WPS4243

POST-CONFLICT TRANSITIONS WORKING PAPER NO. 16

Population Size, Concentration, and Civil War. A Geographically Disaggregated Analysis*

Håvard Hegre

Centre for the Study of Civil War, PRIO (CSCW)

Clionadh Raleigh

CSCW, PRIO & University of Colorado at Boulder

Abstract

Why do larger countries have more armed conflict? This paper surveys three sets of hypotheses forwarded in the conflict literature regarding the relationship between the size and location of population groups: Hypotheses based on pure population mass, on distances, on population concentrations, and some residual state-level characteristics. The hypotheses are tested on a new dataset – ACLED (Armed Conflict Location and Events Dataset) – which disaggregates internal conflicts into individual events. The analysis covers 14 countries in Central Africa. The conflict event data are juxtaposed with geographically disaggregated data on populations, distance to capitals, borders, and road networks. The paper develops a statistical method to analyze this type of data. The analysis confirms several of the hypotheses.

World Bank Policy Research Working Paper 4243, June 2007

The Post-Conflict Transitions Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about post-conflict development (more information about the Post-Conflict Transitions Project can be found at http://econ.worldbank.org/programs/conflict). An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in these papers are entirely those of the authors. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at http://econ.worldbank.org .

* Contact author: Håvard Hegre; CSCW, PRIO, Hausmanns gate 7, N-0187 Oslo, Norway. Email: hhegre@prio.no.

Thanks to Joachim Carlsen for writing a program to create the dataset used in the analysis, to Siri Aas Rustad for research assistance, and to Kristian Gleditsch, Anke Hoeffler, Pat Regan, Mike Ward, Nils Weidmann and Jen Ziemke for valuable comments. The research has been funded by the Research Council of Norway, grant no. 163115/V10.

1 Civil War and Country Size

The most robust empirical finding in country-level studies of civil war is that large countries more frequently have civil war than small countries (Fearon & Laitin, 2003; Collier & Hoeffler, 2004; Hegre & Sambanis, 2006). A country with a population of 10 million inhabitants has an estimated risk that is at least three times higher than one of 1 million inhabitants.[1]

As Sambanis (2003) points out, however, there is little agreement on why large countries have more civil war than small ones (2003:26). The conflict literature has suggested several competing explanations for this relationship. Are large countries more conflict-prone because of the difficulty of projecting power over long distances? Has it to do with governance problems due to the multiple layers of authority necessary in large countries? Or do conflicts stem from the cultural heterogeneity typical of large countries? Or is it simply because there are so many people that might start a bloody quarrel?

Country size, in all these studies, is measured in terms of the size of the population. But the sample of explanations above show that there are several relevant manifestations of size that are closely related at the national level: Most fundamentally, countries may be large in terms of population or in terms of territory. But countries may also be large in terms of the number of distinct cultural or ethnic groups living in the territory, or in terms of the distances over which a government must be able to deploy forces, or the length of the border. It is difficult to distinguish between these variables in empirical studies at the national level, since they are highly correlated when aggregated to this level. Previous studies have largely tested them using state-level population measures – the most notable exception is the study of Buhaug & Rød (2006). The reliance on national-level studies presents an ecological inference problem as the nature of populations and population density in particular areas is assumed to be homogeneous across a state. By disaggregating both the dependent variable of conflict occurrence and the measure of

[1]Calculated on the basis of a parameter estimate for $\ln(\text{population})$ of $.22$, as in Hegre (2003). Other studies obtain larger estimates for this variable.

population density across a state, the ecological inference issue is alleviated as we directly test the propensity of any population group to experience a conflict.

Through disaggregation, we may succeed in supporting explanations based on variables such as the distance from the capital and the overall size of the country's population if we know at which locations conflicts occur. If conflicts are located mainly at some distance from countries' capitals, we might infer that large countries have more conflicts because of the difficulties of projecting governmental power. If they are located in population concentrations irrespective of location relative to the capita, other explanations should be sought.

The paper makes use a new dataset called ACLED (Armed Conflict Location and Events Dataset) to allow for this type of disaggregated analysis. The dataset currently codes the location of all reported conflict events in 14 countries in Central Africa in the 1960–2004 period. The conflict event data are juxtaposed with geographically disaggregated data on populations, distance to capitals, borders, and road networks. The paper suggests some adaptions to a statistical method to allow for analyzing data at this level of analysis.

Related to the size of populations is their distributions. The Democratic Republic of Congo, for instance, is not only characterized by being enormously large, but also shows tremendous variation in population densities. The capital is also located in the far west of the country, and there is a large population concentration in the far east. The disaggregated research design also allows exploring how such patterns of population dispersion – and the resultant political geography of states – affect the risk of conflict events.

2 Population, Geography and Conflict: Location-Specific Factors

We will look into three groups of size-related explanations of civil war. The first simply posits that every citizen has a constant and uniform propensity for becoming involved in a rebellion. The second relates to heterogeneity and military constraints posed by distance. The third set of explanations highlights the importance of population concentrations.

We will review the arguments and formulate empirical implications for them at the appropriate level of analysis: either at the local level (for a village or a small piece of territory), or at the national level. The empirical analysis will attempt to discriminate between the explanations by testing them at the local level and see how much variance between locations is left to be explained by the national-level population size.

2.1 Population-Mass Explanations

The simplest explanation of the national-level relationship between population size and the risk and extent of conflict is based on the assumption of a constant and homogenous `per-capita conflict propensity'. If there is a given probability that a randomly picked individual starts or joins a rebellion, then the risk of rebellion increases with population. Collier & Hoeffler (2002: 11) state the `per-capita propensity' mechanism explicitly:

> Population is likely to be correlated with conflict risk. If two identical areas, each with a conflict risk of p , are treated as a single area, the conflict risk for the single area rises to $2p - p^2$. Since p is small [...], this effect alone would yield an elasticity of conflict risk with respect to population of slightly less than unity (Collier & Hoeffler, 2002: 11).

There are several mechanisms through which a constant per-capita risk of rebellion may emerge. First, a potential rebel group leader can only recruit up to a certain fraction of a population. The larger the recruitment pool, the greater the chances of recruiting a sufficiently large group to initiate a rebellion. To the extent that rebellions occur in the recruits' home locations, the risk of conflict events should be proportional to the population mass at that location. Second, even when insurgents operate away from their home locations, they are likely to target locations that can provide supplies such as food, tools, etc, and that can be taxed. The area should be `economically self-sufficient' (McColl, 1969: 618). Such supplies are, on average, richer where the population is larger. `The struggle is not over the land itself, but rather over the population concentrations' (McColl, 1969: 624). Third, a major part of a rebel group's strategy is to hurt the

government as much as possible. The group will therefore target the most valuable locations controlled by the government. The rebels will attempt to hinder the government from benefiting economically from the large tax base in a populous area, or make the government suffer the psychological and reputational loss of showing a large population that the government cannot protect their territory.

At the local level, the implication of the population-mass explanations is:

Proposition 1 ***Constant Per-Capita Risk of Conflict:*** *The risk of civil war events at a location increases with the size of the population at the location but, controlling for the local effect, is not affected by the size of the population of the country to which the location belongs.*

An interesting implication of this hypothesis is that a continent with few countries would have the same number of conflict events as a similarly sized continent with many countries. The `per-capita propensity' explanation has precisely these implications. At the national level, it implies that the risk of conflict in a country is exactly proportional to the size of its total population.

2.2 Distance, Transportation Costs, and Borders

The probability of conflict events at a location is likely to be dependent on where the location is situated relative to the capital, a rebel group's headquarters, and to international border. Two aspects are particularly important: The relationship between relative locations and preferences, and military factors. We will discuss these in turn.

2.2.1 Heterogeneity of preferences

Alesina & Spolaore (2003: 40–45) develop a model based on the assumption that there are economies of scale in the production of public goods and that the utility individuals derive from the public good decrease with distance. Distance is conceived of in terms of preferences and of physical distance. Cultural and economic background factors such as religion, ethnic affiliation, or dominant occupation tend to be geographically clustered. These background factors determine at least a part of an individuals' preferences. Hence,

geographical distance between two populations is correlated with distance in terms of preferences regarding public policies. Since populous countries are often geographically extensive, they are also likely to be heterogeneous. Alesina & Spolaore argue that economies of scale allow large countries to provide better public goods, but at a given distance from the center of the country, the distance in terms of preferences outweighs the efficiency of the large government. Beyond this distance, population groups will have an incentive to secede. Although not mentioned explicitly by Alesina & Spolaore, such secession attempts may turn into violent conflicts.

Geographic peripherality is often linked with ethnic and political peripherality (see for instance Gurr, 1970). If civil wars are caused by differences over public policies, we may infer that insurgencies should originate in locations far from the capital. Even when the conflict is purely over political power per se, rebel groups are likely to attempt to exploit a local population's resentment toward the government to win their `hearts and minds'.

2.2.2 Military factors

The dispersion of population also has military implications. Although there clearly are economies of scale in defense, (Collier & Hoeffler, 2002: 15; 2004: 572) these are clearly counteracted by the challenge of controlling large territories. States with limited reach may not be able to control activities in territories beyond the established infrastructure of the state (Gurr, 1970; Herbst, 2000; Clapham, 1985). Warfare becomes more difficult the further from the government's main base the front is: Transportation takes longer time and requires better organization, supply lines become more vulnerable to guerrilla attacks, and the local population may become more hostile the further from the capital they are located. Even though a government may establish military bases throughout the territory to minimize the impact of distance, these bases are supported by vulnerable supply lines, and the bases themselves become targets for rebel group activities. Locations far from the capital, then, may be conflict-prone both because the population is likely to have preferences that diverge markedly from the government and because it is difficult for the governments to control distant territories.[2]

[2]If the rebel group is sufficiently strong, it will be able to push the frontline to the capital. In that case, we

Lichbach (1995:156) details the role of geographic isolation for dissident communities as one that results from distance, poor transportation, inaccessible terrain, and fluid boundaries. In addition, he states that it should follow that as distance from national authorities' increases, the collective dissent should increase – and states with poor transportation networks should experience higher levels of dissent. Relatedly, Herbst (2000) describes the political geography of a small country as `favorable' to a government.

The heterogeneity and military explanations have the same empirical implication:

Proposition 2 ***Distance from Capital:*** *The risk of civil war events at a location increases with the distance from the location to the capital of the country.*

Related to this aspect of geographical distribution is the importance of national borders. Rebel groups may operate more easily in border areas since neighboring countries may provide (actively or tacitly allow) safe zones for rebels. Again, it is impossible to distinguish fully between military and cultural factors. Conflicts which begin in areas proximate to borders may also be linked to irredentist movements in defiance of, for instance, the state's ethnic character. Also secessionist movements that are not inspired by neighboring governments are more likely to rise in border areas, since the prospective new state can avoid being an enclave of the former mother country.

Proposition 3 ***Distance from Border:*** *The risk of civil war events at a location is higher in border zones.*

A third location aspect that impacts where conflicts take place is the location of rebel group headquarters. Even though rebel groups often rely on hit-and-run strategies, they are as constrained by logistics as government forces and find it difficult to operate

---

might expect to see a large number of events close to the capital. Simulations reported in Weidmann, Hegre & Raleigh (2006) indicate, however, that in these cases the rebel group tends to be so strong relative to the government that the war ends quickly, with few observable events.

far from their headquarters. Locations far from rebel group headquarters should therefore experience fewer conflict events. In fact, rebel groups are likely to be more constrained by distance than governments since governments typically have had the opportunity to establish military bases throughout the territory before the conflict started.

Proposition 4 ***Distance from Rebel Group Headquarters:*** *The risk of civil war events at a location decreases with the distance from the location to the rebel group's headquarters.*

2.3 Population Concentration

A third set of explanations relate to population concentrations: The conflict-proneness of a local population may be larger the more concentrated is the population in the neighborhood. In addition to the size or mass of the population in the location, the size of the population in the immediate neighborhood is important. There are at least three reasons why this might be so: First, population concentration helps solve coordination problems, second, population concentrations may be more homogenous, and third, concentrated populations are more autonomous than dispersed ones.

Prospective rebel groups face trade-offs between economies of scale and geographical extent just as governments do. Lichbach (1995) and Collier (2000) stress the importance of the collective action and coordination problems prospective rebels face. Accordingly, Lichbach (1995) contends that as the geographic concentration of dissidents increases, collective dissent should also increase. This is directly due to the ability of dissidents to communicate, coordinate mutual expectations, and reduce organizational costs:

As the concentration of dissidents increases, the extent and intensity of interactions among dissidents increases, which in turn increases their communications (e.g. of grievances). [...] [W]ith reduced distance between dissidents, it is easier to administer rewards for compliance and punishment for noncompliance (Lichbach, 1995: 158–159).

In the absence of community, or rebels socialized into common norms (1995: 126), Lichbach argues that rebels will employ a `contract' as one way to overcome the `rebel's dilemma'. Communities which are autonomous, stable and concentrated can forge a

number of different contracts to assure collective dissent. Of those types of contracts, homogeneity in social background allows for lower transaction costs and hence increased cooperation for joint collective dissent – `homogeneity, moreover, facilitates the development of information, trust and norms, and hence reduces the bargaining, monitoring, and enforcement costs of social contracts' (p. 139). As noted above, homogeneity in social background is also a function of geographical concentration. Heterogeneity is likely to vary more with geographical distance than with population size. This serves to increase the importance of local population concentrations.

Herbst (2000:152–154) argue that `hinterland' countries which are large but where most of the population is concentrated in a small number of areas have a relatively low conflict propensity. In the hinterland, population is density is low, and governments can relatively easily control it. Moreover, in such countries, Herbst contends that most political battles will occur in capital areas since this is where the important stakes are located.

Densely populated areas are often dense because they are resource-rich at the outset or are strategically located, e.g. close to a major harbor or a navigable river. Moreover, economies of scale often help such locations to become even richer. Population concentrations then mark locations that are particularly valuable both for the rebels and for the government, and should attract conflict events if hostilities break out.

Yet concentration can also work against a movement as nationally based dissident movements, or small movements in large countries, are more prone to failure because of their inability to permeate the remainder of the state. Lichbach (1995: 160) also notes that a `wide geographic scope can work to the advantage of a dissident movement as it works against the government's ability to repress'. The perceived increase of collective dissent in urban areas is related to the concentration argument as the ability of dissidents to organize is lessened due to proximity. The fewer urban areas in which dissents may gather, the higher the collective dissent (Tilly, 1964 as cited in Lichbach, 1995:162).

Proposition 5 ***Population Concentration:*** *The risk of civil war events at a location increases in the size of population in the immediate geographical neighborhood.*

2.4 Population Concentration and Distance-Related Factors

Distance from the capital and population concentration are factors that are likely to reinforce each other with respect to risk of conflict events.

Political geography in part refers to the ability of governments to penetrate the state. A number of political geographies are at odds with traditional understandings of sovereignty. Herbst (2000) describe political geographies with large territories and non-contiguous areas of high population as the most difficult to control for a government, and Lichbach (1995) argues that the combination of large distances to the capital and population concentrations particularly facilitate dissent. The population geography in such countries can, in some cases, be associated with ethnic complexity, as 70% of African ethnic communities live in spatially distinct ethnic `homelands' (Scarritt & McMillan, 1995). Further, such population patterns are at odds with typical understandings of territorial sovereignty – these states are the most succeptible to fracturing as an uncontrolled population can choose to deny the legitimacy of governments (Herbst, 2000: 146).

A few empirical studies find evidence that concentration and distance reinforce each other. Toft (2003) investigates whether different settlement types lead to increased motives for separatist conflict by employing the MAR ethnic concentrations (urban, concentrated majority, concentrated minority and dispersed). Each group is found to have a different capability of armed rebellion. Urban groups as most able to create networks mobilize the populous and dominate necessary resources; concentrated majorities have similar capabilities. For concentrated minorities, the capabilities are deemed indeterminate and dependent on the context and region. Dispersed minorities are the weakest in terms of ability to create conditions suitable for separatist conflict.

Collier & Hoeffler (2004) test Herbst's hypothesis in their national-level study by calculating a gini coefficient for population distribution. Countries in which the population is evenly distributed throughout the territory will have a score of 0, whereas a country where all the population is concentrated in one of their 20x20km squares will have a score of 1. They consistently obtain a negative estimate for this variable, controlling for other factors, and conclude that population concentration reduces the risk of conflict. Their measure is not likely to capture all aspects of population diffusion,

however, since they cannot distinguish between a country where the population is concentrated in one cluster covering 10% of the territory and one where the population is concentrated in two clusters of 5% each, but with a considerable geographical distance between them.

Population and conflict geography in the Democratic Republic of Congo (DRC) corresponds to these arguments regarding population concentration and dispersion. Concentrations of language-based minorities are evident throughout eastern DRC. Due to the limited access of the government, the close proximity to international borders, and the dense population concentrations, these concentrated minorities have a higher potential of conflict than other, more accessible, sparsely populated areas of DRC. Figure 1 shows the population concentrations in 1990 (CIESIN data) for Central Africa. Heavily populated areas are shaded in deeper tones of red/grey. Civil conflict in DRC has overwhelmingly occurred in the eastern portion of the state, which is the most densely populated area and also geographically peripheral to the capital, Kinshasa. Of the eleven Congolese rebel groups accounted for in the dataset used in this paper, all have operated either exclusively or partially in the eastern and southern areas of DRC.

Proposition 6 ***Concentration and Dispersion:*** *The risk of civil war events at a location increases more strongly in local population concentrations in locations distant from the capital of countries.*

2.5 Residual State-Level Mechanisms

We have pointed out a set of location-specific factors, each of which imply a positive relationship between country size and national-level risk of armed conflict. But it is not certain that such location-specific factors are the only relevant ones. The size of a country itself may affect risk over and beyond what is implied by sheer population size, distance, and population distributions. If the economies of scale with respect to defense are sufficiently large, the risk of conflict events at a location at a given distance from the capital may be lower the larger is the country (Collier & Hoeffler, 2002). Moreover, large countries may rather be *more* conflictual than small ones for several reasons. Fearon & Laitin (2003: 81), for instance, note that insurgency will be favored when potential rebels

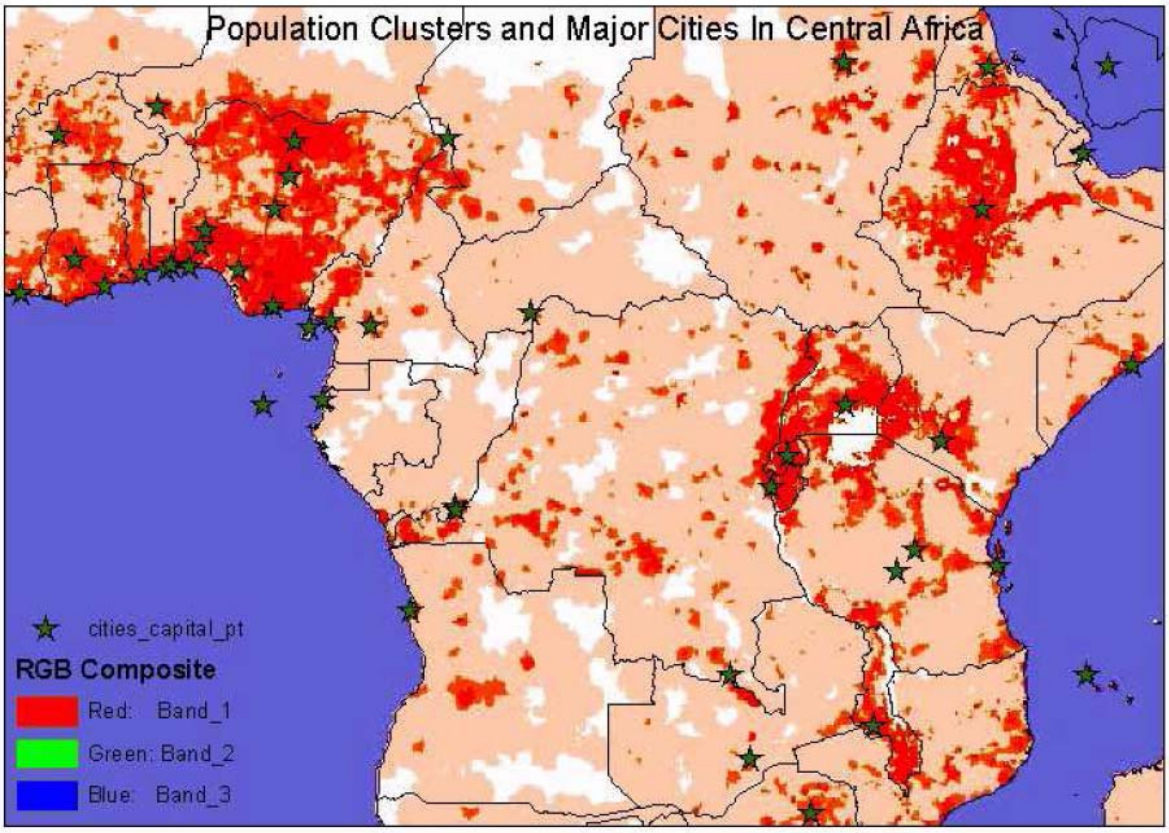


Figure 1: Population concentrations and major cities in Central Africa, 1990.

face a `large country population, which makes it necessary for the center to multiply layers of agents to keep tabs on who is doing what at the local level'. Furthermore, the economies of scale in defense does not unambiguously decrease the risk of conflict, since they also raise the stakes of the political contest. Since controlling the government in a large, wealthy, and powerful country is more attractive to would-be heads of states, they will be more willing to initiate an insurgency to take control over the government, *ceteris paribus.*

We consequently open up for the possibility that the risk of conflict events at a location depends on the size of the country it is located within:

Proposition 7 ***State-Level Effects of Population Size:*** *The risk of civil war events at a location varies with the size of the population of the country to which the location belongs, controlling for the local effects.*

3 Research Design

3.1 Unit of Analysis

To distinguish between the different theoretical statements regarding how population sizes, population concentrations and locations relate to risk of conflict, we need to investigate exactly where conflicts occur. We have created a dataset using a Geographic Information Systems (GIS) program which converted large territories into smaller portions of 8.6 km x 8.6 km, totaling 74 square kilometers. Each of these grid squares are our units of observation (we will refer to them as squares). This approach is similar to that of Buhaug & Rød (2006), with two important differences. First, their squares are much larger (100x100km). Second, they code the dependent variable considerably more crudely than is done in the ACLED dataset described below. Buhaug & Rød (2006) use the `scope' and `location' variables in the Uppsala/PRIO dataset. Although much more suited to geographically disaggregated analysis than other datasets, this location dataset has some limititations, It does not record changes over time in the center location and extent of conflicts, and it reports the total extent of the conflict zone without distinguishing between areas that saw repeated and extensive fighting and those that only experienced scattered activities or individual events far from the center of the conflict.

3.2 Disaggregated Dependent Variable: ACLED

The ACLED dataset (Raleigh & Hegre, 2005) deals with these problems. The dataset takes the PRIO/Uppsala Armed Conflicts Dataset as its point of departure. The dataset is limited to events within conflicts that fall within the Uppsala conflict definition; conflicts involving two parties, one of which is a government, and fighting resulting in at least 25 battle deaths.[3] ACLED is designed to parse out both the temporal and spatial actions of rebels and governments within civil wars.

[3]See the PRIO/Uppsala Armed Conflict Data codebook for more information (Strand, Wilhelmsen & Gleditsch, 2004).

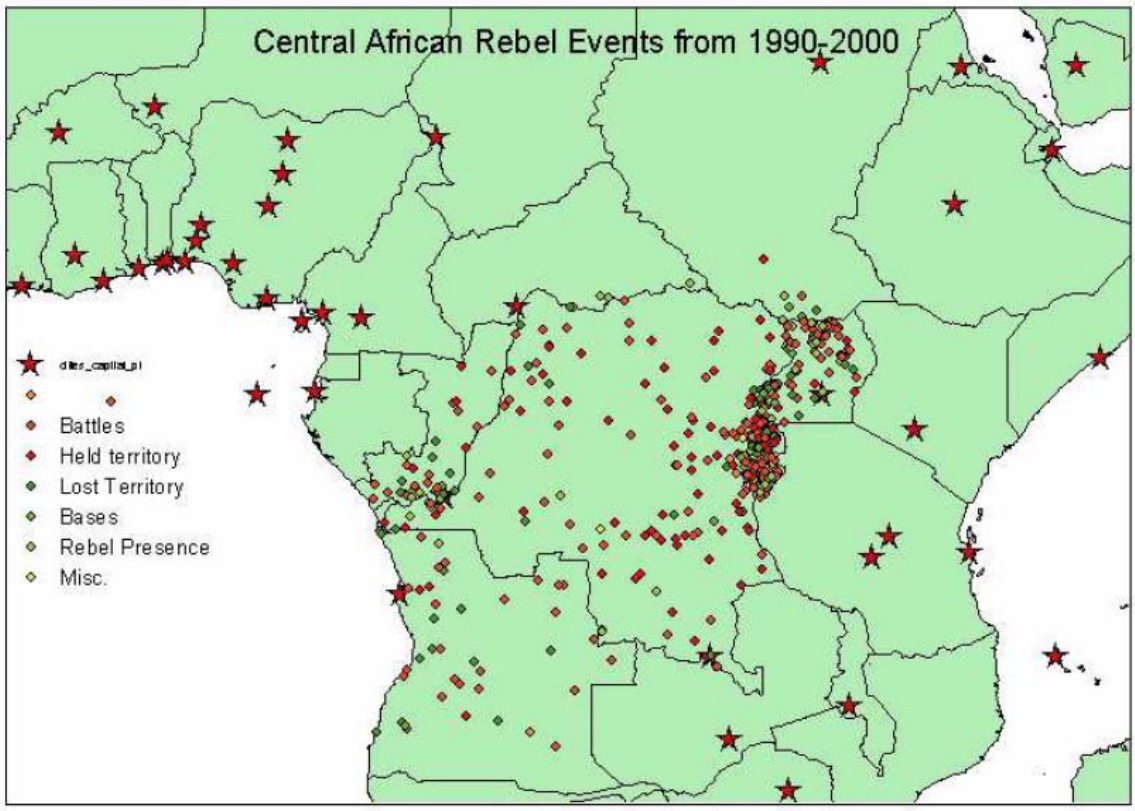


Figure 2: Events in ACLED (Armed Conflict Location and Event Data), 1990s, Central Africa

The fundamental unit of observation in ACLED is the event. Figure 2 illustrates the ACLED data for Central Africa for the 1980s and the 1990s. Each location of a conflict event is represented by a symbol. In several of these locations, multiple events occured over the periods. Events always involve two actors – a rebel group and a government – and are coded to occur at a specific point location and on a specific day. Most of the events are battles, but the dataset also records other activities. The dataset includes information on and distinguishes between six types of events: battles resulting in no change of territory, battles resulting in a transfer of territory to the rebel actor, battles

resulting in government forces recapturing rebel held territory, establishment of a rebel base or headquarters, rebel activity that is not battle related (e.g. presence or the killing of civilians), and territorial transfers. The dataset consists of 4,145 battle events for the 1960–2004 period. In the present analysis, we use 2,530 of these. The remaining events were dropped as they either were in countries not included in the analysis, or because information was missing for one of the key variables. Each conflict event is associated with geographic coordinates and a date of occurrence. This information allows for spatial and temporal modeling of conflict events.

The dataset used in this article covers 14 countries in Central Africa. 6 of them had a conflict in the 1960–2004 period according to the Uppsala/PRIO Armed Conflict Dataset (Gleditsch et al., 2002): Angola, Burundi, Republic of Congo (Brazzaville), Democratic Republic of Congo (Zaire), Rwanda, and Uganda. The remaining 8 did not have a conflict: Cameroon, Central African Republic, Equatorial Guinea, Kenya, Malawi, Tanzania, and Zambia.

3.3 Handling temporal and spatial dependence

Both the squares and the conflicts events are obviously not fully independent – all events within one conflict are related to each other as an action in one location leads to a later retaliation by the opposing party or to further advances in proximate locations. Events in one conflict may also affect the likelihood of other conflicts, such as the spillover of the conflict in Rwanda into Eastern DRC.

The statistical model employed to analyze these data must handle the dependence between observations. We will do this by explicitly modeling the probability of an event in a location as a function of preceding events in the same and in adjacent squares. We can do this since we know both the precise date and the precise geographic location of each event.

We use an adaption of the calendar-time Cox regression model presented in Raknerud & Hegre (1997) for this purpose. In Cox regression, the dependent variable is the transition between `states of nature' -- the transition from peace to conflict in a square. A central concept is the *hazard function,* $\lambda(t)$, which is closely related to the concept of transition probability: $\lambda(t)\Delta t$ is approximately the probability of a transition in the

`small' time interval $(t, \Delta t)$ *given* that the subject under study is at risk of transition at t. The main idea of Cox regression is the assumption that the hazard of war $\lambda_d(t)$ for square d can be factorized into a parametric function of (time-dependent) variables and a non-parametric function of time itself (the baseline hazard):

$$\lambda_d(t) = \alpha(t)\exp\left(\sum_{j=1}^{p} \beta_j X_j^d(t)\right) \qquad (1)$$

In (1), $\alpha(t)$ is the baseline hazard: an arbitrary function of calendar time reflecting unobserved variables at the system level. $X_j^d(t)$ is a (possibly time-dependent) explanatory variable for square d ; β_j is the corresponding regression coefficient; and p is the number of explanatory variables. All legitimate explanatory variables are known prior to t – they must be a part of the history up until immediately before t.

In contrast to ordinary survival analysis, t is *calendar time* here. The model is useful because it allows handling observations that are recorded on the finest possible time-scale to keep track of the succession of events. It also allows for non-stationarity in the underlying baseline probability of conflict events due to changes in latent variables at the system level. Such non-stationarity may be due to several causes: the end of the Cold war, changes over time in the prices and availability of arms, changes over time in the reporting of conflict events in Western media, etc.

Estimating this model involves (i) estimation of the regression coefficients β_j and (ii) estimation of the baseline hazard of war $\alpha(t)$. These two tasks are quite different, since the latter is an unknown function – not a parameter. However, for the specific purpose of inference about conflict, we are mainly interested in the `structural' parameters β . Inferences about β can efficiently be made by conditioning on the time-points of outbreaks of war, $\{t_1, t_2, ..., t_n\}$. This means that we can consider $\{t_1, t_2, ..., t_n\}$ as fixed rather than stochastic, without losing any information about the parameters.

Given that there is an outbreak of war at time t_w , the probability that this war outbreak will happen in square d is:

$$\Pr(\text{war in a square } d \mid \text{a war breaks out at } t_w) = \frac{\exp\left(\sum_{j=1}^{p} \beta_j X_j^d(t)\right)}{\sum_{i \in R_{t_w}} \exp\left(\sum_{j=1}^{p} \beta_j X_j^d(t)\right)} \quad (2)$$

where R_{t_w} is the *risk set* at t_w : the set of squares that are at peace immediately before t_w . The parameters can be interpreted in terms of a *relative* probability of war.

To perform an analysis with this model, we need a data file constructed in the following way: For each t_w – i.e. each day a square war breaks out somewhere -- we take a `snapshot' of the system; we note for all squares the values of the explanatory variables at that particular day. As is seen from expression (2), the square that *did* have an event at t_w is compared to all squares that were at risk of doing so. Thus, all information for the time *between* different t_w 's is ignored. From the combined information about all outbreaks in the period under study, we can estimate the hazard function (1).

A dataset comparing all the 100,000 squares 2,530 times would be forbiddingly large – approximately 250,000,000 lines. It is therefore necessary to analyze a sample of the observations. The observations of positive events contain more information than the non-event observations We therefore sample asymmetrically: We sample all of the transition events and 1.0% of the non-transition events.

3.4 Disaggregated Independent Variables

Local level data on land, population, and elevation is available in the geospatial format of raster files with a resolution of 1km. Using Geographic Information systems (GIS), attributes from raster and point data are associated with the grid square in which they lie. In this way, spatial data is georeferenced to a location that is defined by the grid cell. This process results in a data structure in which each row has within it combined information on a square defined by the grid, the national level information in which is it located, and

the local data on physical geography and population from the raster data. These data can then be imported into statistical programs for analysis.

We aggregate all data up to a grid of 8.6x8.6km squares. Each grid square is assigned attributes of the country it is in along with information from data disaggregated to the level of the individual squares. Figure 3 illustrates this grid as a fictive country somewhat smaller than the average size in our dataset (50x50 squares, or 430x430 km) with a fairly representative but stylized population distribution. The country has three major cities, one of which is the capital, and two smaller ones. A rebel group has its headquarters at the Eastern border.

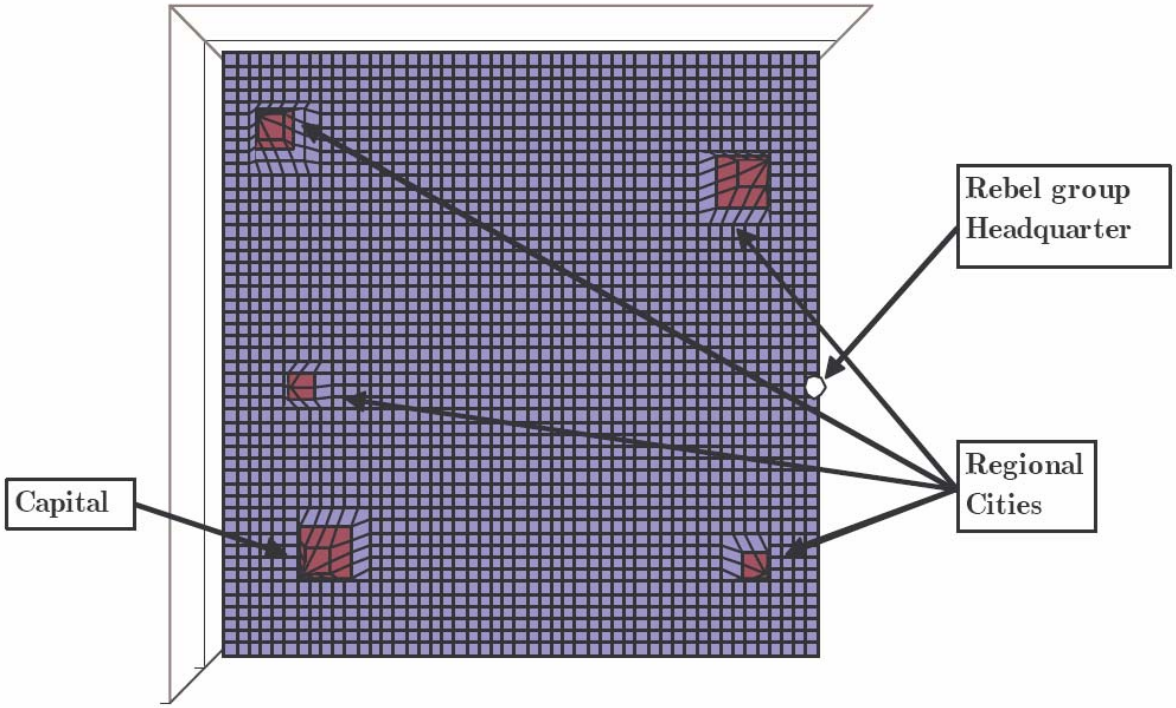


Figure 3: Map of hypothetical country, with locations of capital, rebel group headquarters, and main population concentrations.

The ACLED data for the Central African conflicts were aggregated up to the 8.6x8.6km squares and merged with information on other explanatory variables aggregated to the same level. Squares in which conflict events (not shown in Figure 3)

are located will be coded as conflict squares. The 14 countries in the dataset cover 8 million square kilometers or just over 100,000 squares.

Log population in square

High resolution population data is available through the UN geodata portal. We use the fourth version of population data from a joint UNDP and CIESIN project. The database population count assessments are from a compilation of existing data sources. The sources of error for population counts and distribution are admittedly substantial, particularly in developing areas. In this version, population figures are transformed into a distributive surface to use in spatial analysis. A rasterized (gridded) format is based on interpolating `accessibility weights' by administrative units and assigning population totals based on the assumption that population distribution and densities in Africa are strongly correlated with accessibility. The accessibility index is sum of population totals of the towns in the vicinity weighted by distance. Each raster unit is assigned a population based on interpolating population densities from the accessibility weight and a distance decay function for surrounding areas. Adjustments to the accessibility grid were made for inland water bodies, elevation and protected areas. Although by no means a perfect account of local population, the data remains the most comprehensive and sophisticated spatially referenced population data available. We use estimated total population figures, observed once for every decade since 1960.[4]

In original form, the population count has a resolution of 1km x 1km. For this paper, the population counts were aggregated to the 8.6 x 8.6 km grid square. For observations in the period 1.1.1960–31.12.1964 we used the population figure for 1960. For observations in the period 1.1.1965–31.12.1974 we used the population figure for 1970, etc.

Since the area of each square is identical, the variable also indicates the local population density. The variable was log-transformed in the analysis reported below. 0.5 were added to all observations to avoid non-defined transformations.

[4]The UNDP/CIESIN data are available by continent at http://grid2.cr.usgs.gov/globalpop/Africa/part2.html.

Log population in neighboring squares

To test hypotheses concerning population concentrations in the immediate neighborhood of the squares, information is coded in three variables. The first `ln(population in neighborhood 1st order)' is the sum of the population sizes in the eight immediately contiguous squares. The second is the 2nd order neighborhood population – the sum of the populations in the 16 squares contiguous to these again. The third is the sum of the populations in the 24 3rd order neighbors. The variables were thereafter log-transformed.

Population growth in square

Population growth data was derived from the `log population in square' variable. We coded local population growth as the difference in square population from one period to the next. For observations in the period 1.1.1960–31.12.1964 we coded population growth as missing. For observations in the period 1.1.1965–31.12.1974 we used log population for 1970 minus log population for 1960, etc.

Log population in country

To allow distinguishing between the local-level and national-level mechanisms, we added information for the total population of the country, log-transformed. The data were taken from the *Times Concise Atlas of the World* (2000).

Log area of country

We also added information on the total extent of the country in log square kilometers. The data were taken from the *Times Concise Atlas of the World* (2000).

Distance from Capital

To test hypotheses regarding the distribution of populations, we coded the distance from each square to the capital of the country. The variable was coded as the distance in terms of squares and log-transformed.

Distance from Rebel Group Headquarters

We coded the location of the headquarters of the rebel groups participating in the conflicts under study, and calculated the distance from each square to the most proximate rebel group headquarters (we do not know *a priori* which rebel group or government that will act in a particular square). As the `distance from capital' variable, it was coded as the distance in terms of squares and log-transformed.

Border Square

We coded squares as border squares if a national border runs through it. Such squares belong to more than one country and are not straightforward to code. We coded national-level information for border squares according to the following rule: A border square was considered to belong to the country that was most frequent among the eight neighboring squares. In tie cases, we assigned nationality randomly between the tied countries.

Interaction country-square population

This variable was created to test the population settlement pattern hypothesis. It is an interaction between population count at a location (square) as a portion of the country's total population.

Road type

Road type is a variable by ESRI that is available in the Digitial Chart of the World Data. It is a high resolution dataset at 1:1,000,000 scale and consists of arcs which indicate road mass. A number of different road indicators are available and we choose road line type to use in the analysis. Road type is defined by the following: The reference category (0) points out squares with dual lane/divided highways, other primary roads, or road connectors within urban areas (types 1 or 8 in the ESRI dataset). The second category include secondary roads (type 2), and the third combines squares with informal or tertiary roads (tracks, trails or footpaths) or no road registered at all (types 3 and 0, respectively, in the ESRI dataset). Figure 4 overlays the types of roads in the original dataset before our recategorization. The shaded area represents the portion of Africa for which we code

information for the explanatory variables.[5]

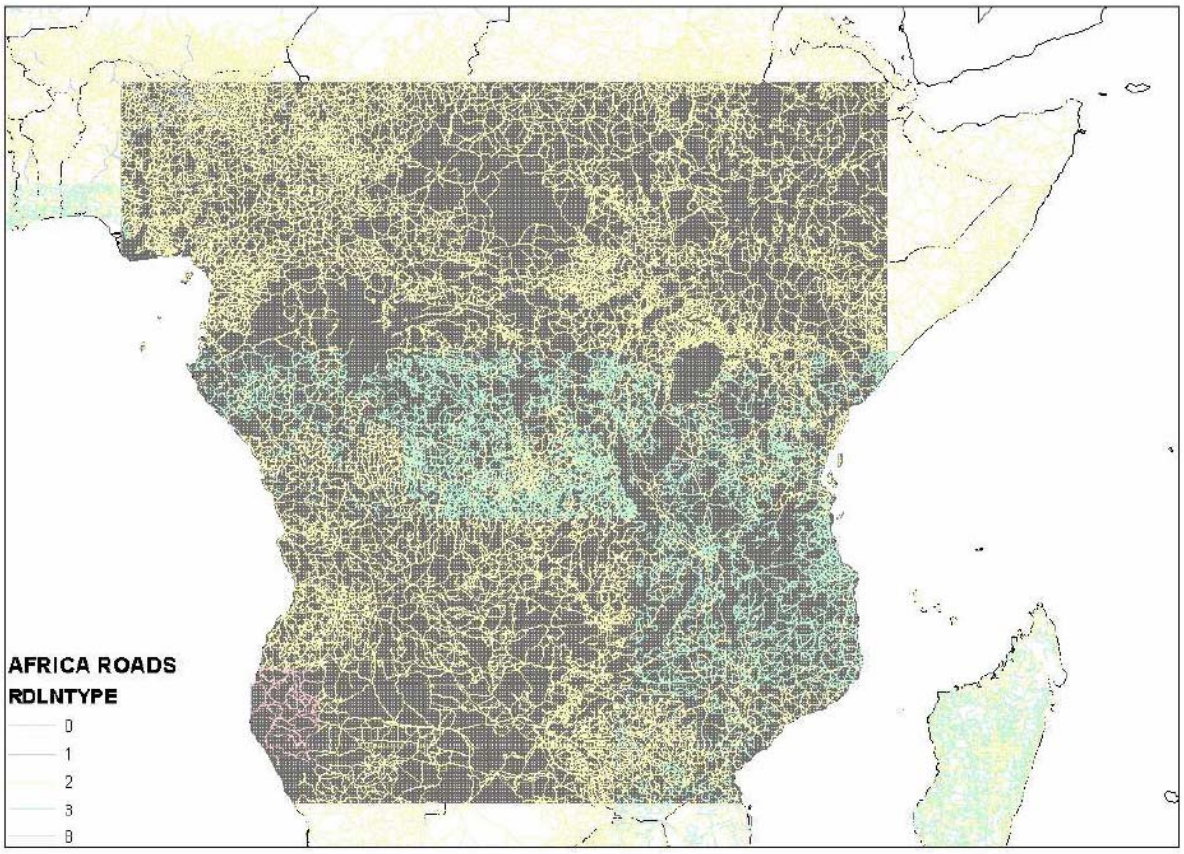


Figure 4: Road network, Central Africa.

3.4.1 Model of Temporal and Spatial Dependence

We coded three variables to account for temporal and spatial dependence:

Proximity of event in square

A fundamental dependence is the dependence between events and previous events in the same location. We calculated the number of days since a conflict event happened in the same location, analogous to the `peace years' variable in country-year setups (e.g. Beck, Katz & Tucker, 1998). As in Raknerud & Hegre (1997), we assume that the effect of the

[5]Further information on the road measure can be obtained at http://atlas.geo.cornell.edu

previous event decreases at a constant rate, and compute a decay function with a half-life of α : $pes = 2^{(-days/\alpha)}$. The variable is called `Proximity of event in square'.

We estimated a set of models with different values for α , corresponding half-lives of $1/4, 1/2, 1, 2,$ and 4 years. $\alpha = 1430$ days or 4 years yielded the highest log likelihood, so we estimated all the models reported below with that as the half-life parameter. We expect a positive parameter estimate, as events are more likely in locations where conflict has already started than in other locations.

Proximity of event in neighboring square (1st and 2nd order)

Events in a location are dependent not only on previous events in the same locations, but also on previous events in other locations. Events in the most proximate locations are presumably the most important. We calculated the number of days since a conflict event happened in 1) first-order neighborhoods – locations immediately adjacent to the unit of observation, and 2) the number of days since conflict in second-order neighborhoods – the squares adjacent to these again. We calculated the decay function with 1430 days as the half-life parameter for both. We refer to these variables as Proximity of event in neighboring square (1st) and Proximity of event in neighboring square (2nd), respectively.

We expect positive parameter estimates for the `Proximity of event in neighboring square (1st, 2nd)' variables, as events are more likely in locations close to where conflict has started than in other locations.

Distance to most recent previous event

The `Proximity of event in neighboring square' variable cannot capture the extent to which events are dependent on geographically more distant events – i.e. events outside the immediate neighborhood. To capture these relationships we calculate the distance (in square units, e.g. 8.6 km) from the unit of observation to the most recent event in the dataset. We log-transform the variable, and expect a negative relationship between the variable and the risk of observing events: Events are most likely to be followed by events in proximate squares, so the risk decreases with distance from the most recent event.

4 Results

4.1 Testing the Hypotheses

In Table 1, we present results for three models omitting information on the locations of Testing the Hypotheses rebel group headquarters. The Table presents three models that vary in terms of how we model population concentrations. In Table 2 we include the headquarters variable for the same three models.[6] We will refer to the models in Table 1 as models A1, A2, and A3, and those in Table 2 as B1, B2, and B3.

In models A1 and B1, the estimates for the `population in square' variable are positive and significant 1 conflict events tend to occur more often in squares that are relatively populous, as posited in Hypothesis 1. The estimate implies that increasing populations in squares by a factor of 2.7 increases the risk of conflict events by 12–13%.

Hypothesis 2 is also clearly supported in Models A1 and B1. Increasing distance from the capital by a factor of 2.7 increases the risk of conflict events by 8–10%.[7] Hypothesis 3 receives strong support from the analyses as well. Controlling for other variables, including the distance from capital, border squares are more than three times as likely to have conflict events as other squares.[8]

Models A2 and B2 allow for investigating the effect of population concentrations. The three `population in neighborhood' variables are significant both individually and jointly in both models. Hypothesis 5 is clearly supported.

In Table 2, the `distance from rebel group headquarters' variable is also included. In Model B1, the estimate for this variable is $-0.107$ – roughly the equivalent to the estimate for the `distance to capital' variable. Conflict events are most frequent close to rebel group headquarters, as stated in Hypothesis 4.

---

[6] Note that the estimates in Tables TableResults1 and TableResults1b are very similar. The inclusion of the `distance to rebel group headquarter' does not affect the results much apart from reducing the log likelihood with more than 5 points.

[7] The estimate may also be interpreted as an elasticity: Increasing distance from the capital by one percent increases risk of conflict events by just less than 0.1 percent relative to the baseline.

[8] $\exp(1.15) = 3.16$.

	Model A1	Model A2	Model A3
Variable	$\hat{\beta}$ $(s.e.)$	$\hat{\beta}$ $(s.e.)$	$\hat{\beta}$ $(s.e.)$
ln(population in square)	0.132^{***} (0.014)	-0.157^{***} (0.038)	-0.215^{***} (0.082)
ln(population in neighborhood)		0.158^{**} (0.081)	0.784^{***} (0.19)
ln(population in neighborhood 2nd order)		0.333^{***} (0.098)	-1.58^{***} (0.26)
ln(population in neighborhood 3rd order)		-0.123^{**} (0.064)	1.87^{***} (0.17)
ln(pop. in square) *ln(distance to capital)			0.0257 (0.026)
ln(pop. in neighborhood) *ln(distance to capital)			-0.197^{***} (0.056)
ln(pop. in neighb. 2nd order) *ln(distance to capital)			0.572^{***} (0.077)
ln(pop. in neighb. 3rd order) *ln(distance to capital)			-0.570^{***} (0.048)
ln(population in country)	0.0493 (0.0343)	0.0492 (0.034)	0.123 (0.039)
ln(distance to capital)	0.0834^{***} 0.024	0.143^{***} (0.024)	2.18^{***} (0.18)
Border square	1.13^{***} (0.072)	1.10^{***} (0.074)	1.14^{***} (0.078)
Road type 0: Primary roads	$ref.cat.$	$ref.cat.$	$ref.cat.$
Road type 1: Secondary roads	-0.628^{***} (0.062)	-0.750^{***} (0.063)	-0.708^{***} (0.067)
Road type 2: Tertiary or no roads	-1.39^{***} (0.081)	-1.64^{***} (0.084)	-1.74^{***} (0.086)
Proximity of previous event in square	6.11^{***} (0.077)	6.18^{***} (0.076)	6.11^{***} (0.078)
Proximity of previous event in neighborhood (1st order)	-0.904^{***} (0.071)	-0.814^{***} (0.073)	-0.619^{***} (0.077)
Proximity of previous event in neighborhood (2nd order)	-0.0097 (0.069)	-0.0054 (0.071)	-0.176 (0.075)
ln distance to previous event	-0.494^{***} (-0.019)	-0.494^{***} (-0.019)	-0.484^{***} (-0.019)
No of. failures	2291	2291	2291
ll null model	-30571.32	-30571.32	-30571.32
ll full model	-22188.49	-22127.27	-22019.07
$^{*}: p < 0.10$, $^{**}: p < 0.05$, $^{***}: p < 0.01$,			

Table 1: Cox Regression Estimation Results, Risk of Conflict Event, Population Size and Concentration Models, 1960–2004

	Model B1	Model B2	Model B3
Variable	$\hat{\beta}$ $(s.e.)$	$\hat{\beta}$ $(s.e.)$	$\hat{\beta}$ $(s.e.)$
ln(population in square)	0.122^{***} (0.015)	-0.179^{***} (0.038)	-0.280^{***} (0.083)
ln(population in neighborhood)		0.192^{**} (0.082)	0.945^{***} (0.19)
ln(population in neighborhood 2nd order)		0.326^{***} (0.099)	-1.80^{***} (0.27)
ln(population in neighborhood 3rd order)		-0.141^{**} (0.065)	2.00^{***} (0.18)
ln(pop. in square) *ln(distance to capital)			0.0378 (0.026)
ln(pop. in neighborhood) *ln(distance to capital)			-0.236^{***} (0.057)
ln(pop. in neighb. 2nd order) *ln(distance to capital)			0.635^{***} (0.079)
ln(pop. in neighb. 3rd order) *ln(distance to capital)			-0.614^{***} (0.0494)
ln(population in country)	0.0756^{**} (0.0353)	0.0789^{**} (0.036)	0.177 (0.041)
ln(distance to capital)	0.0952^{***} 0.024	0.156^{***} (0.024)	2.29^{***} (0.178)
ln(distance to rebel group headquarter)	-0.107^{***} 0.034	-0.115^{***} (0.034)	-0.173^{***} (0.035)
Border square	1.15^{***} (0.072)	1.11^{***} (0.074)	1.14^{***} (0.079)
Road type 0: Primary roads	$ref.cat.$	$ref.cat.$	$ref.cat.$
Road type 1: Secondary roads	-0.650^{***} (0.062)	-0.770^{***} (0.063)	-0.737^{***} (0.067)
Road type 2: Tertiary or no roads	-1.41^{***} (0.081)	-1.65^{***} (0.084)	-1.76^{***} (0.087)
Proximity of previous event in square	6.05^{***} (0.077)	6.12^{***} (0.078)	6.02^{***} (0.079)
Proximity of previous event in neighborhood (1st order)	-0.973^{***} (0.073)	-0.898^{***} (0.077)	-0.738^{***} (0.081)
Proximity of previous event in neighborhood (2nd order)	-0.0530 (0.071)	-0.0953 (0.072)	-0.243 (0.076)
ln distance to previous event	-0.482^{***} (-0.019)	-0.484^{***} (-0.019)	-0.468^{***} (-0.020)
No of. failures	2291	2216	2216
ll null model	-30571.32	-30571.32	-30571.32
ll full model	-22183.54	-22121.65	-22006.83
$^{*}: p < 0.10,\ ^{**}: p < 0.05,\ ^{***}: p < 0.01,$			

Table 2: Cox Regression Estimation Results, Risk of Conflict Event, Population Size and Concentration Models, Including Distance to Rebel Group Headquarter, 1960–2004

Many of the variables in the model are closely related, and the individual coefficients cannot be interpreted on their own. Predictions for the fictive country represented in Figure 3 may help interpreting these estimates. Figure 5 shows how distance alone affects the estimated risk for each location in this map, based on the estimates in Model B1 (this figure disregards variation in population densities). Here, only the distances from the capital and the rebel group headquarter make a difference. The figure shows clearly how the risk increases as the distance from the capital increases and the distance to the rebel group headquarters increases. Both Hypothesis 2 and 4 are supported by these estimates. Note, however, that the magnitude of the effect is relatively small. The largest difference in log relative risk is less than 1 – the per-capita risk of conflict events is two to three times more frequent close to the rebel base than in the capital.

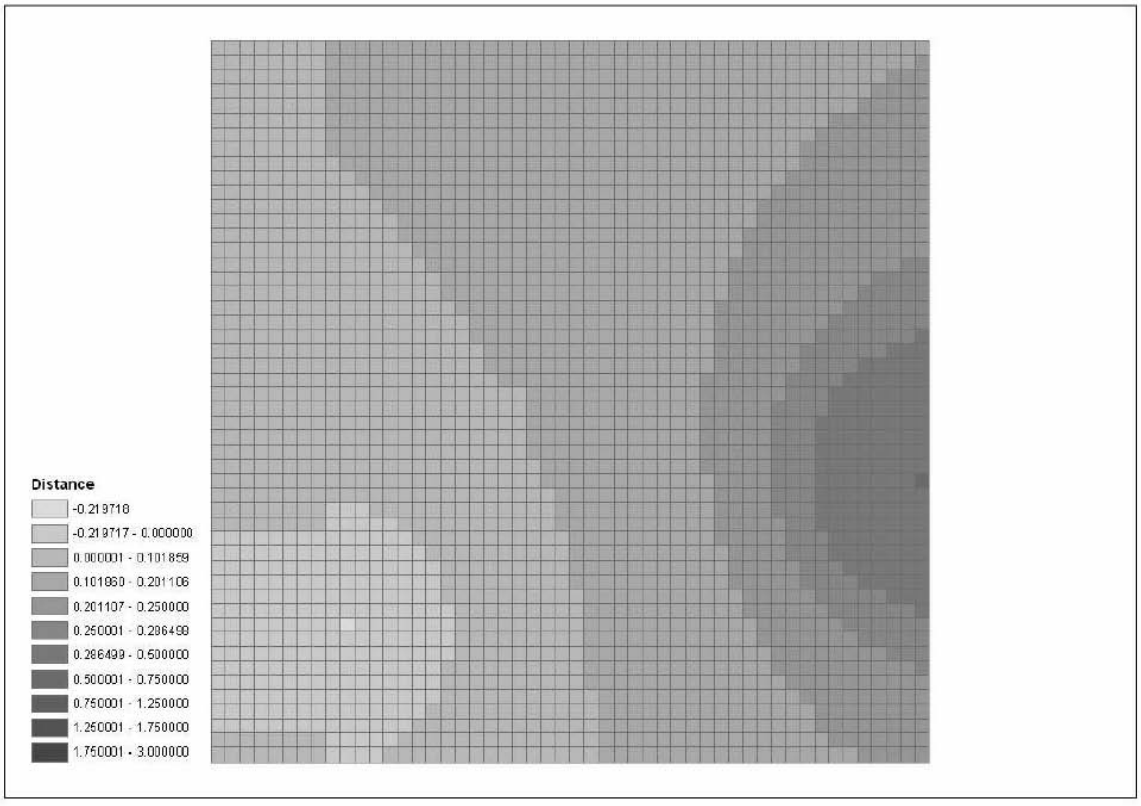


Figure 5: Estimated Per-Capita Relative Risk of Conflict Events for Locations in Figure 3, Distance variables only

Figure 6 shows the predictions from the same model, but taking border squares and population into account. The effect of the population mass variable is reflected as peaks

at the population concentrations. Note how the risk of events is much higher along the borders, and that the risk of events are highest where `ln(population in square') is high. Although the model clearly shows that the risk of events increase with the distance from the capital, the population variable is substantially more important.

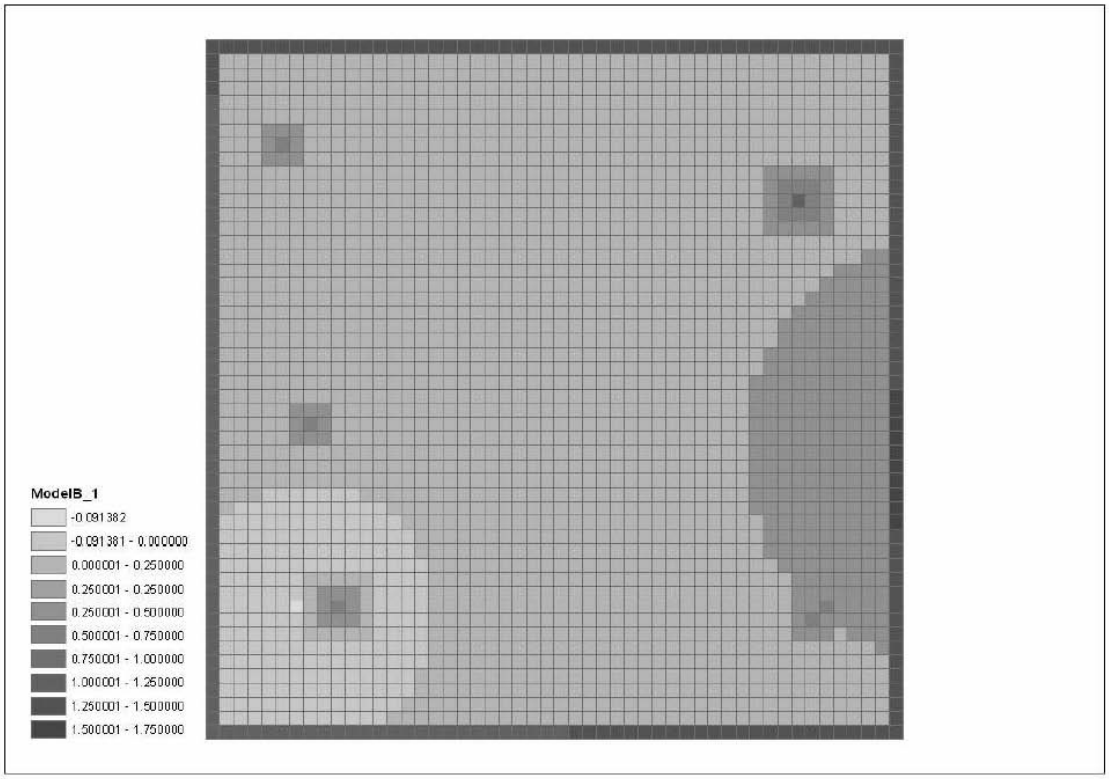


Figure 6: Estimated Relative Risk of Conflict Events for Locations in Figure 3, based on Model B1

In model B2, we test Hypothesis 5, adding the three `ln(population in neighborhood)' variables. The first- and second-order variables are positive and significant, whereas the third-order variable is negative and significant. With these variables in the model, the estimate for the population in square variable is negative. This result, again, may be due to collinearity: Since populations are clustered, populous squares often have populous neighborhoods, and the estimates should be interpreted jointly. Together, they provide ample support for the population concentration hypothesis: Conflict events happen disproportionally in squares close to population

centers. The negative estimate may be interpreted to mean the events occur in the outskirts of such population centers. Note the increase in log likelihood relative to Model B1, from $-22{,}183.54$ to $-22{,}121.65$.

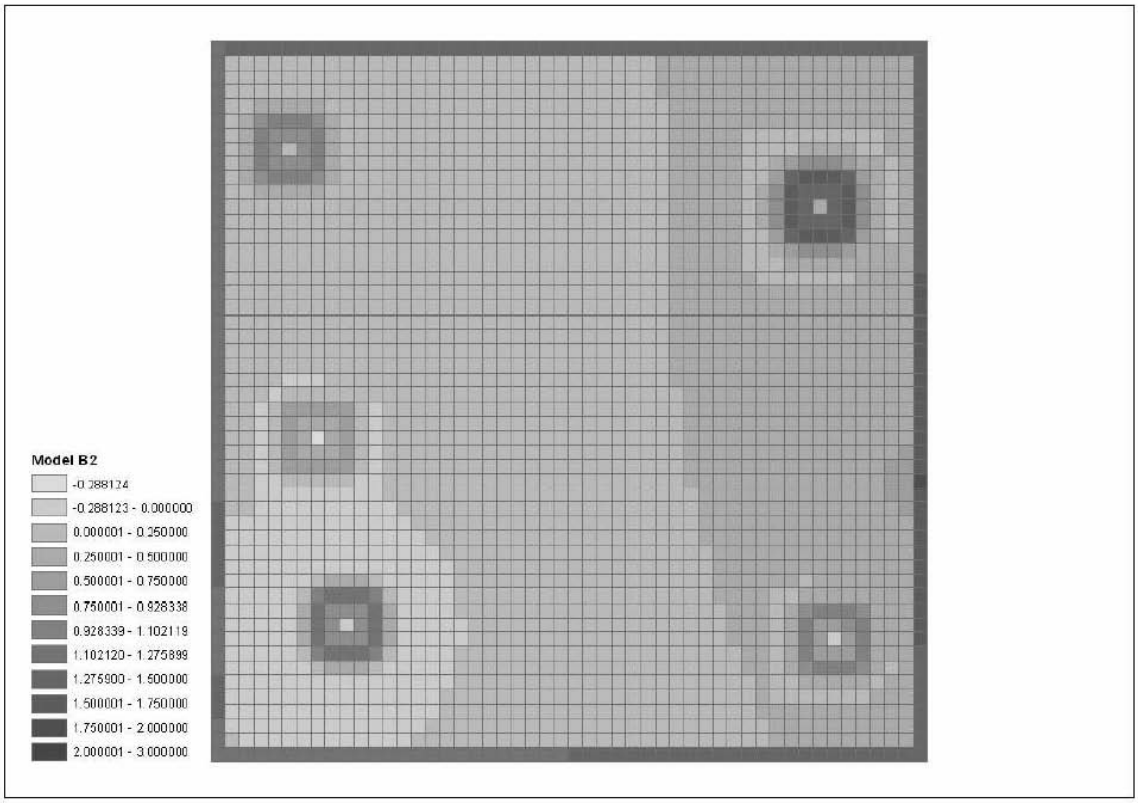


Figure 7: Estimated Relative Risk of Conflict Events for Locations in Figure 3, based on Model B2

Figure 7 helps interpreting the estimates in Model B2. Again, the figure plots the estimated log relative risk for each cell in the fictive country based on Model B2. The inclusion of the population size of the cells in the immediate neighborhood further increases the importance of population relative to distance. The squares close to the major cities are $3-5$ times more likely to experience conflict events than other squares. Conflict events clearly tend to occur close to population concentrations, not in the hinterlands. In particular, the city in the North East region is predicted to be particularly conflict-prone, since it combines a population concentration and a large distance from the capital. This city is the most conflict event-prone location on the map, rivaling even the border squares far from the capital. This testifies to the importance of including

information on the population in the neighborhood in the model, not just the population at the location. These results provide strong support for Hypothesis 6, and reflect the conflict patterns in Herbst's `difficult' political geographies.

In model B3, we test Hypothesis 8 by adding the interaction terms between the four local population variables and the distance from the capital to the model. The interaction terms are significant, and the log likelihood of the model is further reduced to $-22,006.83$.

The predictions from Model B3 are plotted in Figure 8. The population-distance interaction model further strengthens the importance of population concentrations relative to distance.[9]

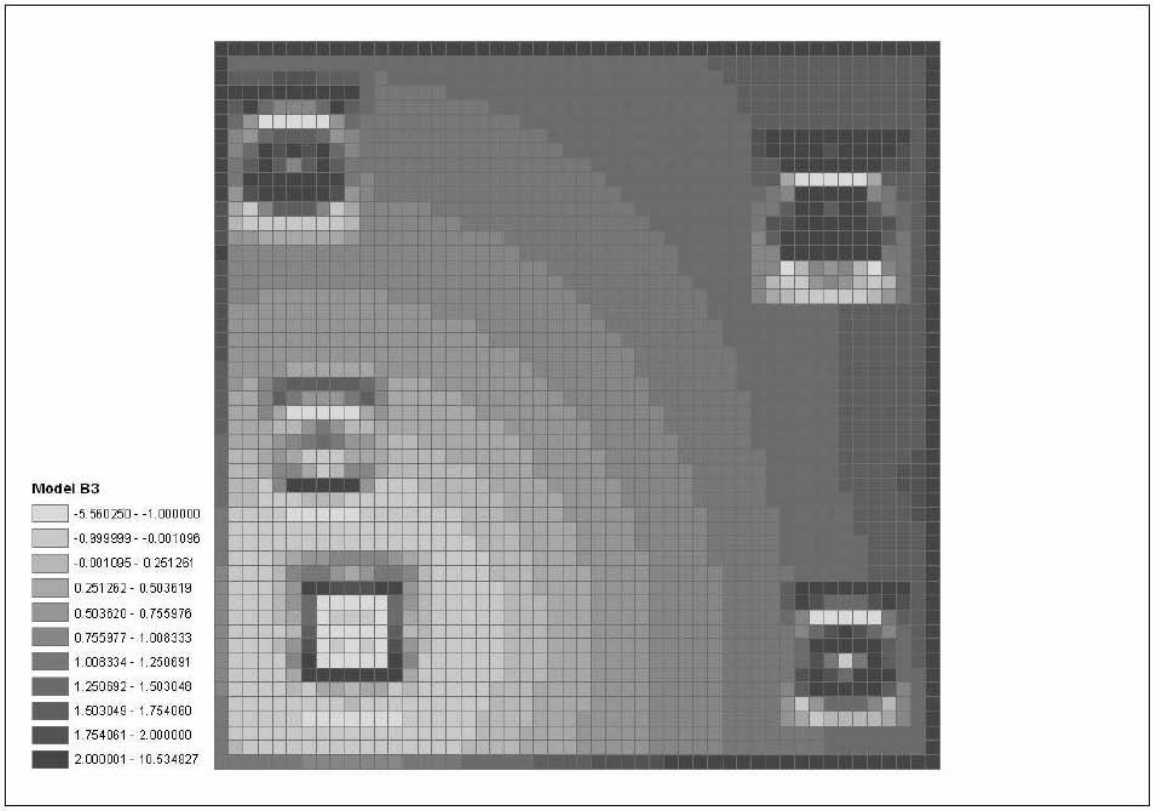


Figure 8: Estimated Relative Risk of Conflict Events for Locations in Figure 3, based on Model B3

[9]Models A3 and B3 suffer somewhat from collinearity problems, reflected as pairs of large positive and large negative estimates for variables that are highly correlated.

The estimate for $\ln(\text{population in country})$ is not statistically significant in most of the models in Tables 1 and 2, controlling for the other variables in the model. There is not much support for Hypothesis 7: The risk of conflict events is clearly larger in populous squares, but is independent of the size of the population in the country. We take this as evidence that we have succeeded in modeling the most important explanations of why large countries have more armed conflicts.

4.2 Results for Control Variables

The three road type variables in Table TableResults1 seek to refine the measure of the government's ability to control the territory at the location. The `primary roads' category was set as the reference category. Conflict events are 47% less likely to happen in squares with secondary roads than in the reference category squares, and 75% less likely to happen in squares with no roads or only informal roads than in squares with primary roads.

The results run counter to our initial expectations – conflicts are assumed to occur in faraway and inaccessible regions. However, the finding may not be so counter-intuitive after all. First, battle events occur where rebel group and army units encounter each other. Such meeting places are normally reached by road. Second, rebel groups tend to target high-value places (villages, military installations, pipelines, mines, etc.), and such places are also often connected by roads. Third, there is also a reporting bias at play here -- media report incidences primarily in accessible areas.

The variables designed to capture spatial and temporal dependence largely obtain the expected estimates. The estimate for the `proximity of previous event in square' variable shows that a square that have experienced conflicts one year ago has a risk of another event 167 times higher than squares with no conflict history. The `ln(distance to previous event)' variable shows that squares that are close to the location of the most recent event in the dataset are much more likely to see events than more distant squares – increasing the distance from an event by one percent decreases the risk of the next event occuring by 1.18 percent.

The estimate for `proximity of previous event in neighborhood, 1st order' is anomalous, on the other hand. The estimate is negative, indicating that the risk of conflict events decrease for a time in the immediate neighborhood of an event. The reason why the estimate turns out this way is not clear, but is most likely due to collinearity.

Both population variables are also associated with the road network variable, since population totals figures were interpolated assuming that population distribution and densities are correlated with accessibility. This introduces another potential for collinearity. The estimate for the road variable pull in the same direction, however: Civil war events are most frequent in the most accessible squares containing primary roads *and* high population concentrations.

5 Conclusion

This paper represents the first attempt to analyze the ACLED data (Raleigh & Hegre, 2005), which codes the dates and exact locations of individual events within a set of internal armed conflicts. In the paper, we have also developed a statistical tool to handle both spatial and temporal dependence, and to allow analyzing the dynamics internal to civil wars. Since we analyze both initial event in a given conflict and the diffusion in time and space given by the subsequent events, our analysis bridges the gap between studies of the onset and the duration of civil war. The unit of analysis in the paper is a 8.6x8.6km square of territory, for which we have coded data on conflict events, population, quality of road network, and location relative to the borders and capitals of countries.

The analysis has illuminated some aspects of the relationship between country size and the risk of internal conflict that cannot be analyzed in national-level studies: Conflict events tend to have frequencies in proportion to the size of the population in a given location, as indicated by a `per capita propensity' hypothesis. However, we also found evidence supporting the hypothesis that conflicts happen predominantly where populations cluster locally. In the sample of Central African conflicts we have studied, conflict events happen more frequently in locations far from the capital of a country and close to the border.

We also find the importance of distance to be considerably less than that of population mass, except when combined with population concentration. Conflict events

do happen in peripheral regions, but the analysis indicates that the picture of African internal conflicts as primarily rural events is inexact. The risk of conflict in a location depends on the value of the location. Here we have proxied the value of the location by the population that resides within it.

Although the effect of distance is moderate relative to population concentration, the results still indicate that countries with populations that are largely concentrated around the capital have fewer internal conflict events than countries with populations that are spread out, or, even more strongly, are also concentrated in locations far from the capital.

Our results might contribute something to the discussion on the usefulness of partition to solve internal armed conflicts (see Kaufmann, 1996 and Sambanis, 2000 for a discussion). The relatively small importance of distance from the capital relative to population mass indicates that partition might not be a very effective solution, in the sense that the total risk of conflict events in the two separated countries will only be marginally smaller. However, territories with large and clearly separated population concentrations might conceivably have perceptibly lower risks of conflict events as two or more countries surrounding each concentration than as one `difficult' political geography (our analysis does not say anything about the risk of interstate war between such entitites, however).

The results for other variables in the model are mainly consistent with our expectations and previous studies such as Buhaug & Gates (2002) and Buhaug & Rød (2006). One exception is the weak effect of distance from the capital in our analysis relative to theirs. This may be explained by their choice to use the distance from the capital to what they define as the geographical center of the conflict area as the distance measure. We, in contrast, use the distance from the capital to each individual conflict event. A more likely explanation for the difference in finding may be that they include data for all of Africa whereas we only analyze Central Africa. In particular, the conflicts in densely populated Rwanda and Burundi weigh more heavily in our analysis than in theirs.

This paper is a first in a series of papers that seek to retest empirical hypotheses using geographically disaggregated data. Up to now, with the exception of the work of Buhaug and associates and some country studies, quantitative studies of civil war have

been limited to national-level analysis. The analysis presented here points to the immense potential inherent in disaggregated analysis. The model presented here is well suited to be extended to test hypotheses regarding the availability of `lootable resources', regarding patterns in the distribution of ethnic groups and/or the geographical distribution of income, or the impact of the location of refugee camps or of lootable resources.

6 References

Alesina, Alberto & Enrico Spolaore, 2003. *The Size of Nations*. Cambridge, MA: MIT Press.

Beck, Nathaniel, Katz, Jonathan N., and Richard Tucker, 1998. `Taking Time Seriously in Binary Time-Series Cross-Section Analysis', *American Journal of Political Science* 42(4):1260-1288.

Buhaug, Halvard & Päivi Lujala, 2005. Accounting for Scale: Measuring Geography in Quantitative Studies of Civil War. *Political Geography* 24(4): 399–418.

Buhaug, Halvard & Jan Ketil Rød, 2006. `Local Determinants of African Civil Wars, 1970-2001', *Political Geography* 25(3): 315–335.

Buhaug, Halvard and Scott Gates, 2002. `The Geography of Civil War', *Journal of Peace Research* 39(4): 417–33.

Clapham, Christopher, 1985. *Third World Politics*. London: Croom Helm.

Collier, Paul, 2000. `Doing Well Out of War: An Economic Perspective', in Mats Berdal & David M. Malone, eds, *Greed & Grievance: Economic Agendas in Civil Wars*. Boulder, CO: Lynne Rienner (91–111).

Collier, Paul & Anke Hoeffler, 2002. `Greed and Grievance in Civil War', *CSAE Working Paper 2002/01. URL:* http://www.csae.ox.ac.uk/workingpapers/pdfs/2002-01text.pdf

Collier, Paul & Anke Hoeffler, 2004. `Greed and Grievance in Civil War', *Oxford Economic Papers* 56(4):563–595.

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

Gurr, Ted Robert, 1970. *Why Men Rebel*. Princeton, NJ: Princeton University Press.

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

Herbst, Jeffrey, 2000. *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton, NJ, Princeton University Press.

Kaufmann, Chaim, 1996. `Possible and Impossible Solutions to Ethnic Civil Wars', *International Security* 20(4):136–175.

Lichbach, Mark I., 1995. *The Rebel's Dilemma.* Ann Arbor, MI., University of Michigan Press.

McColl, Robert W., 1969. `The Insurgent State: Territorial Bases of Revolution'. *Annals of the Association of American Geographers* 59(4):613–631.

Raleigh, Clionadh & Håvard Hegre, 2005. 'Introducing ACLED: An Armed Conflict Location and Event Dataset'. Paper presented to the Conference on "Disaggregating the Study of Civil War and Transnational Violence", University of California Institute of Global Conflict and Cooperation, San Diego, CA, 7–8 March 2005. URL: http://www.prio.no/page/Publication_details//9429/46564.html.

Raknerud, Arvid and Håvard Hegre, 1997. `The Hazard of War: Reassessing the Evidence for the Democratic Peace', J*ournal of Peace Research*, 34(4): 385–404.

Sambanis, Nicholas, 2000. `Partition as a Solution to Ethnic War: An Empirical Critique of the Theoretical Literature', *World Politics* 52(4):437–483.

Sambanis, Nicholas, 2003. `Using Case Studies to Expand the Theory of Civil War'. *Conflict Prevention and Reconstruction Working Paper* No. 5 . May 2003. URL: http://lnweb18.worldbank.org/ESSD/essd.nsf/CPR/WP5

Scarritt, J. and S. McMillan, 1995. `Protest and Rebellion in Africa- Explaining Conflicts between Minorities and the State in the 1980s.', *Comparative Political Studies* 28(3): 323–349.

Strand, Håvard; Lars Wilhelmsen & Nils Petter Gleditsch, 2004. 'Armed Conflict Dataset Codebook Version 3.0', PRIO, 12 October. URL:

http://www.prio.no/page/CSCW_research_detail/Programme_detail_CSCW/9649/42464.html

Times Concise Atlas of the World, 2000. Eight edition. London: Times Books.

Toft, Monica Duffy, 2003. *The Geography of Ethnic Violence*. Princeton, NJ, Princeton University Press.

Weidmann, Nils; Håvard Hegre & Clionadh Raleigh, 2006. Modeling Spatial and Temporal Patterns of Civil War. Typescript, PRIO.