Journal of Economic Growth, 8, 155–194, 2003

Fractionalization

ALBERTO ALESINA

Department of Economics, Harvard University, Cambridge MA 02138, USA

ARNAUD DEVLEESCHAUWER

Department of Economics, Harvard University, Cambridge MA 02138, USA

WILLIAM EASTERLY

New York University and Center for Global Development, 269 Mercer Street, New York NY 10003, USA

SERGIO KURLAT

The World Bank, 1818 H Street, N.W., Washington DC 20433, USA

ROMAIN WACZIARG

Stanford Graduate School of Business, 518 Memorial Way, Stanford CA 94305, USA

We provide new measures of ethnic, linguistic, and religious fractionalization for about 190 countries. These measures are more comprehensive than those previously used in the economics literature and we compare our new variables with those previously used. We also revisit the question of the effects of ethnic, linguistic, and religious heterogeneity on the quality of institutions and growth. We partly confirm and partly modify previous results. The patterns of cross-correlations between potential explanatory variables and their different degree of endogeneity makes it hard to make unqualified statements about competing explanations for economic growth and the quality of government. Our new data, which features the underlying group structure of ethnicities, religions and languages, also allows the computation of alternative measures of heterogeneity, and we turn to measures of polarization as an alternative to the commonly used index of fractionalization.

Keywords: ethnic heterogeneity, growth, government quality

JEL classification: O5, H1

1. Introduction

Ethnic conflict is an important determinant of the political economy of many nations and localities. Many believe that it leads to political instability, poor quality of institutions, badly designed economic policy, and disappointing economic performance.

*Arnaud Devleeschauwer is a FNRS-Bernheim Research Fellow. We thank Francesco Caselli, James Fearon, Oded Galor and two anonymous referees for useful comments. Jessica Seddon Wallack provided excellent research assistance. All remaining errors are ours.

In a cross-country setting, Easterly and Levine (1997) have shown that per capita GDP growth is inversely related to ethnolinguistic fractionalization in a large sample of countries. In particular, they argued that much of Africa's growth failure is due to ethnic conflict, partly as a result of absurd borders left by former colonizers.[1] As a result of that paper, a measure of ethnic fractionalization has become a ''standard'' control in regressions explaining cross-national differences in economic success.[2]

A related literature, early examples being Canning and Fay (1993) and Mauro (1995), has discussed the impact of ethnic fragmentation on government activities and quality of institutions. La Porta et al. (1999), in a broad empirical study of the determinants of the quality of government, suggest that ethnic fractionalization matters, even though variables related to legal origins may be more important. A large literature on US localities show that in more ethnically fragmented communities, public goods provision is less efficient, participation in social activities and trust is lower, and economic success, measured by growth of city size, is inferior.[3] Evidence that trust does not travel well across racial lines is also supported by experimental evidence.[4]

Another related literature, Esteban and Ray (1994) on the theoretical side and Garcia-Montalvo and Reynal-Querol (2002) on the empirical side, discuss which are the best measures of heterogeneity. The traditional measure of ethnic fractionalization is given by the probability that two randomly drawn individuals from the population belong to two different groups. Its theoretical maximum is reached (at the value of 1) when each person belongs to a different group. In contrast, simple measures of polarization reach their maximum when two equally sized groups face each other. We use both measures of fractionalization and of polarization in our empirical work, and discuss how results differ across the two sets of indices.

While existing measure of racial (or ethnic) fragmentation for the United States are reasonably well-accepted, since they are based upon detailed and reliable census data, cross-country measures have been widely debated. Easterly and Levine (1997) use indices based on ethnolinguistic classification provided by sources from the former Soviet Union, the *Atlas Narodov Mira* of 1964. These data rely largely on linguistic distinctions, which may obscure other aspect of ethnicity like racial origin, skin color, etc. Interestingly, studies within the United States do not look at language in the racial classification. If they did, blacks and whites would be classified in the same group. As we discuss below, this example shows that although useful, language is not the only way to look at ethnicity.[5] In Latin America several countries are relatively homogeneous in terms of language spoken, often the one of former colonizers, but much less so if skin color or racial origin is taken into account. The World Bank estimates that the percentage of Afro-Latinos in Latin America is higher than the percentage of African–Americans in the United States. Peoples of indigenous or mestizo background also form a large percentage of the population in most Latin American countries.

Another difficulty in measuring heterogeneity is that ethnic classifications are not set in stone, and are more complex than can be summarized by simple measures. For example, the Oromo in Ethiopia are split into five different groups as a result of regional migrations and intermixing with other groups, suggesting that fractionalization evolves endogenously as a function of migration and intergroup mixing. The infamous Hutu–Tutsi divide in Burundi was thought by some to have been greatly accentuated (some even say created) by

Belgian colonizers, suggesting that fractionalization may also be endogenous because the definition of groups can shift overtime. People of African origin do not have as clear a dividing line from the rest of the population in many Latin American countries as they do in the United States, suggesting that ethnic differences may not be sufficient to fully characterize the degree of heterogeneity. Hence, ethnic classifications are fraught with ambiguities, as we discuss in depth below. Having mentioned this important caveat, our measures of ethnic, linguistic, and religious heterogeneity capture distinctions that may still matter enormously for economic outcomes.

This paper seeks to achieve four goals. First, we provide a new measure of ethnic fragmentation based on a broader classification of groups, taking into account not only language but also other cleavages such as racial characteristics. We provide this measure for many more countries (almost twice as many) than those normally used in the literature, using different sources, and we discuss in detail similarities and differences between our measure and previous ones. We construct three new indices, one based on a broad measure of ethnicity, one based strictly on language, and one based on religion. Another advantage of our new data is that we identify each ethnic, religious, and linguistic group for each country in our sample, allowing us to compute alternative measures of heterogeneity.

Second, we show that indices of fractionalization constructed using measures of ethnicity, language or religion lead to substantially different results when they are entered in regressions to explain growth and government quality.

Third, using our new measures we reexamine the evidence on the effects of ethnic fragmentation on two general areas: economic growth and the quality of institutions and policy. We reach interesting results:

a. On economic growth, we broadly confirm the results by Easterly and Levine (1997). In fact the negative effect of ethnic fragmentation on growth is reinforced with the new data, and we are able to highlight the differences between ethnic, linguistic, and religious fractionalization.

b. On quality of government and policies we make some progress over La Porta et al. (1999). They argued that both legal origin, distance from the equator and ethnolinguistic fractionalization explain the quality of government. In their results, legal origin variables tend to be stronger than ethnolinguistic fractionalization. We argue that results on this point are sensitive to the specification, and one can easily produce reasonable specifications in which ethnic fragmentation ''dominates'' legal origin. We do not intend to argue that ethnic fractionalization ''beats'' legal origin, but more modestly that the pattern of correlation between independent variables makes it very hard to resolve this horse race. Most likely both set of variables are important, and we discuss carefully the patterns of cross-correlation between these variables and the potential channels linking fractionalization to government quality.

c. Ethnic fractionalization is also closely correlated with GDP per capita and geographic variables, like latitude. Ethnic fragmentation is higher in poorer countries that are closer to the equator. This complicates even more the task of

characterizing the role of ethnic fragmentation as a determinant of policy variables, the quality of government and growth. Thus the pattern of cross-correlations between explanatory variables cannot be ignored when drawing conclusions on these issues. As is well known, in many cases the results of cross-country regressions are sensitive to the econometric specification, and this case is no exception. Useful lessons can be learned from this sensitivity, however, as it may inform us as to the channels whereby fractionalization operated to depress growth or reduce the quality of government.

d. While ethnic and linguistic fractionalization are associated with negative outcomes in terms of the quality of government, religious fractionalization is not; in fact, if anything, this measure displays a positive correlation with measures of good governance. This is because measured religious fractionalization tends to be higher in more tolerant and free societies, like the United States, which in fact displays one the of the highest level of religious fractionalization. This result has no bearing, however, on the question of whether specific religious denominations are correlated with better politico-economic outcomes, an issue recently explored by Barro and McLeary (2002).

Finally we explore which indicator of heterogeneity is more correlated with variables of interest. We conclude that the measure of fractionalization traditionally used in the literature performs a bit better than measures of polarization proposed by Garcia-Montalvo and Reynal-Querol (2002).

The paper is organized as follows. In Section 2, we present our new data and new indices of ethnic fractionalization. In Section 3, we present evidence on the relationship between fractionalization and growth in a broad cross-section of countries. In Section 4, we examine how fractionalization relates to the quality of government and institutions. Section 5 summarizes our results using measures of polarization. Section 6 discusses the impact of ethnic fractionalization on economic variables in individual countries. Section 7 concludes.

2. New Measures of Fractionalization[6]

2.1. The Data

Our main goal in gathering data on fractionalization is to clearly distinguish between ethnic, religious, and linguistic heterogeneity. Ethnic and linguistic differences were previously lumped together as part of an ''ethnolinguistic'' fractionalization variable. The data most frequently used in the literature was compiled in the Soviet Union in the early 1960s on the basis of primary country sources, and published in the *Atlas Narodov Mira* in 1964. The ethnolinguistic fractionalization variable (often referred to as ELF) was computed as one minus the Herfindahl index of ethnolinguistic group shares, and reflected the probability that two randomly selected individuals from a population belonged to

different groups.[7] We use the same formula, applied to different underlying data, to compute our measures of fractionalization:

$$\text{FRACT}_j = 1 - \sum_{i=1}^{N} s_{ij}^2, \tag{1}$$

where s_{ij} is the share of group i $(i = 1 \ldots N)$ in country j. Below, we consider alternative measures of heterogeneity, based on the concept of polarization.

A major obstacle to distinguishing between ethnic and linguistic variables is that language is part of the criterion used by ethnologists and anthropologists to define the concept of ethnicity. This is true, for example, in Africa, where racial or physical criteria are seldom used to define ethnic groups. This is not the case, however, in Latin America, where characteristics typically used to distinguish between ethnic groups are racial in nature. To our knowledge, no measures of racial fragmentation exist for a broad cross-section of countries, largely because the underlying data on group size is missing for most countries. Moreover, gathering such data would be fraught with conceptual problems, such as the definition of the physiological characteristics that distinguish races.

One feasible improvement over existing measures, however, is to compile a separate variable for linguistic fractionalization in isolation of any racial of physical characteristics. Our variable ''language'' is based exclusively on data from *Encyclopedia Britannica* (2001), which reports the shares of languages spoken as ''mother tongues,'' generally based on national census data. Other possible sources for language data include the *CIA World Factbook* (which, however, only lists the shares of each language for a few countries) and the Ethnologue project, which lists approximately 6,800 languages.[8] Fractionalization measures constructed from these sources are closely related, as they are based on very similar country source data.[9] Our data includes 1,055 major linguistic groups for 201 countries or dependencies.

We also compute a separate variable for religious fractionalization (''religion''), based on data from the *Encyclopedia Britannica* (2001). The distinctions in this data are perhaps less controversial and subject to arbitrary definitions than the data on linguistic and ethnic fractionalization, since the boundaries of religions are more clear and definitions consistent across countries. Our data cover 294 different religions in 215 countries and dependencies.

Finally, the main variable we focus on is a measure of ethnic fractionalization, ''ethnicity.'' As suggested above, the definition of ethnicity involves a combination of racial and linguistic characteristics. For example, our data on Bolivia involves the following groups: Blancos (10.13 percent), Aymara (30.38 percent), Quechua (30.38 percent), Mestizos (25.32 percent) and others groups (indigenous and Afro, 3.80 percent). This, like the data for most of the rest of Latin America and the Caribbean, is based on racial distinctions rather than linguistic distinctions. In fact, our language data for Bolivia looks very different: Aymara 3.24 percent, Guarani 0.12 percent, Quechua 8.15 percent, Spanish 87.65 percent, Other 0.84 percent.

In contrast, the ethnicity data for some European countries such as Belgium, Luxembourg, and Switzerland largely reflects languages (for example, the ''ethnicities'' we have identified in Switzerland include: German 65 percent, French 18 percent, Italian

10 percent, other Swiss 6 percent and Romansch 1 percent). The same holds for much of Sub-Saharan Africa. These classifications reflect the judgment of ethnologists and anthropologists on the appropriate definition of ethnicity, which to our knowledge remains a rather vague and amorphous concept. It would be wrong to interpret our ethnicity variable as reflecting racial characteristics alone, but it does reflect these characteristics to a greater extent than our language variable, and it should thus be expected to bear a different relationship to economic variables.

An important goal of our collection of ethnicity data was to obtain data on various ethnic groups that was as disaggregated as we could find. This required the use of multiple sources of data, which we painstakingly checked against each other for consistency. The primary source was the *Encyclopedia Britannica* (2001), which was the source of our data in 124 of 190 countries. This was completed with data from the CIA (2000) for 25 countries, Levinson (1998) for 23 cases and Minority Rights Group International (1997) for 13 cases. For France, Israel, the United States, and New Zealand, we directly consulted the national censuses of these countries to come up with ethnicity data as disaggregated as available.[10] The rule we followed for data collection was as follows: if two or more sources for the index of ethnic fractionalization were identical to the third decimal point, we used these sources (this was generally recorded as data sourced from the *Encyclopedia Britannica*). If sources diverged in such a way that the index of fractionalization differed to the second decimal point, we used the source where reported ethnic groups covered the greatest share of the total population. If this was 100 percent in more than one source, we used the source with the most disaggregated data (i.e. the greatest number of reported ethnic groups). In the end, our ethnicity variable covers approximately 650 distinct ethnic groups in 190 countries and dependencies.

2.2. Endogeneity

An important issue to contend with is that of changes in the ethnic fractionalization index through time, i.e. the issue of endogeneity. This is important because our data is from recent sources (generally the early to mid-1990s). If there were major shifts in ethnic composition, using data from the end of our period to explain variables for the 1960–1995 period could lead to endogeneity bias.

Shifts in ethnic composition could stem from changes in the shares of each group or from changes in the definition of the various ethnic groups. Ethnic fractionalization indices are generally taken as exogenous in cross-country regressions, based on the fact that group shares are sufficiently stable that changes only have a minor impact on fractionalization measures. This seems a reasonable assumption at the 30 year horizon of the typical cross-country regression, even though this assumption may be less tenable for a much longer horizon. Think, for instance, of different fertility rates across ethnic groups.

Another problem could occur if the definitions of ethnic groups changed through time, as a function of economic or political variables. The possibility of such changes in definitions has been pointed out by the ''reflexive'' school in ethnology and sociology. According to the reflexive theory of ethnicity and nationality, the boundaries of ethnic groups are changing because individual's self-identification to groups can change as a

Table 1. Sample means of the fractionalization measures (excluding dependencies).

Variable	Number of Observations	Sample Mean
Religion	198	0.439
Ethnic	180	0.435
Language	185	0.385
ELF	112	0.418

result of social, economic or political forces, and ethnicity is therefore endogenous, especially at long horizons.[11] One recent example of this phenomenon is Somalia: prior to the 1991 civil war, this country appeared relatively homogeneous (85 percent Somalis), but during and after the civil war ''clans'' became the dominant dimension of ethnic cleavage. In other words, a political event led to the creation of a new dimension of ethnic cleavage, and self-identification to groups now reflect preexisting clans rather than the Somali ''ethnicity.''[12]

In general, it does not matter for our purposes whether ethnic differences reflect physical attributes of groups (skin color, facial features) or long-lasting social conventions (language, marriage within the group, cultural norms) or simple social definition (self-identification, identification by outsiders). When people persistently identify with a particular group, they form potential interest groups that can be manipulated by political leaders, who often choose to mobilize some coalition of ethnic groups (''us'') to the exclusion of others (''them''). Politicians can also mobilize support by singling out some groups for persecution, where hatred of the minority group is complementary to some policy the politician wishes to pursue (Glaeser, 2002).

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. While the example of Somalia is interesting, in our sample period such examples are rare and ethnic fractionalization displays tremendous time persistence. The problem of the endogeneity of our religious fragmentation variable is more serious. Repressive regimes, especially those with a religious bent, may make it difficult for individuals to be ''counted'' as members of the non-officially sanctioned religion. This phenomenon could introduce a spurious correlation between (lack of) political freedom and religious fragmentation.

2.3. Comparison with Existing Measures

We now compare our measures of linguistic, ethnic, and religious fractionalization with the index of ethnolinguistic fractionalization based on the Soviet data usually used in the literature. First, our indices are available for many more countries, between 190 and 215 compared to 112 for the Soviet index. Table 1 displays the sample means and number of observations for our new indices (excluding dependencies). Table 2 shows the pairwise correlations between these four indices. The Soviet sample is, with very few exceptions, a

Table 2. Pairwise correlations of the fractionalization measures.

	Religion	Ethnic	Language	ELF
Religion	1			
	(198)			
Ethnic	0.142	1		
	(180)	(180)		
Language	0.269	0.697	1	
	(185)	(171)	(185)	
ELF	0.372	0.759	0.878	1
	(111)	(110)	(108)	(112)

subsample of our own. Not surprisingly, the correlations between our ethnic and linguistic index and the Soviet index are fairly high (0.76 and 0.88, respectively). Instead, the religious fractionalization index bears a much lower correlation with the other three indices.

Our data gathering effort can also be related to recent attempts by other scholars to gather cross-country ethnic heterogeneity data. Annett (2001) presents an index of ethnolinguistic fractionalization closely related conceptually to the Soviet data, using data exclusively published in the *World Christian Encyclopedia* (Barrett, 1982), a source distinct from our own. He also presents data on religious fractionalization, but does not attempt to isolate linguistic fractionalization like we do. His data cover 150 countries (compared to 190 for our ethnicity variable and 215 for our religion variable). Perhaps reassuringly given the different sources, for the overlapping sample of countries the correlation between his ethnolinguistic fractionalization variable and our ethnicity variable is 88.85 percent. The correlation between his religious fractionalization variable and our own is 83.66 percent.

Even more recently, Fearon (2002) has gathered detailed data on ethnic groups for 160 countries, from sources that sometimes overlap with ours (he does not present data for religious and linguistic fractionalization). His data is slightly less disaggregated than ours (each country displays on average 5.11 groups in his dataset, versus 5.55 in ours), partly because he restricts attention to groups making up more than 1 percent of the population. These small differences do not greatly impact our respective measures of fractionalization: as Fearon reports (2002, p. 3), referring to our dataset, ''the descriptive statistics for their ethnic measure look broadly similar to those for the measure constructed here.''

Table 3 highlights differences across regions amongst our three indices and ELF. With the exception of East and South East Asia, our ethnic fractionalization index show more fractionalization than the Soviet index. Given the way it is constructed, this is not surprising. Particularly interesting is the case of Latin America, were our ethnic fractionalization index is on average much higher than ELF. This is because, in this region, many ethnically diverse groups (as captured by skin color), often speak the same language as former European colonizers, Spanish, English or Portuguese. So a classification based purely on language shows a much lower degree of fractionalization than one that includes racial characteristics. In fact our linguistic fractionalization index leads to an average of 0.16 versus an average of 0.42 for the ethnicity index. The Soviet index is closer to our

Table 3. Sample means by region.

	Sample Restricted to Countries Available in Soviet Data				Unrestricted Sample		
	ELF	Ethnic	Language	Religion	Ethnic	Language	Religion
Latin America and Carribean	0.265 (23)	0.418 (23)	0.159 (21)	0.367 (23)	0.405 (33)	0.179 (32)	0.442 (40)
Sub-Saharan Africa	0.651 (38)	0.711 (38)	0.689 (37)	0.560 (38)	0.658 (47)	0.625 (47)	0.496 (49)
Eastern and Central Europe	0.315 (2)	0.319 (2)	0.348 (2)	0.512 (2)	0.366 (20)	0.320 (20)	0.491 (20)
Western and Southern Europe	0.147 (17)	0.170 (16)	0.198 (16)	0.285 (16)	0.177 (18)	0.196 (17)	0.311 (20)
Middle East	0.244 (9)	0.431 (8)	0.304 (9)	0.294 (9)	0.453 (13)	0.330 (14)	0.346 (14)
East and South East Asia	0.462 (10)	0.365 (10)	0.460 (10)	0.460 (10)	0.306 (16)	0.353 (17)	0.457 (17)

Note: Number of observations in parentheses.

linguistic index. Note how Sub-Saharan Africa displays the highest index of fractionalization in every single column. Appendix A displays these figures country by country.

Restricting our attention to countries with more than one million inhabitants, according to our data the most ethnically diverse country in the world is Uganda, with a fractionalization index of 0.93. The 13 most ethnically diverse countries are all in Sub-Saharan Africa, followed by Yugoslavia and then seven more SubSaharan African countries. The least ethnically fractionalized countries are South Korea, Japan, and North Korea. Turning to linguistic fractionalization, the most diverse countries are again 18 Sub-Saharan African countries (note that the definition of ethnicity there largely overlaps with linguistic distinctions). They are followed by India, with a linguistic fractionalization index of 0.81. The least diverse countries are South Korea and North Korea, followed by Yemen. Finally, turning to religious fractionalization, the most diverse countries are South Africa, the United States, and Australia, and the least diverse Yemen, Somalia, Morocco, Turkey, and Algeria.

2.4. Additional Uses of the Data

Our data has the potential to free researchers from their exclusive reliance on the fractionalization index. This is because it provides the whole distribution of ethnic, linguistic, and religious groups, instead of just one arbitrary statistic. In contrast, the ELF index forced reliance on fractionalization. With our new data at hand, researchers can write down models, examine the implications of these models with respect to the appropriate measures of heterogeneity, easily calculate them from the data, and test their

models more directly. Thus, our data allows for a much more serious grounding of empirical work in theoretical models. For example, in Section 5, we examine measures of polarization, rather than fractionalization, since many models based on conflict suggest that measures of polarization are more appropriate to capture the intensity of disagreements across groups. Another novel feature is that our dataset contains group names, so users can actually identify the groups. For example, Liberia, one of the most ethnically fractionalized country in our sample, has 13 separate ethnic groups: the Kpelle (18.3 percent), Bassa (13.3 percent), Dan (8.3 percent), Grebo (7.5 percent), Kru (7.3 percent), Ma (7.2 percent), Manding/Vai (7.0 percent), Loma (6.0 percent), Americo-Liberians (5.0 percent), Krahn (4.7 percent), Gola (4.7 percent), Kissi (3.3 percent) and Gbandi (3.0 percent). In contrast, one of the least fractionalized country in our sample (South Korea) only displays two groups: Koreans (99.9 percent) and others (0.1 percent).

To give a sense of the possibilities offered by our new data, Table 4 presents statistics on the number of available ethnic groups for different geographic areas.[13] Counting entries in each country, our dataset has a total of 1,054 entries, corresponding to 650 distinct ethnic groups. The average number of groups per country is highest in Sub-Saharan Africa (7.61 groups per country), and lowest in Latin America (4.22). Sub-Saharan Africa only has one country containing a group that represents more than 90 percent of the total population (out of 44 countries) while 17 out of 28 industrialized countries (including countries in Europe, North America, plus Japan, New Zealand, and Australia) display this

Table 4. Data description by ethnic group and by geographical area.

	World	West	North Africa/ Middle East	Latin America/ Caribbean	Asia	Eastern Europe/Former Soviet Union	Sub-Saharan Africa
No. of countries	190	28	19	34	38	27	44
Total (fraction)		0.15	0.10	0.18	0.20	0.14	0.23
No. of groups	1,054	132	83	146	183	175	335
Total (fraction)		0.13	0.08	0.14	0.17	0.17	0.32
Groups/country	5.55	4.71	4.37	4.29	4.82	6.48	7.61
Max. no. of groups	20	9	8	8	20	12	13
Min. no. of groups	1	2	2	2	1	3	2
Avg. pop share of largest group	0.68	0.82	0.69	0.71	0.76	0.72	0.44
Avg. pop. Share of 2nd largest	0.16	0.09	0.19	0.18	0.14	0.15	0.19
No. of countries with a group $\geq 50\%$	141	25	16	27	34	25	14
Countries with a group $\geq 50\%$	0.13	0.19	0.19	0.18	0.19	0.14	0.04
No. countries with a group $\geq 90\%$	44	17	4	7	13	2	1
Countries with a group $\geq 90\%$	0.23	0.61	0.21	0.21	0.34	0.07	0.02

Note: West includes Australia, New Zealand and Japan, SSA includes Sudan. This table has the same structure as Table 1 in Fearon (2002), to facilitate comparisons.

characteristic. The average population share of the largest group is 44 percent in Sub-Saharan Africa and 82 percent in the industrialized countries.

3. Fractionalization and Growth

In this section we revisit the question of the relationship between fractionalization and long-run growth. For the sake of comparison, we closely follow the specification of Easterly and Levine (1997). We begin in Table 5 by showing the correlation between several economic variables of interest and our three measures of fractionalization: ethnic, linguistic, and religious. Our ethnic variable is highly negatively correlated with GDP per capita growth, schooling and telephones per capita. These correlations are slightly lower for the linguistic measure. The measure of religious fractionalization does not seem to bear any pattern of correlations with the above mentioned variables.

Table 6 is organized in exactly the same way as Easterly and Levine's (1997, Table 4). This table shows that our measure of ethnic fractionalization is inversely related to per capita growth, as shown in column 1. The next three columns show that as one controls for more and more variables, the effect of fractionalization vanishes. The point is that variables such as schooling, telephones per worker, etc., can be understood as channels through which the ethnic fractionalization variable affects growth. Table 7 highlights this by reproducing Table 6 of Easterly and Levine (1997). It shows that ethnic fractionalization is strongly negatively correlated with schooling, financial depth, fiscal surplus, and the log of telephones per worker (these results are the same as in Easterly and Levine except for the fiscal surplus, where Easterly and Levine did not find a significant association). This negative effect of racial fractionalization on infrastructure and

Table 5. Correlations between fractionalization, growth and its determinants.

	Ethnic	Language	Religion	Growth	Income	bmp	Assas	Schooling
Language	0.697	1						
	(171)	(185)						
Religion	0.142	0.269	1					
	(180)	(185)	(198)					
Growth	− 0.471	− 0.305	− 0.103	1				
	(119)	(115)	(119)	(120)				
Log initial	− 0.330	− 0.293	0.049	0.137	1			
income 1960	(118)	(114)	(118)	(119)	(119)			
Black market	0.102	0.096	− 0.041	− 0.260	− 0.277	1		
premium	(96)	(93)	(96)	(91)	(91)	(97)		
Assassinations	− 0.110	− 0.027	− 0.080	− 0.079	− 0.003	− 0.012	1	
	(90)	(89)	(91)	(87)	(87)	(79)	(92)	
Schooling	− 0.459	− 0.387	0.122	0.328	0.816	− 0.225	− 0.117	1
	(97)	(94)	(97)	(91)	(90)	(81)	(71)	(98)
Phones per capita	− 0.356	− 0.248	0.084	0.337	0.895	− 0.271	− 0.080	0.828
	(133)	(128)	(134)	(119)	(118)	(96)	(91)	(97)

Note: Number of observations in parentheses.

Table 6. Ethnic diversity and long-run growth (Dependent variable is growth of per capita real GDP).

Variable	(1)	(2)	(3)	(4)
Dummy for the 1960s	− 0.086	− 0.109	− 0.222	− 0.259
	(− 0.99)	(− 1.24)	(− 2.22)	(− 2.47)
Dummy for the 1970s	− 0.089	− 0.111	− 0.218	− 0.253
	(− 1.02)	(− 1.27)	(− 2.19)	(− 2.42)
Dummy for the 1980s	− 0.109	− 0.131	− 0.236	− 0.269
	(− 1.25)	(− 1.50)	(− 2.36)	(− 2.57)
Dummy variable for	− 0.008	− 0.009	− 0.011	− 0.015
SubSaharan Africa	(− 1.70)	(− 1.99)	(− 2.05)	(− 2.76)
Dummy variable for	− 0.018	− 0.017	− 0.013	− 0.015
Latin America and the Caribbean	(− 4.87)	(− 4.54)	(− 3.55)	(− 4.01)
Log of initial income	0.035	0.041	0.073	0.088
	(1.55)	(1.84)	(2.85)	(3.34)
Log of initial income squared	− 0.003	− 0.003	− 0.005	− 0.007
	(− 1.77)	(− 2.09)	(− 3.24)	(− 4.06)
Log of schooling	0.013	0.013	0.013	0.009
	(3.06)	(3.16)	(3.03)	(1.84)
Assassinations		− 24.728	− 17.654	− 22.55
		(− 2.42)	(− 1.86)	(− 2.46)
Financial depth			0.017	0.013
			(2.89)	(2.12)
Black market premium			− 0.020	− 0.020
			(− 4.14)	(− 4.14)
Fiscal surplus/GDP			0.101	0.163
			(3.06)	(4.26)
Log of telephones per worker				0.007
				(2.52)
Ethnic	**− 0.019**	**− 0.018**	**− 0.009**	**− 0.005**
	(− 2.97)	**(− 2.84)**	**(− 1.41)**	**(− 0.68)**
No. of observations	82; 88; 94	77; 87; 93	44; 71; 74	40; 69; 66
R^2	0.25; 0.22; 0.36	0.24; 0.22; 0.38	0.39; 0.45; 0.52	0.39; 0.51; 0. 58

Notes: *t*-statistics are in parentheses. Estimated using seemingly unrelated regressions: a separate regression for each 10-year period. See the Appendix for definitions and sources.

productive public goods will be discussed in more detail in Section 4. Since ethnic fractionalization affects variables that in turn affect growth, there is a reduced form relationship between these variables and growth. The partial association between growth and fractionalization vanishes once we control for the intermediating variables.

In terms of economic magnitudes, the results in Table 6 suggest that going from complete ethnic homogeneity (an index of 0) to complete heterogeneity (an index of 1) depresses annual growth by 1.9 percentage points (column 1). In other words, up to 1.77 percentage points of the difference in annual growth between South Korea and Uganda can be explained by different degrees of ethnic fractionalization. This effect is reduced as we control for variables that can be interpreted as channels through which ethnic fractionalization affects growth. However, since our regressions contain the log of initial income on the right-hand side, the regressors are to be interpreted as determinants of

Table 7. Ethnicity as a determinant of economic indicators.

Dependent Variable	C	Ethnic	R^2	Number of Observations
Log of schooling	1.963	− 1.394	0.19; 0.23; 0.17	94; 95; 102
	(26.85)	(− 9.83)		
Assassinations	9.79E − 06	6.47E − 06	− 0.01; − 0.06; − 0.02	99; 109; 109
	(1.07)	(0.38)		
Financial depth	0.465	− 0.353	0.22; 0.12; 0.03	95; 103; 106
	(12.42)	(− 5.03)		
Black market premium	0.178	0.104	− 0.01; 0.02; − 0.03	105; 119; 120
	(3.61)	(1.12)		
Fiscal surplus/GDP	− 0.022	− 0.020	− 0.08; − 0.01; − 0.06	56; 94; 100
	(− 4.42)	(− 2.13)		
Log of telephones per worker	4.982	− 3.909	0.26; 0.31; 0.13	98; 105; 95
	(20.72)	(− 9.29)		

Notes: *t*-statistics are in parentheses. Equations estimated using seemingly unrelated regression procedures.

steady-state income levels. Perhaps a better way to get a sense of the magnitude of our effect is to examine implications in terms of steady-state income differences rather than transitional growth rate. The equation in column (1) of Table 6 implies a steady-state level of income that will be 14 percent lower for every 0.10 increase in ethnic fractionalization. If Korea had Uganda's ethnic fractionalization, the income level differential between them would have been reduced by half.

In Tables 8 and 10 we rerun the same regressions as in Table 6, but using religious fractionalization and linguistic fractionalization. While linguistic fractionalization is strongly inversely related to growth, religious fractionalization is not. In fact, as Table 5 already showed religious fractionalization does not seem to be correlated with any of the other right-hand side variable. This contrasts with linguistic fractionalization, especially for telephones per workers and schooling, a result which is confirmed in Tables 9 and 11 and in Section 4. Overall our results are quite similar to those of Easterly and Levine (1997), perhaps even a little stronger when using our new measure of linguistic fractionalization.

The differences in the results between religious and linguistic and ethnic fractionalization are quite suggestive. Religious affiliation is the most endogenous of the three variables. Religions can be banned and individual can relatively easily ''hide'' their religious affiliation to avoid repression. Individuals and families can change from one religion to another far more easily than they can change race (!) or language. In a sense, a higher observed measure of religious fractionalization can be a sign of a more tolerant and democratic form of government. In a more repressive regime, you can hide your religion or conform to the state-imposed religion, but hiding your racial origin, especially if it relates to skin color, is much more difficult. Short of genocide, it is difficult to change the ethnic composition of a country. As early as 1830, Tocqueville (1990) had noted this problem with reference to slavery in America. He wrote that ''there is a natural prejudice that prompts men to despise whoever has been their inferior long after he has become their equal... But amongst the ancients this secondary consequence of slavery had a natural

Table 8. Language diversity and long-run growth (Dependent variable is growth of per capita real GDP).

Variable	(1)	(2)	(3)	(4)
Dummy for the 1960s	− 0.056	− 0.070	− 0.166	− 0.226
	(− 0.63).	(− 0.77).	(− 1.60).	(− 2.13)
Dummy for the 1970s	− 0.058	− 0.072	− 0.162	− 0.219
	(− 0.66)	(− 0.80)	(− 1.57)	(− 2.07)
Dummy for the 1980s	− 0.077	− 0.091	− 0.177	− 0.235
	(− 0.87)	(− 1.00)	(− 1.72)	(− 2.22)
Dummy variable for	− 0.009	− 0.010	− 0.011	− 0.014
SubSaharan Africa	(− 1.81)	(− 2.09)	(− 2.20)	(− 2.53)
Dummy variable for Latin	− 0.023	− 0.022	− 0.018	− 0.019
America and the Caribbean	(− 6.02)	(− 5.78)	(− 4.69)	(− 4.67)
Log of initial income	0.030	0.034	0.062	0.080
	(1.29)	(1.45)	(2.36)	(3.03)
Log of initial income squared	− 0.002	− 0.003	− 0.005	− 0.006
	(− 1.58)	(− 1.75)	(− 2.81)	(− 3.75)
Log of schooling	0.012	0.012	0.011	0.010
	(2.93)	(2.92)	(2.65)	(2.19)
Assassinations		− 18.254	− 10.126	− 16.068
		(− 1.30)	(− 0.76)	(− 1.23)
Financial depth			0.015	0.012
			(2.57)	(1.98)
Black market premium			− 0.023	− 0.020
			(− 4.64)	(− 4.16)
Fiscal surplus/GDP			0.088	0.162
			(2.68)	(4.26)
Log of telephones per worker				0.005
				(1.99)
Language	**− 0.025**	**− 0.024**	**− 0.020**	**− 0.013**
	(− 3.73)	**(− 3.59)**	**(− 3.03)**	**(− 1.85)**
No. of observations	80; 86; 92	75; 85; 91	43; 69; 73	39; 68; 65
R^2	0.24; 0.26; 0.30	0.23; 0.26; 0.31	0.42; 0.48; 0.49	0.42; 0.53; 0.57

Notes: *t*-statistics are in parentheses. Estimated using seemingly unrelated regressions: a separate regression for each period. See the Appendix for definitions and sources.

limit; for the freedman bore so entire a resemblance to those born free that it soon became impossible to distinguish him from them.'' In the United States, instead, skin color differences between blacks and whites makes assimilation more difficult. In other words, skin color becomes an important focal point to characterize lasting differences and perceptions, as also argued by Caselli and Coleman (2002).

4. Fractionalization and Government Quality

One of the reasons why ethnic fractionalization may negatively influence economic success in terms of growth and income levels has to do with the potentially negative effects of ethnic conflict on the quality of policy and institutions. In a sweeping empirical study La Porta et al. (1999) have investigated the determinants of the quality of

Table 9. Determinants of economic indicators.

Dependent Variable	C	Language	R^2	Number of Observations
Log of schooling	1.796	− 1.166	0.19; 0.19; 0.09	91; 92; 99
	(27.75)	(− 9.08)		
Assassinations	8.26E-06	7.44E-06	− 0.02; − 0.06; − 0.02	96; 107; 107
	(1.10)	(0.50)		
Financial depth	0.388	− 0.205	0.09; 0.04; − 0.06	92; 101; 104
	(11.46)	(− 3.01)		
Black market premium	0.194	0.074	− 0.01; 0.01; − 0.04	102; 117; 118
	(4.58)	(0.88)		
Fiscal surplus/GDP	− 0.027	− 0.010	− 0.09; − 0.02; − 0.10	55; 91; 98
	(− 6.40)	(− 1.07)		
Log of telephones per worker	4.453	− 3.118	0.23; 0.24; 0.03	95; 103; 93
	(21.31)	(− 8.05)		

Notes: *t*-statistics are in parentheses. Equations estimated using seemingly unrelated regression procedures.

government and policy outcomes looking and a large number of indicators of policy. They concluded that a country's legal origins are an important determinant of these variables, while the ethnic fractionalization variable (the same as used by Easterly and Levine, 1997) bore a reduced form relationship with government quality. However, fractionalization was typically not significant after controlling for the level of GDP per capita (which however could be endogenous) and latitude.

Table 12 reports a matrix of correlation between all the variables used as potential explanations for the quality of government. Note that our measures of linguistic and ethnic fractionalization are highly correlated with latitude and GDP per capita. Therefore it is quite difficult to disentangle the independent effect of these three variables on the quality of government. While GDP per capita is very likely to be endogenous to the left-hand side variables, so that it is unclear whether one should control for it or not, the other two variables are less endogenous. Also, ethnic fractionalization and latitude are less obviously linked by causal relationships than the same two variables are with income. The correlation between latitude and ethnic fractionalization is quite high, about 0.4. This makes it hard to disentangle the effect of one variable from the other and the result in this type of cross-sectional regressions will depend on the specification. On a priori grounds, while one can think of several reasons why ethnic conflict may affect policy outcomes and institutions, the relationship between latitude and, say, the regulation of economic activity or the protection of property rights seems much less obvious.

The measure of religious fragmentation displays a much lower level of correlation with GDP per capita; in fact this correlation is basically zero. Our ethnic fractionalization variable displays a positive correlation (0.2) with the dummy variables for French legal origins, which according to La Porta et al. (1999) is associated with poor quality of government. This does not help in separating the effects of legal origins from those of fractionalization.

In Table 13 we run a set of regressions along the lines of La Porta et al. (1999). Table 13 is organized as follows. For each left-hand side variable, we present the coefficient on

Table 10. Religious diversity and long-run growth (Dependent variable is growth of per capita real GDP).

Variable	(1)	(2)	(3)	(4)
Dummy for the 1960s	− 0.108	− 0.138	− 0.273	− 0.307
	(− 1.19)	(− 1.51)	(− 2.67)	(− 3.00)
Dummy for the 1970s	− 0.111	− 0.140	− 0.269	− 0.300
	(− 1.22)	(− 1.53)	(− 2.64)	(− 2.94)
Dummy for the 1980s	− 0.131	− 0.160	− 0.285	− 0.316
	(− 1.45)	(− 1.75)	(− 2.80)	(− 3.10)
Dummy variable for	− 0.014	− 0.015	− 0.017	− 0.019
SubSaharan Africa	(− 2.68)	(− 2.98)	(− 3.14)	(− 3.30)
Dummy variable for	− 0.021	− 0.020	− 0.015	− 0.016
Latin America and the Caribbean	(− 5.53)	(− 5.20)	(− 4.11)	(− 4.37)
Log of initial income	0.039	0.047	0.086	0.100
	(1.65)	(1.99)	(3.26)	(3.87)
Log of initial income squared	− 0.003	− 0.003	− 0.006	− 0.008
	(− 1.82)	(− 2.19)	(− 3.61)	(− 4.66)
Log of schooling	0.013	0.013	0.010	0.008
	(2.92)	(2.96)	(2.37)	(1.68)
Assassinations		− 23.630	− 18.235	− 22.956
		(− 2.22)	(− 1.84)	(− 2.49)
Financial depth			0.018	0.012
			(3.05)	(2.11)
Black market premium			− 0.022	− 0.021
			(− 4.48)	(− 4.20)
Fiscal surplus/GDP			0.089	0.172
			(2.76)	(4.58)
Log of telephones per worker				0.007
				(2.88)
Religion	**− 0.004**	**− 0.002**	**0.006**	**0.008**
	(− 0.52)	**(− 0.24)**	**(0.92)**	**(1.16)**
No. of observations	82; 88; 95	77; 87; 94	44; 71; 75	40; 69; 66
R^2	0.20; 0.18; 0.32	0.20; 0.18; 0.34	0.43; 0.44; 0.49	0.43; 0.51; 0.58

Notes: *t*-statistics are in parentheses. Estimated using seemingly unrelated regressions: a separate regression for each period. See the Appendix for definitions and sources.

fractionalization from three regressions.[14] The first one reproduces exactly the full specification of La Porta et al. (1999), including their specification which include the largest number of independent variables, i.e. legal origins, religious variables, latitude, etc. To these variables we have added our measure of ethnic fractionalization. Column 2 present a minimalist specification, which includes only country size and regional dummies. Column 3 adds to this specification income per capita and legal origins variables. For brevity we do not report another column including also the religious variables, but the results (available upon request) are similar to those of column 3. Note that the omitted legal origins variable is the British one. Tables 14 and 15 replicate these regressions with, respectively, the measures of linguistic and religious fractionalization. Several observations are in order.

1. Our index of ethnic fractionalization is significant in the ''minimalist'' regression, column 2, for corruption, bureaucratic delays, infrastructure quality, infant mortality,

Table 11. Religion as a determinant of economic indicators.

Dependent Variable	C	Religion	R^2	Number of Observations
Log of schooling	1.160	0.358	0.01; −0.01; −0.14	94; 95; 103
	(11.99)	(1.91)		
Assassinations	1.77E-05	−1.13E-05	−0.01; −0.06; −0.02	99: 110; 110
	(1.93)	(−0.61)		
Financial depth	0.292	0.012	−0.01; −0.04; −0.17	95; 104; 107
	(7.06)	(0.15)		
Black market premium	0.222	0.004	−0.01; 0.00; −0.05	105; 120; 121
	(4.29)	(0.04)		
Fiscal surplus/GDP	−0.027	−0.008	−0.14; −0.02; −0.08	56; 95; 101
	(−5.25)	(−0.78)		
Log of telephones per worker	2.759	0.321	0.00; −0.12; −0.45	98; 105; 95
	(9.77)	(0.59)		

Notes: *t*-statistics are in parentheses. Equations estimated using seemingly unrelated regression procedures.

Table 12. Correlations of fractionalization measures and the determinants of the quality of government.

	log gnp pc	Latitude	leg_or uk	leg_or soc	leg_or fr	leg_or ger	leg_or scan	Ethnic	Language
Latitude	0.5314								
	(185)								
leg_or uk	−0.0960	−0.2758							
	(184)	(205)							
leg_or soc	−0.0193	0.4426	−0.3223						
	(184)	(205)	(212)						
leg_or fr	−0.1651	−0.2429	−0.6345	−0.3894					
	(184)	(205)	(212)	(212)					
leg_or ger	0.2687	0.1745	−0.1339	−0.0822	−0.1618				
	(184)	(205)	(212)	(212)	(212)				
leg_or scan	0.2817	0.3382	−0.1126	−0.0691	−0.1361	−0.0287			
	(184)	(205)	(212)	(212)	(212)	(212)			
Ethnic	−0.3929	−0.3816	0.0144	−0.1104	0.2085	−0.1561	−0.2324		
	(173)	(183)	(185)	(185)	(185)	(185)	(185)		
Language	−0.3639	−0.2679	0.1483	−0.0741	0.0140	−0.1157	−0.1629	0.6981	
	(174)	(193)	(191)	(191)	(191)	(191)	(191)	(176)	
Religion	0.0269	−0.1138	0.3632	0.0433	−0.3656	0.1012	−0.1481	0.1520	0.2718
	(183)	(205)	(204)	(204)	(204)	(204)	(204)	(185)	(195)

illiteracy, and school attainment. It is significant or nearly significant in column 3 (controlling for GDP per capita) for corruption, infant mortality, and illiteracy. The sign of the coefficient always implies that more fractionalization leads to a lower quality of government. This index is also negatively associated with the share of transfers over GDP, a result consistent with those obtained by Alesina et al. (2001) on a much smaller sample of countries, and by Alesina and Wacziarg (1998) on a large sample of countries but with different data on government spending. It seems that

Table 13. Ethnic fractionalization and the quality of government (only the coefficients on ethnic fractionalization are reported).

Dependent Variable	No. of Obs.	(1)	(2)	(3)
1. Business climate				
Property rights index	141	− 0.028	− 0.573	− 0.262
		(0.089)	(1.189)	(0.676)
		0.582	0.140	0.564
Business regulation index	141	− 0.429	− 0.343	− 0.382
		(1.465)	(0.954)	(1.239)
		0.494	0.196	0.489
2. Corruption and bureaucratic quality				
Corruption	121	1.011	− 2.487**	− 1.317*
		(1.332)	(2.374)	(1.704)
		0.540	0.252	0.517
Bureaucratic delays	59	− 0.896	− 1.969**	− 1.023
		(1.635)	(2.235)	(1.460)
		0.734	0.179	0.671
3. Taxation				
Tax compliance	49	− 0.585	− 0.024	− 0.342
		(1.049)	(0.038)	(0.606)
		0.530	0.127	0.507
Top marginal tax rate	82	10.369	− 3.155	3.260
		(1.495)	(0.509)	(0.445)
		0.202	0.360	0.414
4. Size of the public sector				
SOEs in the economy	103	− 1.815*	− 1.539	− 1.480
		(1.778)	(1.562)	(1.517)
		0.144	0.155	0.264
Public sector employment/ Total population	116	0.017	− 1.367	0.422
		(0.021)	(1.019)	(0.477)
		0.709	0.385	0.721
5. Size of government				
Government consumption/GDP	103	2.935	1.323	2.790
		(1.521)	(0.663)	(1.471)
		0.194	0.250	0.310
Transfers and subsidies/GDP	89	− 0.498	− 7.360**	− 4.984*
		(0.179)	(2.502)	(1.981)
		0.694	0.598	0.724
6. Public goods				
Infrastructure quality	59	− 0.623	− 2.019*	− 0.726
		(1.131)	(1.704)	(0.924)
		0.828	0.169	0.775
Log infant mortality	166	0.442***	1.075***	0.665***
		(3.436)	(4.065)	(3.966)
		0.842	0.481	0.806

Table 13. Continued.

Dependent Variable	No. of Obs.	(1)	(2)	(3)
7. Schooling and literacy				
Illiteracy rate	117	8.991	15.820**	14.090***
		(1.654)	(2.233)	(2.634)
		0.666	0.436	0.636
Log school attainment	101	− 0.056	− 0.568**	− 0.045
		(0.445)	(2.246)	(0.361)
		0.781	0.386	0.779
8. Political rights				
Democracy index	147	− 1.053	− 4.238***	− 2.278*
		(0.951)	(2.906)	(1.797)
		0.545	0.175	0.448
Political rights index	167	− 0.687	− 3.108***	− 2.378***
		(1.150)	(4.148)	(3.135)
		0.518	0.189	0.291

*: Significant at the 10% level.
**: Significant at the 5% level.
***: Significant at the 1% level.
Robust *t*-statistics are in parentheses, adjusted *R*-squared are reported underneath.

Specification (1) includes the log(GNP) for 1970–95, legal origin dummies (Socialist, French, German and Scandinavian), religion variables (Catholic, Muslim and other) and latitude.

Specification (2) includes the log of population in 1960 and regional dummies (SubSaharan Africa, East Asia and Latin America).

Specification (3) includes the log(GNP) for 1970–95, the log of population in 1960, regional dummies (SubSaharan Africa, East Asia and Latin America) and legal origin dummies (Socialist, French, German and Scandinavian). All specifications include a constant.

governments have a much more difficult task achieving consensus for redistribution to the needy in a fractionalized society.

2. The democracy index is inversely related to ethnic fractionalization (when latitude is not controlled for). This result is consistent with theory and evidence presented in Aghion et al. (2002). The idea is that in more fragmented societies a group imposes restrictions on political liberty to impose control on the other groups. In more homogeneous societies, it is easier to rule more democratically since conflicts are less intense.[15]

3. Overall the index of linguistic fractionalization seems to work less well than the index based on ethnicity, in the sense of leading to coefficients that are less robust to changes of specification and more often statistically insignificant.

4. The index of religious fractionalization bears a positive relationship to controlling corruption, preventing bureaucratic delays, tax compliance, transfers, infrastructure quality, lower infant mortality, lower illiteracy, school attainment, democracy, and political rights. Note that this result holds regardless of whether the size of various

Table 14. Linguistic fractionalization and the quality of government (only the coefficients on linguistic fractionalization are reported).

Dependent Variable	Obs.	(1)	(2)	(3)
1. Business climate				
Property rights index	138	0.170	0.139	0.187
		(0.649)	(0.331)	(0.582)
		0.591	0.115	0.569
Business regulation index	138	− 0.099	0.098	− 0.049
		(0.442)	(0.317)	(0.194)
		0.472	0.176	0.470
2. Corruption and bureaucratic quality				
Corruption	120	0.016	− 2.082**	− 1.760**
		(0.024)	(2.387)	(2.595)
		0.511	0.233	0.501
Bureaucratic delays	58	0.085	− 0.707	0.261
		(0.177)	(0.777)	(0.398)
		0.727	0.092	0.666
3. Taxation				
Tax compliance	48	− 0.368	0.042	− 0.248
		(0.750)	(0.070)	(0.454)
		0.491	0.085	0.475
Top marginal tax rate	81	15.744***	− 0.448	6.030
		(2.974)	(0.086)	(0.959)
		0.265	0.350	0.414
4. Size of the public sector				
SOEs in the economy	100	− 0.702	0.044	− 0.020
		(0.858)	(0.049)	(0.023)
		0.129	0.123	0.248
Public sector employment/ Total population	115	− 0.452	− 0.206	0.485
		(0.780)	(0.205)	(0.665)
		0.709	0.369	0.720
5. Size of government				
Government consumption/GDP	100	1.053	0.101	0.672
		(0.596)	(0.049)	(0.335)
		0.181	0.240	0.291
Transfers and subsidies/GDP	86	0.249	− 6.048**	− 2.949
		(0.117)	(2.203)	(1.334)
		0.691	0.586	0.718
6. Public goods				
Infrastructure quality	58	0.290	− 0.566	0.628
		(0.595)	(0.456)	(0.905)
		0.824	0.103	0.775
Log infant mortality	161	0.285**	0.529*	0.244
		(2.334)	(1.937)	(1.605)
		0.832	0.419	0.786

Table 14. Continued.

Dependent Variable	Obs.	(1)	(2)	(3)
7. Schooling and literacy				
Illiteracy rate	111	17.143***	17.727**	13.859**
		(3.590)	(2.221)	(2.060)
		0.693	0.443	0.639
Log school attainment	97	− 0.157	− 0.503*	0.083
		(1.419)	(1.986)	(0.585)
		0.779	0.378	0.774
8. Political rights				
Democracy index	145	− 0.681	− 1.256	− 0.162
		(0.791)	(0.962)	(0.159)
		0.543	0.119	0.432
Political rights index	162	− 0.873*	− 1.247*	− 0.715
		(1.857)	(1.673)	(1.003)
		0.529	0.101	0.235

*: Significant at the 10% level.
**: Significant at the 5% level.
***: Significant at the 1% level.
Robust *t*-statistics are in parentheses, adjusted *R*-squared are reported underneath.

Specification (1) includes the log(GNP) for 1970–1995, legal origin dummies (Socialist, French, German and Scandinavian), religion variables (Catholic, Muslim and other) and latitude.

Specification (2) includes the log of population in 1960 and regional dummies (SubSaharan Africa, East Asia and Latin America).

Specification (3) includes the log(GNP) for 1970–1995, the log of population in 1960, regional dummies (SubSaharan Africa, East Asia and Latin America) and legal origin dummies (Socialist, French, German and Scandinavian). All specifications include a constant.

religious denominations is held constant in the regressions or not. Our interpretation is that observed religious fragmentation is larger in more tolerant countries.

5. The index of ethnic fractionalization loses statistical significance in many of the regressions with the full specification used by of La Porta et al. (1999). This is because these regressions include latitude and, as we argued above, this variable is highly correlated with ethnic fractionalization. The ethnic fractionalization variable remains significant at standard levels even after controlling for latitude in the case of infant mortality and the share of state-owned enterprises. In virtually all other cases the ethnic fractionalization variable retains the ''expected'' sign but it is not statistically different from zero at standard levels of confidence. This reflects the difficulty in disentangling the effects of latitude, per capita income (which again may not belong in the regression due to endogeneity), and fractionalization. The high correlation between fractionalization and latitude, in particular, creates some problems of interpretation. Those who hold a prior that geography affects institutions directly or indirectly will find some support for their view in these results. However the same can be said for those who believe that conflicts amongst groups brings about more difficult and inefficient policy making and that ethnic fractionalization simply happens to be correlated with latitude.

Table 15. Religious fractionalization and the quality of government (Only the coefficients on religious fractionalization are reported).

Dependent Variable	Obs.	(1)	(2)	(3)
1. Business climate				
Property rights index	142	− 0.333	0.673*	− 0.327
		(0.965)	(1.751)	(0.926)
		0.589	0.140	0.569
Business regulation index	142	0.029	0.724**	0.098
		(0.095)	(2.535)	(0.380)
		0.487	0.217	0.485
2. Corruption and bureaucratic quality				
Corruption	122	0.382	1.544**	0.216
		(0.460)	(2.053)	(0.279)
		0.513	0.232	0.490
Bureaucratic delays	59	1.134**	3.190***	2.016***
		(2.389)	(4.550)	(4.242)
		0.736	0.333	0.718
3. Taxation				
Tax compliance	49	0.634	2.288***	1.387***
		(0.919)	(5.119)	(2.862)
		0.526	0.410	0.562
Top marginal tax rate	82	16.736**	− 1.754	4.674
		(2.248)	(0.392)	(0.884)
		0.223	0.358	0.415
4. Size of the public sector				
SOEs in the economy	103	− 0.642	1.091	0.378
		(0.559)	(1.134)	(0.392)
		0.118	0.146	0.246
Public sector empl./Total pop.	117	0.813	1.731	0.643
		(0.776)	(1.609)	(0.723)
		0.709	0.385	0.720
5. Size of government				
Government consumption/GDP	103	5.535**	3.505**	2.567
		(2.109)	(2.202)	(1.067)
		0.210	0.272	0.306
Transfers and subsidies/GDP	89	7.255**	6.020**	4.425
		(2.962)	(2.417)	(1.616)
		0.713	0.591	0.717
6. Public goods				
Infrastructure quality	59	0.595	3.178**	1.669**
		(1.243)	(3.163)	(2.421)
		0.827	0.255	0.797
Log infant mortality	168	− 0.161	− 1.024***	− 0.463**
		(0.863)	(4.598)	(2.502)
		0.824	0.471	0.789

Table 15. Continued.

Dependent Variable	Obs.	(1)	(2)	(3)
7. Schooling and literacy				
Illiteracy rate	118	4.020	− 23.794***	− 17.174***
		(0.567)	(4.084)	(2.989)
		0.664	0.463	0.647
Log school attainment	101	0.083	0.854***	0.336**
		(0.426)	(4.696)	(2.397)
		0.781	0.449	0.791
8. Political rights				
Democracy index	148	− 2.688**	3.172**	0.233
		(2.101)	(2.331)	(0.163)
		0.559	0.150	0.438
Political rights index	169	− 0.690	1.907***	1.226
		(1.004)	(2.754)	(1.480)
		0.516	0.133	0.251

Notes: *: Significant at the 10% level.
**: Significant at the 5% level.
***: Significant at the 1% level.
Robust *t*-statistics are in parentheses, adjusted *R*-squared are reported underneath.

Specification (1) includes the log(GNP) for 1970–95, legal origin dummies (Socialist, French, German and Scandinavian), religion variables (Catholic, Muslim and other) and latitude.

Specification (2) includes the log of population in 1960 and regional dummies (SubSaharan Africa, East Asia and Latin America).

Specification (3) includes the log(GNP) for 1970–95, the log of population in 1960, regional dummies (SubSaharan Africa, East Asia and Latin America) and legal origin dummies (Socialist, French, German and Scandinavian). All specifications include a constant.

6. In many regressions neither latitude nor ethnic fractionalization are significant but they both tend to be when introduced alone. The tables do not show the case in which latitude is entered without ethnic fractionalization, but these results are available upon request.

7. Not surprisingly, since we are using the same data, we confirm results in La Porta et al. (1999) on legal origins. French and Socialist legal origins seem to be negatively associated with measures of government quality.

5. Polarization

Our measures of fractionalization, while the most widespread in the literature, are not the only measures of ethnic, religious, and linguistic heterogeneity available. In particular, scholars have recently started to calculate and use measures of polarization, rather than fractionalization.[16] Holding the ''distance'' between groups constant, polarization is typically maximized when there are two groups of equal size, whereas fractionalization increases when there are many small groups. Additionally, the degree of polarization

should increase as the ''distance'' between groups increases. When it comes to ethnic, linguistic, and religious groups, however, the concept of distance is hard to capture with simple measures, so that researchers have implicitly assumed that the distance between groups is constant across group pairs.[17]

Whether societal conflict is the result of fractionalization or polarization is largely an unresolved question in theory, calling for empirical work. The discussion of whether a country with many relatively small groups is more or less stable than one with only two equally sized groups is an old one, and goes back at least to Madison in the Federalist Papers of 1788 (nos 10 and 11, see Hamilton et al., 1911). Without much of a stretch of Madison's views, one can argue that a polarization measure is, according to him, the appropriate concept to capture heterogeneity. In what follows, we extend our estimates of the effects of ethnic, religious, and linguistic heterogeneity on growth and government quality using measures of polarization.

Esteban and Ray (1994) develop a theory of polarization and axiomatically derive a measure of income polarization with ''desirable'' features.[18] They show that a desirable measure of polarization must take the form:

$$P(s,y) = K \sum_{i=1}^{n} \sum_{j=1}^{n} s_i^{1+\alpha} s_j |y_i - y_j|, \tag{2}$$

where K is a constant, α is a constant between 0 and 1.6, and y_i is the income of group $i = 1 \ldots n$ (polarization here is defined with respect to income—it is presumed there is no within group variation in income). This defines a fairly narrow class of measures, since the only degrees of freedom are K (a scaling factor) and α. In order to compute this measure for ethnically defined groups we need a measure of the distance between groups (the analog of $|y_i - y_j|$ in equation (2)). Lacking such a measure, the empirical literature has had to assume that the distance between every group is the same.[19]

We have repeated all the previous estimations using the measure of polarization based on equation (2) rather than fractionalization. For the sake of brevity, we do not report the estimates, which are available upon request. We set $K = 1$ and considered three values of α—the two extreme values 0 and 1.6, and an intermediate value of 0.8.[20] In a context where we have to assume that $|y_i - y_j|$ is constant for all $i \neq j$, it is easy to show that a value of $\alpha = 0$ leads to a measure that is perfectly correlated with the fractionalization measure presented earlier if and only if every group in a country's population is observed (so that the sum of observed group shares equals 1).[21]

A summary of our results is as follows:

1. The polarization indices, especially when computed with an α around the middle of the feasible range (i.e. around 0.8), are highly correlated with the fractionalization measures. This is due to the assumption of equal distance between all groups within a given country, an assumption required by our lack of data on those distances.

2. We find that polarization indices with α equal to 0.8 perform relatively better than those with an extreme value of 1.6. Thus the index of polarization that works best is the one that is most correlated with fractionalization.

3. In the growth regressions, the results seems substantially weaker using the polarization index. When both indices are used together, the fragmentation index typically remains significant and the polarization index is not.

4. In the quality of government regressions the index of ethnic fractionalization works slightly better than the corresponding index of polarization. In these regressions, because of the high correlation between the two indices, when they are used together the results are generally not easily interpretable.

5. We obtain poor results using linguistic polarization in the quality of government regressions whatever the value of α.

6. As for the fractionalization index, religious polarization is associated with better performances in the areas of regulation, red tape, corruption, tax compliance, transfers and political rights, especially when α has an intermediate value of 0.8. Higher religious polarization is also associated with better public goods, lower infant mortality and illiteracy rates, higher levels of schooling and better infrastructure.

The bottom line is that the measure of polarization produces similar but slightly worse results than the result with fractionalization when we choose a parameter value for α that makes the correlation between the two indices high. When the choose of parameter values makes the polarization index less correlated with fractionalization, the latter is a better predictor of our outcome variables.

Garcia-Montalvo and Reynal-Querol (2002) show that their index of polarization predicts civil wars better than fractionalization. Their results and ours are not contradictory. It may very well be the case that a civil war is more likely when two large equally size group face each other than when many small groups are present. A war is also more likely when the two groups are roughly of equal size than when one is clearly predominant since in that case the dominant one can more easily repress the other one and avoid war. However, the quality of policy may be specially bad when many groups fight over public resources.

6. The Effects of Ethnic Conflict: A Few Examples

A cross-country statistical exercise is a crude way to summarize complex political and economic histories of countries and their constituent ethnic groups. A promising direction for future research, to supplement large-sample studies, would be for economists to do more case histories of development, economic policy, and government quality in ethnically diverse places.

In this sections we briefly examine some individual data points to illustrate salient ethnic divisions as well as the complex history that lies behind our cross-sectional associations. Nigeria has among the highest ethnic and linguistic diversity in the entire sample, and was also ranked as highly diverse by Easterly and Levine (1997). Maier (2000) makes it clear that it would be hard to find a better example of institutional and policy failure leading to

Table 16. Country examples of ethnic fractionalization, growth and government quality.

Country	Average Growth in 1980s	Average Black Market Premium in 1980s	Average Years of Schooling of Labor Force, 1980s	Telephone Lines per 1000 Workers, 1980s	Government Balance to GDP, 1980s	Corruption (0 to 10) Higher Means Less Corruption	Ethnic Fractionalization
Africa							
Botswana	7.0%	16%	3.3	27	11.2%	6.5	0.41
Ethiopia	0.0%	76%		4	− 7.1%	4.3	0.72
Nigeria	− 3.3%	76%		2	0.3%	3.0	0.85
Latin America							
Bolivia	− 3.3%	39%	5.0	46	− 14.4%	2.8	0.74
Chile	1.9%	16%	7.0	84	− 0.2%	5.3	0.19
Guyana	− 2.4%	131%	5.6		− 39.7%	2.0	0.62

underdevelopment. Nigeria has produced $280 billion in oil revenues since the discovery of reserves in the late 1950s, but the average Nigerian is no further out of poverty today than four decades ago. Such egregious failures as the $8 billion state-owned Ajaokuta steel complex, which has yet to produce a bar of steel, give a hint of the breakdown of state institutions. The standard account of Nigeria's ethnic conflict pits the Muslim North versus the Christian South, but this is a simplification. First, the Christian South is divided between the Yoruba and Igbo. Second, there are substantial Southern minority groups living in Northern cities, a situation that has led to recurrent communal violence. Third, fractious ethnic groups in the center of the country and in the oil-rich Niger delta keep small-scale conflict going even out of the limelight of the Hausa/Yoruba/Igbo three-way ethnic war. Table 16 shows that Nigeria has had disastrous economic policies (high black market premia), poor infrastructure (virtually no telephone density) and high corruption.

Ethiopia also has very high ethnic and linguistic diversity (according to both new and old measures), and ethnic conflict has been at the center of Ethiopian history for centuries. Ethiopia has had one of the lowest growth rates in the world over the past half-century and as a result remains one of the least developed nations. It has known various types of regimes, from monarchy to Marxist-Leninist to reformist, but growth has been mediocre to poor under all of them. Political/ethnic conflict and disastrous institutions have partly caused and certainly magnified the effects of major disasters such as famine, AIDS, civil war, and international war, and these disasters have absorbed a high share of the government's paltry aid and tax revenues.

The current government is dominated by the Tigray Peoples's Liberation Front, representing an ethnic group making up only 6 percent of the population. The latter is alleged to own a large number of agricultural, industrial, and financial businesses under the umbrella of the Endowment Fund for the Rehabilitation of Tigray.[22] The current government is attempting to prevent ethnic conflicts by decentralizing power to ethnically defined regions, including the promotion of local languages. However, this strategy remains deeply controversial. At one extreme, some observers see it as a ploy by the Tigrayan ruling elite to divide the potential opposition along ethnic lines, as well as to undercut the national government bureaucracy. The rulers are alleged to have co-opted participants from other

ethnic groups rather than allowed representative organizations to emerge. There are also accusations of a second level of oppression, this time by the dominant majority group in each region oppressing the regional minorities (on some accounts, there are over 80 ethnic groups in Ethiopia, but only nine regions). Some Ethiopians decry the threat to the unity and identity of the country and the ''ethnicization'' of politics. Of course, ethnicization is far from new, given the long-standing Amhara dominance of the state, and its ''colonization'' of other ''nationalities.'' Some see the current government as simply substituting Tigray dominance for Amhara dominance (see Tronvoll, 2000).

More charitable observers see government policy as an honest attempt to address the ethnic divisions that have bedeviled Ethiopia for much of its history (today there continues to be an armed insurgency by the Oromo Liberation Front).[23] These observers see the current government as responding to this history of domination by the Amhara by granting autonomy to the ''nationalities.''

Botswana is an interesting exception to the poor economic outcomes and low quality government in most of Africa. The table shows it had high growth, a low black market premium, a government surplus, and low corruption. While we do not mean to give a monocausal explanation for this success, it is notable that it has relatively low ethnic diversity for Africa. Acemoglu et al. (2003) describe how the Tswana tribes had a long history of cooperation amongst themselves before independence, as well as generally inclusive institutions since.

Ethiopia and Nigeria were already highly diverse relative to Botswana in the old Easterly and Levine (1997) dataset. Even more interesting is the much higher degree of ethnic diversity in some Latin American and Caribbean countries according to our new ethnic fractionalization measure. Among the poorest, most institutionally underdeveloped, and most conflict-ridden societies in this region are Bolivia, Ecuador, and Guyana. All of these score lower on linguistic fractionalization (and hence did not show up as very ethnically diverse in Easterly and Levine, 1997) than on ethnic fractionalization, because of racial differences. A Latin American success story, Chile, continues to show up as relatively homogeneous.

To take Bolivia as an example, whites (about 10 percent of the population) dominated the governments of Bolivia with systematic exploitation of Mestizos, Aymara and Quechua peoples from colonial times to 1952 (Klein, 1992). Six percent of landowners owned 92 percent of the land in 1950. There were feudal anachronisms such as an obligation for Indian tenants to spend part of their time as unpaid servants in the landowner's household. A literacy requirement prevented the majority of the population from voting. The Indians successfully revolted in 1952, redistributed land toward the peasants, and abolished the more obvious exclusionary laws. However, whites continued to dominate politics and economics. Political instability remained endemic after 1952, with frequent military coups overthrowing democratic regimes. Democracy has been restored since 1982, but racial conflict continues. When one of the authors visited Bolivia in early 2002, Indian activists were blockading the main roads surrounding La Paz to articulate various grievances. A meeting of the leaders of the 1952 revolutionary party was conspicuous for its lack of Indian representation. Anecdotally, it appeared that racist sentiments toward the Indians still existed amongst the white elite. Bolivia still has poor growth, high corruption, poor social service delivery, and predatory police and judges.

Guyana shows up as ethnically diverse in our data because of its racial breakdown between Africans, East Indians, Europeans, and others. The Afro-Guyanese and Indo-Guyanese are the predominant groups and are almost numerically equal. Since they have mobilized politically along ethnic lines (supporting two different parties since before independence), any consensus for development has been torn apart by competition for rents between the two groups. As Table 16 shows, Guyana is rated as one of the most corrupt countries in the world, has followed distortionary economic policies, and has had very poor growth outcomes.

Chile, in contrast, is a well known Latin American success story. It has pursued free market reforms since the 1970s under first military and later democratically elected governments. Although it did have political and economic gyrations under Allende in the early 1970s and then a debt crisis and severe political repression under Pinochet in the early 1980s, the last 20 years have shown a high degree of political and economic stability and sustained growth. By the 1980s, Chile had also achieved a high level of schooling and infrastructure (Table 16). There are certainly many causes explaining why reforms were made possible, in particular the authoritarian and repressive nature of the Pinochet regime, which made it easier to eliminate opposition to reform. After the period of repression a considerable amount of consensus emerged on policy. Many other developing countries experienced bloody coups, and did not evolve into peaceful and rapidly growing economies. The difference in Chile was probably due to its higher level of homogeneity. In fact, after Pinochet's departure from power the new democratic regime showed remarkable stability by Latin American standards. The relative ethnic homogeneity of Chilean society may have made achieving support for reform and economic development easier than in Bolivia or Guyana.

7. Conclusion

The question of what makes different countries more or less successful economically and what explains the quality of their policies is one of the most fascinating that economists can ask, but it is also one of the most difficult to answer. Different authors have their own ''favorite'' explanatory variables: from purely ''economic'' ones, to geographic ones, to legal ones, to political, cultural, religious and historical ones. In this paper we have considered closely one such set of variables: measures of ethnic, linguistic and religious fractionalization.

Dealing with this type of variables raises two problems. One is measurement: how to measure ethnicity is a delicate and difficult matter. Moreover, the patterns of correlations between potential explanatory factors makes it difficult to unambiguously answer the question of why certain countries have better policies than others. In this paper we have made some progress on both fronts. First, on the measurement issue we provided a new set of fractionalization variables for a much larger sample of countries than was available before, and we put much effort into solving classification issues using consistent criteria across countries. Second, using these new variables we revisited empirical issues concerning the determinants of growth and of quality of policies and institutions. We concluded that ethnic and linguistic fractionalization variables, more so than religious

ones, are likely to be important determinants of economic success, both in terms of output (GDP growth), other measures of welfare and policy quality (such as the literacy rate, infant mortality, etc.) and the quality of institutions (measured by the extent of corruption, political freedom, etc.). However, it is difficult to evaluate precisely the size of these effects because of the strong correlation of ethnolinguistic fractionalization variables with other potential explanatory variables, especially geographical ones. In the end one has to use theory and priors to interpret our partial correlations.

Notes

1. For a discussion of the political economy of borders, country size and heterogeneity of populations, see Alesina and Spolaore (2003). Note, however, that ethnic conflict was a constant in African history even before colonization, as pointed out by Herbst (2000).
2. See, for example, the general growth empirics exercises of Brock and Durlauf (2001), and Doppelhofer et al. (2000).
3. A partial list of papers in this area include Alesina et al. (1999), Alesina and La Ferrara (2000), Alesina and La Ferrara (2002), Lutmer (2001), Goldin and Katz (1999) and Costa and Kahn (2002). Another line of research has explored the effects of ethnic diversity on civil wars, suggesting that fractionalization does not help predict the incidence of domestic violent conflict once poverty and income levels are controlled for (see Fearon and Lattin, 2003).
4. See Glaeser et al. (2000).
5. Racial classifications follow the census which categorizes Americans into five groups: Whites, Blacks, America Indians, Pacific Islanders, and Hispanics. The country of origin, such as Ireland, Italy, Japan etc., is also available.
6. See http://www.stanford.edu/~wacziarg/papersum.html for our new dataset in electronic format, including the underlying groups structure of ethnicities, languages, and religions by country. Appendix 1 displays the fractionalization indices constructed from these data.
7. For the purpose of cross-country regressions, ELF was used, among many others, in Mauro (1995), Canning and Fay (1993) and Easterly and Levine (1997).
8. http://www.ethnologue.com/
9. However, the Ethnologue data is much more disaggregated than the Encyclopedia Britannica data that we use, as relatively similar dialects are classified there as different languages.
10. See, respectively, INSEE (2000), Israel Central Bureau of Statistics (2001), United States Census Bureau (2001) and Statistics New Zealand (2001).
11. See Hammersley and Atkinson (1995) for a survey of these concepts. The reflexive school of thought, which seems to be associated with the postmodern tradition in sociology, is generally contrasted with the ''primordialist'' school, identified for example with Geertz (1973), which is associated with evolutionist theories.
12. Mozzafar and Scarrit (1999) report data on ethnicity at three distinct levels of ''self-reference'' for Africa. We use their clan data for Somalia, since the rest of their dataset is sufficiently close to our other sources.
13. The table is similar to Table 1 in Fearon (2002) to facilitate comparisons between his data and ours.
14. The coefficient estimates on the other independent variables included in each specification are available in the working paper version of this paper (Alesina et al., 2002) or from the authors upon request.
15. These authors present additional evidence precisely on this point using the data on ethnic fractionalization collected for the present paper. This is consistent with the fact that relatively homogeneous settler colonies like the US, Canada, New Zealand, and Australia had an easier time establishing democracy after independence than the more ethnically diverse former colonies in Latin America and Africa.
16. See, for instance, Garcia-Montalvo and Reynal-Querol (2002).
17. A recent effort to generate an index of ''cultural distance'' appears in Fearon (2002). He noticed that linguists classify languages in trees—two languages on the same branch of a linguistic tree are ''closer'' to each other

than two languages on a different branch. Linguistic trees can therefore be used to approximate the distance between linguistic groups. Efforts to similarly measure the distance between ethnic groups and religions should be at the forefront of the data gathering effort in fractionalization research. We are not aware of such attempts.

18. The axioms imposed on polarization measures are the analogs of the Dalton Axioms for the measurement of inequality. See also Duclos et al. (2002) for a generalization.
19. Garcia-Montalvo and Reynal-Querol (2002) use the following specification:

$$\mathrm{POL}_i = 1 - \sum_{i=1}^{n} \left(\frac{0.5 - s_i}{0.5} \right)^2 s_i.$$

They show that this index can be derived from a model of lobbying. Note that this index reaches a maximum of 1 when there are two equally sized groups in the country.

20. Esteban and Ray (1994) do not point to which value was ''better'' to capture polarization—all values of α in the specified range satisfy the properties that the class of polarization measures should satisfy. There is therefore no *a priori* reason to prefer one value over the other.
21. Thus, in practice a measure based on $\alpha = 0$ does not add anything to our earlier results, so we abstain from commenting on the corresponding results.
22. Abegaz (2001), p. 207.
23. See Marcus (1994) on the complicated history of ethnic groups in Ethiopia.

Appendix

Table A1. Fractionalization data.

Country	Source (Ethnicity Data)	Date (Ethnicity Data)	Ethnic (Source: see column 2. Date: see column 3)	Language (Source: eb. Date: 2001)	Religion (Source: eb. Date: 2001)
Afghanistan	wdm	1995	0.7693	0.6141	0.2717
Albania	wdm	1989	0.2204	0.0399	0.4719
Algeria	eb	1992	0.3394	0.4427	0.0091
American Samoa				0.1733	0.6395
Andorra	eb	1997	0.7139	0.6848	0.2326
Angola	eb	1983	0.7867	0.7870	0.6276
Antigua and Barbuda	eb	1994	0.1643	0.1063	0.6840
Argentina	eb	1986	0.2550	0.0618	0.2236
Armenia	eb	1989	0.1272	0.1291	0.4576
Aruba				0.3889	0.4107
Australia	eb	1986	0.0929	0.3349	0.8211
Austria	lev	1998	0.1068	0.1522	0.4146
Azerbaijan	eb	1995	0.2047	0.2054	0.4899
Bahamas	lev	1989	0.4228	0.1855	0.6815
Bahrain	eb	1991	0.5021	0.4344	0.5528
Bangladesh	eb	1997	0.0454	0.0925	0.2090
Barbados	eb	1990	0.1423	0.0926	0.6934
Belarus	cia	2001	0.3222	0.4666	0.6116
Belgium	cia	2001	0.5554	0.5409	0.2127
Belize	eb	1991	0.7015	0.6303	0.5813
Benin	eb	1992	0.7872	0.7905	0.5544
Bermuda					0.7112

Table A1. Continued.

Country	Source (Ethnicity Data)	Date (Ethnicity Data)	Ethnic (Source: see column 2. Date: see column 3)	Language (Source: eb. Date: 2001)	Religion (Source: eb. Date: 2001)
Bhutan	eb	1993	0.6050	0.6056	0.3787
Bolivia	lev	1998	0.7396	0.2240	0.2085
Bosnia and Herzegovina	eb	1991	0.6300	0.6751	0.6851
Botswana	eb	1983	0.4102	0.4110	0.5986
Brazil	eb	1995	0.5408	0.0468	0.6054
Brunei	wdm	1995	0.5416	0.3438	0.4404
Bulgaria	wdm	1992	0.4021	0.3031	0.5965
Burkina Faso	eb	1983	0.7377	0.7228	0.5798
Burundi	eb	1983	0.2951	0.2977	0.5158
Cambodia	eb	1994	0.2105	0.2104	0.0965
Cameroon	eb	1983	0.8635	0.8898	0.7338
Canada	eb	1991	0.7124	0.5772	0.6958
Cape Verde	eb	1986	0.4174		0.0766
Central African Republic	eb	1988	0.8295	0.8334	0.7916
Chad	eb	1993	0.8620	0.8635	0.6411
Chile	eb	1992	0.1861	0.1871	0.3841
China	eb	1990	0.1538	0.1327	0.6643
Colombia	eb	1985	0.6014	0.0193	0.1478
Comoros	eb	1995	0.0000	0.0103	0.0137
Congo, Dem. Rep. (Zaire)	eb	1983	0.8747	0.8705	0.7021
Congo	eb	1983	0.8747	0.6871	0.6642
Costa Rica	eb	1993	0.2368	0.0489	0.2410
Cote d'lvoire	lev	1998	0.8204	0.7842	0.7551
Croatia	eb	1991	0.3690	0.0763	0.4447
Cuba	eb	1994	0.5908		0.5059
Cyprus	eb	1992	0.0939	0.3962	0.3962
Czech Republic	eb	1991	0.3222	0.3233	0.6591
Denmark	eb	1996	0.0819	0.1049	0.2333
Djibouti	eb	1983	0.7962	0.6558	0.0435
Dominica	eb	1991	0.2003		0.4628
Dominican Republic	eb	1993	0.4294	0.0395	0.3118
East Timor				0.5261	0.4254
Ecuador	eb	1989	0.6550	0.1308	0.1417
Egypt	lev	1998	0.1836	0.0237	0.1979
El Salvador	eb	1993	0.1978		0.3559
Equatorial Guinea	lev	1998	0.3467	0.3220	0.1195
Eritrea	lev	1998	0.6524	0.6530	0.4253
Estonia	eb	1994	0.5062	0.4944	0.4985
Ethiopia	eb	1983	0.7235	0.8073	0.6249
Faroe Islands					0.3147
Fiji	eb	1996	0.5479	0.5479	0.5682
Finland	cia	2001	0.1315	0.1412	0.2531
France	census	1999	0.1032	0.1221	0.4029
French Guiana				0.1154	0.4959
French Polynesia				0.6078	0.5813

Table A1. Continued.

Country	Source (Ethnicity Data)	Date (Ethnicity Data)	Ethnic (Source: see column 2. Date: see column 3)	Language (Source: eb. Date: 2001)	Religion (Source: eb. Date: 2001)
Gabon	eb	1983	0.7690	0.7821	0.6674
Gambia, The	eb	1993	0.7864	0.8076	0.0970
Gaza Strip				0.0104	0.0342
Georgia	eb	1989	0.4923	0.4749	0.6543
Germany	eb	1997	0.1682	0.1642	0.6571
Ghana	eb	1983	0.6733	0.6731	0.7987
Greece	lev	1998	0.1576	0.0300	0.1530
Greenland				0.2188	0.4592
Grenada	eb	1991	0.2661		0.5898
Guadeloupe				0.0933	0.3069
Guam				0.7320	0.4082
Guatemala	cia	2001	0.5122	0.4586	0.3753
Guinea	eb	1990	0.7389	0.7725	0.2649
Guinea-Bissau	eb	1979	0.8082	0.8141	0.6128
Guyana	eb	1993	0.6195	0.0688	0.7876
Haiti	eb	1993	0.0950		0.4704
Honduras	eb	1987	0.1867	0.0553	0.2357
Hong Kong	wdm	1994	0.0620	0.2128	0.4191
Hungary	eb	1993	0.1522	0.0297	0.5244
Iceland	eb	1995	0.0798	0.0820	0.1913
India	eb	2000	0.4182	0.8069	0.3260
Indonesia	eb	1990	0.7351	0.7680	0.2340
Iran	eb	1995	0.6684	0.7462	0.1152
Iraq	eb	1983	0.3689	0.3694	0.4844
Ireland	eb	1995	0.1206	0.0312	0.1550
Isle of Man					0.4729
Israel	census	1995	0.3436	0.5525	0.3469
Italy	eb	1983	0.1145	0.1147	0.3027
Jamaica	eb	1982	0.4129	0.1098	0.6160
Japan	cia	1999	0.0119	0.0178	0.5406
Jersey					0.5479
Jordan	wdm	1993	0.5926	0.0396	0.0659
Kazakhstan	cia	1999	0.6171	0.6621	0.5898
Kenya	cia	2001	0.8588	0.8860	0.7765
Kiribati	eb	1990	0.0511	0.0237	0.5541
Korea, North	eb	1995	0.0392	0.0028	0.4891
Korea, South	eb	1990	0.0020	0.0021	0.6604
Kyrgyzstan	cia	2001	0.6752	0.5949	0.4470
Kuwait	cia	2001	0.6604	0.3444	0.6745
Lao People's Dem. Rep.	eb	1983	0.5139	0.6382	0.5453
Latvia	eb	1996	0.5867	0.5795	0.5556
Lebanon	eb	1996	0.1314	0.1312	0.7886
Lesotho	eb	1986	0.2550	0.2543	0.7211
Liberia	wdm	1992	0.9084	0.9038	0.4883
Libya	eb	1995	0.7920	0.0758	0.0570

Table A1. Continued.

Country	Source (Ethnicity Data)	Date (Ethnicity Data)	Ethnic (Source: see column 2. Date: see column 3)	Language (Source: eb. Date: 2001)	Religion (Source: eb. Date: 2001)
Liechtenstein	eb	1997	0.5726	0.2246	0.3343
Lithuania	eb	1996	0.3223	0.3219	0.4141
Luxembourg	eb	1996	0.5302	0.6440	0.0911
Macau				0.2519	0.5511
Macedonia (former Yug. Rep.)	eb	1994	0.5023	0.5021	0.5899
Madagascar	lev	1985	0.8791	0.0204	0.5191
Malawi	lev	1998	0.6744	0.6023	0.8192
Malaysia	eb	1996	0.5880	0.5970	0.6657
Mali	cia	1996	0.6906	0.8388	0.1820
Malta	lev	1996	0.0414	0.0907	0.1223
Marshall Islands	eb	1988	0.0603	0.0601	0.5207
Martinique				0.0653	0.2336
Mauritania	wdm	1992	0.6150	0.3260	0.0149
Mauritius	eb	1992	0.4634	0.4547	0.6385
Mayotte				0.7212	0.0620
Mexico	eb	1990	0.5418	0.1511	0.1796
Micronesia	eb	1994	0.7005	0.7483	0.6469
Moldova	eb	1989	0.5535	0.5533	0.5603
Monaco	cia	2001	0.6838	0.7305	0.3047
Mongolia	eb	1989	0.3682	0.3734	0.0799
Morocco	wdm	1994	0.4841	0.4683	0.0035
Mozambique	eb	1983	0.6932	0.8125	0.6759
Myanmar (Burma)	eb	1983	0.5062	0.5072	0.1974
Namibia	wdm	1995	0.6329	0.7005	0.6626
Nauru	cia	1995	0.5832	0.6161	0.6194
Nepal	eb	1991	0.6632	0.7167	0.1417
Netherlands Antilles				0.2508	0.3866
Netherlands	lev	1995	0.1054	0.5143	0.7222
New Caledonia				0.6633	0.5462
New Zealand	census	1996	0.3969	0.1657	0.8110
Nicaragua	eb	1991	0.4844	0.0473	0.4290
Niger	eb	1988	0.6518	0.6519	0.2013
Nigeria	eb	1983	0.8505	0.8503	0.7421
Northern Mariana Islands				0.7754	0.4811
Norway	lev	1998	0.0586	0.0673	0.2048
Oman	eb	1993	0.4373	0.3567	0.4322
Pakistan	wdm	1995	0.7098	0.7190	0.3848
Palau	cia	2000	0.4312	0.3157	0.7147
Panama	eb	1992	0.5528	0.3873	0.3338
Papua New Guinea	eb	1993	0.2718	0.3526	0.5523
Paraguay	lev	1998	0.1689	0.5975	0.2123
Peru	eb	1981	0.6566	0.3358	0.1988
Philippines	lev	1998	0.2385	0.8360	0.3056
Poland	lev	1998	0.1183	0.0468	0.1712
Portugal	lev	1998	0.0468	0.0198	0.1438

Table A1. Continued.

Country	Source (Ethnicity Data)	Date (Ethnicity Data)	Ethnic (Source: see column 2. Date: see column 3)	Language (Source: eb. Date: 2001)	Religion (Source: eb. Date: 2001)
Puerto Rico				0.0352	0.4952
Qatar	cia	2001	0.7456	0.4800	0.0950
Reunion				0.1578	0.1952
Romania	lev	1998	0.3069	0.1723	0.2373
Russian Federation	eb	1997	0.2452	0.2485	0.4398
Rwanda	eb	1996	0.3238		0.5066
Saint Lucia	eb	1990	0.1769	0.3169	0.3320
Saint Vincent and Grenadines	eb	1995	0.3066	0.0175	0.7028
Western Samoa	eb	1995	0.1376	0.0111	0.7871
San Marino	eb	1997	0.2927		0.1975
Sao Tome and Principe				0.2322	0.1866
Saudi Arabia	eb	1995	0.1800	0.0949	0.1270
Senegal	eb	1988	0.6939	0.6961	0.1497
Serbia/Montenegro (Yugoslavia)	eb	1991	0.5736		
Seychelles	eb	1983	0.2025	0.1606	0.2323
Sierra Leone	wdm	1993	0.8191	0.7634	0.5395
Singapore	cia	2001	0.3857	0.3835	0.6561
Slovak Republic	eb	1996	0.2539	0.2551	0.5655
Slovenia	cia	1991	0.2216	0.2201	0.2868
Solomon Islands	eb	1986	0.1110	0.5254	0.6708
Somalia	sm	1999	0.8117	0.0326	0.0028
South Africa	lev	1998	0.7517	0.8652	0.8603
Spain	eb	1991	0.4165	0.4132	0.4514
Sri Lanka	cia	2001	0.4150	0.4645	0.4853
St Kitts and Nevis	lev	1998	0.1842		0.6614
Sudan	eb	1983	0.7147	0.7190	0.4307
Suriname	cia	2001	0.7332	0.3310	0.7910
Swaziland	cia	2001	0.0582	0.1722	0.4444
Sweden	lev	1998	0.0600	0.1968	0.2342
Switzerland	cia	2001	0.5314	0.5441	0.6083
Syria	wdm	1993	0.5399	0.1817	0.4310
Taiwan	cia	2001	0.2744	0.5028	0.6845
Tajikistan	cia	2001	0.5107	0.5473	0.3386
Tanzania	eb	1995	0.7353	0.8983	0.6334
Thailand	eb	1983	0.6338	0.6344	0.0994
Togo	eb	1995	0.7099	0.8980	0.6596
Tonga	eb	1995	0.0869	0.3782	0.6214
Trinidad and Tobago	cia	2001	0.6475	0.1251	0.7936
Tunisia	cia	2001	0.0394	0.0124	0.0104
Turkey	cia	2001	0.3200	0.2216	0.0049
Turkmenistan	eb	1997	0.3918	0.3984	0.2327
Tuvalu	eb	1979	0.1629	0.1372	0.2524
Uganda	eb	1983	0.9302	0.9227	0.6332
Ukraine	eb	1998	0.4737	0.4741	0.6157
United Arab Emirates	eb	1993	0.6252	0.4874	0.3310

Table A1. Continued.

Country	Source (Ethnicity Data)	Date (Ethnicity Data)	Ethnic (Source: see column 2. Date: see column 3)	Language (Source: eb. Date: 2001)	Religion (Source: eb. Date: 2001)
United Kingdom	eb	1994	0.1211	0.0532	0.6944
United States	census	2000	0.4901	0.2514	0.8241
Uruguay	eb	1990	0.2504	0.0817	0.3548
Uzbekistan	eb	1995	0.4125	0.4120	0.2133
Vanuatu	eb	1989	0.0413	0.5794	0.7044
Venezuela	eb	1993	0.4966	0.0686	0.1350
Vietnam	eb	1995	0.2383	0.2377	0.5080
Virgin Islands (US)				0.3140	0.6359
West Bank				0.1438	0.3095
Yemen				0.0080	0.0023
Yugoslavia (pre 1991)	eb	1995	0.8092	0.4050	0.5530
Zambia	lev	1998	0.7808	0.8734	0.7359
Zimbabwe	lev	1998	0.3874	0.4472	0.7363

Source Key: eb = Encyclopedia Brittanica, cia = CIA; sm = Scarrit and Mozaffar; lev = Levinson; wdm = World Directory of Minorities; census = national census data.

Table A2. Description of the data from Easterly and Levine, 1997.

Growth of per capita real GDP	Growth rate of real per capita GDP, World Bank (various years)
Dummy for the 1960s, 1970s, 1980s	Dummy variable for 1960s, 1970s, 1980s
Dummy variable for SubSaharan Africa	Dummy variable for SubSaharan African countries. World Bank.
Dummy variable for Latin America and the Caribbean	Dummy variable for Latin America and the Caribbean
Log of initial income	Log of initial income: log of real per capita GDP measured at the start of each decade (1960, 1970, 1980). Summers and Heston (1988)
Log of initial income squared	Log of initial income squared: log of initial real per capita GDP squared. Summers and Heston (1988)
Log of schooling	Log of schooling: log of 1 + average years of school attainment, beginning of each decade (1960, 1970, and 1980). Barro and Lee (1993)
Assassinations	Assassinations: number of assassinations per thousand population, decade average. Banks (1994)
Financial depth	Financial depth: ratio of liquid liabilities of the financial system to GDP, decade average. Liquid liabilities consist of currency held outside the banking system 1 demand and interes-bearing liabilities of banks and nonbank financial intermediaries. King and Levine (1993b)
Black market premium	Black market premium: log of 1 + black market premium, decade average. World Bank (1991) and Pick's Currency Yearbook (various years)

Table A2. Continued.

Fiscal surplus/GDP	Fiscal surplus/GDP: decade average of ratio of central government surplus to GDP, both in local currency, current prices. IMF (various years) *International Financial Statistics* (line 80), and IMF (various years) *Government Finance Statistics* (line L80)
Log of telephones per worker	Log of telephones per worker: log of telephones per 1000 workers. Canning and Fay (1993)
ELF	Index of ethnolinguistic fractionalization, 1960. Measures probability that two randomly selected people from a given country will not belong to the same ethnolinguistic group. *Atlas Narodov Mira* (1964)

Source: This table was extracted from Easterly and Levine (1997)

Table A3. Description of the data from La Porta et al. (1999). Description of the variables.

Variable Name	Description and Source	Number of Observations
Interference with the private sector		
Property rights index	A rating of property rights in each country (on a scale of 1 to 5). The more protection private property receives, the higher the score. The score is based, broadly, on the degree of legal protection of private property, the extent to which the government protects and enforces laws that protect private property, the probability that the government will expropriate private property, and the country's legal protection to private property. Source: *Holmes, Johnson, and Kirkpatrick, 1997*.	149
Business Regulation index	A rating of regulation policies related to opening a business and keeping open a business (on a scale of 1 to 5). Higher score means that regulations are straight-forward and applied uniformly to all businesses and that regulations are less of a burden to business. Source: *Holmes, Johnson and Kirkpatrick, 1997*.	149
Top tax rate	Top marginal tax rate for each country in 1994. Source: *Gwartney, Lawson, and Block, 1996*.	82
Efficiency		
Corruption	Corruption in government index. Low ratings indicate ''high government officials are likely to demand special payments'' and ''illegal payments are generally expected thought lower levels of government'' in the form of ''bribes connected with import and export licenses, exchange controls, tax assessment, policy protection, or loans.'' Scale from 0 to 10. Average of the months of April and October in the monthly index between 1982 and 1995. Source: *Political Risk Services, various years*.	126
Bureaucratic delays	An indicator of bureaucratic delays (red tape). High ratings indicate lower levels of red tape in the bureaucracy of the country. Scale from 0 to 10. The index is published three times per year. The data is the average of the years between 1972 and 1995. Source: *Business Environmental Risk Intelligence's (BERI) Operation Risk Index*.	60
Tax Compliance	Assessment of the level of tax compliance. Scale from 0 to 6, where higher scores indicate higher compliance. Data is for 1995. Source: *World Economic Forum, 1996*.	49

Table A3. Continued.

Variable Name	Description and Source	Number of Observations
Output of public goods		
Log of infant mortality	Logarithm of the number of deaths of infants under one year of age per one thousand live births for the years 1970–1995. Source: *World Bank, World Development Indicators 1997 (WDI)*.	196
Log of school attainment	Log of schooling taken over five year periods (1960–65, 1970–75, and 1980–85). Each value is obtained as the logarithm of (1 + average years of school attainment during the respective period). Source: *Barro and Lee, 1994*.	106
Illiteracy rate	Average of adult illiteracy rate for the years 1990–1995. Adult illiteracy rate is the proportion of adults aged 15 and above who cannot, with understanding, read and write a short, simple statement of their everyday life. 1990–1995. Scale 0 to 100. Source: *WDI*.	128
Infrastructure quality	Assessment of the ''facilities for and ease of communications between headquarters and the operation, and within the country,'' as well as the quality of the transportation. Average data for the years 1972 to 1995. Scale from 0 to 10 with higher scores for superior quality. Source: *BERI's Operation Risk Index*.	60
Size of public sector		
Transfers and subsidies/GDP	Total government transfers and subsidies as a percentage of GDP (scale from 0 to 100). Average for the years 1975–1995. Source: *Gwartney, Lawson, and Block, 1996* (with data from the World Bank and International Montetary Fund).	90
Government consumption/ GDP	Government consumption expenditures as a percentage of GDP (scale from 0 to 100). Average for the years 1975–1995. Government consumption expenditures ''include all spending on goods and services purchased by the government—things like national defense, road maintenance, wages and salaries, office space, and government-owned vehicles. Since it is obtained from the national income accounts, it includes all levels of government spending. It does not include direct transfers and subsidies, since these do not enter into the national income accounts.'' Source: *Gwartney, Lawson, and Block, 1996* (with data from the World Bank and International Monetary Fund).	104
SOEs in the economy	Index of State-Owned Enterprises as a share of the economy (scale from 0 to 10). Higher scores include countries with les government-owned enterprises which are estimated to produce less of the country's output. As the estimated size and breadth of the SOE sector increases, countries are assigned lower ratings. Average of the score for the years 1975–1995. Source: *Gwartney, Lawson and Block, 1996*.	104
Public sector employment/ total population	Average of the ratio of public sector employment in general government to total population for the years 1976–1996. General government employment includes employment in ''all government department offices, organizations and other bodies which are agencies or instruments of the central or local authorities whether accounted for or financed in, ordinary or extraordinary budgets or extra-budgetary funds. They are not solely engaged in administration but also in defense and public order, in the promotion of economic growth and in the provision of education, health and cultural and social services''. Source: *Schiavo-Campo, de Tommaso, and Mukherjee, 1997*.	124

Table A3. Continued.

Variable Name	Description and Source	Number of Observations
Political Freedom		
Democracy index	Average of democracy score for the period 1970–1994. Scale from 0 to 10, with lower values indicating a less democratic environment. Source: *Jaggers and Gurr, 1996*.	161
Political rights index	Index of political rights. Higher ratings indicate countries that come closer ''to the ideals suggested by the checklist questions of: (1) free and fair elections; (2) those elected rule; (3) there are competitive parties or other competitive political groupings; (4) the opposition has an important role and power; and (5) the entities have self-determination or an extremely high degree of autonomy.'' *Source: Freedom House, 1996*.	209
Determinants		
Ethnolinguistic fractionalization	Average value of five different indices of ethonolinguistic fractionalization. Its value ranges from 0 to 1. The five component indices are: (1) index of ethonolinguistic fractionalization in 1960, which measures the probability that two randomly selected people form a given country will not belong to the same ethonolinguistic group(the index is based on the number and size of population groups as distinguished by their ethnic and linguistic status); (2) probability of two randomly selected individuals speaking different languages; (3) probability of two randomly selected individuals do not speak the same language; (4) percent of the population not speaking the official language; and (5) percent of the population not speaking the most widely used language. Sources: *Easterly and Levine, 1997*. The sources of the components of the average index are (1) *Atlas Narodov Mira, 1964;* (2) *Muller, 1964;* (3) *Roberts, 1962;* (4) and (5) *Gunnemark, 1991*.	161
Legal origin	Identifies the legal origin of the Company law or Commercial Code of each country. There are five possible origins: (1) English Common Law; (2) French Commercial Code; (3) German Commercial Code: (4) Scandinavian Commercial Code; and (5) Socialist/Communist laws. Source: *La Porta et al., 1998, extended using ''Foreign Laws: Current Sources of Basic Legislation in Jurisdictions of the World,'' 1989; and CIA World Factbook 1996*.	212
Religion	Identifies the percentage of the population of each country that belonged to the three most widely spread religions in the world in 1980. For countries of recent formation, the data is available for 1990–1995. The numbers are in percent (scale from 0 to 100). The three religions identified here are: (1) Roman Catholic; (2) Protestant; and (3) Muslim. The residual is called ''other religions''. Sources: *Barrett, 1982, Worldmark Encyclopedia of Nations 1995, Statistical Abstract of the World 1995, United Nations, 1995, CIA 1996*.	209
Economic Development		
Latitude	The absolute value of the latitude of the country, scaled to take values between 0 and 1. Source: *CIA 1996*.	209
Log GNP per capita	Logarithm of GNP per capita expressed in current US dollars for the period 1970–1995. Source: *WDL*	186

Source: This table was extracted from La Porta et al., 1999.

References

Abegaz, B. (2001). ''Ethiopia.'' In S. Devarajan, D. Dollar, and T. Holmgren (eds), *Aid and Reform in Africa*, World Bank.

Acemoglu, D., S. Johnson, and J. Robinson. (2003). ''An African Success Story: Botswana.'' In D. Rodrik (ed.), *In Search of Prosperity: Analytical Country Narratives of Economic Growth*. Princeton: Princeton University Press.

Aghion P., A. Alesina, and F. Trebbi. (2002). ''Endogenous Political Institutions.'' Unpublished, Harvard University.

Alesina A., R. Baqir, and W. Easterly. (1999). ''Public Goods and Ethnic Divisions,'' *Quarterly Journal of Economics* 114(4), 1243–1284.

Alesina, A., et al. (2002). ''Fractionalization,'' *NBER Working Paper* No. 9411, December.

Alesina A., E. Glaeser, and B. Sacerdote. (2001). ''Why Doesn't the US Have a European Style Welfare State?'' *Brookings Papers on Economic Activity* Fall.

Alesina, A., and E. La Ferrara. (2000). ''Participation in Heterogeneous Communities,'' *Quarterly Journal of Economics* 115(3), 847–904.

Alesina, A., and E. La Ferrara. (2002). ''Who Trusts Others?'' *Journal of Public Economics* 85, 207–234.

Alesina, A., and E. Spolaore. (2003). *The Size of Nations*. Cambridge, MA: MIT Press.

Alesina, A., and R. Wacziarg. (1998). ''Openness, Country Size and Government,'' *Journal of Public Economics* 69(3), p. 305–321.

Annett, A. (2001). ''Social Fractionalization, Political Instability, and the Size of Government,'' *IMF Staff Papers* 48(3), 561–592.

Atlas Narodov Mira. (1964). Moscow: Miklukho-Maklai Ethnological Institute at the Department of Geodesy and Cartography of the State Geological Committee of the Soviet Union.

Barro, R., and R. McLeary. (2002). ''Religion and Political Economy in an International Panel,'' *NBER Working Paper* No. 8931, May.

Barrett, D. B. (1982). *World Christian Encyclopedia*. New York: Oxford University Press.

Brock, S., and S. Durlauf. (2001). ''Growth Empirics and Reality,'' *World Bank Economic Review* 15(2), 229–272.

Canning, D., and M. Fay (1993). ''The Effects of Transportation Networks on Economic Growth,'' *Columbia University Working Paper*.

Caselli, F., and J. Coleman. (2002). ''On the Theory of Ethnic Conflict,'' unpublished, Harvard University.

Central Intelligence Agency. (2000). *CIA World Factbook*. Washington, DC: CIA Office of Public Affairs.

Costa, D., and M. Kahn. (2002). ''Understanding the Decline in Social Capital, 1952–1998,'' *Kyklos* 56(1), 17–46.

Doppelhofer, G., R. Miller, and X. Sala-i-Martin. (2000). ''Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach,'' *NBER Working Paper* No. 7750, June.

Duclos, J.-Y., J. Esteban, and D. Ray. (2002). ''Polarization: Concepts, Measurement, Estimation,'' *unpublished*, New York University, May.

Easterly, W., and R. Levine. (1997). ''Africa's Growth Tragedy: Policies and Ethnic Divisions,'' *Quarterly Journal of Economics* 111(4), 1203–1250.

Encyclopedia Britannica. (2000). Chicago: Encyclopedia Britannica.

Esteban, J., and D. Ray. (1994). ''On the Measurement of Polarization,'' *Econometrica* 62(4), 819–851.

Fearon, J. (2003). ''Ethnic Structure and Cultural Diversity by Country.'' *Journal of Economic Growth*, 8(2) 195–222.

Fearon, J., and D. Laitin. (2003). ''Ethnicity, Insurgency and Civil War.'', *American Political Science Review*, 97(1) (March), 75–90.

Garcia-Montalvo, J., and M. Reynal-Querol. (2002). ''Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth,'' unpublished, Universitat Pompeu Fabra, September.

Geertz, C. (1973). *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.

Glaeser, E. (2002). ''The Political Economy of Hatred,'' unpublished, Harvard University, June.

Glaeser, E. L., D. I. Laibson, J. A. Sheinkman, C. L. Soutter. (2000). ''Measuring Trust,'' *Quarterly Journal of Economics* 65, 811–846.

Goldin, C., and L. F. Katz. (1999). ''Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910–1940,'' *Journal of Interdisciplinary History* 29(4), 683–723.
Hamilton J., J. Madison, and J. Jay. (1911). *The Federalist*. London: Everyman. (first published in 1788).
Hammersley, M., and P. Atkinson (1995). *Ethnography: Principles in Practice*, 2nd edn., Routledge.
Herbst, J. (2000). *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton, NJ: Princeton University Press.
INSEE. (2000). Recensement de la population 1999.
Israel Central Bureau of Statistics. (2001). 1995 Census of Population and Housing (available on-line).
Klein, H. (1992). *Bolivia: The Evolution of a Multi-Ethnic Society*, 2nd edn., Oxford: Oxford University Press.
La Porta R., et al. (1999). ''The Quality of Government,'' *Journal of Law, Economics and Organization* 15(1), 222–279.
Levinson, D. (1998). *Ethnic Groups Worldwide, A Ready Reference Handbook*. Phoenix: Oryx Press.
Luttmer, Erzo F. P. (2001). ''Group Loyalty and the Taste for Redistribution,'' *Journal of Political Economy* 109(3).
Maier, K. (2000). *This House Has Fallen: Midnight in Nigeria*. New York: Public Affairs.
Marcus, H. (1994). *A History of Ethiopia*. Berkeley: University of California Press.
Mauro P. (1995). ''Corruption and Growth,'' *Quarterly Journal of Economics* 110(3), 681–712.
Minority Rights Group International. (1997). *World Directory of Minorities*. London: Minority Rights Group International.
Mozaffar, S., and J. Scarrit. (1999). ''The Specification of Ethnic Cleavages and Ethnopolitical Groups for the Analysis of Democratic Competition in Contemporary Africa,'' *Nationalism and Ethnic Politics* 5(1), 82–117.
Statistics New Zealand. (2001). 1996 Census (available on-line).
Tocqueville, A. de. (1990). *Democracy in America*, Vol. 2, New York: Vintage Books.
Tronvoll, K. (2000). *Ethiopia: A New Start?* Minority Rights Group International Report.
United States Census Bureau. (2001). 2000 Census (available on-line).