Spatial and Temporal Patterns in Civil Violence: Guatemala 1977-1986

Timothy R. Gulden

Center on Social and Economic Dynamics
Working Paper No. 26
February 2002

ABSTRACT

This paper examines detailed records from the civil conflict in Guatemala between 1977 and 1986. It reveals a number of novel patterns which support the use of complex systems methods for understanding civil violence. It finds a surprising, non-linear relationship between ethnic mix and killing; thereby inviting analysis based on group dynamics. It shows the temporal texture of the conflict to be far from smooth, with a power spectrum that closely resembles that of other, better understood, complex systems. The distribution of incident sizes within the data seems to fall into two distinct sets, one of which, corresponding to "regular" conflict, is Zipf distributed, the other of which includes acts of genocide and is distributed differently. This difference may indicate that that agents of the state were proceeding under different types of orders. These results provide an empirical benchmark for the modeling of civil violence and may have implications for conflict prevention, peace keeping, and the post-conflict analysis of command structures.

The author thanks John Steinbruner, Robert Axtell, Joshua Epstein, Peyton Young, Carol Graham and Ross Hammond for their advice and encouragement.
Tim Gulden is a doctoral candidate at the University of Maryland School of Public Affairs and a Special Guest at the Bookings Center on Social and Economic Dynamics. Email: tgulden@wam.umd.edu.

OVERVIEW

Much of the existing literature examining quantitative aspects of civil violence concentrates on risk factors and searches for correlation between these factors and various indicators of violence. [Bates, 1983; Doyle and Sambanis, 2000; Fearon and Laitin, 1996] The foundation of these studies is generally annual, country level data on conflict deaths [Gurr and Harff, 1996]. While certain types of inferences can legitimately be drawn from such data, it does not lend itself to the study of internal conflict dynamics. This paper examines a substantially more detailed dataset covering the conflict in Guatemala during the ten year period 1977 to 1986. By shifting the basic unit of analysis from the country-year to the municipality-month, many intriguing patterns emerge. These patterns are generally indicative of "complex systems" behavior and point toward the use of new methods for exploring the dynamics of civil violence.

The aim of this paper is not a comprehensive statistical, political, or historical portrait of the Guatemalan conflict – a task which has been ably undertaken by others [Ball, Kobrak and Spirer, 1999; CEH 1999]. Instead, the objective is to uncover patterns in the data which illuminate spatial and temporal dynamics in the conflict and which might be used to guide quantitative modeling of civil and state violence.

THE GUATEMALAN CONFLICT

The history of state repression in Guatemala is, in many respects, particular to Guatemala and the victims of this repression were and are particular people with unique histories of their own. On the other hand, certain patterns can be observed in the data which may be of use in understanding such conflicts in general. This understanding, in turn, may be of use in predicting, preventing and controlling conflicts in the future.

The Guatemalan conflict lasted from 1960 to 1996 with a period of greatly heightened violence in the early 1980's. The state carried out most of the killing during the conflict in an ongoing campaign of repressive terror involving the military, the police, semi-autonomous "death squads" and state organized civilian "civil patrols". [Ball, Kobrak and Spirer, 1999]

Ethnicity played a significant role in the conflict. In the early parts of the conflict, the violence was typically between middle class people of the non-indigenous Ladino group struggling for control of the government. As the conflict progressed, it moved from an urban conflict focused on Guatemala City to a rural counter-insurgency campaign. The victims of state repression shifted at this point (about 1981) from middle class Ladino dissidents to indigenous Mayan peasants who were suspected of aiding rebel groups in the northwestern highlands. The scale and nature of the conflict changed as well, becoming vastly more deadly and including many acts which have been found to meet the formal definition of genocide [CEH, 1999]. It should be noted that the dichotomous division of ethnicity into Ladino and Mayan is probably more clear to the Ladino controlled government than to members of the various Mayan groups, who speak a large number of different languages and do not always consider themselves to be of the same ethnic group.

DATA

This work is based on a remarkable data set constructed jointly by the American Association for the Advancement of Science (AAAS) and the International Center for Human Rights Research (CIIDH) under the direction of Dr. Patrick Ball of AAAS. It documents over 40,000 killings and disappearances in Guatemala between 1960 and 1996. Many of these records include the specific time and place where the incident occurred as well as other detailed information. It is based on an extensive review of Guatemalan press sources over the entire 36 year period and over 5,000 interviews with witnesses.

While there exist other data sets of this sort (for El Salvador, for instance) this is the only record of its kind which is published and generally available for research. It thus provides a fertile ground for the formation of hypotheses (since it is new) but can provide little in the way of confirmation of these hypotheses (since it is unique). It is hoped that research into the spatial and temporal dynamics of violence will spur interest in this kind of disaggregated data and lead to the creation and publication of additional data sets.

This research uses a subset of this data spanning the ten year period of 1977 to 1986. Data are further restricted to killings and disappearances for which the date was known to at least the nearest month. This subset contains 24,000 cases which probably constitutes about 10% of the killings during this 10 year period. This estimate is uncertain because the number of killings overall has been estimated at anywhere from 80,000 to 400,000. The analysis that follows assumes that this sample is relatively unbiased. This is, of course, a risky assumption in spite of the rigor with which the data were collected.

METHODS

Much of the existing quantitative treatment of large scale violence relies on summary statistics which provide information about a conflict over a large span of space (a nation or a conflict zone) and of time (a year or the duration of a conflict). Many of these studies use linear regression and related statistical techniques to correlate violence with other factors in an effort to understand and predict such outbreaks.

Given the richness of this data set, we have taken a different approach. We have tried to preserve the complexity of the data wherever possible and to explore the finer grained data for regularities which might be applicable in other situations. Major tools in this effort included complex queries of the data in Structured Query Language (SQL), spatial analysis and mapping with a geographic information system (GIS), histograms, time series plots, rank/size plots and other, mostly graphical, representations of disaggregated data.

This approach has limitations. In most statistical analyses, one tries to form hypotheses independent of the data and then use the data to test these hypotheses. In this case, an examination of the data was used to construct hypotheses, making it impossible to use the same data to test these hypotheses. The observations which follow are therefore offered not as proven generalizations, but as suggestive patterns with theoretical plausibility. The proof of their generality will have to wait for detailed data from other conflicts.

OBSERVATIONS

Frequency vs. Severity

In the data set, the frequency of killing in a municipality is only weakly correlated with the quantity of killing in that municipality. The coefficient of correlation between frequency of killing (number of months where at least one person was killed in a town) and intensity of killing (number of people killed in the town over the whole study period) is .65. More tellingly, perhaps, the correlation between the number of people killed individually and the number of people killed in groups larger than one is only .31.

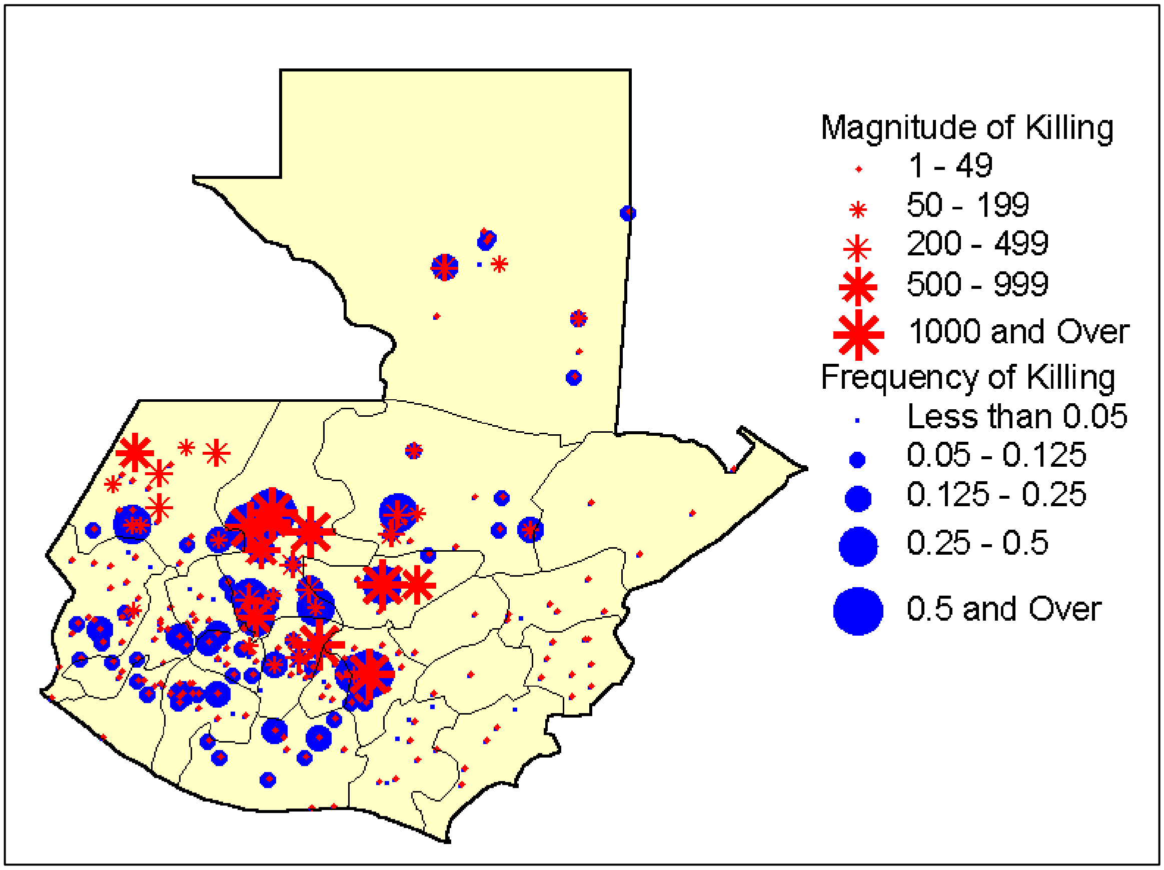

Figure 1. Map of Frequency and Severity of Killing by Municipality

Sources indicate that many of those killed individually were targeted by the government [CEH, 1999]. This observation supports the modeling observations [Epstein, Steinbruner and Parker, 2000] that the removal of leaders is an effective repression technique. We might assume that the government is well aware of this phenomenon and removes leaders (by killing them) in areas where it knows who these leaders are. These are the areas where we see a large number (and a high frequency) of single assassinations. In areas where the government does not know who the leaders are, we see more intensive and indiscriminant killing. This may be a result of a combination of two factors. On the one hand, the government may have killed indiscriminately because it did not know how to choose its targets. On the other hand, insurgent activity may have been able to gain a greater base because the government was less able to repress it through assassination.

The hypothesis that less knowledge on the part of the government can lead to more indiscriminant killing is further supported by the tentative observation that violence was more intense in inaccessible areas. While this is hard to quantify precisely, it appears that massacres were more likely to be carried out in the mountains and away from improved roads.

Ethnic Mix

A second and perhaps more striking observation resulting from the spatial disaggregation of the data is that intensity of killing in a municipality has a somewhat complex relationship to the ethnic mix in that municipality. While the population of Guatemala is fairly evenly divided between the Ladino and Mayan ethnic groups, they are generally segregated at the municipal level. About 76% of the population lives in municipalities which are more than 80% dominated by one group or the other.

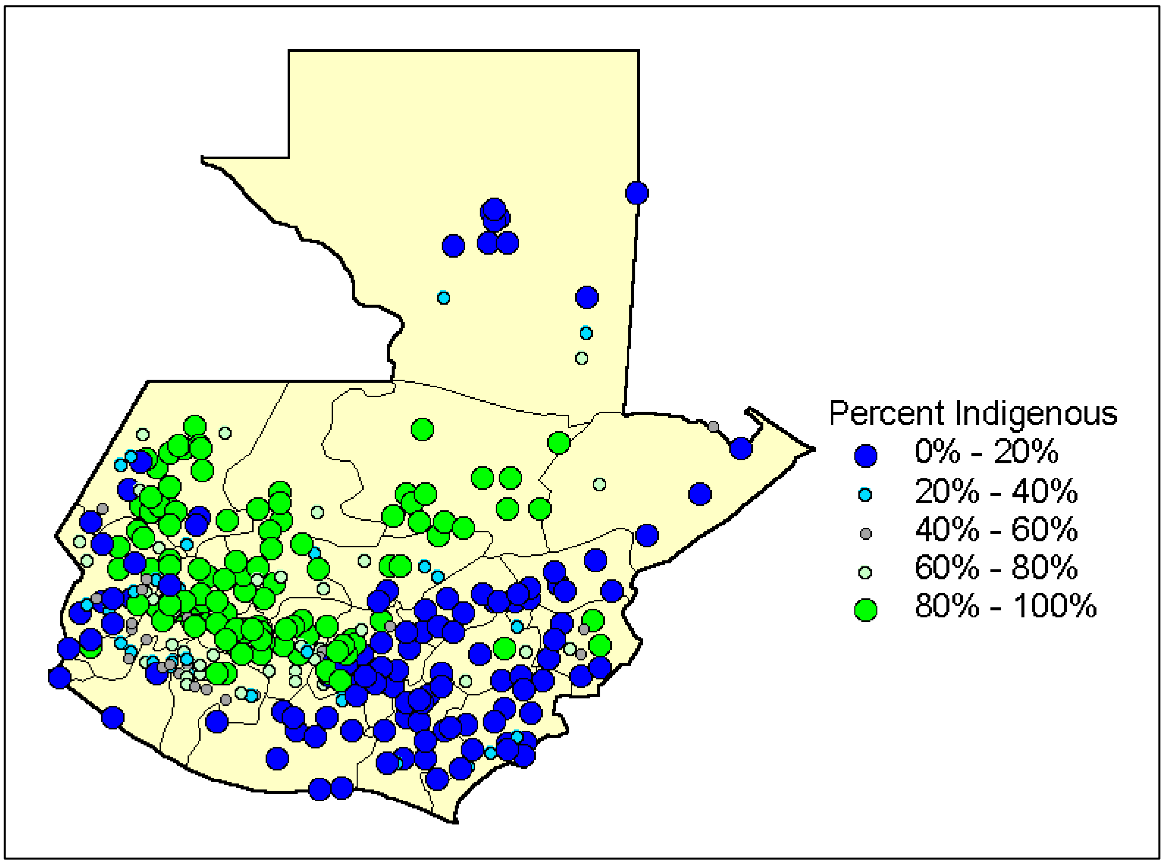

Figure 2 -- Map of Ethnic Distribution in Municipalities

The Mayans live largely in the mountainous northwest section while Ladinos occupy the lower and more agriculturally productive south and east portions of the country. Even within these regions, however, there is significant polarization.

Examination of the quantity of killings within these largely segregated municipalities led to an unexpected finding: the few municipalities where Mayans make up a large, but not overwhelming, majority were the most consistently dangerous. Just over half of the killing took place in municipalities in which the Mayans made up between 80 and 90 percent of the population. This is remarkable because such municipalities make up less than 8% of the municipalities in the country and house just over 8% of the total population (about 17% of the Mayan population). Many more Mayans (45%) live in municipalities where they constitute upward of 90% of the population. Because the number of municipalities is relatively large (n=345) these variations are unlikely to be a pure statistical artifacts (the differences are significant beyond the 99% level). While we might expect violence to increase monotonically with the percentage of Mayan residents, this proves not to be the case.

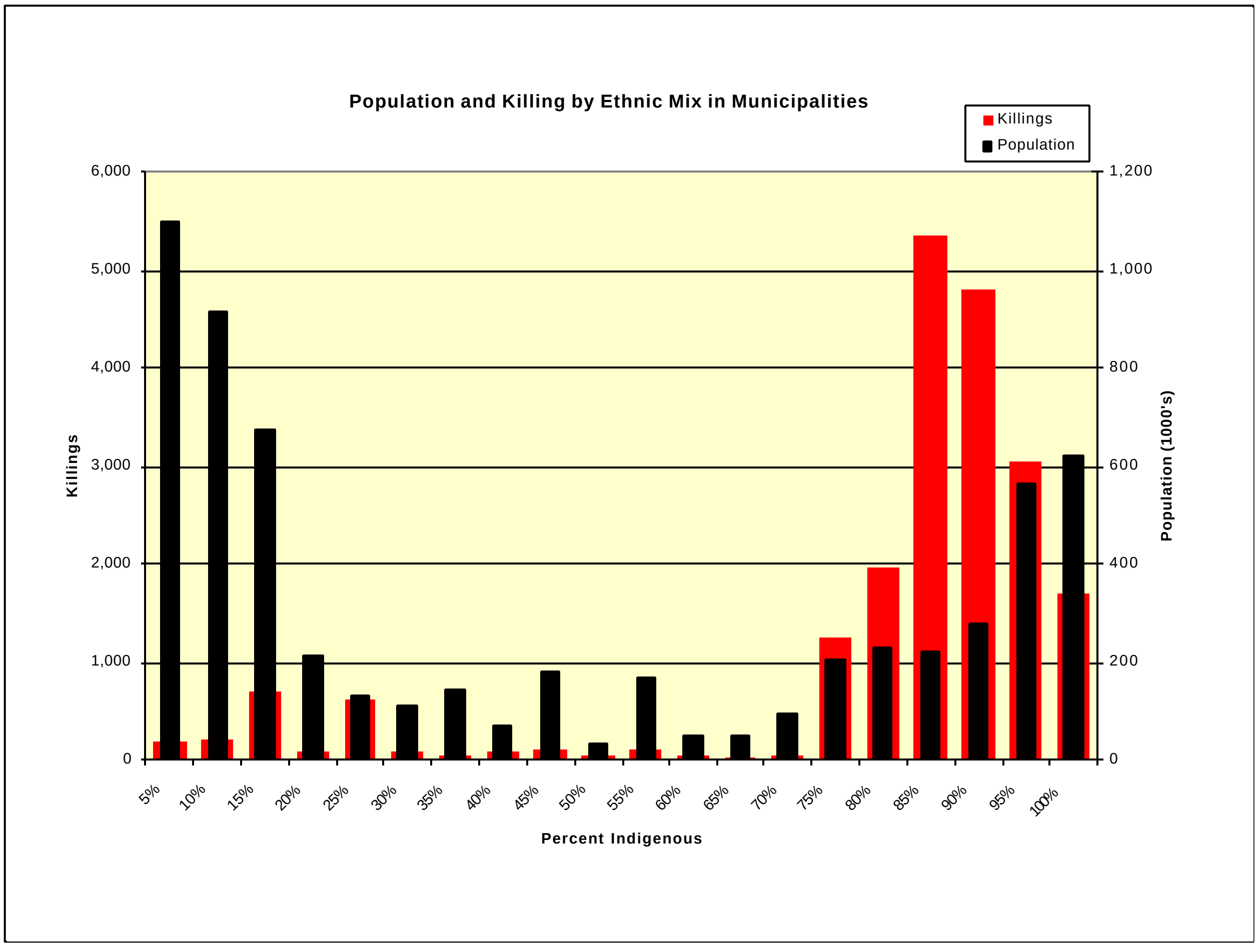

Figure 3 -- Histogram of Ethnicity and Killing.

At least two different mechanisms might explain the fact that more killing took place in municipalities with small but significant Ladino minorities than in municipalities with almost entirely Mayan populations. One thought is that the rate of killing increases with the percentage of Mayan residents up to a point because as this percentage increases, the government knows less about the leadership structure of the insurgency, and is thereby inclined to kill indiscriminately as discussed above. Beyond some point, however, the government may know too little to do anything. This would be a real world example of a little knowledge being a dangerous thing. A lot of knowledge leads to assassination of leaders, a little knowledge leads to indiscriminant killing, and no knowledge leads to no action.

A second mechanism might be based on group dynamics. There may be a threshold concentration that individuals with a minority trait must reach they are considered (or consider themselves) a group. It is possible that, in municipalities where the Ladino population constituted less than 10% of the population, tensions between the groups were substantially less because at some basic level, the Ladino population did not constitute a separate ethnic group.

An examination of the opposite end of the histogram provides some support for this interpretation. We see a similar, though much smaller, bump in the number of killings in the range between 10% and 25% indigenous (i.e. 75% to 90% Ladino). The vast majority of the killing in the conflict was directed against Mayans, and these areas had relatively few Mayans. Therefore, it is not surprising that fewer people were killed in these areas. The basic insight remains the same however. In areas where Mayans constituted less than 10% of the population, they may have been perceived more as individuals than as a threatening group.

Thus, at both ends of this histogram where one group or the other is more than 90% dominant, we see less violence. This may be because, in such communities, people relate as individuals rather than ethnic groups. Such communities might be more tightly knit and better able to avoid government persecution. Also in the middle of the histogram, where neither group is more than 75% dominant, we see relative safety. The area between 75% and 90% dominance, however, seems to be much more volatile. If this observation is born out in the examination of local populations in other conflicts, it could prove to be a useful rule of thumb for peacekeeping operations.

Punctuated Equilibrium

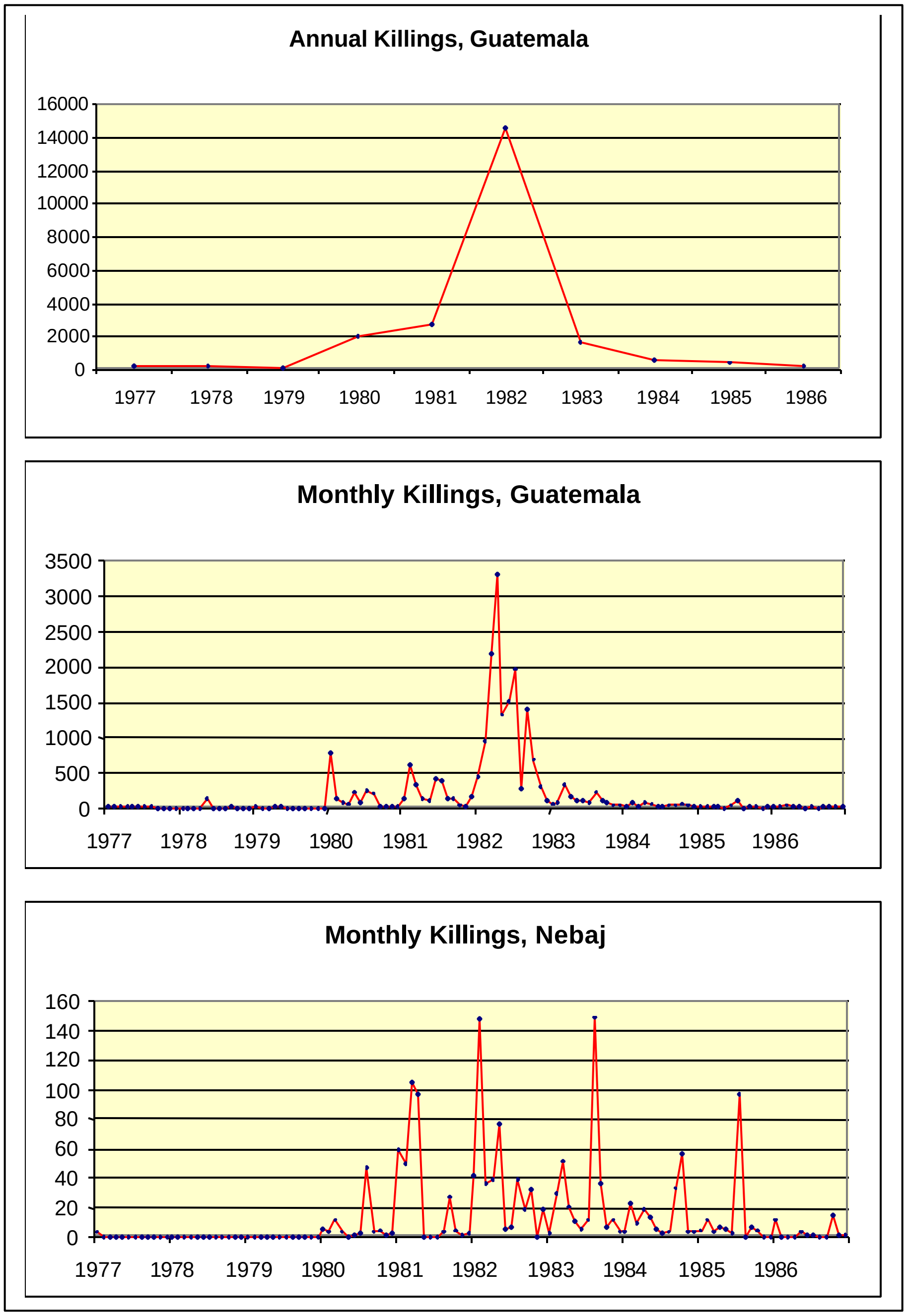

Figure 4 – Time Series Graphs: Annual, Monthly, Monthly for a Single Town

Another striking observation arises when the data are disaggregated with respect to time as well as space. The violence in a given place does not expand and contract smoothly over time. Rather, the pattern of violence is "spiky". A municipality may go for some time without an incident and then experience a major incident, or cluster of incidents. This becomes increasingly apparent as we move from aggregate annual numbers to finer resolutions of time and space.

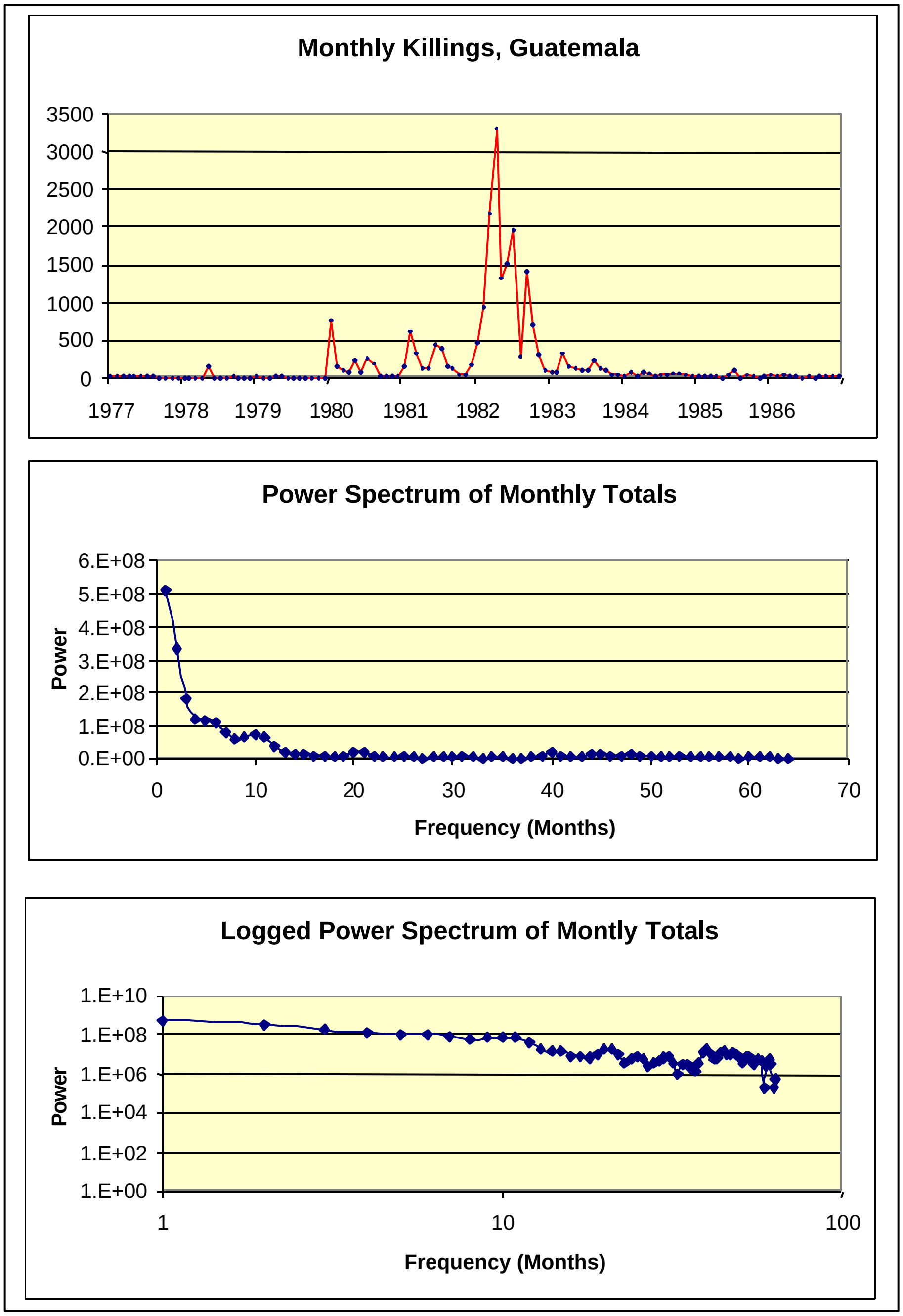

Figure 5 -- Time Series, Power spectrum, Logged Power Spectrum.

An objective measure of the character of such time series data can be obtained by examining its power spectrum. Purely random noise (white noise) has equal power at all frequencies. Complex systems, however, frequently exhibit "pink" noise (sometimes also called "1/f noise"), where the power at a given frequency is inversely proportional to the frequency [Schroeder, 1991]. An examination of the time series of monthly killings in the Guatemala data set (using a Fourier transform) reveals this kind of power law spectrum. In this case the exponent of the power law is not precisely -1 (i.e. $1/f = f^{-1}$), but something closer to -1.4. This exponent provides a kind of signature for a process exhibiting pink noise [Bak, 1997]. Examination of other conflicts may reveal that this signature is consistent from one to the next or that it varies in a way that is informative.

Distribution of Incident Sizes

An examination of the distribution of incident sizes within the data set provides some additional insight into the internal dynamics of the conflict. The conflict can be separated into two parts: a "normal" counterinsurgency and a genocide which was focused in the western highlands in 1981 and 1982. The counterinsurgency is characterized by a "Zipf" distribution of incidents, whereas the genocide follows a different pattern.

A sense of the overall distribution of incident sizes is given by the rank/size (or Pareto) plot presented in Figure 6.

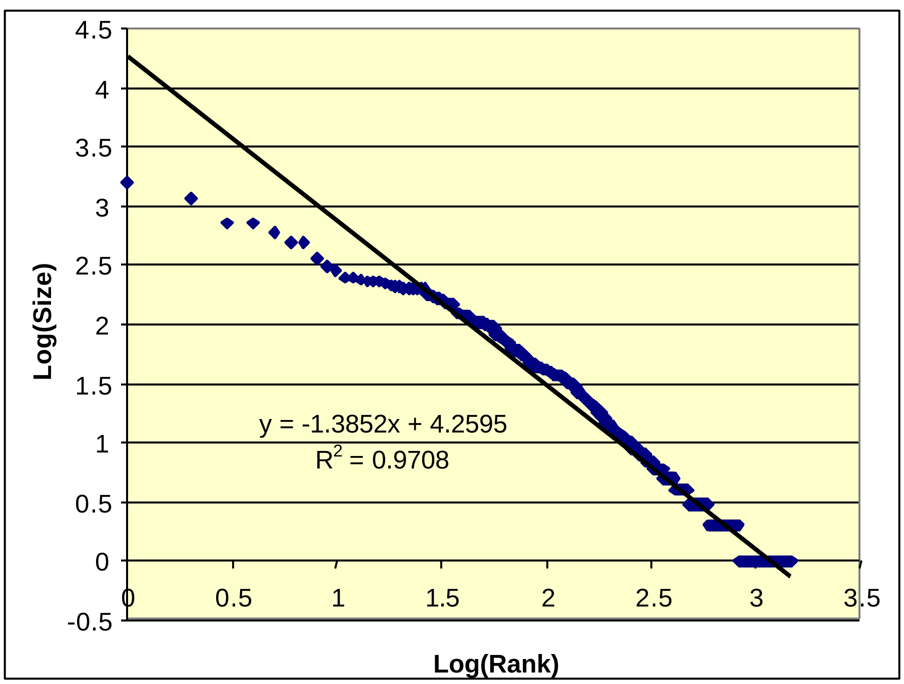

Figure 6 – Rank/Size Plot of Killings per Municipality-Month.

This ordered histogram (on log-log axis) gives only a rough idea of the real distribution for several reasons. First, it combines regular conflict and genocide -- two processes which, I will argue, follow different dynamics. Second, it does not represent killings per incident directly, but rather killings per municipality per month. This is due to data limitations. Both of these problems can be worked around to achieve insight into the workings of the conflict.

To examine the difference between the regular and genocide parts of the conflict, we need to partition the data with respect to both time and space. We saw above that 1981 and 1982 were years of particularly intensive violence. Figure 7 examines this period with respect to spatial distribution by showing the number of massacres per town in 1981 and 1982.

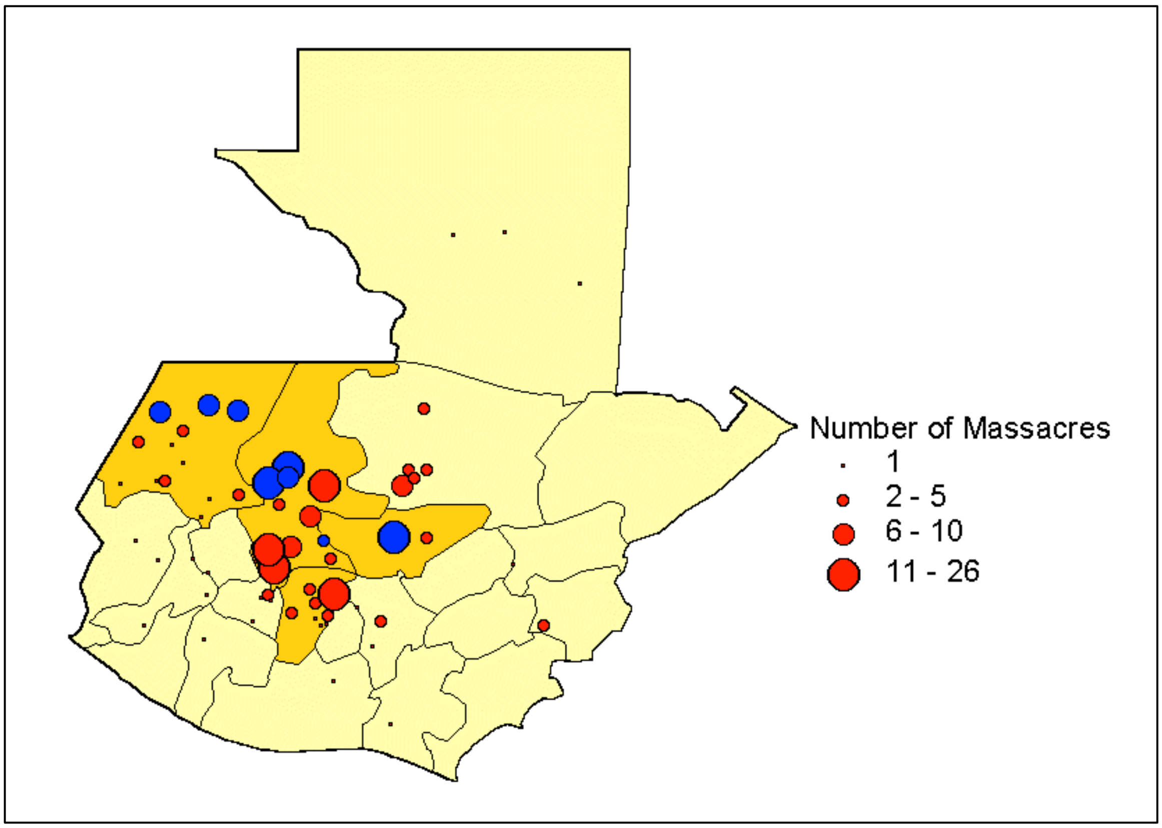

Figure 7 -- Map of Massacres & Genocide.

In its 1999 report, The Guatemalan Commission on Historical Clarification (CEH) documented, with painstaking thoroughness, a number of incidents during which the formal criteria of genocide were met [CEH, 1999]. All of these incidents involved massacres in the highlands between 1981 and 1982. The CEH further acknowledges that many additional incidents of genocide took place but were not formally documented. In Figure 7, the municipalities in which the CEH documented genocide are colored blue.

By taking the number of massacres in a municipality as a proxy for the level of genocide activity in that municipality, we can roughly identify four departments (Huehuetenango, El Quiche, Baja Verapaz, and Chimaltenango) as the focus of the genocide. In order to look for differences between genocide and regular warfare, we separate records from these four highland departments during 1981 and 1982 (the genocide subset), from the rest of the data set (the non-genocide subset).

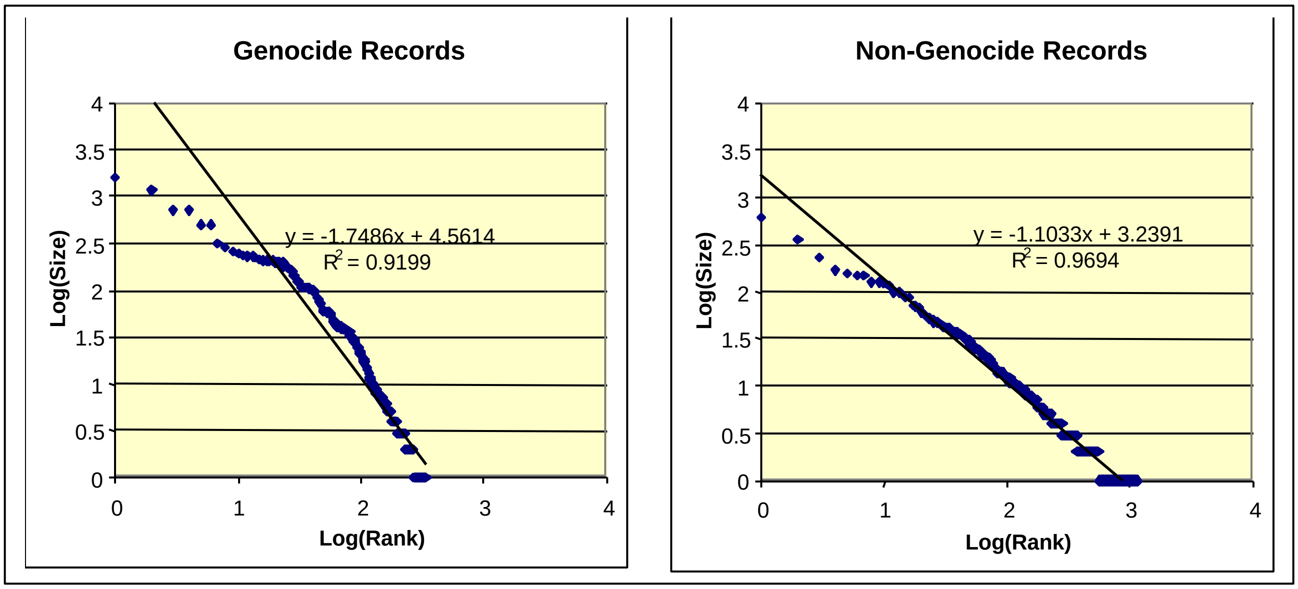

Figure 8 -- Non-Genocide and Genocide Pareto Plots

The municipality-month rank/size plots for these two subsets look significantly different. The non-genocide subset (n=1133) closely approximates a straight line with slope -1.13 in log-log coordinates. This is to say that the distribution can be described by a power law of the form $S=aR^{-1.13}$ (where S is size and R is rank). The genocide set (n=338), on the other hand, is quite concave toward the origin and has a much higher slope (to the extent that it can be described by a power law at all).

The non-genocide subset is actually even closer to the power law distribution than it might appear. The departure in the upper tail is due to two or three "extra" events with size around 250. It is these few events which leave the distribution short at the top end. This is quite different from the overall dataset, where the largest 30 or so events describe a curve with slope much lower than the distribution would require.

Once purged of the genocide related records, the regular conflict data adhere more closely to a power law distribution, but we are still working with the somewhat artificial unit of the municipality-month. While the resolution of the data is not sufficient to examine the exact size distribution at the incident level, it does allow us to estimate the total number of incidents represented by the data – about 3500 in the non-genocide set.

If we think of the municipality-month as an aggregation bin, then the non-genocide, municipality-month set (n=1133) represents an average of 3.05 incidents per bin. By further aggregating the data temporally at the 6 month, 1 year, 2 year, 5 year and 10 year levels, and determining the power law exponent at each of these levels of aggregation, we are able to establish a linear relationship between incidents per bin and the exponent.

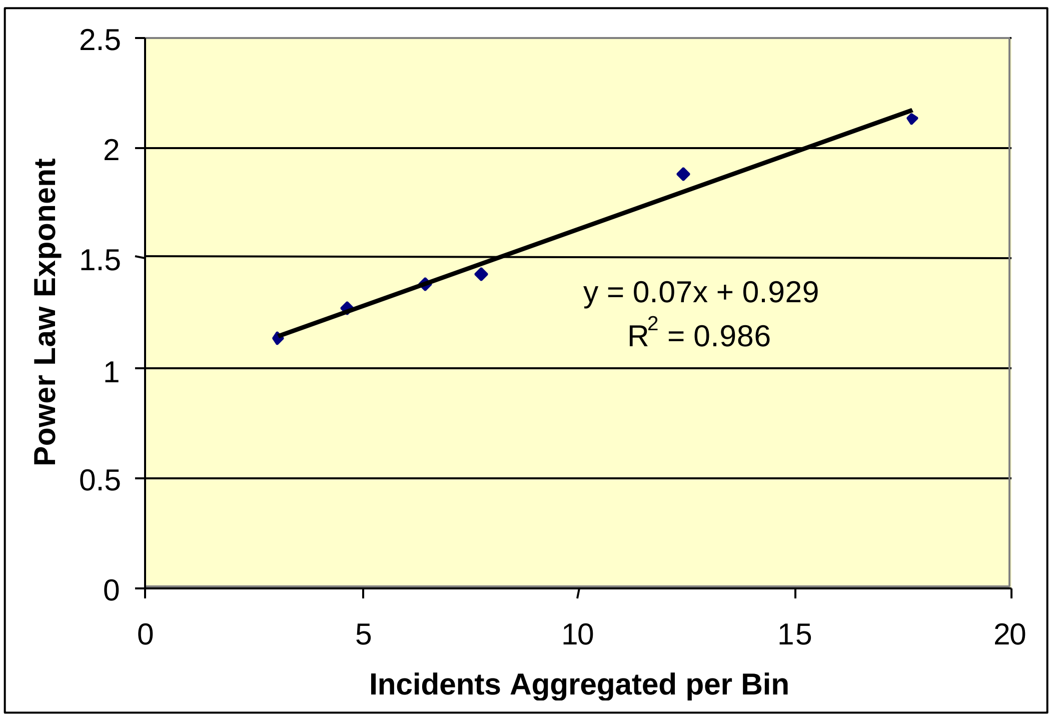

Figure 9 -- Trend of Power Law Exponent

This relationship is nicely described (R^2=0.986) by the linear relationship y = -0.7x - 0.929, where y is the exponent and x is the average number of incidents per bin. From this empirically derived relationship, we can estimate the exponent for the fully dissaggregated case where there is only one incident per bin by simply evaluating the expression at x=1. The resulting value of -1.056 is extremely close to -1, the exponent which defines the so called "Zipf" distribution. The Zipf distribution is characteristic of many processes in the physical and social worlds including city and firm sizes, earthquake magnitudes, certain aspects of Internet traffic, and a host of other phenomena [Bak, 1997].

A random growth rate model is a simple way to create a Zipf distribution and the workings of such a model are suggestive here. The model involves an arbitrary number of objects (in this case, potential incidents), each of which has a size greater than or equal to one (S >= 1). The initial distribution of sizes is not important to the long term behavior of the model, so we will start them all at one. In each model iteration, each object grows or shrinks by a random amount ($S_t = g * S_{t-1}$ Where g is a random variable: $-.1 < g < .1$). A final condition of the model is that no object can become smaller than one (If $g*S_{t-1} < 1$ Then $S_t = 1$).

If, from this set of exponential random walks, a sample is drawn at any arbitrary time (of size N), the sample will be Zipf distributed [Gibrat, 1931; Gabaix 1999]. The largest object in any given sample can be expected to have a size approximately equal to the size of the sample (S ~ N) and the distribution is described by $S_n = N*n^{-1}$. Thus, for N = 1000 the size of the largest object (S_1) could be $1000*1^{-1} = 1000$. The size of the next largest object (S_2) would be $1000 * 2^{-1} = 500$. The size of the smallest (S_{1000}) would be $1000 * 1000^{-1} = 1$.

Interestingly, the same distribution arises independent of the initial distribution of sizes and also independent of the range of growth rates. So long as the growth rate is drawn from a range equally distributed around zero, the distribution will converge toward $S_n = N*n^{-1}$, with larger ranges converging faster.

We can make an analogy here to incidents of violence during a conflict. During a "normal" conflict (say a counterinsurgency like Guatemala's), the objective of the repressive force is not directly to kill people. The objective is to put down the rebellion and secure the power of the state. Killing is a means to this end. The state and its agents therefore operate according to heuristic rules under which the level of killing can vary tremendously depending on the situation. Repression according to heuristic rules can be conceived of as similar to the random growth rate model. Since there is no guide to how much killing is the right amount, each incident unfolds according to the goals and perceptions of the two sides.

This is not to say that there was not central control behind the Guatemalan state forces. There undoubtedly was. However, during the "normal" parts of the conflict, the central orders may have taken the form of rules: "Suppress the insurgents", etc. The objective was specified, but the amount of killing required to meet the objective was probably not specified and was thus dependent on the dynamics of the given situation.

The fact that incidents in the non-genocide subset appear to be Zipf distributed is remarkable because it relates the number of incidents to the sizes of incidents in a more direct way than one might think possible. Given the sizes of the largest few events, one can estimate the number of events and the total number killed. Given the number of events, one can estimate the size of the largest events and the total number killed. Given the total number killed, one can estimate the size of the largest events and the number of events. These estimates would be expected to be rough but a rule which would allow even order of magnitude guesswork would be unexpected and might have considerable prognostic power.

Because the distribution is drawn from a single conflict, there is reason to ask whether the Zipf distribution of incident sizes is a common one in conflicts and even more reason to question whether the exponent of -1 could be expected to remain the same in other situations. This is certainly an open question which awaits empirical verification against other data sets. The random growth model analogy does, however, gives us some reason to believe that it might be typical.

Examining the genocide subsample, on the other hand, reveals a different pattern. As discussed above, the distribution of incidents in the western highlands in 1981 and 1982 (where the CEH identified acts of genocide) was quite different from that resulting from the rest of the conflict. There are far more "middle sized" events where between 10 and 100 people are killed. This is consistent with a different kind of command, a much more direct order to go to a place and kill people. Where the basic logic of normal conflict is to accomplish the objective while taking as little risk as possible (which means avoiding incidents if possible), the basic logic of genocide is to kill some fraction (perhaps 100%) of a given population. It is because of this basic difference in command structure that incidents under normal conflict lack a characteristic size and follow the Zipf distribution, while incidents of genocide tend to have a characteristic size that relates to other factors like the size of a military unit or the size of a village.

If this hypothesis proves consistent with data from other conflicts it would provide several potent tools. First, if the conflict was known to be of the “normal” sort, it might be possible to assume that incidents would be Zipf distributed – providing statistical leverage which has previously been unavailable. Second, the distribution of incidents could provide evidence of the nature of the orders and command structure in a conflict, providing a statistical means of differentiating normal and genocidal warfare.

CONCLUSIONS

Examination of detailed data from the Guatemalan conflict between 1977 and 1986 reveals a number of novel patterns which support the use of complex systems methods for understanding civil violence. The lack of strong correlation between individual and larger scale killings within municipalities provides some support for the notion that the removal of leaders is an effective repression technique. A comparison between the amount of killing in municipalities and the ethnic mix in those municipalities reveals a surprising, non-linear relationship between ethnic mix and killing; thereby inviting analysis based on group dynamics and other complex mechanisms. The temporal texture of the conflict is far from smooth, with a power spectrum that closely resembles that of other, better understood, complex systems. The distribution of incident sizes within the data seems to fall into two distinct sets, one of which (corresponding to "regular" conflict) is Zipf distributed and lacks a characteristic size, the other of which includes acts of genocide and is distributed quite differently -- possibly indicating that agents of the state were proceeding under a different type of orders.

Because of the unique nature of the Guatemala data set, all of the findings in this paper need to be considered as preliminary empirical results. It is hoped, however, that the findings are sufficiently provocative to encourage the compilation and release more data sets of this sort. Many aspects of civil violence seem to depend on the internal dynamics of a conflict, and will not be revealed without a careful examination of detailed data from many conflicts.

References:

Bak, Per (1997): *How Nature Works: The Science of Self Organized Criticality*. New York: Copernicus.

Ball, Patrick (1999): *AAAS/CIIDH database of human rights violations in Guatemala* (ATV20.1). http://hrdata.aaas.org/ciidh/data.html (July 1, 2000).

Ball, Patrick, Paul Kobrak, and Herbert F. Spirer (1999): *State Violence in Guatemala, 1960-1996: A Quantitative Reflection.* http://hrdata.aaas.org/ciidh/data.html (July 1, 2000).

Bates, Robert (1983): "Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa." In *State Versus Ethnic Claims: African PolicyDilemmas*, ed. Donald Rothchild and Victor A. Olorunsola. Boulder, CO: Westview.

Doyle, Michael W. and Nicholas Sambanis (2000): "International Peacebuilding: A Theoretical and Quantitative Analysis." Draft (March 7, 2000).

Epstein, Joshua M., John D. Steinbruner, Miles T. Parker (2001): *Modeling Civil Violence: An Agent-Based Computational Approach*. Brookings Institution Center on Social and Economic Dynamics Working Paper No. 20.

Fearon, James D. and David D. Laitin (1996): "Explaining Interethnic Cooperation," in *American Political Science Review* 90:4, December 1996.

Gabaix, Xavier (1999): "Zipf's Law for Cities: An Explaination," *Quarterly Journal of Economics* CXIV, 739.

Gibrat, R. (1931): *Les Inégalitiés Économiques* Paris: Librarie du Recueil Sirey.

Gurr, Ted Robert and Barbara Harff (1996): *Early Warning of Communal Conflicts and Genocide: Linking Empirical Research to International Responses.* United Nations University Press.

Guatemalan Commission for Historical Clarification (CEH) (1999): *Guatemala: Memoria del Silencio.* http://hrdata.aaas.org/ceh/ (July 1, 2000).

Schroeder, Manfred (1991): *Fractals, Chaos, Power Laws.* New York: W. H. Freeman and Company.

Steinbruner, John D. (2000): *Principles of Global Security.* Washington, DC: Brookings.