Spatial and temporal patterns in civil violence
Guatemala, 1977–1986

Timothy R. Gulden
Center on Social and Economic Dynamics
The Brookings Institution
1775 Massachusetts Avenue, NW
Washington, D.C. 20036
timgulden@yahoo.com

ABSTRACT. Civil violence is a complex and often horrific phenomenon whose characteristics have varied by era, setting, and circumstance. Its objective analysis has rarely been feasible at spatial and temporal scales great enough and resolutions fine enough to reveal patterns useful in prevention, intervention, or adjudication. An extraordinary data set simultaneously meeting scale and resolution criteria was collected during conflict in Guatemala from 1977 through 1986. Reported here is its spatial-temporal analysis; reported as well is a putatively novel method for estimating power-law exponents from aggregate data. Analysis showed that the relationship between ethnic mix and killing was smooth yet highly nonlinear, that the temporal texture of killings was rough, and that the distribution of killing-event sizes was dichotomous, with nongenocidal and genocidal conflict periods displaying Zipf and non-Zipf distributions, respectively. These results add statistical support to claims that the Guatemalan military operated under at least two directives with respect to killing and that one of these effected a genocidal campaign against an indigenous people, the Mayans. Implications for group-behavioral modeling, conflict prevention, peace-keeping intervention, human-rights monitoring, and transitional justice are noted.

Scholars studying civil violence quantitatively have for the most part examined country-level summaries of annual conflict deaths.[1,2,3,4] While various inferences may legitimately be drawn from such country-year summaries, a conflict's internal dynamics cannot be found reflected there. This paper looks for civil-violence patterns in data collected at higher resolution — by municipality and month — in Guatemala from 1977 through 1986.

The aim is not a comprehensive statistical, political, or historical portrait of a particular conflict, a task in this case ably undertaken by others.[5,6,7] Instead, the aim is to discover and characterize spatial and temporal dynamics whose quantitative modeling might improve understanding and management of civil violence — and of state-sponsored violence against civilians.

A conflict in Guatemala

State repression in Guatemala has a long and exceptional history, and the perpetrators, accomplices, opponents, and victims of this repression were and are particular people with personal stories all their own. The conflict they shared in recent times lasted from 1960 to 1996, with a period of greatly heightened violence in the early 1980s. Most killings — the Guatemalan Commission for Historical Clarification (CEH) has estimated over 93 percent[8] — were carried out by the state as part of an ongoing campaign of repressive terror involving the military, the police, semi-autonomous "death squads," and state-organized "civil patrols."[9]

Ethnicity was a factor. Early on, violence typically involved middle-class members of the nonindigenous Ladino group struggling amongst each other for control of the government. This urban conflict, first centered in the capital, Guatemala City, later evolved into a rural insurgency and counter-insurgency. Thus, the targets of state repression shifted, around 1981, from middle-class Ladino dissidents to indigenous highland peasants, essentially all Mayans, suspected of aiding rebel groups. The scale and nature of the conflict changed as well, becoming deadlier and increasingly

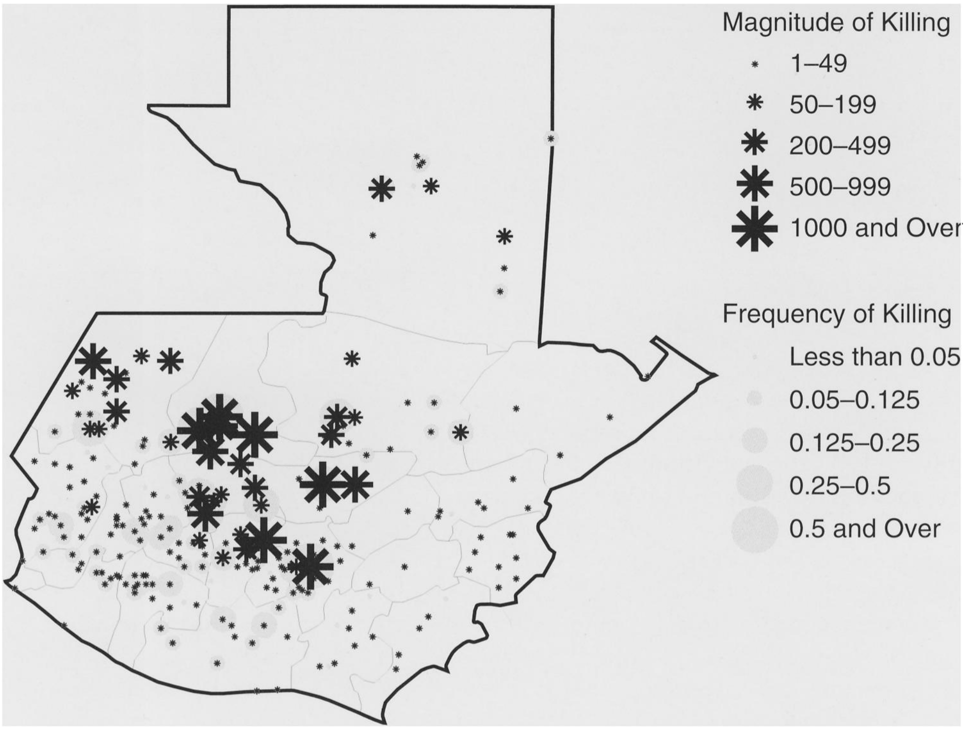

Figure 1. Map of frequency and severity of killing, by municipality.

genocidal. Overall, about 83 percent of the victims were Mayan.[10]

We should note that the dichotomous division of ethnicity into Ladino and Mayan must have been clearer to the Ladinos controlling the government than to members of Mayan groups, who themselves spoke many different languages and did not consistently 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 was based on 36 years of press reports and over 5,000 interviews, and it lists killings and disappearances totaling 37,255 from 1960 through 1996, often including, among other details, site and time.[11] While several comparable data sets do exist, including one for El Salvador, only this one from Guatemala is published and generally available for research.

The research reported here used a subset of these data, one restricted to the ten years from 1977 through 1986 and to killings and disappearances whose dates were known to at least the nearest month. This subset contained 22,559 cases, an indeterminate fraction of the ten-year grand total estimated variously at 80,000 to 400,000.

Methods

Quantitative studies of large-scale violence have typically relied upon summary statistics covering large spans of space and time, such as a nation or conflict zone over a year or for the duration of conflict. Many of these studies have used linear regression and related statistical techniques to correlate violence with other factors.

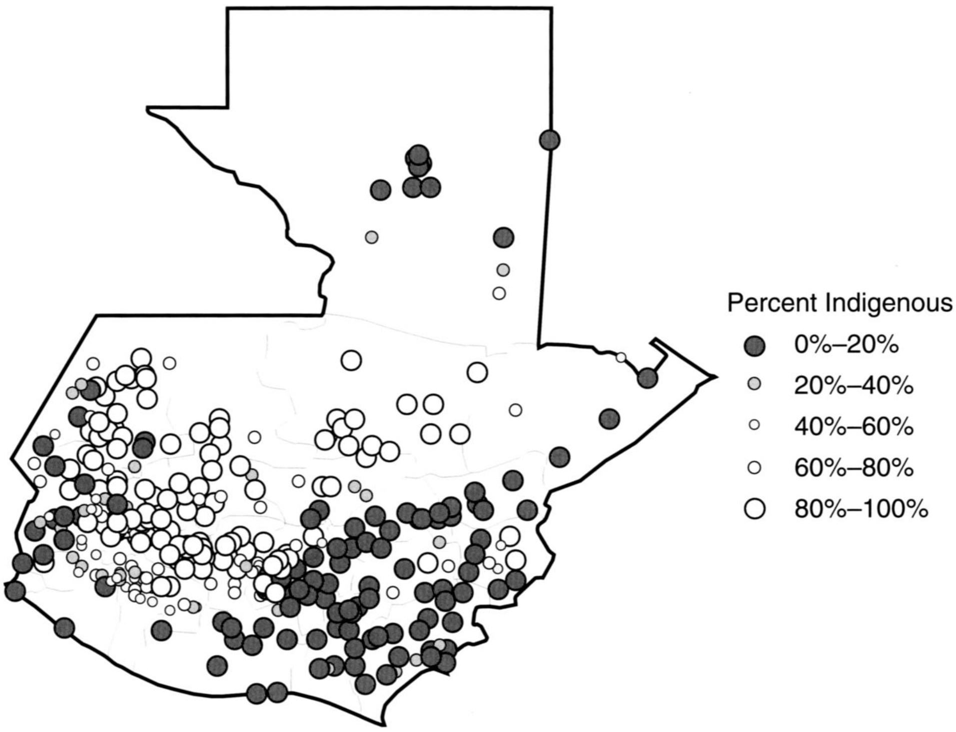

Figure 2. Map of ethnic distribution in municipalities.

This paper details a different approach, one intended to preserve complexity wherever possible while exploring finer-grained data for nonobvious regularities. Major tools in this effort have included complex querying in Structured Query Language (SQL), spatial analysis and mapping in a geographic information system (GIS), analysis of histograms, time-series plots, rank-size (Pareto) plots, and other, mostly graphical, representations of disaggregated data. The analysis of rank-size plots presented here includes a putatively novel method for estimating from aggregated data a power-law exponent for disaggregated data.

This approach has had its limitations. In classical statistical analysis, hypotheses are formed independently of the data in whose context they will be tested. In this case, though, an examination of available data was used to construct hypotheses, making it impossible to employ these same data in a testing mode. The observations which followed cannot then be offered as generalizations but as theoretically plausible ideas deserving challenge before confirmation.

Results

Frequency versus severity. The coefficient of correlation between frequency of killing (number of months where at least one person was killed in a town) and quantity of killing (number of people killed in the town over the whole study period) was 0.65. The correlation between the number of people killed individually and the number of people killed in groups of two or more was 0.31. Many of those killed individually were personally targeted by the government,[12] suggesting a campaign to eliminate an anti-government leadership.[13, 14]

Figure 1 shows the amount of killing which took place in a municipality (black stars) and the frequency of killing in a municipality, as represented by the proportion of the 120 months of the study period when at least one person was killed (gray circles). Note that the northwestern highlands included many places with a great deal of killing (large stars) but a low frequency of killing incidents (small circles). The southwestern

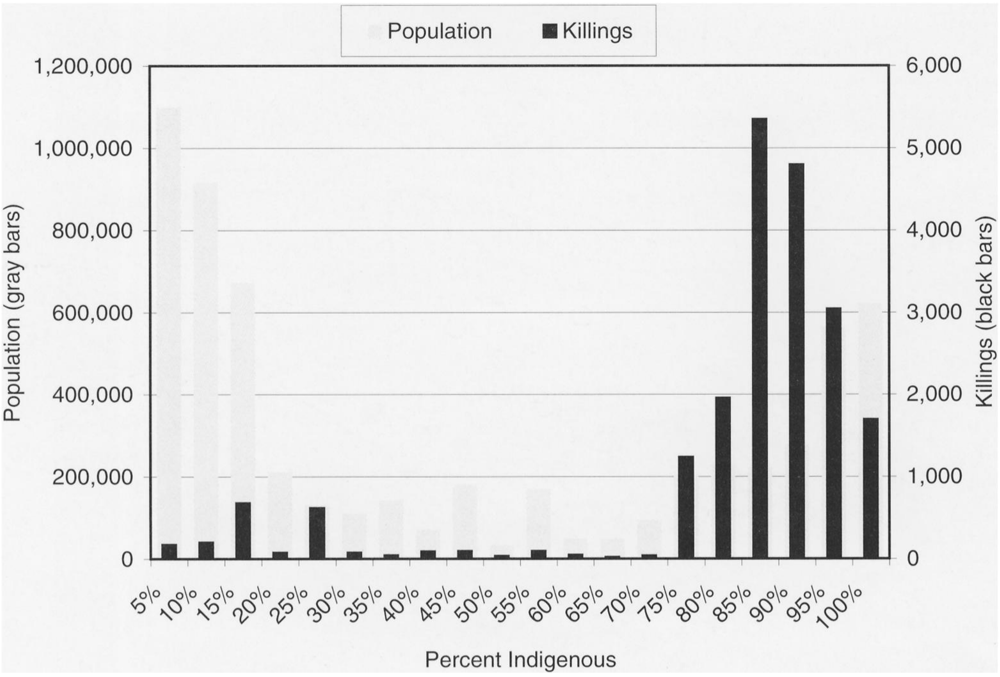

Figure 3. Ethnicity and killing.

region, in contrast, had lower numbers of killings (small stars) but a higher frequency of killing incidents (large circles).

Ethnic mix. The amount of killing in a municipality was correlated with its ethnic mix.

Note, in Figure 2, a paucity of the smaller circles indicating mixed communities. The distribution of ethnic population among municipalities was highly polarized and the distribution U-shaped, as suggested by the gray bars in the histogram presented as Figure 3.

Mayans were living largely in the mountainous northwest, while Ladinos occupied the lower and agriculturally more productive southern and eastern portions of the country. Even within these Ladino-dominated regions, however, there was considerable polarization. Just over half of all killings took place in municipalities in which Mayans made up between 80 and 90 percent of the population, though such municipalities made up less than 8 percent of all municipalities and were home to just over 8 percent of the total population and about 17 percent of all Mayans. Many more Mayans — 45 percent — lived in municipalities in which they constituted at least 90 percent of the citizenry. Though these municipalities saw considerable bloodshed, they did not have as much violence as those that were 80 to 90 percent Mayan. A similar, though much smaller, peak appeared in municipalities that were 10 to 25 percent Mayan.

Thus, at both ends of this histogram, where one group or the other represented more than 90 percent of a local population, and also in the middle, where neither group was more than 75-percent dominant, killings of Mayans, the principal victims, were fewer.

Continuity. Violence in a given place did not expand or contract smoothly over time as one might assume from looking at the annual data presented in the top graph of Figure 4. Rather, the pattern of violence was "spiky." This quality increased as the level of spatial and temporal disaggregation increased, as shown in the middle and bottom graphs of Figure 4.

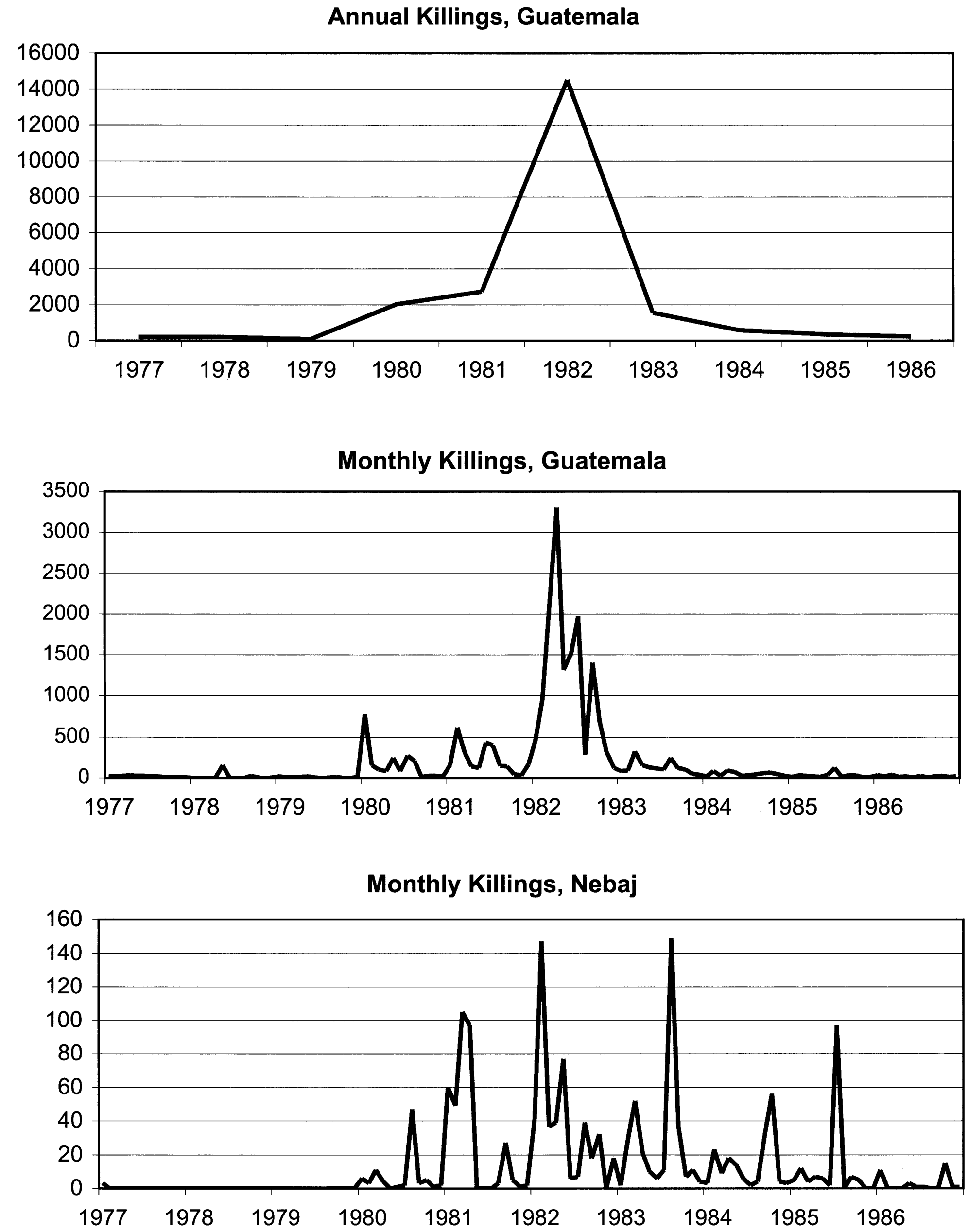

Figure 4. Killings in time series.

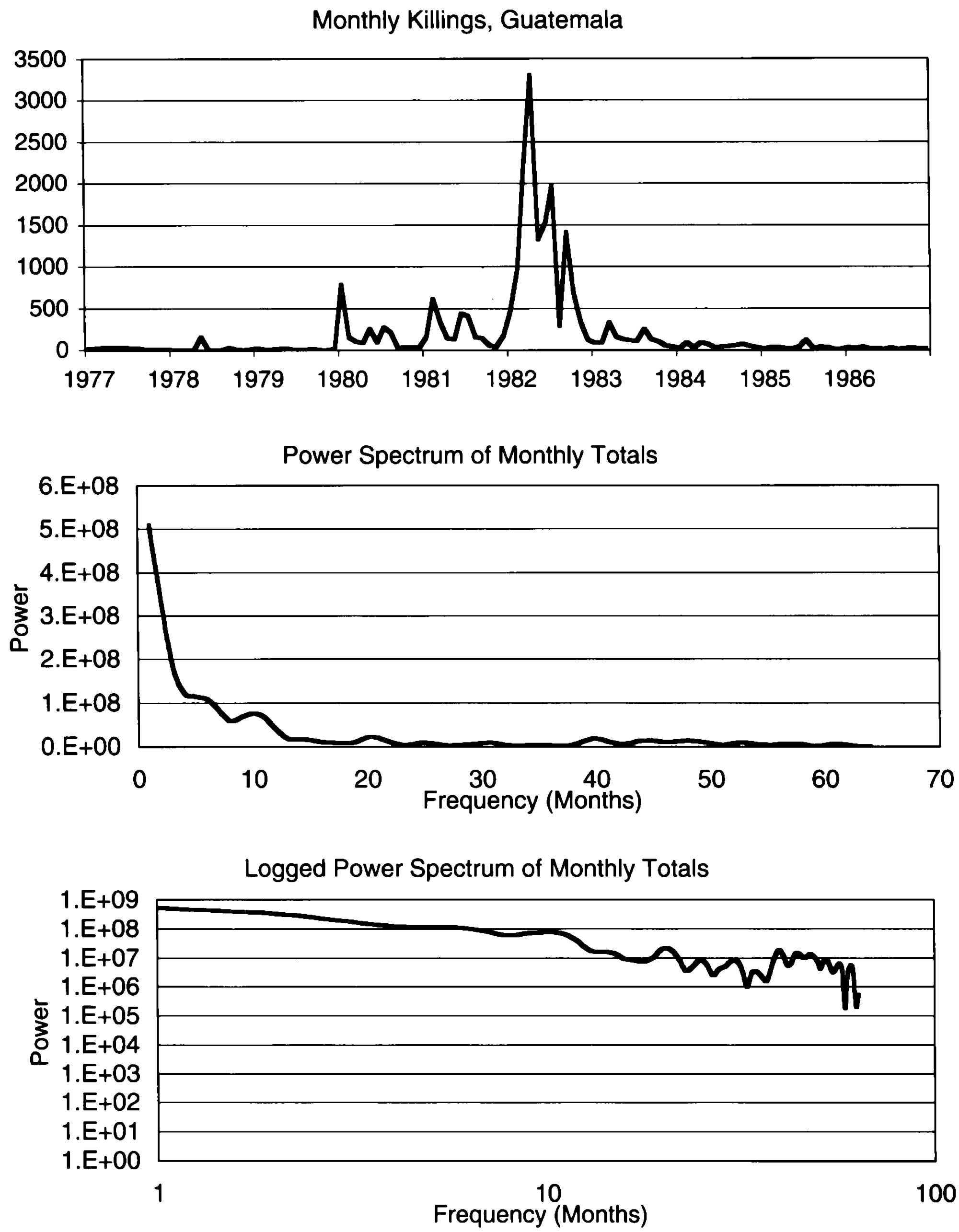

Figure 5. Killings by month, with power-spectrum and logged power-spectrum plots.

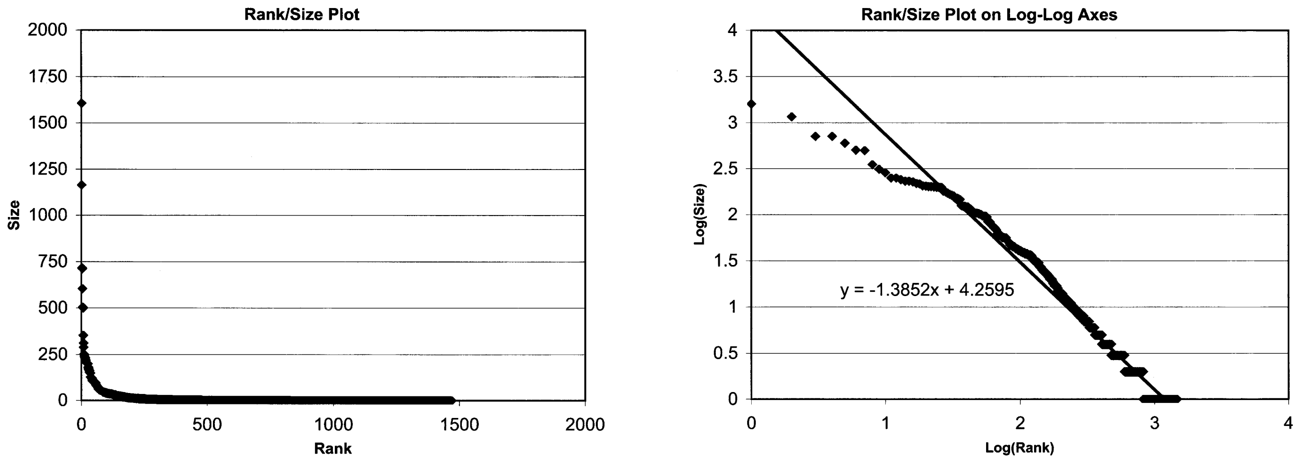

Figure 6. Rank-size plot of killings by municipality-month on normal and log-log axes.

An objective measure of the character of such a time-series data set can be obtained by examining its power spectrum, as might be done in physics to characterize waves, such as sound waves. Mathematical techniques exist, most notably the Fourier transform, to decompose a given signal into a spectrum of sine waves of different amplitudes — different "powers" — which can be combined to reproduce that signal. This technique has been extended to look at time-series phenomena ranging from earthquakes and floods to stock prices. Purely random noise, also known as "white" noise, has equal power at all frequencies. Complex systems, however, often produce "pink" noise, also called "1/f noise," where power at a given frequency is inversely proportional to that frequency.[15] A Fourier transform of the time series of monthly killings reveals just such a power-law spectrum. Figure 5 shows killings by month, along with power-spectrum and logged power-spectrum plots of the killings-by-month time series. The spectrum's exponent is not precisely –1, as might be expected since "1/f" written exponentially is "f^{-1}." Yet it is still close enough, at –1.4, to suggest the signature of pink noise.[16]

Distribution of incident sizes. The conflict under study can be separated into two phenomena: a nongenocidal counterinsurgency and a genocidal campaign, which latter was perpetrated mostly in the western highlands and mostly in 1981 and 1982. The nongenocidal counterinsurgency exhibited a Zipf distribution, a nonrandom scaling observed in many natural and social systems, including chaotic ones, first described by George Kingsley Zipf (1902–1950), an American statistician (see Appendix). The genocidal campaign exhibited a different pattern.

A general sense of the overall distribution of incident sizes is conveyed by the rank-size plot as in Figure 6.

With painstaking thoroughness, the CEH documented a number of incidents meeting formal criteria as genocide. All involved massacres of Mayans in the highlands in 1981 or 1982. Figure 7 depicts the spatial distribution of these events.

Four highland departments were the focus of the government's Mayan genocide. They are shaded gray in Figure 7. Records for these departments during 1981–2 (the genocidal-killings subset), were analyzed separately from and compared to all other records (the nongenocidal-killings subset), as in Figure 8.

The nongenocidal-killings subset included 1,133 municipality-months, while the genocidal-killings subset included 338 municipality-months. When the nongenocidal subset is presented on a rank-size plot on log-log axes it closely approximates a straight line with slope –1.13. In other words, the distribution could be described by a power law of the form $S = \alpha R^{-1.13}$, where S is size and R is rank. The genocidal-killings subset, in contrast, is concave toward the origin and more steeply sloped.

Though these figures at first appear similar, they are

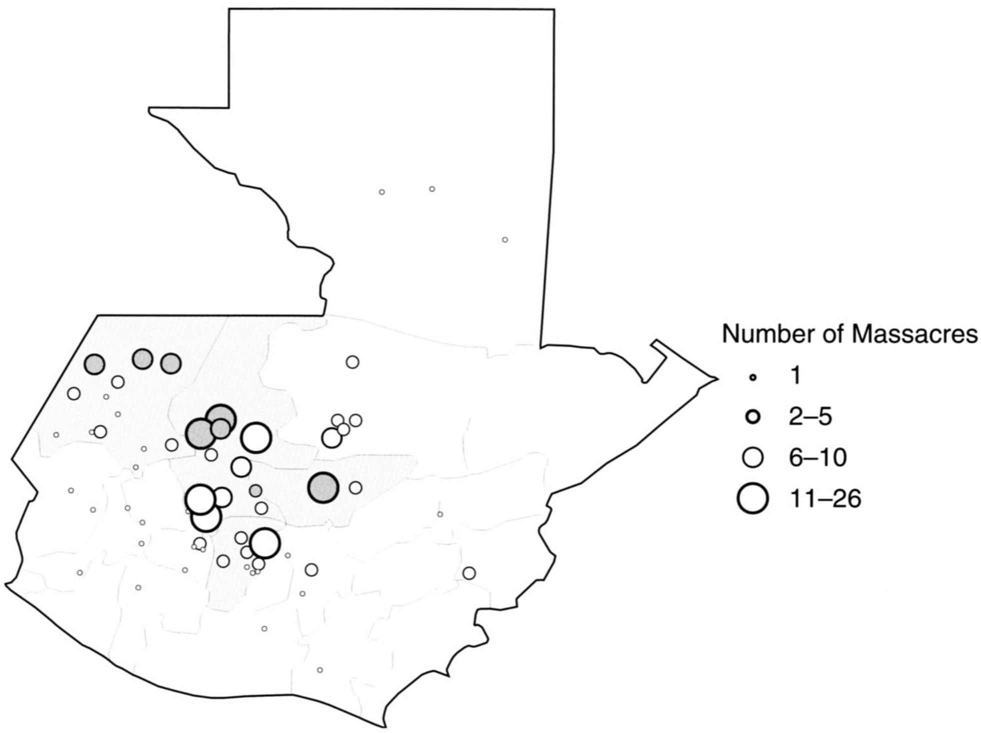

Figure 7. Municipalities with confirmed genocidal massacres (dark circles) and municipalities with massacres for which genocide status was not determined by CEH (light circles), 1981–2.

significantly different, as becomes apparent when analyzed in parts. The top part of the nongenocidal distribution has a slope of –1.01 with a 95-percent confidence interval of –0.99 to –1.03, while the bottom part has a slope of –0.98 with a 95-percent confidence interval of –0.95 to –1.01. Since these confidence intervals overlap, the top and bottom sections cannot be said to have different slopes. In the genocidal distribution, however, the top portion has a slope of –1.00 with a 95-percent confidence interval of –0.95 to –1.05, while the bottom portion has a slope of –2.61 with a 95-percent confidence interval of –2.55 to –2.66. The top and bottom sections of the genocidal distribution do distinctly have different slopes. The genocidal distribution, therefore, can legitimately be said to be concave toward the origin and poorly approximated by a power law.

The nongenocidal distribution can further be analyzed to obtain the power-law exponent which would obtain if the data could completely be disaggregated to the incident level. The data were collected at the level of the reported violation, rather than the incident. While it made possible the recording of a precise number of killings for each municipality month, this practice did not allow reliable numbering by incident. The data as received are sufficient, however, to produce an estimate of about 3,500 total incidents within the 1,133 municipality months of the nongenocidal killings subset. If we think of the municipality-month as an aggregation bin, then the nongenocidal-killings subset represented an average of 3.05 incidents per municipality-month. By further aggregating the data temporally at 6 months, 1 year, 2 years, 5 years, and 10 years and determining the power-law exponent for the bins at each level of aggregation, we are able to establish a linear relationship between incidents per municipality-month and the power-law exponent.

This relationship is nicely described, with R^2 = 0.986, by the linear relationship $y = -0.7x - 0.929$, where y is the exponent and x the average number of incidents per aggregation bin. From this empirically derived relationship we can estimate the exponent for

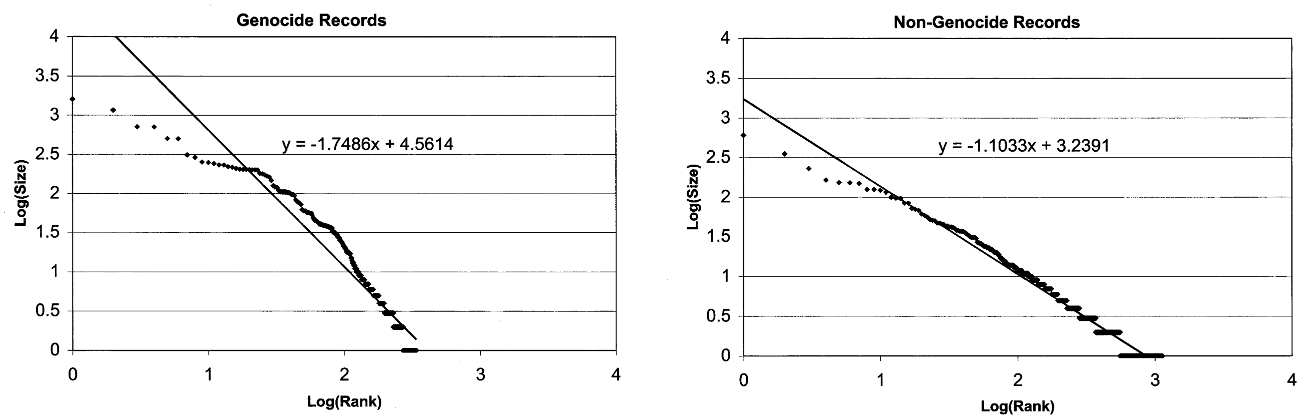

Figure 8. Rank-size plots of the nongenocidal-killings subset and the genocidal-killings subset, by municipality-month.

the fully disaggregated case, with only one incident per aggregation bin, by simply evaluating the expression at $x = 1$. The resulting value, –1.056, is close to –1, the exponent which defines the Zipf distribution. While it seems reasonable that this relationship would be linear — and, empirically, it does appear to be linear — the mathematical justification for this linearity assumption awaits further work.

Discussion

If patterns such as these are the textures of civil violence, what might be said of the larger fabric?

The ethnic mix histogram showed the greatest number of killings in municipalities with predominantly Mayan populations, but fewer killings in the most heavily Mayan areas. This pattern might have been produced by an attempt to clear Mayans from mixed Mayan-Ladino areas but a disinclination to do the same, to attempt "ethnic cleansing," in the Mayan heartland. The ethnic mix histogram also showed a small rise in killings where Mayans were between 10 and 25 percent of the population. Minority populations of such proportions might somehow have threatened or annoyed a local Ladino majority without effectively restraining it. Members of the smallest Mayan populations were safer; they might typically have been individual servants or laborers, with or without families, well known to Ladino employers.

These interpretations are tenuous. They apply to a histogram homogenizing data from nongenocidal and genocidal periods of a long conflict. Further insight would be gained by reexamining these data, if feasible, as a time series. Agent-based modeling, which is well suited to the exploration of group dynamics, might also be used, specifically to test whether a bimodal pattern, such as the one found here, could emerge from a set of simple set of rules for interaction. Comparison to similar data sets from other conflicts in other countries would likewise be helpful, should such data sets become available.

More robust, though less intuitive, was the power-law analysis, which showed that killings in the part of the conflict determined by forensic methods not to have been genocidal followed a Zipf distribution while those in the genocidal part did not. What might have explained this difference, and what might now be learned from it?

During a "normal" counter-insurgency, the objective of repressive force is not directly to kill people. The objective is to put down rebellion and secure the power of the state. Killing is a tactic, 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 considered similar

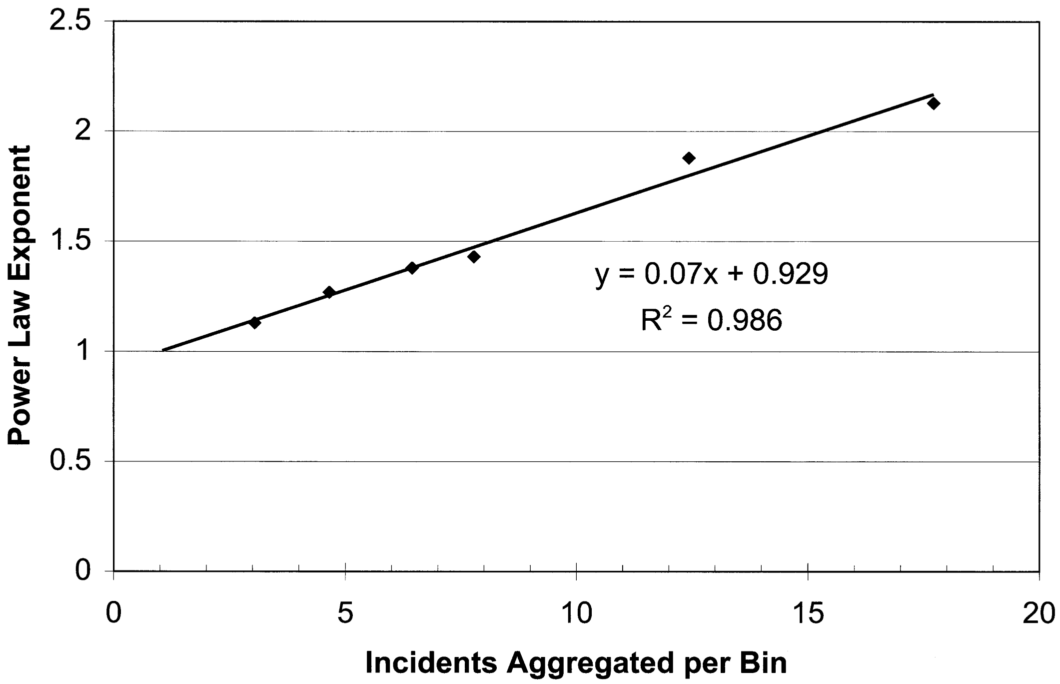

Figure 9. Trend of the power-law exponent for different levels of aggregation in the nongenocidal subset.

to a random-growth-rate model (see Appendix). Since there is no guide to how much killing is the right amount, each incident unfolds according to the goals and perceptions of the sides engaged.

This is not to deny there was central control of the state's counterinsurgency forces. However, the amount of killing required for those forces to meet a nongenocidal objective would have varied according to the dynamics of a given situation.

The finding that incidents in the nongenocidal-killings subset appear to have been Zipf-distributed is remarkable, since it relates the number of incidents more directly to their sizes than might have been thought possible. Given the sizes of the largest 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. It should be noted, however, that the upper end of the distribution in the growth-rate model is much less stable than the lower end — making extrapolation from large events difficult.

Is Zipf distribution of incident sizes commonplace in civil violence? This is a question awaiting empirical verification using other data sets. The random-growth-model analogy does, however, give us some reason to believe it might be.

The genocidal-killings subset, on the other hand, has revealed a different pattern. There were far more "middle-sized" events with 10 to 100 people killed. This pattern suggests a different objective, a different instruction, a more direct order to go to a place and kill people. Killing, in the genocidal case, has shifted, has *been* shifted, or has simply evolved from a tactic to a strategy. Where the basic logic of normal conflict is to accomplish an objective while minimizing risk, which means avoiding incidents if possible, the logic of genocide is to kill the largest feasible fraction — even 100 percent — of a targeted population. Thus, perhaps, nongenocidal killings would lack a characteristic size but come instead to resemble a Zipf distribution, while genocidal killings would occur in sizes derived, roughly, from the numerical strength of a military unit or the population of a village.

The patterns presented here, if supported by other research, could have implications for group-behavioral modeling, conflict prevention, peace-keeping intervention, human-rights monitoring, and transitional justice. These patterns support the notion that behavioral modeling of civil violence using complex-systems methods may be a productive line of policy research.

The author thanks the Brookings Institution Center on Social and Economic Dynamics for providing economic and intellectual support for this research.

APPENDIX. 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 ($S \geq 1$). 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 and $-0.1 < g < 0.1$). A final condition of the model is that no object can become smaller than one, so that if $g * S_{t-1} < 1$ then $S_t = 1$. The result is a collection of exponential random walks bounded by a floor of one.

If, from this set of exponential random walks, a sample of size N is drawn at any arbitrary time, it will be Zipf-distributed. The largest object in any given sample can be expected to have a size approximately equal to the size of the sample ($S \approx N$), and the distribution is described by $S_n = N * n^{-1}$. Thus, for $N = 1000$ the size of the largest object (S_1) would be expected to 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 $1000 * 1000^{-1} = 1$. 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.

References

1. Robert Bates, "Modernization, ethnic competition, and the rationality of politics in contemporary Africa" in *State Versus Ethnic Claims: African Policy Dilemmas,* Donald Rothchild and Victor A. Olorunsola, eds. (Boulder, CO: Westview, 1983).

2. Michael W. Doyle and Nicholas Sambanis, "International peacebuilding: A theoretical and quantitative analysis" (draft paper, March 7, 2000).

3. James D. Fearon and David D. Laitin, "Explaining interethnic cooperation," *American Political Science Review* 90:4, December 1996.

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

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

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

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

8. CEH, 1999.

9. Ball, Kobrak, and Spirer, 1999.

10. CEH, 1999.

11. Christian Davenport and Patrick Ball, "Views to a kill: Exploring the implications of source selection in the case of Guatemalan state terror, 1977–1995," *Journal of Conflict Resolution,* June 2002, 46:3, 42–50.

12. CEH, 1999.

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

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

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

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

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

→ Xavier Gabaix, "Zipf's law for cities: An explanation," *Quarterly Journal of Economics,* CXIV, 739.