Van den Berghe, 1981, 1995). Accordingly, human beings, like other mammals, exhibit nepotistic behavior, such as showing greater loyalty to their immediate family, extended family, or clan, because such behavior implies a higher likelihood of passing on one’s genes successfully to future generations. Thus, ethnicity is viewed as “extended nepotism,” and ethnic differentiation is sustained through cultural markers such as dialects, customs and traditions, norms of social conduct, and so on, which serve to distinguish members of the group from “others.”

Finally, this research contributes to an emerging literature on the genetic structure of human populations and socioeconomic outcomes (Spolaore and Wacziarg, 2009a; Ashraf and Galor, 2013a).<sup>3</sup> In a related paper, Spolaore and Wacziarg (2009b) offer empirical evidence that genetically closer nations have a higher propensity to engage in interstate war. The authors argue that genetic relatedness between two nations reflects greater similarity in preferences over private (rival and excludable) goods, thereby augmenting the likelihood of interstate conflict over such goods. In contrast, the current study examines the role of diversity in national populations for intrastate conflict, where issues of contention, in part, arise due to heterogeneity in preferences over public (rather than private) goods.

### 3 Data and Empirical Framework

This section discusses the empirical strategy used for identifying the reduced-form causal effect of genetic diversity on various conflict outcomes. It first describes the key variables employed by the analysis and then introduces the empirical models that are estimated by the cross-country and panel data regressions.

#### 3.1 Data

##### 3.1.1 Dependent Variables

The main source of data on civil conflicts is the UCDP/PRIO Armed Conflict Dataset (Version 4–2012). As a baseline, the empirical analysis employs the PRIO25 coding of civil conflicts, which is the standard used in the literature. This coding is based on internal armed conflict episodes (including internationalized internal armed conflicts). Each of these episodes comprises one or more years of conflict between the government of a state and one or more internal opposition group(s), in which the number of annual battle-related deaths exceeded 25.

The second measure of conflict employed by the analysis reflects *ethnic* conflicts, as classified by Wimmer et al. (2009) (henceforth WCM 2009). Using the UCDP/PRIO Armed Conflict Dataset as their primary source, these authors apply to these data an ethnic-conflict categorization wherein the opposition groups in a civil conflict explicitly pursue ethnonationalist aims (e.g., attempt to secure ethnonational self-determination, ethnoregional autonomy, or language and other cultural rights) or are motivated by ethnicity-based concerns (e.g., ethnic balance of power in the government, ethnic and racial discrimination), and they recruit fighters and forge political alliances on the basis of ethnic affiliation.

---

<sup>3</sup>See also Desmet et al. (2011) who demonstrate a strong correlation between genetic and cultural distances among European populations to argue that genetic distance can be employed as an appropriate proxy to study the effect of cultural distance on the formation of new political borders in Europe. In addition, Guiso et al. (2009) employ data on genetic distance between European populations as an instrument for measures of trust to estimate its effect on the volume of bilateral trade and foreign direct investment.

Following the common practice in the literature, the analysis focuses on conflict events in the post-1960 period, by which point most of the previous colonies had become independent nation states. Due to constraints on the availability of data for some of the baseline control variables, the sample period for the civil-conflict regressions is 1960–2008. Moreover, since WCM 2009 employ an earlier version (Version 3–2005b) of the UCDP/PRIO Armed Conflict Dataset, the sample period covered by the ethnic-conflict regressions is limited to the 1960–2005 time span.

Depending on the unit of analysis, the dependent variables that are employed capture different dimensions of conflict potential. In the cross-country regressions, the outcome variables record the annual average number of distinct (new) conflicts – rather than the annual average number of conflict episodes – over the relevant sample period. These variables therefore reflect how many different incompatibilities between state actors and armed opposition groups have escalated to a conflict on a yearly basis.

Many conflicts, however, consist of multiple conflict episodes that are often separated by one or more years of inactivity. In panel data regressions, the analysis exploits this temporal dimension of conflict. Specifically, in the conflict incidence (prevalence) regressions, the dependent variable is an indicator, coded 1 for each country-period (either a 1-year or 5-year time span) in which there is at least one active conflict episode observed, and 0 otherwise. The conflict onset regressions, on the other hand, employ one of two different indicators of conflict outbreak. Specifically, for any given country-year, the first outcome measure is coded 1 if and only if a new conflict started that year, whereas the second measure is the standard PRIO2 onset variable, which is coded 1 whenever a (possibly recurrent) conflict episode erupts after at least two years of civil peace.

### 3.1.2 Independent Variable: Genetic Diversity

As with standard measures of ethnolinguistic fractionalization and polarization, *observed* genetic diversity may potentially be endogenous in an empirical model of civil conflict, since it could be tainted by genetic admixtures resulting from movements of populations across space in response to spatial differences in the prevalence of conflict, the nature of political institutions, and the level of economic prosperity.

To circumvent this problem, the current study employs a measure of contemporary genetic diversity introduced by Ashraf and Galor (2013a). Specifically, the measure captures (amongst other dimensions of diversity at the national level, as explained below) the component of intra-group diversity associated with the indigenous groups of a country that is *predicted* by migratory distance from Addis Ababa, Ethiopia to the country’s modern capital city along prehistoric land-connected human migration routes.<sup>4</sup> Exploiting the explanatory power of a serial-founder effect associated with the prehistoric “out of Africa” migration process, the predicted genetic diversity for the native

---

<sup>4</sup>These routes incorporate five obligatory intermediate waypoints: Cairo, Egypt; Istanbul, Turkey; Phnom Penh, Cambodia; Anadyr, Russia; and Prince Rupert, Canada. Rather than using aerial distance from East Africa, the use of these intermediate waypoints ensures that the distance measure more accurately reflects the fact that, in the course of their prehistoric exodus from Africa, humans did not cross large bodies of water.



population of a country is generated by applying the coefficients obtained from an ethnic group-level regression (e.g., Ramachandran et al., 2005; Prugnolle et al., 2005) of observed intra-group genetic diversity on migratory distance from Addis Ababa, Ethiopia in a sample of 53 ethnic groups from the Human Genome Diversity Cell Line Panel, compiled by Human Genome Diversity Project-Centre d’Etudes du Polymorphisme Humain (HGDP-CEPH). According to anthropologists, these 53 ethnic groups have been historically native to their current geographical locations and have also been largely isolated from genetic flows from other ethnic groups.<sup>5</sup>

The observed genetic diversity for each of the 53 HGDP-CEPH ethnic groups is based on the index of *expected heterozygosity*. Like most other measures of population diversity, this index can be interpreted as the probability that two individuals, selected at random from the relevant population, are genetically different from one another with respect to a given spectrum of genetic traits. The index is constructed by geneticists using sample data on allelic frequencies; i.e., the frequency with which a gene variant or allele (e.g., the brown vs. blue variant for the eye color gene) occurs in the population sample.<sup>6</sup> Given allelic frequencies for a particular gene or DNA locus, it is possible to compute a gene-specific heterozygosity statistic (i.e., the probability that two randomly-selected individuals differ with respect to the gene in question), which when averaged across multiple genes, yields the overall expected heterozygosity for the relevant population. More formally, consider the case of a single gene or locus  $l$  with  $k_l$  observed variants or alleles in the population, and let  $p_i$  denote the frequency of the  $i^{th}$  allele. Then, the expected heterozygosity of the population with respect to locus  $l$  is:

$$H_{exp}^l = 1 - \sum_{i=1}^{k_l} p_i^2, \quad (1)$$

which, given allelic frequencies for each of  $m$  different genes or loci, can be averaged across these loci to yield an aggregate expected heterozygosity measure of overall genetic diversity as:

$$H_{exp} = 1 - \frac{1}{m} \sum_{l=1}^m \sum_{i=1}^{k_l} p_i^2. \quad (2)$$

In the absence of systematic and large-scale population movements across geographically distant regions, the genetic diversity of native populations in a given location is a good proxy for the contemporary genetic diversity of that location. While this is to a large extent true for nations in the Old World (i.e., Africa, Europe, and Asia), post-1500 population flows from the

---

<sup>5</sup>For a more detailed description of the HGDP-CEPH Human Genome Diversity Cell Line Panel dataset, the interested reader is referred to Cann et al. (2002). A broad overview of the Human Genome Diversity Project is given by Cavalli-Sforza (2005).

<sup>6</sup>In molecular genetics, an allele is defined as any one of a number of viable DNA codings (formally, a sequence of nucleotides) that occupy a given locus (or position) in a chromosome. Chromosomes themselves are packages for carrying strands of DNA molecules in cells, and they comprise multiple loci that typically correspond to some of the observed discrete units of heredity (or genes) in living organisms. For further elaboration on basic concepts and definitions in genetics, the interested reader is referred to Griffiths et al. (2000).

Old World to the New World has had a considerable impact on the ethnic composition and thus the genetic diversity of national populations in the Americas and Oceania. Rather than limiting the investigation to an Old-World sample of countries, however, the analysis adopts the ancestry-adjusted genetic diversity measure of Ashraf and Galor (2013a) as the independent variable of interest. Using the shares of different groups in the national population, this measure not only accounts for the genetic diversity of sub-national groups that can trace native ancestry to the year 1500 and the diversity of those descended from immigrant settlers, but it also accounts for the additional component of diversity at the national level that arises from the pairwise genetic distances amongst these different sub-national groups.<sup>7</sup> The ancestry-adjusted measure of genetic diversity, however, is subject to potential endogeneity in an empirical model of civil conflict, as it may partly reflect endogenous cross-country migrations in the post-1500 era and their impact on the contemporary ethnic composition of national populations. Section 3.2 describes how this problem is addressed.

### 3.1.3 Control Variables

While the current study documents the previously unexplored role of intra-population genetic diversity as a fundamental determinant of civil and ethnic conflicts, a vast empirical literature has highlighted several other factors. To account for the potentially confounding role of these other determinants, the baseline model includes several variables that commonly feature in empirical studies of the determinants of civil conflict.

**Fractionalization and Polarization.** Whether and how conflicts are related to ethnolinguistic heterogeneity at the national level has been a subject of much debate in the empirical civil conflict literature. Recently, Ashraf and Galor (2013b) have found that genetic diversity is an underlying cause of various existing manifestations of ethnolinguistic fragmentation. This finding suggests that ethnolinguistic heterogeneity could be one of the channels through which genetic diversity influences the potential for conflict, although the residual variation in genetic diversity that is not manifested in measures of ethnolinguistic fragmentation could continue to play a role in explaining civil conflicts through channels associated with inequality, interpersonal trust, and heterogeneity in preferences over public goods and redistributive policies. To investigate this possibility, the analysis controls for the ethnic fractionalization index of Alesina et al. (2003). This index is a commonly used proxy for ethnic fragmentation, measuring the probability that two people, randomly selected from a country’s population, belongs to different ethnic groups. Holding the number of ethnic groups

---

<sup>7</sup>The data on population shares of different groups at the country level are obtained from the World Migration Matrix, 1500-2000 of Putterman and Weil (2010), who compile for each country in their dataset, the share of the country’s population in 2000 that is descended from the population of every other country in 1500. For an in-depth discussion on the construction of the ancestry-adjusted measure of genetic diversity, the reader is referred to Ashraf and Galor (2013a).

constant, the fractionalization index attains its highest value when the population is uniformly distributed across these groups.

There is no compelling theoretical basis, however, for such a configuration of ethnic groups to maximize the potential for conflict in society. Consequently, owing to its more appealing theoretical foundations and stronger predictive power in empirical studies of conflict (Montalvo and Reynal-Querol, 2005; Esteban et al., 2012), the current analysis additionally controls for an index of ethnolinguistic polarization as an alternative measure of diversity. Polarization indices measure the extent to which the composition of groups in a population resembles a perfectly polarized configuration, occurring when the population consists of two groups with equal size. Due to its more comprehensive geographical coverage, the analysis controls for a measure of polarization across linguistic groups in the population from the dataset of Desmet et al. (2012), who provide several such indices constructed at different levels of linguistic aggregation, based on hierarchical language trees listed in the 15th edition of *Ethnologue*. Specifically, the employed measure is ethnolinguistic polarization at the most disaggregated level (level 15). As such, it reflects polarization across contemporary linguistic groups and is conceptually comparable to other polarization indices used in the literature.

**Geographical Factors.** As explained above, the genetic diversity measure exploited by the current analysis is highly correlated with migratory distance from East Africa. This naturally raises the need to control for geographical attributes that may also be correlated with migratory distance and that can reasonably have an impact on conflict through channels unrelated to intra-population diversity. To ensure that genetic diversity is not simply capturing an indirect effect of distance from the equator on conflict (e.g., through a climate-related channel), absolute latitude is included in the regressions as one of the baseline geographical controls. Similarly, mean distance to the nearest waterway is introduced as a control variable to account for its potentially confounding effect on conflict (e.g., through its long-run impact on economic development).

Total land area is also included in the baseline empirical specification to address the concern that, all else equal, larger areas enclosed by national borders may be more conflict prone due to a positive correlation between country size and population size, on the one hand, and a potentially more diverse population in larger geographic regions on the other. Controlling for land area also accounts for the possibility that the measure of genetic diversity, based on migratory distance from East Africa to the capital cities of modern nation states, may be less comparable across countries of different sizes.

The topographical features of a country may also affect the likelihood of civil conflict independently of intra-population diversity. Rough terrains may provide safe havens for rebels and enable them to sustain continued resistance by protecting them from numerically and militarily superior government forces. To the extent that the prevalence of rough terrains is correlated with migratory distance from East Africa, one could be wrongly attributing the effect of terrain

ruggedness to the genetic diversity channel. To address this possibility, the baseline specification includes a control for terrain ruggedness.

Michalopoulos (2012) shows that regions characterized by greater dispersion in land quality and elevation exhibit higher levels of ethnolinguistic diversity, conditional on the average levels of these geographical attributes. Although measures of ethnolinguistic heterogeneity are directly controlled for by the analysis, the aforementioned geographical variables may still explain some unobserved component of diversity in ethnic and cultural traits, thereby exerting some latent influence on conflict potential. To account for such effects, the baseline empirical model controls for the mean and the range in values of both land quality and elevation.<sup>8</sup>

In the process of the prehistoric demic expansion of humans from Africa to the rest of the world, population bottlenecks that occurred at specific inter-continental crossings – a classic example is the Bering Strait – led to discrete spatial differences in observed genetic diversity across these crossings. Therefore, the worldwide variation in genetic diversity exhibits what population geneticists refer to as punctuated “clines,” raising the possibility that the cross-continental variation in genetic diversity may be correlated with cross-continental variation in unobserved – or observed but imprecisely measured – continent-specific characteristics. To ensure that the estimated effect of genetic diversity on conflict potential is not simply reflecting the latent influence of these time-invariant continent-specific cultural, institutional, and geographical factors, a complete set of continent dummies is included as a standard control throughout the analysis.<sup>9</sup>

**Institutional Factors and Development Outcomes.** In addition to the control variables mentioned above, the baseline model considers various proximate determinants of civil conflict that commonly feature in the literature. Specifically, several institutional, demographic, and economic factors may be responsible for the cross-country variation in conflict potential.

A widely accepted hypothesis is that colonial rule may have significantly shaped politics and ethnic divisions in newly-independent states. In particular, the legacy of the colonial period and the identity of former colonizers may have important ramifications for contemporary economic and political institutions, thereby affecting the potential for civil conflict. As such, three control variables are included in the baseline model to account for the legacy of colonialism. Depending on the unit of analysis, these variables reflect either the time-varying presence (in panel data) or the historical presence (in cross-sectional data) of colonial rule by France, the U.K., and any other colonial power. Time-invariant dummies for French and British legal origins are also included in the

---

<sup>8</sup>The country-level aggregate data from Michalopoulos (2012), on the mean of and dispersion in land quality, are based on disaggregated geospatial data on land quality from Ramankutty et al. (2002). The data on elevation come from Nordhaus (2006).

<sup>9</sup>This issue is perhaps more severe for cultural and institutional factors because of the possibility of systematic measurement error at the continent level in variables reflecting cultural and institutional characteristics. The most credible way of addressing all these concerns is to include continent fixed effects throughout the analysis, in order to “soak up” the possibility of both omitted variable bias and the bias arising from systematic measurement error in potentially correlated controls.

model to account for any direct role of legal codes and institutions that may not be fully captured by the measures of colonial experience.

Although many studies find the level of democracy to be an insignificant predictor of civil conflict (e.g., Collier and Hoeffler, 2004; Fearon and Laitin, 2003), some point to a negative association between conflict and democracy (e.g., Esty et al., 1998; Gurr, 2000). There is also evidence that anocratic (hybrid) political regimes tend to face a higher risk of conflict (Fearon and Laitin, 2003). This may be indicative of the notion that weak political legitimacy, and the political grievances associated with it, can lead to more conflicts when the state apparatus is less repressive (as is typically the case for hybrid regimes) than a full-fledged autocratic regime. The baseline model employs three control variables to account for such effects. Two of these capture the type of political regime, reflecting the prevalence of either democracy (a polity score above 5) or autocracy (a polity score below -5) over the time interval covered by each country, with the interval representing either a 1-year or 5-year period in the panel data analysis or the entire sample period in the cross-country analysis.<sup>10</sup> The third control variable is an ordinal index measuring the degree of constraints on the discretionary power of the chief executive, which again, is allowed to vary over time in the panel data analysis but reflects the sample-period average in the cross-country regressions.

Several scholars have argued that oil wealth could foster conflict by funding rebel groups (e.g., Collier and Hoeffler, 2004; Ross, 2006), weakening state institutions, making the state a more attractive target for rebels (e.g., Fearon and Laitin, 2003), or facilitating trade shocks (e.g., Humphreys, 2005). To account for this channel, the per capita value of oil production (in 2000 constant dollars) is included as a baseline control in the model, and it is permitted to vary over time in the panel data analysis but represents the temporal average over the entire sample period in the cross-country regressions. Data on oil wealth comes from the 2013 release of the *Global dataset of oil and gas production and exports (version 2)*, compiled by Ross.

All civil conflict codings are, by necessity, based on a certain death threshold. Most studies control for population size, on the grounds that the scale of conflicts and the resulting number of deaths may be mechanically affected by population size. In addition, all else equal, a larger population may imply a larger recruitment pool for rebels (Fearon and Laitin, 2003). Logged population is therefore introduced as a control in the baseline specification (and allowed to be time-varying in the panel data analysis), although the estimated coefficient associated with this covariate is highly likely to suffer from endogeneity bias.

Income per capita is yet another standard control in civil conflict regressions, motivated by several arguments. One theory is that it proxies for the overall capacity of the state to prevent or fight rebel insurgencies (Fearon and Laitin, 2003). Another notion is that it reflects the opportunity cost for potential fighters to engage in a civil war (Collier and Hoeffler, 1998, 2004). To the extent that a higher standard of living is reflected in rising state revenues, higher income per capita may

---

<sup>10</sup>Anocracy, with a polity score between -5 and 5, serves as the omitted benchmark category.

also incentivize attempts at state capture by armed groups. To account for these effects of the standard of living on conflict potential, the analysis employs data on real income per capita from Maddison’s Historical Statistics (February 2010 version), controlling for time-varying income per capita in the panel data regressions and average per-capita income (over the relevant sample period) in the cross-country regressions.

Appendix B presents the summary statistics for all the variables employed by the baseline model. In addition, detailed information on variable definitions and data sources is provided in Appendix C. Since many of the control variables are endogenous to civil conflict, their estimated coefficients do not permit a causal interpretation. Nonetheless, controlling for these factors is essential for assessing the extent to which the estimated effect of genetic diversity on civil conflict can be attributed to other explanations highlighted in the previous literature.

### 3.2 Empirical Framework

The empirical analysis estimates various specifications to investigate the explanatory power of genetic diversity for three different dimensions of civil conflict in the post-1960 time period. In particular, the cross-country regressions attempt to explain the variation across countries in the frequency of new conflicts, i.e., in the average number of new conflict outbreaks per year. This measure, as mentioned previously, reflects the number of different incompatibilities that have escalated to a full-blown conflict at some point during the sample period. The baseline specification for the cross-country analysis is given by:

$$\ln(1 + ANC_i) = \beta_0 + \beta_1 \hat{GD}_i + \beta_2 EF_i + \beta_3 EP_i + \beta'_4 GEO_i + \beta'_5 INS_i + \beta'_6 DEV_i + \varepsilon_i, \quad (3)$$

where  $ANC_i$  is the average number of new civil conflicts per year in country  $i$ ;  $\hat{GD}_i$  is the ancestry-adjusted measure of genetic diversity;  $EF_i$  and  $EP_i$  are the indices of ethnic fractionalization and ethnolinguistic polarization, respectively;  $GEO_i$ ,  $INS_i$ , and  $DEV_i$  are the respective vectors of control variables for geographical conditions, institutional attributes, and development outcomes, as described in section 3.1.3; and finally,  $\varepsilon_i$  is a country-specific disturbance term. All time-varying right-hand-side variables (i.e., the controls for institutional attributes and development outcomes) reflect their temporal averages over the relevant sample period: 1960–2008 for the PRIO civil conflict regressions and 1960–2005 for the WCM 2009 ethnic conflict regressions.

The baseline empirical model is first estimated using the OLS estimator on a global sample of 143 countries. Since the ancestry-adjusted measure of genetic diversity,  $\hat{GD}_i$ , accounts for the contribution by settlers from the Old World to the diversity of contemporary national populations in the New World, as a result of cross-country migrations in the post-1500 era,  $\hat{GD}_i$  may be subject to a similar endogeneity problem as are the standard measures of ethnic fractionalization and polarization. To mitigate this issue, the baseline cross-country regressions are repeated using the 2SLS estimator, employing the migratory distance of native populations from East Africa as

an instrument for contemporary genetic diversity. The identifying assumption therefore is that migratory distance of native populations is exogenous to conflicts in the post-1960 era, plausibly satisfying the exclusion restriction, conditional on the comprehensive set of controls for geographical and institutional factors as well as outcomes of economic development.

The second dimension of conflict examined by the empirical analysis is the prevalence of civil conflict. Exploiting the panel structure of the data, the analysis investigates the predictive power of genetic diversity for the incidence of an active conflict episode in a given period. Unlike the cross-country analysis, the panel data incidence regressions capture the role of diversity in explaining the duration of conflict, and as such, they do not distinguish between new conflicts – i.e., those associated with a previously unobserved incompatibility – and subsequent episodes of existing conflicts. The following probit specification is adopted for the baseline conflict incidence (or prevalence) regressions:

$$\begin{aligned} CP_{i,t}^* &= \gamma_0 + \gamma_1 \hat{GD}_i + \gamma_2 EF_i + \gamma_3 EP_i + \gamma_4 C_{i,t-1} + \gamma_5' GEO_i + \gamma_6' INS_{i,t-1} + \gamma_7' DEV_{i,t-1} \\ &\quad + \gamma_8' \delta_t + \eta_{i,t} \\ &\equiv \gamma' Z_{i,t} + \eta_{i,t}, \end{aligned} \tag{4}$$

$$C_{i,t} = \mathbf{1}[CP_{i,t}^* \geq D^*], \tag{5}$$

$$Pr(C_{i,t} = 1 | Z_{i,t}) = Pr(CP_{i,t}^* \geq D^* | Z_{i,t}) = \Phi(\gamma' Z_{i,t} - D^*), \tag{6}$$

where  $CP_{i,t}^*$  is the latent variable measuring the potential for conflict in country  $i$  during the 5-year period  $t$ , and it is modeled as a linear function of explanatory variables. Further, the right-hand-side variables  $\hat{GD}_i$ ,  $EF_i$ ,  $EP_i$ , and  $GEO_i$  are as previously defined, but in this case, the time-varying covariates included in  $INS_{i,t-1}$  and  $DEV_{i,t-1}$  correspond to their respective averages over the previous 5-year period,  $\delta_t$  is a vector of period dummies, and  $\eta_{i,t}$  is a country-period-specific disturbance term.<sup>11</sup> Controlling for one-period lags of the time-varying measures, instead of their current values, aims to mitigate the concern that the use contemporaneous measures of time-varying covariates may exacerbate reverse-causality bias in their estimated coefficients.<sup>12</sup> Finally, conflict potential is assumed to also depend on the lagged incidence of conflict,  $C_{i,t-1}$ , which accounts for the possibility that, all else equal, countries with recent conflict experience may exhibit a higher conflict potential in the current period, mainly because of the intertemporal spill-overs that are

<sup>11</sup>The analysis also confirms the robustness of the conflict incidence regression results to using observations at an annual frequency. Naturally, in that analysis, time-dependent covariates are lagged yearly values instead of the lagged 5-year averages, and the time fixed effects represent year dummies.

<sup>12</sup>An alternative method to address the reverse-causality problem in a 5-year panel context is to control for values of the time-varying covariates as measured in the initial year of each period. Although this method would retain the first-period observation for each country, which is dropped under the current specification, it leaves open the possibility that the presence or lack of a conflict in the first year of each period may still exert a direct influence on the right-hand-side variables.

common to most conflict processes – e.g., the self-reinforcing nature of past casualties on both sides of the conflict.<sup>13</sup>

Because latent conflict potential,  $CP_{i,t}^*$ , is unobserved, its level can only be inferred from the binary conflict incidence variable,  $C_{i,t}$ , which indicates whether conflict potential was high enough for the death threshold for a civil conflict to be exceeded in a given period. As is evident from equation 5,  $D^*$  is the corresponding threshold for unobserved conflict potential, and it appears as an intercept in  $\Phi(\cdot)$ , the cumulative distribution function for the country-period-specific disturbance term,  $\eta_{i,t}$ . Due to potential endogeneity of the measure of contemporary genetic diversity, the model is estimated using the instrumental-variables maximum-likelihood probit estimator (IV probit MLE).<sup>14</sup> As in the cross-country analysis, ancestry-adjusted genetic diversity is instrumented with the migratory distance of native populations from East Africa.

The final dimension of conflict that is examined is the onset of civil conflict. Unlike the incidence regressions, the onset model distinguishes between the outbreak of conflict and the continuation of conflict. Specifically, this model focuses solely on the former event, classifying the latter as a non-event akin to years of civil peace. Conceptually, the onset model assesses how genetic diversity affects the degree of instability as measured by the frequency of transitions from peaceful episodes to conflict and vice versa. The baseline model for the onset analysis is similar to the incidence model described by equations (4)-(6), except that now, following the convention in the literature, (i) an annual panel is employed, with the outcome variable taking a value of 1 if the country-year coincides with the first year of a new period of conflict, and 0 otherwise; (ii) lagged conflict incidence is not included in the set of covariates; and (iii) a cubic spline function in peace duration is included in the right-hand side, along with year dummies, to account for temporal dependence (Beck et al., 1998).

What constitutes a new period of conflict depends on the specific onset measure that is employed. The onset analysis explores two different measures. Under the PRIO2 measure, a new period of conflict could be initiated either by a new episode of a recurring (old) conflict or by the onset of an entirely new conflict, as long as this conflict period is preceded by at least 2 years of uninterrupted civil peace. The PRIO-NC measure, on the other hand, is coded 1 for each country-year that coincides with the beginning of an entirely new conflict, regardless of whether this country-year is preceded by an active-conflict year. As in the incidence regressions, the IV probit MLE estimator is employed to identify the effect on conflict onset of ancestry-adjusted genetic diversity, instrumented by migratory distance of native populations from East Africa.

Finally, despite the panel structure of the data, and similar to previous empirical studies on the relationship between conflict and ethnolinguistic heterogeneity, the analyses for both conflict

---

<sup>13</sup>In adopting this strategy, the analysis follows Esteban et al. (2012), who also include the lagged value of the dependent variable as a standard control in their incidence regressions.

<sup>14</sup>In all panel data regressions, robust standard errors of the estimated coefficients are clustered at the country level.



incidence and conflict onset are unable to account for unobserved country-fixed effects, since the measure of genetic diversity is time invariant.

## 4 Empirical Results

This section presents the empirical findings regarding the reduced-form causal effect of genetic diversity on several measures of intrastate conflict. First, the findings from cross-country regressions are presented, and both the number of civil conflicts and that of ethnic conflicts are shown to depend positively on genetic diversity. The discussion then proceeds to the results from regressions exploiting a panel sample of 5-year periods to demonstrate the explanatory power of genetic diversity for the incidence of conflict. These results are shown to be robust to alternative codings of civil and ethnic conflict, as well as conflict between non-state actors and intra-group communal conflict. Next, robustness to annual conflict incidence and onset is demonstrated. The section concludes with the investigation of two potential channels mediating the reduced-form causal effect of genetic diversity on intrastate conflict.

### 4.1 Basic Cross-country Analysis

Before discussing the baseline regressions, Table 1 documents how the bivariate relationships between the annual average number of PRIO25 civil conflicts and various measures of diversity compare with one another. The dependent variable is the natural logarithm of one plus the aforementioned count of distinct conflicts. Genetic diversity appears as a positive and highly significant predictor of civil conflict. The estimated coefficient suggests that a move from 10<sup>th</sup> to 90<sup>th</sup> percentile of the observed diversity distribution is associated with an increase of 0.014 in the average number of new conflicts per year, a relationship that is statistically significant at the 1% level. Given that the cross-country average of the annual frequency of new conflict onsets is around 0.02, the estimated magnitude corresponds to an increase by 70% relative to the cross-country mean. Ethnic and linguistic fractionalization also enter their respective bivariate regressions with a statistically significant positive coefficient, while civil conflict does not appear to be significantly related to religious fractionalization and ethnolinguistic polarization in the absence of other conditioning variables.

In Table 2, the analysis proceeds with regressions that conduct “horse races” between genetic diversity and various combinations of fractionalization and polarization indices. The first column reproduces the unconditional relationship between genetic diversity and conflict, and in the following columns, other diversity indices are introduced – initially, one at a time, and then in pairs. In the final column, all diversity measures, including ethnolinguistic polarization, are introduced simultaneously. The results indicate that the relationship between genetic diversity and civil conflict does not vanish when its potential influence via ethnic, linguistic, or religious heterogeneity is accounted for. Specifically, genetic diversity retains its positive and statistically

significant coefficient throughout the analysis, although when conditioned on ethnic and linguistic fractionalization, genetic diversity is significant at the 5% level, as compared to the 1%-significance level of its unconditional relationship with conflict. Moreover, the estimated coefficients on genetic diversity are somewhat smaller in magnitude, suggesting that ethnic and linguistic fractionalization might be capturing an important – but not the only – mediating channel behind the deep influence of genetic diversity on conflict.<sup>15</sup>

Table 3 presents the results from estimating different variants of the baseline cross-country specification using the OLS estimator. The analysis controls for a successively expanding set of covariates to obtain the full baseline model in column 7. As before, column 1 replicates the bivariate regression on the baseline cross-country sample to provide a comparable benchmark for the remaining specifications. Column 2 introduces all baseline geographical controls, including the continent dummies. In column 3, the analysis additionally controls for political and institutional factors. Columns 4 through 6 check whether accounting for the influence of ethnic fractionalization and polarization on conflict has any qualitative bearing on the coefficient associated with genetic diversity, conditional on all previously mentioned factors. Finally, the last column controls for income and population size.

The results show that, throughout the analysis in Table 3, genetic diversity remains a statistically significant and positive predictor of the average number of civil conflicts per year. The coefficient on genetic diversity is significant at the 5% level across columns, and once conditioned on the baseline set of geographical controls, the relationship becomes stronger in magnitude. This relationship thereafter remains qualitatively intact and quantitatively stable as the remaining sets of covariates are included. Conditional on the full set of controls, a move from the 10<sup>th</sup> to the 90<sup>th</sup> percentile of the ancestry-adjusted genetic diversity distribution leads, on average, to an increase in the number of new conflict onsets per year by 0.026 – an effect that is comparable to a shift from the 50<sup>th</sup> to the 90<sup>th</sup> percentile of the cross-country conflict distribution.<sup>16</sup>

The adjusted- $R^2$  statistic indicates that the baseline model explains about 24 percent of the total variation in the number of conflict onsets per year. Further, the squared partial correlations reported in the bottom two rows suggest that genetic diversity *by itself* explains about 6.6 percent of the cross-country variation in civil conflicts, an explanatory power that cannot be attributed to any of the other right-hand-side variables. In contrast, ethnic fractionalization and polarization combined explain only 0.7 percent of the residual variation in conflict after the partial correlations of the remaining covariates are netted out.

---

<sup>15</sup>The results from a more systematic investigation of this channel are presented in Tables 9 and 10 and will be discussed later on.

<sup>16</sup>The scatter-plot in Figure 1 shows the relationship between genetic diversity and civil conflict, conditional on the full set of baseline controls.

Table A.1 in the appendix repeats the analysis from Table 3, using the average number of ethnic (rather than overall) conflict onsets per year as the dependent variable. Reassuringly, the results are similar to those obtained for overall civil conflict.<sup>17</sup>

**Addressing Endogeneity.** The ancestry adjustment applied to the genetic diversity measure is essential to capture the contribution of cross-country migrations in the post-1500 era to the genetic diversity of contemporary national populations. As previously argued, however, these population flows may have been affected by past conflicts or other historical circumstances that also contribute to civil conflicts in the modern world. Therefore, while the OLS estimates in Table 3 serve as a useful benchmark for the relationship between genetic diversity and conflict, they may suffer from endogeneity bias.

To address this potential endogeneity problem, Table 4 replicates the regressions from Table 3 using the 2SLS estimator, instrumenting ancestry-adjusted genetic diversity with migratory distance of native populations from Addis Ababa, Ethiopia. The estimated 2SLS coefficients on genetic diversity are substantially larger in magnitude and stronger in statistical significance, in comparison to the relevant OLS coefficients from Table 3. Specifically, in the 2SLS case, the estimated effect of genetic diversity under the baseline model in column 7 is significant at the 1% level, and it is twice as large as that implied by the corresponding OLS estimate – the coefficient of interest suggests that a move from the 10<sup>th</sup> to the 90<sup>th</sup> percentile of the ancestry-adjusted genetic diversity distribution leads, on average, to an increase in the number of new conflict onsets per year by 0.052, which corresponds to the difference between the top and bottom deciles of the cross-country conflict distribution.

Table A.2 in the appendix repeats the 2SLS analysis from Table 4, using the average number of ethnic conflicts per year as the outcome variable. As in the OLS analysis, the 2SLS results for ethnic conflict are reassuringly similar to those obtained for overall civil conflict.<sup>18</sup>

**Baseline Analysis Conditional on Linguistic Fractionalization.** The baseline cross-country analysis thus far has adopted the ethnic fractionalization index of Alesina et al. (2003) as the preferred control for fractionalization across sub-national groups in the population, on the grounds

---

<sup>17</sup>The baseline relationship between ethnic conflict and genetic diversity is depicted in on the scatter-plot in Figure A.1. To better assess the relationship of genetic diversity with ethnic conflicts, in comparison to its relationship with overall civil conflicts, columns 1 through 7 in Table A.3 repeat the OLS analysis from Table A.1, using the WCM 2009 coding of overall civil conflicts as the outcome variable. The association of diversity with the number of ethnic conflicts per year is consistently larger, when diversity is increased from the 10<sup>th</sup> to the 90<sup>th</sup> percentile, despite the fact that ethnic conflicts are by definition more rare and, hence, have a lower cross-country sample mean. These results potentially point to a more influential role for genetic diversity in explaining the variation in ethnic conflicts rather than in non-ethnic conflicts.

<sup>18</sup>Once again, to permit a fair comparison of the effects of genetic diversity on ethnic conflicts versus overall civil conflicts, columns 8 to 14 in Table A.3 repeat the 2SLS analysis from Table A.2, using the WCM 2009 coding of overall civil conflicts as the outcome variable. As in the OLS case, the estimated effects on ethnic conflicts are consistently more pronounced than on overall civil conflicts, when diversity is increased from the 10<sup>th</sup> to the 90<sup>th</sup> percentile, indicating that genetic diversity potentially exerts a stronger influence on the potential for ethnic conflicts than on that for non-ethnic conflicts.

that it is the primary index used in empirical work on civil conflict (to proxy for ethnic and cultural fragmentation). Yet, the careful reader might have noticed that, in the bivariate and “horse race” regressions from Tables 1 and 2, the linguistic fractionalization index of Alesina et al. (2003) appears to be somewhat more influential than the commonly used ethnic fractionalization index of the same authors. Although the two alternative measures are highly correlated, before moving on with the latter and more standard measure in the rest of the analysis, it is important to check if the OLS and 2SLS results presented in Tables 3 and 4 are in any way sensitive to the selection of the ethnic (rather than linguistic) fractionalization index as a baseline control. The robustness results, reported in Table A.4, suggest that the coefficient on genetic diversity, when conditioned on linguistic instead of ethnic fractionalization, is in fact not only statistically more significant but also slightly larger in magnitude.

**Robustness to the Elimination of Regions.** As is customary in cross-country empirical studies of civil conflict, the analysis also investigates whether the baseline results are driven by influential observations in a particular region of the world. For this purpose, in Table 5, the analysis drops one geographical region at a time from the regression sample and estimates the baseline model on the remaining sample of countries.<sup>19</sup> Reassuringly, the table reveals that the key findings from the baseline analysis are not qualitatively sensitive to the exclusion of any particular region from the sample.<sup>20</sup>

## 4.2 Explaining Conflict Incidence

Table 6 reports the main conflict incidence regressions, using a panel data structure with 5-year periods to exploit both temporal and cross-country variations in the prevalence of active conflict episodes. The first six columns report the results for PRIO25 civil conflict incidence, and the results for WCM 2009 ethnic conflict incidence are presented in the remaining columns. Throughout the analysis, specifications are estimated using the IV probit MLE estimator, with ancestry-adjusted genetic diversity being instrumented by the migratory distance of native populations from East Africa. Further, in all regressions, robust standard errors are clustered at the country level.

As is evident from the table, not surprisingly, lagged conflict incidence possesses strong predictive power for current conflict. Nevertheless, conditional on whether they experienced a conflict in any given 5-year period, those countries with genetically more diverse populations are significantly more likely to see a conflict in the next 5-year period. For both outcome variables examined, the estimated coefficients on genetic diversity are statistically significant at the 5% level or less, and their magnitudes are rather stable across different specifications.

<sup>19</sup>The eliminated geographical regions are Sub-Saharan Africa, the African continent, Latin America and the Caribbean (LAAC), the Asian continent, Middle East and North Africa (MENA), the European continent, South Asia, and East Asia and the Pacific. Continent dummies are excluded from these regressions in order to permit exploitation of the entire cross-country variation in observables, given the reduction in degrees of freedom from eliminating observations.

<sup>20</sup>Table A.5 presents the results for ethnic conflict under the same robustness check.

For PRIO25 civil conflict incidence, the preferred estimates are reported in column 6. The estimated marginal effect indicates that, on average across countries in the regression sample, a 1 percentage point increase in genetic diversity increases the predicted likelihood of civil conflict by almost 2.5 percentage points, an effect that is statistically significant at the 1% level. Figure 2 plots the likelihood of civil conflict, as predicted by the model in column 6, along the cross-country genetic diversity distribution. It suggests that a move from the 10<sup>th</sup> to the 90<sup>th</sup> percentile of the genetic diversity distribution is associated with a increase in the likelihood of conflict from 19 to 35 percent. Column 12 reports baseline results for ethnic conflict incidence. The estimated marginal effect implies that a 1 percentage point increase in genetic diversity, on average, raises the probability of an ethnic conflict during a 5-year period by more than 2.7 percentage points, an effect that is statistically significant at the 5% level.<sup>21</sup>

**Robustness to Annual Variations in Conflict Incidence.** Although the use of a panel data structure with 5-year periods is common practice for conflict incidence regressions, mainly due to concerns about serial correlation in the error term and reverse causality issues with time-varying explanatory variables, this strategy comes at the cost of ignoring valuable information about the cross-country variation in the duration of conflicts, many of which are relatively short-lived. In response to this concern, Table A.6 establishes that using a panel data structure with an annual frequency of country observations does not qualitatively alter the results for overall and ethnic civil conflict incidence. In terms of the magnitudes of the marginal effects of genetic diversity on conflict, the baseline estimates indicate that, on average, a 1 percentage point increase in diversity would increase the likelihood of a civil conflict by 1.18 percentage points, it would increase the likelihood of an ethnic conflict by 1.25 percentage points.

**Robustness to Alternative Codings and Types of Conflict.** Empirical results in the conflict literature are known to be quite sensitive to alternative codings and types of conflict. A natural question is whether the results uncovered thus far by the current analysis can be extended to large-scale conflicts, i.e., to “civil wars.” The analysis conducted in Table 7 tackles this question. The first three columns employ PRIO1000 as the dependent variable, which unlike PRIO25, adopts a minimum death threshold of 1,000 instead of 25, thereby effectively treating those years of conflict in which annual battle-related casualties remained below 1,000 as non-event years. In columns 4 through 6, the analysis performs robustness checks with respect to the ethnic war coding (henceforth WM 2006-09) of Wimmer and Min (2006), which has been cross-checked and supplemented by the more recent war list of Wimmer and Min (2009).<sup>22</sup> Reassuringly, the analysis in the first six columns of Table 7 indicate that the significance of genetic diversity for conflict is not simply

<sup>21</sup>Figure A.2 plots the predicted likelihood of ethnic conflict as a function of genetic diversity. It suggests that a move from the 10<sup>th</sup> to the 90<sup>th</sup> percentile of the diversity distribution increases the likelihood of ethnic conflict from 11 to 26 percent, an effect that is comparable to those associated with the PRIO25 results.

<sup>22</sup>Unfortunately, in the primary data source for WCM 2009 ethnic conflicts, Wimmer et al. (2009), large scale ethnic conflicts (i.e., ethnic wars) are not separately identified. As such, this particular analysis is restricted to employing

driven by variation in lower-intensity conflicts. The coefficient estimates for genetic diversity are positive and statistically significant at the 5% of less, and for each outcome variable examined, they remain largely stable across specifications.

The civil conflicts analyzed thus far are of the type that involve organized non-state armed groups, on the one hand, and state forces on the other. Although such conflicts are more common and detrimental, intrastate armed conflict, broadly defined, does not always involve state forces. In fact, if genetic diversity in a society contributes to the emergence of cultural, political, or economic grievances between different groups, then one would expect such grievances to be manifested not only as conflicts between *state actors and non-state actors* but also as conflicts between opposing non-state groups.

To explore the plausibility of this argument, the last three columns of Table 7 present the results from regressions using non-state conflict incidence as the outcome variable.<sup>23</sup> A non-state conflict event is coded only if two organized armed groups, neither of which is the government of a state, fought with each other, and the use of armed force resulted in at least 25 battle-related deaths in a year. While these armed groups usually differ in terms of ethnic self-identification, as in the case of the conflict between the Dizi and Surma peoples of Ethiopia, they may also be organized crime syndicates, like the Sinaloa Cartel and the Gulf Cartel in Mexico, without necessarily possessing an ethnic cleavage between them. As is evident from the results, genetic diversity has a positive and statistically significant effect on the incidence of non-state conflicts. The estimated marginal effects are sizable in magnitude, ranging between 4 and 4.9 percentage points in response to a 1 percentage point increase in genetic diversity.

Unlike standard group-based measures of fractionalization or polarization that necessarily assume homogeneity within group boundaries, genetic diversity additionally reflects heterogeneity across individuals belonging to a given ethnic, religious, or linguistic group. As such, genetic diversity should be expected to contribute to not only *intergroup* grievances and conflict but to *intra-group* violence as well. Table 8 offers evidence supporting these priors. The analysis employs the dataset of the Minorities at Risk (MAR) project to construct a country-level indicator, reflecting the incidence of one or more intra-group or communal conflicts between 1990 and 2000, by aggregating information on such events from the ethnopolitical group level to the country level. Given the cross-sectional nature of the analysis and the time coverage of communal conflict events, all time-varying right-hand-side variables are aggregated to the country level by averaging their respective values over the 1990–1999 time period.

The cross-country IV probit regressions in Table 8 uncover an adverse and highly significant effect of genetic diversity, instrumented by migratory distance of native populations from East Africa, on the likelihood of intra-group conflict. The first six columns of the table successively build into the full baseline model of column 7. According to the estimates presented in column 7, a

---

the WM 2006-09 ethnic war coding, which is based on the overall civil war coding of the Correlates of War (COW) Project (rather than the UCDP/PRIO Armed Conflict Dataset).

<sup>23</sup>This variable comes from the UCDP Non-State Conflict Dataset and has a temporal coverage of 1989 to 2008.

1 percentage point increase in diversity across countries in the sample would, on average, raise the probability of observing at least one intra-group conflict in the 1990-2000 time period by almost 10.9 percentage points. Figure 3 plots the predicted relationship between genetic diversity and intra-group conflict, as implied by the baseline model, suggesting that a move from the 10<sup>th</sup> to the 90<sup>th</sup> percentile of the cross-country genetic diversity distribution is associated with an increase in the likelihood of observing an intra-group conflict from 15% to 73%. The control variables added to the baseline model in the last two columns of the table ensure that the qualitative results are not an artifact of the cross-country variation in the number of minority groups, their total share in the population, or the overall number of ethnic groups in the country.

**Robustness to Additional Predictors of Civil Conflict** A few covariates were excluded from the baseline model for the sake of parsimony and in the interest of maximizing the regression samples, given the smaller coverage of the data on some of these variables. Table 9 establishes robustness of the baseline results with respect to these additional controls.

Insurgency-based explanations of civil war argue that conditions that are conducive to the emergence and survival of armed rebel organizations are more important than political grievances or ethnic cleavages. These explanations highlight several factors that can potentially affect the strength of an insurgent band relative to that of a state. Among these are two geographical features proposed by Fearon and Laitin (2003), one of which is the prevalence of mountainous regions in a country. Although the results presented so far do not indicate a systematically significant link between terrain ruggedness and civil conflict, ruggedness may not exactly capture the potential effect associated with the percentage of mountainous terrain. Not only is the latter more commonly used in the literature, it may also be a superior proxy for the availability of safe havens for insurgent groups. Another factor that may facilitate insurgent activity is the presence of a territorial base that is geographically separated from the state’s center. The standard variable in the empirical literature for capturing this argument is an indicator that reflects whether the country has any *noncontiguous territory* that is separated from the region containing the capital city, either by land belonging to a different nation or by at least 100 km of water. These two geographical variables are introduced to the baseline model in column 2, where the ruggedness index is also dropped from the specification. The coefficient on genetic diversity remains positive, relative to the baseline estimate from column 1, but due to the substantial reduction in the sample size, there is a drop in statistical significance to the 10% level.<sup>24</sup>

The next four columns of the table investigate robustness to other distributional indices of diversity. In columns 3 and 4, indices of linguistic and religious fractionalization from Alesina et al. (2003) are introduced one at a time, and in column 5, the analysis controls for the ethnic fractionalization index of Fearon (2003) as an alternative to the one adopted by the baseline model. In support of a theory of inter-group conflict formulated by Esteban and Ray (2011), the recent

---

<sup>24</sup>When the specification in column 1 is estimated on the regression sample from column 2, the coefficient on genetic diversity is significant at the 10% level. These results are not shown in Table 9, but they are available upon request.

empirical study of Esteban et al. (2012) finds that three distributional indices, namely ethnic polarization (P), fractionalization (F), and the Greenberg-Gini index of ethnic difference (G), are all important predictors of civil conflict. As such, in column 6, the analysis simultaneously controls for these indices, taken from the dataset of Esteban et al. (2012). In all cases, the coefficient on genetic diversity remains positive, statistically significant, and quantitatively stable, relative to the baseline estimate from column 1.

To the extent that the transition from hunting and gathering to sedentary agriculture contributed to ethnic heterogeneity, either by augmenting social stratification in human societies or by giving rise to urban centers that served as attractive destinations for migrants, the variation in conflict potential across countries today may be partially rooted in the cross-country variation in the timing of the Neolithic Revolution. Moreover, it is important to confirm that controlling for the timing of the agricultural transition does not affect the quantitative importance of the genetic diversity channel, especially in the light of the positive correlation between the number of years elapsed since the Neolithic transition and migratory distance from East Africa, which serves as the underlying source of variation in genetic diversity. The regression in column 7 performs this robustness check. The results indicate that, while the timing of the Neolithic Revolution enters the regression insignificantly, the coefficient associated with genetic diversity remains statistically significant at the 1% level.

The final robustness check pertains to the potentially confounding role of disease richness in explaining conflict. Specifically, Cervellati et al. (2011) find that persistent exposure to a more unfavorable disease environment is an important determinant of the incidence of civil conflict. As shown in the last column of the table, however, genetic diversity survives the inclusion of this control in the baseline model.

### 4.3 Explaining Conflict Onset

In the civil conflict literature, conflict prevalence and conflict onset are treated as somewhat different phenomena, since the factors contributing to the outbreak of conflicts may not coincide with those that allow them to persist. Moreover, determining the duration of a given conflict is a difficult task with well-known problems. Therefore, establishing robustness of the results with respect to conflict onsets is important to ensure that the results are not driven by subtle coding issues with regard to conflict duration.

As mentioned earlier, two conceptually different onset codings are employed as dependent variables in the onset analysis. The first one is the PRIO2 onset variable, which is coded 1 for the first year of a conflict episode that follows at least two years of civil peace, and 0 otherwise. The second measure, PRIO-NC, is also based on the UCDP/PRIO Armed Conflict Dataset, but it is coded 1 only for the starting year of a new (distinct) conflict.

In addition to the standard control variables in the baseline conflict incidence model, the baseline onset model, presented in Columns 5 and 10 of Table 10, includes (i) an indicator for



ethnic dominance, i.e., whether the largest ethnic group constitutes 45 to 90 percent of the whole population; (ii) a dummy variable for newly independent states; and (iii) an indicator for political instability.<sup>25</sup> Moreover, in addition to year dummies, a cubic spline in the time elapsed since the end of the most recent conflict episode is included in all regressions to account for the possibility of duration dependence.

Turning to the results, across the various onset specifications, the coefficient on genetic diversity is positive and statistically significant at the 5% level or less. The estimated marginal effects from the baseline onset models in columns 5 and 10 suggest that a 1 percentage point increase in genetic diversity leads, on average, to a 0.64 percentage point increase in the likelihood of a PRIO2 conflict onset and a 0.52 percentage point increase in the likelihood of a PRIO-NC onset. Table A.7 replicates the same analysis for ethnic conflict onset, arriving at a qualitatively similar results.

#### 4.4 Evidence on Mediating Channels

This section provides evidence on two channels that could potentially mediate the reduced-form causal effect of genetic diversity on conflict. One channel draws on the role of genetic diversity as a deep determinant of ethnic group formation and differentiation (Ashraf and Galor, 2013b). If the positive effect of genetic diversity on civil conflict, as documented here, is partly a consequence of the contribution of genetic diversity to ethnic divisions, one should expect (i) a positive relationship between ethnic fractionalization and genetic diversity and (ii) the effect of genetic diversity to become smaller in magnitude once the influence of ethnic fractionalization on conflict is accounted for. The results in columns 1 to 3 of Table 11 are consistent with these priors.

In the first column of the table, the number of civil conflicts is regressed on genetic diversity along with the geographical controls from the baseline model, including continent dummies. Like before, genetic diversity is instrumented with the migratory distance of native populations from East Africa. As expected, genetic diversity enters the regression with a positive and statistically significant 2SLS coefficient. In the second column, the analysis demonstrates that genetic diversity is a significant determinant of ethnic fractionalization, even after the potentially confounding effects of ultimate geographical factors are taken into account. The residual scatter-plot in Figure 4 depicts this conditional relationship.

The third column adds ethnic fractionalization to the specification from column 1. At this stage, it is important to note that the other covariates in the original baseline model potentially represent more proximate causes of conflict, and the underlying source of their variations could be rooted in, among other factors, genetic diversity and ethnic fractionalization. Hence, the omission of these endogenous covariates from the current analysis is essential to assess the full

---

<sup>25</sup>Political instability is a binary variable that is coded 1 if there has been any change in the polity score of the country in the past three years. The coefficients associated with the newly independent state dummy are not reported in the tables, as this variable turns out to perfectly predict the outcome variable in the estimation sample and is therefore automatically dropped from the regressions.

extent to which ethnic fractionalization can explain away that part of the residual variation in conflict that was otherwise being attributed to genetic diversity. Turning to the results, although ethnic fractionalization enters the regression with a positive coefficient, it is not distinguishable from zero at the conventional levels of statistical significance. Once fractionalization is controlled for, however, the coefficient on genetic diversity appears somewhat smaller in magnitude, relative to the estimate from the first column. While not conclusive, these findings are suggestive of the possibility that ethnic fractionalization could be one of the mechanisms through which genetic diversity influences civil conflict.

In columns 4 through 6, a similar analysis is repeated for interpersonal trust. As is evident from column 5, countries exhibiting higher genetic diversity exhibit lower levels of interpersonal trust, an effect that is statistically significant at the 1% level. Figure 5 depicts this negative relationship. Importantly, in the conflict regression in column 6, interpersonal trust enters with the expected negative sign (albeit insignificantly) and the coefficient estimate for genetic diversity appears smaller in magnitude, in comparison to the estimated effect of diversity from column 4 where trust is not controlled for.

Table A.8 in the appendix repeats the analysis from Table 11 using the OLS instead of the 2SLS estimator, demonstrating that the findings are not qualitatively sensitive to the instrumentation strategy.

One important caveat in the current analysis is the possible endogeneity of both ethnic fractionalization and trust to civil conflict, which in turn prevents the analysis from drawing stronger conclusions regarding the consistency of the estimated coefficients on these variables and, thus, the true extent of their potential mediating role. In order to assess these mechanisms conclusively, the analysis would need to exploit independent exogenous sources of variation in these variables, a task that remains an interesting avenue for future exploration.

## 5 Concluding Remarks

This research establishes that the emergence, prevalence, and recurrence of civil conflicts in the modern era reflect the long shadow of prehistory. Exploiting variations across national populations, it establishes that genetic diversity, as determined predominantly during the “out of Africa” migration of humans to the rest of the world tens of thousands of years ago, has contributed significantly to the incidence, onset, and frequency of both ethnic and overall civil conflicts in the last half century, accounting for a large set of potential correlates of civil conflict. Importantly, the positive reduced-form causal effect of genetic diversity on civil conflict survives a wide range of robustness checks.

The influence of genetic diversity on civil conflict arguably reflects the adverse effect of genetic diversity on interpersonal trust and cooperation, the potential impact of diversity on income inequality, the potential association between diversity and divergence in preferences for

public goods and redistributive policies, and the contribution of genetic diversity to the degree of fractionalization and polarization across ethnic and linguistic groups in the population. The analysis demonstrates that the reduced-form impact of genetic diversity on conflicts may, indeed, partly operate through interpersonal trust and ethnolinguistic fragmentation. Conclusive evidence on the mediating channels, however, requires independent exogenous sources of variation for each of the hypothesized proximate determinants in the diversity channel, the exploration of which is left for future research.

## References

- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, “Fractionalization,” *Journal of Economic Growth*, June 2003, 8 (2), 155–194.
- Ashraf, Quamrul and Oded Galor, “The “Out of Africa” Hypothesis, Human Genetic Diversity, and Comparative Economic Development,” *American Economic Review*, February 2013a, 103 (1), 1–48.
- and —, “Genetic Diversity and the Origins of Cultural Fragmentation,” *American Economic Review Papers and Proceedings*, May 2013b, 103 (3), 528–533.
- at Risk Project, Minorities, “Minorities at Risk Dataset [Computer File],” <http://www.cidcm.umd.edu/mar/> 2009. College Park, MD: Center for International Development and Conflict Management.
- Bates, Robert H., “Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa,” in Marion E. Doro and Newell M. Schultz, eds., *Governing Black Africa*, New York, NY: Africana Publishing Company, 1986, pp. 153–169.
- Beck, Nathaniel, Jonathan Katz, and Richard Tucker, “Taking Time Seriously: Time-Series-CrossSection Analysis with a Binary Dependent Variable,” *American Journal of Political Science*, October 1998, 42 (4), 1260–1288.
- Brewer, Marilynn B., “In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis,” *Psychological Bulletin*, March 1979, 86 (2), 307–324.
- , “The Social Self: On Being the Same and Different at the Same Time,” *Personality and Social Psychology Bulletin*, October 1991, 17 (5), 475–482.
- , “The Social Psychology of Intergroup Relations: Can Research Inform Practice?,” *Journal of Social Issues*, Spring 1997, 53 (1), 197–211.
- Cann, Howard M., Claudia de Toma, Lucien Cazes, Marie-Fernande Legrand, Valerie Morel, Laurence Piouffre et al., “A Human Genome Diversity Cell Line Panel,” *Science*, April 2002, 296 (5566), 261–262.
- Caselli, Francesco and Wilbur John Coleman II, “On the Theory of Ethnic Conflict,” 2006. NBER Working Paper No. 12125, National Bureau of Economic Research.
- Cavalli-Sforza, L. Luca, “The Human Genome Diversity Project: Past, Present and Future,” *Nature Reviews Genetics*, April 2005, 6 (4), 333–340.

- Cervellati, Matteo, Uwe Sunde, and Simona Valmori**, “Disease Environment and Civil Conflicts,” 2011. IZA Discussion Paper No. 5614.
- Collier, Paul and Anke Hoeffler**, “On Economic Causes of Civil War,” *Oxford Economic Papers*, October 1998, 50 (4), 563–573.
- **and** —, “Greed and Grievance in Civil War,” *Oxford Economic Papers*, October 2004, 56 (4), 563–595.
- den Berghe, Pierre L. Van**, *The Ethnic Phenomenon*, New York, NY: Elsevier Science Publishing, 1981.
- , “Does Race Matter?,” *Nations and Nationalism*, November 1995, 1 (3), 357–368.
- Desmet, Klaus, Ignacio Ortuno-Ortín, and Romain Wacziarg**, “The Political Economy of Linguistic Cleavages,” *Journal of Development Economics*, March 2012, 97 (2), 322–338.
- , **Michel Le Breton, Ignacio Ortuno-Ortín, and Shlomo Weber**, “The Stability and Breakup of Nations: A Quantitative Analysis,” *Journal of Economic Growth*, September 2011, 16 (3), 183–213.
- Eifert, Benn, Edward Miguel, and Daniel N. Posner**, “Political Competition and Ethnic Identification in Africa,” *American Journal of Political Science*, April 2010, 54 (2), 494–510.
- Esteban, Joan and Debraj Ray**, “On the Salience of Ethnic Conflict,” *American Economic Review*, December 2008, 98 (5), 2185–2202.
- **and** —, “A Model of Ethnic Conflict,” *Journal of the European Economic Association*, June 2011, 9 (3), 496–521.
- , **Laura Mayoral, and Debraj Ray**, “Ethnicity and Conflict: An Empirical Study,” *American Economic Review*, June 2012, 102 (4), 1310–1342.
- Esty, Daniel C., Jack A. Goldstone, Ted Robert Gurr, Barbara Harff, Marc Levy, Geoffrey D. Dabelko, Pamela T. Surko, and Alan N. Unger**, “State Failure Task Force Report: Phase II Findings,” Technical Report, Science Applications International Corporation, McLean, VA 1998.
- Fearon, James D.**, “Ethnic and Cultural Diversity by Country,” *Journal of Economic Growth*, June 2003, 8 (2), 195–222.
- **and David D. Laitin**, “Ethnicity, Insurgency, and Civil War,” *American Political Science Review*, February 2003, 97 (1), 75–90.

- Fincher, Corey L and Randy Thornhill**, “Assortative sociality, limited dispersal, infectious disease and the genesis of the global pattern of religion diversity,” *Proceedings of the Royal Society B: Biological Sciences*, November 2008, *275* (1651), 2587–2594.
- Gallup, John L., Jeffrey D. Sachs, and Andrew D. Mellinger**, “Geography and Economic Development,” *International Regional Science Review*, 1999, *22* (2), 179–232.
- Gellner, Ernest**, *Nations and Nationalism*, Ithaca, NY: Cornell University Press, 1983.
- Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand**, “Armed Conflict 1946-2001: A New Dataset,” *Journal of Peace Research*, September 2002, *39* (5), 615–637.
- Griffiths, Anthony J. F., Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, and William M. Gelbart**, *An Introduction to Genetic Analysis*, New York, NY: W. H. Freeman & Co., 2000.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales**, “Cultural Biases in Economic Exchange?,” *Quarterly Journal of Economics*, August 2009, *124* (3), 1095–1131.
- Gurr, Ted Robert**, “Ethnic Warfare on the Wane,” *Foreign Affairs*, May–June 2000, *79* (3), 52–64.
- Hegre, Håvard and Nicholas Sambanis**, “Sensitivity Analysis of Empirical Results on Civil War Onset,” *Journal of Conflict Resolution*, August 2006, *50* (4), 508–535.
- Horowitz, Donald L.**, *Ethnic Groups in Conflict*, Berkeley, CA: University of California Press, 1985.
- , “Structure and Strategy in Ethnic Conflict: A Few Steps toward Synthesis,” in Boris Pleskovic and Joseph E. Stiglitz, eds., *Annual World Bank Conference on Development Economics 1998*, Washington, DC: The World Bank, 1999, pp. 345–370.
- Humphreys, Macartan**, “Natural Resources, Conflict, and Conflict Resolution: Uncovering the Mechanisms,” *Journal of Conflict Resolution*, August 2005, *49* (4), 508–537.
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny**, “The Quality of Government,” *Journal of Law, Economics, and Organization*, March 1999, *15* (1), 222–279.
- Maddison, Angus**, “Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD, [Computer File],” <http://www.ggd.net/maddison/oriindex.htm> 2010.
- Marshall, Monty G. and Keith Jagers**, “Polity IV Project: Political Regime Characteristics and Transitions 1800-2009, [Computer File],” [www.systemicpeace.org/polity/polity4.htm](http://www.systemicpeace.org/polity/polity4.htm) 2009. Center for Systemic Peace.

- Michalopoulos, Stelios**, “The Origins of Ethnolinguistic Diversity,” *American Economic Review*, June 2012, *102* (4), 1508–1539.
- Montalvo, José G. and Marta Reynal-Querol**, “Ethnic Polarization, Potential Conflict, and Civil Wars,” *American Economic Review*, June 2005, *95* (3), 796–816.
- Prugnolle, Franck, Andrea Manica, and François Balloux**, “Geography Predicts Neutral Genetic Diversity of Human Populations,” *Current Biology*, March 2005, *15* (5), R159–R160.
- Putterman, Louis**, “Agriculture, Diffusion, and Development: Ripple Effects of the Neolithic Revolution,” *Economica*, November 2008, *75* (300), 729–748.
- **and David N. Weil**, “Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality,” *Quarterly Journal of Economics*, November 2010, *125* (4), 1627–1682.
- Ramachandran, Sohini, Omkar Deshpande, Charles C. Roseman, Noah A. Rosenberg, Marcus W. Feldman, and L. Luca Cavalli-Sforza**, “Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa,” *Proceedings of the National Academy of Sciences*, November 2005, *102* (44), 15942–15947.
- Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney**, “The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change,” *Global Ecology and Biogeography*, September 2002, *11* (5), 377–392.
- Ross, Michael L.**, “A Closer Look at Oil, Diamonds, and Civil War,” *Annual Review of Political Science*, June 2006, *9*, 265–300.
- , “Oil and Gas Data, 1932-2011,” <http://dvn.iq.harvard.edu/dvn/dv/mlross/faces/study/StudyPage.xhtml?studyId=87422&versionNumber=2> February 2013.
- Shils, Edward**, “Primordial, Personal, Sacred and Civil Ties: Some Particular Observations on the Relationships of Sociological Research and Theory,” *British Journal of Sociology*, June 1957, *8* (2), 130–145.
- Spolaore, Enrico and Romain Wacziarg**, “The Diffusion of Development,” *Quarterly Journal of Economics*, May 2009a, *124* (2), 469–529.
- **and —** , “War and Relatedness,” 2009b. NBER Working Paper No. 15095, National Bureau of Economic Research.
- Sundberg, Ralph, Kristine Eck, and Joakim Kreutz**, “Introducing the UCDP NonState Conflict Dataset,” *Journal of Peace Research*, March 2012, *49*, 351–362.

- Wimmer, Andreas and Brian Min**, “From Empire to Nation-State: Explaining Wars in the Modern World, 1816–2001,” *American Sociological Review*, December 2006, *71* (6), 867–897.
- **and** –, “The Location and Purpose of Wars Around the World: A New Global Dataset, 1816–2001,” *International Interactions*, November 2009, *35* (4), 390–417.
- , **Lars-Erik Cederman, and Brian Min**, “Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set,” *American Sociological Review*, April 2009, *74* (2), 316–337.



TABLE 1: Diversity Measures and Civil Conflict – Bivariate Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Number of new civil conflict onsets per year (PRIO25), 1960-2008						
Genetic diversity (ancestry adjusted)	0.212*** (0.076)					
Ethnic fractionalization (Alesina et al. 2003)		0.024*** (0.008)				
Linguistic fractionalization (Alesina et al. 2003)			0.032*** (0.009)			
Religious fractionalization (Alesina et al. 2003)				0.006 (0.009)		
Ethnolinguistic fractionalization, level 15 (Desmet et al. 2012)					0.026*** (0.008)	
Ethnolinguistic polarization, level 15 (Desmet et al. 2012)						0.007 (0.009)
Effect of moving the diversity measure from 10 <sup>th</sup> to 90 <sup>th</sup> percentile	.014*** (.005)	.017*** (.005)	.025*** (.007)	.004 (.006)	.022*** (.007)	.005 (.006)
Continent dummies	No	No	No	No	No	No
Observations	154	154	154	154	154	154
R-squared	0.032	0.037	0.081	0.002	0.061	0.003

TABLE 2: Diversity Measures and Civil Conflict – “Horse race” Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Number of new civil conflict onsets per year (PRIO25), 1960-2008										
Genetic diversity (ancestry adjusted)	0.212*** (0.076)	0.178** (0.073)	0.146** (0.072)	0.200*** (0.072)	0.157** (0.074)	0.207*** (0.076)	0.183** (0.071)	0.158** (0.069)	0.158** (0.073)	0.160** (0.068)
Ethnic fractionalization (Alesina et al. 2003)		0.021*** (0.007)					0.021*** (0.007)			-0.004 (0.014)
Linguistic fractionalization (Alesina et al. 2003)			0.029*** (0.009)					0.031*** (0.009)		0.029* (0.016)
Religious fractionalization (Alesina et al. 2003)				0.001 (0.009)			-0.003 (0.009)			-0.010 (0.010)
Ethnolinguistic fractionalization, level 15 (Desmet et al. 2012)					0.023*** (0.008)				0.026** (0.011)	0.007 (0.023)
Ethnolinguistic polarization, level 15 (Desmet et al. 2012)						0.005 (0.008)			-0.008 (0.012)	-0.006 (0.015)
Effect of moving genetic diversity from 10 <sup>th</sup> to 90 <sup>th</sup> percentile	.014*** (.005)	.011** (.005)	.009** (.005)	.013*** (.005)	.01** (.005)	.013*** (.005)	.012*** (.005)	.01** (.004)	.01** (.005)	.01** (.004)
Continent dummies	No	No	No	No	No	No	No	No	No	No
Observations	154	154	154	154	154	154	154	154	154	154
Adjusted R-squared	0.026	0.046	0.084	0.019	0.066	0.021	0.041	0.082	0.063	0.066

TABLE 3: Genetic Diversity and Civil Conflict – The Baseline Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS
	Number of new civil conflict onsets per year (PRIO25), 1960-2008						
Genetic diversity (ancestry adjusted)	0.182**	0.322**	0.366**	0.350**	0.390**	0.377**	0.398**
	(0.077)	(0.150)	(0.171)	(0.168)	(0.179)	(0.184)	(0.183)
Ethnic fractionalization				0.011		0.006	0.007
				(0.012)		(0.012)	(0.012)
Ethnoling. polarization					0.013	0.010	0.010
					(0.013)	(0.014)	(0.013)
Absolute latitude		-0.440*	-0.331	-0.225	-0.356	-0.292	0.149
		(0.255)	(0.263)	(0.320)	(0.256)	(0.307)	(0.287)
Land area		1.825	1.709	1.972	1.719	1.862	1.586
		(2.287)	(2.358)	(2.382)	(2.403)	(2.436)	(2.675)
Ruggedness		0.028	0.030	0.036	0.032	0.035	0.056
		(0.044)	(0.044)	(0.047)	(0.045)	(0.047)	(0.047)
Mean elevation		-0.015	-0.017*	-0.018*	-0.018*	-0.018*	-0.020**
		(0.009)	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)
Range of elevation		0.009**	0.009**	0.008*	0.008*	0.008*	0.004
		(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
Mean land suitability		0.018	0.016	0.019	0.019	0.020	0.006
		(0.013)	(0.015)	(0.014)	(0.015)	(0.014)	(0.016)
Range of land suitability		0.014	0.012	0.011	0.014	0.013	0.010
		(0.011)	(0.012)	(0.012)	(0.014)	(0.014)	(0.015)
Mean distance to nearest waterway		0.005	0.007	0.006	0.007	0.007	0.002
		(0.010)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Executive constraints, 1960-2008 average			0.004	0.004	0.004	0.004	0.006*
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Fraction of years under democracy, 1960-2008			-0.015	-0.014	-0.015	-0.014	-0.012
			(0.019)	(0.019)	(0.019)	(0.019)	(0.018)
Fraction of years under autocracy, 1960-2008			-0.006	-0.005	-0.005	-0.005	-0.008
			(0.017)	(0.016)	(0.017)	(0.016)	(0.016)
Value of oil production per capita, 1960-2008 average							0.002**
							(0.001)
Log population, 1960-2008 average							0.003
							(0.003)
Log GDP per capita, 1960-2008 average							-0.015***
							(0.005)
Effect of moving genetic diversity from 10 <sup>th</sup> to 90 <sup>th</sup> percentile	.012**	.021**	.024**	.023**	.025**	.025**	.026**
	(.005)	(.010)	(.011)	(.011)	(.012)	(.012)	(.012)
Continent dummies	No	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	No	No	Yes	Yes	Yes	Yes	Yes
Observations	143	143	143	143	143	143	143
Adjusted R-squared	0.019	0.197	0.192	0.190	0.193	0.188	0.235
Sqr. Partial Correlation (Genetic Diversity)	-	.046	.056	.051	.063	.056	.066
Sum of Sqr. Partial Correlations (other diversity measures)	-	-	-	.006	.010	.007	.007

TABLE 4: Genetic Diversity and Civil Conflict – Addressing Endogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Number of new civil conflict onsets per year (PRIO25), 1960-2008							
Genetic diversity (ancestry adjusted)	0.256*** (0.082)	0.599*** (0.231)	0.737*** (0.267)	0.723*** (0.268)	0.740*** (0.261)	0.744*** (0.270)	0.805*** (0.275)
Ethnic fractionalization				0.007 (0.011)		-0.002 (0.012)	-0.002 (0.013)
Ethnoling. polarization					0.017 (0.012)	0.018 (0.013)	0.019 (0.013)
Executive constraints, 1960-2008 average			0.005* (0.003)	0.005 (0.003)	0.005* (0.003)	0.006 (0.003)	0.008** (0.004)
Fraction of years under democracy, 1960-2008			-0.018 (0.018)	-0.017 (0.018)	-0.018 (0.018)	-0.019 (0.018)	-0.017 (0.017)
Fraction of years under autocracy, 1960-2008			-0.005 (0.016)	-0.005 (0.015)	-0.004 (0.016)	-0.004 (0.015)	-0.007 (0.015)
Value of oil production per capita, 1960-2008 average							0.002* (0.001)
Log population, 1960-2008 average							0.003 (0.003)
Log GDP per capita, 1960-2008 average							-0.016*** (0.004)
Effect of moving genetic diversity from 10 <sup>th</sup> to 90 <sup>th</sup> percentile	.017*** (.005)	.039*** (.015)	.048*** (.017)	.047*** (.017)	.048*** (.017)	.048*** (.018)	.052*** (.018)
Continental and geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	No	No	Yes	Yes	Yes	Yes	Yes
Observations	143	143	143	143	143	143	143
Adjusted R-squared	0.015	0.169	0.143	0.141	0.150	0.142	0.179

TABLE 5: Civil Conflict across Countries – Robustness to the Elimination of Regions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
	Number of new civil conflict onsets per year , 1960-2008							
Genetic diversity (ancestry adjusted)	1.210*** (0.379)	1.256*** (0.393)	0.736*** (0.248)	0.322** (0.162)	0.637*** (0.204)	0.736*** (0.226)	0.392*** (0.146)	0.616*** (0.223)
Ethnic fractionalization	-0.004 (0.019)	-0.007 (0.020)	-0.001 (0.014)	0.014 (0.011)	-0.017 (0.017)	-0.008 (0.014)	0.009 (0.013)	-0.016 (0.016)
Ethnoling. polarization	0.015 (0.015)	0.017 (0.016)	0.011 (0.013)	0.010 (0.010)	0.020* (0.012)	0.038*** (0.015)	0.016 (0.010)	0.027** (0.012)
Executive constraints, 1960-2008 average	0.009 (0.006)	0.011 (0.007)	0.004 (0.003)	0.005 (0.003)	0.008** (0.004)	0.006* (0.003)	0.004 (0.003)	0.009** (0.004)
Fraction of years under democracy, 1960-2008	-0.041 (0.032)	-0.043 (0.034)	-0.015 (0.018)	-0.027** (0.013)	-0.027 (0.018)	0.008 (0.016)	-0.027* (0.015)	-0.022 (0.017)
Fraction of years under autocracy, 1960-2008	-0.015 (0.019)	-0.010 (0.020)	-0.014 (0.017)	-0.008 (0.013)	-0.019 (0.018)	-0.001 (0.015)	-0.011 (0.014)	0.002 (0.018)
Value of oil production per capita, 1960-2008 average	0.000 (0.001)	0.000 (0.001)	0.002* (0.001)	0.001 (0.001)	0.002 (0.001)	0.003** (0.001)	0.002* (0.001)	0.002* (0.001)
Log population, 1960-2008 average	0.000 (0.004)	0.001 (0.004)	0.005* (0.003)	0.002 (0.003)	0.003 (0.003)	0.007** (0.003)	0.002 (0.002)	0.004 (0.003)
Log GDP per capita, 1960-2008 average	-0.014** (0.006)	-0.015*** (0.006)	-0.014*** (0.004)	-0.014*** (0.005)	-0.018*** (0.005)	-0.014*** (0.005)	-0.012*** (0.004)	-0.014*** (0.004)
Continent dummies	No	No	No	No	No	No	No	No
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Omitted Region/Continent	Sub-Sah. Africa	Africa	LAAC	Asia	MENA	Europe	South Asia	E. Asia & Pacific
Observations	102	97	123	104	127	109	137	128

TABLE 6: Genetic Diversity and Conflict Incidence in Panel Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit
	Civil Conflict Incidence (PRIO25)						Ethnic Civil Conflict Incidence (WCM 2009)					
Genetic diversity (ancestry adjusted)	12.540*** (4.215)	10.573** (4.358)	11.034** (4.408)	10.515** (4.361)	11.417** (4.521)	12.833*** (4.807)	15.819** (6.257)	15.048** (6.074)	15.534** (6.195)	14.849** (6.297)	15.872** (6.282)	16.294** (6.952)
Ethnic fractionalization			-0.229 (0.264)		-0.389 (0.300)	-0.297 (0.328)			-0.105 (0.482)		-0.434 (0.517)	-0.420 (0.547)
Ethnoling. polarization			0.213 (0.295)		0.355 (0.329)	0.398 (0.318)				0.572 (0.538)		0.746 (0.586)
Civil Conflict (PRIO), lagged	1.654*** (0.126)	1.617*** (0.125)	1.613*** (0.125)	1.618*** (0.125)	1.611*** (0.125)	1.570*** (0.122)						
Civil Conflict (WCM09), lagged							1.344*** (0.206)	1.336*** (0.215)	1.338*** (0.213)	1.344*** (0.212)	1.338*** (0.213)	1.307*** (0.207)
Absolute latitude	-22.414*** (6.119)	-18.899*** (6.464)	-20.986*** (6.999)	-19.419*** (6.638)	-23.512*** (7.776)	-17.522** (8.546)	-28.942*** (9.371)	-27.176*** (11.060)	-28.301** (11.100)	-28.644*** (11.060)	-33.004*** (12.321)	-29.611** (13.403)
Land area	-28.616 (43.673)	-26.361 (41.668)	-30.655 (42.269)	-26.820 (41.397)	-33.833 (42.295)	-51.506 (48.110)	14.138 (54.321)	27.926 (53.722)	27.117 (54.843)	24.373 (56.526)	15.287 (58.806)	7.663 (64.487)
Ruggedness	0.832 (0.965)	1.473 (0.938)	1.377 (0.952)	1.481 (0.951)	1.345 (0.956)	1.561 (1.020)	-0.475 (1.521)	-0.146 (1.663)	-0.223 (1.623)	-0.274 (1.714)	-0.504 (1.706)	-0.303 (1.885)
Mean elevation	-0.372* (0.215)	-0.510** (0.207)	-0.506** (0.206)	-0.518** (0.210)	-0.516** (0.209)	-0.504** (0.214)	0.111 (0.384)	0.047 (0.389)	0.047 (0.391)	0.054 (0.393)	0.053 (0.396)	0.043 (0.414)
Range of elevation	0.140** (0.065)	0.112 (0.069)	0.120* (0.070)	0.106 (0.067)	0.115* (0.068)	0.115* (0.087)	-0.080 (0.116)	-0.129 (0.130)	-0.125 (0.128)	-0.146 (0.130)	-0.135 (0.129)	-0.178 (0.147)
Mean land suitability	0.227 (0.280)	0.302 (0.275)	0.232 (0.281)	0.359 (0.288)	0.281 (0.289)	0.004 (0.346)	0.239 (0.473)	0.382 (0.476)	0.379 (0.506)	0.570 (0.534)	0.519 (0.543)	0.441 (0.575)
Range of land suitability	0.689*** (0.256)	0.770*** (0.232)	0.818*** (0.241)	0.809*** (0.238)	0.915*** (0.258)	0.775** (0.316)	1.530*** (0.512)	1.852*** (0.510)	1.868*** (0.491)	1.968*** (0.491)	2.070*** (0.499)	2.023*** (0.584)
Mean distance to nearest waterway	0.342* (0.184)	0.551*** (0.198)	0.563*** (0.197)	0.559*** (0.199)	0.587*** (0.198)	0.504** (0.211)	0.166 (0.254)	0.343 (0.283)	0.357 (0.287)	0.381 (0.279)	0.420 (0.280)	0.390 (0.260)
Fraction of years as U.K. colony, lagged	-0.819* (0.420)	-0.840* (0.432)	-0.831** (0.432)	-0.831** (0.418)	-0.871** (0.435)	-0.876* (0.448)		-0.538 (0.391)	-0.535 (0.383)	-0.578 (0.383)	-0.599 (0.367)	-0.614* (0.360)
Fraction of years as French colony, lagged	-4.835 (3.018)	-4.749 (3.008)	-4.858 (3.007)	-4.858 (3.007)	-4.720 (3.007)	-3.964 (2.695)		-0.774 (2.118)	-0.784 (2.137)	-0.809 (2.108)	-0.827 (2.115)	-0.376 (1.908)
Fraction of years as non-U.K./non-French colony, lagged	0.344 (0.682)	0.305 (0.668)	0.305 (0.668)	0.378 (0.691)	0.339 (0.673)	0.419 (0.682)		1.366** (0.661)	1.358** (0.643)	1.515** (0.663)	1.525** (0.642)	1.513** (0.637)
Executive constraints, lagged	0.058 (0.055)	0.062 (0.056)	0.062 (0.056)	0.056 (0.055)	0.062 (0.057)	0.079 (0.057)		0.135** (0.066)	0.138** (0.067)	0.133** (0.066)	0.142** (0.067)	0.165** (0.074)
Fraction of years under democracy, lagged	-0.455* (0.249)	-0.478* (0.258)	-0.478* (0.258)	-0.445* (0.259)	-0.476* (0.259)	-0.483* (0.253)		-0.562** (0.264)	-0.570** (0.267)	-0.552** (0.261)	-0.587** (0.269)	-0.641** (0.258)
Fraction of years under autocracy, lagged	-0.267 (0.171)	-0.267 (0.171)	-0.266 (0.171)	-0.268 (0.170)	-0.267 (0.170)	-0.295* (0.172)		-0.198 (0.287)	-0.193 (0.289)	-0.196 (0.285)	-0.185 (0.286)	-0.164 (0.284)
Value of oil production per capita, lagged						0.023 (0.029)						0.017 (0.038)
Log population, lagged						0.093 (0.069)						0.030 (0.111)
Log GDP per capita, lagged						-0.233** (0.103)						-0.153 (0.187)
Marginal effect	2.538*** (.848)	2.065** (.855)	2.155** (.865)	2.054** (.857)	2.231** (.890)	2.491*** (.947)	2.831*** (1.057)	2.513** (.992)	2.585*** (1.007)	2.463** (1.012)	2.631*** (1.018)	2.716** (1.116)
Continental and geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Observations	1154	1154	1154	1154	1154	1154	1039	1039	1039	1039	1039	1039
Countries	141	141	141	141	141	141	129	129	129	129	129	129
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Wald test chi-squared	7.389	4.995	5.233	4.886	4.927	6.157	10.986	9.948	9.845	9.909	9.874	10.085
Wald test p-value	0.007	0.025	0.022	0.027	0.026	0.013	0.001	0.002	0.002	0.002	0.002	0.001

TABLE 7: Robustness to Alternative Conflict Incidence Measures in Panel Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit
	Civil War Incidence (PRIO1000)			Ethnic Civil War Incidence (WM 2006-09)			Non-state Conflict (UCDP)		
Genetic diversity (ancestry adjusted)	11.912** (5.946)	15.066** (7.035)	14.934** (7.040)	20.779*** (7.429)	24.685*** (8.728)	25.623*** (8.612)	24.974*** (5.392)	24.288*** (6.204)	30.318*** (6.180)
Civil conflict incidence (PRIO25), lagged	1.414*** (0.155)	1.387*** (0.159)	1.393*** (0.157)						
Civil war incidence (WM0609), lagged				1.198*** (0.207)	1.164*** (0.209)	1.133*** (0.217)			
Non-state civil conflict incidence, lagged							1.164*** (0.220)	1.173*** (0.215)	0.940*** (0.208)
Ethnic fractionalization		-0.621 (0.605)	-0.770 (0.618)		-0.692 (0.781)	-0.994 (0.735)		-0.389 (0.559)	0.103 (0.680)
Ethnoling. polarization		0.384 (0.532)	0.369 (0.518)		0.424 (0.708)	0.602 (0.682)		0.921* (0.492)	1.199** (0.490)
Executive constraints, lagged		-0.049 (0.090)	-0.025 (0.089)		0.018 (0.105)	0.030 (0.101)		-0.060 (0.097)	-0.002 (0.097)
Fraction of years under democracy, lagged		-0.287 (0.375)	-0.310 (0.373)		-0.547 (0.551)	-0.529 (0.507)		0.086 (0.343)	-0.163 (0.344)
Fraction of years under autocracy, lagged		-0.328 (0.269)	-0.282 (0.272)		-0.243 (0.331)	-0.234 (0.326)		0.171 (0.283)	0.140 (0.287)
Value of oil production per capita, lagged			0.003 (0.048)			-0.058 (0.065)			-0.028 (0.047)
Log population, lagged			-0.099 (0.111)			-0.176 (0.136)			0.398*** (0.124)
Log GDP per capita, lagged			-0.107 (0.187)			-0.140 (0.248)			-0.289* (0.160)
Marginal effect	1.424* (.747)	1.795** (.898)	1.775** (.894)	1.784** (.726)	2.087** (.864)	2.216*** (.850)	4.233*** (.958)	4.004*** (1.099)	4.891*** (1.159)
Continent and geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	1154	1138	1138	860	860	860	535	535	535
Countries	141	141	141	126	126	126	140	140	140
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Wald test chi-squared	2.511	2.950	2.980	6.854	8.339	10.352	12.990	10.745	19.582
Wald test p-value	0.113	0.086	0.084	0.009	0.004	0.001	0.000	0.001	0.000

TABLE 8: Genetic Diversity and Intra-group Communal Conflict

	(1) IV Probit	(2) IV Probit	(3) IV Probit	(4) IV Probit	(5) IV Probit	(6) IV Probit	(7) IV Probit	(8) IV Probit	(9) IV Probit
	Any Intra-group Communal Conflict, 1990-2000 (MAR)								
Genetic diversity (ancestry adjusted)	23.206*** (7.846)	27.005*** (9.054)	28.393*** (10.410)	28.099*** (10.522)	34.427*** (11.569)	36.846*** (11.377)	41.733*** (12.424)	49.041*** (12.085)	49.574*** (12.031)
Ethnic fractionalization				0.394 (0.850)		-1.111 (0.907)	-0.893 (1.009)	-1.825* (1.003)	-0.880 (1.194)
Ethnoling. polarization					2.388*** (0.803)	2.809*** (0.824)	3.351*** (0.875)	3.320*** (0.858)	2.961*** (0.860)
Executive constraints, 1990-1999 average			0.143 (0.210)	0.146 (0.211)	0.085 (0.212)	0.071 (0.217)	0.066 (0.232)	0.113 (0.232)	0.120 (0.228)
Fraction of years under democracy, 1990-1999			0.283 (0.782)	0.324 (0.784)	0.693 (0.812)	0.673 (0.820)	0.764 (0.868)	0.698 (0.919)	0.714 (0.924)
Fraction of years under autocracy, 1990-1999			-0.067 (0.486)	0.010 (0.524)	-0.107 (0.512)	-0.311 (0.560)	-0.615 (0.602)	-0.845 (0.603)	-0.718 (0.594)
Value of oil production per capita, 1990-1999 average							0.036 (0.086)	-0.015 (0.078)	-0.022 (0.078)
Log population, 1990-1999 average							0.376* (0.206)	0.457** (0.217)	0.446** (0.222)
Log GDP per capita, 1990-1999 average							-0.156 (0.251)	-0.103 (0.245)	-0.112 (0.243)
Population share of all MAR groups in 1990, est. in 1998								0.768 (0.859)	0.354 (0.842)
Number of active MAR groups in country								0.266 (0.188)	0.295 (0.185)
Number of ethnic groups (Fearon, 2003)									-0.084* (0.050)
Marginal effect	8.408*** (2.427)	8.117*** (2.517)	7.817*** (2.819)	7.738*** (2.861)	9.137*** (3.061)	9.772*** (3.053)	10.899*** (3.267)	12.347*** (3.003)	12.217*** (2.939)
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colonial dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	103	103	103	103	103	103	103	103	103
Wald test chi-squared	6.666	7.163	6.673	6.776	6.146	5.963	6.382	6.405	6.872
Wald test p-value	0.010	0.007	0.010	0.009	0.013	0.015	0.012	0.011	0.009

TABLE 9: Robustness to Additional Predictors of Conflict Incidence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit
	Civil Conflict Incidence (PRIO25)							
Genetic diversity (ancestry adjusted)	12.344*** (4.548)	9.591* (5.609)	13.187*** (4.886)	12.916*** (4.704)	12.128*** (4.583)	11.865** (5.533)	13.276*** (5.032)	12.233*** (4.653)
Civil Conflict Incidence, lagged	1.554*** (0.123)	1.550*** (0.144)	1.540*** (0.127)	1.553*** (0.126)	1.550*** (0.122)	1.531*** (0.127)	1.545*** (0.122)	1.552*** (0.122)
Ruggedness	1.646* (0.988)		1.568 (1.077)	1.555 (0.994)	1.677* (0.986)	2.071** (1.052)	1.676* (0.989)	1.700* (0.986)
Log mountainous terrain		0.118 (0.073)						
Noncontiguous state dummy		0.390* (0.235)						
Linguistic fractionalization			0.287 (0.272)					
Religious fractionalization				-0.646** (0.259)				
Ethnic fractionalization (Fearon)					0.099 (0.278)			
Fearon-F (Esteban et al., 2012)						0.259 (0.345)		
Fearon-P (Esteban et al., 2012)						3.503** (1.485)		
Fearon-G/N (Esteban et al., 2012)						-1.439 (1.340)		
Log years since Neolithic Revolution (ancestry adjusted)							-0.100 (0.232)	
Disease richness (Fincher & Thornhill)								-0.002 (0.010)
Marginal effect	2.418*** (0.898)	1.962* (1.162)	2.539*** (0.947)	2.513*** (0.923)	2.402*** (0.916)	2.316** (1.094)	2.616*** (1.002)	2.396*** (0.918)
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1163	873	1117	1154	1148	1083	1155	1163
Countries	142	140	137	141	140	128	141	142
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Wald test chi-squared	6.826	4.06	6.747	6.208	6.332	2.709	6.464	6.778
Wald test p-value	0.009	0.044	0.009	0.013	0.012	0.1	0.011	0.009



TABLE 10: Genetic Diversity and Civil Conflict Onset in Annual Panel Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit
	Civil Conflict Onset (PRIO2)				Onset of a New Civil Conflict (PRIO-NC)					
Genetic diversity (ancestry adjusted)	5.997** (3.020)	8.190** (3.515)	9.107** (3.647)	8.261** (3.618)	8.272** (3.803)	8.615*** (2.985)	9.066** (3.751)	9.554** (3.893)	9.976** (4.055)	9.301** (4.139)
Ethnic fractionalization			-0.421 (0.292)	-0.401 (0.294)	-0.398 (0.295)			-0.300 (0.308)	-0.331 (0.319)	-0.297 (0.324)
Ethnoling. polarization			0.458 (0.281)	0.411 (0.275)	0.393 (0.300)			0.217 (0.286)	0.206 (0.295)	0.092 (0.339)
Ethnic dominance					0.014 (0.104)					0.109 (0.119)
Political instability, lagged					0.124 (0.096)					0.136 (0.121)
Executive constraints, lagged		0.075* (0.041)	0.081* (0.042)	0.101** (0.043)	0.106** (0.044)		0.023 (0.047)	0.026 (0.049)	0.040 (0.050)	0.044 (0.052)
Democracy dummy, lagged		-0.509*** (0.177)	-0.516*** (0.185)	-0.534*** (0.187)	-0.530*** (0.190)		-0.314 (0.207)	-0.326 (0.217)	-0.331 (0.224)	-0.307 (0.229)
Autocracy dummy, lagged		-0.124 (0.140)	-0.121 (0.140)	-0.113 (0.140)	-0.073 (0.140)		-0.293* (0.155)	-0.292* (0.154)	-0.295* (0.151)	-0.252 (0.154)
Value of oil production per capita, lagged				0.052** (0.021)	0.053** (0.021)				0.019 (0.025)	0.021 (0.025)
Log population, lagged				0.002 (0.058)	-0.001 (0.059)				-0.003 (0.055)	-0.001 (0.056)
Log GDP per capita, lagged				-0.206*** (0.074)	-0.195*** (0.074)				-0.178* (0.103)	-0.175* (0.101)
Marginal effect	.472* (.251)	.642** (.306)	.713** (.318)	.641** (.307)	.642** (.323)	.489*** (.191)	.508** (.236)	.536** (.246)	.563** (.263)	.519** (.262)
Continent and geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peace duration cubic splines	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Observations	3907	3904	3904	3904	3900	3372	3369	3369	3369	3365
Countries	116	116	116	116	116	116	116	116	116	116
Period	1961-1999	1961-1999	1961-1999	1961-1999	1961-1999	1961-1999	1961-1999	1961-1999	1961-1999	1961-1999
Time frequency	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual
Wald test chi-squared	3.491	4.210	3.619	3.201	3.166	7.269	5.597	5.544	6.410	5.519
Wald test p-value	0.062	0.040	0.057	0.074	0.075	0.007	0.018	0.019	0.011	0.019

TABLE 11: Mediating Channels – Ethnic Fractionalization and Trust

	(1) 2SLS # New Civil Conflicts per year	(2) 2SLS Ethnic Fractionalization	(3) 2SLS # New Civil Conflicts per year	(4) 2SLS # New Civil Conflicts per year	(5) 2SLS Interpersonal Trust	(6) 2SLS # New Civil Conflicts per year
Genetic diversity (ancestry adjusted)	0.671*** (0.234)	2.075** (0.990)	0.642*** (0.231) 0.014 (0.014)	1.267*** (0.369)	-3.475*** (0.981)	1.205*** (0.398)
Ethnic fractionalization						
Interpersonal trust						-0.018 (0.037)
Absolute latitude	-0.635** (0.248)	-8.980*** (1.814)	-0.511* (0.276)	-0.988** (0.450)	4.113*** (1.585)	-0.915* (0.529)
Land area	1.492 (2.171)	-20.393*** (7.151)	1.773 (2.171)	-0.110 (2.283)	8.465 (6.244)	0.041 (2.291)
Ruggedness	0.028 (0.041)	-0.344 (0.226)	0.033 (0.043)	0.082 (0.067)	-0.084 (0.155)	0.080 (0.065)
Mean elevation	-0.016* (0.009)	0.035 (0.059)	-0.016* (0.009)	-0.046** (0.018)	0.011 (0.044)	-0.045** (0.018)
Range of elevation	0.011** (0.004)	0.027 (0.017)	0.010** (0.004)	0.018*** (0.006)	-0.004 (0.014)	0.018*** (0.006)
Mean land suitability	0.023* (0.013)	-0.281*** (0.080)	0.027** (0.013)	0.035** (0.017)	-0.210*** (0.074)	0.031 (0.019)
Range of land suitability	0.011 (0.012)	0.147* (0.079)	0.009 (0.013)	-0.000 (0.023)	0.051 (0.055)	0.001 (0.023)
Mean distance to nearest waterway	0.004 (0.009)	0.067* (0.040)	0.003 (0.009)	0.015 (0.017)	-0.091** (0.044)	0.014 (0.018)
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	No	No	No
Observations	150	150	150	77	77	77

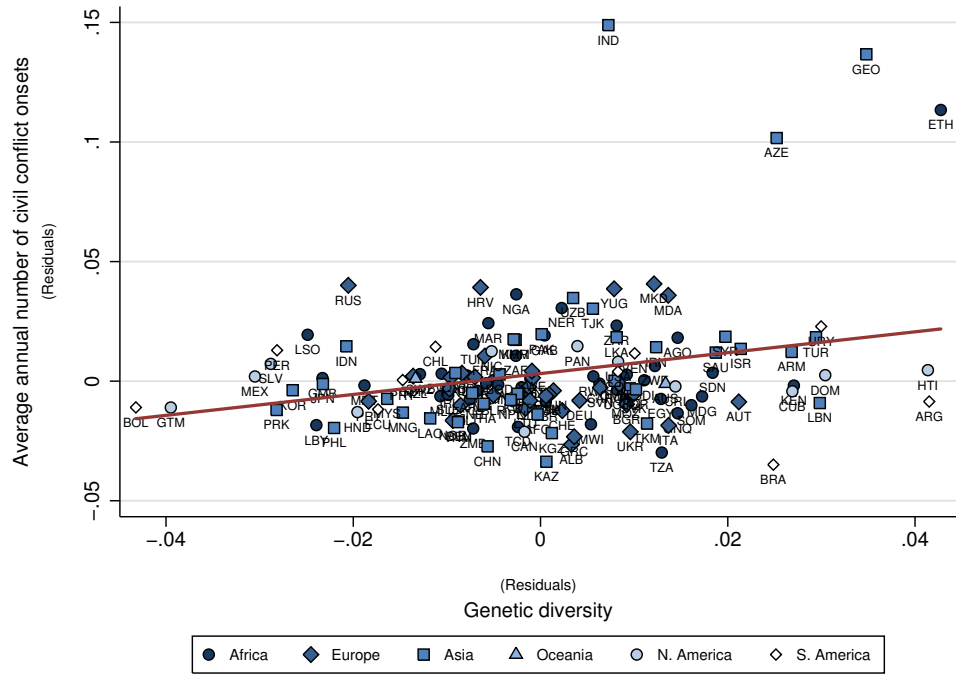


FIGURE 1: Genetic Diversity and the Number of Civil Conflict Onsets per Year

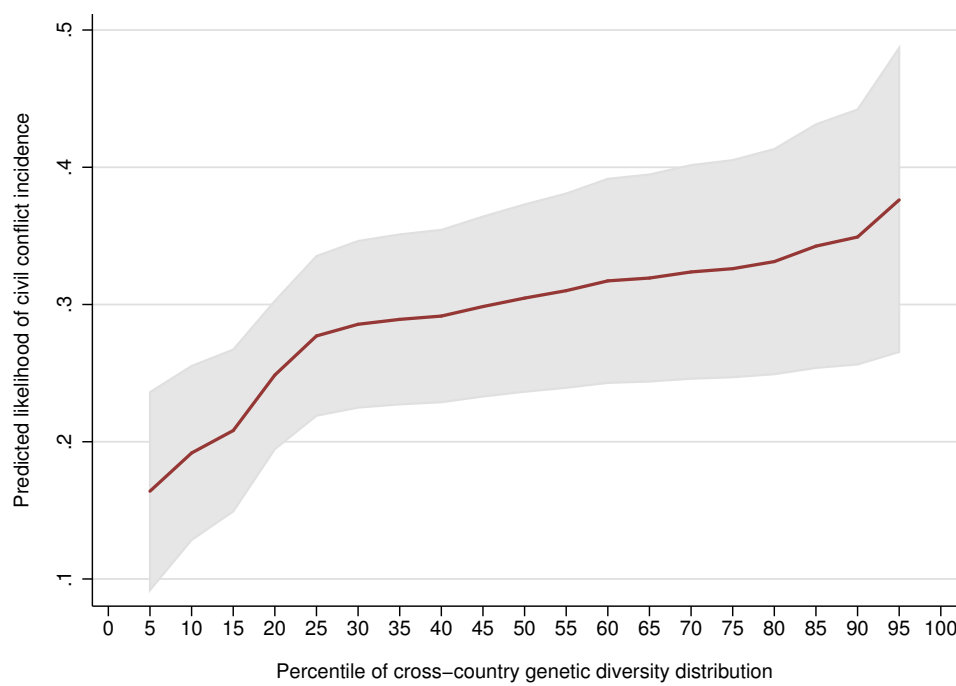


FIGURE 2: Genetic Diversity and the Likelihood of Civil Conflict Incidence

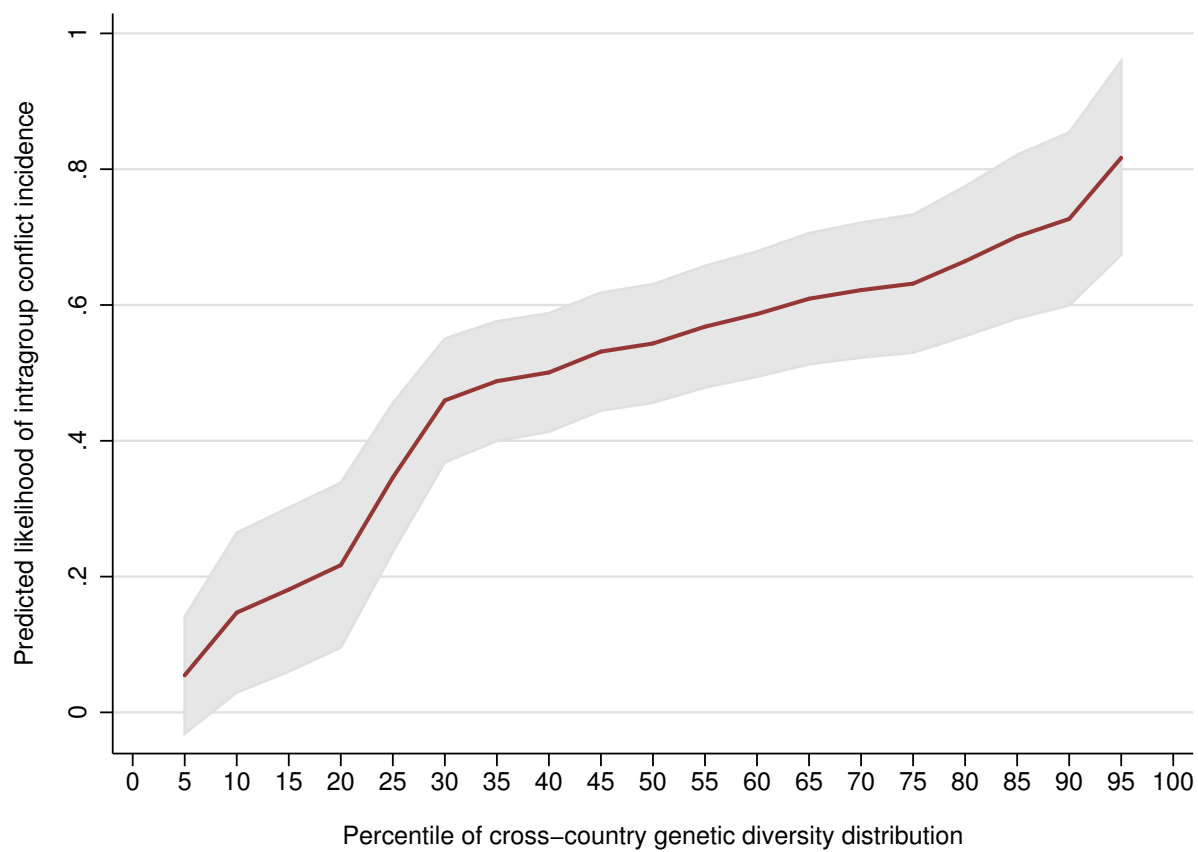
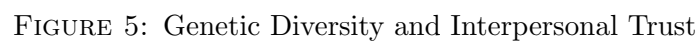
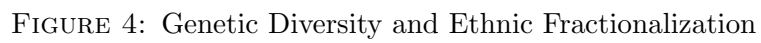


FIGURE 3: Genetic Diversity and the Likelihood of Intra-group Conflict Incidence



## A Supplementary Tables

TABLE A.1: Explaining Ethnic Civil Conflict under the OLS Estimator

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Number of new ethnic civil conflict onsets per year , 1960-2005							
Genetic diversity (ancestry adjusted)	0.217*** (0.082)	0.375** (0.152)	0.385** (0.183)	0.352** (0.176)	0.408** (0.189)	0.366* (0.192)	0.391** (0.191)
Ethnic fractionalization				0.021 (0.014)		0.019 (0.014)	0.016 (0.015)
Ethnoling. polarization					0.012 (0.015)	0.005 (0.016)	0.002 (0.015)
Executive constraints, 1960-2005 average			-0.000 (0.004)	-0.001 (0.004)	-0.000 (0.004)	-0.001 (0.004)	0.003 (0.004)
Fraction of years under democracy, 1960-2005			-0.011 (0.025)	-0.007 (0.024)	-0.011 (0.025)	-0.007 (0.024)	-0.008 (0.023)
Fraction of years under autocracy, 1960-2005			-0.015 (0.020)	-0.013 (0.020)	-0.014 (0.020)	-0.013 (0.020)	-0.015 (0.019)
Value of oil production per capita, 1960-2005 average							0.003** (0.001)
Log population, 1960-2005 average							-0.000 (0.003)
Log GDP per capita, 1960-2005 average							-0.021*** (0.005)
Effect of moving genetic diversity from 10 <sup>th</sup> to 90 <sup>th</sup> percentile	.014*** (.005)	.024** (.010)	.025** (.012)	.023** (.011)	.027** (.012)	.024* (.012)	.025** (.012)
Continental and geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	No	No	Yes	Yes	Yes	Yes	Yes
Observations	141	141	141	141	141	141	141
Adjusted R-squared	0.024	0.165	0.139	0.146	0.138	0.140	0.207

TABLE A.2: Genetic Diversity and Ethnic Civil Conflict – Addressing Endogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Number of new ethnic civil conflict onsets per year, 1960-2005							
Genetic diversity (ancestry adjusted)	0.362*** (0.090)	0.707*** (0.254)	0.753** (0.296)	0.718** (0.294)	0.758*** (0.290)	0.736** (0.300)	0.795*** (0.297)
Ethnic fractionalization				0.016 (0.012)		0.010 (0.014)	0.005 (0.014)
Ethnoling. polarization					0.017 (0.014)	0.013 (0.016)	0.011 (0.014)
Executive constraints, 1960-2005 average			0.001 (0.004)	0.001 (0.004)	0.001 (0.004)	0.001 (0.004)	0.005 (0.004)
Fraction of years under democracy, 1960-2005			-0.014 (0.022)	-0.011 (0.022)	-0.014 (0.022)	-0.012 (0.022)	-0.013 (0.021)
Fraction of years under autocracy, 1960-2005			-0.014 (0.019)	-0.013 (0.018)	-0.014 (0.018)	-0.013 (0.018)	-0.015 (0.017)
Value of oil production per capita, 1960-2005 average						0.003** (0.001)	0.003** (0.001)
Log population, 1960-2005 average						-0.000 (0.003)	-0.000 (0.003)
Log GDP per capita, 1960-2005 average						-0.022*** (0.005)	-0.022*** (0.005)
Effect of moving genetic diversity from 10 <sup>th</sup> to 90 <sup>th</sup> percentile	.024*** (.006)	.046*** (.016)	.049** (.019)	.047** (.019)	.049*** (.019)	.048** (.019)	.052*** (.019)
Continent and geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	No	No	Yes	Yes	Yes	Yes	Yes
Observations	141	141	141	141	141	141	141



TABLE A.3: Benchmark for Ethnic Civil Conflict Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
	Number of new civil conflict onsets per year, 1960-2005													
Genetic diversity (ancestry adjusted)	0.219** (0.110)	0.598*** (0.191)	0.615*** (0.231)	0.593** (0.227)	0.642*** (0.241)	0.623** (0.247)	0.665*** (0.240)	0.283** (0.119)	0.908*** (0.316)	0.978*** (0.359)	0.958*** (0.362)	0.983*** (0.354)	0.984*** (0.370)	1.100*** (0.352)
Ethnic fractionalization				0.014 (0.015)		0.008 (0.017)	0.007 (0.018)				0.009 (0.015)		-0.001 (0.017)	-0.004 (0.019)
Ethnoling. polarization					0.014 (0.018)	0.011 (0.020)	0.008 (0.018)					0.019 (0.017)	0.019 (0.019)	0.018 (0.017)
Executive constraints, 1960-2005 average			-0.004 (0.005)	-0.004 (0.005)	-0.004 (0.005)	-0.004 (0.005)	-0.000 (0.005)			-0.002 (0.005)	-0.003 (0.005)	-0.003 (0.005)	-0.003 (0.005)	0.002 (0.005)
Fraction of years under democracy, 1960-2005			-0.012 (0.030)	-0.009 (0.029)	-0.012 (0.030)	-0.010 (0.030)	-0.010 (0.028)			-0.015 (0.028)	-0.013 (0.027)	-0.014 (0.027)	-0.015 (0.027)	-0.016 (0.026)
Fraction of years under autocracy, 1960-2005			-0.028 (0.026)	-0.026 (0.025)	-0.027 (0.026)	-0.026 (0.025)	-0.030 (0.025)			-0.027 (0.024)	-0.027 (0.023)	-0.026 (0.024)	-0.026 (0.023)	-0.029 (0.022)
Value of oil production per capita, 1960-2005 average							0.003* (0.002)							0.003* (0.002)
Log population, 1960-2005 average							0.002 (0.004)							0.002 (0.004)
Log GDP per capita, 1960-2005 average							-0.026*** (0.008)							-0.027*** (0.007)
Effect of moving genetic diversity from 10 <sup>th</sup> to 90 <sup>th</sup> percentile	.014** (.007)	.039*** (.013)	.041*** (.015)	.039*** (.015)	.042*** (.016)	.041*** (.016)	.044*** (.016)	.019** (.008)	.06*** (.021)	.064*** (.024)	.063*** (.024)	.065*** (.023)	.065*** (.024)	.072*** (.023)
Continental and geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Observations	141	141	141	141	141	141	141	141	141	141	141	141	141	141
Adjusted R-squared	0.012	0.237	0.220	0.217	0.218	0.213	0.277	0.010	0.219	0.197	0.194	0.198	0.191	0.245

TABLE A.4: The Baseline Analysis under Linguistic Fractionalization

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	2SLS	2SLS	2SLS
Number of new civil conflict onsets per year (PRIO25), 1960-2008						
Genetic diversity (ancestry adjusted)	0.407*** (0.145)	0.417*** (0.147)	0.449*** (0.145)	0.774*** (0.271)	0.769*** (0.266)	0.838*** (0.270)
Linguistic fractionalization	0.017 (0.011)	0.015 (0.012)	0.016 (0.012)	0.019** (0.009)	0.015 (0.009)	0.015 (0.010)
Ethnoling. polarization		0.006 (0.013)	0.006 (0.013)		0.011 (0.013)	0.011 (0.013)
Executive constraints, 1960-2008 average	0.003 (0.004)	0.003 (0.004)	0.005 (0.004)	0.004 (0.003)	0.004 (0.003)	0.006* (0.003)
Fraction of years under democracy, 1960-2008	-0.010 (0.017)	-0.011 (0.017)	-0.009 (0.017)	-0.013 (0.018)	-0.013 (0.018)	-0.011 (0.017)
Fraction of years under autocracy, 1960-2008	-0.005 (0.015)	-0.004 (0.015)	-0.007 (0.015)	-0.002 (0.015)	-0.002 (0.015)	-0.003 (0.014)
Value of oil production per capita, 1960-2008 average			0.002* (0.001)			0.002** (0.001)
Log population, 1960-2008 average			0.003 (0.003)			0.003 (0.003)
Log GDP per capita, 1960-2008 average			-0.015*** (0.005)			-0.016*** (0.004)
Effect of moving genetic diversity from 10 <sup>th</sup> to 90 <sup>th</sup> percentile	.027** (.012)	.027** (.013)	.029** (.013)	.05*** (.018)	.05*** (.017)	.055*** (.018)
Continental and geographical controls	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	139	139	139	139	139	139
Adjusted R-squared	0.213	0.208	0.253	0.169	0.168	0.205
Sqr. Partial Correlation (Genetic Diversity)	.064	.066	.08	.064	.066	.08
Sum of Sqr. Partial Correlations (other diversity measures)	.022	.016	.018	.022	.016	.018

TABLE A.5: Ethnic Civil Conflict across Countries – Robustness to the Elimination of Regions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
	Number of new ethnic civil conflict onsets per year, 1960-2005							
Genetic diversity (ancestry adjusted)	0.875** (0.348)	0.972*** (0.364)	0.516** (0.238)	0.274** (0.134)	0.634*** (0.183)	0.768*** (0.238)	0.389*** (0.135)	0.586*** (0.211)
Ethnic fractionalization	0.017 (0.019)	0.007 (0.020)	0.007 (0.016)	0.020* (0.011)	-0.016 (0.018)	-0.009 (0.016)	0.011 (0.015)	-0.004 (0.019)
Ethnoling. polarization	0.016 (0.015)	0.021 (0.016)	0.000 (0.015)	0.006 (0.010)	0.018 (0.014)	0.033* (0.018)	0.015 (0.013)	0.018 (0.013)
Executive constraints, 1960-2005 average	0.011 (0.007)	0.014* (0.008)	0.002 (0.004)	0.000 (0.003)	0.005 (0.004)	0.004 (0.004)	0.002 (0.003)	0.003 (0.005)
Fraction of years under democracy, 1960-2005	-0.062 (0.038)	-0.070* (0.040)	-0.010 (0.023)	-0.022 (0.014)	-0.023 (0.023)	-0.003 (0.018)	-0.027 (0.020)	-0.026 (0.023)
Fraction of years under autocracy, 1960-2005	-0.022 (0.019)	-0.017 (0.020)	-0.023 (0.019)	-0.027* (0.015)	-0.028 (0.022)	-0.012 (0.018)	-0.019 (0.016)	-0.029 (0.020)
Value of oil production per capita, 1960-2005 average	0.001 (0.001)	0.002 (0.001)	0.003** (0.001)	0.002* (0.001)	0.002 (0.001)	0.004** (0.002)	0.003** (0.001)	0.003*** (0.001)
Log population, 1960-2005 average	-0.001 (0.004)	0.000 (0.004)	0.003 (0.004)	0.000 (0.002)	0.001 (0.003)	0.004 (0.003)	-0.000 (0.003)	0.002 (0.003)
Log GDP per capita, 1960-2005 average	-0.022*** (0.006)	-0.024*** (0.006)	-0.019*** (0.005)	-0.010* (0.005)	-0.022*** (0.006)	-0.019*** (0.007)	-0.014*** (0.005)	-0.018*** (0.005)
Continent dummies	No	No	No	No	No	No	No	No
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Omitted Region/Continent	Sub-Sah. Africa	Africa	LAAC	Asia	MENA	Europe	South Asia	E. Asia & Pacific
Observations	101	96	121	103	126	107	135	126

TABLE A.6: Robustness to Annual Variations in Overall Civil and Ethnic Conflict Incidence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit
	Civil Conflict Incidence (PRIO25)						Ethnic Civil Conflict Incidence (WCM 2009)					
Genetic diversity (ancestry adjusted)	10.632*** (3.521)	10.135*** (3.721)	10.751*** (3.757)	10.042*** (3.726)	11.180*** (3.850)	12.154*** (3.974)	12.926** (6.120)	13.282** (6.005)	13.703** (6.098)	12.947** (6.132)	13.881** (6.092)	13.106** (6.589)
Ethnic fractionalization			-0.349 (0.230)		-0.546** (0.269)	-0.496* (0.299)			-0.038 (0.559)		-0.316 (0.604)	-0.378 (0.624)
Ethnoling. polarization				0.243 (0.258)	0.431 (0.295)	0.470 (0.288)				0.462 (0.507)	0.596 (0.546)	0.617 (0.545)
Civil Conflict (PRIO), lagged	2.690*** (0.116)	2.654*** (0.114)	2.644*** (0.114)	2.656*** (0.114)	2.642*** (0.114)	2.622*** (0.116)						
Civil Conflict (WCM), lagged							2.482*** (0.275)	2.485*** (0.280)	2.491*** (0.280)	2.494*** (0.279)	2.487*** (0.282)	2.441*** (0.282)
Executive constraints, lagged		0.031 (0.035)	0.035 (0.036)	0.029 (0.035)	0.035 (0.036)	0.043 (0.037)		0.107* (0.062)	0.110* (0.063)	0.106* (0.063)	0.113* (0.063)	0.130* (0.068)
Democracy dummy, lagged		-0.272*** (0.137)	-0.296*** (0.139)	-0.262** (0.138)	-0.291*** (0.141)	-0.306*** (0.140)		-0.434*** (0.211)	-0.438*** (0.210)	-0.425*** (0.206)	-0.451*** (0.208)	-0.509*** (0.199)
Autocracy dummy, lagged		-0.274*** (0.112)	-0.273*** (0.114)	-0.276*** (0.111)	-0.273*** (0.111)	-0.281*** (0.114)		-0.077 (0.228)	-0.074 (0.230)	-0.075 (0.226)	-0.065 (0.226)	-0.051 (0.228)
Value of oil production per capita, lagged						0.004 (0.021)						0.015 (0.039)
Log population, lagged						0.059 (0.057)						-0.014 (0.109)
Log GDP per capita, lagged						-0.115 (0.071)						-0.123 (0.175)
Marginal effect	1.045*** (.381)	.976** (.395)	1.04*** (.403)	.963** (.393)	1.077*** (.412)	1.177*** (.434)	1.239** (.591)	1.232** (.586)	1.264** (.592)	1.198** (.586)	1.284** (.591)	1.246* (.645)
Continental and geographical controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin and colony dummies	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Observations	5797	5797	5797	5797	5797	5797	4855	4840	4840	4840	4840	4840
Countries	141	141	141	141	141	141	129	129	129	129	129	129
Period	1960-2008	1960-2008	1960-2008	1960-2008	1960-2008	1960-2008	1960-2005	1960-2005	1960-2005	1960-2005	1960-2005	1960-2005
Time frequency	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual
Wald test chi-squared	10.889	9.398	10.174	9.780	10.272	11.501	10.646	10.890	10.700	10.713	10.541	11.934
Wald test p-value	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

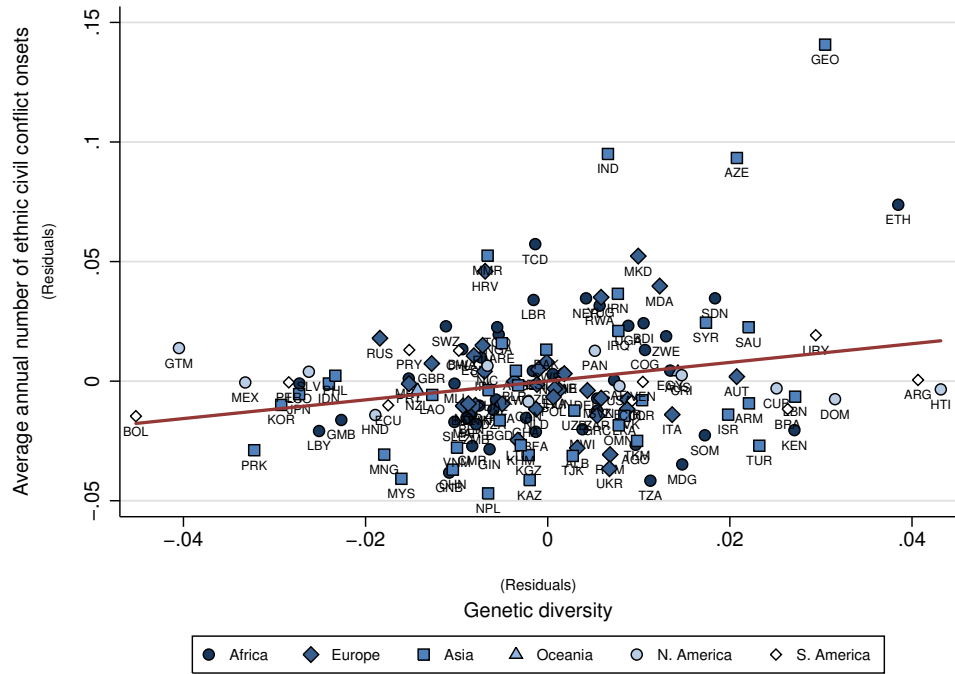


FIGURE A.1: Genetic Diversity and the Number of Ethnic Civil Conflict Onsets per Year

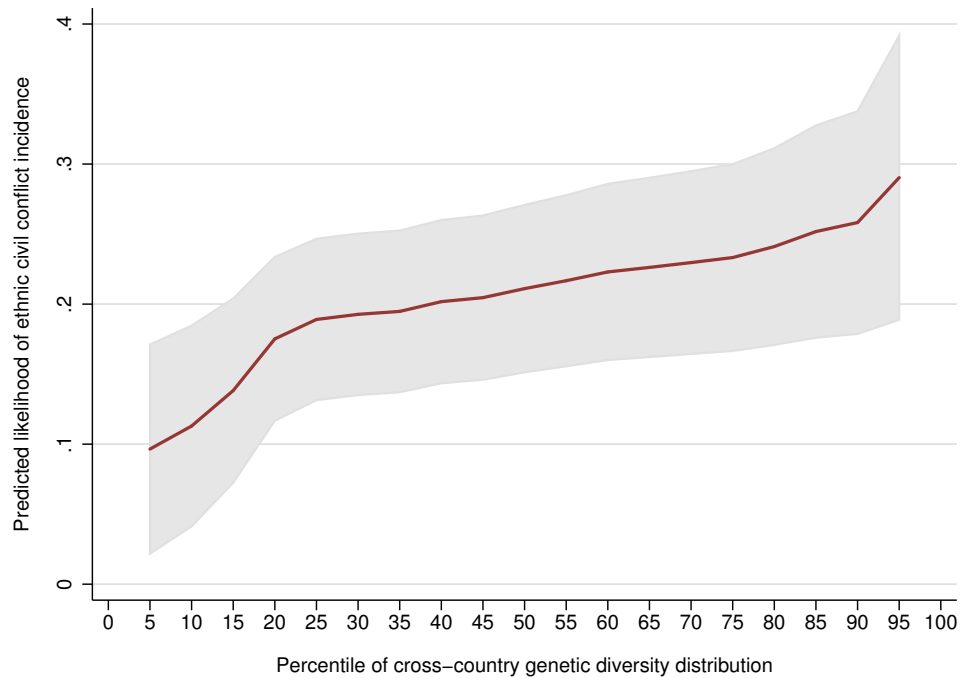


FIGURE A.2: Genetic Diversity and the Likelihood of Ethnic Civil Conflict Incidence

TABLE A.7: Genetic Diversity and Ethnic Civil Conflict Onset in Annual Panel Data

	(1)	(2)	(3)	(4)	(5)
	IV Probit	IV Probit	IV Probit	IV Probit	IV Probit
Ethnic Civil Conflict Onset (WCM 2009)					
Genetic diversity (ancestry adjusted)	8.570*** (3.223)	10.498*** (3.839)	10.646*** (4.000)	10.286** (4.240)	10.929** (4.485)
Ethnic fractionalization			0.196 (0.433)	0.143 (0.439)	0.063 (0.435)
Ethnoling. polarization			0.283 (0.377)	0.311 (0.414)	0.334 (0.430)
Ethnic dominance					0.029 (0.152)
Political instability, lagged					0.102 (0.142)
Executive constraints, lagged		0.054 (0.054)	0.055 (0.054)	0.074 (0.058)	0.079 (0.059)
Democracy dummy, lagged		-0.454** (0.183)	-0.431** (0.187)	-0.505*** (0.196)	-0.508** (0.199)
Autocracy dummy, lagged		-0.381** (0.183)	-0.382** (0.180)	-0.395** (0.187)	-0.377** (0.188)
Value of oil production per capita, lagged				0.025 (0.030)	0.025 (0.029)
Log population, lagged				0.045 (0.102)	0.037 (0.104)
Log GDP per capita, lagged				-0.215 (0.148)	-0.209 (0.146)
Marginal effect	.466** (.195)	.582** (.248)	.578** (.248)	.637** (.286)	.680** (.301)
Continental and geographical controls	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Peace duration cubic splines	Yes	Yes	Yes	Yes	Yes
Lagged colony dummies	No	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	Yes	Yes	Yes
Observations	3010	3007	3007	3007	3004
Countries	106	106	106	106	106
Period	1960-1999	1960-1999	1960-1999	1960-1999	1960-1999
Time frequency	Annual	Annual	Annual	Annual	Annual
Wald test chi-squared	8.576	8.645	8.171	7.693	8.609
Wald test p-value	0.003	0.003	0.004	0.006	0.003

## B Descriptive Statistics

TABLE A.8: Mediating Channels – Ethnic Fractionalization and Interpersonal Trust (OLS Results)

	(1) OLS # New Civil Conflicts per year	(2) OLS Ethnic Fractionalization	(3) OLS # New Civil Conflicts per year	(4) OLS # New Civil Conflicts per year	(5) OLS Interpersonal Trust	(6) OLS # New Civil Conflicts per year
Genetic diversity (ancestry adjusted)	0.327** (0.142)	1.545** (0.763)	0.298** (0.140)	0.816** (0.331)	-1.708** (0.850)	0.756** (0.348)
Ethnic fractionalization			0.019 (0.015)			
Interpersonal trust						-0.035 (0.044)
Absolute latitude	-0.517** (0.256)	-8.799*** (1.940)	-0.353 (0.292)	-0.763 (0.493)	3.230** (1.558)	-0.650 (0.572)
Land area	1.816 (2.168)	-19.892** (7.738)	2.187 (2.168)	0.490 (2.806)	6.118 (7.753)	0.703 (2.719)
Ruggedness	0.024 (0.043)	-0.350 (0.238)	0.031 (0.046)	0.069 (0.072)	-0.032 (0.183)	0.068 (0.069)
Mean elevation	-0.014 (0.009)	0.038 (0.062)	-0.014 (0.009)	-0.037* (0.020)	-0.022 (0.052)	-0.038* (0.019)
Range of elevation	0.010** (0.004)	0.026 (0.018)	0.009** (0.005)	0.016** (0.007)	0.004 (0.013)	0.016** (0.007)
Mean land suitability	0.022* (0.013)	-0.282*** (0.085)	0.027** (0.014)	0.036** (0.018)	-0.215** (0.081)	0.029 (0.022)
Range of land suitability	0.007 (0.012)	0.141* (0.083)	0.005 (0.012)	-0.002 (0.025)	0.057 (0.058)	0.000 (0.025)
Mean distance to nearest waterway	0.005 (0.009)	0.069 (0.042)	0.004 (0.010)	0.015 (0.020)	-0.088* (0.046)	0.011 (0.019)
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	No	No	No
Observations	150	150	150	77	77	77
Adjusted R-squared	0.146	0.442	0.152	0.175	0.443	0.169



TABLE B.1: Summary Statistics for Baseline Variables in the Cross-country Regressions

[Panel A: Overall Civil Conflict Regressions]						[Panel B: Ethnic Civil Conflict Regressions]					
	Obs	Mean	SD	Min	Max		Obs	Mean	SD	Min	Max
Log(1+ # new civ. conf. onsets per year), PRIO25	143	0.020	0.030	0.000	0.186	Log(1+ # new ethnic civ. conf. onsets per year), WCM 2009	141	0.018	0.033	0.000	0.182
Genetic diversity	143	0.728	0.027	0.628	0.774	Genetic diversity	141	0.727	0.027	0.628	0.774
Ethnic fractionalization	143	0.461	0.259	0.002	0.930	Ethnic fractionalization	141	0.460	0.259	0.002	0.930
Ethnoling. polarization	143	0.450	0.241	0.000	0.957	Ethnoling. polarization	141	0.446	0.240	0.000	0.957
Absolute latitude (x 0.001)	143	0.028	0.017	0.001	0.064	Absolute latitude (x 0.001)	141	0.028	0.017	0.001	0.064
Land area (x 0.001)	143	0.001	0.002	0.000	0.016	Land area (x 0.001)	141	0.001	0.002	0.000	0.016
Ruggedness (x 0.001)	143	0.122	0.117	0.004	0.585	Ruggedness (x 0.001)	141	0.124	0.118	0.004	0.585
Mean elevation (x 0.001)	143	0.585	0.537	0.001	2.837	Mean elevation (x 0.001)	141	0.591	0.538	0.001	2.837
Range of elevation (x 0.001)	143	1.685	1.370	0.040	6.176	Range of elevation (x 0.001)	141	1.706	1.367	0.043	6.176
Mean land suitability	143	0.388	0.248	0.003	0.951	Mean land suitability	141	0.393	0.246	0.003	0.951
Range of land suitability	143	0.717	0.266	0.000	0.999	Range of land suitability	141	0.726	0.256	0.002	0.999
Mean distance to nearest waterway (x 0.001)	143	0.359	0.463	0.014	2.386	Mean distance to nearest waterway (x 0.001)	141	0.363	0.465	0.020	2.386
Ever a U.K. colony	143	0.245	0.431	0.000	1.000	Ever a U.K. colony	141	0.241	0.429	0.000	1.000
Ever a French colony	143	0.196	0.398	0.000	1.000	Ever a French colony	141	0.199	0.400	0.000	1.000
Ever a non-U.K./non-French colony	143	0.308	0.463	0.000	1.000	Ever a non-U.K./non-French colony	141	0.305	0.462	0.000	1.000
British Legal Origin	143	0.238	0.427	0.000	1.000	British Legal Origin	141	0.241	0.429	0.000	1.000
French Legal Origin	143	0.462	0.500	0.000	1.000	French Legal Origin	141	0.454	0.500	0.000	1.000
Executive constraints	143	3.939	1.884	1.000	7.000	Executive constraints	141	3.903	1.891	1.000	7.000
Fraction of years under democracy	143	0.377	0.384	0.000	1.000	Fraction of years under democracy	141	0.368	0.388	0.000	1.000
Fraction of years under autocracy	143	0.393	0.337	0.000	1.000	Fraction of years under autocracy	141	0.403	0.343	0.000	1.000
Log(1+Value of oil production per capita)	143	2.507	2.758	0.000	9.748	Log(1+ Value of oil production per capita)	141	2.335	2.652	0.000	9.743
Log population	143	9.163	1.405	5.863	13.841	Log population	141	9.179	1.363	6.579	13.822
Log GDP per capita	143	8.051	1.057	6.210	9.943	Log GDP per capita	141	8.003	1.042	6.180	9.910
Africa dummy	143	0.322	0.469	0.000	1.000	Africa dummy	141	0.319	0.468	0.000	1.000
Asia dummy	143	0.273	0.447	0.000	1.000	Asia dummy	141	0.270	0.445	0.000	1.000
North America dummy	143	0.084	0.278	0.000	1.000	North America dummy	141	0.085	0.280	0.000	1.000
South America dummy	143	0.070	0.256	0.000	1.000	South America dummy	141	0.071	0.258	0.000	1.000
Oceania Dummy	143	0.014	0.118	0.000	1.000	Oceania Dummy	141	0.014	0.119	0.000	1.000

TABLE B.2: Summary Statistics for Baseline Variables in the 5-yearly Panel Regressions

[Panel A: Overall Civil Conflict Regressions]						[Panel B: Ethnic Civil Conflict Regressions]					
	Obs	Mean	SD	Min	Max		Obs	Mean	SD	Min	Max
Civil conflict incidence	1154	0.273	0.446	0.000	1.000	Ethnic civil conflict incidence (WCM 2009)	1039	0.183	0.387	0.000	1.000
Genetic diversity	1154	0.726	0.028	0.628	0.774	Genetic diversity	1039	0.730	0.023	0.650	0.774
Ethnic fractionalization	1154	0.463	0.263	0.002	0.930	Ethnic fractionalization	1039	0.462	0.269	0.002	0.930
Ethnoling. polarization	1154	0.437	0.241	0.000	0.957	Ethnoling. polarization	1039	0.444	0.235	0.000	0.957
Civil conflict incidence (PRIO25), lagged	1154	0.263	0.441	0.000	1.000	Civil conflict incidence (WCM 2009), lagged	1039	0.254	0.436	0.000	1.000
Absolute latitude (x 0.001)	1154	0.026	0.017	0.001	0.064	Absolute latitude (x 0.001)	1039	0.027	0.017	0.001	0.064
Land area (x 0.001)	1154	0.001	0.002	0.000	0.016	Land area (x 0.001)	1039	0.001	0.002	0.000	0.016
Ruggedness (x 0.001)	1154	0.119	0.109	0.004	0.585	Ruggedness (x 0.001)	1039	0.123	0.113	0.004	0.585
Mean elevation (x 0.001)	1154	0.574	0.497	0.001	2.837	Mean elevation (x 0.001)	1039	0.568	0.502	0.001	2.837
Range of elevation (x 0.001)	1154	1.739	1.401	0.040	6.176	Range of elevation (x 0.001)	1039	1.619	1.267	0.043	6.176
Mean land suitability	1154	0.388	0.250	0.003	0.951	Mean land suitability	1039	0.393	0.251	0.003	0.929
Range of land suitability	1154	0.721	0.271	0.000	0.999	Range of land suitability	1039	0.711	0.262	0.002	0.999
Mean distance to nearest waterway (x 0.001)	1154	0.339	0.424	0.014	2.386	Mean distance to nearest waterway (x 0.001)	1039	0.354	0.442	0.020	2.386
Fraction of years as U.K. colony, lagged	1154	0.011	0.090	0.000	1.000	Fraction of years as U.K. colony, lagged	1039	0.012	0.094	0.000	1.000
Fraction of years as French colony, lagged	1154	0.003	0.030	0.000	0.600	Fraction of years as French colony, lagged	1039	0.004	0.032	0.000	0.600
Fraction of years as non-U.K./non-French colony, lagged	1154	0.005	0.057	0.000	1.000	Fraction of years as non-U.K./non-French colony, lagged	1039	0.004	0.055	0.000	1.000
British Legal Origin	1154	0.240	0.427	0.000	1.000	British Legal Origin	1039	0.267	0.442	0.000	1.000
French Legal Origin	1154	0.511	0.500	0.000	1.000	French Legal Origin	1039	0.457	0.498	0.000	1.000
Executive constraints	1154	3.767	2.243	1.000	7.000	Executive constraints	1039	3.747	2.240	1.000	7.000
Fraction of years under democracy, lagged	1154	0.341	0.458	0.000	1.000	Fraction of years under democracy, lagged	1039	0.326	0.454	0.000	1.000
Fraction of years under autocracy, lagged	1154	0.441	0.469	0.000	1.000	Fraction of years under autocracy, lagged	1039	0.457	0.472	0.000	1.000
Log(1+Value of oil production per capita), lagged	1154	2.096	2.674	0.000	10.662	Log(1+Value of oil production per capita), lagged	1039	1.907	2.603	0.000	10.662
Log population, lagged	1154	9.219	1.457	4.905	14.062	Log population, lagged	1039	9.252	1.433	5.839	14.062
Log GDP per capita, lagged	1154	7.915	1.073	5.954	10.536	Log GDP per capita, lagged	1039	7.864	1.098	5.954	10.271
Africa dummy	1154	0.336	0.473	0.000	1.000	Africa dummy	1039	0.366	0.482	0.000	1.000
Asia dummy	1154	0.259	0.438	0.000	1.000	Asia dummy	1039	0.279	0.449	0.000	1.000
North America dummy	1154	0.097	0.296	0.000	1.000	North America dummy	1039	0.108	0.310	0.000	1.000
South America dummy	1154	0.085	0.279	0.000	1.000	South America dummy	1039	0.000	0.000	0.000	0.000
Oceania Dummy	1154	0.000	0.000	0.000	0.000	Oceania Dummy	1039	0.000	0.000	0.000	0.000

TABLE B.3: Pairwise Correlations among the Baseline Variables in the Cross-sectional Civil Conflict Regressions

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Log(1+ # new civ. conf. onsets per year), PRIO25	1.00													
2 Genetic diversity	0.16	1.00												
3 Ethnic fractionalization	0.20	0.20	1.00											
4 Ethnoling. polarization	0.05	0.10	0.24	1.00										
5 Absolute latitude (x 0.001)	-0.14	0.08	-0.55	0.20	1.00									
6 Land area (x 0.001)	0.18	-0.06	0.00	0.02	0.11	1.00								
7 Ruggedness (x 0.001)	0.06	-0.20	-0.26	0.02	0.19	-0.11	1.00							
8 Mean elevation (x 0.001)	0.14	-0.01	0.12	0.13	-0.05	0.05	0.64	1.00						
9 Range of elevation (x 0.001)	0.29	-0.32	0.10	0.10	-0.03	0.36	0.41	0.60	1.00					
10 Mean land suitability	-0.02	-0.23	-0.35	-0.34	0.06	-0.25	0.28	-0.12	-0.14	1.00				
11 Range of land suitability	0.19	-0.20	0.00	-0.16	0.16	0.28	0.15	0.20	0.43	0.11	1.00			
12 Mean distance to nearest waterway (x 0.001)	0.21	0.19	0.26	0.14	0.00	0.44	-0.02	0.41	0.33	-0.40	0.13	1.00		
13 Ever a U.K. colony	0.05	0.24	0.25	0.18	-0.22	0.10	-0.20	-0.02	-0.02	-0.24	-0.15	0.02	1.00	
14 Ever a French colony	0.06	0.14	0.25	-0.06	-0.24	0.09	-0.18	-0.15	-0.10	-0.19	-0.03	0.12	-0.08	1.00
15 Ever a non-U.K./non-French colony	0.12	-0.35	0.17	-0.30	-0.51	0.05	-0.06	0.01	0.21	0.14	-0.03	-0.15	-0.06	-0.06
16 British Legal Origin	0.06	0.16	0.19	0.09	-0.19	0.10	-0.13	0.06	0.01	-0.14	0.00	0.00	0.71	-0.15
17 French Legal Origin	-0.01	-0.12	0.18	-0.13	-0.41	-0.08	-0.15	-0.08	0.02	-0.04	-0.16	-0.08	-0.30	0.32
18 Executive constraints	-0.10	-0.21	-0.35	-0.01	0.50	0.05	0.17	-0.09	-0.03	0.24	0.28	-0.31	-0.03	-0.28
19 Fraction of years under democracy	-0.16	-0.21	-0.37	-0.04	0.47	0.04	0.07	-0.19	-0.07	0.25	0.22	-0.33	-0.07	-0.27
20 Fraction of years under autocracy	0.05	0.24	0.22	0.08	-0.28	0.02	-0.20	-0.01	-0.02	-0.31	-0.29	0.21	0.08	0.24
21 Log(1+ Value of oil production per capita)	0.00	0.00	0.01	0.18	0.10	0.28	-0.23	-0.18	0.11	-0.39	-0.05	0.08	0.08	0.00
22 Log population	0.31	-0.19	-0.12	-0.11	0.06	0.47	0.04	0.08	0.52	0.12	0.59	0.09	0.04	-0.02
23 Log GDP per capita	-0.29	-0.15	-0.49	0.09	0.66	0.12	0.04	-0.27	-0.09	0.00	0.08	-0.30	-0.13	-0.30
24 Africa dummy	0.11	0.49	0.52	-0.05	-0.56	-0.08	-0.37	0.04	-0.23	-0.28	-0.18	0.19	0.27	0.38
25 Asia dummy	0.15	-0.14	-0.10	0.25	0.05	-0.03	0.28	0.23	0.30	-0.14	-0.06	0.16	0.09	-0.06
26 North America dummy	-0.09	-0.34	-0.05	-0.23	-0.11	0.12	0.14	-0.09	0.06	0.31	-0.12	-0.10	-0.06	0.04
27 South America dummy	-0.09	-0.50	-0.01	-0.15	-0.17	0.11	-0.07	0.06	0.30	0.02	0.22	-0.08	-0.16	-0.14
28 Oceania Dummy	-0.08	-0.04	-0.10	-0.12	0.04	0.18	-0.01	-0.05	-0.01	-0.06	0.12	-0.04	0.21	-0.06
15 Ever a non-U.K./non-French colony	1.00													
16 British Legal Origin	-0.05	1.00												
17 French Legal Origin	0.42	-0.52	1.00											
18 Executive constraints	-0.07	0.12	-0.29	1.00										
19 Fraction of years under democracy	-0.04	0.09	-0.20	0.92	1.00									
20 Fraction of years under autocracy	-0.05	-0.12	0.15	-0.85	-0.80	1.00								
21 Log(1+ Value of oil production per capita)	-0.03	-0.09	0.08	-0.13	-0.09	0.24	1.00							
22 Log population	0.04	0.08	-0.07	0.11	0.06	-0.06	0.06	1.00						
23 Log GDP per capita	-0.17	-0.12	-0.15	0.60	0.60	-0.39	0.44	0.01	1.00					
24 Africa dummy	0.09	0.25	0.20	-0.43	-0.42	0.31	-0.19	-0.21	-0.63	1.00				
25 Asia dummy	-0.20	0.03	-0.19	-0.22	-0.28	0.22	0.14	0.18	0.00	-0.42	1.00			
26 North America dummy	0.41	-0.05	0.18	0.07	0.10	-0.14	-0.06	0.00	0.05	-0.21	-0.19	1.00		
27 South America dummy	0.41	-0.15	0.30	0.10	0.18	-0.16	0.13	0.08	0.10	-0.19	-0.17	-0.08	1.00	
28 Oceania Dummy	-0.08	0.21	-0.11	0.19	0.19	-0.14	0.10	-0.02	0.18	-0.08	-0.07	-0.04	-0.03	1.00

TABLE B.4: Pairwise Correlations among the Baseline Variables in the 5-yearly Panel Sample

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Civil conflict incidence (PRIO25)	1.00													
2 Genetic diversity	-0.03	1.00												
3 Ethnic fractionalization	0.17	0.19	1.00											
4 Ethnoling. polarization	0.03	0.09	0.23	1.00										
5 Civil conflict incidence (PRIO25), lagged	0.63	-0.04	0.16	0.04	1.00									
6 Absolute latitude (x 0.001)	-0.23	0.09	-0.58	0.15	-0.22	1.00								
7 Land area (x 0.001)	0.02	-0.05	0.01	0.06	0.00	0.16	1.00							
8 Ruggedness (x 0.001)	0.02	-0.23	-0.30	0.01	0.02	0.20	-0.09	1.00						
9 Mean elevation (x 0.001)	0.09	-0.03	0.13	0.14	0.09	-0.06	0.09	0.58	1.00					
10 Range of elevation (x 0.001)	0.21	-0.35	0.09	0.11	0.21	-0.01	0.39	0.39	0.60	1.00				
11 Mean land suitability	-0.04	-0.25	-0.34	-0.38	-0.02	0.03	-0.25	0.32	-0.11	-0.12	1.00			
12 Range of land suitability	0.13	-0.21	0.02	-0.15	0.12	0.17	0.27	0.14	0.21	0.43	0.11	1.00		
13 Mean distance to nearest waterway (x 0.001)	0.10	0.22	0.28	0.11	0.07	-0.02	0.53	-0.11	0.37	0.25	-0.41	0.11	1.00	
14 Fraction of years as U.K. colony, lagged	-0.01	0.07	0.07	0.04	0.02	-0.09	-0.03	-0.05	0.04	-0.04	-0.03	-0.04	0.02	1.00
15 Fraction of years as French colony, lagged	-0.05	0.06	0.09	0.00	-0.04	-0.07	-0.01	-0.08	-0.03	-0.03	-0.08	-0.02	0.06	0.01
16 Fraction of years as non-U.K./non-French colony, lagged	0.03	0.04	0.01	-0.03	0.01	-0.07	0.00	-0.01	0.02	0.02	0.01	-0.04	0.01	0.00
17 British Legal Origin	0.12	0.19	0.22	0.16	0.11	-0.16	0.07	-0.11	0.09	0.02	-0.12	0.00	0.06	0.19
18 French Legal Origin	0.07	-0.13	0.17	-0.09	0.05	-0.36	-0.10	-0.15	-0.07	0.02	-0.04	-0.16	-0.06	-0.10
19 Executive constraints	-0.07	-0.18	-0.29	0.00	-0.06	0.41	0.04	0.15	-0.07	0.04	0.17	0.24	-0.21	0.05
20 Fraction of years under democracy, lagged	-0.11	-0.17	-0.31	-0.03	-0.10	0.37	0.02	0.10	-0.13	0.01	0.18	0.19	-0.21	0.00
21 Fraction of years under autocracy, lagged	-0.04	0.22	0.18	0.06	-0.05	-0.19	0.02	-0.13	0.02	-0.05	-0.22	-0.22	0.16	-0.04
22 Log(1+Value of oil production per capita), lagged	0.00	-0.01	0.02	0.23	0.01	0.11	0.27	-0.17	-0.11	0.14	-0.37	-0.05	0.05	-0.03
23 Log population, lagged	0.22	-0.17	-0.15	-0.08	0.22	0.15	0.47	0.10	0.13	0.53	0.12	0.60	0.10	-0.09
24 Log GDP per capita, lagged	-0.22	-0.14	-0.45	0.07	-0.21	0.63	0.10	0.06	-0.27	-0.07	0.00	0.08	-0.31	-0.09
25 Africa dummy	0.03	0.52	0.52	-0.02	0.01	-0.53	-0.10	-0.39	0.05	-0.27	-0.28	-0.18	0.26	0.11
26 Asia dummy	0.22	-0.16	-0.10	0.25	0.21	0.03	-0.02	0.24	0.17	0.27	-0.14	-0.08	0.03	-0.01
27 North America dummy	-0.03	-0.32	-0.05	-0.22	-0.02	-0.08	0.13	0.16	-0.10	0.06	0.32	-0.12	-0.09	-0.04
28 South America dummy	-0.01	-0.51	0.00	-0.15	-0.01	-0.17	0.11	-0.07	0.08	0.33	0.02	0.24	-0.08	-0.04
15 Fraction of years as French colony, lagged	1.00													
16 Fraction of years as non-U.K./non-French colony, lagged	0.07	1.00												
17 British Legal Origin	-0.06	0.01	1.00											
18 French Legal Origin	0.11	0.03	-0.57	1.00										
19 Executive constraints	-0.06	0.00	0.11	-0.22	1.00									
20 Fraction of years under democracy, lagged	-0.08	-0.01	0.07	-0.14	0.88	1.00								
21 Fraction of years under autocracy, lagged	0.06	-0.02	-0.10	0.07	-0.77	-0.68	1.00							
22 Log(1+Value of oil production per capita), lagged	-0.04	-0.06	-0.10	0.11	-0.08	-0.04	0.11	1.00						
23 Log population, lagged	-0.07	-0.02	0.08	-0.15	0.20	0.14	-0.15	0.08	1.00					
24 Log GDP per capita, lagged	-0.09	-0.09	-0.14	-0.08	0.51	0.51	-0.33	0.45	0.07	1.00				
25 Africa dummy	0.15	0.07	0.23	0.16	-0.35	-0.35	0.25	-0.22	-0.27	-0.58	1.00			
26 Asia dummy	-0.06	0.00	0.09	-0.20	-0.12	-0.15	0.13	0.15	0.21	-0.03	-0.42	1.00		
27 North America dummy	-0.04	-0.03	-0.05	0.14	0.07	0.09	-0.12	-0.04	-0.02	0.08	-0.23	-0.19	1.00	
28 South America dummy	-0.03	-0.02	-0.17	0.30	0.09	0.15	-0.15	0.15	0.05	0.13	-0.22	-0.18	-0.10	1.00

TABLE B.5: List of Countries with at least One New Civil Conflict Onset (PRIO25) over the period 1960-2008

	New Civil Conflict Onsets			Civil Conflict		New Civil Conflict Onsets			Civil Conflict
	Total	Avg. Number	Incidence			Total	Avg. Number	Incidence	
	Number	per year	% years			Number	per year	% years	
<b>Africa</b>					<b>Asia (cont.)</b>				
Burkina Faso	1	0.0204	2.0		Nepal	1	0.0204	28.6	
Cameroon	1	0.0204	6.1		Oman	1	0.0204	14.3	
Central African Republic	1	0.0204	6.1		Philippines	1	0.0204	81.6	
Chad	1	0.0204	69.4		Saudi Arabia	1	0.0204	2.0	
Congo, Rep.	1	0.0204	10.2		Syrian Arab Republic	1	0.0204	10.2	
Cte d'Ivoire	1	0.0204	6.1		Thailand	1	0.0204	30.6	
Egypt, Arab Rep.	1	0.0204	12.2		Bangladesh	1	0.0263	47.4	
Gabon	1	0.0204	2.0		Iran, Islamic Rep.	2	0.0408	53.1	
Ghana	1	0.0204	6.1		Myanmar	2	0.0408	98.0	
Guinea	1	0.0204	4.1		Sri Lanka	2	0.0408	49.0	
Liberia	1	0.0204	14.3		Turkey	2	0.0408	51.0	
Madagascar	1	0.0204	2.0		Tajikistan	1	0.0556	33.3	
Mali	1	0.0204	8.2		Uzbekistan	1	0.0556	16.7	
Mauritania	1	0.0204	8.2		Indonesia	3	0.0612	67.3	
Senegal	1	0.0204	18.4		Pakistan	3	0.0612	26.5	
Somalia	1	0.0204	38.8		Azerbaijan	2	0.1111	33.3	
Togo	1	0.0204	2.0		Georgia	3	0.1667	27.8	
Tunisia	1	0.0204	2.0		India	10	0.2041	83.7	
Sierra Leone	1	0.0208	20.8						
Algeria	1	0.0213	38.3		<b>Europe</b>				
Burundi	1	0.0213	36.2		France	1	0.0204	4.1	
Rwanda	1	0.0213	25.5		Romania	1	0.0204	2.0	
Uganda	1	0.0213	68.1		Spain	1	0.0204	20.4	
Kenya	1	0.0217	2.2		United Kingdom	1	0.0204	44.9	
Gambia, The	1	0.0227	2.3		Croatia	1	0.0556	16.7	
Zimbabwe	1	0.0227	20.5		Macedonia, FYR	1	0.0556	5.6	
Lesotho	1	0.0233	2.3		Moldova	1	0.0556	5.6	
Guinea-Bissau	1	0.0286	5.7		Serbia and Montenegro	3	0.0612	6.1	
Mozambique	1	0.0294	47.1		Russian Federation	6	0.1224	32.7	
Djibouti	1	0.0313	15.6						
Morocco	2	0.0408	32.7		<b>N. America</b>				
South Africa	2	0.0408	46.9		Dominican Republic	1	0.0204	2.0	
Sudan	2	0.0408	75.5		El Salvador	1	0.0204	28.6	
Angola	2	0.0588	88.2		Haiti	1	0.0204	6.1	
Niger	3	0.0612	14.3		Mexico	1	0.0204	4.1	
Congo, Dem. Rep.	4	0.0816	34.7		Nicaragua	1	0.0204	24.5	
Nigeria	4	0.0816	12.2		Panama	1	0.0204	2.0	
Ethiopia	8	0.1633	91.8		United States	1	0.0204	14.3	
<b>Asia</b>					<b>S. America</b>				
Afghanistan	1	0.0204	61.2		Chile	1	0.0204	2.0	
Cambodia	1	0.0204	61.2		Colombia	1	0.0204	91.8	
Iraq	1	0.0204	79.6		Peru	1	0.0204	42.9	
Israel	1	0.0204	100.0		Uruguay	1	0.0204	2.0	
Malaysia	1	0.0204	16.3		Venezuela, RB	1	0.0204	6.1	

TABLE B.6: List of Countries with at least One New Ethnic Civil Conflict Onset (WCM 2009) over the period 1960-2005

	Avg. Number of New Ethnic Conflict Onsets per year	Ethnic Incidence % years		Avg. Number of New Ethnic Conflict Onsets per year	Ethnic Incidence % years
<b>Africa</b>			<b>Asia (cont.)</b>		
Cameroon	0.0217	2.2	Turkey	0.0217	47.8
Central African Republic	0.0217	4.3	Bangladesh	0.0294	55.9
Congo, Rep.	0.0217	21.7	Philippines	0.0435	78.3
Cte d'Ivoire	0.0217	6.5	Syrian Arab Republic	0.0435	10.9
Mali	0.0217	10.9	Afghanistan	0.0652	21.7
Senegal	0.0217	30.4	Indonesia	0.0652	93.5
South Africa	0.0217	17.4	Iraq	0.0652	82.6
Zimbabwe	0.0244	19.5	Pakistan	0.0652	26.1
Angola	0.0323	45.2	Iran, Islamic Rep.	0.0870	47.8
Mozambique	0.0323	51.6	Myanmar	0.1087	100.0
Congo, Dem. Rep.	0.0435	10.9	Azerbaijan	0.1333	26.7
Niger	0.0435	13.0	India	0.1739	80.4
Togo	0.0435	4.3	Georgia	0.2000	20.0
Burundi	0.0455	36.4			
Rwanda	0.0455	25.0	<b>Europe</b>		
Uganda	0.0455	47.7	Spain	0.0217	28.3
Liberia	0.0652	26.1	United Kingdom	0.0217	60.9
Nigeria	0.0652	13.0	Russian Federation	0.0652	30.4
Sudan	0.0652	71.7	Serbia and Montenegro	0.0652	6.5
Chad	0.1087	67.4	Moldova	0.0667	6.7
Ethiopia	0.1304	67.4	Croatia	0.0714	28.6
			Macedonia, FYR	0.0769	7.7
<b>Asia</b>			<b>N. America</b>		
Lao PDR	0.0217	4.3	Guatemala	0.0217	67.4
Lebanon	0.0217	34.8	Mexico	0.0217	2.2
Sri Lanka	0.0217	45.7	Nicaragua	0.0217	19.6
Thailand	0.0217	6.5			

## C Variable Definitions and Sources

### *Conflict Variables*

Depending on the time frequency of the sample, a conflict variable is coded as follows:

- **Cross-sectional regressions:** Whenever sufficient information is available for the conflict events underlying an outcome variable, the dependent variable is equal to the natural logarithm of (1+ the number of new conflict onsets in the country over the sample period). This happens to be the case for all measures used in cross-sectional regressions except for MAR intra-group conflicts.
- **5-yearly panel regressions:** The outcome variable is coded 1 for any 5-year period with at least one ongoing conflict and 0 otherwise (conflict incidence during the period).
- **Annual panel regressions:** (a) The outcome variable is coded 1 for any year with an ongoing conflict and 0 otherwise (annual incidence regressions) and (b) The outcome variable is coded 1 for the starting year when a new conflict or a new episode of a conflict, and zero otherwise (annual onset regressions).

Below is the list of conflict codings that underlie the outcome variables employed in the empirical analyses.

#### 1. UCDP/PRIO Civil Conflicts

**PRIO25:** It is the main civil conflict coding employed in the analysis. Civil conflict is defined as a contested incompatibility that concerns government and/or territory where the use of armed force between the government of a state and any armed opposition organization, i.e, any non-governmental group of people having announced a name for their group and using armed force to influence the outcome of the stated incompatibility. Civil conflict events include those with and without intervention from other states. Massacres and genocides are not included because the victims are neither organized nor armed; communal riots and pogroms are excluded because the government is not directly involved. *PRIO25* identifies an active civil conflict for a dyad-year if the conflict satisfies the above mentioned criteria *and* the armed clashes resulted in at least 25 battle-related deaths in that year within the dyad in question. In cross-country regressions, the outcome measure is the number of distinct *PRIO25* conflicts, each of which represents a different incompatibility and is assigned a unique conflict ID in the source dataset. *Source: The UCDP/PRIO Armed Conflict Dataset, version 4–2012.*

**PRIO1000 civil war:** A civil war in a dyad-year is defined as a large scale civil conflict between the parties which resulted in at least 1000 battle-related deaths during that year. See *PRIO25* for the definition of a civil conflict. *Source: The UCDP/PRIO Armed Conflict Dataset, version 4–2012.*

**PRIO2 onset:** It is coded 1 whenever a *new episode* of a *PRIO25* civil conflict breaks out after at least two years of civil peace, i.e., without any active civil conflict. It is coded 0 otherwise. The variable is directly borrowed from the dataset *Onset of Intrastate Armed Conflict, 1946–2011* which contains a country-year version of the *UCDP/PRIO Armed Conflict Dataset v.4–2012* structured for quantitative analysis.

**PRIO-NC onset:** It is coded 1 whenever an entirely new *PRIO25* civil conflict (with a unique conflict ID) erupts in a given year, and coded 0 otherwise. This variable differs from *PRIO2 onset* in that the latter codes a new episode of an old civil conflict, which was inactive (according to *PRIO25* definition) for a while, as a new onset event. The variable is directly borrowed from the dataset *Onset of Intrastate Armed Conflict, 1946–2011* which contains a country-year version of the *UCDP/PRIO Armed Conflict Dataset v.4–2012* structured for quantitative analysis.

2. **WCM2009 Ethnic Civil Conflicts:** This measure is borrowed from Wimmer et al. (2009). The authors classify a civil conflict as an *ethnic civil conflict* if armed organizations (1) explicitly pursue ethno-nationalist aims (e.g., ethno-national self-determination, ethno-regional autonomy, language and other cultural rights) or be motivated by ethnicity-based concerns like ethnic balance of power in the government or ethnic and racial discrimination; and (2) recruit fighters and forge alliances on the basis of ethnic affiliations. This dataset is based on the civil conflict definition adopted in a previous version of UCDP/PRIO Armed Conflicts Dataset which has a time coverage only until 2005.
3. **WM2006-09 Ethnic Civil War:** It is based on the ethnic civil war classification of Wimmer and Min (2006). An ethnic civil war is a civil war –resulting a total of at least 1,000 battle-deaths during that year– where the lines of conflict are defined in ethnic terms and/or there is significant recruitment on the basis of ethnic networks. The list of wars used by Wimmer and Min (2006) is cross-checked and updated with a more recent dataset introduced by Wimmer and Min (2009). The time coverage of the original data is only until 2001. The basic criteria for what constitutes a civil war is based on Correlates of War (COW) definition, but authors made some modifications and additions to the original COW war list. For details, see Wimmer and Min (2009).
4. **UCDP Non-state Conflicts:** A non-state conflict is defined by the Uppsala Conflict Data Program (UCDP) as the use of armed force between two organized armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year. *Source: UCDP Non-State Conflict dataset , version 2.3-2010.*
5. **MAR Intra-group Conflicts:** A dummy for whether in a given country there was a communal conflict anytime during 1990 to 2000 between the factions within any of the



ethnopolitical minority groups that were identified to be at risk by the Minorities at Risk (MAR) Project. *Source: Minorities at Risk dataset, version 2/2009.*

### ***Diversity and Polarization Measures***

1. **Genetic diversity (Ancestry Adjusted):** This variable is due to Ashraf and Galor (2013a). It measures the expected heterozygosity (genetic diversity) of a country’s population, predicted by migratory distances from Addis Ababa (Ethiopia) to the year 1500 CE locations of the ancestral populations of the country’s component ethnic groups in 2000 CE, as well as by pairwise migratory distances between these ancestral populations. The source countries of the year 1500 CE ancestral populations are identified from the *World Migration Matrix, 1500–2000*, discussed in Putterman and Weil (2010), and the modern capital cities of these countries are used to compute the aforementioned migratory distances. The measure of genetic diversity is then calculated by applying (i) the regression coefficients obtained from regressing expected heterozygosity on migratory distance from Addis Ababa at the ethnic group level, using the worldwide sample of 53 ethnic groups from the Human Genome Diversity Cell Line Panel, (ii) the regression coefficients obtained from regressing pairwise  $F_{st}$  genetic distances on pairwise migratory distances between these ethnic groups, and (iii) the ancestry weights representing the fractions of the year 2000 CE population (of the country for which the measure is being computed) that can trace their ancestral origins to different source countries in the year 1500 CE. The expected heterozygosities, geographical coordinates, and pairwise  $F_{st}$  genetic distances of the 53 ethnic groups are from Ramachandran et al. (2005). The ancestry weights are from the *World Migration Matrix, 1500–2000*
2. **Ethnic, religious and linguistic fractionalization (Alesina et al., 2003):** Fractionalization indices, constructed by Alesina et al. (2003), that capture the probability that two individuals, selected at random from a country’s population, will belong to different ethnic, religious and linguistic groups respectively.
3. **Ethnolinguistic fractionalization (Desmet et al., 2012):** It measures the probability that two individuals, selected at random from a country’s population, will belong to different ethnolinguistic groups as identified by the linguistic classification in the 15th edition of *Ethnologue*. Desmet et al. (2012) compute different fractionalization indices for linguistic group profiles at 15 different levels of the language tree. We use ELF(15), the measure that corresponds to the most contemporary –lowest– profile of linguistic divisions (level 15).
4. **Ethnolinguistic polarization (Desmet et al., 2012):** It is an index of polarization among ethnolinguistic groups in a country as identified by the linguistic classification in the 15th edition of *Ethnologue*. Desmet et al. (2012) compute different polarization measures for linguistic group profiles corresponding to 15 different levels of the language tree. We use POL(15), the measure that reflects the most contemporary profile of linguistic groups (level

15). The exact formulation of the polarization index follows (Montalvo and Reynal-Querol, 2005). Accordingly,

$$POL(15) = 4 \sum_{i=1}^N \pi_i^2 (1 - \pi_i) \quad (C.1)$$

where  $N$  is the total number of contemporary linguistic groups in a country and  $\pi_i$  is the population share of group  $i$ .

5. **Fearon-P, Fearon-F and Fearon-G/N Indices (Esteban et al., 2012):** These three measures are based on the classification of ethnic and ethnoreligious groups provided by Fearon (2003).  $P$  is a measure of polarization computed as

$$P = \sum_{i=1}^m \sum_{j=1}^m n_i^2 n_j d_{ij}$$

where  $m$  is the number of groups in the country,  $n_i$  and  $n_j$  are population shares of group  $i$  and group  $j$ , and  $d_{ij}$  is the intergroup distance between the two groups.  $G$  is the so called Greenberg-Gini index defined as

$$G = \sum_{i=1}^m \sum_{j=1}^m n_i n_j d_{ij}$$

Due to the underlying theory of conflict in Esteban et al. (2012), the authors use  $G/N$  which is Greenberg-Gini index normalized by population size.

$F$  is ethnic fractionalization measure which assumes that the intergroup distances in  $G$  index are all equal to one. Formally:

$$F = \sum_{i=1}^m n_i (1 - n_i)$$

6. **Number of ethnic groups (Fearon, 2003) :** The variable is based on the classification of ethnic groups in Fearon (2003). For each country the variable records the natural logarithm of (1+ number of ethnic groups).
7. **Number and population share of active MAR groups:** The first variable shows for each country the number of politically-active communal groups included in the Minorities at Risk Dataset with a current (as of the 2004-2006 update of the dataset) population of at least 500,000. The second variable is the population of all MAR groups combined as a fraction of the total country population. Population figures for individual groups are based on the 1998 MAR estimates of 1990 group populations. For more information on the inclusion criteria

of ethno-political groups in the MAR dataset, see the Minorities at Risk Codebook (2007).  
*Source: Minorities at Risk dataset, version 2/2009.*

8. **Ethnic dominance:** A dummy indicating whether the largest ethnic group in a country constitutes 45 to 90% of the population. The variable is borrowed from Hegre and Sambanis (2006) whose primary source is Collier and Hoeffler (2004).

### ***Geographical Variables***

1. **Migratory distance:** The great circle distance from Addis Ababa (Ethiopia) to the country's modern capital city along a land-restricted path forced through one or more of five intercontinental waypoints, including Cairo (Egypt), Istanbul (Turkey), Phnom Penh (Cambodia), Anadyr (Russia) and Prince Rupert (Canada). Distances are calculated using the Haversine formula and are measured in units of 1,000 km. The geographical coordinates of the intercontinental waypoints are from Ramachandran et al. (2005), while those of the modern capital cities are from the CIA's *World Factbook*.
2. **Absolute latitude:** The absolute value of the latitude of a country's approximate geodesic centroid, as reported by the CIA's *World Factbook*.
3. **Land area:** Total land area of a country in millions of square kilometer as reported by the World Bank's *World Development Indicators*.
4. **Ruggedness:** The degree of terrain roughness of a country, calculated using geospatial surface undulation data reported by the G-ECON project (Nordhaus, 2006) at a 1-degree resolution, which is based on more spatially disaggregated elevation data at a 10-minute resolution from New et al. (2002). The measure is the average degree of terrain roughness across the grid cells within a country. The interested reader is referred to the G-ECON project web site for additional details.
5. **Mean and the range of elevation:** The measures are calculated using geospatial elevation data reported by the G-ECON project (Nordhaus 2006) at a 1-degree resolution, which, in turn, is based on similar but more spatially disaggregated data at a 10-minute resolution from New et al. (2002). The mean of elevation is thus the average elevation across the grid cells within a country. The range of elevation is the difference between the maximum and minimum elevation across the grid cells within the country. The interested reader is referred to the G-ECON project web site for additional details.
6. **Mean and range of land suitability:** Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH.

This index was initially reported at a half-degree resolution by Ramankutty et al. (2002). Formally, Ramankutty et al. (2002) calculate the land suitability index ( $S$ ) as the product of climate suitability ( $S_{clim}$ ) and soil suitability ( $S_{soil}$ ), i.e.,  $S = S_{clim} \times S_{soil}$ . The climate suitability component is estimated to be a function of growing degree days ( $GDD$ ) and a moisture index ( $\alpha$ ) gauging water availability to plants, calculated as the ratio of actual to potential evapotranspiration, i.e.,  $S_{clim} = f_1(GDD)f_2(\alpha)$ . The soil suitability component, on the other hand, is estimated to be a function of soil carbon density ( $C_{soil}$ ) and soil pH ( $pH_{soil}$ ), i.e.  $S_{soil} = g_1(C_{soil})g_2(pH_{soil})$ . The functions,  $f_1(GDD)$ ,  $f_2(\alpha)$ ,  $g_1(C_{soil})$ , and  $g_2(pH_{soil})$  are chosen by Ramankutty et al. (2002) by empirically fitting functions to the observed relationships between cropland areas,  $GDD$ ,  $\alpha$ ,  $C_{soil}$ , and  $pH_{soil}$ . For more details on the specific functional forms chosen, the interested reader is referred to Ramankutty et al. (2002). Since Ramankutty et al. (2002) report the land suitability index at a half-degree resolution, Michalopoulos (2012) aggregates the index to the country level by averaging land suitability across grid cells within a country. Likewise, the range of land suitability is the difference between the maximum and minimum values of the index across grid cells within each country. This study employs the country-level aggregate measures reported by Michalopoulos (2012) as the control for the mean and range of land suitability in the baseline regression specifications.

7. **Mean distance to nearest waterway:** The distance, in thousands of km, from a GIS grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells of a country. This variable was originally constructed by Gallup et al. (1999) and is part of Harvard University's CID Research Datasets on General Measures of Geography.
8. **Log mountainous terrain:** Log(% mountainous terrain) in a country. The data source is Fearon and Laitin (2003) who use the codings of geographer A.J. Gerard.
9. **Non-contiguous state:** Refers to countries with territory holding at least 10,000 people and separated from the land area containing the capital city either by land or by 100 kilometers of water. *Source: Fearon and Laitin (2003).*
10. **Disease richness:** The total number of different types of infectious diseases in a country, as reported by Fincher and Thornhill (2008), based on the Global Infectious Disease and Epidemiology Network (GIDEON) online database. See Fincher and Thornhill (2008) for additional details.

### ***Institutional and Political Factors***

1. **Executive constraints:** The 1960–2000 mean of an index, reported annually as a 7-point categorical variable (from 1 to 7) by the Polity IV dataset, quantifying the extent of institutionalized constraints on the decision-making power of chief executives (Marshall and Jaggers, 2009).

2. **Colonial history variables:** They indicate for a given period whether a country had been under the rule of some colonial power in that period. In regressions we distinguish among colonial rules under the U.K., France and some other colonizer. The codings are done by the authors based on various sources including but not limited to the CIA World Factbook, Britannica, Library of Congress and the state.gov website.
3. **Legal origin dummies:** Two dummy variables indicating that legal origin of the Company Law or Commercial Code of a country is the English Common Law or the French Commercial Code. Legal origins of countries are reported by La Porta et al. (1999).
4. **Democratic, anocratic and autocratic regimes:** Country-years with a combined Polity IV score between 5.5 and 10 are classified as democracies. Those with a score ranging from -5.5 to 5.5 are anocracies. Finally, those country-years with a score ranging between -10 and -5.5 are defined as autocratic regimes (Marshall and Jaggers, 2009).
5. **Political instability:** Dummy variable indicating whether there was a change in Polity IV score in the past three years. The variable is borrowed from Hegre and Sambanis (2006) whose primary source is Fearon and Laitin (2003).

### *Other Control Variables*

1. **GDP per capita:** Logged GDP per capita, in 1990 International Geary-Khamis dollars. The relevant variables for the five-year and cross-sectional samples are the yearly average value over a given five-year episode and the average over the period 1960-1999 respectively. *Source: Maddison, Angus. 2010. Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD.*
2. **Population:** Logged population size. The relevant variables for the five-year and cross-sectional samples are the yearly average value over a given five-year episode and the average over the period 1960-1999 respectively. *Source: Maddison, Angus. 2010. Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD.*
3. **Value of oil production per capita:** Per capita value of oil production (in 2000 constant dollars) in each country. Data on oil wealth comes from the 2013 release of the *Global dataset of oil and gas production and exports (version 2)* compiled by Ross (2013).
4. **Interpersonal trust:** The fraction of total respondents within a given country, from five different waves of the World Values Survey conducted during the time period 1981-2008, that responded with “Most people can be trusted” (as opposed to “Can’t be too careful”) when answering the survey question “Generally speaking, would you say that most people can be trusted or that you can’t be too careful in dealing with people?”

5. **Neolithic Transition Timing (Ancestry Adjusted):** Neolithic transition timing is the number of thousand years elapsed, until the year 2000, since the majority of the population residing within a country’s modern national borders began practicing sedentary agriculture as the primary mode of subsistence. This measure, reported by Putterman (2008) , is compiled using a wide variety of both regional and country-specific archaeological studies as well as more general encyclopedic works on the transition from hunting and gathering to agriculture during the Neolithic. The reader is referred to the author’s web site for a detailed description of the primary and secondary data sources employed by the author in the construction of this variable. The measure used in this study is an ancestry adjusted version. It is the cross-country weighted average of Neolithic transition timing, where the weight associated with a given country in the calculation represents the fraction of the year 2000 CE population (of the country for which the measure is being computed) that can trace its ancestral origins to the given country in the year 1500 CE. The ancestry weights are obtained from the *World Migration Matrix, 1500–2000* of Putterman and Weil (2010).