UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 09/04

Ethnic Diversity and Local Conflicts

By Michael Bleaney and Arcangelo Dimico

April 2009

2009 DP 09/04

Ethnic Diversity and Local Conflicts

Michael Bleaney and Arcangelo Dimico[1]

University of Nottingham

Abstract

We hypothesise that, given the typically uneven distribution of ethnic groups within a country, ethnic diversity leads to greater local polarization and more frequent, but smaller, conflicts that involve only some ethnic groups. These conflicts can be overlooked if the number of fatalities is small. Our empirical work exploits data on the proportion of a country affected by a conflict, and we control for country size, poverty, geography and natural resource endowments. We show that, consistent with the hypothesis, at the margin ethnic diversity makes conflict more probable, but also makes it more likely to be localized. This finding is robust to persistence in the incidence and extent of conflict. This potentially explains the apparent lack of correlation between ethnic fractionalization and the incidence of conflict found in previous research that uses a higher threshold number of fatalities to define a conflict.

JEL No.: O4

Keywords: conflict, ethnicity, fractionalization, polarization.

[1] Corresponding author: Arcangelo Dimico, School of Economics, University of Nottingham, e-mail:lexad5@nottingham.ac.uk

1) Introduction

There is considerable evidence that greater ethnic diversity is associated with weaker economic performance (Alesina *et al*., 2003; Easterly and Levine, 1997; Montalvo and Reynal-Querol, 2005b), but why? The usual answer is that more ethnically diverse societies suffer more serious internal conflicts. Barro (1991) shows that revolutions and coups have a strong negative effect on growth. It is unclear, however, that more ethnically diverse societies are more conflict-ridden in reality. There is some not entirely unambiguous evidence that ethnic diversity makes conflicts harder to stop (Collier *et al*., 2004; Fearon, 2004); there is no evidence that it makes conflicts more likely to start (Collier and Hoeffler, 1998, 2004; Fearon and Laitin, 2003, Collier *et al*., 2009). Montalvo and Reynal-Querol (2005a) find that ethnic diversity does not affect the overall incidence of conflict, but that ethnic polarization does (a country is said to be polarized if the second largest group makes up a substantial share of the population and is not much smaller than the largest group).

In this paper we pursue the idea that polarization between two substantial ethnic groups matters, but at the sub-national as well as the national level. Different regions of ethnically diverse societies typically have very different ethnic composition, because ethnic groups tend to be geographically concentrated (consider, for example, the former Yugoslavia). Consequently, more ethnically diverse countries may have pockets of polarization that tend to give rise to ethnic tensions and conflicts. Moreover, these conflicts may not be very contagious to neighbouring regions if the neighbouring regions have a very different ethnic composition. If these conflicts do not result in many fatalities, they may not have been counted as such in previous research.[2] As a result, the impact of ethnic diversity on the incidence of conflict may have been significantly underestimated.

In this paper we use data on the proportion of a country affected by conflict provided by Political Instability Task Force (2008) [PITF]. To our knowledge this geographical dimension of conflict has not previously been explored. The data cover all countries from 1955 to 2007. Since the PITF data impose no minimum threshold on the number of fatalities required to identify a conflict, they should capture significant conflicts of any magnitude. Using these data, we test our hypotheses that (a) ethnic diversity increases the incidence of conflicts, and (b) it causes these conflicts to be more localized. We find strong empirical support for both hypotheses.[3]

The paper is organised in the following way. In Section 2 we discuss measures of ethnic diversity. In Section 3 we introduce the data and describe our econometric approach. Section 4 presents our main results. In Section 5 we test the robustness of our results controlling for a possible persistence in the in extensiveness of the conflict. The paper ends with a short conclusion.

---

[2] Criteria which have been used to define conflicts include a minimum of 1000 deaths annually (Singer and Small, 1982; 1994; Collier and Hoeffler, 2004) and a minimum of 100 deaths annually and 1000 cumulatively (Fearon and Laitin, 2003). Montalvo and Reynal-Querol (2005a) consider various criteria, all with a minimum of 1000 cumulative deaths.

[3] There is also some information on the geographical extent of conflict in the Peace Research Institute Oslo data set, but this is less detailed since it relates only to the maximum extent of any conflict, and not its extent in the specified year. The data are also puzzling since in quite a few cases they seem to imply that considerably more than 100% of the country is affected (applying the formula for the area of a circle from the radius of the conflict given in the data).

2) Measuring Ethnic Diversity

The most commonly used measure of ethnic diversity is ethnic fractionalization, which is equal to one minus a Herfindahl index of concentration:

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

where π_i represents the population share of group *i*. This has a minimum of zero, when there is just one group, and a maximum of one. In practice it is strongly negatively correlated with the share of the largest group.

Esteban and Ray (1994) argue that conflict is most likely when the largest group is faced by a substantial minority group. Based on this, Montalvo and Reynal-Querol (2005a) use the following index of ethnic polarization:

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

Polarization reaches a maximum of one when there are just two equally sized groups, and is in practice strongly positively correlated with the share of the second largest group. Polarization diminishes at high levels of fractionalization, and also as the share of the dominant group becomes very large. It is strongly positively correlated with fractionalization when the two largest groups form a large share of the population, but tends to be negatively correlated with fractionalization if the share of the two largest groups is relatively low.[4]

The available data sets refer only to polarization (or fractionalization) at the national level. The relationship between this number and local polarization can be complex. Polarization at the local level will depend not just on the national population shares of each ethnic group but also on how geographically concentrated they are. A plausible hypothesis is that the probability of conflict is greater when the national polarization measure is higher, but also when local polarization is higher relative to national polarization. If two large groups are geographically separated, it is possible that polarization could be low at the local level but high at the national level. On the other hand, if ethnic fractionalization is high, local polarization could be considerably higher than national polarization. For example, suppose that there are ten groups each representing exactly 10% of the population. Then $F = 0.90$ and $P = 0.36$, which is not very high. If, however, these groups are unequally distributed across the country, so that in each region there are only five groups each representing 20% of the population, then local polarization is 0.64 in every region, which implies that the average local polarization is considerably greater than the national polarization figure. Thus the combination of high ethnic fractionalization and an uneven regional distribution of ethnic groups is likely to result in relatively high polarization levels in at least some regions. If polarization at the local as well as the national level increases

[4] If there are only two groups, P is precisely equal to twice F. If there are n groups of equal size, then as n tends to infinity, F tends to one but P tends to zero. In our sample the share of the two largest groups averages 81%. For cases where the share exceeds 81%, the correlation between F and P is 0.61, but for cases where the share is less than 81%, the correlation is only 0.09. See Montalvo and Reynal-Querol (2005a) for graphs of the relationship between F and P.

the incidence of conflict, then high ethnic fractionalization will also tend to be associated with an increased incidence of local conflicts. We test this hypothesis using data on the proportion of the country affected by conflict in any given year.

To investigate the relationship between *average* local polarization and ethnic fractionalization more formally, consider cases where at least one ethnic group is not represented in any given region. We analyse two possibilities: (1) where the regional population shares are formed from the national population shares by adding the missing group's share entirely to that of one other group; and (2) where the share of the missing group is redistributed equally amongst all the represented groups.

In the first case, as shown in the Appendix, the merging of any two groups with shares π_i and π_j will change polarization by

$$\Delta P = 4\pi_i \pi_j [2 - 3(\pi_i + \pi_j)] \tag{3}$$

which is positive if the two groups' shares sum to less than two-thirds, but negative if they sum to more than two-thirds. Once the number of groups exceeds three, and if the two largest groups make up less than two-thirds of the population, then all regions will have greater polarization than the national measure. Thus in this case it is highly likely that average local polarization exceeds national polarization when ethnic fractionalization is high.

In the second case, it is shown in the Appendix that elimination of group j and redistribution of its share equally amongst the n remaining groups results in a change of polarization of:

$$\Delta P = 4am^2[3F - 1 - a(1 + m + 3m^2) + a^2 m(1 + 3m + 2m^2)] \tag{4}$$

where $m = n^{-1}$, $a = n\pi_j$ and F is the index of ethnic fractionalization defined in equation (1). The parameter a reflects the relative size of the excluded group j (which is equal to the average of the others when $a = n/(n+1)$). Equation (4) is more likely to be positive when F is high, but it is also negatively related to the size of the eliminated group (a). For a=1, equation (5) reduces to

$$\Delta P = 4m^2(3F - 2 + 2m^3)\,. \tag{5}$$

This is always satisfied if $F > ⅔$, which is likely if $n \geq 3$ and the largest group's share is less than 50%. Thus, when regional shares are constructed by setting one group's share to zero and adding the remainder equally to all the other groups' national shares, average local polarization will exceed national polarization when F is high.

On the other hand, if there are only two groups (so $F \leq 0.50$), deviations of regional population shares from the national population shares imply that the population-weighted average of regional polarization is *below* the national figure, although regional polarization will be greater than the national figure in some regions.[5]

[5] For example, if there are two equally sized regions, then average regional polarization is smaller than national polarization by $2x^2$, where x is the deviation of a group's regional population share from its national

The implication of the above is that the *average* of regional polarization across the country tends to be higher relative to the national figure when ethnic fractionalization is high. Thus higher ethnic fractionalization implies a greater probability of pockets of local polarization that may give rise to local conflicts. In the absence of regional data on ethnic composition, we use ethnic fractionalization as a proxy for the difference between average regional polarization (*LP*) and national polarization (*NP*). This deviation tends to be negative when *F* is low and positive when *F* is high.

We hypothesize that the incidence of conflict is increasing in *NP* and in *LP*, but that the proportion of the country affected is decreasing in *LP*, for given *NP*, because if *NP* is low, some regions may be very unpolarized, or have ethnic composition very different from those where there is conflict, and so remain unaffected. Thus we expect the extent of conflict to be negatively related to the degree of ethnic fractionalization.

3) Data and Model

Our data source is Political Instability Task Force (2008), which identifies a total of 145 conflicts globally during the period 1955-2007. Of these, 80 are defined as ethnic conflicts and 65 as non-ethnic. Since we are interested in the relationship between ethnic diversity and the overall pattern of conflict, we ignore this distinction. A conflict is defined to occur

average. This is the result of the non-linearity of the polarization index, which becomes more sensitive to variations in the size of the largest group as that size deviates from 0.5.

if the government is challenged by an organised group[6], whether or not there are fatalities involved. The total number of conflict-year episodes is 970.

The PITF also records how much of the country is affected by a conflict. The coding rule is "*based on source materials about how much of the country is directly or indirectly affected by fighting or revolutionary protest in a given year. A province, region, or city is "directly affected" if fighting/terrorist attacks/revolutionary protest occur there at any time during the year. It is "indirectly affected" if the area has significant spillover effects from nearby fighting, for example refugee flows, curtailment of public services, martial law imposed. If open conflict expands or contracts during the course of the year, code according to its greatest extent*". The extent of the conflict is recorded in one of five categories according to the proportion of the surface area of the country affected: (1) less than 10% and no significant cities; (2) less than 10% and at least one significant provincial city; (3) between 10% and 25% and/or the capital city; (4) between 25% and 50%; and (5) over 50%. Since data on the extent of the conflict are not given for genocides, we exclude genocides from the conflict data.[7]

Because the only difference between the first two categories is the involvement (or not) of significant cities, we merge these two categories into one. Thus our coding rule (denoted MAGAREA) is:

- 0 if no conflict occurred;

[6] This definition excludes genocides that do not involve the government and a rebel group. The PITF data also include information on the incidence of genocides, but not on the proportion of the country affected by them.
[7] In a later section we show that this makes little difference to our equation for the incidence of conflict.

- 1 if less than 10% of the country is directly or indirectly affected;

- 2 if between 10% and 25% of the country or the capital city is directly or indirectly affected

- 3 if between 25% and 50% of the country and/or most major urban areas are directly or indirectly affected;

- 4 if more than 50% of the country is directly or indirectly affected.

Table 1 shows the distribution of the variable in our dataset. Conflicts occurred in 13.78% of all country-years. Their extent is fairly evenly spread across the categories, with less extensive conflicts being somewhat more common.

Table 1: Descriptive Statistics on the Extent of Conflict

MAGAREA	Frequency	Percentage	Cumulative Percentage
0	6,070	86.22	86.22
1	273	3.88	90.10
2	309	4.39	94.49
3	197	2.80	97.29
4	191	2.71	100.00
Total	7040	100	

Table 2: Descriptive Statistics for Types of Conflicts

Variable	Obs	Mean	Std. Dev.	Min	Max
Extent if the conflict is ethnic	588	2.039	0.990	1	4
Extent if the conflict is non-ethnic	382	2.740	1.083	1	4

Table 2 shows that ethnic conflicts tend to be somewhat smaller in extent than non-ethnic conflicts. Since, as Table 3 shows, the geographical extent of the conflict is positively correlated with the number of fatalities, Table 2 suggests that defining conflicts by a minimum number of fatalities, as in most previous research, is likely to exclude more ethnic than non-ethnic conflicts.

Table 3: Fatalities and Extent of Conflict

Number of Fatalities by Extent of Conflict		
Fatalities	Mean MAGAREA	Frequency
Less than 100	1.705	61
100 to 1000	1.866	351
1000 to 5000	2.434	371
5000 to 10000	3.390	87
More than 10000	3.056	71

Source: Political Instability Task Force; ***some more explanation needed**

Finally, in Table 4 we show how ethnic fractionalization and polarization vary with the extent of conflict. On average fractionalization is larger for country-years with any extent of conflict than for country-years with no conflict, and the same is true for polarization; but polarization tends to be slightly higher for extensive conflicts (MAGAREA = 3 or 4) than for less extensive ones (MAGAREA = 1 or 2), whereas the opposite is true for fractionalization. In either case, if a threshold number of fatalities is used to define a conflict, effectively shifting some country-years with MAGAREA > 0 into the MAGAREA = 0 category, this will bias upwards the average measure of ethnic diversity for country-years with no conflict, and thus bias downwards the effect of fractionalization or polarization on the incidence of conflict.

Table 4: Ethnic Fractionalization and Extent of Conflict

Magarea	Ethnic frac.	Ethnic Pol	Freq.
0	0.3686	0.5019	5913
1	0.5654	0.5750	272
2	0.5636	0.5656	309
3	0.4453	0.6082	197
4	0.4836	0.6095	190
Total	0.3905	0.5136	6881

We initially estimate a simple ordinal probit model just for the cases where MAGAREA > 0, to check whether ethnic fractionalization significantly affects the extent of conflicts. The problem with a simple ordinal probit model is that we have pre-selected the cases where a conflict occurs, which may introduce some bias into the estimates. In econometric terms, we have a selection problem, and as a consequence, a correction for the non-zero expectation of the error term in the outcome equation is needed in order to eliminate the omitted-variable bias. Heckman (1979) proposes a simple two-step estimator in order to tackle possible selection biases. He proposes to estimate a first step in which we predict the probability that such an observation is selected (in this case, whether a conflict occurs). In the second step, we estimate the outcome equation (the extent of conflict) conditional on the probability of being selected. Heckman (1979) shows that the two steps estimator provides consistent estimates under the assumption that disturbances are normally jointly distributed.

So, in order to check the reliability of results in the simple ordinal probit model, we also estimate a Heckman two-step estimator. In the first step we estimate a probit model for the incidence of conflicts which will allow us to evaluate variables which affect the occurrence of conflicts. Using predicted values from this first stage model, we calculate the inverse Mills ratio as a ratio of the probability distribution function to the cumulative distribution function. In the second stage we estimate an ordinal probit model that controls for selection bias by incorporating the inverse Mills ratio as an explanatory variable. The *t*-statistic of this ratio represents a test of the null that the two-step procedure is unnecessary. Bootstrap standard errors are computed in order to take into account of the sampling error related to the predicted variable.

The baseline ordinal probit model can be written as:

$$y_i * = \beta X'_{i.} + \mu_i \qquad \text{for } i = 1, .., n,$$

where y_i* is the latent variable and *X'* are some independent variables. Ethnic fractionalization, ethnic polarization, income per capita, population, oil exports and alluvial deposits of diamonds are some of the variables we use in our benchmark model. Data on ethnic fractionalization are from Fearon and Laitin (2003) [FL], and ethnic polarization data are from Montalvo and Reynal-Querol (2005a).[8] Oil export is a dummy variable which is coded one if the ratio of oil exports to manufacturer exports is greater than one third, and zero otherwise. This variable is also from FL. Data on income per capita and population are provided by the World Bank. We use data on diamond deposits from Lujala *et al.* (2005) to construct a dummy variable which is one if the deposits are alluvial and economically exploitable, and zero otherwise. In the selection equation, we also use the proportion of the country with mountainous terrain (source: FL), which has been shown in previous research to be associated with an increased likelihood of conflict.

4) Results

Table 5 presents results for the ordinal probit model. The effect of fractionalization in Model 1 is significant at a 1% level and its sign is negative. Therefore, consistent with our hypothesis, societies which are more fractionalized tend to experience less extensive conflicts. Marginal effects suggest that the probability that less than 1/10 of the country is affected by the conflict is 17% for minimally fractionalized societies, and 33% for societies maximally fractionalized. The extent of conflict (as a proportion of the country) is also negatively related to a country's size, as measured by its population, and its per capita income level, but tends to be larger for oil exporters.

[8] The polarization data are not available for all countries, so we impute the missing numbers using data on the share of the second largest group provided by FL. The correlation between polarization and the share of the second largest group is 0.77.

In Model 2 we replace the index of ethnic fractionalization with polarization, which is not significant, but when we enter polarization together with the fractionalization index[9] (Model 3) the former also becomes significant. For a given degree of fractionalization, a greater degree of polarization increases the extent of conflict.

Table 5: Ordinal Probit Model

Dependent Variable: Extent of Conflict			
Estimation Method: Ordinal Probit	Model 1	Model 2	Model 3
Population (t-1)	-0.294***	-0.297***	-0.284***
	(-9.62)	(-9.60)	(-9.07)
GDP per capita (t-1)	-0.226***	-0.177***	-0.242***
	(-5.33)	(-4.51)	(-5.62)
Oil Export Dummy	0.179*	0.188**	0.213**
	(1.89)	(2.02)	(2.28)
Alluvial Diamonds Dummy	0.103	0.0072	0.105
	(1.11)	(0.08)	(1.14)
Ethnic Fractionalization	-0.600***		-0.665***
	(-4.15)		(-4.66)
Ethnic Polarization		0.281	0.469**
		(1.38)	(2.28)
Cutoff point 1	-5.496***	-4.696***	-5.275***
	(-13.10)	(-11.09)	(-11.97)
Cutoff point 2	-4.568***	-3.778***	-4.343***
	(-10.99)	(-8.98)	(-9.94)
Cutoff point 3	-3.917***	-3.136***	-3.688***
	(-9.49)	(-7.50)	(-8.48)
Observations	959	959	959

1) Robust z-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

[9] The correlation between the imputed measure of polarization and ethnic fractionalization is 0.49.

Table 5 shows the results using the two-step estimator. Panel A shows the outcome equation, where the dependent variable is the extent of conflict, and Panel B shows the selection equation, where the dependent variable is the incidence of conflict.

In Panel A the significant negative coefficient on the inverse Mills ratio indicates a negative correlation between factors which affect the incidence of conflicts and those which influence the extent of conflicts. In Model 1 of Panel A ethnic fractionalization has a highly significant negative coefficient, as does population, but per capita GDP is now insignificant. Exploitable deposits of alluvial diamonds appear to be associated with more extensive conflicts. In Model 2, fractionalization is replaced by polarization, which also has a significant negative coefficient. In Model 3, when both are entered together, both have significant negative coefficients, although fractionalization appears to be the more important. Thus greater ethnic diversity tends to be associated with less geographically extensive conflicts.

Table 6: Two-Step Heckman Estimates (Outcome Eq.)

	PANEL A: SECOND-STEP REGRESSION (OUTCOME EQUATION)		
Dependent Variable: Extent of Conflict	Model 1	Model 2	Model 3
GDP per capita (t-1)	0.0567	0.324**	0.152
	(0.68)	(2.49)	(1.29)
Population (t-1)	-0.526***	-0.680***	-0.605***
	(-8.35)	(-6.80)	(-6.59)
Oil Export Dummy	-0.065	-0.296**	-0.177
	(-0.58)	(-2.12)	(-1.32)
Alluvial Diamonds Dummy	0.205**	0.136	0.251***
	(2.19)	(1.46)	(2.52)
Ethnic Fractionalization	-1.120***		-1.029***
	(-6.53)		(-6.04)
Ethnic Polarization		-1.389***	-0.810**
		(-3.28)	(-2.26)
Inverse Mills Ratio	-1.186***	-1.776***	-1.510***
	(-4.06)	(-4.10)	(-3.69)
Cutoff point 1	-7.528***	-8.074***	-8.440***
	(-11.77)	(-8.28)	(-9.28)
Cutoff point 2	-6.591***	-7.150***	-7.502***
	(-10.31)	(-7.36)	(-8.25)
Cutoff point 3	-5.932***	-6.503***	-6.843***
	(-9.29)	(-6.69)	(-7.51)
Observations	959	959	959

1) Bootstrap (1000 reps) z-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1
2) The omitted variable is Mountainous Terrain.

Table 6 (continued): Selection Equation

Dependent Variable: Incidence of Conflict	***PANEL B: FIRST-STEP (SELECTION EQUATION)***		
	Model 1	Model 2	Model 3
GDP per capita (t-1)	-0.297***	-0.360***	-0.325***
	(-12.58)	(-14.91)	(-12.85)
Population (t-1)	0.242***	0.270***	0.263***
	(15.70)	(16.80)	(16.34)
Oil Export Dummy	0.320***	0.370***	0.353***
	(5.32)	(6.21)	(5.86)
Alluvial Diamonds Dummy	-0.0246	-0.0217	-0.0578
	(-0.42)	(-0.37)	(-1.00)
Ethnic Fractionalization	0.616***		0.368***
	(8.07)		(4.20)
Ethnic Polarization		1.002***	0.823***
		(11.23)	(8.05)
Mountainous Terrain	0.140***	0.0936***	0.105***
	(8.68)	(5.86)	(6.43)
Constant	-1.617***	-1.564***	-1.858***
	(-7.66)	(-7.42)	(-8.45)
Observations	6631	6631	6631

1) Robust z-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Panel B of Table 6 shows that the incidence of conflict is positively correlated with both ethnic fractionalization and ethnic polarization, and also with poverty, country size, being an oil exporter and having mountainous terrain. Polarization is more significant than fractionalization, consistent with the results of Montalvo and Reynal-Querol (2005a).

Table 7 reports marginal effects for Model 3 in Panel A of Table 6. All of the variables with significant negative coefficients, including ethnic fractionalization, make small conflicts (MAGAREA = 1 or 2) more likely and large conflicts (MAGAREA = 3 or 4) less likely, and *vice versa* for variables with significant positive coefficients.

Table 7: Marginal Effects

	MARGINAL EFFECTS			
	Pr(magarea=1) 0.2343	Pr(magarea=2) 0.3451	Pr(magarea=3) 0.2170	Pr(magarea=4) 0.2035
GDP per capita	-0.0470	-0.0121	0.0181	0.0410
	(-1.33)	(-1.36)	(1.32)	(1.36)
Population	0.187***	0.0484***	-0.0720***	-0.163***
	(5.99)	(3.88)	(-5.29)	(-6.55)
Oil Exporters*	0.0570	0.0105*	-0.0225	-0.0449
	(1.26)	(1.96)	(-1.23)	(-1.43)
Diamonds*	-0.0727***	-0.0264**	0.0261***	0.0730**
	(-2.85)	(-2.03)	(3.07)	(2.46)
Ethnic Frac.	0.318***	0.0823***	-0.122***	-0.278***
	(5.74)	(3.81)	(-5.08)	(-6.26)
Ethnic Polarization	0.250**	0.0647**	-0.0963**	-0.219**
	(2.12)	(2.03)	(-2.10)	(-2.17)
Inverse Mills Ratio	0.467***	0.121***	-0.180***	-0.408***
	(3.45)	(3.51)	(-3.28)	(-3.88)
Observations	959	959	959	959

1) Pr(magarea=1): Probability that less than 1/10 of the country is affected
2) Pr(magarea=3): Probability that a share between 1/10 and 1/4 is affected
3) Pr(magarea=3): Probability that a share between 1/4 and 1/2 is affected
4) Pr(magarea=4): Probability that more than 1/2 of the country is affected
5) (*) dy/dx is for discrete change of dummy variable from 0 to 1

5) Robustness Check

In this Section we test the robustness of our results in three ways: (1) by using a different exclusion restriction for the outcome equation; (2) by allowing for the persistence of conflicts; and (3) by testing the effect of ethnic diversity on the incidence of conflict using different data.

An alternative outcome equation

Table 8 shows the second-step estimates if we exclude the oil export dummy rather than mountainous terrain from the outcome equation, as we did in Table 6, keeping the selection equation unchanged. Ethnic fractionalization is still negative and significant at a 1% level, as is population. Ethnic polarization is the only variable which loses significance. The variable is marginally significant in Model 2, but not significant in Model 3. The effect of the share of mountainous terrain is insignificant (except in Model 2) which justifies our original decision of constraining its coefficient to be zero.

Table 8: Substituting Mountainous Terrain for the Oil Export Dummy in Table 6A

	PANEL A: SECOND STEP REGRESSION (OUTCOME EQUATION)		
	Dependent Variable: Extent of Conflict		
Estimation Method: 2-Step Heckman	Model 1	Model 2	Model 3
GDP per capita (t -1)	0.0073	0.039	0.0099
	(0.08)	(0.40)	(0.11)
Population (t -1)	-0.488***	-0.473***	-0.494***
	(-6.23)	(-5.78)	(-6.26)
Alluvial Diamond Dummy	0.194**	0.114	0.212**
	(2.02)	(1.23)	(2.22)
Ethnic Fractionalization	-1.020***		-0.862***
	(-4.63)		(-5.00)
Ethnic Polarization		-0.638*	-0.442
		(-1.95)	(-1.39)
Mountainous Terrain	0.021	0.090**	0.042
	(0.47)	(2.48)	(1.07)
Inverse Mills Ratio	-0.977***	-0.741**	-0.949***
	(-2.66)	(-2.22)	(-2.85)
Cutoff point 1	-7.132***	-6.084***	-7.235***
	(-8.34)	(-7.35)	(-8.12)
Cutoff point 2	-6.195***	-5.158***	-6.297***
	(-7.26)	(-6.25)	(-7.08)
Cutoff point 3	-5.536***	-4.509***	-5.637***
	(-6.51)	(-5.48)	(-6.34)
Observations	959	959	959

1) Bootstrap (1000 reps) z-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1.
2) In this Table Mountainous Terrain replaces the Oil Export Dummy in Table 6A.

Persistence of conflicts

Conflicts tend to be highly persistent: if a conflict occurred in country *j* in year *t*-1, it is much more likely that country *j* also experiences a conflict in year *t*. It is probable that the extent of conflict is also serially correlated. To allow for this, we re-estimate the model including lagged dummy variables indicating the state of the dependent variable in the previous year (whether or not there was a conflict for the selection equation, and the extent of the conflict for the outcome equation). The results are shown in Table 9.

As expected, these lagged variables turn out to be highly significant. In the outcome equation (Panel A), the coefficients of the lagged extent increase with MAGAREA (the omitted category is MAGAREA(t-1) = 1), indicating that the extent is positively serially correlated. In Models 1 and 3 both ethnic fractionalization and population have significant negative coefficients as before, but ethnic polarization is no longer significant either by itself (Model 2) or in combination with fractionalization (Model 3), unlike in Table 6.

In the selection equation (Panel B), both ethnic fractionalization and ethnic polarization still have a significant positive coefficient when entered individually (Models 1 and 2), but polarization dominates when they are entered together (Model 3), and fractionalization is no longer significant in this case, again unlike in Table 6.

These results show that our finding that ethnic diversity increases the incidence of conflicts but reduces their extent is robust to allowing for persistence in the incidence and extent of conflicts.

Table 9: Two-Step Heckman Estimates (Outcome Eq.)

PANEL A: SECOND STEP REGRESSION (OUTCOME EQUATION)

Dependent Variable: Extent of War	Model 1	Model 2	Model 3
GDP per capita (t-1)	-0.0747	-0.0456	-0.0792
	(-1.55)	(-1.07)	(-1.63)
Population (t-1)	-0.125***	-0.130***	-0.122***
	(-3.32)	(-3.56)	(-3.33)
Oil Export Dummy	0.0710	0.0643	0.0771
	(0.69)	(0.59)	(0.72)
Alluvial Diamonds Dummy	0.0491	0.00191	0.0493
	(0.42)	(0.02)	(0.44)
10-25 % of country affected by conflict (t-1)	1.853***	1.844***	1.856***
	(10.93)	(11.02)	(11.51)
25-50 % of country affected by conflict (t-1)	2.852***	2.866***	2.849***
	(12.97)	(13.52)	(13.76)
> 50 % of country affected by conflict (t-1)	4.106***	4.119***	4.104***
	(13.98)	(14.54)	(14.92)
Ethnic Fractionalization	-0.345**		-0.353**
	(-2.07)		(-2.06)
Ethnic Polarization		0.00672	0.104
		(0.03)	(0.40)
Inverse Mills Ratio	-0.380***	-0.372***	-0.375***
	(-5.38)	(-5.11)	(-4.48)
Cutoff point 1	-1.145*	-0.790	-1.093*
	(-1.95)	(-1.40)	(-1.84)
Cutoff point 2	0.463	0.812	0.516
	(0.76)	(1.39)	(0.84)
Cutoff point 3	1.548**	1.893***	1.601**
	(2.45)	(3.16)	(2.53)
			.
Observations	953	953	953

1) Bootstrap (1000 reps) z-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 9 (continued): Two-Step Heckman Estimates (Selection Eq.)

PANEL B: FIRST STEP (SELECTION EQUATION)			
Dependent Variable: Incidence of Conflict	Model 1	Model 2	Model 3
GDP per capita (t-1)	-0.144***	-0.173***	-0.157***
	(-4.53)	(-5.85)	(-4.88)
Population (t-1)	0.103***	0.116***	0.113***
	(3.87)	(4.33)	(4.19)
Oil Export Dummy	0.130	0.154	0.146
	(1.18)	(1.40)	(1.33)
Alluvial Diamonds Dummy	-0.000547	0.00597	-0.0146
	(-0.01)	(0.06)	(-0.14)
Dummy for Conflict in Year t-1	3.301***	3.291***	3.288***
	(44.88)	(44.69)	(44.61)
Ethnic Fractionalization	0.285**		0.177
	(2.18)		(1.21)
Ethnic Polarization		0.469***	0.386**
		(3.02)	(2.19)
Mountainous Terrain	0.0703***	0.0479*	0.0541*
	(2.64)	(1.73)	(1.96)
Constant	-2.217***	-2.189***	-2.331***
	(-6.49)	(-6.61)	(-6.72)
Observations	6564	6564	6564

1) Robust z-statistics in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Alternative conflict data

Finally we check the robustness of our results using alternative sources of data on the incidence of conflict. In Table 10 we provide some descriptive statistics on alternative datasets for the incidence of conflict. Sambanis (2004), Fearon and Laitin (2003), the Political Instability Task Force, and the Peace and Research Institute Oslo are the sources of our data. For the Political Instability Task Force we consider two different samples: one that excludes genocides (labelled PITF), and one that includes them (labelled PITF + Genocide).[10] For data from the Peace Research Institute Oslo we use the sample with the lowest death-threshold (25 deaths per year). Table 10 shows that the number of observations varies from 6194 in Sambanis (2004) to 7040 in the PITF dataset. The pairwise correlations between these data sets all lie within the range 0.70 to 0.83, which is high, but not so high as to generate identical results for the empirical analysis of conflict incidence.

Table 11 reproduces Model 3 of Table 8B for the incidence of conflict, including both fractionalization and polarization as regressors, using these different data sources. Polarization is always more significant than fractionalization, and is significant at at least the 10% level in every case, even after allowing for the persistence of conflicts. All data sets show conflicts to be significantly more likely in poorer, more populous states, but mountainous terrain is not always significant. In short, the determinants of incidence appear similar across data sets.

[10] Our previous analysis used PITF rather than PITF + Genocides for lack of data on the extent of genocides.

Table 10: Descriptive Statistics for Different Sources of Data

Descriptive Statistics For the Incidence of Conflict					
Variable	Obs	Mean	Std. Dev.	Min	Max
PITF + Genocide	7040	0.1571023	0.363923	0	1
PITF	7040	0.1380682	0.344996	0	1
PRIO (25 deaths)	6916	0.1625217	0.3689554	0	1
Fearon & Laitin	6273	0.1442691	0.3513904	0	1
Sambanis	6194	0.146432	0.3535674	0	1

Correlation For the Incidence of Conflict					
	PITF + genocide	Pitf	PRIO(25)	Fearon	Sambanis
PITF + Genocide	1.0000				
PITF	0.9225	1.0000			
PRIO (25 deaths)	0.6989	0.7189	1.0000		
Fearon and Laitin	0.7499	0.7536	0.7410	1.0000	
Sambanis	0.7992	0.8053	0.7395	0.8341	1.0000

Table 11: Incidence of Conflict With Alternative Datasets

	Dependent Variable: Incidence of Conflict				
Estimation Method: Probit	***PITF***	***PITF + Genocide***	***Fearon***	***Sambanis***	***PRIO (25 deaths)***
Log GDP per Capita (t-1)	-0.157***	-0.167***	-0.156***	-0.158***	-0.131***
	(-4.88)	(-5.20)	(-4.54)	(-4.75)	(-4.93)
Log Population (t-1)	0.113***	0.113***	0.127***	0.102***	0.116***
	(4.19)	(4.21)	(4.59)	(3.93)	(5.80)
Ethnic Fractionalization	0.177	0.174	0.227	0.152	0.268**
	(1.21)	(1.17)	(1.56)	(0.99)	(2.34)
Ethnic Polarization	0.386**	0.313*	0.361**	0.361*	0.475***
	(2.19)	(1.87)	(2.05)	(1.94)	(3.46)
Oil Export Dummy	0.146	0.103	0.0902	0.119	0.169**
	(1.33)	(0.95)	(0.73)	(1.06)	(2.03)
Alluvial Diamonds Dummy	-0.0146	-0.0323	-0.0033	-0.0902	-0.116
	(-0.14)	(-0.32)	(-0.03)	(-0.86)	(-1.45)
Mountainous Terrain	0.0541*	0.0519*	0.0670**	0.0365	0.0319
	(1.96)	(1.89)	(2.38)	(1.42)	(1.51)
PITF Incidence (t-1)	3.288***				
	(44.61)				
PITF+Genocide Incidence (t-1)		3.387***			
		(46.10)			
Fearon-Laitin Incidence (t-1)			3.381***		
			(42.29)		
Sambanis Incidence (t-1)				3.232***	
				(43.56)	
PRIO (25) Incidence (t-1)					2.681***
					(45.45)
Constant	-2.331***	-2.186***	-2.504***	-2.101***	-2.290***
	(-6.72)	(-6.45)	(-6.87)	(-5.92)	(-7.92)
Observations	6564	6564	6027	5800	6543

Robust z statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

6) Conclusions

Previous research has failed to find much evidence that ethnic diversity increases the incidence of civil conflict, although Montalvo and Reynal-Querol (2005a) show that ethnic polarization (a sizeable second largest ethnic group) is associated with a significantly higher incidence of conflict. Because the regional distribution of ethnic groups within a country can be very uneven, an ethnically fractionalized country may be quite polarized in some localities, even if its national polarization measure is not particularly high. This may give rise to local conflicts that tend not to spread to other areas with very different ethnic composition. These local conflicts may not cause sufficient fatalities (either annually or cumulatively) to be coded as conflicts by some criteria that have been used in academic research, in which case the impact of ethnic diversity on the incidence of conflict will have been underestimated.

Using a new data set that records the proportion of the country affected by a conflict in each year, and imposing no minimum threshold on the number of fatalities required to identify a conflict, we have demonstrated that ethnic diversity does lead to a greater incidence of conflicts, but also that it makes these conflicts more localised and less likely to spread to the whole country. The incidence of conflicts is more closely associated with polarization than with fractionalization, but the extent of the conflict is more closely negatively correlated with ethnic fractionalization. Our conjecture is that the latter effect reflects the greater frequency of pockets of local polarization when fractionalization is high.

Variables and Sources of Data

Incidence of Conflict: Political Instability Task Force. Conflicts are deemed to occur when an organised group challenges the government, with or without fatalities.

Extensiveness of Conflict and Number of Fatalities: Political Instability Task Force (2008). The PITF uses a categorical variable to proxy the share of the country affected by the conflict. More details are available in the paper.

Ethnic Fractionalization, Largest Ethnic Group, Second Largest Group, Mountainous Terrain, the Ratio of Oil Export to Manufacturer Exports: Fearon and Laitin (2003)

GDP per Capita and Population: World Bank (WDI (2008))

Alluvial diamonds: Lujala *et al.* (2005).

Ethnic Polarization: Montalvo and Reynal-Querol (2005a).

Appendix

Merging of two groups

If groups i and j merge, then

$$\Delta P = 4[(\pi_i + \pi_j)^2(1-\pi_i-\pi_j) - \pi_i^2(1-\pi_i) - \pi_j^2(1-\pi_j)] \qquad \text{(A1)}$$

which simplifies to equation (3) above.

Elimination of one group with its share equally divided between the remaining groups

Let the eliminated group be group 1, and let there be n other groups.
Writing $m = 1/n$, then

$$\Delta P = 4\sum_{i=2}^{n+1}[(\pi_i + m\pi_1)^2(1-\pi_i-m\pi_1) - \pi_i^2(1-\pi_i)] - 4\pi_1^2(1-\pi_1) \qquad \text{(A2)}$$

Expanding this gives

$$\Delta P = 4\sum_{i=2}^{n+1}(2m\pi_i\pi_1 - 3m\pi_i^2\pi_1 - 3m^2\pi_i\pi_1^2 + m^2\pi_1^2 - m^3\pi_1^3) - 4\pi_1^2(1-\pi_1) \quad \text{(A3)}$$

Substituting

$$\sum_{i=2}^{n+1} 1 = 1/m, \sum_{i=2}^{n+1}\pi_i = 1-\pi_1, \sum_{i=2}^{n+1}\pi_i^2 = 1 - F - \pi_1^2,$$

we obtain

$$\Delta P = 4m\pi_1(3F-1) - 4(1+m+3m^2)\pi_1^2 + 4(1+3m+2m^2)\pi_1^3 \quad \text{(A4)}$$

Substituting $\pi_1 = am$ yields equation (4).

References

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

Barro, Robert J. (1991), "Economic Growth in a Cross-Section of Countries," *Quarterly Journal of Economics*, 106 (2), 407-433.

Collier, Paul, and Anke Hoeffler (1998) "On Economic Causes of Civil War," *Oxford Economic Papers*, October, 50(4), 563-573

Collier, Paul, and Anke Hoeffler (2004) "Greed and Grievance in Civil War," *Oxford Economic Papers*, October, 56 (4), 563-595.

Collier, Paul, Anke Hoeffler and Måns Söderbom (2004), "On the Duration of Civil War," *Journal of Peace Research*, 41(3), 253-273.

Collier, Paul, Anke Hoeffler and Dominic Rohner (2009), "Beyond Greed and Grievance: Feasibility and Civil War," *Oxford Economic Papers*, January, 61(1), 1-27

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

Esteban, Joan and Debraj Ray (1994) "On the Measurement of Polarization," *Econometrica*, 62(4), pp. 819-51.

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

Fearon, James D. (2004), "Why Do Some Civil Wars Last So Much Longer than Others?" *Journal of Peace Research*, 41 (3), 275-301.

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

Heckman, J. (1979) "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161

Hudson, **Michael C. and Charles L. Taylor** (1972) "*World Handbook of Political and Social Indicators*," Second Edition, Yale University Press.

Lujala, Paivi, Nils P. Gleditsch, and Elisabeth Gilmore (2005) "A Diamond Curse?: Civil War and a Lootable Resource," *Journal of Conflict Resolution*, August, 49 (4), 538-56

Montalvo, Josè and Marta Reynal-Querol (2005a) "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95(3), 769-813.

Montalvo, Josè and Marta Reynal-Querol (2005b) "Fractionalization, Polarization and Economic Development," *Journal of Development Economics*, 76(2), 293-323.

Peace Research Institute Oslo, "Armed Conflict Dataset: Version 4-2007", www.prio.no/CSCW/Datasets/Armed-Conflict/UCDP-Prio/

Political Instability Task Force, "Internal Wars and Failure of Governance, 1955-2006" 2008, http://globalpolicy.gmu.edu/pitf/pitfcode.htm

Sambanis, Nicholas (2004), "What is Civil War? Conceptual and Empirical Complexities of an Operational Definition," *Journal of Conflict Resolution*, vol. 48(6), pages 814-59

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

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