Greed and Grievance in Civil War

Paul Collier
The World Bank, CEPR and CSAE

Anke Hoeffler
CSAE, Oxford

1. Introduction[1]

Civil war is both a tragedy and an impediment to development, affecting most of the world's poorest countries. It is now far more common than international conflict: of the 27 major armed conflicts listed in the Stockholm International Peace Research Institute Yearbook for 1999 (SIPRI 1999), all but two were internal.

In Section 2 we compare two contrasting motivations for rebellion: greed and grievance. Most rebellions are ostensibly in pursuit of a cause, supported by a narrative of grievance. However, since grievance-assuagement through rebellion is a public good which a government will not supply, economists would predict that such rebellions would be rare. Empirically, many rebellions appear to be linked to the capture of resources: diamonds in Angola and Sierra Leone, drugs in Colombia, and timber in Cambodia. We set up a simple rational choice model of greed-rebellion and contrast its predictions with those of a simple grievance model. A second empirical regularity is that some countries are prone to repeat conflict. This may be either because their underlying characteristics make them highly conflict-prone, or because of a feedback effect whereby conflict generates grievance which in turn generates further conflict. We show why such a feedback effect might be present in greed-motivated rebellions as well as in grievance-rebellions. Finally, we consider an integrated model in which the motivation for rebellion is both greed and grievance.

In Section 3 we discuss the construction of a comprehensive data set of 161 countries for each of the eight, five-year periods between 1960 and 1999, giving a total of 1288 potential observations. For 73 of these observations the society was at peace at the start of the period but experienced civil war during it. In Section 4 we use logit regressions to explain these collapses into civil war in terms of characteristics at the start of the period. We perform non-nested tests on the greed and grievance models. Although the greed model substantially outperforms the grievance model, we cannot reject the hypothesis that the grievance model adds to explanatory power. An integrated model, which incorporates some features of grievance into the greed model, outperforms all other models. An economic calculus of the costs and opportunities for the control of primary commodity exports appears to be the main systematic initial impetus to rebellion, with an additional effect from fear of domination by an ethnic majority. After peace has been restored, the legacy of conflict-induced grievance enables rebel movements to restart conflict by drawing on the support of their diasporas. We show that the results are robust to the inclusion of a wide range of alternative variables, and to tests for random and fixed effects and to correction for rare events bias. The last section concludes.

2. Theories of Rebellion

[1] Previous versions of this paper have benefited from presentations at CERDI, NYU, LSE, Oxford, Princeton, the World Bank and a CEPR conference at Lisbon. We would like to thank participants for comments, and especially Todd Sandler for helpful written suggestions.

Civil war occurs as a result of rebellion. Hence, the phenomenon to be explained is the emergence of a rebel organization. A rebellion is somewhat analogous to three other types of organization: protest movements, armies, and organized crime. In popular and political analysis, the most common analogue to rebellion is protest. The theory of protest movements, (Kuran, 1989), focuses on the coordination problem. In effect, by joining each other on the streets, people can create an election. However, the analogy is not very close. Protest is not a sustained economic activity: the participants are not full-time employees of the protest organization, and they risk little. Indeed, this is the essence of Kuran's `prairie fire' model: protest only takes off if the risks fall fast enough (through increasing participation) to make it safe. By contrast, participants in a rebellion must be prepared to fight over a prolonged period against organized force which aims to kill them. In this, a rebel organization is more analogous to a regular army than to a protest movement. Like an army it must solve immense problems of hierarchy and cohesion in order to get people to risk their own death in order to further the military objective of the group. However, unlike both an army and a protest organization, a rebellion must generate revenue in order to feed and pay its workers. The payroll of an army is financed out of taxation which the army itself does not raise. A protest movement does not have a significant payroll. The rebel organization must generate income despite not being directly productive, and in this respect rebellion is like crime. Indeed the analogy with crime is standard in the present economic theory of rebellion For example, Grossman (1999) states `in such insurrections the insurgents are indistinguishable from bandits or pirates' (p.269). However, the very scale of rebellion makes it distinctive from other crime. Rebel organizations usually have between 500 and 5,000 workers, whereas most criminals are self-employed or work in small groups. Hence, within crime, the closest analogy is that with *organized* crime. Even organized crime is usually a rather small scale activity, but the largest groups are approximately comparable in size to the smaller rebel organizations. The recent theory of organized crime (e.g. Konrad and Skaperdas, 1998), explains its larger scale than other crime by scale economies in extortion, eventually counterbalanced by policing.

The analogies with the economic theories of organized crime and protest movements form the basis for the `greed-rebellion' theory of Section 2.1. In this model rebellion is a distinctive type of organized crime, although subject to constraints similar to those facing an incipient protest movement. In Section 2.2 we present a contrasting political model of rebellion in which we put aside the economic considerations and focus instead upon various forms of grievance. Section 2.3 compares the two models and considers various ways in which rebellion might combine both greed and grievance.

2.1. A Model of Greed-Rebellion

Consider a rebellion which, like organized crime, generates its income from extortion. The rebels menace legitimate economic activities and exact a tribute. Empirically, rebellion is nevertheless distinctive from other types of organized crime, if only in terms of scale and casualties. We are concerned only with those rebellions which are sufficiently large and

sustained, and come into sufficient conflict with government forces, to generate at least 1,000 battle-related deaths. This is the conventional empirical definition of civil war, and this is the phenomenon which we will be seeking to explain. Hence, rebellion is distinctive from other crime in the scale of organized violence. We suggest that rebellion is distinctive because the object of extortion is different. The typical object of criminal extortion is a street of shopkeepers. The criminal organization is small because the scale economies in menacing shopkeepers are modest, and because the rate of predation cannot get very high before businesses relocate. Rebellion is at the apex of organized crime because the object of extortion is not a street of shopkeepers but the export of primary produce. As a result, both the scale economies of menace and the sustainable rate of extortion, are atypically high. Primary commodity exports are sustainable targets for predation because their production is intensive in irreversible and immobile assets, and because produce must be transported to a port. The owners of irreversible and immobile assets such as land, trees and mines, receive rents which can be expropriated without curtailing the activity, whereas similar predation of the incomes generated by mobile factors would simply produce relocation. As primary commodities are transported to a port they are exposed to predation at many geographic `choke points'. The government itself typically levies taxes at the tightest choke point, namely the port, but rebels may attack at any point along the transport route.

We begin with a model which, for a country of given population, analyzes how the risk of such a predatory rebellion is affected by variations in the level of income and its structure. Corresponding to our subsequent empirical analysis, we will be seeking to explain only the initiation of conflict, rather than its duration.

We first specify a function for the revenue which a rebellion can generate from predation upon primary commodity exports, while being opposed by government forces which are protecting those exports. The gross revenue function has two components, the tribute which can be exacted conditional upon a successful threat of military force, and a military contest function. The tribute is assumed to be a function of the value of primary commodity exports: the tribute increases in this base for predation, but at a diminishing rate. It is convenient to specify primary commodity exports as the product of income, y, and the share of income constituted by such exports, n. For the military contest function, we follow the specification of Konrad and Skaperdas (1998), in which military outcomes reflect the balance of opposing forces, rebel, r, and government, g. Hence:

$$R_r = (ny)^a [r/(r+g)]. \qquad (1)$$

We next introduce an insight from the economic theory of protest movements (Kuran, 1989). Government forces are not simply trying to impede rebels from predation, they are trying to kill them. If the rebel force is too small, when it attacks the choke points, the government forces which it encounters will turn from defense to attack. In Kuran's model, there is a tipping point of participation above which the protest becomes viable. Here, we assume that there is a threshold size of rebel forces relative to government forces, below which predation of primary commodity

exports is too dangerous to be viable. We introduce this as a survival constraint: rebel forces must exceed some fraction of government forces in order to engage in predation without suffering punishment. Thus,

$$r \geq bg. \qquad (2)$$

The rebel cost function, C_r, is its wage bill:

$$C_r = rw_r. \qquad (3)$$

Where w_r = rebel wage rate.

The rebel wage is assumed to be linear in the level of income, y:

$$w_r = jy. \qquad (4)$$

The rebel leader thus chooses the size of the rebel laborforce so as to maximize net revenue, (1)-(3), subject to (2). Rebellion will occur when net revenue is non-negative, which we refer to as the financial viability constraint:

$$R_r - C_r \geq 0. \qquad (5)$$

We next endogenize the initial government defense effort, g, which precedes the initiation of rebellion. The government cost function, C_g, mirrors that of the rebels:

$$C_g = gw_g \qquad (6)$$

$$w_g = ly. \qquad (7)$$

We assume that the government sets its precautionary military expenditure as fraction of its revenue, R_g. Reflecting observed fiscal patterns, government revenue is assumed to be elastic in income, and in the share of income from primary commodity exports. As discussed, the same factors make natural resource exports atypically taxable and lootable.[2]

$$R_g = dy^q + gny. \qquad q>1 \qquad (8)$$

Prior to rebellion, the government devotes a given share of its expenditure, n, to military expenditure. Hence, prior to rebellion, government forces, g, are:

[2] Government military expenditure may also be increasing in the proportion of natural resource exports because the latter are recognized as being vulnerable to predation. This would simply reinforce the results below on the non-monotonic effect of natural resource exports on the risk of conflict.

$$g = n(dy^{q-1} + gn)/l. \quad (9)$$

The minimum size of a viable rebellion is that at which the survival constraint is binding. If the rebellion is financially viable at this size, the rebel leader may choose to expand it further. However, since there are diminishing marginal returns to rebel labor, the rebellion will only be worth expanding beyond the size imposed by the survival constraint if it is also financially viable at that size. Hence, financial viability at the size which just satisfies the survival constraint is the condition for the initiation of a rebellion.

Substituting, and rearranging, the financial viability condition is:

$$l n^{a}/(1+b) \geq n(d j y^{q-a} + n j g y^{1-a}) \quad (10)$$

Equation (10) is thus the key condition for the initiation of greed-rebellion.

Now consider whether the rebellions so initiated are likely to become sufficiently large to be recorded as civil wars: that is, whether they generate at least 1,000 combat-related deaths. For this a key consideration is whether the incipient rebellion generates a phase of `arms race' in which both government and rebel forces grow in response to each other. If there is such a phase, then we will assume that the rebellion is large enough to generate a civil war.

The sequence of rebellion is as follows. In the pre-history of the rebel movement it builds its forces until they reach the level which just satisfies the survival condition. At this stage rebel forces can start operations, and so combat deaths commence. We assume that the initial response of the government to rebellion is to increase its expenditure upon its military forces and focus on how rebel forces respond to this increase. If rebel forces respond by contracting, then it is possible that the rebellion gets snuffed out before combat-related mortalities reach the critical level of 1,000 at which the conflict is classified as a civil war. By contrast, if the rebel group initially responds by increasing its own forces, then there is a phase of mutual escalation, although this phase may end with the bankruptcy of the rebellion. The Konrad-Skaperdas (KS) military combat function predicts how the optimal size of the rebel organization changes in response to an increase in government forces: rebel forces will increase if $r>g$, and decrease if $r<g$. However, at the point at which the rebel organization commences predation its forces are below their optimal size: the rebel organization builds its forces from zero, and once they satisfy the survival constraint, operations commence. At this point the government expands its own forces. Hence, regardless of whether the *optimal* size of the rebel organization increases or reduces in response to the expansion of government forces, the actual size must increase in line with the rise in the minimum size necessary to satisfy the survival constraint. The rebel organization expands its forces at least in reaction to this increase and perhaps additionally in a continuing adjustment towards the optimum. In response, the government further expands its own forces.[3] Hence, once (10) is satisfied, not only does the rebellion come into existence, but

3 Collier (2000) further analyzes this escalation.

it enters an arms race phase with government forces. While this race may end with the bankruptcy of the rebellion, we assume that in the interim it will have caused sufficient combat deaths to meet the criterion used in our empirical analysis. Condition (10) is thus both a necessary condition for rebellion and a sufficient condition for civil war. We now derive from it specific predictions as to how the risk of rebellion will be affected by the level of income and its structure.

First, consider the level of income. The LHS of (10) is not a function of y. Since $q>1>a$, the derivative of the RHS w.r.t. y is strictly positive:

$$dRHS/dy = (q-a)ndjy^{q-a-1} + (1-a)nnjgy^{-a} > 0. \quad (11)$$

Hence, the conditions under which rebellion is profitable become more restrictive the higher is income. Restated, higher income is predicted to reduce the risk of conflict.

Next, consider the structure of income. Differentiating (10) w.r.t. n yields:

$$d(10)/dn = [a\,l n^{a-1}/(1+b)] - njgy^{1-a}. \quad (12)$$

In the neighborhood of $n=0$, (10) is negative and (12) is positive, so that as n increases from zero the risk of conflict increases. However, (12) is not monotonic in n. Setting (12) to zero and solving for n yields a critical value, n^*, below which the conditions for rebellion become easier as n increases, and above which they become tighter. Hence, n^* denotes that intermediate level of primary commodity dependence at which the risk of rebellion is at its peak:

$$n^* = \{[(1+b)njgy^{1-a}]/a\,l\}^{1-a}. \quad (13)$$

The above analysis applies to a country of given population. Now consider how the risk will be affected by variations in population size. If the risk in each of two identical territories is p, the risk that there will be at least one conflict in the two territories is $p + p(1-p)$. The elasticity of risk with respect to population, controlling for n and y, will therefore be $(1-p)$. Globally, over the period 1960-1999 the mean risk of conflict per period was around 0.06, so that the expected elasticity of conflict with respect to population is around 0.94.

Thus far, the model has four testable predictions. The risk of conflict should be decreasing in per capita income, increasing in the share of natural resource exports in GDP at low levels, decreasing at high levels, and be slightly less than unit elastic in population. We now introduce three refinements which introduce further testable propositions: relative military advantage, rebel costs of recruitment, and start-up finance.

Relative military advantage

The military combat function included in (1) is a convenient simple special case of the KS function. The more general form of the KS combat function is:

$$r/(r+g+k), \qquad (14)$$

where k denotes the relative military advantage of the two forces. For $k=0$, the special case we have adopted above, neither force has technological superiority. For $k>0$ government forces have the advantage, and conversely for $k<0$. Relative military advantage affects the necessary conditions for financial viability. Trivially, the greater is the relative military advantage of the government, (the higher is the value of k), the lower is the value of the combat function, (14), and hence the lower is gross and net revenue from rebellion. Rebellion is less likely the greater is the relative military advantage of the government. We endogenize relative military advantage with respect to geography, cohesion and motivation.

Rebels need a refuge, and two geographic features, mountains and forests, are commonly supposed to make counter-insurgency more difficult. We measure these as the proportion of a country's terrain which is mountainous, m, and the proportion which is forested, L. Additionally, Herbst (2000) has suggested that states such as Zaire are prone to rebellion because the population is dispersed around the edges of the country, making government military control more difficult. We measure this by the geographic dispersion of the population, Y. We predict that the risk of conflict is increasing in these three measures.

In war, the size of a fighting force is probably less important than its willingness to fight. Military history offers many instances in which cohesion and motivation offset numerical inferiority. The government army has the advantage that it predates the rebellion and so has had time to build cohesion and motivation. The rebel organization must create these attributes as it recruits. The need for cohesion places a simple constraint upon the composition of rebel recruits: the organization cannot afford diversity. If recruitment spans ethnic and religious divides it will be more difficult to forge the resulting laborforce into a cohesive fighting force. The strategy of homogenous recruitment will sometimes come at a cost, since in socially fractionalized societies it will reduce the size of the recruitment pool and so raise labor costs for any given size of recruitment. Thus, the more fractionalized the society, the more costly is rebel recruitment (or the less cohesive is the rebel force). We thus predict that social fractionalization should reduce the incidence of civil war. We measure social fractionalization, f, by the extent to which a society is divided by ethnicity and religion.

In addition to cohesion, an army needs motivation: rebel forces must be persuaded to want to kill the enemy. The Leninist theory of rebel organization, which many rebel groups follow even if they are not Marxist, is that initially the population from which forces are recruited does not *realize* that it is oppressed, so that an awareness of grievance has to be built by the rebellion. Many rebel organizations invest substantial resources in this process of indoctrination. For example, the highly successful Eritrean People's Liberation Front routinely withdrew troops from the front line for periods of six months for ideological training. Thus, subjective grievance

is consciously generated by conflict to enhance efficiency, rather than being either an accidental by-product of conflict, or a prior cause of conflict. The greed model departs from Leninist theory in assuming that the ability of a rebel organization to inculcate subjective grievance is unrelated to objective grounds for grievance. This assumption is partly a convenience to distinguish the model as sharply as possible from the subsequent grievance model, but also has some a priori claim to credibility. Many rebel groups target their recruitment on children and drug addicts, and such selectivity, in conjunction with the phenomenon of relative deprivation, whereby a majority of the population considers itself to be unfortunate, may enable indoctrination to be effective in generating grievance regardless of objective circumstances. This theory of military motivation has two testable predictions. First, objective causes of grievance will not explain the risk of conflict. We return to this in Section 2.2 where we measure objective grounds for grievance. Secondly, since conflict produces military cohesion, rebel military advantage will be greater in societies which have recently experienced conflict. We measure this by the time, t, which has elapsed since the end of any previous civil war.

The model implies that in ethnically diverse societies, rebellion will *appear* to be caused by ethnic differences even though diversity actually reduces the risk of conflict. Where rebellion occurs in such societies, the need for homogeneity in recruitment will make it ethnically specific, and the need for the rebel organization to generate a subjective sense of grievance may focus the discourse on ethnicity.

The hypothesized effects of geography, cohesion and motivation on relative military advantage are summarized in (15):

$$k = k(m, L, Y, f, t). \qquad (15)$$
$$\quad + \; + \; + \; - \; -$$

Rebel costs of recruitment

In the basic model set out above, rebel and government labor costs are treated as being symmetrical: for both the wage is simply linear in per capita income. However, this ignores an important difference between the rebel force and the government army. Our focus is on the *initiation* of rebellion, and in this phase the rebel force must *grow*, and probably needs to grow rapidly to reach the survival constraint at minimum cost. By contrast, the government army is already in steady state. We would expect this to make the labor costs of the rebel organization more sensitive to the state of the labor market than those of the government. The tighter is the labor market, controlling for the level of per capita income, the more costly would be rebel recruitment, and so the lower would be the risk of conflict. We measure the state of the labor market through three proxies: education, population growth, and economic growth. Our education variable is the proportion of young males enrolled in secondary education, s. This both directly measures an important alternative activity to rebellion for that part of the population most likely to be recruited, and proxies income-earning opportunities for the group. The

population growth rate ($\dot{p}$), proxies the change in labor supply, and will consequently proxy the change in pressure on fixed assets, notably land. The growth rate of income proxies the demand for labor. Income growth decomposes into $\dot{p}$ and the growth in per capita income, $\dot{y}$, and since the former is already directly included, we measure income growth net of population growth. The difficulty of rebel recruitment is assumed to be increasing in s and, $\dot{y}$, while being decreasing in $\dot{p}$. To reduce problems of endogeneity, these variables are all measured prior to the period of prediction. In (16) we introduce this as a modification to the rebel wage function, (4):

$$w_r = j\,y.\,\dot{}\,(s,\ \dot{y},\ \dot{p}) \qquad (16)$$
$$\qquad\qquad +\ +\ -$$

Since this change in the wage function affects the financial viability of rebellion, we predict that the risk of conflict decreases with s and $\dot{y}$ and increases with $\dot{y}$.

Start-up finance

We now consider how rebel organizations might acquire start-up finance. The survival condition imposes a minimum size on rebel forces below which they cannot be operational in resource predation. This implies that there are threshold start-up costs. Since rebellions may not be able to raise funding from conventional sources, they must look elsewhere. We consider two potential sources of finance, foreign governments and diasporas.

Foreign governments may provide pump-priming finance for rebellions which subsequently become self-sustaining. An unusually clear example was the financing of Renamo in Mozambique by the government of Southern Rhodesia. The latter collapsed shortly after Renamo was established, but Renamo sustained itself as a rebel force for many years before evolving into a political party. To test the proposition empirically, we need some proxy for the willingness of foreign governments to finance military opposition to the incumbent government. Among the possible proxies, we regard the end of the Cold War as providing the clearest test of the proposition. Undoubtedly, during the Cold War the two major powers tried to enlist other governments as allies, and each power had an incentive to destabilize those governments which allied with the opposing power. Hence, we predict that with the end of the Cold War, the supply of foreign power finance for rebellions declined and that this should have reduced the risk of conflict. We test this by introducing a dummy variable, D_1, for the post-Cold War period. We might note that this is a controversial prediction. For example, Kaplan, (2000) argues that the end of the Cold War `lifted the lid' off previously suppressed conflict.

A further potentially important source of start-up finance for rebellion is a diaspora living in OECD countries. Such diasporas are usually much richer than the population in their country of origin. They are better-placed for collective action: emigrants have a cultural incentive to create diaspora organizations which can then discipline free-riding. They do not suffer the consequences of the conflicts they finance. As with grievance among the local population, in the greed-model grievance among the diaspora is assumed to be manufactured by the rebel organization rather than being an original cause of conflict. Hence, the diaspora increases the risks of conflict renewal, but not the initial risk of conflict. We measure the size of diasporas in the USA relative to the population in their country of origin. We test whether this variable, d, is significant as an initial cause of conflict, and by interacting it with the time since any previous conflict, d/t, whether it is a significant risk factor post-conflict. The theory of greed-conflict predicts that only the latter will be significant.

2.2. A model of grievance-rebellion

In a grievance-rebellion the objective is not predation, but rather to assuage grievance. This is indeed the standard characterization of rebellion in the popular literature. The most proximate economic theory is, as noted above, that of protest movements. Within this framework, a rebellion might be thought of as a protest movement which has failed to escalate into mass participation. Recall that rebellions seldom have more than 5,000 participants, whereas successful protest movements are generally many times this size. In Kuran's model, the early phase of a protest movement is generated by a few ardent adherents who are insensitive to the risks involved. It escalates if this induces a cascade of participation by those who are initially deterred by fear of government reprisals. A theory of rebellion as a failed protest movement must thus explain two stages: the initial impetus for protest, and the failure of escalation.

Consider, first, the circumstances in which some people feel sufficiently strongly to protest, to the extent of being insensitive to the risks of government reprisal. Evidently, the impetus for such protest should be rooted in objective grievances, some of which are potentially measurable. We distinguish three of them: inter-group hatred, political exclusion, and vengeance.

Inter-ethnic or inter-religious hatreds, h, are probably the most common popular explanation for civil conflict. A possible example of such a conflict might be Bosnia. There is indeed evidence that such hatreds exist, and since many conflicts are inter-ethnic or inter-religious, it is widely presumed that the hatreds are the cause of the conflict. Although such hatreds are usually not directly observed, they can evidently only occur in societies which are multi-ethnic or multi-religious. Inter-group hatred may be monotonic in the extent of social fractionalization, or the relationship may be a quadratic: for example, societies with two groups may have a higher incidence of inter-group hatred than societies with many groups. Hence,

$$h = h(f),\ h(0) = 0;\ h(f>0) > 0 \qquad (17)$$

A second presumed cause of grievance is political exclusion, j. The quantitative political science literature has already explored the relationship between conflict and the political rights of a society, q, as the latter range from dictatorial repression to full representative democracy. Econometric studies have found that other than when repression is very severe, it tends to increase the risk of conflict (Gleditsch and Hegre, 1997). However, even democracies may generate grievance if one voting block is able to forge a persistent majority and uses its power to disadvantage a minority. Normally, such persistent majorities are illusive because excluded groups are able to make offers which split the majority. However, one circumstance in which a stable winning coalition can form is where political allegiance is pre-determined by ethnic identity and one ethnic group constitutes a majority of the population. Whether such a group uses its power to extract transfers from the minority depends in part upon the size of the majority. Conditional upon power, large majorities have less incentive to be exploitative than small majorities. For example, if there are fixed costs to inter-group transfers, there is a critical size of minority below which the minority is not worth exploiting. The circumstances in which one or more ethnic groups are permanently politically exploited by a dominant group thus depend upon the largest group having a sufficiently large share of the population to control the political process, but not being so large that exploitation is not worthwhile. The precise range of population shares over which the largest ethnic group exercises such dominance cannot be determined *a priori*, although simple models would bound the minimum at 50% and the maximum at strictly less than 100%. The grievance model postulates that a dummy variable for societies characterized by such ethnic dominance, D_2, will be significant and positive, increasing the risk of conflict. A possible example of grievance-rebellion driven by an excluded ethnic minority might be the conflict in Sri Lanka, where the Tamil Tigers claim to defend the interests of the 12% of the population who are Tamil.

A second circumstance of political exclusion is where the poor are marginalized from the political process. As Hirshleifer (1991) shows, normally the poor will succeed in using the political contest to ameliorate their economic position. A high degree of economic inequality, i, is therefore some indication that the poor are atypically politically marginalized. The `rage of the poor' at high inequality is indeed probably the single most popular explanation for conflict after that of inter-ethnic hatred, and may be exemplified by the Castro rebellion in Cuba. A final circumstance of persistent political exclusion is where a rich minority is heavily taxed by the majority. The rich may contest the government or, following Buchanan and Faith (1987), they might attempt to secede from the fiscal jurisdiction of the state. A possible example of the former might be the Contra rebellion in Nicaragua, and of the latter, the eventually successful secession of the Eritrean region from the much poorer state of Ethiopia. Since there is more incentive for the majority to tax the rich the larger is the share of income or wealth accruing to them, the `rage of the rich' might also be expected to increase with economic inequality.

Hence, grievance due to political repression is postulated to be a function of the general level of political rights, the ethnic composition of the society, and the degree of inequality:

$$j = j(q, D_2, i) \qquad (18)$$

Finally, much of the case study literature dwells upon history: current rebellion is motivated by the desire to revenge atrocities committed during a previous conflict. Corresponding to the greed theory of induced conflict, we assume that the longer the period since a previous conflict, t, the less strong may be the demand for grievance assuagement for such atrocities. Thus, the intensity of grievance, G, is assumed to be a function of these three components:

$$G = G(h, j, t). \qquad (19)$$

Now consider the second stage in protest-rebellion, the failure of the initial protest to escalate into successful political revolution. In the Kuran model the protest movement succeeds in generating political change if, as people join the protest, the resulting reduction in the risk of punishment attracts enough new entrants further to reduce the risk, yielding a cascade. The key parameter is the elasticity of participation with respect to the risk of punishment. Kuran suggests that this elasticity will be greater the more homogenous the society. The participation cascade breaks if there are gaps in the distribution of preferences such as might occur if the society is fractionalized. Hence, social fractionalization enters twice in a model of rebellion as aborted protest. It is both an impetus for the initial cadre of protestors, and an impediment to political revolution. If rebellions are aborted protests they should therefore be significantly more common in fractionalized societies.

However, unlike a protest, a rebellion which generates a civil war is a sustained, full-time effort. Adherents must eat, and so, as with a greed-rebellion, the rebel organization is constrained to be financially viable. Here the grievance-rebellion faces a severe test. Grievance-assuagement is predominantly a public good and evidently it is one which will not be provided by the government. Its financing will therefore face acute difficulties of free-riding. The willingness of non-participants to support the rebel organization financially may be affected by their own level of grievance, so our grievance proxies might affect the risk of conflict both through generating hard core adherents, and by increasing the scope for funding. A further possible source of grievance-related funding is diasporas. There is some evidence that diasporas are more grievance-conscious than the populations from which they originated. Hence, it is possible that diasporas are willing to finance the initiation of rebellion even when local populations are not sufficiently concerned to do so themselves. We have already discussed how this proposition can be tested. However, the collective action problem is so severe in the temporary provision of public goods by non-governmental means, that their supply will usually be negligible. Thus, the principle prediction of an economic theory of grievance rebellion is that such a rebellion will not occur, regardless of the intensity of the grievance. We test this prediction using the grievance proxies in (19).

2.3 Greed-Grievance Interactions

We now compare the greed and grievance models of conflict and consider a synthesis.

The greed model postulates that the cause of initial conflict is an economic calculus of relative military advantage, the government's ability to finance defense expenditure, the scale of primary commodity exports, and the costs of rebel recruitment. If a conflict occurs, the rebel organization will generate subjective grievance. Post-conflict, until this gradually decays, it will increase the risk of subsequent conflict. The extent of this post-conflict risk depends upon the size of diasporas, since they are able to finance rebellion. By contrast, the grievance model postulates that the cause of initial conflict is not an economic calculus but rather is a protest generated by objective grievances: ethnic or religious hatreds, inequality, oppression, or historical vengeance. Rebellions are protests which fail to cascade into non-violent revolution. The scope for rebels to find a refuge, proxied by the three geography variables, can reasonably be added to the grievance model without upsetting its essentially non-economic spirit.

Greed and grievance can co-habit. Where the conditions for greed-rebellion exist but those for grievance-rebellion do not, a group initially motivated by grievance may become dependent upon primary commodity predation for survival, thus transforming itself into a greed-rebellion. Conversely, greed-rebellions need to manufacture subjective grievance for military cohesion and may find an objective grievance an effective basis for generating it. Hence, the presence of primary commodity exports may sustain rebellions which are motivated by objective grievance, while the presence of objective grievance may sustain rebellions motivated by predation. Such interdependence may make case study evidence difficult to interpret.

The two models are evidently not nested, relying largely upon distinct variables, and in Section 4 we compare their predictive power using standard tests. However, one variable, ethnic and religious fractionalization, is included in both models with an opposite predicted sign and so is of particular interest. The grievance model naturally sees ethnic and religious difference as a potential source of hatreds as compared with a homogenous society, and as an impediment to non-violent revolution. It thus predicts that the risk of conflict will be lower in homogenous societies. The greed model sees fractionalization as tightening the constraint on rebellion: the need for organizational cohesion imposes homogeneity and so if the society is diverse the costs of rebel recruitment are raised.

An integrated model of conflict would allow both greed and grievance to initiate conflict. Objective grievance could make recruits more willing to join, and so lower the costs of rebellion. Potentially, diasporas could provide funding for initial rebellions, motivated by objective grievance, rather than just financing subsequent conflicts motivated by induced grievance, as proposed by the greed model. The greed and grievance models thus potentially nest into an integrated model which combines both sets of variables as causes of initial, as well as subsequent conflicts. Testing these models is the agenda for the rest of the paper.

3. Data

Our empirical analysis incorporates several new data sets constructed for the study. In this data

section we present a brief description of the war variable and the various proxies for greed and grievance. The data source for the war variable and all other variables is Hoeffler and Sambanis (2000) where a more detailed data description can be found.

3.1. War data

We analyze the risk of civil war using a panel data set for 161 countries. For each country we have potentially one observation for each of the eight sub-periods 1960-64, 1965-69, ..., 1995-99. For 73 of these observations the country was at peace at the beginning of the period but a civil war started during it. We predict only war-starts, not their continuation.[4] Hence, once civil war has commenced, the possibility for a further initiation of conflict only occurs once peace has been re-established. In fact, most of the countries which had one conflict went on to have further conflicts. Indeed, 47 of the 73 wars were in countries which had already had at least one previous civil war since 1945. We define a civil war as an internal conflict in which at least 1000 battle related deaths (civilian and military) occurred per year. This definition has become standard in the literature following the seminal data collection exercise and analysis by Singer and Small (1982, 1994). Table 1 lists the 73 observations for which a war started during a sub-period. Our new war data set primarily updates that of Singer and Small from 1992 to the end of the decade.

3.2. Rebel Revenue Sources

The three main sources of rebel revenue are primary commodity exports, diasporas and, during the Cold War, foreign powers.

Primary commodity exports are measured as a ratio of GDP at the beginning of each sub-period. This data was obtained from various World Bank sources.

In order to proxy the size of the diaspora we used US immigration data. The US Bureau of the Census provides detailed figures on the size of the foreign born population.[5] Inter-census figures were interpolated. To capture the relative size of the population we divided the US immigration figures by the total population of the country of origin.

The end of the Cold War was proxied by a dummy variable which took the value of unity for the periods 1990-94 and 1995-99.

3.3 Relative Military Advantage

Military advantage reflects geography, rebel cohesion and rebel motivation.

[4] In our earlier work (Collier and Hoeffler, 1998), we used a much smaller sample of wars and conflated war starts with war duration, using a tobit procedure. We now regard this approach as flawed since duration appears to be determined by rather different factors from starts.
[5] Source: http://www.census.gov/population

Using geographic data we calculated a Gini coefficient measuring the dispersion of the population in a country. Analogous to the income Gini coefficient, the Gini coefficient of population concentration will be high if a large proportion of the country's population is concentrated in a relatively small area of the country. The more evenly the population is dispersed across the country the lower is the Gini coefficient of population concentration. For the calculation of this Gini coefficient we used population data per 400km^2 cells.[6] We also experimented with two further measures of population dispersion, population density and the proportion of the population living in urban areas, neither of which were significant. The data source for both series is WDI 1998.

Data on forest coverage was obtained from the Food and Agriculture Organization (FAO)[7]. For each period this provided an estimate of the percentage of a country's land area covered in forests and woods. We could find no satisfactory existing data set on mountainous terrain. The study commissioned a new data set from Dr. Gerrard, a specialist on the subject. The measure allows for the ruggedness of the terrain, rather than simply relying upon altitude and is available from the World Bank.[8]

We measure social fractionalization by ethno-linguistic and religious fractionalization. The ethno-linguistic fractionalization data was obtained directly from the Atlas Naradov Mira (1964) which lists the ethno-linguistic groups for each country. Based on the Altas Naradov Mira data an ethno-linguistic fractionalization index can be calculated.[9] For each country it measures the probability that two randomly selected people do not speak the same language. An index of 0 means that the entire population speaks the same language while a higher index indicates a higher degree of linguistic heterogeneity. The maximum value of the index is 100. Using the same concept we also constructed a religious fractionalization index which measures the probability that two randomly selected people do not share the same religious affiliation.[10] The ethno-linguistic data was measured at about 1960 and the religious data was measured in 1970 and 1980. Since these measures only change very slowly over time we use the linguistic data for all sub-periods and the 1970 data on religion for the observations 1960-70 and the 1980 data for 1980-99. For 1975 we use the average of the 1970 and 1980 data. Social fractionalization can potentially be equal to, additive or multiplicative in religious and ethnic fractionalization. For example, if all the religious divisions coincide with some or all of the ethnic divisions, then the overall fractionalization of the society may be no greater than that measured by ethnic fractionalization. With such coincident divisions, social diversity would simply by the maximum of the underlying ethnic and religious diversity measures. A second possibility is that the religious

[6] We would like to thank Uwe Deichman (World Bank) for extracting the original data from the Geographic Information System (GIS).
[7] Source: http://www.fao.org/forestry
[8] We would like to thank John Gerrard, University of Birmingham, for the compilation of this variable.
[9] We would like to thank Tomila Lankina, University of Oxford, for translating the data entries from the Atlas Naradov Mira. The first study to use this measure in the economics literature is Mauro (1995).
[10] Most of the data was kindly made available by Robert Barro (Barro 1997) and for some countries we used the data from the original source (Barrett 1982).

divisions occur within ethnic groups (or conversely). In this case, social diversity would be approximated by the sum of the underlying measures. A third possibility is that cleavages are cross-cutting, so that ethnic groups are divided by religion and religious groups by ethnicity. In this case social diversity would be approximated by the product of the underlying measures. More precisely, since the index for each measure ranges 0-100, cross-cutting cleavages would be proxied by the product of the two indices plus the maximum of the two underlying indices. It is necessary to add the maximum of the underlying indices to the product, since otherwise, a society which is (say) ethnically diverse but homogenous in religion would appear to be as homogenous as a society that was homogenous in both ethnicity and religion. In the analysis of Section 4 we use only this third measure of social fractionalization, demonstrating in Table 8 that it dominates both the other approaches to aggregating ethnic and religious diversity.

The time since the previous civil conflict is measured in months. If a country experienced a civil war we measure the peace period as the number of months between the end of the civil war and the beginning of the sub-period, 1960, 1965, ..., 1995. If a country never experienced a civil war we measure the peace period since the end of World War II, i.e. in 1960 these countries have 172 months of peace and they accumulate 60 additional peace months in every sub-period.

3.4 Opportunity Cost

We use four measures of the opportunity cost of recruitment: secondary school enrolment rates for men, average income, population growth and income growth.

Male schooling is measured by the gross male secondary school enrollment ratio. This was obtained from the World Bank World Development Indicators (WDI) 1998. Gross male enrollment ratios are defined as the ratio of total male enrollment, regardless of age, to the male population of the age group that officially corresponds to the level of secondary schooling. Thus, the gross enrollment ratio can exceed 100 percent.

Income is measured as the per capita GDP at the beginning of each period. In order to be able to compare income data over time and across countries we mainly use the real PPP adjusted figures as provided in the Penn World Tables Mark 5.6 (PWT). For the 1995 data we used the growth rates from the real PPP adjusted GDP per capita data from the WDI 1998 in order to update the PWT data which only provide data up to 1992.

Population growth is measured as the average annual growth rate of the population in the previous five years. The main source of the population data is WDI 1998. Income growth is measured as the average annual growth rate of real per capita GDP in the previous five years. The income data used to calculate the growth rates is as described above. We combine population growth and per capita income growth into a single variable, which is the weighted difference between them. We experimented with weights, choosing that which gave the variable the highest t-statistic, this being when the population growth rate was given three times the

weight of per capita income growth.

3.5 Sources of Grievances

Here we concentrate on two sources of grievances, political exclusion and inequality, since the variables which proxy sources of hatred (social fractionalization) and vengeance (previous conflicts) have already been discussed above.

Data on political exclusion is available in the Polity III data set (Jaggers and Gurr 1995). We concentrate on the democracy variable which characterizes the general openness of political institutions. The democracy score ranges from 0 to 10 where 10 denotes a highly open regime. In addition we tried the autocracy score from the Polity III data set and a measure of political openness published by Freedom House ("Gastil Index").[11]

We constructed a measure of ethnic dominance based on the ethno-linguistic data from the Atlas Naradov Mira. This dummy takes a value of one if the largest linguistic group makes up between 45 and 90 percent of the total population and zero otherwise. We constructed a number of different dummies to proxy ethnic domination. As reported below, we found that when ethnic dominance was defined by this range the economic and statistical significance of the variable was at a maximum.

Income inequality is measured as in Deininger and Squire (1996) either as the Gini coefficient of income distribution or as the ratio of top quintile's share of income to bottom quintile's share. We proxy land inequality by a Gini coefficient of land distribution, for a more detailed description please refer to Deininger and Squire (1998).

In Table 2 we present some descriptive statistics for the main variables of interest, distinguishing between the peace observations, the war observations and the entire sample. The war observations are on average characterized by a higher proportion of previous civil wars, lower opportunity costs of war (lower secondary school enrolment rates, lower per capita income, lower economic growth and higher population growth) larger populations, higher ethnic fractionalization, more mountainous terrain, a lower democracy score and higher income inequality. However, war observations have a similar average share of primary commodity exports in GDP, similar religious fractionalization, a similar incidence of ethnic dominance and have a smaller diaspora than peace observations.

We now turn to the regression analysis to examine the relationship between these possible causes of conflict and the risk of civil war.

4. Empirical Results

[11] Source: http://www.freedomhouse.org/ratings

Our empirical analysis attempts to predict the risk that a civil war will start during a five-year sub-period, through a logit regression in which the explanatory variables are characteristics at the start of the sub-period.

We start with the greed model (see Table 3). Because per capita income and enrolment in secondary schooling are highly correlated, they cannot be used in the same regression. The first four columns include secondary schooling but not per capita income, which permits a sample of 688 episodes of which 43 are war observations. The final column replicates the core regression using per capita income instead of secondary schooling, which permits a sample of 747 episodes of which 47 are war observations.

The first column omits the variables which proxy the effect of previous civil war. The two opportunity cost proxies are significant with the expected signs. A higher gross secondary school enrolment rate for males reduces the risk of war. The difference between income growth per capita and population growth, both measured for the previous five year period, decreases the risk of war. As predicted in (12), the effect of primary commodity exports on the risk of war is significant but non-linear. The highest risk of war is at a share in GDP of about 25 percent. The positive linear effect on risk suggests that there is a predatory element in rebellion. The negative quadratic effect suggests that the increased tax revenue eventually augments the capacity of the government to defend itself sufficiently to offset the attraction of enhanced loot. The elasticity of the risk of conflict with respect to population is approximately unity, as predicted. A higher degree of fractionalization makes societies significantly safer, high cross-cutting ethno-linguistic and religious fractionalization decrease the risk of civil war. Two of the technology parameters, the dispersion of the population and mountainous terrain, are also significant. More mountainous countries are more likely to experience a civil war, while a higher concentration of the population decreases the risk of conflict.

We then introduce the effect of previous conflicts. We hypothesize that a previous civil war may increase the risk of experiencing a civil war. Initially, we test this hypothesis by including a dummy which equals one if the country experienced a civil war in the past. Here we include any civil wars which occurred between the end of World War II and 1995. The coefficient on this previous war dummy is highly significant, previous wars increase the risk of new civil war. However, such a dummy could equally be picking up unobserved country-specific effects. We then introduce the length of the peace period measured in months since the end of the last civil war. In column 3 we add this peace duration variable to the model. The coefficient is negative and significant, i.e. the longer the peace period the lower the risk of conflict. The coefficient on the previous war dummy is now insignificant. Thus, the peace period is a more precise measure of the effect than is the previous war dummy. This indicates that the risk decays after conflict, as might grievance, rather than being a proxy for unobserved country-specific effects. In column 4 we therefore drop the insignificant previous war dummy, leaving a model in which greed causes initial conflict; conflict causes some effect such as grievance; and this grievance causes further

conflict.[12] In the last column we present an alternative specification in which we include the average income per capita as a measure of opportunity costs instead of schooling. The results are very similar, although the overall performance of the regression deteriorates.

In Table 4 we investigate whether the effect of post-conflict grievance on the risk of conflict is reinforced by the diaspora, as suggested by the analytic model. Since the data set on the size of diasporas in the USA reduces the number of countries on which we have data, the sample size is consequentially radically reduced from the 747 observations and 47 wars which is our maximum sample in the previous analysis. In order to preserve sample size we therefore retreat to a more parsimonious version of the model, dropping four sample-constraining peripheral explanatory variables: ethnic and religious fractionalization, geographic concentration of the population, the extent to which the terrain is mountainous, and the rate of growth in the previous five year period. 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 595 observations (containing 29 war observations). However, all the included explanatory variables remain significant.

On this sample, we then test for the effect of a diaspora. We measure the size of the diaspora relative to the resident population in the country of origin. To allow for a fading post-conflict effect, we interact this measure with the number of months since the previous conflict, dividing the former by the latter. This variable, dias/peace in Table 4, is added to the regression, the result being shown in the first column. The variable is positive and significant, indeed, slightly more significant than the number of months of peace by itself, although the latter also remains significant. A large diaspora considerably increases the risk of further conflict. If we compare the post-conflict society with the largest diaspora against that with the smallest, with other variables at their mean values, after five years of peace the risk of renewed conflict is around six times greater.

However, while this result is consistent with the analytic model, it is also open to a more anodyne interpretation. Evidently, diasporas are to an extent endogenous to the intensity of conflict: when civil war occurs, many people leave and settle in the USA. Hence, the size of the diaspora might simply be proxying the intensity of conflict. The result may therefore simply show that more severe conflicts have a higher risk of renewed conflict. To test for this we decomposed observed diasporas into a component which is exogenous to the intensity of conflict and a residual component which is therefore endogenous to its intensity. For this decomposition we estimated a simple migration model, reported in the Appendix, Table A3. The size of the diaspora in a census year is predicted to be a function of its size in the previous

[12] We examined this time effect 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, Hoeffler and Söderbom (1998) support the results presented in this paper.

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. We then replace the diaspora data used in the model with estimated diaspora size in all cases where the observed diaspora is for a year subsequent to a conflict. Thus, all post-conflict observations of diasporas are estimates which are purged of any effect from the intensity of conflict. The difference between these estimates and the actual figures are 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 second column of Table 4. The purged measure of the diaspora remains significant, and the size of the coefficient is only slightly altered. Further, its coefficient is not significantly different from that on the endogenous diaspora measure. Had the effect of the diaspora been simply a proxy for the intensity of conflict, neither of these would have been the case. The purged variable would have been insignificant, and the coefficient on the endogenous measure would have been larger. This suggests that the substantial effect of the diaspora on the risk of conflict renewal is indeed due to its financial contribution to war start-up.

In Table 5 we turn to the examination of a rebellion which is motivated only by grievance. In the first column we examine the relationship between ethnic dominance, social fractionalization, democracy and the duration of peace. As in the greed model, we control for geographic military advantage by including the dispersion of the population and mountainous terrain. Since we are not including any lagged variables we can use 884 observations of which 57 observations experienced an outbreak of civil war. The results suggest that societies which are dominated by one large ethnic group (45-90 percent of the population belonging to the same ethnic group) are more likely to experience conflict. Greater openness of political institutions reduces the risk of conflict. In the second column we add the gini coefficient of income inequality as an explanatory variable and in column 3 we add the gini coefficient of land inequality as an alternative measure of inequality. Neither measure is significant. Note that the sample size is reduced when we include the income inequality measures. However, we are still analyzing a substantial number of war occurrences (40 in column 2 and 38 in column 3).[13] All three grievance models have very low explanatory power, the regressions only have an R^2 of 0.14 or lower.

We now turn to the question of which model, greed 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). This non-nested 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, greed and grievance.

(1) p=f(greed)
(2) p=f(grievance)

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

Based on these logit regressions we calculate the predicted probabilities and add these predicted values, $\hat{p}^{\text{greed}}$ and $\hat{p}^{\text{grievance}}$ to our alternative models.

(1) p=f(greed, $\hat{p}^{\text{grievance}}$)
(2) p=f(grievance, $\hat{p}^{\text{greed}}$)

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 greed model we reject the greed model in favor of the grievance model. If $\hat{p}^{\text{greed}}$ is significant in the grievance model we reject the grievance model in favor of the greed model. As can be seen in columns 2 and 3 of Table 6 $\hat{p}^{\text{grievance}}$ is significant in the greed model and $\hat{p}^{\text{greed}}$ is significant in the grievance model. Thus, we conclude that we cannot choose between the two models, grievance adds explanatory power to the greed model and vice versa.[14] Since we find neither model dominates the other, we next investigate the combination of the two models as presented in column 3 of Table 6.

In this combined model all variables apart from democracy are significant and the overall fit is reasonable (pseudo R^2 of about 0.3). Since both the grievance and greed 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 both the greed and the grievance models.[15]

Although the combined model is superior to the greed and grievance models, one variable, democracy, is completely insignificant.[16] Dropping it yields the model in column 4. Without the democracy variable the coefficients and standard errors of all of the other variables remain virtually unchanged. Based on a likelihood ratio test we cannot reject the hypothesis of a zero coefficient on the democracy variable.[17] While all variables in the model of column 4 are significant, that for mountainous terrain is only significant at the ten percent level and in column 5 we investigate a more parsimonious model in which it is dropped. Since this model is nested within the two previously presented models, we again use likelihood ratio tests to compare it to those of columns 3 and 4. The tests narrowly favor this model over a model which includes

[14] When we tested an alternative specification of the greed model which includes income rather than schooling, we found that grievance did not seem to add explanatory power to the greed model, i.e. the coefficient $\hat{p}^{\text{grievance}}$ was insignificant.

[15] Using the same sample as for the combined model (n=665) we obtain the following results: Greed model versus combined model, 2 degrees of freedom, Likelihood Ratio Test statistic 6.34, p-value 0.042; grievance model versus combined model, 5 degrees of freedom, LRT statistic 41.76, p-value 0.000.

[16] We tried different specifications to test for the effect of political repression 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.

[17] Using the same sample (n=665) we obtained a log likelihood of –111.25 for the restricted model. Thus, the LRT statistic is 2.01, 1 degree of freedom, p-value 0.16.

mountainous terrain.[18] No further reduction in the model is accepted, and no additions of variables included in our previous models are accepted. We refer to the model in column 5 as our baseline model.

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 the Appendix, Table A5. At the median of all variables a country would have a 1.4 percent probability of experiencing a civil war, while at the mean of all variables the risk of a war-start is higher, about 14 percent. Our model predicts that a country with the worst characteristics (lowest per capita income, lowest GDP growth and highest population growth, a primary commodity export share of 0.26, the largest population, the lowest fractionalization, ethnic dominance, a geography Gini coefficient of zero and only one month of peace) would have a near-certain risk of war while a country with the best characteristics would be a very safe society. We now calculate how the change in one variable (while the others are assumed to take their mean values) affects the probability of experiencing a civil war. Whether a country is characterized by ethnic dominance has a considerable impact on the risk of experiencing a civil war. Countries without a dominant ethnic group have a risk of war of about 10 percent, however, the risk is doubled if countries have a dominant ethnic group. If the male secondary school enrollment rate is 10 percentage points higher than the average, the risk of war is reduced by about four percentage points (a decline in the risk from 14% to 10%). Higher income growth reduces the risk of conflict: an additional percentage point on the growth rate reduces the risk of war by one percentage point (a decline from 14% to 13%). A reduction in population growth has a larger effect on the risk of war, a one percentage point decrease implies a 2.5 percentage point reduction in the risk of conflict (from 14% to 11.5%). The effect of primary commodity exports is very considerable: at the risk maximizing value of the primary commodity export share (0.26), the risk of civil war is about 23 percent, while a country with no natural resource exports only has a probability of a war-start of 0.5 percent. A society characterized by the maximum social fractionalization has a very small risk of experiencing a civil war (about three percent), however, the risk is about eight times higher for a country which is socially homogenous. Directly after a civil war there is a high probability of a re-start, the risk is about 42 percent. This risk declines over time, however, sustained peace only slowly decreases the risk of experiencing a new conflict. Five years after the end of the civil war the risk is still about 37 percent. The elasticity of the risk of conflict with respect to population is 0.95, close to the analytic expectation of 0.94 discussed in Section 2. Lastly, we examine the impact of the dispersion of the population on the risk of conflict. Countries with a highly concentrated population are very safe from conflict while countries which are characterized by a homogenously dispersed population have a very high risk of civil war (about 60 percent).

In column 6 of Table 6 we run the baseline model with average per capita income instead of

[18] First, using the same sample (n=665) we test the combined model (column 3) versus the baseline model (column 5). We obtain a LRT statistic of 3.24, 2 degrees of freedom, p-value 0.198. Second, we test the reduced model (column 4) against the baseline model (column 5). We obtain a LRT statistic of 2.03, 1 degree of freedom, p-value 0.16.

schooling. Recall that since we have more data on income than on schooling we are able to use a larger sample which includes four additional war observations. The variables are all significant at conventional levels, although the overall explanatory power is slightly lower. Using this maximum sample we analyze the effect of external assistance as causes of war. Adding a post cold-war dummy (which takes a value of one for the periods 1990-94 and 1995-99) in column 7 we find that that the risk of civil war was significantly higher during the cold war period. This is consistent with the prediction that the super-powers eased the financing constraint on rebellion.[19] Thus, we find some weak evidence that external assistance increased the risk of civil war during the cold war.

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, and indeed, all those models which include a feedback effect from previous conflict, we accept the hypothesis that we can pool across the observations.[20] 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 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.

We analyzed 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.[21] However, as discussed above, the dummies for 1990-94 and 1995-99 are jointly significant and support consolidation into a single dummy for the post Cold War period in the variant of the baseline model.

Finally, in the last column of Table 7 we use a recently developed correction method for rare events data (King and Zeng, 2000). Our data is characterized by a relatively small number of events (wars), only about six percent of the observations are characterized by a civil war outbreak. The results presented in King and Zeng (2000) suggest that standard logit estimation underestimates the probability of an event occurring when the events are rare. We used their correction procedure but found the differences between the standard logit results and the rare

[19] This result only holds for the augmented sample which includes 47 civil wars. Due to data constraints the baseline model excludes four of these wars all of which started during the cold war period. The wars we have to exclude are Chad (1980-84), China (1965-69), Jordan (1970-74) and Liberia (1985-89).

[20] The null-hypothesis (rho=0) is not rejected: the chi-squared statistic is 0 (p = 0.998). When the feedback effect from previous conflict is excluded, pooling is marginally rejected: the chi-squared statistic is 2.02 (p=0.155).

[21] The LRT statistic is 2.1, 6 restrictions, p=0.91.

events corrected results to be negligible. Using the rare events logit procedure, all coefficients on the variables have the same signs and are significant at the same levels. The mean of the predicted probabilities obtained from the standard logit regression is 0.063 and the predicted probabilities from the rare events logit regression have a mean of 0.068. The standard deviations and the extreme values of the predicted probabilities obtained from the standard logit and the rare events logit were also similar. Thus, contrary to King and Zeng (2000), we do not find that the logit substantially underestimates the probability of the event occurring.

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, the proportion of the population living in urban areas and the proportion of young men aged 15 to 29. In Tables 8 and 9 we present a more detailed analysis of ethnic fractionalization. In Table 8 we demonstrate that the measure of cross-cutting cleavages (the product of religious and ethnic fractionalization plus the maximum of each component separately), 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 Table 9 we investigate alternative definitions of ethnic dominance. In the first column we include the ratio of the largest ethnic group to the total population in the model and find its coefficient to be insignificant. We then construct a number of different ethnic dominance dummies, taking the value one if 45-90, 45-85, 45-95, 30-90, 40-90, or 50-90 percent of the total population belong to the same ethnic group. As can be seen, the relationship between ethnic dominance is most significant when the largest ethnic group makes up 45-90 percent of the total population.[22]

5. Conclusion

We started from two simple models of rebellion. Greed-rebellion was motivated by predation of the rents from primary commodity exports, subject to an economic calculus of costs and a military survival constraint. Grievance-rebellion was motivated by hatreds which might be intrinsic to ethnic and religious differences, or reflected objective resentments such as domination by an ethnic majority, political repression, or economic inequality. Both of these models could allow for technological constraints upon rebellion such as differences in geography. We then introduced the possibility of a feedback effect, whereby conflict increased the risk of subsequent conflict through generating grievance. In grievance-rebellion this added a simple destabilizing loop: grievance causes conflict which then causes further grievance. In greed-rebellion the feedback was more complex because induced grievance only raised the risk of further conflict if it augmented the potential financial rewards to rebellion. We postulated that such a loop might arise if grievance enabled rebellions to tap into the resources of diasporas. Finally, we allowed for interdependence between greed and grievance. Greed-rebellions need to generate grievance for military cohesion, grievance rebellions might be driven to predation to raise finance.

[22] We also experimented with various other definitions of ethnic dominance, such as the dummy taking a value of one if the largest ethnic group made up 50-75, 50-80, 50-85, 50-90 percent of the population. Neither of these definitions provided a better fit than our preferred definition based on 45-90 percent.

Using a comprehensive panel data set of conflict over the period 1960-99 we examine the risk of civil war using logit regressions. Starting with the distinct greed and grievance models, we find that most of the proxies for objective grievance are insignificant and that the best-performing grievance model has very low explanatory power. By contrast, the simple greed model performs well. The extent of primary commodity exports is the largest single influence on the risk of conflict. Strikingly, ethnic and religious fractionalization makes a society safer, as predicted by the greed model, rather than more dangerous as predicted by the grievance model. We find that a war substantially increases the risk of subsequent war, although this risk fades over time. Thus, although objective grievance is not a powerful primary cause of conflict, conflict may generate grievances which become powerful additional risk factors. However, we can show that by far the strongest effect of war on the risk of subsequent war works through diasporas. After five years of post-conflict peace, the risk of renewed conflict is around six times higher in the societies with the largest diasporas in America than in those without American diasporas. Presumably this effect works through the financial contributions of diasporas to rebel organisations.

Finally, we test for an integrated greed-grievance model. We find that only one of the potential sources of objective grievance significantly adds to the explanatory power of the greed model, namely ethnic dominance. Societies in which the largest ethnic group has between 45% and 90% of the population have around double the risk of conflict of other societies. Presumably, this is because such societies have both the power and the incentive to exploit their minorities.

Our results thus contrast with conventional beliefs about the causes of conflict. A stylized version of these beliefs would be that grievance begets conflict which begets grievance which begets further conflict. With such an analysis, the only point of intervention is to reduce the level of objective grievance. Our model suggests that what is actually happening is that opportunities for primary commodity predation cause conflict, and that the grievances which this generates induce diasporas to finance further conflict. The policy intervention points here are reducing the absolute and relative attraction of primary commodity predation, and reducing the ability of diasporas to fund rebel movements.

References

Barro, R. J. 1997. *Determinants of Economic Growth.* Cambridge, MA: MIT Press.

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

Buchanan, J.M. and R.L. Faith. 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, (forthcoming).

Collier, P. and A. E. Hoeffler. 1998. On the Economic Causes of Civil War, *Oxford Economic Papers*, 50:563-73.

Collier, P., A. E. Hoeffler, and M. Söderbom. 1998. On the Duration of Civil War and Postwar Peace. Centre for the Study of African Economies, Discussion Paper 98/8.

Davidson R. and J.G. MacKinnon. 1981. Several Tests for model specification in the presence of alternative hypotheses. *Econometrica* 49:781-793.

Deininger, K., and L. Squire. 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* 57:249-287.

Department of Geodesy and Cartography of the State Geological Committee of the USSR. 1964. Atlas Naradov Mira. Moscow.

Gerrard, A.J.W. 2000. 'What is a Mountain?' Mimeo. DECRG, World Bank.

Gleditsch, Nils Petter, and Havard Hegre. 1997. Peace and Democracy: Three Levels of Analysis, *Journal of Conflict Resolution*, 41:283-310.

Grossman, H. I. 1999. Kleptrocracy and Revolutions. *Oxford Economic Papers* 51:267-283.

Herbst, Jeffrey. 2000. *States and Power in Africa*, Princeton, Princeton University Press.

Hirshleifer, J. 1991. 'The Paradox of Power'. *Economics and Politics* 3:171-200.

Hoeffler, A. E. and N. Sambanis. 2000. Data Notes - World Bank Project on 'The Economics of Civil Wars'. Mimeo.

Kaplan, Robert, D. 2000. *The Coming Anarchy*, New York, Random House.

Keith Jaggers and Ted Robert Gurr. 1995. 'Tracking Democracy's Third Wave with the Polity III Data', *Journal of Peace Research* 32:469-482.

King, G., and L. Zeng. 2000. Logistic Regression in Rare Events Data. http://gking.harvard.edu/preprints.shtml

Konrad, Kai, A. and Stergios Skaperdas. 1998. 'Extortion', *Economica*, 65:461-77.

Kuran, T. 1989. 'Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution'. *Public Choice* 61:41-74.

Londregan, J.B. and K. T. Poole. 1996. 'Does High Income Promote Democracy?' *World Politics* 49:1-30.

Mauro, P. 1995. 'Corruption and Growth'. *The Quarterly Journal of Economics* 110:681-712.

Sachs, J.-D. A.M. Warner. 1995. 'Natural Resource Abundance and Economic Growth'. National Bureau of Economic Research Working Paper: 5398.

The Stockholm International Peace Research Institute. 1993. *Yearbook of World Armaments and Disarmaments*. Oxford: Oxford University Press.

Table 1: Outbreaks of War

country	year	war started	prev. war	country	year	war started	prev. war
Afghanistan	1975-79	1	0	Mozambique	1960-64	1	0
Afghanistan	1990-94	1	1	Mozambique	1975-79	1	1
Algeria	1960-64	1	1	Myanmar/Burma	1965-69	1	1
Algeria	1990-94	1	1	Myanmar/Burma	1980-84	1	1
Angola	1960-64	1	1	Nicaragua	1975-79	1	0
Angola	1975-79	1	1	Nicaragua	1980-84	1	1
Azerbaijan	1990-94	1	0	Nigeria	1965-69	1	0
Bosnia	1990-94	1	0	Nigeria	1980-84	1	1
Burundi	1970-74	1	0	Pakistan	1970-74	1	0
Burundi	1984-89	1	1	Peru	1980-84	1	0
Burundi	1990-94	1	1	Philippines	1970-74	1	1
Cambodia	1970-74	1	1	Romania	1984-89	1	0
Chad	1980-84	1	0	Russia	1990-94	1	0
China	1965-69	1	1	Russia	1995-99	1	1
Colombia	1980-84	1	1	Rwanda	1960-64	1	1
Dominican Republic	1965-69	1	0	Rwanda	1990-94	1	1
El Salvador	1975-79	1	0	Somalia	1980-84	1	0
Ethiopia	1970-74	1	1	Somalia	1984-89	1	1
Georgia	1990-94	1	0	Sri Lanka	1970-74	1	0
Guatemala	1965-69	1	1	Sri Lanka	1980-84	1	1
Guatemala	1970-74	1	1	Sudan	1960-64	1	0
Guatemala	1975-79	1	1	Sudan	1980-84	1	1
Guinea-Bissau	1960-64	1	0	Tajikistan	1990-94	1	0
India	1980-84	1	1	Turkey	1990-94	1	0
Indonesia	1975-79	1	1	Uganda	1965-69	1	1
Iran	1970-74	1	1	Uganda	1980-84	1	1
Iran	1975-79	1	1	Vietnam	1960-64	1	1
Iran	1980-84	1	1	Yemen	1990-94	1	0
Iraq	1970-74	1	1	Yemen AR	1960-64	1	1
Iraq	1984-89	1	1	Yemen PR	1984-89	1	0
Iraq	1990-94	1	1	Yugoslavia	1990-94	1	0
Jordan	1970-74	1	0	Yugoslavia	1995-99	1	1
Laos	1960-64	1	1	Zaire	1960-64	1	1
Lebanon	1975-79	1	1	Zaire	1990-94	1	1
Liberia	1984-89	1	0	Zaire	1995-99	1	1
Liberia	1990-94	1	1	Zimbabwe	1970-74	1	0
Morocco	1975-79	1	1				

Table 2: Descriptive Statistics

	sample (n=1174)					no civil war (n=1101)					civil war (n=73)				
	mean	st.dev.	min	max	n	mean	st.dev.	min	max	n	mean	st.dev.	min	max	n
war starts	0.06				1174	0				1101	1				73
previous war	0.34				1174	0.32				1101	0.64				73
male secondary schooling	43.3	31	0.3	147	965	44.3	31	0.3	147	902	30.2	26.6	1	102	63
GDP per capita (const. US$)	4056	4317	222	33946	1014	4200	4388	257	33946	958	1644	1353	222	5832	56
GDP per capita growth t-1	1.62	3.83	-22.08	14.41	864	1.72	3.78	-22.1	14.41	815	-0.05	4.28	-10.66	7.09	49
population growth t-1	2.15	1.54	-6.38	16.62	991	2.12	1.57	-6.38	16.62	929	2.50	1.01	-0.013	5.81	62
primary commodity exports/GDP	0.17	0.19	0	2.14	1090	0.17	0.19	0	2.14	1023	0.14	0.1	0.01	0.48	67
population (millions)	25.4	94.5	0.014	1200	1161	24.5	92.9	0.014	1200	1089	40.0	116.4	0.54	715	72
ethnic fractionalization	40	28	0	93	1107	39	28	0	93	1039	50	29	4	90	68
religious fractionalization	36	24	0	79	1147	36	24	0	79	1078	36	25	0	69	69
geographic dispersion	0.57	0.23	0	0.97	1028	0.57	0.23	0	0.97	961	0.60	0.15	0.24	0.92	67
mountainous terrain	15.8	20.73	0	94.3	1174	15.12	20.28	0	94.3	1101	26.73	24.2	0	81	73
ethnic dominance (45-90%)	0.47		0	1	1107	0.47				1039	0.47				68
democracy	3.90	4.21	0	10	987	4.21	4.21	0	10	987	1.67	2.58	0	9	67
diaspora	0.02	0.03	0	0.19	654	0.02	0.03	0	0.19	654	0.004	0.005	0	0.16	37
Income Inequality (Gini)	41.35	10.27	17.83	63.2	721	40.65	10.27	17.83	63.2	721	41.97	10.30	23.38	62.3	44

Table 3: Greed Model

	1	2	3	4	5
male secondary schooling	-0.036 (0.011)***	-0.037 (0.011)***	-0.028 (0.012)**	-0.027 (0.012)**	
ln GDP per capita					-0.822 (0.288)***
(GDP growth - 3*population growth) t-1	-0.080 (0.036)**	-0.074 (0.036)**	-0.084 (0.036)**	-0.086 (0.036)**	-0.101 (0.035)***
primary commodity exports/GDP	34.088 (8.186)***	34.109 (8.423)***	32.147 (8.497)***	32.058 (8.465)***	22.923 (6.903)***
(primary commodity exports/GDP)^2	-67.792 (18.683)***	-67.172 (18.679)***	-62.307 (18.954)***	-62.091 (18.962)***	039.192 (14.710)***
ln population	1.026 (0.191)***	0.960 (0.194)***	0.832 (0.198)***	0.832 (0.197)***	0.574 (0.146)***
social fractionalization	-0.0002 (0.0001)**	-0.0003 (0.0001)***	-0.0003 (0.0001)***	-0.0003 (0.0001)***	-0.0003 (0.0001)***
geographic dispersion	-3.517 (1.142)***	-3.888 (1.181)***	-3.384 (1.200)***	-3.289 (1.180)***	-1.568 (1.051)
mountainous terrain	0.019 (0.009)**	0.013 (0.009)	0.014 (0.010)	0.015 (0.010)	0.012 (0.009)
previous war dummy		1.016 (0.405)***	0.252 (0.527)		
Peace duration			-0.004 (0.002)**	-0.004 (0.001)***	-0.004 (0.001)***
n	691	691	688	688	747
no of wars	43	43	43	43	47
pseudo R^2	0.25	0.27	0.29	0.29	0.26
log likelihood	-120.98	-117.63	-114.32	-114.43	-129.44

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

Table 4: Greed – Diaspora Model

	1	2
ln GDP per capita	-1.032 (0.285)***	-1.037 (0.286)***
primary commodity exports/GDP	23.98 (8.542)***	23.78 (8.56)***
(primary commodity exports/GDP)^2	-48.163 (22.1)***	-47.75 (22.12)**
ln population	0.318 (1.156)**	0.319 (1.155)**
peace duration	-0.002 (0.001)***	-0.002 (0.001)*
diaspora/peace duration	648.77 (367.57)*	
predicted diaspora/peace		687.09 (393.12)*
(diaspora-predicted diaspora)/peace duration		768.9 (562.42)
n	595	595
no of wars	29	29
pseudo R^2	0.23	0.23
log likelihood	-89.05	-89.01

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

Table 5: Grievance Model

	1	2	3
social fractionalization	0.00003 (0.00007)	0.00000 (0.00009)	0.00008 (0.00009)
ethnic dominance (45-90%)	0.263 (0.290)	0.534 (0.342)	0.567 (0.374)
democracy	-0.117 (0.046)***	-0.091 (0.051)**	-0.138 (0.052)***
geographic dispersion	0.095 (0.787)	-0.276 (0.958)	-0.152 (1.039)
mountainous terrain	0.015 (0.006)**	0.011 (0.008)	0.015 (0.006)**
income inequality		0.014 (0.016)	
land inequality			-0.150 (1.159)
Peace duration	-0.005 (0.001)***	-0.004 (0.001)***	-0.005 (0.001)***
n	884	614	620
no of wars	57	40	38
pseudo R^2	0.13	0.11	0.14
log likelihood	-184.17	-132.11	-122.61

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

Table 6: Combined Greed and Grievance Model

	1	2	3	4	5	6	7
male secondary schooling	-0.026 (0.012)**		-0.030 (0.012)***	-0.031 (0.012)***	-0.036 (0.011)***		
ln GDP per capita						-1.007 (0.281)***	-1.006 (0.283)***
(GDP growth - 3*population growth) t-1	-0.079 (0.036)**		-0.085 (0.037)**	-0.090 (0.037)**	-0.085 (0.036)**	-0.103 (0.035)***	-0.113 (0.035)***
primary commodity exports/GDP	31.902 (8.333)***		33.576 (8.514)***	34.220 (8.610)***	32.089 (8.375)***	22.983 (6.806)***	23.025 (6.795)***
(primary commodity exports/GDP)^2	-61.491 (18.489)***		-64.234 (18.873)***	-65.676 (19.186)***	-62.511 (18.964)***	-39.293 (14.505)***	-39.026 (14.394)***
ln population	0.840 (0.203)***		0.912 (0.211)***	0.934 (0.210)***	0.946 (0.204)***	0.625 (0.148)***	0.678 (0.153)***
social fractionalization	-0.0003 (0.0001)**	0.00003 (0.0001)	-0.0003 (0.0001)**	-0.0003 (0.0001)**	-0.0003 (0.0001)***	-0.0004 (0.0001)***	-0.0003 (0.0001)***
democracy		-0.042 (0.058)	-0.043 (0.058)				
ethnic dominance (45-90%)		0.895 (0.373)**	0.909 (0.477)***	0.882 (0.374)***	0.847 (0.370)**	0.623 (0.348)*	0.676 (0.351)**
geographic dispersion	-2.443 (1.270)**	0.023 (1.016)	-3.670 (1.208)***	-3.600 (1.200)***	-3.891 (1.172)***	-1.851 (1.006)*	-2.020 (1.023)*
mountainous terrain	0.003 (0.011)	0.0004 (0.009)	0.016 (0.010)*	0.016 (0.010)*			
peace duration	0.001 (0.003)	0.0004 (0.0015)	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***
post-cold war dummy							-0.980 (0.499)**
grievance residual	0.897 (0.481)*						
greed residual		1.066 (0.200)***					
n	665	665	665	688	688	747	747
no of wars	43	43	43	43	43	47	47
pseudo R^2	0.30	0.30	0.30	0.31	0.30	0.27	0.28
log likelihood	-112.27	-111.02	-110.97	-111.61	-113.03	-128.71	-126.47

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

Table 7: Robustness Tests of the Baseline Model

	1	2	3	4
	Random Effects	Fixed Effects	Pooled Logit plus Time dummies	Rare Events Logit
male secondary schooling	-0.036 (0.011)***	0.0001 (0.033)	-0.034 (0.012)***	-0.034 (0.011)***
(GDP growth - 3*population growth) t-1	-0.085 (0.036)**	-0.063 (0.067)	-0.105 (0.041)***	-0.080 (0.036)**
primary commodity exports/GDP	32.089 (8.375)***	34.885 (14.945)**	32.978 (8.648)***	29.399 (8.667)***
(primary commodity exports/GDP)^2	-62.511 (18.964)***	-63.363 (28.824)**	-65.483 (19.822)***	-55.982 (22.781)***
ln population	0.946 (0.204)***	0.373 (1.374)	0.960 (0.210)***	0.891 (0.155)***
social fractionalization	-0.0003 (0.0001)***	-0.007 (0.005)	-0.0003 (0.0001)***	-0.0003 (0.0001)***
ethnic dominance (45-90%)	0.847 (0.370)**		0.913 (0.381)**	0.807 (0.350)**
geographic dispersion	-3.891 (1.172)***	78.555 (80.216)	-3.936 (1.196)***	-3.722 (1.041)***
Peace duration	-0.004 (0.001)***	0.01 (0.003)***	-0.004 (0.001)***	-0.004 (0.001)***
T70-74			-0.960 (0.841)	
T75-79			0.145 (0.776)	
T80-84			-0.002 (0.777)	
T85-89			0.477 (0.749)	
T90-94			-0.820 (0.855)	
T95-99			-0.229 (0.168)	
n	688	135	668	688
no of wars	43	43	43	43
pseudo R^2			0.33	
log likelihood	-113.03	-112.27	-107.44	-113.03

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

Table 8: Social Fractionalization

	1	2	3	4	5	6	7	8	9	10
	Baseline Model	Baseline Model plus ethnic fractionalization	ethnic fractionalization only	Baseline Model plus religious fractionalization	religious fractionalization only	ethnic and religious fractionalization	Baseline Model plus ethnic and religious fractionalization	social frac.= ethnic plus religious fractionalization	social frac.= ethnic minus religious fractionalization	social frac.= ethnic times religious fractionalization
male secondary schooling	-0.036 (0.011)***	0.037 (0.012)***	-0.025 (0.001)***	-0.037 (0.012)***	-0.034 (0.011)***	-0.036 (0.113)***	-0.037 (0.116)***	-0.034 (0.011)***	-0.021 (0.010)**	-0.036 (0.011)***
(GDP growth - 3*population growth) t-1	-0.085 (0.036)**	-0.083 (0.037)**	-0.093 (0.035)***	-0.084 (0.037)**	-0.078 (0.036)**	-0.081 (0.036)**	-0.084 (0.037)**	-0.086 (0.036)**	-0.082 (0.036)**	-0.085 (0.036)**
primary commodity exports/GDP	32.089 (8.375)***	32.700 (8.425)***	23.442 (7.213)***	32.019 (8.390)***	28.368 (7.872)***	29.447 (8.060)***	32.862 (8.454)***	28.809 (8.003)***	21.120 (6.822)***	32.147 (8.381)***
(primary commodity exports/GDP)^2	-62.511 (18.964)***	-64.283 (19.044)***	-44.962 (16.501)***	-62.791 (19.035)***	-56.891 (18.260)***	-58.464 (18.633)***	-64.450 (19.020)***	-56.049 (18.491)***	-42.019 (15.923)***	-62.621 (18.970)***
ln population	0.946 (0.204)***	0.925 (0.206)***	0.749 (0.177)***	0.941 (0.205)***	0.828 (0.186)***	0.892 (0.201)***	0.924 (0.206)***	0.904 (0.199)***	0.643 (0.164)***	0.946 (0.204)***
social fractionalization	-0.0003 (0.0001)***	-0.004 (0.001)***		-0.0003 (0.0002)*			-0.0005 (0.0003)*	-0.014 (0.005)***	0.004 (0.007)	-0.0003 (0.0001)***
ethnic fractionalization		0.009 (0.010)	-0.0138 (0.007)**			-0.007 (0.007)	0.011 (0.013)			
religious fractionalization				-0.006 (0.014)	-0.026 (0.009)***	-0.023 (0.009)**	0.004 (0.018)			
ethnic dominance (45-90%)	0.847 (0.370)**	0.754 (0.381)**	0.923 (0.369)***	0.816 (0.375)**	0.773 (0.372)**	0.833 (0.379)**	0.756 (0.381)**	0.940 (0.369)***	0.923 (0.373)***	0.847 (0.370)**
geographic dispersion	-3.891 (1.172)***	-4.099 (1.201)***	-2.530 (1.042)**	-4.070 (1.242)***	-4.129 (1.222)***	-4.129 (1.230)***	-4.027 (1.245)***	-3.711 (1.163)***	-2.581 (1.053)***	-3.894 (1.172)***
peace duration	-0.004 (0.001)***	-0.0004 (0.001)***	-0.005 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***
N	688	688	668	688	688	688	688	688	688	688
no of wars	43	43	43	43	43	43	43	43	43	43
pseudo R^2	0.30	0.30	0.27	0.30	0.29	0.29	0.30	0.29	0.26	0.30
log likelihood	-113.03	-112.60	-118.56	-112.93	-114.42	-113.96	-112.57	-114.62	-118.98	-113.00

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

Table 9: Ethnic Dominance

	1	2	3	4	5	6	7
	Largest group as a proportion of the total population	45-85%	45-90%	45-95%	30-90%	40-90%	50-90%
male secondary schooling	-0.034 (0.011)***	-0.036 (0.011)***	-0.036 (0.011)***	-0.032 (0.011)***	-0.039 (0.011)***	-0.038 (0.012)***	-0.035 (0.011)***
(GDP growth - 3*population growth) t-1	-0.079 (0.036)**	-0.084 (0.036)**	-0.085 (0.036)**	-0.083 (0.036)**	-0.085 (0.036)**	-0.089 (0.037)**	-0.084 (0.036)**
primary commodity exports/GDP	31.444 (8.422)***	31.351 (8.289)***	32.089 (8.375)***	30.546 (8.296)***	31.918 (8.285)***	30.802 (8.218)***	31.292 (8.265)***
(primary commodity exports/GDP)^2	-62.283 (19.037)***	-61.182 (18.712)***	-62.511 (18.964)***	-59.257 (18.837)***	-63.073 (18.656)***	-60.491 (18.433)***	-61.363 (18.776)***
ln population	0.842 (0.196)***	0.909 (0.201)***	0.946 (0.204)***	0.870 (0.196)***	0.951 (0.204)***	0.915 (0.200)***	0.907 (0.198)***
social fractionalization	-0.0004 (0.0001)***	-0.0003 (0.0001)***	-0.0003 (0.0001)***	-0.0003 (0.0001)***	-0.0004 (0.0001)***	-0.0004 (0.0001)***	-0.0003 (0.0001)***
ethnic dominance	-1.223 (1.076)	0.695 (0.373)**	0.847 (0.370)**	0.371 (0.554)	0.752 (0.385)**	0.736 (0.374)**	0.659 (0.373)*
geographic dispersion	-3.907 (1.186)***	-4.042 (1.190)***	-3.891 (1.172)***	-3.891 (1.172)***	-4.155 (1.182)***	-3.931 (1.170)***	-3.686 (1.150)***
peace duration	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***
N	688	688	688	688	688	688	688
No of wars	43	43	43	43	43	43	43
Pseudo R^2	0.28	0.29	0.30	0.28	0.29	0.29	0.29
Log likelihood	-115.05	-113.98	-113.03	-115.45	-113.69	-113.67	-114.14

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

Table 9A: Ethnic Dominance - Descriptive Statistics

N=688	mean	std. dev.	min	max
Largest group as a proportion of the total population	0.70	0.25	0.18	1
45-85%	0.33	0.47	0	1
45-90%	0.44	0.50	0	1
45-95%	0.77	0.42	0	1
30-90%	0.60	0.49	0	1
40-90%	0.55	0.50	0	1
50-90%	0.39	0.49	0	1

Country	year	war started	previous war
Dominican Rep.	1965	1	0
Nigeria	1965	1	0
Burundi	1970	1	0
Pakistan	1970	1	0
Sri Lanka	1970	1	0
Zimbabwe	1970	1	0
El Salvador	1975	1	0
Nicaragua	1975	1	0
Somalia	1980	1	0
Peru	1980	1	0
Romania	1985	1	0
Turkey	1990	1	0
Guatemala	1965	1	1
Myanmar/Burma	1965	1	1
Uganda	1965	1	1
Iraq	1970	1	1
Iran	1970	1	1
Guatemala	1970	1	1
Philippines	1970	1	1
Ethiopia	1970	1	1
Morocco	1975	1	1
Iran	1975	1	1
Guatemala	1975	1	1
Mozambique	1975	1	1
Angola	1975	1	1
Indonesia	1975	1	1
Nicaragua	1980	1	1
Colombia	1980	1	1
Iran	1980	1	1
Myanmar/Burma	1980	1	1
India	1980	1	1
Sri Lanka	1980	1	1
Sudan	1980	1	1
Uganda	1980	1	1
Nigeria	1980	1	1
Burundi	1985	1	1
Iraq	1985	1	1
Somalia	1985	1	1
Burundi	1990	1	1
Algeria	1990	1	1
Rwanda	1990	1	1
Zaire	1990	1	1
Zaire	1995	1	1
		43	31

Table A2: Descriptive Statistics

	sample (n=688)				no civil war (n=645)				civil war (n=43)			
	mean	st.dev.	min	max	mean	st.dev.	min	max	mean	st.dev.	min	max
War starts	0.06											
Previous war	0.36				0.33				0.72			
Male secondary schooling	44.5	31	1	147	45.6	31	1	147	27.3	21.7	3	87
GDP per capita (const. US$)	4025	4065	222	18993	4182	4137	290	18993	1673	1360	222	5832
GDP per capita Growth t-1	1.63	3.51	-13.1	14.41	1.72	3.43	-13.1	14.41	0.25	4.46	-10.66	7.09
Population growth t-1	2.09	1.19	-6.38	6.61	2.05	1.2	-6.38	6.61	2.73	0.64	0.47	4.09
Primary commodity exports/GDP	0.16	0.14	0	0.94	0.16	0.14	0	0.94	0.16	0.1	0.02	0.48
Population (millions)	30				30				40			
Ethnic fractionalization	41	29	0	93	40	29	0	93	50	30	4	90
Religious fractionalization	36	25	0	79	36	25	0	79	34	25	2	69
Geographic dispersion	0.6	0.21	0	0.97	0.6	0.21	0	0.97	0.57	0.15	0.24	0.91
Mountainous terrain	15.8	20.19	0	82.2	15.06	19.73	0	82.2	26.81	23.77	2.4	74.5
Ethnic dominance (45-90%)	0.28				0.27				0.4			
Democracy	4.26	4.31	0	10	4.43	4.34	0	10	1.92	2.9	0	9
Diaspora	0.01	0.02	0	0.16	0.01	0.02	0	0.16	0.003	0.02	0	0.16
Income Inequality (Gini)	41.35	10.04	20.97	62.3	41.14	10.0	20.97	62.3	44.33	10.29	23.38	62.3

Table A3: A simple Migration model

	Dependent Variable: $Diaspora_{it}$
$Diaspora_{i, t-1}$	1.163 (0.045)***
Ln GDP per capita $_{i, t-1}$	-0.002 (0.001)**
War dummy $_{i, t-1}$	0.003 (0.003)
N	216
R^2	0.76

The regression is based on the following model:

$dias_{it}=f(dias_{i, t-1}, lnGDP_{i, t-1}, war_{i, t-1}, T_t)$

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.

Dias $est_{it}=x_{it}*\hat{}ß$

For countries which experienced a previous civil war[23] 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. We took the averages of this corrected diaspora data measured in 1960 and 1970 (1970 and 1980, 1980 and 1990) in order to obtain values for 1965, 1975 and 1985. For 1995 we use the observations measured in 1990.

In Table A4 we present some descriptive statistics of the original diaspora data, the corrected diaspora data and the difference between the two variables.

[23] Here we only consider wars after 1960.

Table A4: Diaspora – Descriptive Statistics

N=704	mean	st.dev.	min	max
diaspora	0.0161	0.0286	0.00003	0.1948
predicted diaspora	0.0162	0.0283	0.00003	0.1948
diaspora-predicted diaspora	-0.0001	0.0043	-0.0200	0.0625

Table A5: Marginal Probabilities

Variabl	Coeff	Mean of X	median	mean	worst	best	no ethnic dom	ethnic dom	plus 10 education	plus gy1	less 2% popg1
education	-0.037	44.485	-1.460	-1.624	-0.011	-5.366	-1.624	-1.624	-1.989	-1.624	-1.624
gy1-	-0.085	-4.696	0.433	0.397	2.804	-2.857	0.397	0.397	0.397	0.226	-0.114
exports	32.089	0.158	3.594	5.054	8.022	0.000	5.054	5.054	5.054	5.054	5.054
exports2	-62.511	0.025	-0.784	-1.551	-3.907	0.000	-1.551	-1.551	-1.551	-1.551	-1.551
lnpop	0.946	17.233	14.909	16.301	19.777	11.256	16.301	16.301	16.301	16.301	16.301
fractionaliz.	-0.0003	1785.878	-0.607	-0.607	-0.004	-2.372	-0.607	-0.607	-0.607	-0.607	-0.607
ethnic dom.	0.847	0.440	0.000	0.373	0.847	0.000	0.000	0.847	0.373	0.373	0.373
geog dispersion	-3.891	0.575	-2.366	-2.238	0.000	-3.778	-2.238	-2.238	-2.238	-2.238	-2.238
peace	-0.004	347.670	-1.507	-1.488	-0.004	-2.534	-1.488	-1.488	-1.488	-1.488	-1.488
constant	-16.443		-16.443	-16.443	-16.443	-16.443	-16.443	-16.443	-16.443	-16.443	-16.443
X*betahat			-4.231	-1.826	11.081	-22.093	-2.199	-1.351	-2.191	-1.996	-2.337
probability			**0.014**	**0.139**	**1.000**	**0.000**	**0.100**	**0.206**	**0.101**	**0.120**	**0.088**

Variable	Coeff	Mean of X	Exports= 0.26	Exports= 0	min fraction.	max fraction.	min peace	5 yrs peace	10yrs peace	geog dis- persion=1	geog persion=0
education	-0.037	44.485	-1.624	-1.624	-1.624	-1.624	-1.624	-1.624	-1.624	-1.624	-1.624
gy1-	-0.085	-4.696	0.397	0.397	0.397	0.397	0.397	0.397	0.397	0.397	0.397
exports	32.089	0.158	8.343		5.054	5.054	5.054	5.054	5.054	5.054	5.054
exports2	-62.511	0.025	-4.226		-1.551	-1.551	-1.551	-1.551	-1.551	-1.551	-1.551
lnpo	0.946	17.233	16.301	16.301	16.301	16.301	16.301	16.301	16.301	16.301	16.301
fractionaliz.	-0.0003	1785.878	-0.607	-0.607	-0.004	-2.372	-0.607	-0.607	-0.607	-0.607	-0.607
ethnic dom.	0.847	0.440	0.373	0.373	0.373	0.373	0.373	0.373	0.373	0.373	0.373
geog dispersion	-3.891	0.575	-2.238	-2.238	-2.238	-2.238	-2.238	-2.238	-2.238	0.000	-3.891
peace	-0.004	347.670	-1.488	-1.488	-1.488	-1.488	-0.004	-0.257	-0.514	-1.488	-1.488
constant	-16.443		-16.443	-16.443	-16.443	-16.443	-16.443	-16.443	-16.443	-16.443	-16.443
X*betahat			-1.212	-5.329	-1.223	-3.590	-0.342	-0.595	-0.851	0.412	-3.479
probability			**0.229**	**0.005**	**0.227**	**0.027**	**0.415**	**0.356**	**0.299**	**0.602**	**0.030**

The probability of a war-start was calculated in the following way:

probability=exp(X^ ß)/(1+exp(X^ ß))