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

Paper presented to the conference on
'Mapping the Complexity of Civil Wars',
Zürich, September 15–17, 2005.

September 6, 2005

Abstract

The paper surveys a set of hypotheses forwarded in the conflict literature regarding the relationship between the size and location of population groups. The hypotheses are tested on a new dataset called ACLED that 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.

*Contact author: Håvard Hegre; hhegre@prio.no. Thanks to Joachim Carlsen for programming tools to generate several of the variables used in the analysis. The research has been funded by the Research Council of Norway.

1 Civil War and Country Size

The most robust empirical finding in country-level studies of civil war is that large countries often have more civil war than small countries (Fearon & Laitin, 2003; Collier & Hoeffler, 2004; Hegre & Sambanis, 2005; ...). Size, in all these studies, is measured in terms of population. As Sambanis (2003) points out, however, there is little agreement on why populous countries have more civil war than small ones (2003:26). This paper surveys a set of hypotheses forwarded in the conflict literature regarding the relationship between the size and location of population groups. The hypotheses are tested on a new dataset called ACLED that 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.

2 Population, Geography and Conflict

Several qualitative and quantitative studies within civil war literature have addressed the issue of geography and expanded upon its role within conflict. Geographic variables are often limited to populations, terrain and the reach of state power. Various arguments have been made, some contradictory, about the impact of these geographical variables on the onset and duration of civil war. This paper will focus on the population findings to emphasize how disaggregated data can further explain the links between large populations and conflict.

We will review a set of theoretical arguments and formulate empirical implications of them at two levels of analysis: a local level (e.g. for a village or a small piece of territory) and at the country level. The empirical analysis will attempt to discriminate between the explanations by testing them at the local level.

2.1 Absolute Population Size

The simplest explanation of the country-level relationship between population size and the risk and extent of conflict is based on the assumption of a constant '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 (0.07 at the mean), this effect alone would yield an elasticity of conflict risk with respect to population of slightly less than unity (Collier & Hoeffler, 2002: 11)

Similarly, they explain the finding that larger counties tend to have longer running conflicts by stating that

> [m]ore populous countries tend to have more rebellions, though not necessarily more than proportionately to their population, and so may have several under way at the same time. A conflict is coded as continuing if any rebellion is continuing, so that this alone will tend to produce a correlation between country size and the duration of conflict... Hence, the results do not imply that a continent divided into many countries would have shorter wars than an otherwise identical continent divided into few countries (Collier, Hoeffler & Söderbom, 2004:263).

Since this study investigates conflict duration, they do not state whether a continent with few countries would also have the same number of conflict *onsets* as a similarly sized continent with many countries. The 'per-capita propensity' explanation has precisely these implications. At the country level, it implies that the risk of conflict in a country increases with the size of its population. At the local level, the implication is:

Hypothesis 1 ***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.*

2.2 State-Level Mechanisms

The 'per-capita propensity' explanation is obviously too naïve, and very few researchers would rule out that populous countries may be different on the government side than small countries. Collier & Hoeffler (2002: 15), for instance, note that there may be economies of scale with respect to defense: A large country may have a lower per-capita propensity of conflict since the government of the large country will tend to be more effective and powerful relative to the rebel group. They interpret their country-level results as support of this hypothesized mechanism: 'The elasticity of the risk of conflict with respect to population is less than unity, consistent with modest economies of scale in defense'. At the local level, the corresponding implication is:

Hypothesis 2 ***Economies of scale in defense:*** *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, decreases with the size of the population of the country to which the location belongs.*

However, large countries may be more conflictual than small countries for several reasons. First, the average distance between the capital or the infrastructure of the state and the population clusters of the country will tend to be larger. We will return to this below. These distances may not only be geographical.

Fearon & Laitin (2003: 81), for instance, note that insurgency will be favored when potential rebels face

> (d) 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, and, also, increases the number of potential recruits to an insurgency for a given level of income.

The last part of this statement is the 'per-capita propensity' explanation, but the first in fact states that the elasticity of conflict with respect to population should be more than unity. At the local level,

Hypothesis 3 ***Heterogeneity:*** *The risk of civil war events at a location increases with the size of the population at the location and, controlling for the local effect, increases with the size of the population of the country to which the location belongs.*

Other factors might also reinforce this empirical expectation. One relates to heterogeneity: Countries with large populations (or extensive territories) tend to be more culturally or ethnically heterogenous. To the extent that differences in policy preferences cause civil wars, the chance is higher in larger countries that there exist groups who are willing to take up arms to protect their culture or ethnic group against a government controlled by a different group.

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*.

2.3 Population Diffusion

The state-level mechanisms are dependent on the distribution of populations within the country. Economies of scale in defense are clearly counteracted by the challenge of controlling large territorires. 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). Populous countries are also often geographically extensive. This is likely to add to the heterogeneity of the country. Preferences tend to be geographically clustered, so that geographical distance between two populations is correlated with distance in terms of preferences regarding public policies.

Alesina & Spolaore (2003: 40–45) develop a model based the assumptions that there are economies of scale in the production of a public good and that the utility individuals derive from the public good is decreasing with distance. Distance is conceived of in terms of both preferences and in terms of physical distance. The economies of scale allows large countries to provide better public goods, but at a given distance from the centre 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.

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.

Geographic peripherality is often linked with ethnic and political peripherality in the work of for instance Gurr (1970). These connections are difficult to test without specific information regarding the spatial character of ethnic groups within the state and neighboring states, but it is still reasonable to believe this reinforces the impact of geographical distance from the center.

Hypothesis 4 ***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.*

Collier & Hoeffler (2004) test Herbst's (2000) hypothesis in their country-level study by calculating a gini coefficient for population distribution. Countries in which 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.

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. 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 secession 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.

Hypothesis 5 ***Distance from Border:*** *The risk of civil war events at a location is higher in border zones.*

2.4 Population Concentration

Prospective rebel groups face trade-offs between economies of scale and geographical extent as do governments. 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). And this facilitates the bargaining of the terms of a contract. [...] As dissidents become more visible to one another, they are better able to monitor contributors and non-contributors [...] Finally, with 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. This serves to increase the importance of local population concentrations.

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 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).

Relatedly, 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.

Hypothesis 6 ***Population Concentration:*** *The risk of civil war events at a location is increasing in the size of population in the immediate geographical neighborhood*

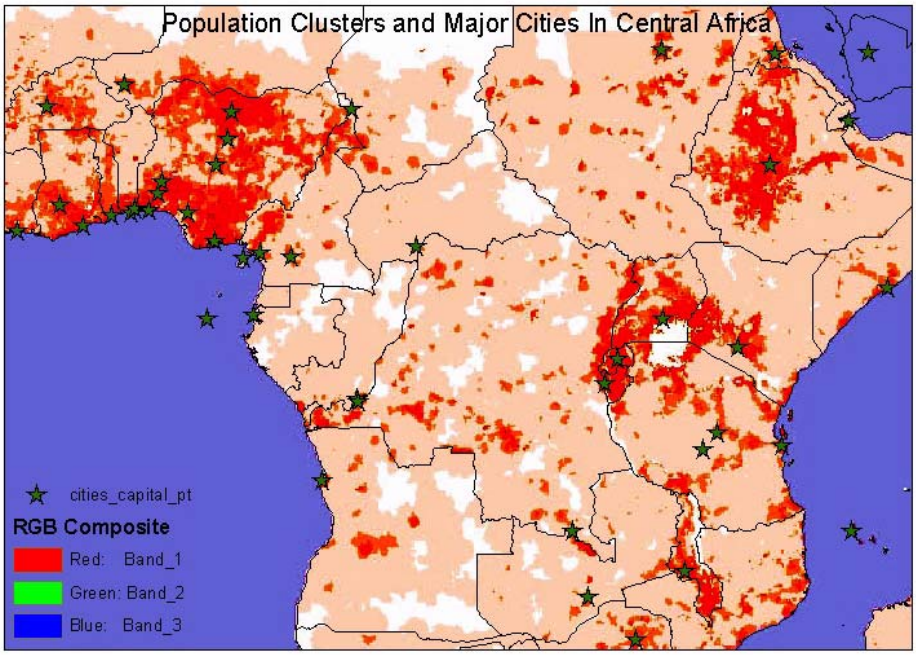

Figure 1: Population concentrations and major cities in Central Africa, 1990.

Hypothesis 7 ***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.*

Figure 1 shows the population concentrations in 1990 (CIESIN data) for Central Africa. Heavily populated areas are shaded in deeper tones of red/grey. Major cities are represented by stars on the map. The map shows population distribution at 1 km^2 resolution. In the analysis below, we aggregate up to 8.6*8.6 km squares.

2.5 Geographical Diffusion of Conflict

Lichbach (1995) addresses temporal and spatial diffusion as a construct of community relations. For example, he contends that collective dissent can be a product of successful dissent by group j at time t which will result in continuing the dissent pattern by group j in $t + 1$. Although not easily tested with our current data setup, Lichbach and others note clustering or 'cycles' of collective dissent by groups. However, in further tests of our data, it will be possible to test whether success at location x at time t leads to government reprisals in

location x at $t+1$.

Although these potential explanations involve theorizing on a local level, previous studies have largely tested them using state-level population measures. The only exception is the study of Buhaug & Rød (2005). This presents an ecological inference problem within the models 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 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.

2.6 Population Growth

Pressure on scarce resources due to population growth and in-migration is often quoted as a source of social tension and of violent conflict, e.g. Diamond (2005). André & Platteau (1998) explicitly cite land scarcity as source of the genocide in Rwanda:

> The 1994 events provided a unique opportunity to settle scores, or to reshuffle land properties, even among Hutu villagers... It is not rare, even today, to hear Rwandans argue that a war is necessary to wipe out an excess of population and to bring numbers into line with the available land resources (André & Platteau, 1998, cited in Diamond, 2005).

Hypothesis 8 ***Population Growth:*** *The risk of civil war events at a location increases with increasing population growth.*

Hypothesis 9 ***Population Density:*** *The risk of civil war events at a location increases the higher the population concentration.*

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 squares are our units of observation. We will refer to them as squares. This approach is similar to that of Buhaug & Rød (2005), with two important differences. First, their squares are much larger (100x100km). Second, they code the dependent variable considerable more crudely than is done in the ACLED dataset described below. This dataset records for each square the occurrence and date of conflict events, if any.

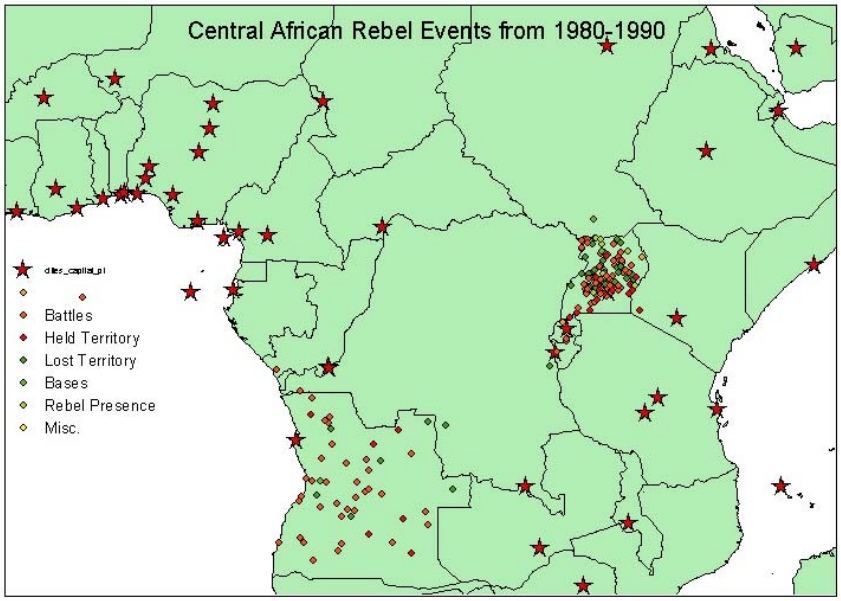

Figure 2: Events in ACLED (Armed Conflict Location and Event Data), 1980s, Central Africa

Our dataset cover 14 countries in Central Africa. 6 of them had a conflict in the 1960–2004 period according to the PRIO/Uppsala 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. The 14 countries cover 8 million square kilometers or just over 100,000 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.

3.2 Disaggregated Dependent Variable: ACLED

The PRIO/Uppsala dataset contains two variables (coded by Buhaug) that report the location and extent of each conflict: "In order to specify the geographic location of each conflict, every observation is assigned a conflict center point by its geographical coordinates (latitude and longitude). The conflict center is

fixed, so as to represent the geographic mid-point of all significant battle-zones during the conflict, including territory occupied by the opposition actors." (2002: 421). This is the dataset used in the analysis by Buhaug & Rød.

Although much more suited to geographically disaggregated analysis than other datasets, the PRIO/Uppsala dataset has some limititations, First, the dataset does not record changes over time in the center location and extent of conflicts. Second, the dataset 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.

We use the Armed Conflict Location and Events Dataset (Raleigh & Hegre, 2005). The dataset – called ACLED – 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.[1] ACLED is designed to parse out both the temporal and spatial actions of rebels and governments within civil wars. Rebel and government actions (or events) are recorded by date and type for the duration of the conflict. At present, twelve central and western African civil wars in eight countries are covered. 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 fundamental unit of observation in ACLED is the event. 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 records on 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.

For this paper, the ACLED data for the Central African conflicts was aggregated up to the 8.6x8.6km squares and merged with information on other explanatory variables aggregated to the same level. Figures 2 and 3 illustrate the ACLED data for Central Africa for the 1980s and the 1990s. Each location of a conflict event are represented by a symbol. In several of these locations, multiple events occured over the periods. Squares in which conflict events are located will be coded as conflict squares.

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

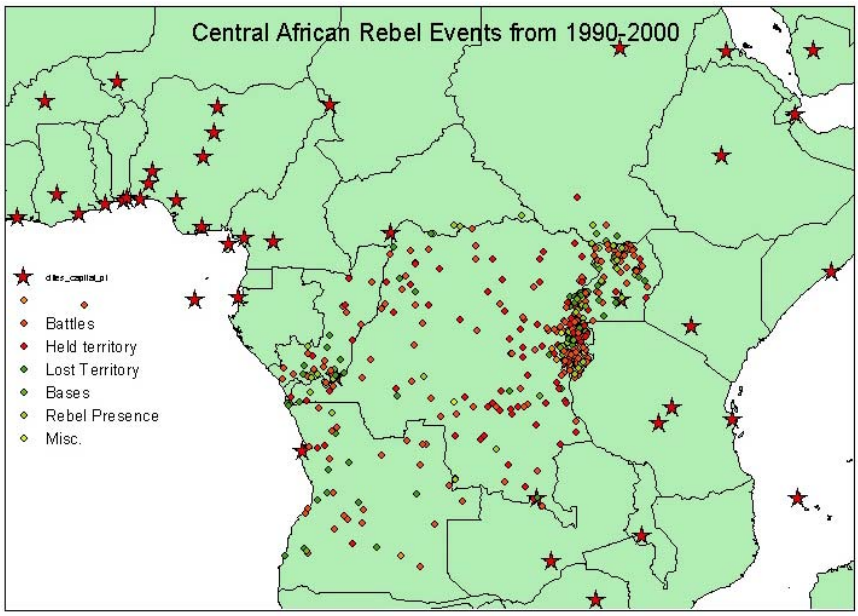

Figure 3: Events in ACLED (Armed Conflict Location and Event Data), 1990s, Central Africa

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):

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

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:

$$\begin{aligned} &\Pr(\text{war in a square } d | \text{a war breaks out at } t_w) \qquad (2) \\ &= \frac{\exp\left(\sum_{j=1}^{p} \beta_j X_j^d(t)\right)}{\sum_{i \notin R_{t_w}} \exp\left(\sum_{j=1}^{p} \beta_j X_j^d(t)\right)} \end{aligned}$$

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.[2]

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. Information on resources is also available as point data. 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 country 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.

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

[2] In the next version of the paper, we will sample half of the transition events in order to validate the model with out-of-sample predictions.

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

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.

Log population in neighboring squares To test hypotheses concerning population concentrations in the immediate neighborhood of the squares, we calculated the sum of the population sizes in the eight neighboring squares. The variable was 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 country-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.

Border Square We coded squares as border squares if a national border runs through it. Such squares belong to more than one countries. We coded country-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.

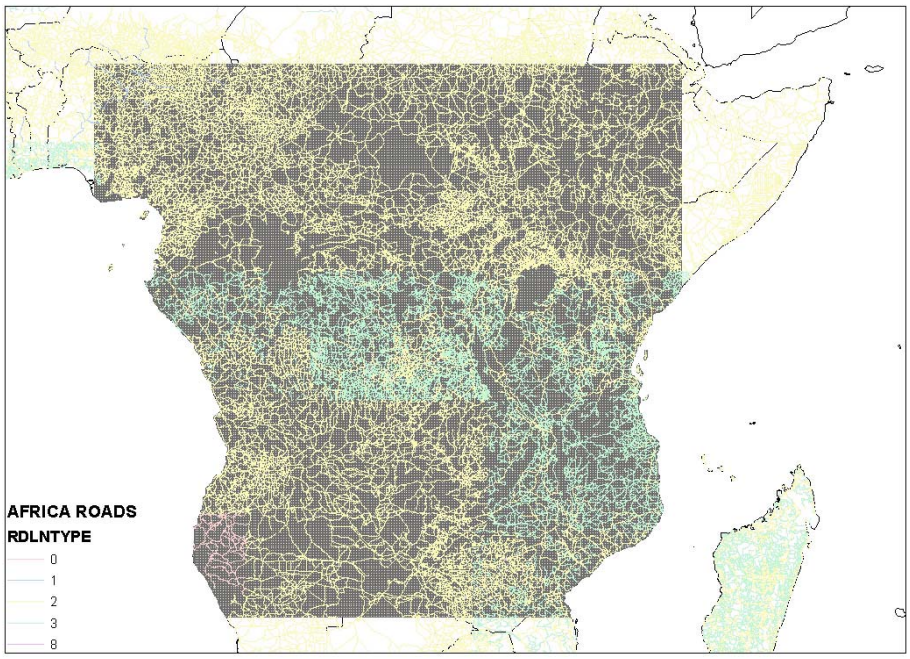

Figure 4: Road network, Central Africa.

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.

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

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) assume that the effect of the previous event decrease at a constant rate, and compute a decay function with 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

Table 1 presents estimated parameters for three models designed to test Hypotheses 1– 7. In the first model, the estimate for the 'population in square'

	Model 1	Model 2	Model 3
Variable	$\widehat{\beta}$ $(s.e.)$	$\widehat{\beta}$ $(s.e.)$	$\widehat{\beta}$ $(s.e.)$
ln(population in square)	0.122^{***} (0.014)	-0.132^{***} (0.021)	-0.667 (0.073)
ln(population in neighborhood)		0.340^{***} (0.028)	1.335^{***} (0.101)
ln(pop. in square) *ln(distance to capital)			0.0157^{***} (0.0212)
ln(pop. in neighborhood) *ln(distance to capital)			-0.0271^{***} (0.028)
ln(distance to capital)	0.118^{***} 0.024	0.161^{***} (0.024)	1.756^{***} (0.101)
ln(population in country)	0.0152 (0.034)	-0.0126 (0.035)	0.0565 (0.0386)
Border square	1.20^{***} (0.074)	1.15^{***} (0.074)	1.20^{***} (0.074)
Road type 0: Primary roads	$ref.cat.$	$ref.cat.$	$ref.cat.$
Road type 1: Secondary roads	-0.576^{***} (0.063)	-0.632^{***} (0.063)	-0.627^{***} (0.065)
Road type 2: Tertiary or no roads	-1.35^{***} (0.082)	-1.47^{***} (0.081)	-1.54^{***} (0.082)
Proximity of previous event in square	6.10^{***} (0.078)	6.10^{***} (0.076)	6.11^{***} (0.076)
Proximity of previous event in neighborhood (1st order)	-0.872^{***} (0.072)	-0.782^{***} (0.072)	-0.698^{***} (0.0734)
Proximity of previous event in neighborhood (2nd order)	0.0748 (0.074)	-0.0273 (0.075)	-0.0755 (0.077)
ln distance to previous event	-0.514^{***} (-0.020)	-0.504^{***} (-0.020)	-0.499^{***} (-0.020)
No of. failures	2216	2216	2216
ll null model	-29585.24	-29585.24	-29585.24
ll full model	-21593.90	-21535.44	-21479.43

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

variable is positive and significant – conflict events tend to occur more often in squares that are relatively populous. The estimate implies that increasing populations in squares by a factor of 2.7 increases the risk of conflict events by 12%. The estimate for ln (population in country) is not statistically significant, controlling for the other variables in the model. These two estimates support Hypothesis 1 rather than Hypotheses 2 and 3: The risk of conflict events is clearly larger in populous squares, but is independent of the size of the population in the country. Two caveats should be noted regarding the conclusions on the state-level mechanisms, however. First, the distance from the capital is part of the model. This variable is also likely to partly measure heterogeneity. Second, if Hypotheses 2 and 3 both are true, they will counteract each other. A failure to observe a relationship between the national population variable and the risk of conflict in an event may be because they balance each other exactly.

Hypothesis 4 is clearly supported in Model 1. Increasing distance from the capital by a factor of 2.7 increases the risk of conflict events by 12%. The estimate may also be interpreted as an elasticity: Increasing distance from the capital by one percent increases risk of conflict events by 0.12 percent relative to the baseline.

Hypothesis 5 also receives strong support from the analysis. 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.

The three road type variables in Table 1 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 44% less likely to happen in squares with secondary roads than in the reference category squares, and 74% 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.) for other types of events, 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

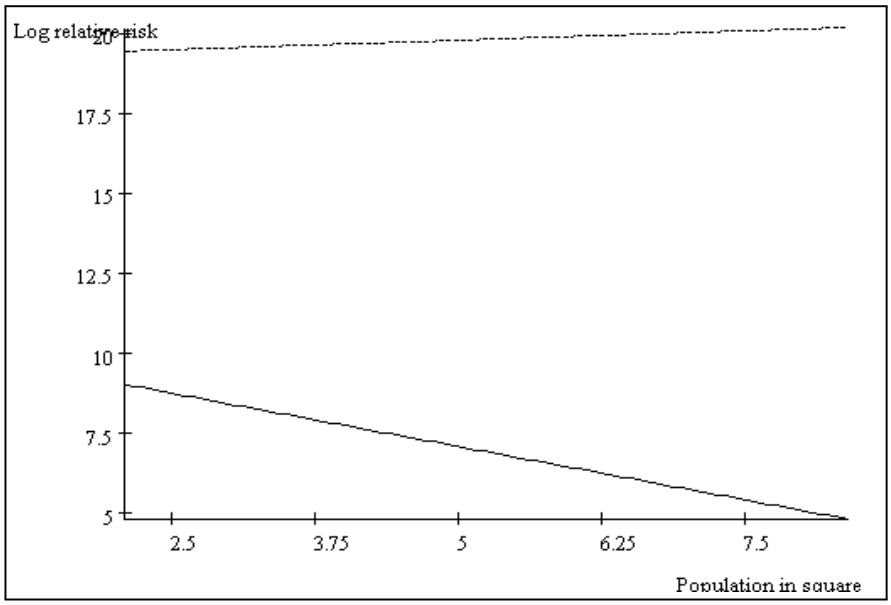

Figure 5: Log relative risk of conflict event as function of population in square, central (solid line) and peripheral squares (dotted line).

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.

In model 2, we test Hypothesis 6. The estimate for the 'ln(population in neighborhood)' variable is positive and significant. With this variable 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.

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.

In model 3, we test Hypothesis 7 by adding the interaction terms between the two local population variables and the distance from the capital to the model. The interaction terms are significant, but the parameter estimates are quite small in magnitude. In figure 5, the estimated log relative risk of war is

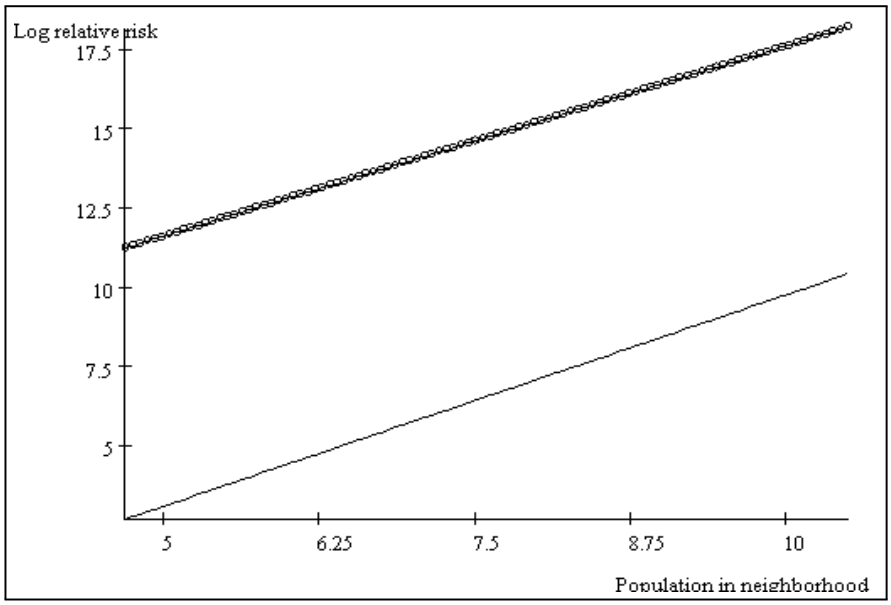

Figure 6: Log relative risk of conflict event as function of population in the immediated neighborhood square, central (solid line) and peripheral squares (dotted line).

plotted as function of log population in square (x axis) for a square adjacent to capital (solid line) and for a square with log distance 5 from the capital (i.e. 1500 km, which is the 90% percentile in the dataset). In both cases, the size of the population in the immediate neighborhood was held at the median (7.8). The figure shows that the (counter-intuitive) negative relationship between the population in the square is restricted to the immediate neighborhood of the capital. In the periphery, the risk of conflict events is independent of the size of the population.

In Figure 6, the estimated log relative risk of conflict event is plotted as function of log population in the neighboring squares (x axis) for a square adjacent to capital (solid line) and for a square with log distance 5 from the capital (i.e. 1500 km, which is the 90% percentile in the dataset). In both cases, the size of the population in the square itself was held at the median (5.4). The figure shows that the the interaction term has no substantial impact of the risk of conflict events.

In Figure 7, the estimated log relative risk of conflict event is plotted as function of log distance from the capital (x axis) for four types of squares: The black solid line represents a square with both population in square and population in the neighborhood at the 10% percentile (2.1 and 4.7, respectively). The black dotted line represents a square with the same square population, but with population in the neighborhood at the 90% percentile (10.5). The gray

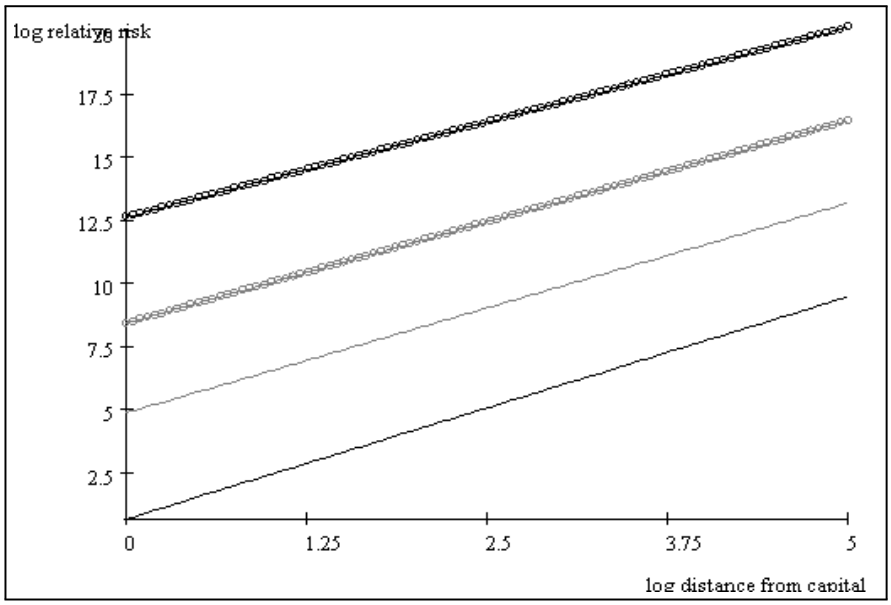

Figure 7: Log relative risk of conflict event as function of log distance from capital

solid line represents a square with log population in square at the 90% percentile (8.4) and log population in the neighborhood at the 10% percentile. Finally, the gray dotted line represents a square with both log population in square and log population in the neighborhood at the 90% percentile.

Figure 7 shows that when interpreted jointly, the risk of conflict events increases with distance from the capital for all constellations of local populations. It also shows that differences between squares in terms of log population in neighborhood have a much larger impact on the risk of conflict events than differences within the square itself. Moving from the 10th percentile to the 90th for this variable increases the risk several hundred times.

In Table 2, we test Hypotheses 8 and 9 by adding the local population growth variable to Model 2 as well as the size of the country in which the square is located. The estimate for the local population growth variable is negative and significant – conflict events are more likely to occur where population growth is low. The estimate for the size of the country is positive and significant, whereas the estimate for the total population of the country is negative. The two country-level variable estimates may be interpreted as one for population density: holding the area of a country constant, an increase in total population means an increase in population density. Hence, the estimate for the country-level population is also an estimate for population density, controlling for our other population distribution variables. This estimate is negative, implying that high-density countries have less conflict events than low-density countries.

	Model 4
Variable	$\widehat{\beta}$ $s.e.$
ln(population in square)	0.115^{***} (0.22)
ln(population in neighborhood)	0.334^{***} (0.029)
Change in ln(pop. in square)	-0.817^{***} (0.15)
ln(distance to capital)	0.160^{***} (0.028)
Border square	1.14^{***} (0.077)
ln(population in country)	-0.0712^{***} (0.040)
ln(size of country in km^2)	0.0246^{***} (0.035)
Road type 0: Primary roads	$ref.cat.$
Road type 1: Secondary roads	-0.534^{***} (0.065)
Road type 2: Tertiary roads	-1.33^{***} (0.084)
Proximity of previous event in square	6.18^{***} 0.077
Proximity of previous event in neighborhood (1st order)	-0.745^{***} 0.075
Proximity of previous event in neighborhood (2nd order)	0.0029 0.078
ln distance to previous event	-0.492^{***} -0.020
No of. failures ll null model ll full model	2134 -28409.61 -20514.68

Table 2: Cox Regression Estimation Results, Risk of Conflict Event, Population Growth and Density Model, 1960–2004

Both of these estimates run counter to Hypotheses 8 and 9: Against the background of our model, population growth and density rather decreases than increases the risk of conflict events at a location.

5 Conclusion

This paper represents the first step 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 allows bridging 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 country-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 in where populations cluster locally. In addition, conflict events happen predominantly in locations far from the capital of a country and close to the border.

Our set of findings correspond well to Lichbach's expectation regarding population clusters. Population clusters facilitate communication between dissidents, and hence helps solving prospective rebels collective action problems. The findings also confirm the argument that rebellion is facilitated by relative isolation, as argued by Lichbach, Gurr, and others.

We also tested whether population density within a country or population growth within a square affects the risk of conflict events, as predicted by theories regarding the effects of increasing scarcity of land. We do not find much support for these hypotheses, however, beyond the finding that conflcits tend to occur where many people live. This finding may have several causes, 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 (2005).

This paper is a first in a series of paper 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 has been limited to country-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', or regarding patterns in the distribution of ethnic groups.

6 References

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

André, Catherine & Jean-Philippe Platteau, 1998. 'Land Relations Under Unbearable Stress: Rwanda Caught in the Malthusian Trap', *Journal of Economic Behaviour and Organization* 34:1–47.

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, 2005. *Exit, Voice and Violence. Determinants of Territorial and Governmental Conflict, 1946–99.* Paper presented at the 13th National Conference in Political Science. Hurdalsjøen, 5–7 January. Http://www.statsvitenskap.uio.no/konferanser/nfkis/cr/buhaug.pdf.

Buhaug, Halvard & Päivi Lujala, 2005. "Accounting for Scale: Measuring Geography in Quantitative Studies of Civil War." *Political Geography* 24(00): 000–000. [In press.]

Buhaug, H. and S. 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 & Anke Hoeffler, 2004. 'Greed and Grievance in Civil War', *Oxford Economic Papers* 56(4):563-595.

Collier, Paul et al., 2003. *Breaking the Conflict Trap. Civil War and Development Policy.* Oxford: Oxford University Press. Http://econ.worldbank.org/prr/CivilWarPRR/. 0821354817.

Collier, Paul; Anke Hoeffler & Måns Söderbom, 2004. 'On the Duration of Civil Wars', *Journal of Peace Research* 41(13):253–273.

Diamond, Jared, 2005. *Collapse: How Societies Choose to Fail or Succeed.* Viking Books.

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

Hegre, Håvard & Nicholas Sambanis, 2005. 'Sensitivity Analysis of the Empirical Literature on Civil War Onset'. Paper presented to the Norwegian National Political Science conference, Hurdalsjøen, January 5–7.

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

Kalyvas, Stathis N., 2003. 'The Ontology of Political Violence: Action and Identity in Civil Wars', *Perspectives on Politics* 1:3, 475-494.

Kalyvas, Stathis N., 2005. 'Techniques of Violence in Greece and Vietnam'. Paper presented at PRIO, Oslo, Norway, 20-21 August, 2004.

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

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. URL: http://www.prio.no/page/CSCW_publication_detail//9793/46564.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, 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

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, 2004. *The Geography of Ethnic Violence.* Princeton, NJ, Princeton University Press.