Oxford Economic Papers 56 (2004), 563–595 **563**
doi:10.1093/oep/gpf064

Greed and grievance in civil war

By Paul Collier* and Anke Hoeffler†

*Centre for the Study of African Economies, University of Oxford
†Centre for the Study of African Economies, University of Oxford, 21 Winchester Road, Oxford OX2 6NA; e-mail: anke.hoeffler@economics.ox.ac.uk

We investigate the causes of civil war, using a new data set of wars during 1960–99. Rebellion may be explained by atypically severe grievances, such as high inequality, a lack of political rights, or ethnic and religious divisions in society. Alternatively, it might be explained by atypical opportunities for building a rebel organization. While it is difficult to find proxies for grievances and opportunities, we find that political and social variables that are most obviously related to grievances have little explanatory power. By contrast, economic variables, which could proxy some grievances but are perhaps more obviously related to the viability of rebellion, provide considerably more explanatory power.

1. Introduction

Civil war is now far more common than international conflict: all of the 15 major armed conflicts listed by the Stockholm International Peace Research Institute for 2001 were internal (SIPRI, 2002).

In this paper we develop an econometric model which predicts the outbreak of civil conflict. Analogous to the classic principles of murder detection, rebellion needs both motive and opportunity. The political science literature explains conflict in terms of motive: the circumstances in which people want to rebel are viewed as sufficiently rare to constitute the explanation. In Section 2 we contrast this with economic accounts which explain rebellion in terms of opportunity: it is the circumstances in which people are able to rebel that are rare. We discuss measurable variables which might enable us to test between the two accounts and present descriptive data on the 79 large civil conflicts that occurred between 1960 and 1999. In Section 3 we use econometric tests to discriminate between rival explanations and develop an integrated model which provides a synthesis. Section 4 presents a range of robustness checks and Section 5 discusses the results.

This analysis considerably extends and revises our earlier work (Collier and Hoeffler, 1998). In our previous theory, we assumed that rebel movements incurred net costs during conflict, so that post-conflict pay-offs would be decisive. The core of the paper was the derivation and testing of the implication that high post-conflict pay-offs would tend to justify long civil wars. We now recognize that

this assumption is untenable: rebel groups often more than cover their costs during the conflict. Here we propose a more general theory which juxtaposes the opportunities for rebellion against the constraints. Our previous empirical analysis conflated the initiation and the duration of rebellion. We now treat this separately. This paper focuses on the initiation of rebellion.[1] Our sample is expanded from a cross-section analysis of 98 countries during the period 1960–92, to a comprehensive coverage of 750 five-year episodes over the period 1960–99, enabling us to analyse double the number of war starts. Further, we expand from four explanatory variables to a more extensive coverage of potential determinants, testing for robustness to select a preferred specification.

2. Rebellion: approaches and measures

2.1 Preferences, perceptions, and opportunities

Political science offers an account of conflict in terms of motive: rebellion occurs when grievances are sufficiently acute that people want to engage in violent protest. In marked contrast, a small economic theory literature, typified by Grossman (1991, 1999), models rebellion as an industry that generates profits from looting, so that 'the insurgents are indistinguishable from bandits or pirates' (Grossman, 1999, p.269). Such rebellions are motivated by greed, which is presumably sufficiently common that profitable opportunities for rebellion will not be passed up.[2] Hence, the incidence of rebellion is not explained by motive, but by the atypical circumstances that generate profitable opportunities. Thus, the political science and economic approaches to rebellion have assumed both different rebel motivation—grievance versus greed—and different explanations—atypical grievances versus atypical opportunities.

Hirshleifer (1995, 2001) provides an important refinement on the motive-opportunity dichotomy. He classifies the possible causes of conflict into preferences, opportunities, and perceptions. The introduction of perceptions allows for the possibility that both opportunities and grievances might be wrongly perceived. If the perceived opportunity for rebellion is illusory—analogous to the 'winners' curse'—unprofitability will cause collapse, perhaps before reaching our threshold for civil war. By contrast, when exaggerated grievances trigger rebellion, fighting does not dispel the misperception and indeed may generate genuine grievances.

Misperceptions of grievances may be very common: all societies may have groups with exaggerated grievances. In this case, as with greed-rebellion, motive would not explain the incidence of rebellion. Societies that experienced civil war would be distinguished by the atypical viability of rebellion. In such societies rebellions

[1] On the analysis of the duration of civil war see Collier *et al.* (2004) and Fearon (2004).

[2] By the 'Machiavelli Theorem' (Hirshleifer, 2001, p.10–11) no one will pass up a profitable opportunity to exploit someone else.

would be conducted by viable not-for-profit organizations, pursuing misperceived agendas by violent means.

Greed and misperceived grievance have important similarities as accounts of rebellion. They provide a common explanation—'opportunity' and 'viability' describe the common conditions sufficient for profit-seeking, or not-for-profit, rebel organizations to exist. On our evidence they are observationally equivalent since we cannot observe motives. They can jointly be contrasted with the political account of conflict in which the grievances that both motivate and explain rebellion are assumed to be well-grounded in objective circumstances such as unusually high inequality, or unusually weak political rights. We now turn to the proxies for opportunities and objective grievances.

2.2 Proxies for opportunity

The first step in an empirical investigation of conflict is a clear and workable definition of the phenomenon. We define civil war as an internal conflict with at least 1,000 combat-related deaths per year. In order to distinguish wars from massacres, both government forces and an identifiable rebel organization must suffer at least 5% of these fatalities. This definition has become standard following the seminal data collection of Small and Singer (1982) and Singer and Small (1994). We use an expanded and updated version of their data set that covers 161 countries over the period 1960–99 and identifies 79 civil wars, listed in Table 1. Our task is to explain the initiation of civil war using these data.[3]

We now consider quantitative indicators of opportunity, starting with opportunities for financing rebellion. We consider three common sources: extortion of natural resources, donations from diasporas, and subventions from hostile governments.[4]

Klare (2001) provides a good discussion of natural resource extortion, such as diamonds in West Africa, timber in Cambodia, and cocaine in Colombia. In Table 2, we proxy natural resources by the ratio of primary commodity exports to GDP for each of the 161 countries. As with our other variables, we measure at intervals of five years, starting in 1960 and ending in 1995. We then consider the subsequent five years as an 'episode' and compare those in which a civil war broke out ('conflict episodes') with those that were conflict-free ('peace episodes'). The descriptive statistics give little support to the opportunity thesis: the conflict episodes were on average slightly less dependent upon primary commodity exports than the peace episodes. However, there is a substantial difference in the dispersion. The peace episodes tended to have either markedly below-average or markedly above-average dependence, while the conflict episodes were grouped around the mean.[5] Possibly if

[3] Later in the paper (Table 6) we examine whether our results are sensitive to this definition of civil war. Sambanis (2002) comes to a similar conclusion.

[4] We list the data sources and definitions in the Appendix.

[5] The standard deviation of primary commodity exports is 0.11 for the conflict episodes and 0.19 for the peace episodes.

Table 1 Outbreaks of war

Country	Start of the war	End of the war	Previous war	Gdp sample	Secondary schooling sample
Afghanistan	04/78	02/92			
Afghanistan	05/92	Ongoing	*		
Algeria	07/62	12/62	*		
Algeria	05/91	Ongoing	*	*	*
Angola	02/61	11/75			
Angola	11/75	05/91	*	*	*
Angola	09/92	Ongoing	*	*	*
Azerbaijan	04/91	10/94			
Bosnia	03/92	11/95			
Burma/Myanmar	68	10/80	*	*	*
Burma/Myanmar	02/83	07/95	*	*	*
Burundi	04/72	12/73		*	*
Burundi	08/88	08/88	*	*	*
Burundi	11/91	ongoing	*	*	*
Cambodia	03/70	10/91	*		
Chad	03/80	08/88		*	
China	01/67	09/68	*	*	
Columbia	04/84	ongoing	*	*	*
Congo	97	10/97		*	*
Cyprus	07/74	08/74		*	
Dominican Rep.	04/65	09/65		*	*
El Salvador	10/79	01/92		*	*
Ethiopia	07/74	05/91		*	*
Georgia	06/91	12/93			
Guatemala	07/66	07/72	*	*	*
Guatemala	03/78	03/84	*	*	*
Guinea-Bissau	12/62	12/74			
India	08/65	08/65	*	*	*
India	84	94	*	*	*
Indonesia	06/75	09/82	*	*	*
Iran	03/74	03/75		*	*
Iran	09/78	12/79	*	*	*
Iran	06/81	05/82	*	*	*
Iraq	09/61	11/63	*		
Iraq	07/74	03/75	*	*	*
Iraq	01/85	12/92	*	*	*
Jordan	09/70	09/70		*	
Laos	07/60	02/73	*		
Lebanon	05/75	09/92	*		
Liberia	12/89	11/91		*	
Liberia	10/92	11/96	*		
Morocco	10/75	11/89	*	*	*
Mozambique	10/64	11/75			
Mozambique	07/76	10/92	*	*	*
Nicaragua	10/78	07/79		*	*
Nicaragua	03/82	04/90			
Nigeria	01/66	01/70		*	*

(*continued*)

Table 1 Continued

Country	Start of the war	End of the war	Previous war	Gdp sample	Secondary schooling sample
Nigeria	12/80	08/84	*	*	*
Pakistan	03/71	12/71		*	*
Pakistan	01/73	07/77			
Peru	03/82	12/96		*	*
Philippines	09/72	12/96	*	*	*
Romania	12/89	12/89		*	*
Russia	12/94	08/96			
Russia	09/99	Ongoing	*		
Rwanda	11/63	02/64			
Rwanda	10/90	07/94	*	*	*
Sierra Leone	03/91	11/96		*	*
Sierra Leone	05/97	07/99	*	*	
Somalia	04/82	05/88		*	*
Somalia	05/88	12/92	*	*	*
Sri Lanka	04/71	05/71		*	*
Sri Lanka	07/83	ongoing	*	*	*
Sudan	10/63	02/72			
Sudan	07/83	ongoing	*	*	*
Tajikistan	04/92	12/94			
Turkey	07/91	ongoing		*	*
Uganda	05/66	06/66		*	*
Uganda	10/80	04/88	*	*	*
Vietnam	01/60	04/75	*		
Yemen	05/90	10/94			
Yemen, Arab Rep.	11/62	09/69	*		
Yemen, People's Rep.	01/86	01/86	*		
Yugoslavia	04/90	01/92			
Yugoslavia	10/98	04/99	*		
Zaïre/Dem. Rep. of Congo	07/60	09/65			
Zaïre/Dem. Rep. of Congo	09/91	12/96	*	*	*
Zaïre/Dem. Rep. of Congo	09/97	09/99	*	*	*
Zimbabwe	12/72	12/79		*	*

Note: Previous wars include war starts 1945–1994.

natural resources are sufficiently abundant, as in Saudi Arabia, the government may be so well-financed that rebellion is militarily infeasible. This offsetting effect may make the net effect of natural resources non-monotonic.[6] The observed pattern may also reflect differences between primary commodities (which we defer to Section 3). Further, primary commodities are associated with other characteristics that may cause civil war, such as poor public service provision, corruption and economic mismanagement (Sachs and Warner, 2000). Potentially, any increase in conflict risk may be due to rebel responses to such poor governance rather than to financial opportunities.

[6] Collier (2000) provides an illustrative formal model of such a non-monotonic relationship.

Table 2 Descriptive statistics

	Sample ($n = 1167$)	No civil war ($n = 1089$)	Civil war ($n = 78$)
War starts	0.067	0	1
Primary commodity exports/GDP	0.168	0.169	0.149
GDP *per capita* (const. US$)	4061	4219	1645
Diaspora (relative to population of country of origin)	0.017	0.018	0.004
Male secondary schooling (% in school)	43.42	44.39	30.3
GDP *per capita* growth (average for previous 5 years)	1.62	1.74	−0.23
Previous war (% with war since 1945)	20.8	18.5	53.8
Peace duration (months since last conflict)	327	334	221
Forest cover (%)	31.11	31.33	27.81
Mountainous terrain (%)	15.82	15.17	24.93
Geographic concentration of the population (Gini)	0.571	0.569	0.603
Population density (inhabitants per square km)	150	156	62
Population in urban areas (%)	45.11	46.00	32.7
Ethnic fractionalization (index, 0–100)	39.57	38.64	52.63
Religious fractionalization (index, 0–100)	36.09	35.98	37.70
Polarization $\alpha = 1.6$ (index, 0–0.165)	0.077	0.077	0.076
Democracy (index, 0–10)	3.91	4.07	1.821
Ethnic dominance (% with main ethnic group 45–90%)	0.465	0.465	0.452
Income inequality (Gini)	0.406	0.406	0.410
Land inequality (Gini)	0.641	0.641	0.631

Note: We examine 78—rather than the 79—war starts as listed in Table 1 because Pakistan experienced two outbreaks of war during 1970–74. We only include one of these war starts to avoid double counting.

A second source of rebel finance is from diasporas. Angoustures and Pascal (1996) review the evidence, an example being finance for the Tamil Tigers from Tamils in north America. We proxy the size of a country's diaspora by its emigrants living in the US, as given in US Census data. Although this neglects diasporas living in other countries, it ensures uniformity in the aggregate: all diasporas are in the same legal, organizational, and economic environment. We then take this emigrant population as a proportion of the population in the country of origin. In the formal analysis we decompose the diaspora into that part induced by conflict and that which is exogenous to conflict, but here we simply consider the crude numbers. These do not support the opportunity thesis: diasporas are substantially smaller in the conflict episodes.

A third source of rebel finance is from hostile governments. For example, the government of Southern Rhodesia pump-primed the Renamo rebellion in Mozambique. Our proxy for the willingness of foreign governments to finance military opposition to the incumbent government is the Cold War. During the Cold War each great power supported rebellions in countries allied to the opposing

power. There is some support for the opportunity thesis: only eleven of the 79 wars broke out during the 1990s.

We next consider opportunities arising from atypically low cost. Recruits must be paid, and their cost may be related to the income forgone by enlisting as a rebel. Rebellions may occur when foregone income is unusually low. Since non-economists regard this as fanciful we give the example of the Russian civil war. Reds and Whites, both rebel armies, had four million desertions (the obverse of the recruitment problem). The desertion rate was ten times higher in summer than in winter: the recruits being peasants, income foregone were much higher at harvest time (Figes, 1996). We try three proxies for foregone income: mean income *per capita*, male secondary schooling, and the growth rate of the economy. As shown in Table 2, the conflict episodes started from less than half the mean income of the peace episodes. However, so many characteristics are correlated with *per capita* income that, depending upon what other variables are included, the proxy is open to other interpretations. Our second proxy, male secondary school enrollment, has the advantage of being focused on young males—the group from whom rebels are recruited. The conflict episodes indeed started from lower school enrollment, but this is again open to alternative interpretation: education may affect the risk of conflict through changing attitudes. Our third measure, the growth rate of the economy in the preceding period, is intended to proxy new income opportunities. Conflict episodes were preceded by lower growth rates. This is consistent with evidence that the lower is the rate of growth, the higher is the probability of unconstitutional political change (Alesina *et al.*, 1996).[7] Although the three proxies are all consistent with atypically low forgone income as an opportunity, low income could also be interpreted as an objective economic grievance.

The opportunity for rebellion may be that conflict-specific capital (such as military equipment) is unusually cheap. We proxy the cost of such capital by the time since the most recent previous conflict: the legacy of weapon stocks, skills, and organizational capital will gradually depreciate. Empirically, peace episodes are preceded by far longer periods of peace than conflict episodes (Table 2). While this supports the opportunity thesis, it could also be interpreted as reflecting the gradual decay of conflict-induced grievances.

Another dimension of opportunity is an atypically weak government military capability. An unambiguous indicator is if the terrain is favorable to rebels: forests and mountains provide rebels with a safe haven. We measure the proportion of a country's terrain that is forested, using FAO data. We could find no equivalent data

[7] The economic growth literature concentrates on the analysis of political instability as a determinant of economic growth (see for example Barro 1991, 1997). Alesina *et al.* (1996) estimate a simultaneous equation system of economic growth and political instability. They present support for the hypothesis that political instability reduces growth. Lower growth does not seem to cause political instability, defined as the number of government changes. However, when they define political instability more narrowly as unconstitutional government changes they find that lower growth rates are a causal factor of political instability.

on mountainous terrain: proxies such as altitude tend to misclassify both plateaus and rugged uplands. We therefore commissioned a new index from a specialist, John Gerrard. The descriptive statistics (Table 2) suggest that terrain may matter: in conflict episodes 25% of the terrain is mountainous, versus only 15% in peace episodes, although there is no difference in forest cover. Geographic dispersion of the population may also inhibit government capability: Herbst (2000) suggests that Zaire is prone to rebellion because its population lives around the edges of the country. We measure dispersion by calculating a Gini coefficient of population dispersion.[8] In fact, the concentration of the population is slightly lower prior to peace episodes (0.57) than prior to war episodes (0.6). Similarly, low population density and low urbanization may inhibit government capability. Empirically, prior to war episodes both population density and urbanization are low (Table 2).

A final source of rebel military opportunity may be social cohesion. Ethnic and religious diversity within organizations tends to reduce their ability to function (Easterly and Levine, 1997, Alesina *et al.*, 1999, Collier, 2001). A newly formed rebel army may be in particular need of social cohesion, constraining recruitment to a single ethnic or religious group. A diverse society might in this case reduce the opportunity for rebellion by limiting the recruitment pool. The most widely used measure of ethnic diversity is the index of ethno-linguistic fractionalization. This measures the probability that two randomly drawn people will be from different ethnic groups. We could find no measure of religious fractionalization, but we constructed one equivalent to that of ethnic fractionalization using data from Barrett (1982). If ethnic and religious divisions are cross-cutting, social fractionalization is multiplicative rather than additive. We could find no data relating religious and ethnic divisions and so we construct a proxy that measures the maximum potential social fractionalization.[9] The thesis that social cohesion enhances opportunity is not supported by the descriptive statistics: conflict episodes have atypically high fractionalization. This seems more consistent with a grievance interpretation, to which we now turn.

2.3 Proxying objective grievances

We consider four objective measures of grievance: ethnic or religious hatred, political repression, political exclusion, and economic inequality.

[8] For the calculation of the Gini coefficient we used the population data per 400 km^2 cell. Analogous to the income Gini coefficient, the Gini coefficient of population dispersion will be high if the population is concentrated in a relatively small area of the country.

[9] If there were e equally sized ethnic groups and r equally sized religious groups, maximum potential social fractionalization would be measured simply by the product er. Since both the underlying indices of ethnic and religious fractionalization range on the scale 0–100, their product is zero if there is either religious or ethnic homogeneity whereas there is social homogeneity only if both indices are zero. We therefore measure social fractionalization as the product of the underlying indices plus whichever index is the greater.

Ethnic and religious hatreds are widely perceived as a cause of civil conflict. Although such hatreds cannot be quantified, they can evidently only occur in societies that are multi-ethnic or multi-religious and so our proxies measure various dimensions of diversity. Our previously discussed measures of fractionalization are pertinent: inter-group hatreds must be greater in societies that are fractionalized than in those which are homogenous. However, arguably the source of inter-group tension is not diversity but polarization.[10] Fortunately, the allowable class of measures of polarization is quite limited. We adopt a general measure due to Esteban and Ray (1994)

$$P = K \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i^{1+\alpha} \pi_j d \tag{1}$$

where π_i denotes the percentage of people that belong to group i in the total population, $i = 1, \ldots, n$. This measure of polarization depends on the parameters K and α. K does not change the order, but is used for population normalization. Esteban and Ray show that α is bounded between zero and 1.6. We calculate the polarization measure for three different values of α, 0, 0.8 and 1.6, using primary data on ethnic composition.[11] In addition we investigate the variant of the Esteban-Ray measure proposed by Reynal-Querol. These measures indeed distinguish polarization from fractionalization: their correlation coefficient ranges between 0.39 ($\alpha = 1.6$) and 1.0 ($\alpha = 0$).[12] The descriptive data does not suggest that polarization is important: conflict and peace episodes have very similar mean values (Table 2).

We measure political repression using the Polity III data set (see Jaggers and Gurr, 1995). This measure of political rights ranges 0–10 on an ascending ordinal scale. Political rights differ considerably between conflict and peace episodes. We also investigate the Polity III measure of autocracy, and a measure of political openness published by Freedom House (the 'Gastil Index'). The quantitative political science literature has already applied these measures to conflict risk. Hegre *et al.* (2001) find that repression increases conflict except when it is severe.

Even in democracies a small group may fear permanent exclusion. A potentially important instance is if political allegiance is based on ethnicity and one ethnic group has a majority. The incentive to exploit the minority increases the larger is the minority, since there is more to extract (Collier, 2001). Hence, a minority

[10] The link from polarization to conflict is proposed by Esteban and Ray (1999) and Reynal-Querol (2000) and is common in the popular literature.

[11] Our data source was *Atlas Narodov Mira*, USSR (1964). The Esteban-Ray measure includes a coefficient d that denotes the degree of antagonism between two different ethnic groups. Obviously, in large samples such as we are using this is not observed. Following Reynal-Querol (2000) we assume that the distance between any two ethnic groups is unity whereas that within the group is zero, so that d has the properties: $d = 1$ if $i \neq j$ and $d = 0$ if $i = j$.

[12] For $\alpha = 0$ the polarization measure is equal to the Gini coefficient.

may be most vulnerable if the largest ethnic group constitutes a small majority. We term this ethnic dominance. In Table 2 we define it as occurring if the largest ethnic group constitutes 45–90% of the population. On this definition it does not appear important: it is as common in peace episodes as in conflict episodes.

The opening page of Sen's *On Economic Inequality* (Sen, 1973) asserts that 'the relation between inequality and rebellion is indeed a close one'. The poor may rebel to induce redistribution, and rich regions may mount secessionist rebellions to pre-empt redistribution.[13] We measure income inequality by the Gini coefficient and by the ratio of the top-to-bottom quintiles of income. We measure asset inequality by the Gini coefficient of land ownership. The data are from Deininger and Squire (1996, 1998). Inequality is slightly higher prior to the conflict episodes.

2.4 Scale effects

Our measures of opportunity, such as primary commodity exports, income, and school enrolment, are scaled by measures of country size. For given values of these variables, opportunities should be approximately proportional to size. Grievance might also increase with size: public choices diverge more from the preferences of the average individual as heterogeneity increases.[14] We are, however, able to control for three aspects of heterogeneity: ethnic, religious and income diversity. Empirically, the conflict episodes had markedly larger populations than the peace episodes.

3. Regression analysis

As set out above, the proxies for opportunity and objective grievances are largely distinct and so can be compared as two non-nested econometric models. There is, however, no reason for the accounts to be exclusive and the aim of our econometric tests is to arrive at an integrated model which gives an account of conflict risk in terms of all those opportunities and grievances that are significant.

We now attempt to predict the risk that a civil war will start during a five-year episode, through a logit regression. Our dependent variable, civil war start, takes a value of one if a civil war started during a five year episode (1965–69, . . . , 1994–99). Episodes that were peaceful from the beginning until the end are coded zero. Ongoing wars are coded as missing observations as to not conflate the analysis of civil war initiation and duration.[15] If a war ended and another one started in the same period we coded these events as one. Some of our explanatory variables are

[13] This is analogous to the theory of tax exit proposed by Buchanan and Faith (1987).

[14] Mounting diversity is the offset to scale economies in the provision of public goods in the model of optimal county size proposed by Alesina and Spolaore (1997).

[15] This approach contrasts with our initial work (Collier and Hoeffler, 1998) in which we used a tobit procedure to study the duration of civil war (on a much inferior data set) and argued that the same factors that determined duration would determine the risk of initiation. Collier *et al.* (2004) establishes that this is wrong: initiation and duration are radically different processes.

time invariant. Those that are not are measured either for the first year of the period (e.g. 1965) or during the preceding five years (e.g. growth during 1960–64) in order to avoid endogeneity problems. Our results rest on how societies that experienced an outbreak of war differed from those that sustained peace.

We start with the opportunity model (see Table 3). The first regression (column 1) excludes *per capita* income and diasporas. Because *per capita* income and enrollment

Table 3 Opportunity model

	1	2	3	4	5
Primary commodity exports/GDP	18.149 (6.006)***	18.900 (5.948)***	16.476 (5.207)***	17.567 (6.744)***	17.404 (6.750)***
(Primary commodity exports/GDP)^2	−27.445 (11.996)***	−29.123 (11.905)***	−23.017 (9.972)**	−28.815 (15.351)*	−28.456 (15.366)*
Post-coldwar	−0.326 (0.469)	−0.207 (0.450)	−0.454 (0.416)		
Male secondary schooling	−0.025 (0.010)**	−0.024 (0.010)**			
Ln GDP *per capita*			−0.837 (0.253)***	−1.237 (0.283)***	−1.243 (0.284)***
GDP growth	−0.117 (0.044)***	−0.118 (0.044)***	−0.105 (0.042)***		
Peace duration	−0.003 (0.002) $p = 0.128$	−0.004*** (0.001)	−0.004 (0.001)***	−0.002 (0.001)	−0.002 (0.001)
Previous war	0.464 (0.547) $p = 0.396$				
Mountainous terrain	0.013 (0.009) $p = 0.164$	0.014 (0.009)	0.008 (0.008)		
Geographic dispersion	−2.211 (1.038)**	−2.129 (1.032)**	−0.865 (0.948)		
Social fractionalization	−0.0002 (0.0001) $p = 0.109$	−0.0002 (0.0001) $p = 0.122$	−0.0002 (0.0001)**		
Ln population	0.669 (0.163)***	0.686 (0.162)***	0.493 (0.129)***	0.295 (0.141)**	0.296 (0.141)**
Diaspora/peace				700.931 (363.29)**	
Diaspora corrected/peace					741.155 (387.636)*
(Diaspora-diaspora corrected)/peace					823.941 (556.024)
N	688	688	750	595	595
No of wars	46	46	52	29	29
Pseudo R^2	0.24	0.24	0.22	0.25	0.25
Log likelihood	−128.49	−128.85	−146.86	−93.27	−93.23

Notes: All regressions include a constant. Standard errors in parentheses. ***, **, * indicate significance at the 1, 5, and 10% level, respectively.

in secondary schooling are highly correlated, they cannot be used in the same regression ($\rho = 0.8$). Our diaspora measure is available only for 29 war episodes and so we explore it as an addendum. The variables included in the first regression permit a sample of 688 episodes (from 123 countries), including 46 wars.

Primary commodity exports are highly significant. Although their effect is non-linear, the risk of conflict peaks when they constitute around 33% of GDP,[16] which is a high level of dependence. The other proxy for finance, the end of the Cold War, has the expected sign but is insignificant. The foregone earnings proxies are also both significant with the expected sign: secondary schooling and growth both reduce conflict risk. Our proxy for the cost of conflict-specific capital is the number of months since any previous conflict (back to 1945). To distinguish between this interpretation and the danger that the proxy might be picking up fixed effects, we add a dummy variable that is unity if there was a previous conflict post-1945. Our proxy has the expected sign and is on the borderline of significance, while the dummy variable is completely insignificant. When the dummy variable is dropped (column 2) the proxy becomes highly significant and no other results are changed. The proxies for military advantage also have the expected sign and are marginally significant: mountainous terrain, population dispersion and social fractionalization. Finally, the coefficient on population is positive and highly significant.

The third column replaces secondary schooling with *per capita* income. This permits a larger sample—750 episodes (from 125 countries) including 52 wars. *Per capita* income is highly significant with the expected negative sign. However, the change of variable and the expansion of sample have little effect on the other results—social fractionalization becomes significant and population dispersion loses significance. There is little to chose between these two variants of the model—secondary schooling gives a slightly better fit, but *per capita* income permits a slightly larger sample.

In the last two columns of Table 3 we introduce our diaspora variable. Since many observations are missing, the number of war episodes with complete data is radically reduced. In order to preserve sample size we therefore retreat to a more parsimonious version of the model. We drop four sample-constraining peripheral explanatory variables: social fractionalization, population dispersion, mountainous terrain, and the rate of growth in the previous episode. The remaining explanatory variables are thus *per capita* GDP, primary commodity exports, population, and the number of months since the previous conflict. Even with these data-restoring deletions, the sample size is reduced to 29 war episodes (and 595 observations). However, all the included explanatory variables remain significant. The size of the diaspora is not directly significant in the initiation of conflict (column 4). However, it is significant when interacted with the number of months since the previous conflict. 'Diaspora/peace' divides the size of the diaspora by the time since a

[16] We differentiate the probability of civil war with respect to primary commodity exports and find that the risk is at its maximum at 33% of primary exports in GDP ($18.149/(2*27.445) = 0.33$).

previous conflict. The variable is positive and significant: a large diaspora considerably increases the risk of repeat conflict.

While this result may indicate that diasporas increase the risk of conflict through their finance of rebel organizations, it is also open to a more anodyne interpretation. Diasporas are endogenous to the intensity of conflict: when civil war occurs, people emigrate to the USA. Hence, the size of the diaspora might be proxying the intensity of conflict. The result may therefore be spurious: intense conflicts may have a higher risk of repetition. To test for this we decomposed observed diasporas into a component which is exogenous to the intensity of conflict and a residual endogenous component. For this decomposition we estimated a migration model, reported in Appendix 1. The size of the diaspora in a census year is predicted to be a function of its size in the previous census, time, *per capita* income in the country of origin, and whether there was a war in the intervening period. This model predicts the size of the diaspora with reasonable accuracy. For years subsequent to a conflict we replace the actual data on the size of the diaspora with an estimate from this regression. Thus, all post-conflict observations of diasporas are estimates which are purged of any effect of the intensity of conflict. The difference between actual and estimated figures is then used as an additional variable, measuring that part of the diaspora which is potentially endogenous to the intensity of conflict. Both of these measures are then introduced into the regression in place of the previous single measure of the diaspora. The results are reported in the final column of Table 3. The purged measure of the diaspora remains significant, and the size of the coefficient is only slightly altered (it is not significantly different from that on the endogenous diaspora measure). This suggests that there is indeed a substantial causal effect of the diaspora on the risk of conflict renewal. The result also guides our interpretation of why the risk of conflict repetition declines as peace is maintained. Recall that in principle this could be either because hatreds gradually fade, or because 'rebellion-specific capital' gradually depreciates. How might diasporas slow these processes? Diasporas preserve their own hatreds: that is why they finance rebellion. However, it is unlikely that the diaspora's hatreds significantly influence attitudes among the much larger population in the country of origin. By contrast, the finance provided by the diaspora can offset the depreciation of rebellion-specific capital, thereby sustaining conflict risk.

In Table 4 we turn to objective grievance as the explanation of rebellion, dropping all the economic measures of opportunity.[17] We retain the number of months since a previous conflict, since (subject to our discussion above) this can be interpreted as proxying fading hatreds. In the first column we also exclude the inequality measures due to considerations of sample size. This enables a very large sample of 850 episodes and 59 civil wars.

[17] We retain the two geographic measures, population dispersion and mountainous terrain. Although their exclusion does not affect the results, non-economists often find the proposition that geographic opportunity affects conflict plausible and inoffensive, while contesting the role of economic opportunity. We retain population size as a scale variable.

Table 4 Grievance model

	1	2	3
Ethnic fractionalization	0.010	0.011	0.012
	(0.006)*	(0.007)*	(0.008)
Religious fractionalization	−0.003	−0.006	−0.004
	(0.007)	(0.008)	(0.009)
Polarization $\alpha = 1.6$	−3.067	−4.682	−6.536
	(7.021)	(8.267)	(8.579)
Ethnic dominance (45–90%)	0.414	0.575	1.084
	(0.496)	(0.586)	(0.629)*
Democracy	−0.109	−0.083	−0.121
	(0.044)***	(0.051)*	(0.053)**
Peace duration	−0.004	−0.003	−0.004
	(0.001)***	(0.001)***	(0.001)***
Mountainous terrain	0.011	0.007	−0.0001
	(0.007)	(0.009)	(0.009)
Geographic dispersion	−0.509	−0.763	−1.293
	(0.856)	(1.053)	(0.102)
Ln population	0.221	0.246	0.300
	(0.096)**	(0.119)**	(1.133)**
Income inequality		0.015	
		(0.018)	
Land inequality			0.461
			(1.305)
N	850	604	603
No of wars	59	41	38
Pseudo R^2	0.13	0.11	0.17
Log likelihood	−185.57	−133.46	−117.12

Notes: All regressions include a constant. Standard errors in parentheses. ***, **, * indicate significance at the 1, 5 and 10% level, respectively.
Column 1: the two measures of fractionalization and ethnic dominance are not jointly significant.

The four proxies for ethnic and religious tension are surprisingly unimportant in view of the attention that the phenomenon attracts. Ethnic fractionalization is significant at 10% with the expected sign. Religious fractionalization and polarization are insignificant with the wrong sign, and ethnic dominance is insignificant. Nor are the three measures jointly significant.[18] Democracy is highly significant with the expected sign—repression increases conflict risk. The time since the previous conflict is again highly significant, but we have suggested that this is more likely to be proxying rebellion-specific capital than grievance. In the second and third columns we introduce income inequality and land inequality respectively. Although the sample size is reduced, it is still substantial—over 600

[18] At this stage we measure polarization with $\alpha = 1.6$ and define ethnic dominance as occurring when the largest ethnic group constitutes 45–90% of the population. These specifications are justified in Section 4 where we investigate robustness to alternative definitions.

episodes of which 41 (income) and 38 (land) are wars. Neither variable is close to significance.[19] All three grievance models have very low explanatory power with a pseudo R^2 of 0.17 or lower.

We now turn to the question of which model, opportunity or grievance, provides a better explanation of the risk of civil war. Since the two models are non-nested, i.e. one model is not a special case of the other, we use the J-test as suggested by Davidson and MacKinnon (1981). As shown in the first two columns of Table 5, we find that we cannot reject one model in favor of the other.[20] Thus, we conclude that while the opportunity model is superior, some elements of the grievance model are likely to add to its explanatory power. We therefore investigate the combination of the two models as presented in column 3 of Table 5.

Since this combined model includes income inequality and a lagged term, our sample size is much reduced (479 observations). In column 4 we drop inequality (which is consistently insignificant). Omitting inequality increases the sample size to 665. In this combined model neither democracy, ethnic and religious fractionalization nor the post-Cold War dummy are significant. Other variables are statistically significant or close to significance and the overall fit is reasonable (pseudo R^2 of 0.26). Since both the grievance and opportunity models are nested in the combined model, we can use a likelihood ratio test to determine whether the combined model is superior. We can reject the validity of the restrictions proposed by the grievance model, but not by the opportunity model.[21]

Although the combined model is superior to the opportunity and grievance models, several variables are completely insignificant and we drop them sequentially. First we exclude the post-Cold War dummy, then religious fractionalization,

[19] We also tried the ratio of the income shares of the top to the bottom quintiles. This was also insignificant.

[20] The J-test is based on the following artificial nesting procedure. First we explain the risk of civil war, p, in terms of the two different models, opportunity and grievance.
(1) $p = \text{f(opportunity)}$
(2) $p = \text{f(grievance)}$
Based on these logit regressions we calculate the predicted probabilities and add these predicted values, $\hat{p}^{\text{opportunity}}$ and $\hat{p}^{\text{grievance}}$ to our alternative models.
(1) $p = \text{f(opportunity, } \hat{p}^{\text{grievance}})$
(2) $p = \text{f(grievance, } \hat{p}^{\text{opportunity}})$
According to the J-test the significance of the coefficients of these added variables enables us to choose between the two different models. If $\hat{p}^{\text{grievance}}$ is significant in the opportunity model we reject the opportunity model in favor of the grievance model. If $\hat{p}^{\text{opportunity}}$ is significant in the grievance model we reject the grievance model in favor of the opportunity model. As can be seen in columns 1 and 2 of Table 5, $\hat{p}^{\text{grievance}}$ is significant in the opportunity model and $\hat{p}^{\text{opportunity}}$ is significant in the grievance model.

[21] Using the same sample as for the combined model ($n = 665$) we obtain the following results: Opportunity model versus combined model, 5 degrees of freedom, Likelihood Ratio Test (LRT) statistic 7.85 ($p = 0.165$); grievance model versus combined model, 6 degrees of freedom, LRT statistic 29.64 ($p = 0.000$).

Table 5 Combined opportunity and grievance model

	1	2	3	4	5	6	7
Primary commodity exports/GDP	19.107		37.072	23.385	18.937	16.773	50.608
	(5.996)***		(10.293)***	(6.692)***	(5.865)***	(5.206)***	(14.09)***
(Primary commodity exports/GDP)^2	−30.262		−69.270	−36.335	−29.443	−23.800	−131.00
	(12.008)***		(21.697)***	(12.998)***	(11.781)***	(10.040)**	(42.93)***
Post-coldwar	−0.208		−0.873	−0.281			
	(0.457)		(0.644)	(0.459)			
Male secondary schooling	−0.021		−0.029	−0.022	−0.031		−0.034
	(0.011)**		(0.013)**	(0.011)**	(0.010)***		(0.011)***
Ln GDP *per capita*						−0.950	
						(0.245)***	
(GDP growth) $t-1$	−0.108		−0.045	−0.108	−0.115	−0.098	−0.113
	(0.044)***		(0.062)	(0.045)**	(0.043)***	(0.042)**	(0.046)***
Peace duration	−0.0003	0.0005	−0.0003	−0.003	−0.004	−0.004	−0.003
	(0.002)	(0.0014)	(0.0015)	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Mountainous terrain	0.005	0.001	0.005	0.015			
	(0.010)	(0.008)	(0.012)	(0.009)			
				$p=0.11$			
Geographic dispersion	−1.976	0.053	−4.032	−1.962	−2.487	−0.992	−2.871
	(1.049)*	(1.101)	(1.490)***	(1.149)*	(1.005)**	(0.909)	(1.130)***
Ln population	0.489	−0.022	0.927	0.697	0.768	0.510	1.123
	(0.193)**	(0.136)	(0.250)***	(0.181)***	(0.166)***	(0.128)***	(0.226)***
Social fractionalization	−0.0002		−0.0008	−0.0005	−0.0002	−0.0002	−0.0003
	(0.0001)***		(0.0003)**	(0.0003)	(0.0001)**	(0.0001)***	(0.0001)***
				$p=0.11$			
Ethnic fractionalization		0.008	0.041	0.023			
		(0.007)	(0.019)**	(0.015)			
Religious fractionalization		−0.005	0.015	0.014			
		(0.008)	(0.020)	(0.019)			
Polarization		−9.338	−25.276	−15.992			

		(8.734)	(13.390)*	(10.518)			
Ethnic dominance (45–90%)		1.210	2.020	1.592	0.670	0.480	0.769
		(0.648)*	(0.915)**	(0.746)**	(0.354)*	(0.328)	(0.369)**
						$p = 0.14$	
Democracy		−0.036	−0.018	−0.042			
		(0.054)	(0.062)	(0.054)			
Income inequality			0.025				
			(0.024)				
Grievance predicted value	0.765						
	(0.413)*						
Opportunity predicted value		1.044					
		(0.211)***					
Primary commodity exports/GDP* oil dummy							−28.275
							(9.351)***
(Primary commodity exports/GDP)^2* oil dummy							106.459
							(38.704)***
N	665	665	479	665	688	750	654
No of wars	46	46	32	46	46	52	45
Pseudo R^2	0.24	0.25	0.24	0.26	0.24	0.22	0.30
Log likelihood	−126.69	−125.29	−89.55	−124.60	−128.21	−146.84	−114.20

Notes: All regressions include a constant. Standard errors in parentheses. ***, **, * indicate significance at the 1, 5, and 10% level, respectively.

then democracy,[22] then polarization, then ethnic fractionalization and finally mountainous terrain, yielding the baseline model of column 5 and its variant with *per capita* income replacing secondary enrolment in column 6. No further reduction in the model is accepted, and no additions of variables included in our previous models are accepted. The baseline model and its variant yield very similar results although the variant has less explanatory power and two variables lose significance (ethnic dominance and geographic dispersion).

Our baseline model allows us to calculate the change in the probability of war-starts for different values of the explanatory variables. We present these calculations in Appendix Table A2. At the mean of all variables the risk of a war-start is about 11.5%. Our model predicts that a hypothetical country with all the worst characteristics found in our sample would have a near-certain risk of war, while one with all the best characteristics would have a negligible risk. We now calculate how each variable affects the risk of civil war (keeping all other variables at their mean values).

The effect of primary commodity exports on conflict risk is both highly significant and considerable. At peak danger (primary commodity exports being 33% of GDP), the risk of civil war is about 22%, while a country with no such exports has a risk of only 1%. The effect is sufficiently important to warrant disaggregation into different types of commodities. We categorized primary commodity exports according to which type of product was dominant: food, non-food agriculture, oil, other raw materials, and a residual category of 'mixed'.[23] Of the many potential disaggregations of primary commodity exports permitted by this data, only one was significant when introduced into our baseline regression, namely oil versus non-oil. The results are reported in column 7. We add variables that interact the primary commodity export share and its square with a dummy variable that takes the value of unity if the exports are predominantly oil. Both variables are significant: oil exports have a distinct effect on the risk of conflict. However, the effect is modest: at the average value of primary commodity exports oil has the same effect as other commodities. Low levels of oil dependence are somewhat less risky than other commodities and high levels of dependence are somewhat more risky. The disaggregation slightly reduces the sample size, does not change the significance of any of the other variables, and substantially improves the overall fit of the model.[24]

Recall that the other proxies for financial opportunities, the Cold War and diasporas, are not included in this baseline. The end of the Cold War does not

[22] We tried different specifications to test for the effect of political repression by investigating non-linear effects, by including the autocracy score instead of the democracy score, and by using the difference between the two variables as suggested by Londregan and Poole (1996). We also tried the Freedom House measure of political freedom, but neither of these alternative political repression measures were found to be significant.

[23] We would like to thank Jan Dehn for providing us with the data that enabled this disaggregation.

[24] Furthermore, using data from Dehn (2000) we investigated whether contemporaneous export price changes altered the risk of conflict. We could not find any evidence to support this hypothesis.

appear to have had a significant effect. Diasporas are excluded from the baseline purely for considerations of sample size. In the parsimonious variant in which they are included, their effect on the risk of repeat conflict is substantial: after five years of peace, switching the size of the diaspora from the smallest to the largest found in post-conflict episodes increases the risk of conflict six-fold.

The proxies for earnings foregone have substantial effects. If the enrollment rate for secondary schooling is ten percentage points higher than the average the risk of war is reduced by about three percentage points (a decline in the risk from 11.5% to 8.6%). An additional percentage point on the growth rate reduces the risk of war by about one percentage point (a decline from 11.5% to 10.4%). Our other proxy for the cost of rebellion is also highly significant and substantial. Directly after a civil war there is a high probability of a re-start, the risk being about 32%. This risk declines over time at around one percentage point per year.

The only measures of rebel military advantage that survive into the baseline are population dispersion and social fractionalization. Consistent with Herbst's hypothesis, countries with a highly concentrated population have a very low risk of conflict, whereas those with a highly dispersed population have a very high risk (about 37%). Consistent with the hypothesis that cohesion is important for rebel effectiveness, social fractionalization makes a society substantially safer: a maximally fractionalized society has a conflict risk only one quarter that of a homogenous society.

Only one of the proxies for grievance survives into the baseline regression, namely ethnic dominance. If a country is characterized by ethnic dominance its risk of conflict is nearly doubled. Thus, the net effect of increased social diversity is the sum of its effect on social fractionalization and its effect on ethnic dominance. Starting from homogeneity, as diversity increases the society is likely to become characterized by ethnic dominance, although this will be reversed by further increases in diversity. The risk of conflict would first rise and then fall. Note that while these measures in combination are superficially similar to the hypothesized effect of polarization, our measure of polarization itself is insignificant.

Finally, the coefficient on the scale variable, population, is highly significant and close to unity: risk is approximately proportional to size. We have suggested that proportionality is more likely if conflict is generated by opportunities than by grievances.

4. Robustness checks

We now test these baseline results for robustness. We consider the sensitivity both to data and to method. With respect to data, we investigate the effect of outlying observations, and of different definitions of the dependent and independent variables. With respect to method, we investigate random effects, fixed effects and rare events bias.

We investigate outlying observations using two different methods. First, we inspect the characteristics of the 46 conflict episodes used in the baseline regression

and second, we use a systematic analysis of influential data points. Since our sample is unbalanced as between events and non-events, the potential problems of outliers arises predominantly among the 46 conflict episodes. Of these conflict episodes, 24 were first-time conflicts and 22 were repeat conflicts.

First, the classification of events in Romania in 1989, and in Iran in 1974, 1978, and 1981 as civil wars is in various respects questionable. They are, on our analysis highly atypical of conflict episodes. Both had secondary school enrolments much higher than the other conflict episodes, and Iran also had an atypically high primary commodity export share. In Table 6 column 1 we drop these doubtful observations. No results are overturned, but the performance of the regression improves and all variables are now significant at the 1% or 5% level.

There are four observations of highly negative growth: Angola in 1970–74, Zaïre (now the Democratic Republic of the Congo) in 1990–95, Iran in 1975–79 and Iraq in 1980–84. All of these growth collapses appear to be genuine, and they occur in different countries. We now check whether the result that the growth rate affects conflict risk is dependent upon these four observations, deleting them along with Iran and Romania (Table 6, column 2). Growth remains significant, and its coefficient is only slightly reduced. Hence, we can conclude that the increased risk of conflict due to slow growth is not confined to episodes of growth collapse, but is a more continuous relationship.

We next analyse whether our regression results are sensitive to the inclusion of influential data points. Based on the methods developed by Pregibon (1981)[25] we examined which observations may be influential and investigated whether omitting these observations from our baseline model changed our results. We find three influential observations: Congo 1995–99, Iran 1970–74, and Romania 1985–89. However, when we omitted these three observations from our regression, the overall fit of the regressions improved (from previously $R^2 = 0.24$ to $R^2 = 0.29$) and all of the coefficients remain statistically significant. (Table 6, column 3).

We now investigate the possibility that a few countries with a high commodity export ratio account for the non-monotonic relationship to conflict risk. This might imply that the reduction in conflict risk only occurred at extreme values of commodity dependence. Four peaceful countries have particularly high values of primary dependence: Saudi Arabia, Guyana, Oman, and Trinidad and Tobago. In Table 6 column 4 we present our baseline model excluding these four high primary commodity exporters. The non-monotonic relationship between primary commodity exports and the risk of conflict remains significant, as do all other results.

We next turn to questions of variable definition. The most contentious aspect of the dependent variable is distinguishing between whether a country has a single long war or multiple shorter wars interrupted by periods of peace. In the above analysis we have been guided by the judgement of the political scientists who built

[25] Long (1997) pp.98–101 provides a discussion of influence in limited dependent variable models.

the original data sets. Some peace periods are, however, quite short and it might be better to conceptualize these as interludes in a single war. We first reclassified all those wars that were separated by peace periods of less than one month as continuous wars (Table 6, column 5). The baseline results are not altered by this redefinition. We then reclassified those wars separated by less than a year as continuous wars (Table 6, column 6). The only result to be affected is that the growth rate becomes marginally insignificant ($p = 0.12$), although its coefficient is little changed.[26]

We investigated how robust our results are to the definition of ethnic dominance and social fractionalization. In the baseline we define ethnic dominance as the largest ethnic group constituting 45–90% of the population. We investigate other definitions that either vary the range of the population or use the share of the largest group regardless of its size. As the range is changed from 45–90% the significance level and the coefficient are both reduced, while if the definition is changed more radically to being the population share of the largest group it is completely insignificant. We also find that 'social fractionalization', our measure of cross-cutting cleavages, dominates the other possible aggregation procedures for ethnic and religious diversity. When this measure of fractionalization is included with the ethnic and religious diversity indices either together or individually, it is significant whereas the underlying indices are not significant.

In the baseline we use only the most extreme measure of polarization over the range proposed by Esteban and Ray (1994). However, if this measure is replaced by either the lower bound ($\alpha = 0$), or the central measure ($\alpha = 0.8$) the results are unaffected: polarization remains insignificant and the other variables remain significant. We also experimented with the alternative measure proposed by Reynal-Querol (2002), and with the number of ethnic groups, but with the same result.[27]

In Table 7 we investigate a number of different estimation issues. We concentrate on the analysis of random effects, fixed effects, time effects, and a correction for rare events. We re-estimated our models using random effects. For the baseline model we find that the panel data estimator is not different from the pooled estimator, i.e. we accept the hypothesis that we can pool across the observations.[28] The estimation of fixed effects logits was only possible on a very small sub-sample of the observations. The countries for which the dependent variable does not vary over time (the majority of countries experienced only peace) cannot be included in

[26] We also examined the effect of time since the previous conflict in more detail by including the natural logarithm of the peace variable or its square, however, a linear decay term provides a better fit. Note that the measure of peace since the end of the civil war is somewhat imprecise since we only measure it from the end of the war to the initial year of each sub-period. A duration model of post-war peace would allow a more detailed analysis of this peace effect, however, the duration model results in Collier *et al.* (2004) support the results presented in this paper.

[27] All of these robustness checks are presented in Collier and Hoeffler (2002).

[28] A LRT provides a χ^2 statistic of 0 ($p = 0.998$). Thus, we cannot reject the null-hypothesis that the panel data and pooled estimator provide the same results.

Table 6 Robustness checks

	1 Excluding Iran and Romania	2 Excluding Iran and Romania and growth collapses	3 Excluding influential data points	4 Excluding high primary commodity exporters	5 Peace periods of shorter than one month are treated as continuous wars	6 Peace periods of shorter than one year are treated as continuous wars
Primary commodity exports/GDP	19.696	19.029	28.745	18.771	19.147	22.686
	(6.608)***	(6.671)***	(7.862)***	(6.063)***	(5.939)***	(6.718)***
(Prim. com. exports/GDP)^2	−34.090	−33.250	−59.818	−28.466	−30.150	−39.053
	(14.356)**	(14.609)**	(17.781)***	(12.299)**	(12.031)***	(14.405)***
Male secondary schooling	−0.035	−0.037	−0.041	−0.031	−0.031	−0.031
	(0.011)***	(0.011)***	(0.011)***	(0.010)***	(0.010)***	(0.010)***
(GDP growth) $t-1$	−0.140	−0.100	−0.137	−0.122	−0.102	−0.071
	(0.047)***	(0.052)**	(0.046)***	(0.044)***	(0.044)***	(0.047)
Peace duration	−0.004	−0.003	−0.004	−0.004	−0.003	−0.003
	(0.001)***	(0.001)***	(0.0011)	(0.001)***	(0.001)***	(0.001)***
Geographic dispersion	−2.114	−2.272	−2.890	−2.449	−2.541	−2.953
	(1.080)**	(1.090)**	(1.136)***	(1.008)**	(1.012)***	(1.049)***

Social fractionalization	−0.0002	−0.0002	−0.0003	−0.0002	−0.0002	−0.0002
	(0.0001)**	(0.0001)**	(0.0001)***	(0.0001)**	(0.0001)**	(0.0001)***
Ethnic dominance	0.727	0.732	0.655	0.647	0.732	0.741
	(0.368)**	(0.370)**	(0.372)*	(0.354)*	(0.357)**	(0.362)**
Ln population	0.747	0.743	0.899	0.772	0.782	0.832
	(0.174)***	(0.175)***	(0.195)***	(0.168)***	(0.167)***	(0.176)***
N	674	671	685	662	686	683
No. of wars	42	39	43	46	44	41
Pseudo R^2	0.25	0.22	0.29	0.21	0.23	0.21
Log likelihood	−118.40	−116.17	−114.04	−122.23	−126.33	−122.23

Notes: All regressions include a constant. Standard errors in parentheses. ***, **, * indicate significance at the 1, 5, and 10% level, respectively.
Column 2: We exclude the following three growth collapses: Angola 1970–74, Iraq 1980–84, and Zaire 1990–94.
Column 3: We exclude the following three influential data points: Iran 1970–74, Romania 1985–89, Congo 1995–99.
Column 4: We exclude the countries with the highest primary commodity export to GDP ratio, namely Saudi Arabia, Guyana, Oman, and Trinidad and Tobago. Their average primary commodity export to GDP ratio is 0.504 (sample average 0.158).
Column 5: We exclude the following war starts: Angola 1975 and Somalia 1988.
Column 6: We exclude the following war starts: Angola 1975, Mozambique 1976, Sierra Leone 1997, Somalia 1988 and Zaïre/Democratic Rep. of Congo 1997.

Table 7 Estimation issues

	1 Random effects	2 Fixed effects	3 Pooled logit plus time dummies	4 Rare events logit
Primary commodity exports/GDP	18.937 (5.865)***	35.850 (14.436)***	18.895 (5.988)***	17.161 (6.535)***
(Primary commodity exports/GDP)^2	−29.443 (11.782)***	−65.967 (26.964)***	−29.815 (12.098)***	−25.594 (14.355)*
Male secondary schooling	−0.032 (0.010)***	0.007 (0.033)	−0.031 (0.010)***	−0.029 (0.010)***
(GDP growth) *t*-1	−0.115 (0.043)***	−0.045 (0.072)	−0.129 (0.047)***	−0.110 (0.040)***
Peace duration	−0.004 (0.001)***	0.011 (0.002)***	−0.004 (0.001)***	−0.004 (0.001)***
Geographic dispersion	−2.487 (1.005)***	115.363 (74.562)	−2.447 (1.018)**	−2.394 (1.085)**
Social fractionalization	−0.0002 (0.0001)**	−0.007 (0.006)	−0.0002 (0.0001)**	−0.0002 (0.0001)**
Ethnic dominance (45–90%)	0.670 (0.354)*		0.682 (0.359)*	0.644 (0.336)*
Ln population	0.768 (0.166)***	0.010 (1.410)	0.762 (0.170)***	0.726 (0.151)***
*T*70-74			0.725 (0.602)	
*T*75-79			0.578 (0.608)	
*T*80-84			1.137 (0.602)*	
*T*85-89			−0.013 (0.757)	
*T*90-94			0.802 (0.677)	
*T*95-99			−0.492 (0.921)	
N	688	145	688	688
No of wars	46	44	46	46
Pseudo R^2			0.26	
Log likelihood	−128.21	−38.18	−124.30	

Notes: All regressions include a constant. Standard errors in parentheses. ***, **, * indicate significance at the 1, 5, and 10% level, respectively

the analysis. Although the fixed effects test is very severe, the non-monotonic effect of primary commodity exports remains significant. Were the effect of primary commodity exports dependent only upon cross-section data, it might suggest that the variable was proxying some other characteristic such as geography.

However, the fixed effects regression uses only changes in primary commodity dependence, and so reduces the scope for alternative interpretations.[29]

We analysed whether time effects matter by including time dummies in the model. Based on a log likelihood ratio test we cannot reject the hypothesis that the time dummies are zero.[30]

Finally, in the last column of Table 7 we use a recently developed correction method for rare events data (King and Zeng, 2001). The event we predict (war) occurs in only about 7% of our observations. King and Zeng show that standard logit estimation tends to underestimate the probability of rare events. We therefore used their correction procedure. The differences between the standard logit results and the rare events corrected results are negligible with all variables significant at the same levels. The mean of the predicted probabilities obtained from the rare events logit regression is 0.072. Thus, we find that the corrected results are very similar to the logit results.

We examined a number of different model specifications. We found that none of the following geographic and demographic characteristics were significant: forest coverage, population density and the proportion of young men aged 15 to 29.[31] We also investigated the potential endogeneity of income to civil war. Evidently since we are measuring income prior to war the endogeneity only arises if a country has more than one war. Since the first war will have reduced income, for subsequent wars the correlation between income and war could in principle reflect this reverse causation. To control for this we re-estimated excluding repeat wars. The income variable remained highly significant.

5. Interpretation and conclusion

Using a comprehensive data set of civil wars over the period 1960–99 we used logit regressions to predict the risk of the outbreak of war in each five-year episode. We find that a model that focuses on the opportunities for rebellion performs well,

[29] We also investigated the effect of commodity prices. Since prices are exogenous, they can be entered contemporaneous with the episode being predicted, whereas our value-based proxy has to be lagged. We experimented with both the level of export prices and with the change in prices from the previous period. However, in either form when added to the baseline regression the variable was insignificant. The fact that lagged values of exports are significant even in the fixed effects regression suggests that rebels do respond to changes in values, but the response is evidently not so rapid as to give rise to an in-period price response. Potentially, the effect on conflict risk captures the 'voracity effect' predicted by Lane and Tornell (1999) whereby an increase in the price of a natural resource export would induce more than the increment in value to be devoted to conflict. Our results suggest that there may be such an effect but that it is lagged.

[30] The LRT statistic is 7.83, 6 restrictions ($p = 0.251$).

[31] The proportion of the population living in urban areas was statistically significant when we excluded the geographic concentration of the population. However, when we included both proxies for the concentration of the population, the geographic concentration measure remained statistically significant while the proportion of the population living in urban areas was marginally insignificant ($p = 0.11$).

whereas objective indicators of grievance add little explanatory power. The model is robust to a range of tests for outliers, redefinitions, and alternative specifications.

One factor influencing the opportunity for rebellion is the availability of finance. We have shown that primary commodity exports substantially increase conflict risk. We have interpreted this as being due to the opportunities such commodities provide for extortion, making rebellion feasible and perhaps even attractive. An alternative explanation would be that primary commodity dependence worsens governance and so generates stronger grievances. However, we are controlling for economic performance—the level, growth, and distribution of income—and for political rights (which appear not to affect the risk of conflict). While we would not wish to discount the possibility of an effect working through corruption (for which we cannot control), there is plenty of case study evidence supporting the extortion interpretation. Another source of finance for which there is good case study evidence is diasporas. We have found that diasporas substantially increase the risk of conflict renewal, and it is hard to find an alternative explanation for this result.

A second factor influencing opportunity is the cost of rebellion. Male secondary education enrollment, *per capita* income, and the growth rate all have statistically significant and substantial effects that reduce conflict risk. We have interpreted them as proxying earnings foregone in rebellion: low foregone earnings facilitate conflict. Even if this is correct, low earnings might matter because they are a source of grievance rather than because they make rebellion cheap. However, if rebellion were a protest against low income, we might expect inequality to have strong effects, which we do not find.

A third aspect of opportunity is military advantage. We have found that a dispersed population increases the risk of conflict, and there is weaker evidence that mountainous terrain might also advantage rebels. It remains possible that these are correlated with unmeasured grievances.

Most proxies for grievance were insignificant: inequality, political rights, ethnic polarization, and religious fractionalization. Only 'ethnic dominance'—one ethnic group being a majority—had adverse effects. Even this has to be considered in combination with the benign effects of social fractionalization: societies characterized by ethnic and religious diversity are safer than homogenous societies as long as they avoid dominance. We have suggested that diversity makes rebellion harder because it makes rebel cohesion more costly. It would be difficult to argue that diversity reduced grievance.

Finally, the risk of conflict is proportional to a country's population. We have suggested that both opportunities and grievances increase with population. Thus, the result is compatible with both the opportunity and grievance accounts. However, grievances increase with population due to rising heterogeneity. Yet those aspects of heterogeneity that we are able to measure are not associated with an increased risk of conflict. Hence, a grievance account of the effect of population would need to explain why unobserved, but not observed, heterogeneity increases conflict risk.

One variable, the time since a previous conflict, has substantial effects: time heals. Potentially, this can be interpreted either as opportunity or grievance. It may reflect the gradual depreciation of rebellion-specific capital, and hence an increasing cost of rebellion, or the gradual erosion of hatred. However, we have found that a large diaspora slows the 'healing' process. The known proclivity of diasporas to finance rebel groups offsets the depreciation of rebellion-specific capital, and so would be predicted to delay 'healing'. The diaspora effect thus lends support to the opportunity interpretation.

Opportunity as an explanation of conflict risk is consistent with the economic interpretation of rebellion as greed-motivated. However, it is also consistent with grievance motivation as long as perceived grievances are sufficiently widespread to be common across societies and time. Opportunity can account for the existence of either for-profit, or not-for-profit, rebel organizations. Our evidence does not therefore imply that rebels are necessarily criminals. But the grievances that motivate rebels may be substantially disconnected from the large social concerns of inequality, political rights, and ethnic or religious identity.

Acknowledgements

We would like to thank Todd Sandler and two anonymous referees for very helpful comments. Anke Hoeffler gratefully acknowledges funding from the Research Council of Norway (NFR).

References

Alesina, A., Baqir, R., and Easterly, W. (1999). 'Public goods and ethnic divisions', *Quarterly Journal of Economics*, **114**, 1243–84.

Alesina, A., Oetzler, S., Roubini, N., and Swagel, P. (1996). 'Political instability and economic growth', *Journal of Economic Growth*, **1**, 189–211.

Alesina, A. and Spolaore, E. (1997). 'On the number and the size of nations', *Quarterly Journal of Economics*, **112**, 1027–56.

Angoustures, A. and Pascal, V. (1996). 'Diasporas et financement des conflits', in F. Jean and J.-C. Rufin (eds), *Economie des Guerres Civiles*, Hachette, Paris.

Barro, R.J. (1991). 'Economic growth in a cross section of countries', *The Quarterly Journal of Economics*, **106**, 407–43.

Barro, R.J. (ed.) (1997). *Determinants of Economic Growth*, MIT Press, Cambridge, MA.

Barrett, D.B. (ed.) (1982). *World Christian Encyclopedia*, Oxford University Press, Oxford.

Buchanan, J.M. and Faith, R.L. (1987). 'Secession and the limits of taxation: towards a theory of internal exit', *American Economic Review*, **77**, 1023–31.

Collier, P. (2000). 'Rebellion as a quasi-criminal activity', *Journal of Conflict Resolution*, **44**, 839–53.

Collier, P. (2001). 'Ethnic diversity: an economic analysis of its implications', *Economic Policy*, **32**, 129–66.

Collier, P. and Hoeffler, A. (1998). 'On the economic causes of civil war', *Oxford Economic Papers*, **50**, 563–73.

Collier, P. and Hoeffler, A. (2002). 'Greed and grievance in civil war', CSAE Working Paper, WPS 2002–01, Oxford University.
http://www.economics.ox.ac.uk/CSAEadmin/workingpapers/pdfs/2002–01text.pdf.

Collier, P., Hoeffler, A., and Söderbom, M. (2004). 'On the duration of civil war.' *Journal of Peace Research*, **41**, 253–73.

Davidson R. and MacKinnon, J.G. (1981). 'Several tests for model specification in the presence of alternative hypotheses', *Econometrica*, **49**, 781–93.

Dehn, J. (2000). 'Commodity price uncertainty, investment and shocks: implications for economic growth', D.Phil. thesis, University of Oxford.

Deininger, K. and Squire, L. (1996). 'A new data set measuring income inequality', *World Bank Economic Review*, **10**, 565–91.

Deininger, K. and Squire, L. (1998). 'New ways of looking at old issues: inequality and growth', *Journal of Development Economics*, **57**, 249–87.

Easterly, W. and Levine, R. (1997). 'Africa's growth tragedy: policies and ethnic divisions.' *Quarterly Journal of Economics*, **113**, 203–49.

Esteban, J.-M. and Ray, D. (1994). 'On the measurement of polarization.' *Econometrica*, **62**, 819–51.

Esteban, J.-M. and Ray, D. (1999). 'Conflict and distribution', *Journal of Economic Theory*, **87**, 379–415.

Figes, O. (1996). *A People's Tragedy: The Russian Revolution 1891–1924*, Pimlico, London.

Fearon, J. (2004) 'Why do some civil wars last so much longer than others?' *Journal of Peace Research*, **41**, 275–301.

Gerrard, A.J.W. (2000). 'What is a mountain?', mimeo, DECRG, World Bank, Washington, DC.

Grossman, H.I. (1991). 'A general equilibrium model of insurrections', *American Economic Review*, **81**, 912–21.

Grossman, H.I. (1999). 'Kleptocracy and revolutions', *Oxford Economic Papers*, **51**, 267–83.

Hegre, H., Ellingsen, T., Gates, S., and Gleditsch, N.-P. (2001). 'Toward a democratic civil peace? Democracy, political change, and civil war, 1816–1992', *American Political Science Review*, **95**, 33–48.

Herbst, J. (2000). *States and Power in Africa*, Princeton University Press, Princeton, NJ.

Hirshleifer, J. (1995). 'Theorizing about conflict', in K. Hartley and T. Sandler (eds), *Handbook of Defense Economics*, Vol. 1, Elsevier Science, Amsterdam, 165–89.

Hirshleifer, J. (2001). *The Dark Side of the Force: Economic Foundations of Conflict Theory*, Cambridge University Press, Cambridge.

Jaggers, K. and Gurr, T.R. (1995). 'Tracking democracy's third wave with the Polity III data', *Journal of Peace Research*, **32**, 469–82.

King, G. and Zeng, L. (2001). 'Logistic regression in rare events data', *Political Analysis*, **9**, 137–63.

Klare, M.T. (2001). *Natural Resource Wars: The New Landscape of Global Conflict*, Metropolitan Books, New York.

Lane A. and Tornell, P.R. (1999). 'The voracity effect', *American Economic Review*, **89**, 22–46.

Londregan, J.B. and Poole, K.T. (1996) 'Does high income promote democracy?', *World Politics*, **49**, 1–30.

Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables*, Sage Publications, London.

Mauro, P. (1995). 'Corruption and growth', *The Quarterly Journal of Economics*, **110**, 681–712.

Pregibon, D. (1981). 'Logistic regression diagnostics', *The Annals of Statistics*, **9**, 705–24.

Reynal-Querol, M. (2000). 'Religious conflict and growth: theory and evidence'. Ph.D. thesis, London School of Economics and Political Science.

Reynal-Querol, M. (2002). 'Ethnicity, political systems and civil war', *Journal of Conflict Resolution*, **46**, 29–54.

Sachs, J. and Warner, A.M. (2000). 'Natural resource abundance and economic growth', in G.M. Meier and J.E. Rauch (eds), *Leading Issues in Economic Development*, 7th ed., Oxford University Press, Oxford.

Sambanis, N. (2002). 'What is a civil war? Conceptual and empirical complexities of an operational definition', mimeo, Yale University.

Sen, A. (1973). *On Economic Inequality*, Clarendon Press, Oxford.

Singer, D.J. and Small, M. (1994). 'Correlates of war project: international and civil war data, 1816–1992', data file, Inter-University Consortium for Political and Social Research, Ann Arbor, MI.

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

The Stockholm International Peace Research Institute (2002). *Yearbook of World Armaments and Disarmaments*, Oxford University Press, Oxford.

Summers, R. and Heston, A. (1991). 'The Penn World Table (Mark 5): an expanded set of international comparisons, 1950–1988', *The Quarterly Journal of Economics*, **99**, 327–68.

USSR (1964). *Atlas Narodov Mira*, Department of Geodesy and Cartography of the State Geological Committee of the USSR. Moscow.

World Bank (2000). *World Development Indicators*, Washington DC, data file.

Appendix 1

1. A simple migration model

Our estimation of migration is based on the following model

$$\underset{}{dias_{it}} = \underset{(0.045)^{***}}{1.163 \cdot dias_{it}} - \underset{(0.001)^{**}}{0.0002 \cdot \ln GDP_{i,t-1}} + \underset{(0.03)}{0.003 \cdot war_{i,t-1}} + \underset{(0.002)}{0.003 \cdot T_{80}} + \underset{(0.002)}{0.005 \cdot T_{90+}} \underset{(0.008)}{0.013}$$

Where *dias* denotes diaspora which is measured as the ratio of emigrants in the USA to the total population of the country of origin. The variable *war* is a war dummy, measured at $t-1$ it takes a value of one if the country experienced a civil war in the previous period. The method of estimation is OLS. The data is measured at the beginning of each decade,

i.e. 1960, 1970, 1980, and 1990. The regression includes time dummies, T, which are jointly significant.

Based on this simple migration model we estimated the size of the diaspora at time t.

$$\hat{diaspora}_{it} = x_{it} \cdot \hat{\beta}$$

For countries which experienced a previous civil war we used these estimated values to correct for a possible endogeneity problem. We replaced a total of 64 observations. For countries which did not experience a civil war we use the actual diaspora data. In order to obtain values for 1965 we took the averages of this corrected diaspora data measured in 1960 and 1970, and analogously for the values for 1975 and 1985. For 1995 we use the observations measured in 1990.

Appendix 2

1. Calculating the marginal probabilities

In our regressions we estimate the probability of a war breaking out during a five-year period, and the model can be written in the following general form

$$Y_{it} = a + bX_{it} + cM_{i,t-1} + dZ_i + u_{it} \qquad (A2.1)$$

where t and i are time and country indicators. The dependent variable is a dummy variable indicating whether a war broke out during the five-year period, so that Y_{it} is the log odds of war. The explanatory variables are either measured at the beginning of the period (for example, income *per capita*, primary commodity exports/GDP, population), or during the previous five-year period (for instance, *per capita* income growth, or are time invariant or changing slowly over time (for example, social fractionalization).

The expected probability $\hat{p}_{it}$ of a war breaking out can be calculated by using the estimated coefficients obtained from equation (A1.1):

$$\hat{a} + \hat{b}X_{it} + \hat{c}M_{i,t-1} + \hat{d}Z_i = \hat{W}_{it} \qquad (A2.2)$$

$$\hat{p}_{it} = \frac{e^{\hat{W}_{it}}}{(1 + e^{\hat{W}_{it}})} \cdot 100 \qquad (A2.3)$$

Appendix 3

1. Data sources

Democracy The degree of openness of democratic institutions is measured on a scale of zero (low) to ten (high). Source: http://www.cidcm.umd.edu/polity/index.html. The data are described in Jaggers and Gurr (1995).

Table A2 Marginal probabilities

Variable	Coefficient	Mean of X	(1) At the mean	(2) Worst	(3) Best	(4) primary commodity/ GDP = 0.33	(5) 10% extra men in school	(6) 1% extra growth	(7) Min. peace	(8) Max. fraction-alization	(9) Ethnic dominance
Primary commodity exports/GDP	18.937	0.158	2.992	6.060	0	6.060	2.992	2.992	2.992	2.992	2.992
(Primary commodity exports/GDP)^2	−29.443		−0.735	−3.015	0	−3.015	−0.735	−0.735	−0.735	−0.735	−0.735
Male secondary schooling	−0.032	44.489	−1.406	−0.032	−4.645	−1.406	−1.722	−1.406	−1.406	−1.406	−1.406
(GDP growth)$_{t-1}$	−0.115	1.618	−0.186	1.508	−1.660	−0.186	−0.186	−0.417	−0.186	−0.186	−0.186
Peace duration	−0.004	347.5	−1.286	−0.004	−2.19	−1.286	−1.286	−1.286	−0.004	−1.286	−1.286
Geographic dispersion	−2.487	0.602	−1.497	0.000	−2.415	−1.497	−1.497	−1.497	−1.497	−1.497	−1.497
Social fractionalization	−0.0002	1790	−0.376	−0.004	−1.465	−0.376	−0.376	−0.376	−0.376	−1.465	−0.376
Ethnic dominance (45–90%)	0.670	0.439	0.294	0.670	0	0.294	0.294	0.294	0.294	0.294	0.670
Ln population	0.768	30,500,000	13.230	16.049	9.136	13.230	13.230	13.230	13.230	13.230	13.230
Constant	−13.073		−13.073	−13.073	−13.073	−13.073	−13.073	−13.073	−13.073	−13.073	−13.073
$\hat{W}$			−2.043	8.160	−16.312	−1.255	−2.359	−2.273	−0.761	−3.132	−1.667
$\hat{p}$			0.115	1.000	0.000	0.222	0.086	0.093	0.318	0.042	0.159

Diaspora We used the data on the foreign born population from the U.S. Bureau of the Census and divided these numbers by the total population in the country of origin. http://www.census.gov/population/

Ethnic dominance Using the ethno-linguistic data from the original data source (USSR, 1964) we calculated an indicator of ethnic dominance. This variable takes the value of one if one single ethno-linguistic group makes up 45 to 90% of the total population and zero otherwise. We would like to thank Tomila Lankina for the translation of the original data source.

Forest coverage We used the FAO measure of the proportion of a country's terrain which is covered in woods and forest. Source: http://www.fao.org/forestry

GDP per capita We measure income as real PPP adjusted GDP *per capita*. The primary data set is the Penn World Tables 5.6 (Summers and Heston, 1991). Since the data are only available from 1960–92 we used the growth rates of real PPP adjusted GDP *per capita* data from the World Bank's World Development Indicators 1998 in order to obtain income data for the 1990s. These GDP per cpaita data were used to calculate the average annual growth rate over the previous five years.

Geographic dispersion of the population We constructed a dispersion index of the population on a country by country basis. Based on population data for 400 km^2 cells we generated a Gini coefficient of population dispersion for each country. A value of 0 indicates that the population is evenly distributed across the country and a value of 1 indicates that the total population is concentrated in one area. Data is available for 1990 and 1995. For years prior to 1990 we used the 1990 data. We would like to thank Uwe Deichman of the World Bank's Geographic Information System Unit for generating this data. He used the following data sources: Center for International Earth Science Information Network (CIESIN), Columbia University; International Food Policy Research Institute (IFPRI); and World Resources Institute (WRI). 2000. Gridded Population of the World (GPW), Version 2. Palisades, NY: IESIN, Columbia University. Available at http://sedac.ciesin.org/plue/gpw.

Inequality Inequality was either measured as income inquality (source: Deininger and Squire, 1996) or as inequality in land ownership (source: Deininger and Squire, unpublished). Both inequality measures are provided as a Gini coefficient.

Male secondary school enrolment rates We measure male secondary school enrolment rates as gross enrolment ratios, i.e. the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. Secondary education completes the provision of basic education that began at the primary level, and aims at laying the foundations for lifelong learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers. Source: World Bank Development Indicators, 1998.

Mountainous terrain The proportion of a country's terrain which is mountainous was measured by John Gerrard, a physical geographer specialized in mountainous terrain. His measure is based not just on altitude but takes into account plateaus and rugged uplands. The data are presented in Gerrard (2000).

Peace duration This variable measures the length of the peace period (in months) since the end of the previous civil war. For countries which never experienced a civil war we measure the peace period since the end of World War II.

Population Population measures the total population, the data source is the World Bank's World Development Indicators 1998.

Primary commodity exports/GDP The ratio of primary commodity exports to GDP proxies the abundance of natural resources. The data on primary commodity exports and GDP were obtained from the World Bank. Export and GDP data are measured in current US dollars.

Social, ethnolinguistic, and religious fractionalization We proxy social fractionalization in a combined measure of ethnic and religious fractionalization. Ethnic fractionalization is measured by the ethno-linguistic fractionalization index. It measures the probability that two randomly drawn individuals from a given country do not speak the same language. Data are only available for 1960. In the economics literature this measure was first used by Mauro (1995). Using data from Barrett (1982) on religious affiliations we constructed an analogous religious fractionalization index. Following Barro (1997) we aggregated the various religious affiliations into nine categories: Catholic, Protestant, Muslim, Jew, Hindu, Buddhist, Eastern Religions (other than Buddhist), Indigenous Religions, and no religious affiliation.

The fractionalization indices range from zero to 100. A value of zero indicates that the society is completely homogenous whereas a value of 100 would characterize a completely heterogeneous society.

We calculated our social fractionalization index as the product of the ethno-linguistic fractionalization and the religious fractionalization index plus the ethno-linguistic or the religious fractionalization index, whichever is the greater. By adding either index we avoid classifying a country as homogenous (a value of zero) if the country is ethnically homogenous but religiously divers, or vice versa.

War data A civil war is defined as an internal conflict in which at least 1,000 battle related deaths (civilian and military) occurred per year. We use mainly the data collected by Small and Singer (1992) and according to their definitions (Singer and Small, 1984) we updated their data set for 1992–99.