Does Ethnic Fractionalization Matter for Development?

Jeremy Majerovitz*

Undergraduate Honors Thesis
Stanford University

May 7, 2015

Abstract

An existing literature finds that ethnic fractionalization (a measure of ethnic diversity) is negatively correlated with various development outcomes, including growth and public goods provision (Alesina and La Ferrara, 2005). This paper attempts to deal with issues of endogeneity in work related to ethnic fractionalization. I develop an instrument for fractionalization based on the arbitrary construction of African borders, and also use a control function approach to deal with omitted variables bias. Both methods generate evidence that the endogeneity of ethnic fractionalization causes substantial bias in regressions of growth on fractionalization, and the true effect is substantially smaller, although very difficult to estimate precisely. I develop a game theoretic model to explain how to reconcile micro-level evidence of harmful ethnic favoritism with macro-level effects of ambiguous sign.

1 Introduction

A large literature in political economy has suggested that ethnic fractionalization[1] has negative effects on various development outcomes, including public goods provision, logged GDP per capita, and economic growth (Alesina and La Ferrara, 2005). Easterly and Levine (1997) argue that differences in ethnic fractionalization "explain between one-fourth and two-fifths

*Email address: jmajic@stanford.edu. I thank Alberto Alesina, Klaus Desmet, James Fearon, Matthew Jackson, Saumitra Jha, Francesco Trebbi, and Romain Wacziarg for helpful comments. I thank Marcella Alsan for sharing her country-level data on her TseTse Suitability Index. I thank Marcelo Clerici-Arias for organizing the honors program and for his helpful comments on this paper. I am particularly indebted to Pascaline Dupas, my thesis advisor and Economics major advisor, whose comments have greatly improved this paper and whose advice and example have greatly improved my skill as an economist.

[1]Ethnic fractionalization is defined as the probability that two individuals, selected at random from a country's population, will be of different ethnicities. Fractionalization is by far the most common measure of ethnic diversity used in the existing literature, which is why I focus on it in this paper.

of the East Asia-Africa growth differential and may fully account for some extreme country cases." Subsequent work using updated data has supported this early finding (Alesina et al., 2003; Posner, 2004a).

This literature typically treats ethnic fractionalization as exogenous, and when it does not it typically employs a simple control function approach to deal with a small set of possible omitted variables. Alesina et al. (2003) explain this decision, noting: "The bottom line is that while we recognize that ethnic fractionalization could to some extent be endogenous, and that the previous literature has probably underplayed this point, we do not believe this is a very serious problem at the horizon of 20-30 years which characterizes our cross-country work." Since much of the literature focuses on growth regressions with controls for initial income, the argument goes, we have dealt with the effect of development on fractionalization and thus have a well-identified regression.

This approach is unsatisfactory. As has been noted in prior literature (Alesina and La Ferrara, 2005; Michalopoulos, 2012), ethnic fractionalization is endogenous to a variety of factors: I will argue that most of these omitted variables are not adequately dealt with by the minimalist control function used in the prior literature. This paper takes up the call put forward in Alesina and La Ferrara (2005) to deal with endogeneity in order to identify the causal effect of ethnic fractionalization on development outcomes. I find that dealing with the major threats to identification results in an estimate of fractionalization's effect on growth that is substantially smaller than in the "naive" specification, and this estimate is often statistically insignificant. In order to establish the robustness of this result, I use a variety of techniques to deal with identification threats, in order to show that the effect of ethnic fractionalization can be reduced or eliminated in various ways. I present estimates from an instrumental variables strategy that exploits the arbitrary nature of African borders; I also use control function approaches controlling for slave exports or controlling for geographic variables. I discuss what identification threats these methods are intended to deal with in order to further explain the story behind the regression results. In general, I find that naive estimates of the effect of ethnic fractionalization substantially overstate the true causal effect, and secondly I find that, since most of the variation in ethnic fractionalization is endogenous, it is very difficult to obtain estimates that are both well identified and precisely estimated.

These results may seem deeply counterintuitive, especially given the literature at the micro level that finds harmful effects of ethnic competition and ethnic favoritism on productivity and policy (Burgess et al., 2013; Marx et al., 2014; Hjort, 2014), as well as myriad examples of ethnic politics appearing to lead to poor policy making. I thus discuss how to reconcile the non-effect at the cross-country level with the apparent relevance of ethnicity within-country. I argue that the answer lies in the instrumentalist view of ethnicity that has been supported by much of the literature. In essence, political coalitions form around ethnic identities because people can be easily mobilized along these lines, and thus within-country we observe inter-ethnic conflict. However, what matters is the number and size of the political coalitions. I present a model in which an increase in ethnic fractionalization has ambiguous effects on the quality of government policy. This suggests that the observation of harmful ethnic favoritism at the micro level does not imply that ethnic diversity is harmful at the macro level, and that extrapolating from micro-level evidence requires increased attention to general equilibrium effects.

This paper proceeds as follows. Section 2 describes the various threats to identification. Section 3 discusses the instrumental variables approach used in the paper. Section 4 presents the empirical results. Section 5 develops a simple formal model of how ethnic fractionalization could have a positive, negative, or null effect on outcomes even though inter-ethnic competition occurs. Section 6 concludes.

2 Identification Threats

Although the literature has typically treated ethnic fractionalization as exogenous or close to exogenous, ethnic diversity is actually highly endogenous. Regressions of development on ethnic fractionalization suffer from reverse causality and omitted variables bias. The reverse causality comes from the fact that states ethnically homogenize their populations as part of the process of state-building and development. Weber (1976) details how a central government based in Paris convinced French peasants, who initially identified regionally rather than nationally, to identify as ethnically French through education and language policy. Balcells (2013) corroborates this idea that a strong state can bring about homogeneity through schooling policy; she finds that there is a strong Catalan ethnic identity in Spanish Catalonia but not in French Catalonia, and argues that this is explained by the differences in the strength of the Spanish and French states at the time that the population was becoming literate. The state-building process is also a nation-building process, and as a result ethnic diversity slowly decreases as a result of development. [2]

In the context of a typical growth regression, which is concerned with rates of change rather than with levels, our concern about reverse causality instead becomes a concern about omitted variables bias. Since more developed countries were able to reduce their level of ethnic diversity, initial income is an omitted variable. However, it is not clear that controlling for initial income is sufficient. Ethnic homogenization through language and education policy is a gradual process requiring a strong central government to implement; we would thus expect fractionalization to be caused by institutions or state capacity even beyond being caused by income. I will remain agnostic about what is the best measure of the historical centralization, state capacity, and other factors that lead ethnic homogenization; however, recent work has demonstrated that controlling for political centralization, proxied for by the population of the largest city in 1900, halves the coefficient on fracitionalization in a growth regression (Weese, 2011).

Other omitted variables are also a major problem. Michalopoulos (2012) finds a suite of geographic variables that explain a sizable portion of ethnic fractionalization, such as distance from the equator, dispersion of elevation and agricultural suitability, and distance

[2] Another possible channel of reverse causality runs through border drawing. Poorer countries were less powerful on the international stage, and are thus more likely to have their borders drawn by a more powerful country (perhaps a colonizer) that may not have drawn borders so as to create ethnically homogeneous countries.

from the coast. These geographic variables almost certainly have effects on development and growth running through channels other than just ethnic fractionalization, for example, Dell et al. (2012) find that temperature shocks have persistent effects on development and growth. Whether we take the position that geography has a direct effect on development or that it affects development through its effect on history and/or institutions is not relevant here. Nunn (2008) found sizable negative effects of the slave trade on contemporary economic development in Africa, and also found that the slave trade caused greater ethnic fractionalization. Although Nunn argues that this increased fractionalization is a channel through which the effect of the slave trade is carried, assuming that not all of the effect of the slave trade went through increased ethnic fractionalization means that slave exports is an important omitted variable to control for.

Beyond the sources of bias listed above, there may be further forms of omitted variables bias or channels of reverse causality not yet known by researchers. The problem is that ethnicity is an endogenous variable: it is a result of human decisions about how to identify themselves. In fact, even though certain biological characteristics are immutable, such as skin color, the decision of what differences are salient and which are not is a decision (we don't, for example, consider blondes and brunettes ethnically distinct). The ethnic categories of the United States is illustrative of this issue. In the United States, ethnic groups are defined largely based on very broad categories: black, white, Asian, Latino, etc. Yet in other countries ethnic groups are often defined by much subtler distinctions. Kenya is one of the most ethnically fractionalized countries in the world, but applying American ethnic categories would make the country almost perfectly homogeneous; the various ethnic groups of Kenya would all count as black. Ethnic identity and ethnic diversity are the result of human decisions, and, since we do not yet fully understand all of the factors affecting those decisions, it may not be correct to assume that a simple OLS regression will be sufficient to achieve identification.

3 Instrument

Given the endogeneity of ethnic fractionalization, we would like to find an instrument that satisfies our exclusion restriction of not directly affecting development and related outcomes. A natural place to look is Africa. A number of papers have used the arbitrary, colonizer-drawn borders in Africa as a source of identification (Posner, 2004b; Miguel, 2004; Michalopoulos and Papaioannou, 2013a). Because African borders were drawn arbitrarily by Europeans at the Conference of Berlin who did not know where ethnic groups were located, and because these initial borders have largely survived into the present day, we can treat African borders as exogenous. We can then determine a component of ethnic fractionalization resulting from the arbitrary nature of African borders, and use this component as an instrument for ethnic fractionalization. This will give us a valid causal estimate of the effect of ethnic fractionalization, although it will be a local average treatment effect for countries in Africa whose ethnic fractionalization is affected by the arbitrary nature of their borders. Although the instrument will at best give us a LATE, Africa is a highly relevant setting in which to study the effects of ethnic fractionalization, as Africa is the continent with the highest level and standard deviation of ethnic fractionalization (Fearon, 2003), and because Africa's poor growth performance has been attributed in part to its high level of ethnic fractionalization, due in part to its arbitrary borders (Easterly and Levine, 1997).

3.1 Theory

Extracting a component of ethnic fractionalization caused by the arbitrary nature of African borders is easier said than done. To get the best estimates possible, one wants to exploit as much of the variation as possible while still ensuring that the exclusion restriction is satisfied. To do this, I develop a procedure that predicts ethnic fractionalization based on the Murdock Map of African ethnic groups and relevant ethnicity-level and country-level characteristics, subtracts this from the true Murdock Map based measure of ethnic fractionalization, and treats the deviations as the instrument. I use data from Michalopoulos and Papaioannou (2013b), which provides ethnicity-country observations with data on country area, ethnic group area, ethnic-country group area, and some geographical variables.

To further motivate the construction of the instrument, consider the following pair of structural equations, corresponding to the first and second stage of the instrumental variables regression.

$$Y = \alpha + \beta ELF + u \tag{1}$$
$$ELF = \gamma + \delta Z + \xi \tag{2}$$

Assume that u and ξ are correlated, but Z is uncorrelated with u or ξ ($\alpha, \beta, \gamma,$ and δ are constant coefficients). Since ξ is part of ELF and is correlated with u, ELF is correlated with u and thus OLS will give biased and inconsistent estimates. However, we can use Z as an instrument and get consistent estimates using IV regression. What we want, then, is to actually obtain some such variable Z.

To construct such a Z, we consider the data from the Murdock Map, taken from Michalopoulos and Papaioannou (2013b). Here, we have observations at the ethnicity-country level in Africa, with some ethnic groups entirely contained by one country and others split between two or more countries. We will use data on the size of these ethnic groups, the ethnic-country subdivisions, and the country, as well as relevant geographic data. Consider the variable *SHARE*, representing the area of the ethnic-country group divided by the total area of the ethnic group. If an ethnicity is unsplit, then it will have a value of *SHARE* equal to one; if it is split, its ethnic-country groups will have values of *SHARE* between zero and one, but summing to one. Since we are exploiting the arbitrary nature of African borders, we will argue that, controlling for a few key variables, the variable *SHARE* (or more precisely, $SHARE^2$) is uncorrelated with u. More specifically, let us assume that

$$SHARE^2 = \theta X + v \tag{3}$$

where u and v are uncorrelated. Assume for the moment that we know what is in X, which is a matrix of controls, as well as a a constant. Then, introducing the subscript j to index over countries and i to index over ethnic groups, we have:

$$\widehat{SHARE}_{ij}^{2} = \hat{\theta} X_{ij} \tag{4}$$

$$AREA\ ELF_j = 1 - \sum_i (SHARE_{ij}^{2} \times \frac{ETHNIC\ AREA_i^{2}}{COUNTRY\ AREA_j^{2}}) \tag{5}$$

$$\widehat{AREA\ ELF}_j = 1 - \sum_i (\widehat{SHARE}_{ij}^{2} \times \frac{ETHNIC\ AREA_i^{2}}{COUNTRY\ AREA_j^{2}}) \tag{6}$$

$$Z_j = \widehat{AREA\ ELF}_j - AREA\ ELF_j \tag{7}$$

where $\hat{\theta}$ is obtained by estimating (3) with OLS. Letting X_j denote the set of X_{ij} associated with country j, we get:

$$Z_j = \sum_i (SHARE_{ij}^{2} - \widehat{SHARE}_{ij}^{2}) \times \frac{ETHNIC\ AREA_i^{2}}{COUNTRY\ AREA_j^{2}} \tag{8}$$

$$E\left(Z_j \mid X_j\right) = \sum_i E\left(((\theta - \hat{\theta})X_{ij} + v_{ij}) \times \frac{ETHNIC\ AREA_i^{2}}{COUNTRY\ AREA_j^{2}} \mid X_j\right) \tag{9}$$

$$= \sum_i Cov\left((\theta - \hat{\theta})X_{ij} + v_{ij}, \frac{ETHNIC\ AREA_i^{2}}{COUNTRY\ AREA_j^{2}} \mid X_j\right) \tag{10}$$

which is simply the covariance of the estimated residuals (as opposed to the true v_{ij}) from (3) and $\frac{ETHNIC\ AREA_i^{2}}{COUNTRY\ AREA_j^{2}}$, conditional on X_j. If we include $\frac{ETHNIC\ AREA_i^{2}}{COUNTRY\ AREA_j^{2}}$ in X_{ij} then we get orthogonality between these two terms, and thus have:

$$E\left(Z_j \mid X_j\right) = 0 \tag{11}$$

$$= E\left(Z_j\right) \tag{12}$$

We now can establish our main result, that Z_j is uncorrelated with u_j. Recalling our identifying assumption, that v_{ij} is uncorrelated with u_j, we get:

$$
\begin{aligned}
Cov\left(Z_j, u_j\right) &= E\left(Z_j u_j\right) - E\left(Z_j\right)E\left(u_j\right) && (13)\\
&= E\left(Z_j u_j\right) && (14)\\
&= E\Big[E\left(Z_j u_j \mid X_j\right)\Big] && (15)\\
&= E\left[\sum_i E\left(u_j((\theta-\hat{\theta})X_{ij}+v_{ij}) \times \frac{\textit{ETHNIC AREA}_i^{\,2}}{\textit{COUNTRY AREA}_j^{\,2}} \mid X_j\right)\right] && (16)\\
&= E\left[\sum_i Cov\left((\theta-\hat{\theta})X_{ij}+v_{ij}, u_j\frac{\textit{ETHNIC AREA}_i^{\,2}}{\textit{COUNTRY AREA}_j^{\,2}} \mid X_j\right)\right] && (17)\\
&= E\left[\sum_i \frac{\textit{ETHNIC AREA}_i^{\,2}}{\textit{COUNTRY AREA}_j^{\,2}} \times Cov\left((\theta-\hat{\theta})X_{ij}+v_{ij}, u_j \mid X_j\right)\right] && (18)\\
&= E\left[\sum_i \frac{\textit{ETHNIC AREA}_i^{\,2}}{\textit{COUNTRY AREA}_j^{\,2}} \times Cov\left((\theta-\hat{\theta})X_{ij}, u_j \mid X_j\right)\right] && (19)\\
&= E\left[\sum_i \frac{\textit{ETHNIC AREA}_i^{\,2}}{\textit{COUNTRY AREA}_j^{\,2}} \times Cov\left((\theta-\hat{\theta}), u_j \mid X_j\right)X_{ij}\right] && (20)\\
&= E\left[\sum_i \frac{\textit{ETHNIC AREA}_i^{\,2}}{\textit{COUNTRY AREA}_j^{\,2}} \times 0 \times X_{ij}\right] && (21)\\
&= 0 && (22)
\end{aligned}
$$

where (21) is a result of the fact that, conditioning on X_{ij}, an omitted variable such as u_j can only bias the estimation of $\hat{\theta}$ if it is correlated with the error term v_{ij}, and u_j is uncorrelated with v_{ij} by our identifying assumption. We have thus established that if our identifying assumption is true, then our proposed instrument Z_j meets the exclusion restriction.

3.2 Functional Form for Deriving the Instrument

With our key result established, we can return to the topic we earlier ignored: What goes in X_{ij}? We have, for technical reasons, already chosen to put in $\frac{\textit{ETHNIC AREA}_i^{\,2}}{\textit{COUNTRY AREA}_j^{\,2}}$. This variable, however, gives some intuition into what else should be included in X_{ij}. With borders falling in arbitrary places, what matters for predicting $\textit{SHARE}^{\,2}$ is the overall area of the ethnic group relative to the area of the country. Beyond this, it is difficult to say more about the appropriate functional form to relate $\frac{\textit{ETHNIC AREA}_i}{\textit{COUNTRY AREA}_j}$ to $\widehat{\textit{SHARE}}_{ij}^{\,2}$. Deriving this functional form theoretically is somewhat unwieldy and would require making assumptions about the

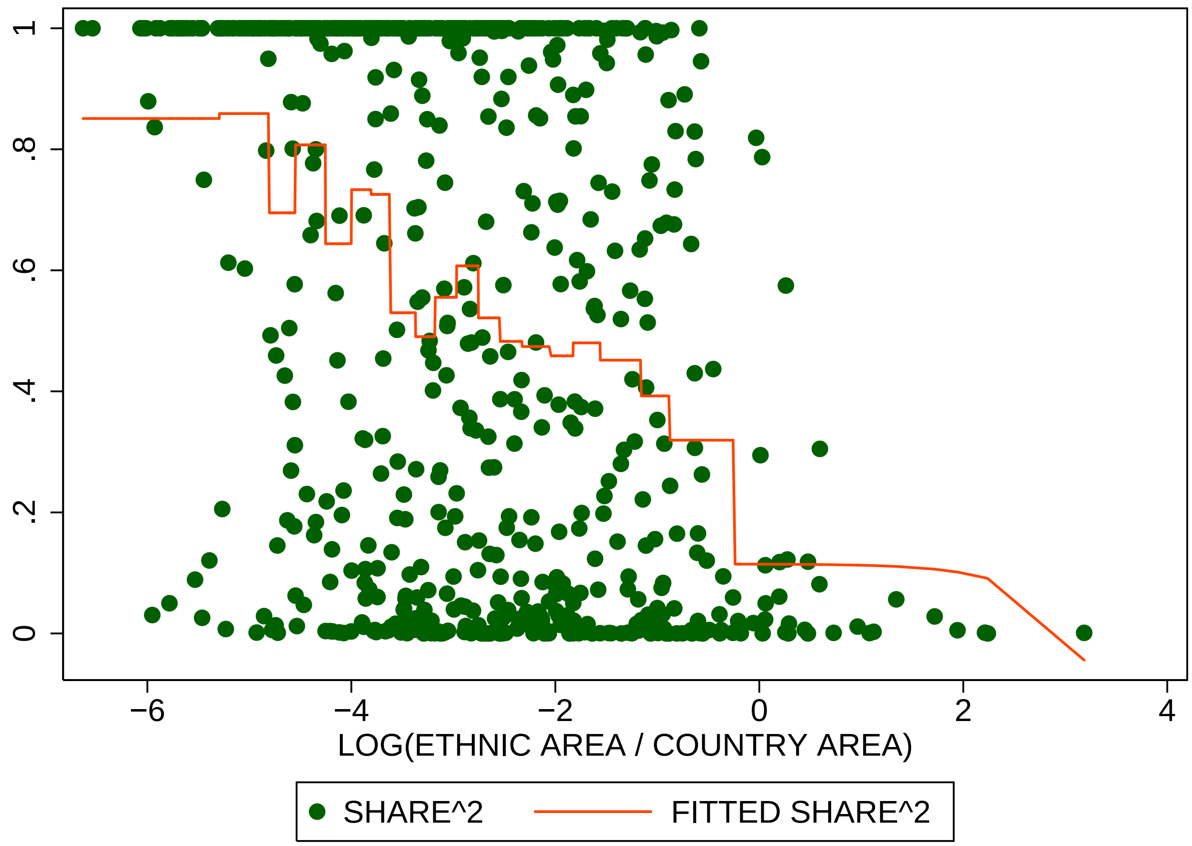


Figure 1: Plot of "Zeroth Stage" Relationship Between $\frac{ETHNIC\ AREA_i}{COUNTRY\ AREA_j}$ and $SHARE_{ij}^2$.

shape of countries and ethnic homelands. Graphical analysis does not reveal an obvious and compelling functional form.

Instead, I opt for a non-parametric specification. I divide the sample into twenty quantiles based on each observation's value of $\frac{ETHNIC\ AREA_i}{COUNTRY\ AREA_j}$. I then regress $SHARE_{ij}^2$ on $\frac{ETHNIC\ AREA_i^2}{COUNTRY\ AREA_j^2}$ and dummies for these quantiles. The resulting fit is shown in Figure ; I use a logarithmic scale for plotting the x-axis because the plot has long tails. This specification is quite flexible, and thus should give a fairly close approximation to the true functional form. Given this flexibility, an important note is that once the proper necessary variables have been put into X_{ij}, adding additional unnecessary regressors to X_{ij} will not lead to a violation of the exclusion restriction. This is because once X_{ij} has the necessary regressors included all of the remaining variation in $SHARE_{ij}^2$ comes from the arbitrary element of African borders, and thus should not be predicted by new regressors. Thus, extra regressors should not be statistically significant, nor should they change the estimated coefficient on fractionalization.

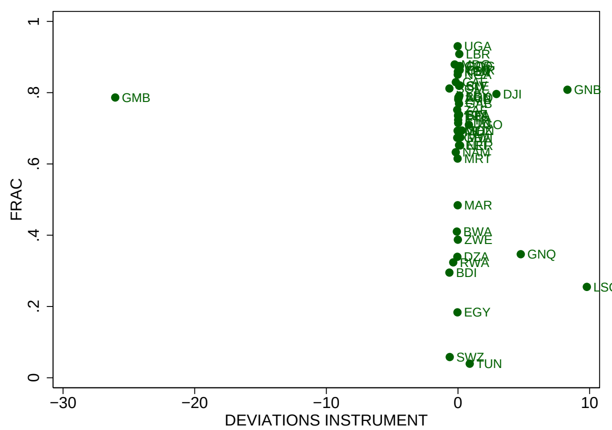
3.3 First Stage

With our instrument in hand, we can look at the first stage regression. Figure 2 examines the first stage graphically, using data on ethnic fractionalization from Alesina et al. (2003). Of immediate note is that the instrument has some major outliers. Looking into these outliers, one finds that they are very small countries, such as Guinea-Bissau, Lesotho, The Gambia, and Togo; these countries are major outliers because they have ethnic groups that are very large relative to the size of the country (with only a small share of that ethnic group falling inside the country), and the fit of the predictive function used to generate the instrument does poorly for these extreme ethnic-country observations. As a result, these countries take on extreme and often absurd values of the instrument (since the instrument is defined as actual minus predicted area-based fractionalization, it should not ever take on an absolute value greater than one). I thus drop all countries with a value of the instrument greater in absolute value to 0.8; since the instrument represents the amount of fractionalization contributed by the arbitrariness of African borders, such values of the instrument are unreasonable. The result is a better fit in the first stage, but with the remaining outlier of Somalia (interestingly, while Somalia is coded as having high fractionalization in the Alesina et al. (2003) data, it is coded as having very low fractionalization in the Fearon 2003 data). However, since growth data is not available for Somalia, it too ends up dropped from the sample. Thus, in the final panel of figure 2, and in all future tables and analysis, I use the instrument only for countries for which there is both data on growth from the Penn World Table (Heston et al., 2012) and which have values of the instrument between -0.8 and 0.8.

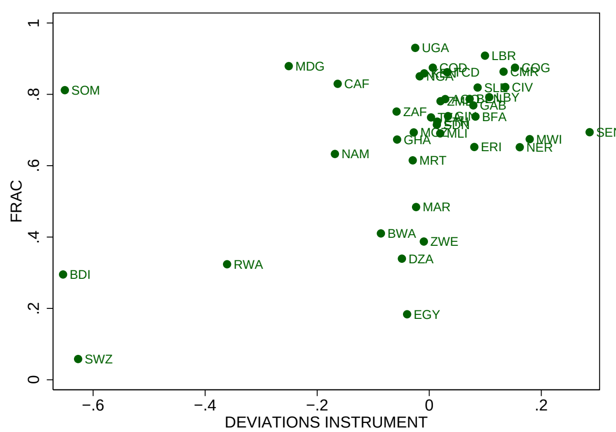
I next show regression results for the first stage, using the restricted sample. Table 1 looks at the first stage, progressively adding in geographic controls from Michalopoulos (2012). The instrument has a strongly significant effect on fractionalization, and does not appear to suffer from omitted variable bias, as the effect of the instrument is relatively stable to the inclusion of controls. Relatedly, Table 2 looks at whether the instrument can be predicted by the same set of controls. The table does not provide any evidence that the instrument is correlated with these controls. Thus, the instrument appears to be predictive of ELF as well as plausibly exogenous.

4 Results

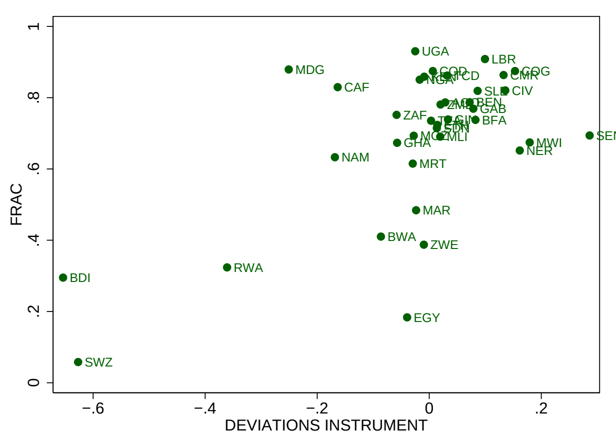
Since many early papers on ethnic fractionalization have focused on the effects on growth (Easterly and Levine, 1997; Alesina et al., 2003; Posner, 2004a), it is useful to focus on growth as a starting point for analyzing the causal effect of fractionalization on development. I use data on ethnic fractionalization from Alesina et al. (2003), and obtain my data on growth from the latest edition of the Penn World Table (Heston et al., 2012). I restrict the sample window from 1970 through 2010, as this provides a balanced panel for the thirty seven countries in the IV sample. All standard errors are clustered by country. In subsection 4.3, I apply the methods discussed here to other important outcomes that have been found in the previous literature to be negatively affected by ethnic fractionalization.



(a) First Stage Without Restrictions on the Instrument



(b) First Stage with the Instrument Trimmed



(c) First Stage with the Instrument Trimmed, Observations for Which Growth Data is Available

Figure 2: ELF vs. Deviations of Area Based ELF

Table 1: This shows the effects of the instrument on ELF with and without controls.

	(1) FRAC	(2) FRAC	(3) FRAC	(4) FRAC
DEVIATIONS INSTRUMENT	0.682***	0.621***	0.621***	0.558**
	(0.145)	(0.140)	(0.145)	(0.175)
ABS LATITUDE		-0.0102**	-0.0102**	-0.00684
		(0.00357)	(0.00342)	(0.00856)
DIST COAST			0.000000154	0.000216*
			(0.0000691)	(0.0000907)
SD ELEVATION				0.478
				(0.235)
SD AGRIC SUIT				0.341
				(0.225)
ELEVATION				-0.128
				(0.146)
AGRIC SUIT				0.0497
				(0.144)
PRECIPITATION				0.00158
				(0.00112)
TEMPERATURE				0.00143
				(0.0233)
MIG DIST ETHIOPIA				0.0202
				(0.0335)
Constant	0.703***	0.830***	0.829***	0.299
	(0.0269)	(0.0391)	(0.0504)	(0.701)
Observations	37	37	37	37
R^2	0.381	0.543	0.543	0.702
Adjusted R^2	0.363	0.516	0.502	0.588

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Dependent variable for all columns is the instrument used in later regressions.

	(1)	(2)	(3)	(4)
ABS LATITUDE		-0.00308	-0.00373	-0.0111
		(0.00458)	(0.00406)	(0.00956)
DIST COAST			-0.0000664	-0.0000402
			(0.0000922)	(0.0000916)
SD ELEVATION				-0.0890
				(0.335)
SD AGRIC SUIT				0.349
				(0.319)
ELEVATION				-0.206
				(0.186)
AGRIC SUIT				-0.177
				(0.172)
PRECIPITATION				0.000271
				(0.000817)
TEMPERATURE				-0.00107
				(0.0231)
MIG DIST ETHIOPIA				0.0107
				(0.0233)
Constant	-0.0262	0.0123	0.0586	0.331
	(0.0313)	(0.0673)	(0.0517)	(0.750)
Observations	37	37	37	37
R^2	0.000	0.018	0.034	0.388
Adjusted R^2	0.000	-0.010	-0.023	0.184

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.1 OLS Results

Before using the instrument, I first use a control function approach to deal with omitted variables bias. I first deal with omitted geographic variables, extending the work of Michalopoulos (2012). I then move on to deal with potential omitted variable bias due to the slave trade in Africa, based on Nunn (2008) and Nunn and Puga (2012). The results provide strong evidence of omitted variable bias, and suggest that the true effect of ethnic fractionalization on growth is substantially smaller than naive specifications suggest. Finally, I use disease related variables to provide suggestive evidence of reverse causality, building on the work of Acemoglu et al. (2001) and Alsan (2015).

4.1.1 Controlling for Geography

In this section, I use OLS and introduce geographic controls from Michalopoulos (2012). Michalopoulos finds that these geographic variables are major determinants of ethnic fractionalization. Since geography is exogenous, introducing these controls should not change the estimate of ethnic fractionalization's effect if the naive estimates are well identified. Table 3 shows the results. The estimates are reduced substantially by the inclusion of geographic controls, particularly absolute latitude, distance from the coast, and a dummy variable for Africa. All of these variables are predictors of ethnic fractionalization, and there is reason to believe that they have independent effects on growth. For example, Dell et al. (2012) find causal effects of temperature shocks on growth, and argue that this may explain some of the poor performance of countries close to the equator, and Nunn (2008) finds that the slave trade explains a substantial portion of Africa's underdevelopment, thus suggesting potential reasons other than just fractionalization that African countries have performed worse in terms of growth. Given the correlations between these three predictors and fractionalization, as well as their posited independent effect on development (countries far from the coast are thought to have greater difficulties engaging in trade), all three of these variables should produce omitted variables bias in favor of finding a negative effect of fractionalization; controlling for them should eliminate this bias.

Moreover, although the negative effect of fractionalization has been reduced substantially by the inclusion of geographic controls, that still does not mean we have achieved identification. The remaining variation in fractionalization is not well identified. We have already argued that it is likely polluted by reverse causality, and will next argue that it is potentially polluted by other omitted variables, such as the effects of the slave trade.

4.1.2 Controlling for Slave Trade Related Variables

An additional source of possible omitted variables bias is the slave trade. Nunn (2008) finds that the slave trade increased fractionalization and also harmed economic development. It seems likely that the slave trade harmed economic development through channels other than just ethnic fractionalization; if this is true then the slave trade will be a source of omitted variable bias, biasing our regressions towards finding a negative effect of fractionalization on development. To control for the slave trade, I estimate specifications based on Nunn (2008)

Table 3: Dependent variable is growth of GDP per capita. Estimated using Ordinary Least Squares. Standard errors clustered by country.

	(1)	(2)	(3)
FRAC	-0.0314***	-0.0197*	-0.0164*
	(0.00485)	(0.00816)	(0.00661)
ABS LATITUDE		0.0000132	-0.000335
		(0.0000927)	(0.000337)
DIST COAST		-0.00000349	-0.00000773
		(0.00000447)	(0.00000401)
AFRICA		-0.00889*	-0.00466
		(0.00393)	(0.00486)
SD ELEVATION			0.0118
			(0.00712)
SD AGRIC SUIT			-0.0126
			(0.0104)
ELEVATION			-0.00805
			(0.00785)
AGRIC SUIT			-0.00591
			(0.00600)
PRECIPITATION			-0.000136***
			(0.0000395)
TEMPERATURE			-0.000235
			(0.000602)
MIG DIST ETHIOPIA			-0.000765
			(0.000571)
EUROPE			0.00990
			(0.00515)
AMERICAS			0.0216*
			(0.00943)
EAST ASIA PACIFIC			0.0293***
			(0.00568)
Constant	0.0320***	0.0298***	0.0581*
	(0.00242)	(0.00531)	(0.0259)
Observations	6161	4910	4910
R^2	0.008	0.010	0.017

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Dependent variable is growth of GDP per capita. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012). Standard errors clustered by country.

	(1)	(2)	(3)	(4)	(5)
FRAC	-0.0314***	-0.0255***	-0.0432**	-0.0197	-0.0231
	(0.00485)	(0.00703)	(0.0138)	(0.0304)	(0.0314)
RUGGEDNESS		-0.000519			-0.00863
		(0.00144)			(0.00491)
RUGGED*AFRICA		-0.000800			
		(0.00270)			
AFRICA		-0.00706			
		(0.00569)			
SLAVE INTENSITY				-0.00276	-0.00505
				(0.00318)	(0.00401)
Constant	0.0320***	0.0325***	0.0354**	0.0318**	0.0520***
	(0.00242)	(0.00350)	(0.0104)	(0.0103)	(0.0133)
Observations	6140	6140	1927	1927	1927
R^2	0.008	0.009	0.010	0.008	0.006

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and on Nunn and Puga (2012), which finds that ruggedness of terrain had negative effects on development in most countries but positive effects in Africa, because ruggedness impeded the slave trade. Ruggedness is also closely related to the standard deviation of elevation, which was found in Michalopoulos (2012) to increase ethnic fractionalization. I thus first run a specification controlling for ruggedness interacted with an Africa dummy variable. I then focus specifically on Africa, controlling for the "slave intensity" variable (based on slave exports per unit of area) from Nunn (2008), using four variables representing distances from slave ports as instruments for slave intensity (these instruments come from Nunn 2008). Finally, I introduce ruggedness into this specification.

Table 4 shows the results. It appears that the slave trade is indeed a substantial source of omitted variables bias. Two things are important to note here. First, even though controlling for variables related to the slave trade has reduced the size of the estimated negative effect, we once again have reason to believe that there are still further sources of bias in our regressions (for example, there may still be reverse causality polluting our estimates). Second, we have begun to see the difficulty of obtaining precise estimates when we attempt to deal with endogeneity. Once we control for slave intensity, the estimate effect of fractionalization shrinks and becomes statistically insignificant. However, the estimated effect size is still non-negligible, but the standard errors are very large, corresponding to slightly more than three percentage points for a movement from full homogeneity to full heterogeneity. This issue of lack of precision will continue to be a problem for us in our IV regressions.

4.1.3 Reverse Causality Concerns

I next turn towards finding suggestive evidence of reverse causality. To do this I look at two disease related variables which have been found to have had substantial effects on economic development. These are the TseTse Suitability Index (TSI) from Alsan (2015), and settler mortality from Acemoglu et al. (2001). The TseTse fly was found in Alsan (2015) to have caused lower development in African countries where it was prevalent, through its lethal effect on livestock. Acemoglu et al. (2001) find that higher settler mortality lowered development in colonized countries by reducing the number of European settlers who emigrated to these countries.

The reason these two variables are particularly interesting is because although they both have substantial negative effects on development, the time frame of the two stories is quite different. If there is reverse causality from development outcomes to ethnic fractionalization, then we would expect TSI to have a significant effect on fractionalization, while the effect of settler mortality would be more muted, because settler mortality's effect is more recent and thus has had less time to affect fractionalization (fractionalization is thought to be slow to change; Alesina et al. 2003). Because both TSI and settler mortality are correlated with a number of other geographic variables such as distance from the equator, which are known to be correlated with fractionalization, I include geographic controls taken from Michalopoulos (2012). Table 5 shows the results. Indeed, TSI does appear to have a robust effect on ethnic fractionalization in the expected direction, while logged settler mortality does not keep a significant effect once controls are included. This provides suggestive evidence of a strong channel of reverse causality flowing from development to ethnic homogeneity.

4.2 IV Results

Finally, we turn to the use of the instrument. The results are in Table 6. The first-stage F-statistic shows that the instrument is strong. In order to tighten the standard errors, I add in controls for absolute latitude, as well as African regional dummies from Nunn (2008). Although these shrink the size of the standard errors, they are still relatively large. A Hausman test rejects the null hypothesis of no endogeneity (p=0.08 for no controls and p=0.04 with controls), however, sizable effects cannot be ruled out. However, it is clear that the naive estimates overstate the negative effect of fractionalization. Dealing with the endogeneity of ethnic fractionalization is an important issue, and deserves greater attention in the existing literature.

Table 5: TSI from Alsan (2015) for the African sample and settler mortality data from Acemoglu et al. (2001) for their base sample.

	(1) FRAC	(2) FRAC	(3) FRAC	(4) FRAC	(5) FRAC	(6) FRAC
TSI	0.169***	0.188***			0.162***	0.184*
	(0.0257)	(0.0510)			(0.0337)	(0.0732)
ABS LATITUDE		-0.00670		-0.0169***		-0.00124
		(0.00832)		(0.00465)		(0.00827)
SD ELEVATION		0.444		0.241		0.794**
		(0.331)		(0.130)		(0.268)
SD AGRIC SUIT		0.438		0.447		0.334
		(0.281)		(0.245)		(0.235)
ELEVATION		0.203		-0.349**		0.185
		(0.178)		(0.124)		(0.230)
AGRIC SUIT		-0.144		0.0109		0.0870
		(0.140)		(0.137)		(0.176)
PRECIPITATION		0.000541		0.0000757		0.0000354
		(0.000930)		(0.000749)		(0.000977)
TEMPERATURE		0.0194		-0.0184*		0.0312
		(0.0237)		(0.00857)		(0.0214)
DIST COAST		0.0000359		0.000162		0.000107
		(0.0000708)		(0.0000839)		(0.000116)
MIG DIST ETHIOPIA		0.0567		-0.0112**		0.0751*
		(0.0305)		(0.00372)		(0.0308)
LN(SETT. MORTALITY)			0.0968***	0.0447	-0.00994	-0.0151
			(0.0192)	(0.0242)	(0.0278)	(0.0300)
Constant	0.676***	-0.151	0.0851	1.091**	0.750***	-0.693
	(0.0239)	(0.732)	(0.0918)	(0.375)	(0.155)	(0.652)
Observations	47	45	87	75	36	35
R^2	0.488	0.676	0.230	0.535	0.468	0.779

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Dependent variable is growth of GDP per capita. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders. Standard errors clustered by country.

	(1)	(2)	(3)	(4)
FRAC	-0.0446**	-0.0353*	-0.0189	-0.00101
	(0.0144)	(0.0144)	(0.0174)	(0.0120)
ABS LATITUDE		0.000605		0.000727
		(0.000419)		(0.000514)
AFRICA - NORTH		0.00500		0.0171
		(0.0121)		(0.0147)
AFRICA - SOUTH		0.00600		0.0119
		(0.00811)		(0.0102)
AFRICA - EAST		0.00483		0.00771
		(0.00618)		(0.00693)
AFRICA - CENTRAL		0.0132		0.0108
		(0.00672)		(0.00727)
Constant	0.0364**	0.0174	0.0188	-0.00974
	(0.0109)	(0.0121)	(0.0123)	(0.0118)
Observations	1517	1517	1517	1517
R^2	0.008	0.011	0.005	0.008
FirstStageF			22.63	30.44

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.3 Other Outcomes

I have repeated the analysis discussed above using the natural logarithm of GDP per capita as an outcome. The results are reported in the appendix. The general conclusions from the analysis on growth carry over to the analysis for GDP per capita. In every instance, the negative effect of ethnic fractionalization on logged GDP per capita found in the "naive" specification was reduced substantially when some method (either the inclusion of controls or the use of the instrument) was used to deal with endogeneity issues. In fact, the coefficient frequently flipped signs to suggest a beneficial effect of fractionalization, although the effect in these cases was never statistically significant. These results confirm the general conclusion of the growth results: naive estimates of the effects of fractionalization on GDP per capita likely overstate the true effects of fractionalization, but the true effect is difficult to estimate precisely in a well-identified way.

I also repeat the analysis using a number of measures of public goods provision. I select these measures to follow as closely as possible those used by La Porta et al. (1999), (Alesina et al., 2003), and Desmet et al. (2012) to measure the effect of fractionalization on public goods provision. I select the year at which to measure the statistics based on a desire to use the most recent year possible while also maximizing sample coverage. The results are reported in the appendix. The conclusions are broadly similar to those found for growth and for GDP per capita: controlling for omitted variables makes the estimated effect of fractionalization on public goods provision less negative, and the null effect is quite plausible.

4.4 Robustness Checks

4.4.1 Measurement Error

One potential concern in the OLS specifications is measurement error. If ethnic fractionalization is measured with error, then including controls which explain a portion of true ethnic fractionalization will bias our estimates towards zero, regardless of any omitted variables bias that may exist. In order to partially deal with this concern, I repeat the specification from Table 3, instrumenting the Alesina et al. (2003) measure of ethnic fractionalization with the Fearon (2003) measure [3]. The results are reported in the appendix. The results are broadly similar to the results in Table 3, and I do not find any evidence that measurement error is driving my results. Moreover, the regressions using my instrument should not be biased by any measurement error in the Alesina et al. (2003) index.

4.4.2 Controlling for Income

In the existing literature, as in most cross-country growth regressions, logged GDP per capita is usually included as a control in regressions of growth on fractionalization. Although there is some theoretical justification for including GDP per capita in the regression, I exclude it in

[3] This will deal with any measurement error that is uncorrelated between the two indices; to the extent that there is measurement error that is correlated between the two indices then instrumenting will not be sufficient

my main set of regressions, based on concerns that including it may falsely bias the effect of fractionalization towards zero. If fractionalization does indeed cause lowered GDP per capita, then controlling for GDP per capita will cut off any effect of fractionalization on growth that runs through GDP per capita; moreover, if fractionalization is measured with error then the coefficient on GDP per capita could pick up the effect of fractionalization to the extent that GDP per capita provides information about true fractionalization not provided in the index of fractionalization included in our regression. For robustness, however, I also rerun the growth regressions including logged GDP per capita as a control: the results are reported in the appendix. The results are broadly similar to those in the main specification, although the regressions using slave trade controls see substantially larger standard errors and a somewhat less dramatic reduction in the magnitude of the coefficients after the addition of controls.

4.5 Connection with "Bad Borders"

The instrument I construct is intimately connected with the notion of "bad" borders: borders which were arbitrarily drawn on maps in a way that does not match meaningful boundaries on the ground. This connection is important for understanding what LATE we are estimating. Many authors and commentators, such as Herbst (2000), have argued that arbitrarily drawn borders are a key explanation for political and economic failures in Africa. The variation I exploit is a result of such bad borders: the countries with high values of my instrument are countries whose borders were arbitrarily drawn to have higher than expected fractionalization. Thus, the IV regression identifies an LATE that coincides with the effect of bad borders on ethnic fractionalization, and thus provides us an answer to the question of whether bad borders that yielded higher ethnic fractionalization are responsible for a lack of growth in Africa. This is a particularly useful LATE to identify, since bad borders yielding excessive fractionalization is a frequently raised concern.

Beyond giving us insight into the LATE estimated by the IV regression, this understanding of the instrument as being driven by bad borders also helps us think about potential omitted variables bias in the instrument. In particular, since the instrument reflects bad borders, if bad borders are correlated with development separately from the effect of the instrument on development through fractionalization, then we have a problem of omitted variables bias. Here, we could either have a problem because the assumption that the placement of borders is random fails (i.e. perhaps some places have better-drawn borders for non-arbitrary reasons) or because of a direct negative effect of bad borders on development (e.g. bad borders split up economic centers leading to poor outcomes). Whichever of these is at play, we know from Alesina et al. (2011) that two measures of artificiality of borders, one based on the splitting of ethnic groups by borders and one based on the straightness of the borders, appear to be negatively correlated with development. More recent work shows in particular that conflict is substantially more likely among African ethnic groups that were partitioned to lie on both sides of a border (Michalopoulos and Papaioannou, 2011), backing up our belief that bad borders are bad for development. Given this, we would thus expect the IV regressions to be, if anything, biased downwards, since the instrument is also picking up the negative correlation between bad borders and development. Since the key result of our

IV regressions is that OLS gives heavily negative biased results, and that appropriate controls and an instrument cause this negative coefficient to shrink or disappear, this direction of bias, if anything, reinforces our results.

5 Game-Theoretic Model

5.1 Motivation

At this point, an advocate of the hypothesis that ethnic diversity harms development might turn to the within-country evidence for support. Prior literature has taken micro-level evidence of harmful ethnic favoritism and ethnic politics as evidence for a negative macro-level effect of ethnic diversity on development (Alesina et al. (2003); Easterly and Levine (1997)). Since there is reasonably convincing micro-level evidence of ethnic favoritism (see, for example, Burgess et al. 2013; Marx et al. 2014; Hjort 2014), one might argue that we can use this evidence to show that ethnic diversity harms development, regardless of any issues with the cross-country evidence.

However, this conclusion does not necessarily follow. In order to make this point robustly, in this section I develop a straightforward formal model in which political coalitions endogenously form along ethnic lines. I then show how in this model it is possible for an increase in ethnic diversity to be associated with an improvement in policy, a worsening in policy, or no effect on policy, even though there is rent-seeking political activity taking place along ethnic lines. I do not rely on the often discussed positive effects of diversity on productivity (Alesina and La Ferrara, 2005); instead I focus exclusively on how ethnic diversity affects policy through its effects on coalition formation.

5.2 Model

Consider a model in which members of different ethnic groups choose whether or not to "loot". Looting is a form of rent seeking, and can be taken to mean different things in different settings; looting might mean deciding to start a civil war or it might mean simply enacting a government policy that benefits oneself rather than the optimal policy. The decision of whether or not to loot is intended to capture the idea that ethnicity may lead to a lack of national cooperation; rather than all cooperating to better the country as a whole, different ethnic groups may jostle with each other for power. In the model, individuals can band together to attempt to loot in political groups; this increases their probability of success but also requires that they share the benefits (for simplicity we will assume that benefits must be shared equally among all members). We assume that political group formation requires the consent of all members of the political group being formed.[4]

[4]To make this more rigorous, we must describe the entry into political groups in terms of strategies played by the agents. We require agents to list themselves and all of the other agents who they would like in their political group: if all of those people have the same list, then the political group is formed and the agent receives the appropriate payoff. If the political group does not form properly, then we do not

Each individual, once she has chosen her political group, has two options: she may loot or she may not loot. Her probability of success is a weakly increasing function of the size of her political group, denoted $F(N_i)$, where N_i is the political group size of individual i. We normalize the payoff to not looting to be one and the payoff to unsuccessful looting to be zero. The collective payoff to a successful looting is denoted θ, and must be split evenly among all members of the political group; the payoff to a successful individual is thus $\frac{\theta}{N_i}$.

To capture losses associated with rent-seeking behavior, we assume that there is some societal loss associated with looting, which is spread equally across society. Each unit of looting is associated with β units of losses, where $\beta > 1$. Thus, each individual faces the cost $\frac{\beta}{N_{total}} \sum_j \frac{\theta}{N_j} F(N_j)$, where N_{total} refers to the size of the entire country's population.

Finally, we assume there is some cost to inter-ethnic cooperation. We assume the communication cost between two individuals i and j depends solely on the ethnic group of i and the ethnic group of j. Co-ethnics face no communication cost, while non-coethnics face a positive communication cost that depends on their ethnic groups.[5] Each individual pays a cost equal to the average communication cost between them and their fellow members, which we denote $\bar{\delta}_i$.[6]

To summarize, the expected payoffs are given as follows:

$$EU_i = \begin{cases} 1 - \bar{\delta}_i - \frac{\beta}{N_{total}} \sum_j \frac{\theta}{N_j} F(N_j) & \text{if Doesn't Loot} \\ \frac{\theta}{N_i} F(N_i) - \bar{\delta}_i - \frac{\beta}{N_{total}} \sum_j \frac{\theta}{N_j} F(N_j) & \text{if Loots} \end{cases}$$

We will consider Coalition-Proof Nash Equilibria, as defined in Bernheim et al. (1987). We will also impose the additional condition that all individuals in a political group must choose the same action; since this condition could only be violated when agents are indifferent, this restriction will yield a subset of the possible equilibria that will be non-empty as long as the overall set of equilibria is non-empty. This restriction is also more realistic; if only some group members are actually engaged in looting, the presence of non-looters in the group is not likely to contribute to the success of the looting, so $F(N_i)$ should really only count the number of looters in the group.

We now assume a continuum of agents for each ethnic group (this is equivalent to assuming that all the ethnic groups are large) and assume that $F(N)$ is right continuous.

give the agent any of the positive benefits of being in the political group, however, we do make that agent pay the interethnic communication cost associated with the group they wanted to form, and we also figure out what the largest political group they could have formed from their list is, and we impose the harms from looting onto society that would have arisen from that political group (the interethnic communication cost and the social harm from looting are described in later paragraphs). Doing this will, under our use of the coalition-proof equilibrium concept, make our game well-behaved in that it will look like a game where political groups are formed by unanimous consent of the group members.

[5]This assumption, in some form, is standard in the literature (see for example Alesina et al. 2004; Esteban and Ray 2011).

[6]This could represent a number of different things; perhaps different ethnicities simply dislike interacting with each other, perhaps it is easier to communicate with co-ethnics, or perhaps it is easier to organize co-ethnics and punish those who do not contribute to the looting effort.

Theorem 1. *In equilibrium, for any given individual who wishes to join a political group, there will be a political group that includes some of her co-ethnics that she weakly prefers over any political group that includes none of her co-ethnics.*

Proof. If the individual's best option is not to loot, then this result follows trivially. If the individual's best option is to loot, she can do so in a group with co-ethnics or in a group without any co-ethnics. Since we have assumed a continuum of agents and right continuity of $F(N)$, we know that the utility she gets from joining the best group that includes her co-ethnics will be the utility that her co-ethnics who are already a member of that group get. Since the co-ethnics are already in the group, that means that being in the group must be weakly preferred to all other options (this logic is also what ensures that there will be a group of her co-ethnics who are looting; if looting is the best option than at least one of her co-ethnics will be looting). This proves the proposition. $\square$

In the case where looters must play a strict best response, then we can strengthen this result.[7]

Theorem 2. *If all looters are playing a strict best response, looting groups will be composed of one or more whole ethnic groups.*

Proof. By the common utility function of co-ethnics, the right continuity of $F(N)$, and the assumption of a continuum of agents, we also know that if one ethnic group member is looting in a certain group, then any other co-ethnics could get the same utility by doing the same thing. Thus, if two co-ethnics, one of whom is a looter, are doing different things, they must be indifferent between the two actions. However, this would violate the assumption of strict best response for looters. Thus all co-ethnics must be doing the same thing if one of them is looting, and thus all looting groups must be composed of entire ethnic groups. $\square$

These two results are interesting because they highlight the idea that political groups will form along ethnic lines, which seems to happen in the real world. The results are driven by the lower costs of group membership for co-ethnics, something which has been suggested by a number of prior papers (Miguel and Gugerty, 2005; Habyarimana et al., 2007), has been used extensively in prior theoretical work (Alesina and Spolaore, 1997; Alesina et al., 2004; Esteban and Ray, 2011), and is consistent with the large literature which finds political groups forming along ethnic lines (Eifert et al., 2010). The results are not theoretically difficult to derive; they follow almost immediately from the assumptions. They are interesting, however, because they show that making the assumption that, in one form or another, is standard in the literature (non-coethnics face cooperation costs that co-ethnics do not) is sufficient to yield ethnic coalition building.[8] Moreover, the model still allows for the creation of political coalitions consisting of more than one ethnic group, if

[7] We will not assume in general that looters play strict best responses, and in fact we will examine cases where some looters are playing a weak best response. However, this result gives us a stronger sense of how inter-ethnic communication costs lead to the formation of political groups along ethnic lines.

[8] This is related to the results of Alesina and Spolaore (1997), who model nation formation as a tradeoff between the costs of diversity and returns to scale.

it turns out that the probability of looting success will be sufficiently improved by inter-ethnic cooperation. This explains the result of Posner (2004b), who argues that Chewas and Tumbukas are allies in Zambia and adversaries in Malawi because of the relative feasibility of their ethnic groups as political groups in each country, as well as similar results in Posner (2005) when looking at the formation of ethnic coalitions in Zambia.

It is worthwhile to briefly talk about efficiency in the model. The model allows for Pareto Inefficient outcomes (e.g. when all individuals are looting), but there may also be times when it appears that the outcome is efficient, even though there is looting (e.g. when exactly one political group is looting). This might cause some readers to discount the value of the model, as we expect rent-seeking (which is what "looting" is intended to represent) to be Pareto Inefficient. However, although within the model such an outcome is Pareto efficient, putting the model in a broader context (i.e. imagining payoffs as cash and allowing for transfers) reveals that such an outcome would only be constrained efficient: it is efficient within the mechanisms available in the model, but since total surplus decreases with looting, a social planner with the ability to make transfers could achieve a Pareto improvement by having no looting and making the appropriate transfers.

5.3 Effects of Diversity

We now turn to analyzing the effects of ethnic diversity within the model. To do this, we will think about the effect of splitting an existing ethnic group into two ethnic groups. It turns out that, in the model, the effects of diversity are ambiguous. We will demonstrate this through three theorems. From now on, we will assume that an equilibrium exists. We first consider the case where inter-ethnic communication costs are very large.

Theorem 3. *If inter-ethnic communication is prohibitively costly, and a particular ethnic group is composed either entirely of looters or entirely of non-looters, then splitting that ethnic group will weakly decrease looting and weakly increase the sum of expected payoffs across agents.*

Proof. There are two cases: either all members of the ethnic group in question are looting or they are all not looting. If they are all not looting, then splitting will not matter; their rewards to looting have simply decreased even more, since the maximum possible group size of co-ethnics has shrunk (members of the new ethnic subgroups can only form viable coalitions with their subgroup members). If the whole ethnic group is looting (not necessarily in a single political group), then the splitting will limit the possible group sizes. This may have no effect, but may make looting unprofitable if $\frac{F(N)}{N}$ is convex in the relevant region. Thus, looting will weakly decrease after splitting. Since communication costs are zero they will never be paid in equilibrium, and since $\beta > 1$, looting strictly decreases the sum of expected payoffs. Thus, since looting weakly decreases, the sum of expected payoffs across agents weakly increases. □

This theorem could apply even in cases of extreme changes in ethnic diversity. For example, under a supermajority rule (looting will be successful if and only if more than two thirds of the population participates), moving from perfect homogeneity to perfect polarization (two equally sized groups) will reduce looting for the appropriate θ and β, and the same will happen when moving from perfect homogeneity to perfect heterogeneity (all different ethnic groups).

Theorem 4. *If inter-ethnic communication is prohibitively costly, and a particular ethnic group has both looters and non-looters, then splitting that ethnic group may increase looting and decrease the sum of expected payoffs across agents.*

Proof. Consider an $F(N)$ such that, for our ethnic group of interest, looting is sure to be successful with two thirds or more of the ethnic group participating, just under three quarters sure for group sizes in between one half (inclusive) and two thirds (exclusive) of the ethnic group, and not possible for group sizes under one half. If the ethnic group is sufficiently small as a share of the total population, then for β sufficiently small and θ sufficiently large, our initial equilibrium is that a lucky two thirds of the ethnic group will form a group and loot, and the remaining third will find looting unprofitable. If we then split the ethnic group into equally sized halves, then the new equilibrium will have each new ethnic subgroup form a political group and loot (for θ sufficiently large). Since there is no inter-ethnic communication and looting is welfare decreasing, the sum of expected payoffs across all agents in the country is decreased. $\square$

These two results demonstrate how, in an environment in which inter-ethnic coalitions are not viable, the effects of ethnic diversity are ambiguous. We will now look at an opposite extreme: the case where inter-ethnic distances are very small.

Theorem 5. *As the inter-ethnic communication cost between two groups approaches some value δ, the associated equilibria approach*[9] *allowed equilibria under communication cost δ. In other words, the set of equilibria is upper hemicontinuous in inter-ethnic communication cost.*[10]

Proof. Let the sequence δ_n be the distance between two given ethnic groups, and let $\delta_n \to \delta$. Fix all other primitives of the game. Let A_n be an equilibrium associated with δ_n, and let $A_n \to A$. We wish to show that A is an equilibrium under δ. For this, since what matters is the mapping between strategy profiles and payoffs, we examine how payoffs behave as $\delta_n \to \delta$. Since the expected payoff is linear in ethnic communication costs, it is continuous in δ_n, and thus the expected payoffs (for each strategy profile) associated with δ_n approach those associated with δ. The one difference is that in the limit, there will be equality between some payoffs for which there was an inequality at each point along the sequence. This does not, however, matter for whether A is a valid equilibrium under δ: although the equality of payoffs may mean that some additional equilibria are now possible, indifference means that agents will still be willing to enact the limiting equilibrium. Similarly, some

[9] We have not defined a metric between equilibria: one possible choice is the number of agents who we would have to take from one political group and put into another political group in order to get from one equilibrium to the other.

[10] Lower hemicontinuity will not hold in general, as one can see easily by analyzing $F(N_i) = N_i/N_{total}$.

coalitional deviations that were previously not self-enforcing may now become self-enforcing; however those deviations will all be deviations that no agent strictly prefers and thus coalition proofness does not require us to enact those deviations. Thus A is a valid equilibrium under the inter-ethnic communication cost δ. □

From this, we get the following corollary:

Corollary 6. *If inter-ethnic communication is sufficiently cheap, and there is the same level of looting in all possible equilibria in the counterfactual country where all agents have the same ethnicity, then splitting will have approximately no effect on the level of looting or the sum of expected payoffs across agents.*[11]

Proof. As the cost of inter-ethnic communication approaches zero, the initial equilibrium will approach an equilibrium that would have also been a possible equilibrium if inter-ethnic communication were costless. By Theorem 5, the equilibrium after the split will also approach an equilibrium drawn from the set of equilibria that would have been possible if inter-ethnic communication were costless (i.e. the country with perfect homogeneity). Since we have assumed all such equilibria have the same level of looting, this means that the equilibrium after the split will have the same level of looting as the initial equilibrium. Since communication costs are negligible, there is also approximately no effect on the sum of expected payoffs across agents. □

Corollary 6 suggests that if ethnic differences are sufficiently small, then ethnic diversity will not have aggregate effects. However, since we only observe equilibrium outcomes, it is difficult to tell which ethnic distances are truly large or small. Posner (2004b) observes that in Zambia, where Chewas and Tumbukas are political allies, members of these ethnic groups consider the other group to be very ethnically similar or even identical, while in Malawi group members emphasize ethnic distinctions; Posner argues that these differences are the result of the relative political viability of independent Chewa and Tumbuka coalitions in each country. The fact that such inter-ethnic coalition building is viable in this context suggests that ethnic distances may be small in other contexts, particularly in areas where the pre-colonial state was weak, as was the case in Zambia Posner (2005). Ethnic distances may be larger between ethnicities where a strong sense of national ethnic identity was fostered by a strong state, as was the case with the French Weber (1976).

This discussion also raises an interesting possibility. If ethnic differences in general are in fact largely superficial, then it may be the case that, in each country, individuals will find some difference that they will make salient. For example, in Britain ethnic differences between the English, Welsh, Irish, and Scottish are salient; in the United States these groups would all be categorized as white. Individuals could also cleave along other lines, such as region, wealth, or religion.

More generally, Theorems 3 and 4 and Corollary 6 show that the effects of increased ethnic diversity is quite ambiguous, even as Theorems 1 and 2 show that the model will generally

[11] Additional technical conditions may be required for the limit arguments used in the proof to hold, but we will simply ignore these pathological cases.

yield some form of ethnic politics. Thus, the micro-level evidence of ethnic favoritism must not be viewed as evidence of a negative effect of ethnic diversity.

6 Conclusion

This paper intends to convey two main messages. First, there is reason to believe, both due to economic argument and econometric evidence, that endogeneity of ethnic fractionalization is an important issue, and naive regressions will overstate the negative effect of fractionalization. In fact, the econometric evidence suggests that a null effect of ethnic fractionalization is well within the plausible range of true causal effects. The micro-level evidence of ethnic favoritism is also insufficient to demonstrate a negative effect of fractionalization, as there are models consistent with ethnic favoritism with ambiguous effects of fractionalization. Second, it is very difficult to obtain estimates of the effects of fractionalization that are both well-identified and precise. Most of the variation in fractionalization is not exogenous. Moreover, cross-country regressions are frequently underpowered because there are not many countries, ethnic fractionalization does not vary quickly over time, and some identification strategies (e.g. the instrument I developed in this paper) require restricting the sample to an even smaller set of countries.

These problems are common in political economy, but that does not mean we should ignore them. We are better served by obtaining the most accurate estimates we can and acknowledging the limitations of what we know. This paper has cast doubt on claims that ethnic fractionalization is a major determinant of poor growth, although the large standard errors leave room for sizable, although reduced, effects. Nonetheless, the micro-level evidence suggests ethnicity plays an important role in political economy, regardless of whether diversity has aggregate effects on development. Solving these dual econometric issues of identification and lack of statistical power, as well as developing models to connect micro-level evidence to macro-level effects, is an important area for future research.

References

Acemoglu, Daron, Simon Johnson, and James A Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, December 2001, *91* (5), 1369–1401.

Alesina, Alberto and Eliana La Ferrara, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, September 2005, *43* (3), 762–800.

_ **and Enrico Spolaore**, "On the Number and Size of Nations," *Quarterly Journal of Economics*, 1997, *112*, 1027–1056.

_ **, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg**, "Fractionalization," *Journal of Economic Growth*, 2003, *8*, 155–194.

— , **Reza Baqir, and Caroline Hoxby**, "Political Jurisdictions in Heterogeneous Communities," *Journal of Political Economy*, April 2004, *112* (2), 348–396.

— , **William Easterly, and Janina Matuszeski**, "ARTIFICIAL STATES," *Journal of the European Economic Association*, April 2011, *9* (2), 246–277.

Alsan, Marcella, "The Effect of the TseTse Fly on African Development," *American Economic Review*, January 2015, *105* (1), 382–410.

Balcells, Laia, "Mass Schooling and Catalan Nationalism," *Nationalism and Ethnic Politics*, October 2013, *19* (4), 467–486.

Bernheim, B.Douglas, Bezalel Peleg, and Michael D Whinston, "Coalition-Proof Nash Equilibria I. Concepts," 1987.

Burgess, Robin, Edward Miguel, Remi Jedwab, and Ameet Morjaria, "The Value of Democracy: Evidence from Road Building in Kenya," *NBER Working Paper 19398*, 2013.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken, "Temperature Shocks and Economic Growth: Evidence from the Last Half Century," *American Economic Journal: Macroeconomics*, July 2012, *4* (3), 66–95.

Desmet, Klaus, Ignacio Ortu no Ortín, and Romain Wacziarg, "The political economy of linguistic cleavages," *Journal of Development Economics*, March 2012, *97* (2), 322–338.

Easterly, William and Ross Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *The Quarterly Journal of Economics*, November 1997, *112* (4), 1203–1250.

Eifert, Benn, Edward Miguel, and Daniel N Posner, "Political Competition and Ethnic Identification in Africa," *American Journal of Political Science*, 2010, *54*, 494–510.

Esteban, Joan and Debraj Ray, "Linking conflict to inequality and polarization," *American Economic Review*, 2011, *101*, 1345–1374.

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

Habyarimana, James, Macartan Humphreys, Daniel N Posner, and Jeremy M Weinstein, "Why Does Ethnic Diversity Undermine Public Goods Provision?," *The American Political Science Review*, 2007, *101*, 709–725.

Herbst, Jeffrey, *States and Power in Africa*, Vol. 15 2000.

Heston, Alan, Robert Summers, and Bettina Aten, "Penn World Table Version 7.1," 2012.

Hjort, J., "Ethnic Divisions and Production in Firms," *The Quarterly Journal of Economics*, October 2014.

La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny, "The quality of government," *Journal of Law, Economics, and Organization*, March 1999, *15* (1), 222–279.

Marx, Benjamin, Thomas Stoker, and Tavneet Suri, "There Is No Free House: Ethnic Patronage and Diversity in a Kenyan Slum," 2014.

Michalopoulos, Stelios, "The Origins of Ethnolinguistic Diversity," *American Economic Review*, June 2012, *102* (4), 1508–1539.

_ **and Elias Papaioannou**, "The Long-Run Effects of the Scramble for Africa," 2011.

_ **and** _ , "National Institutions and Subnational Development in Africa," *The Quarterly Journal of Economics*, October 2013, *129* (1), 151–213.

_ **and** _ , "Pre-Colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 2013, *81*, 113–152.

Miguel, Edward, "Tribe or Nation?: Nation Building and Public Goods in Kenya versus Tanzania," *World Politics*, 2004, *56* (3), 327–362.

_ **and Mary Kay Gugerty**, "Ethnic diversity, social sanctions, and public goods in Kenya," *Journal of Public Economics*, December 2005, *89* (11-12), 2325–2368.

Nunn, Nathan, "The Long-Term Effects of Africa's Slave Trades *," *Quarterly Journal of Economics*, February 2008, *123* (1), 139–176.

_ **and Diego Puga**, "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 2012, *94*, 20–36.

Posner, Daniel N, "Measuring Ethnic Fractionalization in Africa," *American Journal of Political Science*, October 2004, *48* (4), 849–863.

_ , "The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi," *American Political Science Review*, November 2004, *98* (04), 219–21.

_ , *Institutions and ethnic politics in Africa*, New York: Cambridge University Press, 2005.

Weber, Eugen, *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914* 1976.

Weese, Eric, "Endogeneity of Linguistic Fragmentation and Implications," 2011.

7 Appendix

7.1 Measurement Error

Table 7: Dependent variable is growth of GDP per capita. Fractionalization index from Alesina et al. (2003) instrumented for using index from Fearon (2003). Standard errors clustered by country.

	(1)	(2)	(3)
FRAC	-0.0320***	-0.0174	-0.0173
	(0.00651)	(0.0112)	(0.00981)
ABS LATITUDE		0.0000283	-0.000348
		(0.000101)	(0.000366)
DIST COAST		-0.00000379	-0.00000763
		(0.00000459)	(0.00000403)
AFRICA		-0.00924*	-0.00463
		(0.00423)	(0.00485)
SD ELEVATION			0.0119
			(0.00744)
SD AGRIC SUIT			-0.0124
			(0.0105)
ELEVATION			-0.00823
			(0.00839)
AGRIC SUIT			-0.00596
			(0.00589)
PRECIPITATION			-0.000136***
			(0.0000393)
TEMPERATURE			-0.000248
			(0.000634)
MIG DIST ETHIOPIA			-0.000778
			(0.000583)
EUROPE			0.00984
			(0.00530)
AMERICAS			0.0217*
			(0.00941)
EAST ASIA PACIFIC			0.0292***
			(0.00591)
Constant	0.0313***	0.0286***	0.0593*
	(0.00303)	(0.00637)	(0.0288)
Observations	4910	4910	4910
R^2	0.008	0.010	0.017

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.2 Other Outcomes

For most outcomes, only the coefficient on fractionalization is displayed. To determine what is being controlled for in each column, note that the tables follow the same template as Tables 8, 9, and 10.

7.2.1 Logged GDP Per Capita

Table 8: Dependent variable is logged GDP per capita in 2010. Estimated using Ordinary Least Squares.

	(1)	(2)	(3)
FRAC	-2.631***	-0.426	-0.603
	(0.349)	(0.398)	(0.347)
ABS LATITUDE		0.0331***	-0.0429*
		(0.00550)	(0.0181)
DIST COAST		-0.000512***	-0.000789***
		(0.000146)	(0.000209)
AFRICA		-0.996***	-1.156***
		(0.189)	(0.269)
SD ELEVATION			-0.389
			(0.275)
SD AGRIC SUIT			1.086
			(0.582)
ELEVATION			-0.767**
			(0.272)
AGRIC SUIT			-0.754*
			(0.328)
PRECIPITATION			-0.00548**
			(0.00209)
TEMPERATURE			-0.115***
			(0.0323)
MIG DIST ETHIOPIA			0.0138
			(0.0299)
EUROPE			0.392
			(0.287)
AMERICAS			0.183
			(0.547)
EAST ASIA PACIFIC			0.146
			(0.324)
Constant	10.01***	8.508***	14.06***
	(0.171)	(0.318)	(1.381)
Observations	161	130	130
R^2	0.273	0.659	0.750

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Dependent variable is logged GDP per capita in 2010. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012).

	(1)	(2)	(3)	(4)	(5)
FRAC	-2.639***	-1.357***	-1.298*	1.252	1.002
	(0.349)	(0.355)	(0.554)	(1.525)	(1.709)
RUGGEDNESS		-0.174*			-0.486**
		(0.0723)			(0.173)
RUGGED*AFRICA		0.0398			
		(0.115)			
AFRICA		-1.669***			
		(0.252)			
SLAVE INTENSITY				-0.300	-0.422*
				(0.154)	(0.210)
Constant	10.01***	10.15***	8.312***	7.926***	9.072***
	(0.171)	(0.184)	(0.400)	(0.442)	(0.348)
Observations	160	160	47	47	47
R^2	0.274	0.536	0.123	.	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Dependent variable is logged GDP per capita in 2010. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders.

	(1)	(2)	(3)	(4)
FRAC	-1.370*	-0.610	-0.165	0.426
	(0.675)	(0.675)	(1.254)	(0.945)
ABS LATITUDE		0.0298		0.0335
		(0.0263)		(0.0273)
AFRICA - NORTH		0.492		0.859
		(0.515)		(0.657)
AFRICA - SOUTH		0.625		0.802
		(0.453)		(0.536)
AFRICA - EAST		-0.0116		0.0752
		(0.295)		(0.310)
AFRICA - CENTRAL		0.691		0.620
		(0.473)		(0.497)
Constant	8.344***	7.161***	7.518***	6.341***
	(0.502)	(0.681)	(0.894)	(0.831)
Observations	37	37	37	37
R^2	0.101	0.328	0.023	0.292
FirstStageF			22.02	25.47

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.2.2 Child and Infant Mortality

Table 11: Dependent variable is logarithm of under 5 mortality rater per 1,000 live births in 2011. Estimated using Ordinary Least Squares.

	(1)	(2)	(3)
FRAC	2.350***	0.441	0.503
	(0.282)	(0.319)	(0.274)
Observations	184	142	142
R^2	0.278	0.734	0.811

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Dependent variable is logarithm of under 5 mortality rater per 1,000 live births in 2011. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012).

	(1)	(2)	(3)	(4)	(5)
FRAC	2.365***	1.204***	1.252**	-0.161	-0.208
	(0.281)	(0.299)	(0.417)	(0.854)	(0.814)
Observations	183	183	52	52	52
R^2	0.283	0.514	0.225	0.164	0.039

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Dependent variable is logarithm of under 5 mortality rater per 1,000 live births in 2011. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders.

	(1)	(2)	(3)	(4)
FRAC	1.056*	-0.119	-0.293	-1.059
	(0.487)	(0.272)	(0.717)	(0.766)
Observations	41	40	41	40
R^2	0.158	0.711	.	0.628
FirstStageF			5.145	6.769

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Dependent variable is logarithm of infant mortality rater per 1,000 live births in 2011. Estimated using Ordinary Least Squares.

	(1)	(2)	(3)
FRAC	2.152***	0.408	0.463
	(0.267)	(0.308)	(0.260)
Observations	184	142	142
R^2	0.264	0.709	0.796

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Dependent variable is logarithm of infant mortality rater per 1,000 live births in 2011. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012).

	(1)	(2)	(3)	(4)	(5)
FRAC	2.166***	1.166***	1.034**	-0.140	-0.157
	(0.266)	(0.290)	(0.365)	(0.740)	(0.706)
Observations	183	183	52	52	52
R^2	0.268	0.473	0.208	0.131	0.012

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16: Dependent variable is logarithm of infant mortality rater per 1,000 live births in 2011. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders.

	(1)	(2)	(3)	(4)
FRAC	0.893*	-0.110	-0.264	-1.031
	(0.405)	(0.237)	(0.611)	(0.698)
Observations	41	40	41	40
R^2	0.156	0.688	.	0.578
FirstStageF			5.145	6.769

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.2.3 Clean Water and Sanitation

Table 17: Dependent variable is percent of population with access to an improved water source in 2007. Estimated using Ordinary Least Squares.

	(1)	(2)	(3)
FRAC	-30.75***	-6.298	-6.109
	(4.421)	(5.826)	(5.766)
Observations	179	140	140
R^2	0.233	0.542	0.597

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Dependent variable is percent of population with access to an improved water source in 2007. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012).

	(1)	(2)	(3)	(4)	(5)
FRAC	-30.94***	-13.55**	-36.52***	-27.46	-23.46
	(4.418)	(4.576)	(7.722)	(22.32)	(21.39)
Observations	178	178	50	50	50
R^2	0.237	0.438	0.271	0.328	0.357

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Dependent variable is percent of population with access to an improved water source in 2007. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders.

	(1)	(2)	(3)	(4)
FRAC	-31.90**	-20.19	21.98	6.153
	(11.71)	(14.22)	(37.06)	(30.82)
Observations	40	39	40	39
R^2	0.169	0.427	.	0.360
FirstStageF			4.999	5.809

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: Dependent variable is percent of population with access to improved sanitation facilities in 2007. Estimated using Ordinary Least Squares.

	(1)	(2)	(3)
FRAC	-55.13***	-12.40	-14.66
	(7.889)	(10.29)	(10.23)
Observations	178	140	140
R^2	0.226	0.619	0.682

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: Dependent variable is percent of population with access to improved sanitation facilities in 2007. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012).

	(1)	(2)	(3)	(4)	(5)
FRAC	-55.50***	-20.43*	-55.84***	14.75	14.03
	(7.875)	(8.918)	(14.43)	(33.38)	(31.28)
Observations	177	177	51	51	51
R^2	0.230	0.481	0.273	0.216	0.010

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: Dependent variable is percent of population with access to improved sanitation facilities in 2007. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders.

	(1)	(2)	(3)	(4)
FRAC	-62.15***	-27.30*	-31.47	-25.68
	(15.61)	(13.47)	(19.39)	(19.07)
Observations	41	40	41	40
R^2	0.287	0.721	0.217	0.721
FirstStageF			5.145	6.769

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.2.4 Road Density

Table 23: Dependent variable is kilometers of roads per thousand people in 1996. Estimated using Ordinary Least Squares.

	(1)	(2)	(3)
FRAC	-8.586**	-3.804	-1.516
	(2.980)	(4.433)	(3.619)
Observations	124	100	100
R^2	0.079	0.171	0.378

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 24: Dependent variable is kilometers of roads per thousand people in 1996. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012).

	(1)	(2)	(3)	(4)	(5)
FRAC	-8.586**	-8.064*	1.460	22.93	17.94
	(2.980)	(3.634)	(0.939)	(14.65)	(11.94)
Observations	124	124	40	40	40
R^2	0.079	0.116	0.004	.	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 25: Dependent variable is kilometers of roads per thousand people in 1996. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders.

	(1)	(2)	(3)	(4)
FRAC	-0.0105	3.080	-6.480	-5.203
	(2.767)	(4.787)	(7.886)	(9.791)
Observations	30	30	30	30
R^2	0.000	0.224	.	0.181
FirstStageF			17.02	18.64

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.2.5 Measles Immunization

Table 26: Dependent variable is percent of children ages 12-23 months immunized against measles in 2011. Estimated using Ordinary Least Squares.

	(1)	(2)	(3)
FRAC	-15.40***	-12.85**	-11.98*
	(3.808)	(4.692)	(5.017)
Observations	184	142	142
R^2	0.102	0.289	0.346

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 27: Dependent variable is percent of children ages 12-23 months immunized against measles in 2011. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012).

	(1)	(2)	(3)	(4)	(5)
FRAC	-15.48***	-4.035	-29.43***	-0.765	-1.338
	(3.811)	(4.075)	(6.355)	(17.03)	(16.52)
Observations	183	183	52	52	52
R^2	0.103	0.226	0.230	.	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 28: Dependent variable is percent of children ages 12-23 months immunized against measles in 2011. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders.

	(1)	(2)	(3)	(4)
FRAC	-39.58***	-27.87***	2.335	21.45
	(6.243)	(6.390)	(34.50)	(42.20)
Observations	41	40	41	40
R^2	0.290	0.435	.	0.127
FirstStageF			5.145	6.769

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.2.6 Schooling

Table 29: Dependent variable is logarithm of (1 + average years of school attainment) in 2010. Estimated using Ordinary Least Squares.

	(1)	(2)	(3)
FRAC	-0.702***	-0.0738	-0.0795
	(0.106)	(0.124)	(0.131)
Observations	141	124	124
R^2	0.211	0.505	0.655

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 30: Dependent variable is logarithm of (1 + average years of school attainment) in 2010. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012).

	(1)	(2)	(3)	(4)	(5)
FRAC	-0.702***	-0.365***	-0.372*	0.878	0.663
	(0.106)	(0.106)	(0.177)	(0.534)	(0.474)
Observations	141	141	37	37	37
R^2	0.211	0.411	0.057	.	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 31: Dependent variable is logarithm of (1 + average years of school attainment) in 2010. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders.

	(1)	(2)	(3)	(4)
FRAC	-0.272	-0.00748	-0.0944	0.139
	(0.238)	(0.315)	(0.454)	(0.469)
Observations	32	32	32	32
R^2	0.024	0.248	0.014	0.244
FirstStageF			37.53	42.24

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7.3 Growth Regressions with Controls for Initial Income

Table 32: Dependent variable is growth of GDP per capita. Estimated using Ordinary Least Squares. Standard errors clustered by country. Includes control for log of initial GDP per capita.

	(1)	(2)	(3)
FRAC	-0.0330***	-0.0198*	-0.0180*
	(0.00629)	(0.00855)	(0.00694)
LN(GDP PER CAPITA)	-0.000744	-0.00354*	-0.00578*
	(0.00122)	(0.00178)	(0.00229)
ABS LATITUDE		0.000138	-0.000576
		(0.000120)	(0.000366)
DIST COAST		-0.00000561	-0.0000126*
		(0.00000480)	(0.00000528)
AFRICA		-0.0118*	-0.0102
		(0.00462)	(0.00599)
SD ELEVATION			0.00790
			(0.00720)
SD AGRIC SUIT			-0.00562
			(0.0106)
ELEVATION			-0.0118
			(0.00850)
AGRIC SUIT			-0.0101
			(0.00631)
PRECIPITATION			-0.000156***
			(0.0000442)
TEMPERATURE			-0.000943
			(0.000708)
MIG DIST ETHIOPIA			-0.000619
			(0.000547)
EUROPE			0.0122*
			(0.00582)
AMERICAS			0.0216*
			(0.00942)
EAST ASIA PACIFIC			0.0279***
			(0.00576)
Constant	0.0390**	0.0577***	0.135**
	(0.0128)	(0.0157)	(0.0414)
Observations	6161	4910	4910
R^2	0.008	0.011	0.019

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 33: Dependent variable is growth of GDP per capita. SLAVE INTENSITY instrumented for using distance from slave ports as in Nunn (2008). Data from Nunn and Puga (2012). Standard errors clustered by country. Includes control for log of initial GDP per capita.

	(1)	(2)	(3)	(4)	(5)
FRAC	-0.0330***	-0.0289***	-0.0456**	-0.0295	-0.0391
	(0.00629)	(0.00765)	(0.0154)	(0.0268)	(0.0270)
LN(GDP PER CAPITA)	-0.000715	-0.00332*	-0.00309	-0.00490	-0.00763
	(0.00122)	(0.00138)	(0.00324)	(0.00424)	(0.00595)
RUGGEDNESS		-0.00110			-0.00848
		(0.00153)			(0.00559)
RUGGED*AFRICA		-0.000528			
		(0.00273)			
AFRICA		-0.0123*			
		(0.00613)			
SLAVE INTENSITY				-0.00205	-0.00381
				(0.00328)	(0.00448)
Constant	0.0387**	0.0642***	0.0593	0.0707	0.112*
	(0.0128)	(0.0149)	(0.0303)	(0.0364)	(0.0561)
Observations	6140	6140	1927	1927	1927
R^2	0.008	0.010	0.011	0.010	0.013

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 34: Dependent variable is growth of GDP per capita. First two columns estimated using OLS on the IV sample for the purpose of comparison. All other columns estimated using an instrument for ELF based on the randomness of African borders. Standard errors clustered by country. Includes control for log of initial GDP per capita.

	(1)	(2)	(3)	(4)
FRAC	-0.0469**	-0.0382*	-0.0189	-0.000367
	(0.0158)	(0.0157)	(0.0181)	(0.0146)
LN(GDP PER CAPITA)	-0.00254	-0.00648	-0.000762	-0.00559
	(0.00371)	(0.00427)	(0.00344)	(0.00435)
ABS LATITUDE		0.000720		0.000837
		(0.000494)		(0.000582)
AFRICA - NORTH		0.00719		0.0201
		(0.0125)		(0.0149)
AFRICA - SOUTH		0.00891		0.0149
		(0.00860)		(0.0108)
AFRICA - EAST		0.00390		0.00716
		(0.00654)		(0.00723)
AFRICA - CENTRAL		0.0162		0.0132
		(0.00853)		(0.00909)
Constant	0.0563	0.0634	0.0243	0.0276
	(0.0317)	(0.0315)	(0.0323)	(0.0345)
Observations	1517	1517	1517	1517
R^2	0.008	0.013	0.005	0.010
FirstStageF			28.33	31.45

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$