Bleaney, Michael; Dimico, Arcangelo

Working Paper

Incidence, onset and duration of civil wars: A review of the evidence

CREDIT Research Paper, No. 09/08

Provided in Cooperation with:
The University of Nottingham, Centre for Research in Economic Development and International Trade (CREDIT)

Suggested Citation: Bleaney, Michael; Dimico, Arcangelo (2009) : Incidence, onset and duration of civil wars: A review of the evidence, CREDIT Research Paper, No. 09/08, The University of Nottingham, Centre for Research in Economic Development and International Trade (CREDIT), Nottingham

This Version is available at:
https://hdl.handle.net/10419/65452

CREDIT Research Paper

No. 09/08

Incidence, Onset and Duration of Civil Wars: A Review of the Evidence

by

Michael Bleaney and Arcangelo Dimico

Abstract

We investigate the determinants of onset, duration and incidence of civil wars, and their sensitivity to different coding rules. Whatever the coding rule used, incidence of civil war is largely determined by poverty, country size, mountainous terrain and ethnic diversity. Poverty reduces the opportunity cost of rebellion, and mountainous terrain makes rebellions harder to defeat, while ethnic diversity creates potential conflict, since some linguistic groups may feel under-represented.

Continuation and onset of civil war are non-overlapping subsets of incidence, depending on whether civil war occurred in the previous year. We exploit this aspect to test whether the determinants of continuation and onset are statistically significantly different. For four out of five coding rules, we cannot reject the null hypothesis of common determinants of civil war onset and continuation. This is the main contribution of the paper.

Ethnic diversity is particularly high in Africa, where many civil wars occur, but we show that the significance of ethnic diversity is not just an unidentified Africa effect. Our estimated model works as well for Africa as for other areas.

There is evidence that ethnic polarization matters as well as fractionalization. We show that polarization understates the propensity for conflict in societies with high ethnic diversity, where the polarization measure tends to be low, but otherwise outperforms fractionalization as a predictor of conflict. Mountainous terrain is generally not significant when ethnic polarization is included.

Centre for Research in Economic Development and International Trade, University of Nottingham

CREDIT Research Paper

No. 09/08

Incidence, Onset and Duration of Civil Wars: A Review of the Evidence

by

Michael Bleaney and Arcangelo Dimico

Outline

The Authors: Michael Bleaney and Arcangelo Dimico, School of Economics, University of Nottingham

Introduction

In recent years there has been considerable academic interest in the question of when and why countries experience civil wars (e.g. Collier and Hoeffler 1998, 2002, 2004, 2007; Collier *et al*., 2004, 2009; Elbadawi & Sambanis, 2002; Fearon, 2004, 2005; Fearon and Laitin, 2003; Hegre & Sambanis 2006; Montalvo and Reynal-Querol, 2005, 2008), and also in their severity as measured by casualties (Lacina, 2006; Lujala, 2009). Nevertheless this research has failed to reach definitive conclusions. Collier and Hoeffler and their collaborators have tended to emphasise economic motivations for rebellion, and the role of oil and gemstones has been investigated by various authors (Lujala *et al*. (2005); Lujala, 2010; Ross, 2006). Since many conflicts have an ethnic dimension, measures of ethnic diversity have figured commonly in civil war regressions, with mixed results. Montalvo and Reynal-Querol (2005) argue that civil war tends to be associated with ethnic polarization rather than fractionalization, and Schneider and Wiesehomeier (2008) consider the interaction of ethnic diversity with political structures. Fearon (2005) and Fearon and Laitin (2003) claim that political factors dominate either of these explanations.

One reason for the variation in results is that there is no uniform criterion for determining whether a country is experiencing a civil war in a given year, as is discussed at length by Sambanis (2004), who also examines the robustness of the determinants of the onset of a civil war to different coding rules. A second source of variation is that studies have typically investigated only one of three possible aspects of civil war: *incidence* (whether civil war *occurs* at a given date in a given country), *onset* (whether a civil war *starts* in a given period), and *duration* (whether a civil war continues, having already occurred in the previous period). Most studies focus on onset, but a few consider incidence (Elbadawi and Sambanis, 2002; Lujala *et al*., 2005; Montalvo and Reynal-Querol, 2005, 2008) or duration (Buhaug *et al*., 2009; Collier *et al*., 2004; Fearon, 2004). A further complication is that some data sets used to analyze onset and duration identify separate conflicts in a given country, more than one of which may be ongoing at any particular date.

There is a widespread impression that the determinants of civil war look rather different according to the data set used and the aspect (incidence, onset or duration) analyzed. The main purpose of this article is to test this rigorously. In the main, we find the opposite – that the determinants of civil war are remarkably consistent across data sets and aspects. In the second half of the paper we look more deeply into the ethnic element. We show that the role of ethnic diversity is not just an unidentified sub-Saharan African effect (although ethnic diversity is particularly high in sub-Saharan Africa and conflict is more frequent there), and we also demonstrate that both polarization and fractionalization matter.

Our analysis is based on two principles. One is that a country which is experiencing a civil war is in a worse state, from an economic and social point of view, than one which is not. In that case, the single most important aspect of civil war for study is incidence (those who prefer to focus on onset and duration may regard incidence as a special case where the determinants of onset and duration are the same, as we discuss below). The other principle is that onset and continuation of civil wars may be considered as subsets of incidence: if a civil war occurs in country *j* in year *t*, the country was either in a state of civil war at time *t-1* (continuation), or it was not (onset). Although empirical studies of onset greatly outnumber those of duration, it is not immediately obvious why this is the case: even though switches of state from peace to war and *vice versa* are rare, it matters a great deal whether a civil war lasts for ten years or twenty. Collier *et al*. (2004) and Fearon (2004) conclude that there are substantial differences in the determinants of onset and duration, but it is possible that this conclusion simply reflects sampling error. Since onset and continuation are subsets of incidence, standard statistical tests of equality of coefficients will tell us whether the differences in coefficients are statistically significant. This is one of the main questions which we wish to investigate: are the determinants of incidence, onset and duration significantly different?

The second purpose of our investigation is to determine how much our results vary across coding rules for defining civil wars. We use five different coding rules that are representative of those studied by previous researchers, as discussed further below.

Our results are reassuring. The main determinants of the incidence of civil war – poverty, a large population, mountainous terrain and ethnic diversity – are consistently significant across coding rules, even though there can be substantial disagreements between the rules about when a country is in a state of civil war (the pairwise correlation averages 0.72). Moreover, in the majority of cases, the influence of these variables is similar whether or not there was a civil war going on in the previous period. In other words, we do not find strong evidence that these factors affect onset and duration significantly differently.

The paper is structured as follows. In the next section we describe our empirical approach, introducing the variables used and the sources of our data. In Section Three we present our main empirical results for incidence. In Section Four we test for significant differences in the determinants of onset and continuation of civil wars. In Section Five we consider whether the relatively high frequency of conflict in sub-Saharan Africa explains the ethnic diversity effect. In Section Six we examine the ethnic diversity effect in more detail, distinguishing between fractionalization and polarization. Some brief conclusions are presented in Section Seven.

2) Empirical Approach and Data Description

Our basic model is a probit analysis of the incidence of civil war, where the dependent variable (Y_{jt}) takes the value one if a civil war occurs in country j in year t, and zero otherwise. If X_{kjt} is a k-vector of explanatory variables, then the following equation is estimated by maximum likelihood:

$$\Pr(Y_{jt} = 1) = \Phi(X_{kjt}\beta_k) \qquad (1)$$

where $\Phi(.)$ represents the cumulative normal distribution function.

We estimate equation (1) using five different definitions of civil war: those of Sambanis (2004), Fearon & Laitin (2003) [FL], the Political Instability Task Force (2008) [PITF], the UCDP/PRIO Armed Conflict Dataset [UCDP/PRIO] (Harbom and Wallensteen, 2007), and Collier *et al*. (2009) [CHR]. All of these sources define a civil war as a conflict between the government and an organized rebel group, but they differ mainly with respect to the death-threshold applied to define a civil war.

Of the five sources, CHR use the strictest death threshold, based on data for the Correlates of War (COW) project.[1] A civil war is only recorded when the battle-related deaths exceed 1,000 per year. Over the period 1960-2004 this gives 554 war-year episodes and 96 onsets.[2] Sambanis and FL have a lower annual death threshold, but combine it with a cumulative-death criterion. Sambanis codes a conflict if annual deaths exceed 100 and the cumulative number of deaths is above 1,000, and if there is no three-year period during which the conflict causes fewer than 500 deaths. This definition yields 920 war-year episodes and 118 onsets over the period 1945-1999.[3] Fearon and Laitin use similar criteria to Sambanis, but without the minimum of 500 deaths in a three-year period. They count 918 war-years and 111 onsets over the period 1945-1999 is 111[4]. The PITF[5] and the UCDP/PRIO do not use any strict battle-related death-threshold. In order to code a conflict the PITF requires a mobilization threshold (at least 1,000 people), and the condition that in at least one year the conflict-related death toll exceeds 100 fatalities.

[1] CHR use data from Gleditsch (2004) who updated the COW datasets. The COW project was initiated by Singer and Small (1982, 1994) who defined a civil war when the battle-related deaths are above 1,000 per year. This is the definition which is predominant in civil war studies.
[2] See Table 1 in Collier et al. (2009). The number of onsets reported in their paper is lower because some conflicts in the Table are not used for the estimation of the baseline model. For example, if two wars break out in the same 5-year period they only code one war start.
[3] Version A of Sambanis (2004). Version A is the one with observation for ongoing wars are dropped.
[4] In their paper, FL argue that over the period 1945-1999 they code 122 civil wars. We downloaded a replication of their data set from Fearon's website (www.stanford.edu/~jfearon/) and we only count 111 onsets. Actually, for Russia 1956 the onset was coded 4 and not 1. Given the stricter requirements in Sambanis (2004), one would expect that Sambanis would record a smaller number of conflicts compared to FL. This inconsistency reflects the fact that the number of fatalities varies greatly among sources (e.g. for the war in Iraq the US Army reports less than one third of the fatalities reported by Amnesty International).

[5]The PITF provides three different datasets: 1) ethnic conflicts; 2) revolutions; 3) adverse regime changes. We consider ethnic conflicts and revolutions as civil wars, but not adverse regime changes.

Over the period 1948-2004 the PITF codes 134 onsets[6] and for each conflict it provides data on the intensity, geographical extent and number of people mobilized. The total number of war-year episodes is 1007. The UCDP/PRIO requires only that the battle-related deaths exceed 25 per year[7]. It codes the highest number of war-year episodes (1170) and onsets (249) over the period 1945-2004.

Panel A of Table 1 provides some simple descriptive statistics about these five sources of data. The mean incidence of civil war varies between 0.154 for UCDP/PRIO and 0.083 for CHR. The picture is broadly the same when we make the comparison only for the 1960-99 period covered by all data sets (Panel B). In this case the mean incidence varies between 0.181 for UCDP/PRIO and 0.093 for CHR. In both Panel A and Panel B, CHR is something of an outlier, with much lower incidence than the others, presumably because of its high annual death threshold (1,000).

Panel C of Table 1 shows pairwise correlations for incidence. The outlier CHR is least well correlated with other sources of data (0.65 in average), whereas the correlation between the other pairs averages 0.76. Panels D and E show pairwise correlations for the ends of wars ($Y_{jt} = 0$ and $Y_{jt-1} = 1$) and their onset respectively. Both of these are less highly correlated than incidence, the average correlation being 0.59 for the ends of wars and only 0.42 for onsets. The lower correlation of onsets most probably reflects the fact that wars are more likely to end than to start suddenly (e.g. with a peace agreement), so that different coding rules make less difference. These numbers imply that it will be harder to get consistent results across coding rules for onset or duration than for incidence, for which the correlation is higher.

[6] Actually, the PITF (2009) provides data for 81 ethnic wars and 66 revolutionary wars. However, some civil wars are classified both as a revolution and as an ethnic war. After dropping conflicts which are coded twice, and conflicts which break out after 2004, we are left with 134 onsets.

[7] The UCDP/PRIO provides a second dataset in which a cumulative 1,000 battle related deaths is used as a criterion to code a civil war. This second dataset has an average 0.72 correlation with other sources and it is likely to provide results which are much closer to other datasets. We use the dataset with the lowest death-requirement because it provides a more stringent test of the effect of different coding rules.

Table 1: Comparison Among Data Sets

Panel A: Descriptive Statistics

Variable	Obs	Mean	Conflict-Episodes
PITF 1948-2004	7538	0.134	1,007
UCDP/PRIO 1945-2004	7594	0.154	1,170
Sambanis 1945-1999	6691	0.137	920
FL 1945-1999	6586	0.139	918
CHR 1960-2004	6549	0.083	554

Panel B: Descriptive Statistics: 1960-1999 for All Data Sets

Variable	Obs	Mean	Conflict-Episodes
PITF	5253	0.163	855
UCDP/PRIO (25 deaths)	5253	0.181	950
Sambanis	5253	0.162	849
FL	5253	0.160	843
CHR	5253	0.093	487

Panel C: Pairwise Correlation for Incidence

	PITF	UCDP/PRIO	Sambanis	FL	CHR
PITF	1.0000				
UCDP/PRIO	0.7163	1.0000			
Sambanis (2004)	0.8056	0.7384	1.0000		
FL	0.7490	0.7434	0.8333	1.0000	
CHR	0.6675	0.6060	0.6928	0.6505	1.0000

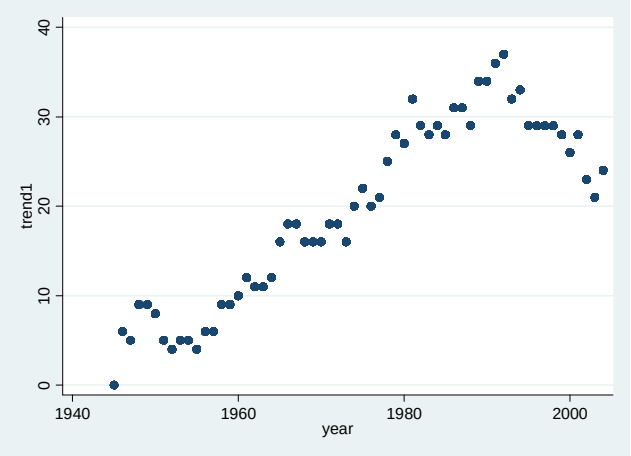
Panel D: Pairwise Correlation for Peace Onsets

	PITF	UCDP/PRIO	Sambanis	FL	CHR
PITF	1.0000				
UCDP/PRIO	0.5247	1.0000			
Sambanis	0.6008	0.5840	1.0000		
FL	0.5338	0.5713	0.7530	1.0000	
CHR	0.5130	0.5145	0.6656	0.6057	1.0000

Panel E: Pairwise Correlation for War Onsets

	PITF	UCDP/PRIO	Sambanis	FL	CHR
PITF	1.0000				
UCDP/PRIO	0.3181	1.0000			
Sambanis	0.5343	0.3653	1.0000		
FL	0.4291	0.3669	0.6343	1.0000	
CHR	0.4097	0.1975	0.5480	0.4430	1.000

Figure 1: Incidence of Civil War over Time



The incidence of civil war has varied not only across countries but also across time. Figure 1 shows the global incidence of civil war over fifty years (PITF data).[8] Incidence increases sharply up until the early 1990s, and then declines. This reflects fluctuations in duration rather than in onsets (Hegre, 2004). It is unclear what has driven this pattern, so we represent it by a time trend with a structural break after 1990. We do this by including a normal linear time trend together with a trend that begins only in 1990 and an intercept dummy that is one up to 1990 and zero afterwards.

Most other explanatory variables that have been used have more cross-country variation than time-series variation, so their principal role is to indicate what type of country is vulnerable to civil war. There are numerous possibilities – for example Hegre and Sambanis (2006) consider 88 variables in their sensitivity analysis of onsets, although this number includes regional and time dummies. Aside from the usual controls (such as country size and level of development), research has tended to concentrate on factors such as geography, natural resource endowments, political systems and ethnic diversity. Some variables more or less disqualify themselves because of limited coverage. We present results including the following variables:

[8]The pattern is similar in all sources of data.

1) Per capita real GDP at purchasing power parity (previous year in logarithms). The Source is Penn World Tables [PWT] 6.2 from 1950, and Fearon and Laitin (2003) before 1950;
2) Population (previous year in logarithms) – source as (1);
3) Ethnic diversity – we use ethnic fractionalization as given in Fearon and Laitin (2003), primarily because it has better coverage than ethnic polarization from Montalvo and Reynal-Querol (2005), but we consider polarization at a later stage;
4) Politics – we use a binary measure of anocracy which takes the value one if the polity2 index of Marshall and Jaggers (2000) is either below 5 or above -5 (± 5 excluded) in the previous year. The variable is meant to capture the non-linear relationship between conflict and democracy discussed in Hegre et *al*. (2001); this variable works better than a dummy for a high level of democracy (+6 or above on the polity2 index);
5) Natural resources – data on oil fields are collected from Lujala *et al*. (2007). The variable is coded 1 starting since the year of production of the earliest field (the one which was put into production first). Gas fields and unidentified oil fields are not taken in consideration for the construction of the dummy. Janse (2007) and Janse and Sheahan (1995) are the sources for surface deposits of diamonds;
6) Geography – the share of mountainous terrain as given in Fearon and Laitin (2003).

It can certainly be argued that our choice of variables does not do justice to the reasoning behind their inclusion. Natural resources do not just consist of oil and diamonds; other aspects of geography might matter besides mountains; and political factors cannot be fully captured in an anocracy index. We have used these variables because they have wide coverage and have been claimed to be significantly correlated with conflict in previous research.

3) Empirical Results

Table 2 presents our basic results for the incidence of civil war. The predicted probability of civil war ranges between 5.6% per year for CHR and 12.6% for PRIO. Seven variables are always significant at the 0.01 level, with the same sign across all data sets.

These are: (1) GDP per capita (–); (2) population (+); (3) mountainous terrain (+); (4) ethnic fractionalization (+); (5) the onshore oil dummy; (6) anocracy; and (7) the linear time-trend (+). The cold war dummy and the post-1990 time trend are not significant for CHR, but negative and significant at the 0.01 level in the other four cases. Surface diamond deposits have a coefficient that is significant at the 0.05 level only in one case, and then the sign of the coefficient (negative) is the opposite of that expected.

Table 2: Probit Model for Incidence

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit Model	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.303***	-0.247***	-0.292***	-0.296***	-0.341***
	(-15.85)	(-13.75)	(-14.59)	(-14.73)	(-15.03)
Log Population	0.158***	0.138***	0.142***	0.227***	0.0929***
	(9.61)	(8.96)	(8.47)	(13.57)	(5.10)
Share of Mountainous Terrain	0.623***	0.381***	0.344***	0.680***	0.676***
	(6.46)	(4.23)	(3.43)	(6.87)	(6.40)
Ethnic Fractionalization	0.787***	0.864***	0.603***	0.661***	0.555***
	(10.47)	(12.11)	(7.63)	(8.36)	(6.23)
Anocracy Dummy	0.468***	0.427***	0.520***	0.438***	0.488***
	(9.27)	(8.87)	(9.98)	(8.15)	(8.27)
Onshore Oil Dummy	0.362***	0.376***	0.301***	0.268***	0.173***
	(7.79)	(8.68)	(6.32)	(5.61)	(2.96)
Surface Diamond Deposits Dummy	0.0443	-0.106*	-0.134**	0.0680	-0.0170
	(0.76)	(-1.86)	(-2.11)	(1.10)	(-0.23)
Cold War Dummy (1 if year <1991)	-0.637***	-0.493***	-0.850***	-0.807***	0.123
	(-5.56)	(-4.42)	(-7.16)	(-6.59)	(0.81)
Linear Time Trend	0.0247***	0.0241***	0.0227***	0.0206***	0.0176***
	(10.43)	(11.31)	(9.91)	(9.03)	(5.39)
Post-1990 Time Trend	-0.0190***	-0.0184***	-0.0201***	-0.0187***	-0.00291
	(-6.08)	(-6.24)	(-6.09)	(-5.58)	(-0.70)
Constant	-0.984***	-1.146***	-0.440*	-1.309***	-0.727***
	(-4.41)	(-5.27)	(-1.87)	(-5.54)	(-2.61)
Observations	7009	7093	6268	6398	6088
Predicted Probability	0.0933	0.1211	0.1048	0.0934	0.0568
Pseudo R-Squared	0.1953	0.1596	0.1623	0.2031	0.1564
Area under ROC Curve	0.8086	0.7801	0.7909	0.8144	0.8007

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

The problem with the regressions in Table 2 is that they take no account of the persistence of conflicts. The best guess about which countries will experience civil wars next year is not based on any of the variables in Table 2, but on which ones are suffering a civil war this year. Persistence also means that the *z*-statistics in Table 2 are exaggerated by a pseudo-replication effect.[9] In econometric terms the error term in country *j* at time *t* is correlated with the error at time *t-1*. In order to control for first- and second-order serial correlation in Table 3 we enter a control for past incidence of civil war (war_{t-1} and war_{t-2}). The first lag of the dependent variable will be significant if war is more likely if it occurred in the previous year; the second lag will be significant if war is more likely after only one year of peace than after more years of peace, and *vice versa*. As expected, war_{t-1} and war_{t-2} have highly significant and positive coefficients, indicating that a conflict at time *t-1* or *t-2* increases the probability of conflict at time *t*, and the *z-statistics* of other variables are reduced.[10] Nevertheless per capita GDP and the linear time trend are still significant at the 0.01 level in all cases; ethnic fractionalization is significant at the 0.01 level in four out of five cases, and at the 0.05 level in the fifth case (Sambanis); mountainous terrain is significant at the 0.01 level in one case, at the 0.05 level in the three cases, and not significant in the fifth case (Sambanis); and population is significant at the 0.05 level in three cases and at the 0.10 level in the other two cases. Onshore oil fields are a more marginal case, since they are significant at the 0.05 level in two cases and not significant in the other three cases. Diamonds and anocracy never reach the 0.05 level of significance.

In Table 4, we show that the results are similar if we confine the sample to the same years (1960-99) for each data set. The *z-statistics* are a bit smaller than in Table 3, because of the reduction in the number of observations, but otherwise the picture is similar.

---

[9] This effect increases with the frequency of the data, because of the multiplication of observations.

[10] Further lags of war tend not to be significant. See the Appendix for results with more lags, and using the cubic spline procedure of Beck *et al*. (1998). These results differ little from those shown in Table 3.

Table 3: Probit Model for Incidence – Controlling For War (t -1) and War(t -2)

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit Model	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.146***	-0.110***	-0.141***	-0.134***	-0.176***
	(-5.49)	(-4.52)	(-5.05)	(-4.47)	(-5.85)
Log Population	0.0646**	0.0518**	0.0483*	0.0803**	0.0550*
	(2.26)	(2.28)	(1.75)	(2.56)	(1.75)
Share of Mountainous Terrain	0.328**	0.250**	0.224	0.368**	0.434***
	(2.07)	(1.97)	(1.43)	(2.18)	(2.66)
Ethnic Fractionalization	0.349***	0.499***	0.335**	0.356***	0.376***
	(2.79)	(4.78)	(2.55)	(2.61)	(2.66)
Anocracy Dummy	0.0954	0.107	0.157*	0.120	0.176*
	(1.07)	(1.45)	(1.73)	(1.22)	(1.87)
Onshore Oil Dummy	0.169**	0.156**	0.0807	0.0815	0.0451
	(2.31)	(2.50)	(1.07)	(0.99)	(0.51)
Surface Diamond Deposits Dummy	0.0416	-0.0526	-0.0409	0.00486	-0.00550
	(0.38)	(-0.58)	(-0.37)	(0.04)	(-0.05)
Cold War Dummy (1 if year <1991)	0.0859	-0.126	-0.0642	-0.0240	-0.00249
	(0.39)	(-0.71)	(-0.26)	(-0.09)	(-0.01)
Linear Time Trend	0.0103***	0.0110***	0.0115***	0.00944***	0.0138***
	(2.80)	(3.76)	(3.43)	(2.66)	(3.12)
Post-1990 Time Trend	-0.00285	-0.00892**	-0.00669	-0.00339	-0.00722
	(-0.51)	(-2.00)	(-1.08)	(-0.51)	(-1.17)
War (t–1)	3.042***	2.032***	2.810***	2.964***	2.634***
	(22.06)	(26.15)	(22.75)	(21.59)	(19.28)
War (t–2)	0.317**	0.957***	0.509***	0.549***	0.469***
	(2.22)	(12.11)	(3.96)	(3.87)	(3.30)
Constant	-2.207***	-1.959***	-1.861***	-2.334***	-1.942***
	(-5.59)	(-5.76)	(-4.56)	(-5.13)	(-4.36)
Observations	6730	6840	6037	6165	5783
Predicted Probability	0.0425	0.0709	0.0482	0.0418	0.0231
Pseudo R-Squared	0.7379	0.6146	0.7109	0.7600	0.6594
Area under ROC Curve	0.9679	0.9442	0.9608	0.9683	0.9577

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

Table 4: Probit Model for Incidence – Same Time Period

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit Model	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.142***	-0.118***	-0.135***	-0.127***	-0.173***
	(-4.77)	(-4.44)	(-4.59)	(-4.02)	(-5.57)
Log Population	0.0628*	0.0454*	0.0647**	0.103***	0.0416
	(1.96)	(1.80)	(2.14)	(2.95)	(1.28)
Share of Mountainous Terrain	0.388**	0.282*	0.236	0.397**	0.370**
	(2.31)	(1.93)	(1.40)	(2.14)	(2.10)
Ethnic Fractionalization	0.446***	0.503***	0.361**	0.391***	0.380**
	(3.17)	(4.27)	(2.51)	(2.63)	(2.55)
Anocracy Dummy	0.0795	0.154*	0.214**	0.197*	0.186*
	(0.80)	(1.81)	(2.17)	(1.83)	(1.84)
Onshore Oil Dummy	0.139*	0.170**	0.0383	0.0106	0.0701
	(1.68)	(2.39)	(0.47)	(0.12)	(0.76)
Surface Diamond Deposits Dummy	-0.0447	-0.112	-0.0623	-0.0391	-0.0394
	(-0.37)	(-1.10)	(-0.53)	(-0.30)	(-0.31)
Cold War Dummy (1 if year <1991)	0.0735	-0.0401	0.0442	0.103	-0.147
	(0.29)	(-0.19)	(0.17)	(0.35)	(-0.54)
Linear Time Trend	0.00906**	0.00882**	0.00840**	0.00631	0.0141***
	(2.05)	(2.39)	(1.98)	(1.39)	(3.16)
Post-1990 Time Trend	-0.00191	-0.00569	-0.00360	5.78e-05	-0.00943
	(-0.28)	(-1.01)	(-0.54)	(0.01)	(-1.32)
War (t–1)	2.999***	2.045***	2.927***	3.149***	2.593***
	(20.22)	(23.20)	(20.77)	(19.04)	(18.35)
War (t–2)	0.340**	0.991***	0.409***	0.408**	0.514***
	(2.20)	(11.06)	(2.79)	(2.38)	(3.51)
Constant	-2.182***	-1.882***	-2.080***	-2.645***	-1.693***
	(-4.91)	(-4.85)	(-4.76)	(-5.31)	(-3.54)
Observations	5228	5217	5096	5204	5056
Predicted Probability	0.0517	0.0822	0.0560	0.0470	0.0255
Pseudo R-Squared	0.7364	0.6262	0.7217	0.7753	0.6573
Area under ROC Curve	0.9680	0.9466	0.9640	0.9727	0.9555

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

4) Onset and Duration

In this section we investigate whether the determinants of the incidence of conflict differ according to whether conflict was occurring in the previous year or not. We do this by adding to the regression interactions of the lagged dependent variable (war_{t-1}) with other

regressors, as in Elbadawi and Sambanis (2002). If these interaction terms are statistically significant, it suggests that the variable in question affects the probability of continuation of war (implied by the lagged dependent variable taking the value one) differently from the probability of onset (implied by the lagged dependent variable taking the value zero). If the interaction terms are collectively not significant, then the null hypothesis that the variables in question affect onset and continuation similarly cannot be rejected. For these purposes we drop the dummies for anocracy and surface diamonds, which were not significant in Table 3.

Table 5 shows the results of this exercise, with lagged war interacted with per capita GDP, population, the share of mountainous terrain and ethnic fractionalization. The F-test that the five interaction variables have zero coefficients is rejected at the 0.05 level only in the case of the UCDP/PRIO data. In the case of the UCDP/PRIO data, the rejection of the null is attributable to the ethnic fractionalization coefficient, which is close to zero (0.681 – 0.643) for continuation but strongly positive for onset (0.681). For none of the other data sets, however, is there any indication of a difference in the ethnic fractionalization coefficient. Per capita GDP interacted with the lagged dependent variable always has a positive coefficient, which is significant at the 0.05 level in two cases, and at the 0.10 level in one case. This suggests that poverty is slightly more important for onset than for continuation.

Table 5 – Separating Onset and Continuation

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.178***	-0.137***	-0.182***	-0.159***	-0.214***
	(-6.65)	(-5.07)	(-6.41)	(-5.36)	(-7.04)
Log Population	0.0876***	0.0631**	0.0399	0.0731**	0.0595*
	(2.87)	(2.42)	(1.36)	(2.13)	(1.92)
Share of Mountainous Terrain	0.315	0.198	0.254	0.415**	0.401**
	(1.63)	(1.31)	(1.43)	(2.23)	(2.13)
Ethnic Fractionalization	0.408***	0.681***	0.319**	0.301*	0.526***
	(2.76)	(5.57)	(2.07)	(1.85)	(3.15)
Onshore Oil Dummy	0.206**	0.109	0.0766	0.0742	0.0959
	(2.30)	(1.40)	(0.84)	(0.76)	(0.87)
Cold War Dummy (1 if year < 1991)	0.114	-0.115	-0.104	-0.0919	-0.121
	(0.50)	(-0.66)	(-0.45)	(-0.38)	(-0.49)
Linear Time Trend	0.00994**	0.00975***	0.00894**	0.00547	0.0155***
	(2.46)	(3.02)	(2.32)	(1.37)	(3.02)
Post-1990 Time Trend	-0.00220	-0.00864*	-0.00816	-0.00558	-0.0101
	(-0.38)	(-1.91)	(-1.31)	(-0.84)	(-1.61)
War (t–1)	2.648***	1.608**	1.345*	1.663*	1.726*
	(3.16)	(2.36)	(1.70)	(1.83)	(1.88)
War (t–2)	0.333**	0.924***	0.460***	0.474***	0.490***
	(2.41)	(11.70)	(3.59)	(3.34)	(3.52)
Time trend * War(t–1)	-0.000329	0.00348	0.00778	0.0133**	-0.00773
	(-0.06)	(0.78)	(1.32)	(2.05)	(-1.09)
Constant	-2.212***	-1.872***	-1.340***	-1.866***	-1.682***
	(-5.31)	(-4.91)	(-3.14)	(-3.86)	(-3.65)
Variable Included in the F-Test					
GDP * War (t–1)	0.125*	0.0855	0.150**	0.0894	0.210**
	(1.69)	(1.41)	(2.04)	(1.06)	(2.37)
Population * War (t–1)	-0.0471	-0.00785	0.0244	0.0392	-0.0119
	(-0.74)	(-0.16)	(0.39)	(0.54)	(-0.15)
Mountainous T. * War (t–1)	-0.0547	0.0916	-0.119	-0.357	-0.151
	(-0.17)	(0.34)	(-0.37)	(-0.97)	(-0.42)
Ethnic Frac. * War (t–1)	-0.0337	-0.643***	-0.0570	0.0938	-0.355
	(-0.13)	(-2.87)	(-0.21)	(0.32)	(-1.12)
Onshore Oil * War (t–1)	-0.144	0.112	-0.0124	0.0279	-0.155
	(-0.91)	(0.84)	(-0.08)	(0.16)	(-0.82)
Observations	6787	6898	6076	6206	5832
F-test (war_{t-1}*$variable$= 0) p-value	0.3664	0.0043	0.4090	0.6719	0.1051
Predicted Probability	0.0409	0.0692	0.0478	0.0423	0.0220
Pseudo R-Squared	0.7374	0.6156	0.7109	0.7616	0.6588
Area under ROC Curve	0.9676	0.9447	0.9610	0.9694	0.9581

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

Now let us consider what happens if we estimate probit models for the onset and continuation of civil wars separately. For onset, we discard all observations where *war(t-1) = 1*, and consider only the probability of a war starting in period *t*, conditional on peace in period *t-1*. For continuation, we discard all observations where *war(t-1) = 0*, and consider only the probability of a war continuing in period *t*, conditional on war in period *t-1*. The results are shown in Tables 6 and 7 respectively.

In Table 6 (onset), per capita GDP, population, mountainous terrain and ethnic fractionalization always have the same sign as in Table 3, but occasionally do not reach the 0.05 level of significance (Sambanis for population; PITF, UCDP/PRIO and Sambanis for mountainous terrain; and Sambanis and FL for ethnic fractionalization). The onshore oil dummy is significant at the 0.05 level in one case only (PITF). In Table 7 (continuation), many variables are insignificant because of the smallness of the sample. Apart from the time trends, ethnic fractionalization is most frequently significant, with a positive coefficient. Even if they are not generally significant in Table 7, the variables that are significant in Table 3 tend to have the same sign in Table 7. Note the difference in predicted probabilities (below 0.03 for Table 6; above 0.85 for Table 7), which indicates how persistent civil wars are.

Taken in isolation, Table 7 appears to justify the conclusions of Collier *et al*. (2004) and Fearon and Laitin (2004) that GDP per capita, population and mountainous terrain have no significant effect on the duration of civil war. According to our approach, this is a sampling error problem. These variables have similar coefficients in Tables 6 and 7, as does also ethnic fractionalization, but sampling error is increased by dividing the whole set of country-year observations into two subsets, which makes the coefficients not infrequently insignificant. Our results in Table 5 show that this division of the sample is unjustified, once we have allowed for the shift effect on the probability of civil war of its occurrence in the previous year.

Table 6: Probit Model for Incidence if War (t-1)=0

	Dependent Variable: Onset of Civil War				
Estimation Method: Probit Model	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.176***	-0.134***	-0.180***	-0.161***	-0.207***
	(-6.41)	(-4.88)	(-6.32)	(-5.23)	(-6.58)
Log Population	0.0773**	0.0723***	0.0406	0.0685**	0.0622*
	(2.45)	(2.69)	(1.33)	(1.98)	(1.80)
Share of Mountainous Terrain	0.375*	0.185	0.249	0.451**	0.428**
	(1.84)	(1.18)	(1.32)	(2.32)	(2.15)
Ethnic Fractionalization	0.298**	0.614***	0.243	0.206	0.378**
	(2.00)	(4.86)	(1.54)	(1.27)	(2.23)
Anocracy Dummy	0.217**	0.247***	0.285***	0.246**	0.280***
	(2.04)	(2.88)	(2.78)	(2.26)	(2.59)
Onshore Oil Dummy	0.226**	0.110	0.0813	0.0750	0.116
	(2.49)	(1.41)	(0.89)	(0.76)	(1.04)
Surface Diamond Deposits Dummy	0.184	-0.00576	0.0591	0.155	0.0212
	(1.56)	(-0.06)	(0.48)	(1.21)	(0.15)
Cold War Dummy (1 if year <1991)	-0.0534	-0.388**	-0.172	-0.153	-0.0200
	(-0.20)	(-2.16)	(-0.67)	(-0.57)	(-0.08)
Linear Time Trend	0.00960**	0.00716*	0.00541	0.00184	0.00897
	(2.05)	(1.89)	(1.23)	(0.41)	(1.46)
Post-1990 Time Trend	-0.00477	-0.0112**	-0.00441	-0.00138	-0.00324
	(-0.70)	(-2.18)	(-0.63)	(-0.19)	(-0.44)
War (t–2)	0.262	0.745***	0.494**	0.303	0.713***
	(1.01)	(6.07)	(2.39)	(1.02)	(3.66)
Constant	-2.015***	-1.684***	-1.273***	-1.733***	-1.734***
	(-4.47)	(-4.22)	(-2.77)	(-3.36)	(-3.55)
Observations	5787	5742	5174	5297	5264
Predicted Probability	0.0117	0.0238	0.0146	0.0116	0.0098
Pseudo R-Squared	0.0733	0.0883	0.0599	0.0580	0.0999
Area under ROC Curve	0.7382	0.7375	0.7128	0.7109	0.7743

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

Table 7: Probit Model for Incidence if War (t-1)=1

	Dependent Variable: Duration of Civil War				
Estimation Method: Probit Model	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.0626	-0.0739	-0.0500	-0.109	-0.0553
	(-0.87)	(-1.28)	(-0.70)	(-1.27)	(-0.62)
Log Population	0.0195	0.00713	0.0728	0.117*	0.00158
	(0.33)	(0.16)	(1.28)	(1.77)	(0.02)
Share of Mountainous Terrain	0.275	0.407*	0.164	0.0657	0.396
	(0.95)	(1.72)	(0.57)	(0.19)	(1.17)
Ethnic Fractionalization	0.470**	0.142	0.396	0.541**	0.204
	(1.97)	(0.68)	(1.61)	(2.01)	(0.66)
Anocracy Dummy	-0.0904	-0.0954	-0.0183	-0.0633	-0.0965
	(-0.67)	(-0.85)	(-0.13)	(-0.41)	(-0.62)
Onshore Oil Dummy	0.0668	0.216*	0.0443	0.100	-0.0677
	(0.48)	(1.86)	(0.32)	(0.62)	(-0.42)
Surface Diamond Deposits Dummy	-0.120	0.0173	-0.136	-0.267	0.130
	(-0.64)	(0.11)	(-0.70)	(-1.23)	(0.59)
Cold War Dummy (1 if year < 1991)	0.276	0.339	-0.149	-0.232	0.0671
	(0.81)	(1.20)	(-0.44)	(-0.63)	(0.15)
Linear Time Trend	0.00940	0.0188***	0.0240***	0.0279***	0.0212***
	(1.39)	(3.27)	(3.80)	(3.77)	(2.79)
Post-1990 Time Trend	0.00105	-0.00464	-0.0175*	-0.0180	-0.0146
	(0.11)	(-0.60)	(-1.81)	(-1.61)	(-1.30)
War (t–2)	0.342**	1.042***	0.421***	0.532***	0.265
	(2.05)	(9.28)	(2.67)	(3.14)	(1.51)
Constant	0.577	-0.253	0.0297	-0.00155	0.568
	(0.68)	(-0.38)	(0.04)	(-0.00)	(0.59)
Observations	943	1098	863	868	519
Predicted Probability	0.9197	0.8718	0.9148	0.9426	0.8597
Pseudo R-Squared	0.0351	0.1548	0.0672	0.1066	0.0428
Area under ROC Curve	0.6278	0.7665	0.6752	0.7174	0.6334

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

5) Is Africa Different?

Since independence, most sub-Saharan African countries have experienced at least one civil war, and the incidence of civil war is higher there than in other regions of the world (see Table 8). Averaged over the five data sets, sub-Saharan Africa represents 24.4% of the country-year observations, but 32.6% of those identified as experiencing a civil conflict. The difference is particularly marked if a high death threshold is used, as is

shown by the figures for CHR data, which indicates that sub-Saharan Africa is particularly prone to conflicts that result in a high number of battle deaths annually.

Table 8 – Incidence of Civil War in sub-Saharan Africa and elsewhere

Panel A: Sub-Saharan Africa			
Variable	Obs	Mean	Std. Dev.
PITF	1798	0.1863	0.3894
UCDP/PRIO	1804	0.1951	0.3964
Sambanis	1552	0.1842	0.3878
FL	1593	0.1682	0.3741
CHR	1754	0.1242	0.3300

Panel B: Rest of the Sample			
Variable	Obs	Mean	Std. Dev.
PITF	5740	0.1170	0.3215
UCDP/PRIO	5790	0.1412	0.3483
Sambanis	5139	0.1233	0.3288
FL	4993	0.1301	0.3365
CHR	4814	0.0697	0.2548

As well as being relatively poor, sub-Saharan Africa is characterized by a relatively high degree of ethnic diversity. The ethnic fractionalization measure averages 0.65 for sub-Saharan Africa and 0.30 for the rest of the world, and there is a relatively high positive correlation between a sub-Saharan Africa (SSA) dummy and ethnic fractionalization (0.53). This raises the possibility that the ethnic diversity effect which we have estimated so far may in fact be an unidentified SSA effect. More generally, it is also interesting to know whether different factors determine the incidence of civil wars in SSA, compared with elsewhere.

To address the first question, we add an SSA dummy to the Table 3 regression. The results are shown in Table 9. The SSA dummy always has a positive coefficient, but it is never close to statistical significance. The ethnic fractionalization coefficient tends to be smaller than in Table 3, and its z-statistic is reduced by a slightly larger percentage, reflecting the effect of multicollinearity on the precision of the estimates. Nevertheless

the variable is still significant at the 0.05 level in two cases, and at the 0.10 level in two other cases. Thus the estimated role of ethnic diversity is not an unexplained effect of higher incidence of civil conflict in sub-Saharan Africa.

In Table 10, we add interactions of the SSA dummy with the four main explanatory variables, and perform an F-test similar to that in Table 5. The F-statistic is never significant, but in every case ethnic fractionalization interacted with the SSA dummy is significantly negative at the 0.10 level, indicating that ethnic diversity is less important in explaining which countries are prone to conflict *within* SSA. One possible reason for this is that all sub-Saharan countries tend to have relatively high values of ethnic diversity.

Table 9: Probit Model for Incidence – Controlling for Sub-Sahara Africa

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit Model	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.139***	-0.0986***	-0.128***	-0.122***	-0.159***
	(-5.02)	(-3.89)	(-4.37)	(-3.82)	(-5.02)
Log Population	0.0683**	0.0587**	0.0570**	0.0883***	0.0652**
	(2.35)	(2.57)	(1.98)	(2.74)	(2.05)
Share of Mountainous Terrain	0.350**	0.291**	0.269*	0.415**	0.470***
	(2.10)	(2.21)	(1.65)	(2.35)	(2.77)
Ethnic Fractionalization	0.303**	0.431***	0.225	0.256*	0.264*
	(2.16)	(3.63)	(1.49)	(1.69)	(1.74)
Anocracy Dummy	0.0951	0.108	0.169*	0.128	0.180*
	(1.07)	(1.45)	(1.85)	(1.30)	(1.92)
Onshore Oil Dummy	0.188**	0.178***	0.119	0.119	0.0869
	(2.47)	(2.73)	(1.52)	(1.40)	(0.94)
Surface Diamond Deposits Dummy	0.0330	-0.0641	-0.0619	-0.0165	-0.0313
	(0.30)	(-0.70)	(-0.56)	(-0.14)	(-0.27)
Cold War Dummy (1 if year < 1991)	0.0909	-0.112	-0.0464	-0.00745	0.00746
	(0.41)	(-0.62)	(-0.19)	(-0.03)	(0.03)
Linear Time Trend	0.0105***	0.0110***	0.0117***	0.00960***	0.0156***
	(2.83)	(3.74)	(3.46)	(2.67)	(3.48)
Post-1990 Time Trend	-0.00306	-0.00896**	-0.00682	-0.00348	-0.00818
	(-0.54)	(-1.98)	(-1.09)	(-0.51)	(-1.27)
War (t–1)	3.040***	2.032***	2.808***	2.966***	2.627***
	(22.04)	(26.13)	(22.74)	(21.57)	(19.20)
War (t–2)	0.318**	0.958***	0.508***	0.548***	0.475***
	(2.22)	(12.13)	(3.97)	(3.86)	(3.35)
Sub-Saharan Africa Dummy	0.0796	0.114	0.174	0.165	0.196*
	(0.74)	(1.30)	(1.53)	(1.40)	(1.72)
Constant	-2.317***	-2.143***	-2.095***	-2.551***	-2.261***
	(-5.53)	(-6.04)	(-4.77)	(-5.19)	(-4.76)
Observations	6730	6840	6037	6165	5783
Predicted Probability	0.0423	0.0706	0.0478	0.0414	0.0227
Pseudo R-Squared	0.7380	0.6149	0.7115	0.7604	0.6603
Area under ROC Curve	0.9679	0.9442	0.9609	0.9684	0.9581

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

Table 10: Interaction of Variables with Sub-Saharan Africa

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.154***	-0.113***	-0.134***	-0.123***	-0.173***
	(-5.63)	(-4.35)	(-4.61)	(-3.89)	(-5.85)
Log Population	0.0574*	0.0425*	0.0456	0.0985***	0.0488
	(1.83)	(1.78)	(1.52)	(2.83)	(1.52)
Share of Mountainous Terrain	0.165	0.0527	0.210	0.360*	0.413**
	(0.84)	(0.34)	(1.15)	(1.72)	(2.05)
Ethnic Fractionalization	0.512***	0.634***	0.346**	0.397**	0.421**
	(3.01)	(4.64)	(2.08)	(2.20)	(2.22)
Onshore Oil Dummy	0.200**	0.204***	0.0986	0.0859	0.0236
	(2.37)	(2.85)	(1.15)	(0.90)	(0.23)
Cold War (1 if year < 1991)	0.0725	-0.114	0.0171	0.114	-0.156
	(0.31)	(-0.62)	(0.07)	(0.40)	(-0.58)
Linear Time Trend	0.0119***	0.0122***	0.0115***	0.00817**	0.0188***
	(3.03)	(3.95)	(3.22)	(2.17)	(3.89)
Post-1990 Time Trend	-0.00380	-0.00903**	-0.00564	-0.00141	-0.0124*
	(-0.65)	(-1.97)	(-0.87)	(-0.20)	(-1.88)
War (t–1)	3.029***	2.030***	2.811***	2.973***	2.610***
	(22.52)	(26.26)	(22.71)	(21.49)	(19.10)
War (t–2)	0.314**	0.947***	0.507***	0.543***	0.487***
	(2.24)	(12.05)	(3.94)	(3.81)	(3.39)
Sub-Sahara Africa	-0.712	-0.295	-0.103	0.532	-0.790
	(-0.67)	(-0.37)	(-0.10)	(0.44)	(-0.72)
Time trend * SSAfrica	-0.00330	-0.00277	0.00377	0.00851	-0.00705
	(-0.49)	(-0.51)	(0.50)	(1.04)	(-0.95)
Constant	-2.134***	-1.927***	-1.982***	-2.710***	-1.900***
	(-5.03)	(-5.41)	(-4.52)	(-5.51)	(-4.21)
Variable Included in the F-Test					
GDP* SSAfrica	0.0547	0.0353	0.0286	0.00255	0.0676
	(0.55)	(0.44)	(0.29)	(0.02)	(0.66)
Population*SSAfrica	0.100	0.0567	0.0531	-0.0151	0.121
	(1.26)	(0.91)	(0.66)	(-0.18)	(1.50)
Montainous T.* SSAfrica	9.47e-05	0.00265	-0.00308	-0.00333	-0.00389
	(0.02)	(0.79)	(-0.80)	(-0.77)	(-0.99)
Ethnic Frac.* SSAfrica	-0.706**	-0.596*	-0.782*	-0.727*	-0.656*
	(-2.00)	(-1.84)	(-1.87)	(-1.85)	(-1.79)
Onshore Oil * SSAfrica	-0.0195	-0.0539	0.110	0.171	0.156
	(-0.10)	(-0.33)	(0.57)	(0.85)	(0.76)
Observations	6787	6898	6076	6206	5832
F-test (war_{t-1}*variable= 0) *p*-value	0.4037	0.2179	0.5825	0.4821	0.4038
Predicted Probability	0.0425	0.0710	0.0483	0.0414	0.0232
Pseudo R-Squared	0.7376	0.6139	0.7102	0.7609	0.6586
Area under ROC Curve	0.9678	0.9440	0.9603	0.9685	0.9582

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

6) Ethnic Fractionalization or Ethnic Polarization?

Montalvo and Reynal-Querol (2005) [MRQ] claim that civil conflict is associated with ethnic polarization rather than fractionalization, and that fractionalization becomes insignificant when polarization is included in a regression for the incidence of conflict. Polarization is meant to capture the idea that conflict is most likely if two sizeable groups are competing for domination. If there are n groups, and group i represents share s_i of the population, then polarization is defined as

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

whereas fractionalization is

$$EF = 1 - \sum_{i=1}^{n} s_i^2 \qquad (3)$$

Both measures are defined over the interval (0, 1). Ethnic fractionalization increases as group shares get smaller, so that it is a straightforward index of ethnic diversity. As can be seen from equation (2), polarization is maximized when group shares are close to 50%, which requires two approximately equally sized groups. As is illustrated in the graphs in MRQ, the relationship between fractionalization and polarization is an inverted U-shape, with polarization maximized at intermediate levels of fractionalization. This is because group shares can be a long way from 50% either because one group is very large (*EF* is small) or because all groups are small (*EF* is large).

The data on ethnic polarization given in MRQ do not cover the full range of countries, but we use data on the share of the second largest group from Fearon and Laitin (2003) to impute missing values (the correlation between *EP* and the share of the second largest group is 0.77). Table 11 shows the effect of replacing *EF* by *EP*. Ethnic polarization is always significant at the 0.01 level, and in four out of five cases its *z*-statistic is higher than that of ethnic fractionalization in Table 3. Surprisingly, mountainous terrain loses significance when ethnic polarization is included.

We next test the MRQ hypothesis that ethnic fractionalization is insignificant when included together with polarization. Table 12 shows the results, with anocracy and

diamonds omitted. Although in two cases (Sambanis and FL) *EP* is significant and *EF* is not, as MRQ suggest, in the other three cases *EF* is highly significant, and indeed more significant than *EP*. This suggests that the relevant aspects of ethnic diversity are not completely captured either by polarization or fractionalization.

Table 11: Incidence (Using Imputed Ethnic Polarization)

Dependent Variable: Incidence					
	Model 1	Model 2	Model 3	Model 4	Model 4
Estimation Method: Probit Model	PITF	UCDP/PRIO	Sambanis	FL	CH
Log GDP per capita	-0.168***	-0.139***	-0.162***	-0.157***	-0.197***
	(-6.66)	(-6.01)	(-6.19)	(-5.57)	(-6.80)
Log Population	0.0807***	0.0733***	0.0648**	0.0972***	0.0771**
	(2.79)	(3.19)	(2.31)	(3.10)	(2.43)
Share of Mountainous Terrain	0.182	0.0796	0.0734	0.205	0.277
	(1.08)	(0.59)	(0.44)	(1.14)	(1.57)
Ethnic Polarization	0.503***	0.592***	0.482***	0.579***	0.544***
	(3.08)	(4.45)	(2.85)	(3.20)	(2.96)
Anocracy, Lagged	0.0962	0.111	0.161*	0.118	0.186**
	(1.09)	(1.52)	(1.79)	(1.21)	(1.99)
Onshore Oil Dummy	0.144**	0.123**	0.0525	0.0571	0.0272
	(1.96)	(1.98)	(0.69)	(0.69)	(0.31)
Surface Diamond Deposits Dummy	0.0608	-0.0167	-0.0221	0.0132	0.0225
	(0.56)	(-0.19)	(-0.21)	(0.11)	(0.19)
Cold War Dummy (1 if year < 1991)	0.0204	-0.179	-0.128	-0.0912	-0.0204
	(0.09)	(-1.02)	(-0.53)	(-0.35)	(-0.08)
Time Trend	0.00958***	0.0103***	0.0111***	0.0089**	0.0117***
	(2.64)	(3.56)	(3.35)	(2.53)	(2.67)
Post-1990 Time Trend	-0.00364	-0.00950**	-0.00763	-0.00423	-0.00651
	(-0.66)	(-2.15)	(-1.25)	(-0.64)	(-1.06)
War (t–1)	3.045***	2.031***	2.801***	2.955***	2.627***
	(22.29)	(26.10)	(22.64)	(21.51)	(19.24)
War (t–2)	0.308**	0.960***	0.508***	0.549***	0.467***
	(2.18)	(12.13)	(3.96)	(3.90)	(3.29)
Constant	-2.192***	-1.929***	-1.863***	-2.360***	-2.009***
	(-5.53)	(-5.75)	(-4.57)	(-5.24)	(-4.51)
Observations	6730	6840	6037	6165	5783
Predicted Probability	0.0420	0.0706	0.0477	0.0409	0.0232
Pseudo R-Squared	0.7380	0.6138	0.7110	0.7605	0.6596
Area under ROC Curve	0.9686	0.9448	0.9614	0.9699	0.9593

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

Table 12: Incidence (Including Both Ethnic Polarization and Fractionalization)

Dependent Variable: Incidence					
	Model 1	Model 2	Model 3	Model 4	Model 4
Estimation Method: Probit Model	PITF	UCDP/PRIO	Sambanis	FL	CH
Log GDP per capita	-0.160***	-0.125***	-0.156***	-0.153***	-0.181***
	(-6.24)	(-5.16)	(-5.75)	(-5.30)	(-6.18)
Log Population	0.0801***	0.0622***	0.0513*	0.0918***	0.0749**
	(2.84)	(2.82)	(1.92)	(3.05)	(2.50)
Share of Mountainous Terrain	0.206	0.138	0.114	0.213	0.279*
	(1.28)	(1.07)	(0.72)	(1.23)	(1.68)
Ethnic Polarization	0.369**	0.408***	0.386**	0.446**	0.369*
	(2.02)	(2.80)	(2.07)	(2.29)	(1.77)
Ethnic Fractionalization	0.274**	0.353***	0.213	0.233	0.306**
	(1.98)	(3.18)	(1.51)	(1.59)	(1.96)
Onshore Oil Dummy	0.152**	0.149**	0.0709	0.0733	0.0123
	(2.06)	(2.38)	(0.94)	(0.88)	(0.14)
Cold War Dummy (1 before 1990)	0.0494	-0.124	-0.0720	-0.0169	-0.120
	(0.23)	(-0.71)	(-0.30)	(-0.06)	(-0.51)
Time Trend	0.0103***	0.0109***	0.0119***	0.00945***	0.0131***
	(2.83)	(3.78)	(3.57)	(2.68)	(2.99)
Post-1990 Time Trend	-0.00373	-0.00880**	-0.00685	-0.00324	-0.0101*
	(-0.68)	(-2.00)	(-1.12)	(-0.49)	(-1.69)
War (t–1)	3.036***	2.029***	2.813***	2.971***	2.619***
	(22.68)	(26.24)	(22.80)	(21.64)	(19.33)
War (t–2)	0.306**	0.944***	0.496***	0.533***	0.486***
	(2.20)	(12.00)	(3.87)	(3.77)	(3.42)
Constant	-2.311***	-2.062***	-1.885***	-2.439***	-2.034***
	(-5.90)	(-6.20)	(-4.69)	(-5.49)	(-4.64)
Observations	6787	6898	6076	6206	5832
Predicted Probability	0.0425	0.0707	0.0482	0.0411	0.0235
Pseudo R-Squared	0.7372	0.6136	0.7096	0.7604	0.6570
Area under ROC Curve	0.9675	0.9445	0.9599	0.9688	0.9579

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

In order to probe further into this issue, we can allow the *EF* coefficient to take different values for different levels of *EF*. There are three distinct regions in the relationship between *EF* and *EP*. If $EF < 0.3$, there is a very strong positive correlation between *EF* and *EP*, which reflects the fact that when there are only two groups, *EP* is equal to exactly twice *EF*. On the other hand, when $EF > 0.7$, there is an equally strong negative correlation between *EF* and *EP*. In the intermediate region ($0.3 < EF < 0.7$), the graphs in MRQ show that *EF* and *EP* are almost uncorrelated. This intermediate region includes

almost all of the observations of $EP > 0.5$. The fact that both *EF* and *EP* are significant suggests that there is a non-linearity in the relationship between *EF* and conflict (otherwise *EP* would not be significant), but of a form not well captured by *EP* (otherwise *EF* would not be significant). By allowing the *EF* coefficient to differ according to whether it is above or below 0.7, we can capture the non-linearity in the relationship between *EF* and conflict without resort to *EP*. Then the *EP* coefficient will only be significant if polarization makes a difference for the intermediate range of *EF*. In other words, we will be able to test whether *EP* is significant in Table 12 because *EP* genuinely matters in countries with intermediate levels of *EF*, or only because it helps to explain why conflict is less frequent than expected at high levels of *EF*. If the latter is true, then it is less clear that polarization is the important factor, and there may be some other reason for the non-linear relationship between *EF* and conflict.

The results of this exercise are shown in Table 13. Ethnic polarization is at least significant at the 0.05 level, but ethnic fractionalization has a particularly small coefficient when it is less than or equal to 0.7. Accordingly, in Table 14, we assume that *EF* only has an effect above 0.7. Then *EP* is always significant at the 0.01 level, and (*EF*–0.7) is always significant at the 0.05 level (except in Model 3). The fit in Table 14 is slightly better than in Table 12. This suggests that polarization is the important factor for low and intermediate levels of ethnic diversity, but that it underpredicts the likelihood of conflict at high levels of diversity, where polarization declines. In another paper (Bleaney and Dimico, 2009) we suggest that countries with high ethnic diversity are likely to be be quite polarized at the local level, even if the national polarization measure is not high, and we show that conflicts tend to involve a smaller proportion of the country when ethnic diversity is high.

Finally, we consider whether including ethnic polarization in preference to mountainous terrain, which tends not to be significant in Table 12, makes a difference to the tests shown in Table 5 for the difference between the probability of onset and continuation of conflict. Table 15 shows the results. In Table 15 the *F*-statistics all tend to be slightly less significant than in Table 5, but the statistic is still significant at the 0.01 level for the

PRIO data set, with ethnic fractionalization having much less effect on continuation than onset. Nevertheless the general picture is much as in Table 5.

Table 13: Separating Different Levels of Ethnic Fractionalization

Dependent Variable: Incidence of Civil War					
	Model 1	Model 2	Model 3	Model 4	Model 4
Estimation Method: Probit Model	PITF	UCDP/PRIO	Sambanis	FL	CH
Log GDP per capita	-0.158***	-0.124***	-0.154***	-0.150***	-0.179***
	(-6.13)	(-5.14)	(-5.67)	(-5.19)	(-6.08)
Log Population	0.0759***	0.0613***	0.0472*	0.0877***	0.0692**
	(2.64)	(2.74)	(1.74)	(2.85)	(2.26)
Share of Mountainous Terrain	0.238	0.145	0.142	0.257	0.319*
	(1.46)	(1.10)	(0.88)	(1.45)	(1.88)
Ethnic Polarization	0.472**	0.433***	0.489**	0.571***	0.517**
	(2.42)	(2.79)	(2.50)	(2.73)	(2.27)
(Ethnic Frac. – 0.7)*(Dummy=1 if EF>0.7)	1.215	0.577	1.151	1.508*	1.522*
	(1.53)	(0.85)	(1.38)	(1.75)	(1.73)
(Dummy=1 if EF≤0.7)*(EF – 0.7)	0.117	0.317**	0.0598	0.0329	0.0879
	(0.65)	(2.09)	(0.32)	(0.16)	(0.42)
Onshore Oil Dummy	0.154**	0.150**	0.0745	0.0736	0.0156
	(2.09)	(2.38)	(0.99)	(0.88)	(0.18)
Cold War Dummy (1 if year <1991)	0.0327	-0.128	-0.0936	-0.0430	-0.141
	(0.15)	(-0.73)	(-0.39)	(-0.16)	(-0.59)
Linear Time Trend	0.0103***	0.0110***	0.0120***	0.00945***	0.0132***
	(2.85)	(3.79)	(3.60)	(2.68)	(3.03)
Post-1990 Time Trend	-0.00410	-0.00889**	-0.00735	-0.00378	-0.0107*
	(-0.74)	(-2.01)	(-1.19)	(-0.56)	(-1.76)
War (t–1)	3.031***	2.028***	2.812***	2.969***	2.611***
	(22.63)	(26.24)	(22.85)	(21.58)	(19.26)
War (t–2)	0.309**	0.943***	0.495***	0.533***	0.493***
	(2.22)	(12.00)	(3.88)	(3.77)	(3.48)
Constant	-2.212***	-1.836***	-1.824***	-2.398***	-1.952***
	(-5.58)	(-5.54)	(-4.53)	(-5.37)	(-4.42)
Observations	6787	6898	6076	6206	5832
Predicted Probability	0.0423	0.0707	0.0479	0.0408	0.0232
Pseudo R-Squared	0.7374	0.6136	0.7099	0.7608	0.6576
Area under ROC Curve	0.9682	0.9445	0.9606	0.9694	0.9598

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

Table 14: Including Ethnic Fractionalization only if > 0.7

Dependent Variable: Incidence of Civil War					
	Model 1	Model 2	Model 3	Model 4	Model 4
Estimation Method: Probit Model	PITF	UCDP/PRIO	Sambanis	FL	CH
Log GDP per capita	-0.161***	-0.131***	-0.156***	-0.151***	-0.180***
	(-6.26)	(-5.53)	(-5.78)	(-5.23)	(-6.18)
Log Population	0.0767***	0.0639***	0.0477*	0.0880***	0.0701**
	(2.67)	(2.85)	(1.75)	(2.85)	(2.28)
Share of Mountainous Terrain	0.235	0.134	0.139	0.256	0.313*
	(1.44)	(1.02)	(0.86)	(1.45)	(1.83)
Ethnic Polarization	0.540***	0.626***	0.524***	0.591***	0.567***
	(3.35)	(4.81)	(3.15)	(3.29)	(3.08)
(Ethnic Frac. – 0.7)*(Dummy if EF>0.7)	1.499**	1.317**	1.296*	1.588**	1.717**
	(2.22)	(2.36)	(1.84)	(2.25)	(2.35)
Onshore Oil Dummy	0.150**	0.138**	0.0722	0.0724	0.0147
	(2.05)	(2.21)	(0.97)	(0.88)	(0.17)
Cold War Dummy (1 if year < 1991)	0.0202	-0.166	-0.102	-0.0473	-0.145
	(0.09)	(-0.95)	(-0.42)	(-0.18)	(-0.61)
Linear Time Trend	0.0102***	0.0107***	0.0120***	0.00942***	0.0130***
	(2.85)	(3.71)	(3.61)	(2.69)	(3.01)
Post-1990 Time Trend	-0.00425	-0.00940**	-0.00750	-0.00384	-0.0106*
	(-0.77)	(-2.13)	(-1.22)	(-0.57)	(-1.75)
War (t–1)	3.030***	2.030***	2.811***	2.968***	2.610***
	(22.63)	(26.28)	(22.86)	(21.58)	(19.24)
War (t–2)	0.311**	0.944***	0.495***	0.533***	0.495***
	(2.24)	(12.01)	(3.89)	(3.78)	(3.49)
Constant	-2.264***	-1.972***	-1.851***	-2.413***	-1.996***
	(-5.80)	(-6.03)	(-4.64)	(-5.46)	(-4.57)
Observations	6787	6898	6076	6206	5832
Predicted Probability	0.0423	0.0710	0.0479	0.0409	0.0232
Pseudo R-Squared	0.7374	0.6129	0.7099	0.7608	0.6576
Area under ROC Curve	0.9685	0.9444	0.9607	0.9694	0.9603

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

Table 15: Onset and Continuation with Ethnic Polarization

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.187***	-0.148***	-0.191***	-0.175***	-0.225***
	(-7.23)	(-5.51)	(-6.92)	(-6.11)	(-7.61)
Log Population	0.0971***	0.0721***	0.0490*	0.0857***	0.0742**
	(3.23)	(2.79)	(1.71)	(2.61)	(2.44)
Ethnic Polarization	0.301	0.398**	0.300	0.477**	0.418*
	(1.41)	(2.33)	(1.38)	(2.11)	(1.69)
Ethnic Fractionalization	0.306*	0.547***	0.222	0.133	0.390**
	(1.75)	(3.94)	(1.25)	(0.69)	(1.97)
Onshore Oil Dummy	0.194**	0.0993	0.0672	0.0549	0.0779
	(2.11)	(1.25)	(0.73)	(0.55)	(0.70)
Cold War Dummy (if year < 1991)	0.0563	-0.139	-0.137	-0.127	-0.143
	(0.26)	(-0.82)	(-0.60)	(-0.53)	(-0.57)
Linear Time Trend	0.00983**	0.00998***	0.00891**	0.00541	0.0153***
	(2.44)	(3.07)	(2.30)	(1.35)	(2.97)
Post-1990 Time Trend	-0.00315	-0.00928**	-0.00858	-0.00604	-0.0104*
	(-0.56)	(-2.07)	(-1.40)	(-0.92)	(-1.65)
War (t–1)	2.584***	1.599**	1.252	1.550*	1.718*
	(3.06)	(2.29)	(1.57)	(1.72)	(1.81)
War (t–2)	0.310**	0.919***	0.453***	0.470***	0.486***
	(2.25)	(11.60)	(3.53)	(3.30)	(3.47)
Time trend * War (t–1)	-0.000585	0.00355	0.00728	0.0127**	-0.00781
	(-0.11)	(0.81)	(1.25)	(1.98)	(-1.12)
Constant	-2.219***	-1.960***	-1.385***	-1.922***	-1.787***
	(-5.17)	(-5.10)	(-3.17)	(-3.97)	(-3.73)
Variables included in the F-Test					
GDP * War (t–1)	0.0998	0.0709	0.131*	0.0755	0.199**
	(1.38)	(1.18)	(1.78)	(0.87)	(2.22)
Population * War (t–1)	-0.0397	0.00310	0.0330	0.0471	-0.00779
	(-0.60)	(0.06)	(0.51)	(0.60)	(-0.09)
Ethnic Pol. * War (t–1)	0.362	0.0429	0.268	0.0828	0.0115
	(0.89)	(0.13)	(0.63)	(0.18)	(0.02)
Ethnic Frac. * War (t–1)	-0.0814	-0.643***	-0.0893	0.136	-0.349
	(-0.28)	(-2.68)	(-0.30)	(0.44)	(-1.03)
Onshore Oil * War (t–1)	-0.130	0.119	-0.0174	0.0358	-0.137
	(-0.80)	(0.89)	(-0.11)	(0.19)	(-0.72)
Observations	6787	6898	6076	6206	5832
F-test (war_{t-1}*variable= 0) *p*-value	0.4586	0.0092	0.3644	0.8656	0.1665
Predicted Probability	0.0411	0.0683	0.0478	0.0423	0.0221
Pseudo R-Squared	0.7378	0.6163	0.7114	0.7620	0.6586
Area under ROC Curve	0.9676	0.9455	0.9609	0.9694	0.9578

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

7) Conclusions

Because of the existence of alternative data sources and variations in coding rules, it is possible to get a different picture of the causes of conflict from different data sets. Nevertheless we have shown that the main determinants of the incidence of civil war – poverty, a large population, mountainous terrain and ethnic diversity – consistently stand out as statistically significant across all data sets. Splitting incidence into onsets and continuing wars, we demonstrated that for most data sets the estimated equations were not significantly different, once we had controlled for the persistence of civil war. Thus, although conflicts can last a long time, the factors which tend to set civil wars going are similar to the ones that keep them going, and there is a danger of simply introducing sampling error by analyzing these two aspects separately.

We have also investigated the role of ethnic diversity in more detail. Although sub-Saharan Africa has proved particularly susceptible to conflict, and tends to have high ethnic diversity, we showed that the ethnic diversity coefficient is robust to the inclusion of a sub-Saharan Africa dummy. We failed to confirm the finding of Montalvo and Reynal-Querol (2005) that ethnic polarization dominates fractionalization. The evidence suggests that neither is a sufficient measure of diversity as it affects the probability of conflict. In a more detailed investigation, we found that polarization is the important factor at low or moderate levels of ethnic diversity (up to 0.7 on the fractionalization measure), but that it underpredicts the propensity to conflict at high levels of diversity, when polarization drops off quite dramatically.

References

Beck, Nathaniel, Jonathan N. Katz & Richard Tucker, 1998. 'Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable', *American Journal of Political Science* 42(4): 1260–88.

Bleaney, M.F. & Arcangelo Dimico, 2009. 'Ethnic Diversity and Local Conflicts', *University of Nottingham School of Economics* Discussion Paper no. 09/04.

Buhaug, H., S. Gates & P. Lujala, 2009. 'Geography, Rebel Capability, and the Duration of Civil Conflict', *Journal of Conflict Resolution* 53(4): 544-569.

Collier, Paul & Anke Hoeffler, 1998. 'On Economic Causes of Civil War', *Oxford Economic Papers*, 50(3): 563–73.

Collier, Paul & Anke Hoeffler, 2002. 'On the Incidence of Civil War in Africa', *Journal of Conflict Resolution* 46 (1), 13-28.

Collier, Paul & Anke Hoeffler, 2004. "Greed and Grievance in Civil War," *Oxford Economic Papers* 56, 563–95.

Collier, Paul & Anke Hoeffler, 2007. 'Civil War', in T. Sandler and K. Hartley (eds) *Handbook of Defense Economics*, Elsevier, Amsterdam (712–739).

Collier, P., A. Hoeffler and D. Rohner, 2009. 'Beyond Greed and Grievance: Feasibility and Civil War', *Oxford Economic Papers* 61(1):1-27.

Collier, P., A. Hoeffler & M. Söderbom, 2004. 'On the Duration of Civil War', *Journal of Peace Research* 41(3): 253–73.

Elbadawi, I. & N. Sambanis, 2002. 'How Much Civil War Will We See? Explaining the Prevalence of Civil War', *Journal of Conflict Resolution* 46(2): 307-334.

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

Fearon, J., 2004. 'Primary Commodity Exports and Civil War', *Journal of Peace Research* 41(1): 75-301.

Fearon, J., 2005. 'Why Do Some Wars Last So Much Longer Than Others?' *Journal of Conflict Resolution* 49(4): 483-507.

Gleditsch, Nils Petter; Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg & Håvard Strand, 2002. 'Armed Conflict 1946–2001: A New Dataset', *Journal of Peace Research* 39(5): 615–637.

Gleditsch, Nils Petter & Kristian Skrede , 2004. 'A Revised List of Wars Between and Within Independent States, 1816-2002', *International Interactions* 30(2): 231-262.

Harbom, Lotta & Peter Wallensteen, 2007. 'Armed Conflict, 1989–2006', *Journal of Peace Research* 44(5): 623–634.

Hegre, Håvard, Tanja Ellingsen, Scott Gates & Nils Petter Gleditsch, 2001. 'Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816–1992', *American Political Science Review* 95(1): 33–48.

Hegre, H., 2004. 'The Duration and Termination of Civil War', *Journal of Peace Research* 41 (3): 243-252.

Hegre, H. & N. Sambanis, 2006. 'Sensitivity Analysis of the Empirical Literature on Civil War Onset', *Journal of Conflict Resolution* 50(4), 508–535.

Janse, A. J. A., 2007. 'Global Rough Diamond Production from 1870 to 2005', *Gems & Gemology*, 2007(Fall): 136.

Janse, A.J.A. & Patricia A. Sheahan, 1995. 'Catalogue of World Wide Diamond and Kimberlite Occurrences: A Selective and Annotative Approach', *Journal of Geochemical Exploration* 53(1): 73-111.

Lacina, Bethany, 2006. 'Explaining the Severity of Civil Wars', *Journal of Conflict Resolution* 50(2): 276-289.

Lujala, Päivi, 2009. 'Deadly Combat Over Natural Resources: Gems, Petroleum, Drugs and the Severity of Armed Conflict', *Journal of Conflict Resolution* 53(1): 50-71.

Lujala, Päivi, 2010. 'The Spoils of Nature: Armed Civil Conflict and Rebel Access to Natural Resources', *Journal of Peace Research*, forthcoming.

Lujala, Päivi, Jan Ketil Rød, & Nadja Thieme, 2007. 'Fighting over oil: Introducing a new dataset', *Conflict Management and Peace Science* 24(3): 239-56.

Marshall, M. & Jaggers, K., 2000. *Polity IV Project: Codebook and Datafiles*. http://www.bsos.umd.edu/cidcm/inscr/polity.

Montalvo, Garcia-Jose & Marta Reynal-Querol, 2005. 'Ethnic Polarization, Potential Conflict, and Civil Wars', *American Economic Review* 95(3): 796-816

Montalvo, Garcia-Jose & Marta Reynal-Querol, 2008. 'Discrete Polarization with an Application to the Determinants of Genocides', *The Economic Journal* 118(532): 1835-1865.

Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, September 2006.

Political Instability Task Force, 2009. '*Internal Wars and Failures of Governance, 1955-2007*', Center for Global Policy in the School of Public Policy at George Mason University.

Ross, Michael, 2006. 'A Closer Look at Oil, Diamonds, and Civil War', *Annual Reviews in Political Science*, 9: 265-300.

Sambanis, N., 2004. 'What Is A Civil War? Conceptual and Empirical Complexities', *Journal of Conflict Resolution* 48(4): 814 – 858.

Schneider, Gerald & Nina Wiesehomeier, 2008. 'Rules That Matter: Political Institutions and the Diversity-Conflict Nexus', *Journal of Peace Research* 45(2): 183-203.

Singer, J.D. and M. Small, 1982. '*Resort to Arms: International and Civil War, 1816–1980*', Sage, Beverly Hills, CA.

Singer, J.D. and M. Small, 1994. 'Correlates of War Project: International and Civil War Data, 1816–1992', *Inter-University Consortium for Political and Social Research*, Ann Arbor, MI.

APPENDIX (not for publication)

This Appendix shows the effect of estimating the Table 3 regression with up to six lags of the dependent variable (Table A1) and using the spline procedure of Beck *et al*. (1998) (Table A2). In both cases the results are very similar to those shown in Table 3.

Table A1: Probit Model for Incidence – With Six Lags of the Dependent Variable

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit Model	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	UCDP/PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.140***	-0.0829***	-0.126***	-0.112***	-0.145***
	(-4.91)	(-3.20)	(-4.25)	(-3.57)	(-4.72)
Log Population	0.0563*	0.0440*	0.0392	0.0669**	0.0493
	(1.90)	(1.88)	(1.31)	(2.03)	(1.48)
Share of Mountainous Terrain	0.383**	0.304**	0.226	0.418**	0.555***
	(2.30)	(2.24)	(1.30)	(2.27)	(3.23)
Ethnic Fractionalization	0.357***	0.480***	0.382***	0.351**	0.404***
	(2.77)	(4.38)	(2.75)	(2.47)	(2.74)
Anocracy, Lagged	0.110	0.0997	0.186*	0.189*	0.144
	(1.18)	(1.26)	(1.94)	(1.83)	(1.41)
Onshore Oil Dummy	0.184**	0.140**	0.0987	0.0762	0.0613
	(2.37)	(2.10)	(1.19)	(0.86)	(0.65)
Surface Diamond Deposits Dummy	0.0604	-0.00939	-0.0161	0.0286	0.0415
	(0.54)	(-0.10)	(-0.13)	(0.23)	(0.34)
Cold War Dummy (1 if year <1991)	0.0933	-0.163	-0.413	-0.270	-0.0725
	(0.36)	(-0.77)	(-1.38)	(-0.83)	(-0.27)
Linear Time Trend	0.00967**	0.00899***	0.0143***	0.00857**	0.0138***
	(2.32)	(2.67)	(3.72)	(2.09)	(2.62)
Post-1990 Time Trend	-0.00285	-0.0101*	-0.0164**	-0.00942	-0.00924
	(-0.44)	(-1.94)	(-2.21)	(-1.17)	(-1.34)
War (t-1)	3.058***	1.930***	2.909***	3.106***	2.635***
	(20.73)	(22.47)	(21.04)	(19.48)	(18.21)
War (t - 2)	0.233	0.570***	0.233	0.312	0.124
	(1.05)	(5.06)	(1.10)	(1.24)	(0.64)
War (t – 3)	-0.00342	0.376***	0.409*	0.150	0.641***
	(-0.01)	(3.00)	(1.66)	(0.57)	(3.07)
War (t – 4)	0.0301	0.0335	-0.192	-0.358	0.119
	(0.13)	(0.23)	(-0.83)	(-1.38)	(0.50)
War (t – 5)	-0.0805	0.236*	-0.0123	0.311	-0.541**
	(-0.35)	(1.68)	(-0.06)	(1.19)	(-2.27)
War (t – 6)	0.173	0.136	0.105	0.160	0.230
	(0.94)	(1.12)	(0.52)	(0.72)	(1.31)
Constant	-2.200***	-2.054***	-1.699***	-2.165***	-2.119***
	(-5.20)	(-5.66)	(-3.86)	(-4.43)	(-4.58)
Observations	6113	6231	5428	5553	5150
Pseudo R-Squared	0.7401	0.6304	0.7310	0.7745	0.6656
Area under ROC Curve	0.9689	0.9502	0.9644	0.9714	0.9610

Table A2: Probit Model for Incidence – With Cubic Splines and War (t-2)

	Dependent Variable: Incidence of Civil War				
Estimation Method: Probit Model	Model 1	Model 2	Model 3	Model 4	Model 5
	PITF	PRIO	Sambanis	FL	CHR
Log GDP per capita	-0.133***	-0.0982***	-0.127***	-0.124***	-0.160***
	(-4.80)	(-3.92)	(-4.30)	(-3.87)	(-5.21)
Log Population	0.0572**	0.0472**	0.0444	0.0778**	0.0374
	(2.01)	(2.09)	(1.61)	(2.53)	(1.15)
Share of Mountainous Terrain	0.318**	0.242*	0.205	0.323*	0.361**
	(1.98)	(1.86)	(1.29)	(1.86)	(2.09)
Ethnic Fractionalization	0.331***	0.448***	0.325**	0.333**	0.387***
	(2.64)	(4.24)	(2.49)	(2.45)	(2.71)
Anocracy, Lagged	0.0811	0.0851	0.126	0.0985	0.115
	(0.92)	(1.14)	(1.40)	(1.00)	(1.20)
Onshore Oil Dummy	0.149**	0.137**	0.0622	0.0713	-0.00768
	(2.00)	(2.16)	(0.81)	(0.86)	(-0.08)
Surface Diamond Deposits Dummy	0.0363	-0.0427	-0.0352	0.00468	0.0254
	(0.33)	(-0.47)	(-0.32)	(0.04)	(0.21)
Cold War Dummy (1 if year <1991)	0.0636	-0.152	-0.118	-0.101	-0.137
	(0.30)	(-0.88)	(-0.52)	(-0.40)	(-0.60)
Linear Time Trend	0.0143***	0.0157***	0.0180***	0.0178***	0.0227***
	(3.31)	(4.88)	(4.79)	(4.21)	(4.35)
Post-1990 Time Trend	-0.00397	-0.0109**	-0.00987	-0.00625	-0.0124*
	(-0.67)	(-2.36)	(-1.62)	(-0.89)	(-1.95)
War (t -1)	3.120***	2.250***	2.492***	2.903***	1.799***
	(12.24)	(12.03)	(11.32)	(11.87)	(4.05)
War (t -2)	0.305**	0.939***	0.325**	0.450***	0.161
	(1.98)	(9.00)	(2.22)	(2.83)	(0.93)
Years since last war	0.0842	0.244**	-0.0522	0.0481	-0.428
	(1.00)	(2.52)	(-0.55)	(0.56)	(-1.26)
Spline 1	0.000110	-0.000016	-0.000015	-0.000019	-0.0324
	(1.43)	(-0.39)	(-0.23)	(-0.21)	(-0.98)
Spline 2	0.00254	0.0126***	0.000328	0.00180	0.00627
	(1.01)	(3.02)	(0.12)	(0.83)	(0.71)
Spline 3	-0.00181	-0.0082***	-0.00067	-0.00172	0.000037
	(-0.79)	(-2.88)	(-0.30)	(-0.86)	(0.01)
Spline 4	0.000276	0.00284**	0.000527	0.00101	-0.000688
	(0.24)	(2.15)	(0.38)	(0.81)	(-0.30)
Spline 5	0.000211	-0.00082	-0.000224	-0.000539	0.000215
	(0.30)	(-1.36)	(-0.26)	(-0.64)	(0.16)
Spline 6	-0.000528	0.00027	0.00012	0.00027	0.000082
	(-0.92)	(0.84)	(0.22)	(0.39)	(0.12)
Constant	-2.364***	-2.265***	-1.539***	-2.355***	-0.870
	(-4.90)	(-5.56)	(-3.15)	(-4.32)	(-1.27)
Observations	6730	6840	6037	6165	5783
Pseudo R-Squared	0.7396	0.6190	0.7139	0.7625	0.6671
Area under ROC Curve	0.9684	0.9458	0.9621	0.9697	0.9605

Note to Tables A1 and A2: Robust z-statistics in parentheses: * p<0.01, ** p<0.05, * p<0.1