LICOS Centre for Institutions and Economic Performance
Centre of Excellence

LICOS Discussion Paper Series

Discussion Paper 256/2010

The intensity of the Rwandan genocide: Fine measures from the gacaca records

Marijke Verpoorten

Katholieke Universiteit Leuven

LICOS Centre for Institutions and Economic Performance
Huis De Dorlodot
Deberiotstraat 34 – mailbox 3511
B-3000 Leuven
BELGIUM

TEL:+32-(0)16 32 65 98
FAX:+32-(0)16 32 65 99
http://www.econ.kuleuven.be/licos

The intensity of the Rwandan genocide: fine measures from the gacaca records[*]

Marijke Verpoorten[†]

University of Leuven

March 2010

Abstract

This article demonstrates how fine continuous and categorical measures of genocide intensity can be derived from the records of the Rwandan transitional justice system. The data, which include the number of genocide suspects and genocide survivors across 1484 administrative sectors, are highly skewed and contain a non-negligible number of outlying observations. After deriving nine proxies of genocide intensity from the data, various sets of these proxies are subjected to skewness-adjusted Robust Principal Component Analysis (ROBPCA), yielding four distinct continuous indices of genocide intensity. The effect of survival bias on these indices is reduced by augmenting the set of genocide proxies subjected to ROBPCA with the distance from an administrative sector to the nearest mass grave. Finally, the administrative sectors are divided into distinct categories of low, moderate and high genocide intensity by means of Local Indicators of Spatial Auto-Correlation (LISA) that allow identifying significant high-high and low-low clusters of genocide intensity.

1 Introduction

The micro-level research on armed conflict has exploded over the past decade. Besides a steady increase in the number of studies, there have been considerable improvements in

[*]The data set will be made available upon publication of the article

[†]Research scholar of the fund for Scientific Research - Flanders, Belgium (FWO), Center for Economic Studies, Centre for Institutions and Economic Performance LICOS, KULeuven. e-mail: marijke.verpoorten@econ.kuleuven.be

methodology. In particular, scholars have increasingly devoted attention to identifying rich micro-level measures of conflict intensity (e.g. Restrepo, Spagat, and Vargas (2006), Raleigh and Hegre (2005)). This is no coincidence because the identification of micro-level causes and consequences of armed conflict stands or falls with the conflict intensity measure used. This article aims to promote the use of rich micro-level conflict intensity measures in two ways. First, it provides easy access to the data released by the gacaca courts, i.e. the transitional Rwandan justice system in charge of judging 1994 genocide suspects. Second, it presents fine continuous and categorical measures of genocide intensity, which are derived from the gacaca data using spatial autocorrelation analysis and recent advancements in principal component analysis that are well suited for highly skewed data with outlying observations.

The 2005-2007 gacaca data released include 5 types of information: (1) the number of accused persons living in the country; (2) the number of genocide survivors living in the country; (3) the number of accused persons who are not living in the country; (4) the number of persons who committed genocide and who passed away. The two latter types are only available at the district level. The first two types of information are available for 1484 sectors, which are, after the cell, the lowest codified administrative unit in Rwanda. The data was released in 2007 in pdf format on the website of gacaca[1]. After converting the data into spreadsheet format, we subject them to a critical examination. In particular, we evaluate their overall reliability through a comparison with data from other sources, including the number of persons imprisoned (Office of the Prosecutor (2002)), an estimate of the number of perpetrators by Straus (2004), and a 2006 census of genocide survivors (Government of Rwanda (2008)). Such a critical examination is required because, as gacaca proceeded, its operation was criticized for lack of objectivity due to political manipulation (Longman (2009), Pitsch (2002), Wolters (2005)).

After this overall data quality check, we transform the data in several ways to obtain rich measures of genocide intensity. We proceed in four steps. First, we combine the data with information on 1994 sector level population size in order to derive genocide proxies, e.g. the number of genocide suspects as a proportion of the 1994 population. Second, we detect outliers by means of skewness-adjusted box-and-whisker diagrams (Hubert and der Veeken (2008)). Third, we subject different sets of genocide proxies to skewness-adjusted Robust Principal Component Analysis (ROBPCA), proposed by Hubert, Rousseeuw, and Verdonck (2009), and retain the first PCs as indices of genocide intensity. At this point,

[1] http://www.inkiko-gacaca.gov.rw/

we make a correction for survival bias, by augmenting the set of genocide proxies with the distance of a sector to the nearest mass grave. Finally, we revert to Local Indicators of Spatial Association (LISA) in order to identify sectors belonging to spatial high-high clusters and low-low clusters (Anselin (1995)). In this way, we obtain a non-arbitrary categorization of the sectors according to genocide intensity. Such categorization may be useful given that categorical variables, in particular dummies, are preferred for some applications, e.g. summary statistics across conflict intensity or interaction effects in a regression analysis.

Given that the proposed genocide intensity measures are at the sector level, they can be matched with existing nationally representative household surveys that use the sector as a sample unit, e.g. the Demographic and Health Surveys (DHS 1992, 2000, 2005) and the Integrated Household Living Conditions Surveys (IHLCS, 2000/2001 and 2005/2006). Combining these household surveys with a fine measure of genocide intensity can mean an important step forward in the research on the causes and consequences of the Rwandan genocide. So far, there have been a number of empirical micro-level studies on both the causes and consequences of the Rwandan genocide, but to the best of our knowledge, only Yanagizawa (mimeo) has used the gacaca data described in this article. The transformation of the data presented here goes at least four steps further, by (1) using robust techniques for outlyingness in skewed data, (2) correcting for survival bias, (3) providing other measures of genocide intensity, besides participation, e.g. measures that take into account excess mortality amongst Tutsi, and by (4) deriving categorical measures of genocide intensity in a non-arbitrary way.

Section 2 provides an overview and a first quality check of the data. Section 3 defines proxies of genocide intensity and detects outliers. Section 4 constructs genocide indices by subjecting different sets of genocide proxies to skewness-adjusted ROBPCA. Section 5 derives a categorical variable for genocide intensity using LISA. Section 6 concludes.

2 The available data

2.1 Overview

In 2005, the gacaca courts were stepping in the first phase of their activities, i.e. the phase of collecting information. During weekly sessions with compulsory attendance of all community members, lists were made of victims, suspects and survivors[2]. Part of the results achieved

[2] Attendance was initially voluntary, but after problems with low attendance in the pilot phases, the law was revised, making attendance compulsary (Longman (2009)).

during this phase were made public in the course of 2007. The released sector level data include the number of genocide suspects in a sector, classified in three groups, and the number of genocide survivors, classified in five groups[3].

- Genocide suspects
  - Category 1: accused of planning, organizing or supervising the genocide, and committing sexual torture
  - Category 2: accused of killings or other serious physical assaults
  - Category 3: accused of looting or other offences against property
- Genocide survivors
  - Widowed
  - Orphaned
  - Disabled
  - Male
  - Female

The first category of alleged genocide perpetrators has to be referred to national criminal courts. The gacaca courts are charged with judging the two remaining categories. However, if a third category offender and the victim have agreed on an amicable settlement, the offender is no longer prosecuted by the gacaca court.A person cannot be classified in several categories at the same time, therefore if someone stole (Category 3) but also killed (Category 2), he is classified in the higher category (Category 2). Widowed, orphaned and disabled survivors may overlap with either male or female survivors, which include persons old enough to testify. In general, these survivors are Tutsi although it is not excluded that Hutu, related to Tutsi by inter-ethnic marriage are also considered as genocide survivors.

2.2 Reliability

The first column of Table 1 gives the nationwide total number of suspects and survivors. The sum of category 1 and 2 suspects is close to 510,000. Given that on average 20% of suspects are acquitted, this would mean that category 1 and 2 count approximately 400,000

[3] The exact legal definitions of the suspect categories can be found in appendix.

genocide perpetrators. But, adding about 100,000 perpetrators who passed away by 2005, their number increases again to about half a million (Government of Rwanda (2005)). This implies an active participation to the genocide of almost 20% of the adult Hutu population in 1994, or 40% of the adult male Hutu population[4].

Is this a plausible figure? Compared to the work of Straus (2004), who puts forward an estimate of 175,000 to 210,000 perpetrators, this is at the high end. Straus (2004) underpins his estimate with detailed fieldwork in five administrative communes[5] and in-depth interviews with prisoners. From his fieldwork and interviews he takes a best estimate of 30-35 perpetrators per administrative cell over the course of the genocide and multiplies this with the number of cells in Rwanda in which genocide took place (5,852). Despite the large effort undertaken in collecting first-hand data, it is difficult to assess the reliability of the estimate put forward by Straus (2004) mainly because of a large number of untestable assumtions underlying the estimate. However, the number of genocide suspects emerging from the gacaca is also at the high end compared to the number of detainees and accused persons not detained. In 2000, the government held 109,499 detainees on genocide charges, while the number of accused persons not detained was 49,066 (Office of the Prosecutor (2002), Government of Rwanda (2005)).

According to critics of the gacaca courts, at least three reasons may have caused overreporting of the accused. First, late Human Rights Watch adviser Alison Des Forges argued that the concession programme, which requires the naming of all those who participated along with the accused in return for a lighter sentence, led to a multiplication of names. Second, Longman (2009) claims that, over time, gacaca was undermined by government manipulation, aiming at a conviction of the largest possible number of Hutu in order to exclude much of the Hutu from holding public office. Third, several sources, including the Rwandan government, acknowledge that gacaca became a means of taking personal revenge on enemies, which contributed to the steep rise of the number of accused as gacaca proceeded. On the other hand, most sources evaluating gacaca also acknowledge that individuals may have escaped from accusation due to intimidation of witnesses, including murder or attempted murder of potential gacaca witnesses.

Another way to assess the reliability of the data is by looking at the number of survivors

[4] According to the 1991 census, Rwanda had 2,813,232 citizens between 18 and 54, of which approximately 2,530,000 Hutu. Based on an annual average growth of 3%, the Hutu population in 1994 would have been close to 2,750,000 (Straus (2004), Verpoorten (2005)).

[5] At the time of the genocide, Rwanda counted 145 administrative communes, encompassing on average 44 administrative cells, which are the lowest codified adminsitrative unit in Rwanda.

reported by gacaca. The sum of male and female genocide survivors amounts to approximately 250,000. This is higher than the estimate of 150,000 survivors, based on counting in refugee camps immediately after the genocide (Prunier (1998)). In contrast, it is far lower than the reported 335,718 survivors in the census of survivors executed by the Rwandan government in 2006 (Government of Rwanda (2008)). However, apart from Tutsi living in the country at the time of the genocide, this census also includes Tutsi who escaped ethnic violence in neighboring countries, in particular Congo, as well as Tutsi who came back from living in exile abroad, especially Uganda.

The assessment of the quality of the gacaca data remains tentative, because the alternative data sources referred to are not flawless and comparison with the gacaca data is blurred because different definitions are applied for identifying survivors and suspects. In any case, the above discussed reasons for over- and under-reporting of accused urge for a cautious interpretation of the gacaca data. In this respect, it is noteworthy that for the purpose of constructing a genocide intensity index on a less to more scale, over- and under-reporting only matters to the extent that they are nonrandomly distributed across sectors.

3 Genocide Intensity Proxies

3.1 Definition

We define sector level genocide intensity as the death toll of genocide in a sector relative to the sector's population size. The gacaca data do not allow calculating the death toll directly. However, we assume that the available data on alleged genocide participation as well as survival of Tutsi relatives (e.g. widowed and orphaned genocide survivors) provide valuable information on genocide intensity[6]. In order to calculate the number of accused and the number of survivors proportional to 1994 population size, we match the gacaca data with the 1991 population census and calculate 1994 population size, projecting forward from the 1991 population census using 1978-1991 commune level population growth rates[7].

We retain the proportions of the 1994 population belonging to category 1 and 2 genocide suspects as our first two genocide intensity proxies, denoted respectively by $GE1$ and $GE2$. Table 1 lists average shares of respectively 1.1% and 6.2%. "Proxies" is an appropriate designation because sectors with similar levels of $GE1$ and $GE2$ may have experienced different

[6] For examples of indirect mortality estimates from the survival of close relatives, we refer to Hill and Trussell (1977).

[7] Sector level 1978-1991 population growth rates are not available. Communes are one unit higher up the ranking of the administrative subdivision.

levels of genocide intensity. The reasons are threefold. First, category 2 is an aggregation of different accusations, including murder, attempted murder, and involuntary murder (forced participation). The weight of each of these accusations may differ across sectors. Second, the number of innocent people accused may differ across sectors. Finally, the average number of victims killed per perpetrator may vary across sectors.

Other proxies may capture some of this remaining variation. To start with, a large number of category 3 suspects may, ceteris paribus, point to a large passive participation to the genocide. Passive participation, or - put otherwise - low resistance to the genocide may have increased the average number of victims per killer. Table 1 lists an average share of 4.6% category 3 suspects, from now onwards referred to as the third genocide intensity proxy, $GE3$.

The number of category 1-3 survivors, i.e. respectively widowed, orphaned and disabled survivors are indicators of excess mortality among the targeted population. Taken proportional to 1994 population size, we refer to them as genocide intensity proxies 4-6 ($GE4$, $GE5$ and $GE6$). The sector level mean of these proxies equal respectively 0.4%, 1.1% and 0.2%. These figures are very low because, on average, Tutsi accounted for less than 10% of the 1994 population.

For some purposes, it may be more appropriate to take the widowed, orphaned and disabled survivors as a share of the 1994 Tutsi population instead of the 1994 total population[8]. Given that we don't have information on the sector level size of the Tutsi population prior to the genocide, we use the total number of Tutsi who survived as a proxy (the sum of category 4 & 5 survivors). The resulting shares are the final three genocide proxies: $GE7$, $GE8$ and $GE9$, with means of respectively 12.4%, 38.9% and 4.7%. The high proportion of orphans is due to the fact that category 4 and 5 survivors only include those persons old enough to testify in trials.

The standard boxplots of the nine genocide intensity proxies $GE1 - GE9$ are given in Figure 1. All variables have a highly right-skewed distribution. This is in line with the fact that genocide intensity was very unequally distributed across sectors, mainly because the proportion of Tutsi across sectors in Rwanda was very uneven, but also because support

[8] For example, when the interest lies in studying the causes of genocide, genocide intensity is best captured as excess mortality among the targeted population, i.e. Tutsi. In contrast, when interest lies in studying the consequences of genocide for the total population, genocide intensity may best be captured as genocide-induced excess mortality among the total population.

In the former case, genocide intensity can be high even in areas with a very low number of Tutsi provided that the death toll among Tutsi was high. In the latter case, genocide intensity can be high even in areas with a relatively low death toll among Tutsi provided that the proportion of Tutsi in the population was high.

for the genocide from the local administration and civilians varied across communes and provinces (Des Forges (1999)).

3.2 Outliers

The presence of outliers, stemming from real rare events or incidental (systematic) error, amplifies the skewness of the distribution. It has been demonstrated that a high number of outlying observations can results in misleading statistics derived from the data, e.g. the sample mean and variance, making commonly used techniques such as OLS regression analysis and classical Principal Components Analysis (PCA) very sensitive to the presence of outliers (Barnett and Lewis (1993)). Detection of outliers as well as the use of outlier-robust techniques are often required to double-check results. To avoid arbitrariness in labelling extreme values as outliers, we turn to a procedure of outlier detection for skewed data (Hubert and der Veeken (2008)). For normal distributions, standard boxplot like those presented in Figure 1 can be used for detecting outliers. The whiskers of a standard boxplot are given by

$$[Q1 - 1.5IQR,\ Q3 + 1.5IQR],$$

with $Q1$ the first quartile, $Q3$ the third quartile and IQR the interquartile range for a univariate continuous variable $X_n = \{x1, x2, ..., xn\}$. When the original variables are skewed, too many points tend to be flagged as outlying according to the standard boxplot whiskers. In order to identify outliers in skewed data, it is more appropriate to adjust the whiskers to

$$[Q1 - 1.5e^{-4MC}IQR,\ Q3 + 1.5e^{3MC}IQR],$$

with MC the medcouple defined as:

$$MC(X_n) = med_{xi<medn<xj}h(x_i, x_j),$$

$medn$ the sample median, and

$$h(x_i, x_j) = \frac{(x_j - medn) - (medn - x_i)}{x_j - x_i}$$

Using these definitions, we derive the skewness-adjusted whiskers of the genocide proxies $G1 - G9$. The values exceeding these whiskers are identified as outliers. Table 1 gives the values of the skewness-adjusted whiskers and the corresponding number of outliers. On average, we find 11.5 outliers per genocide proxy. In total, 67 sectors have outlying observations for one or more of the nine genocide proxies. An examination of spatial autocorrelation of sectors with outliers shows that they are not the results of spatially correlated over- or under-reporting. More precisely, Moran's I, equalling 0.0025, is not significantly different from zero. Columns 6 & 7 of Table 1 provide the mean and standard deviation of the genocide proxies when excluding the outlying observations. Note that the standard deviations become much smaller.

In the next section we provide summary statistics by province for the genocide proxies, both including and excluding the outlier observations. However, when turning to constructing indices, rather than throwing away observations, we use a method that is robust to the presence of outliers.

3.3 Province level summary statistics

Table 2 provides the province level averages of the genocide proxies $GE1-GE9$, with Panel A including the outlying observations and Panel B excluding them. The pattern that emerges is similar across the two panels. To facilitate spotting provinces with high genocide intensity, for each of the proxies, the four highest values are put in bold. Butare stands out with top-4 values for eight genocide proxies. Kibuye and Gitarama have respectively seven and six proxies with top-4 values. In contrast, Gikongoro only has two proxies with values among the highest four (see Figure 2 for an administrative map of Rwanda).

The fact that Gitarama features higher than Gikongoro across the genocide proxies goes against common knowledge on genocide intensity in these provinces. Gikongoro not only had a larger share of Tutsi in its population (12.8% compared to 9.2%, see Table 3), but Gikongoro is also known for a much worse genocide record than Gitarama. In particular, Gikongoro had several large-scale massacres, whereas Gitarama had many sites of strong resistance of the population against the genocide (Des Forges (1999)). This difference in the unfolding of genocide is reflected by the distance of a sector to the nearest mass grave, which is 7.5 km for Gikongoro and 10.2 km for Gitarama (Table 3). The upside-down ranking of Gikongoro and Gitarama may therefore stem from survival bias, i.e. the genocide proxies are biased downwards in areas where more Tutsi families were completely exterminated. The next section attempts to attenuate survival bias.

4 Genocide indices

4.1 Method

The challenge we face is to aggregate the information embodied in the genocide intensity proxies $GE1 - GE9$ into a meaningful index of excess mortality. To overcome arbitrariness and safeguard maximum variation, several studies have persuasively argued for the use of principal component analysis (PCA) (e.g. Filmer and Pritchett (2001)). PCA has the desirable property of reducing the dimensionality of a data set while retaining maximum variation in the data set. More precisely, from a set of variables, PCA extracts orthogonal linear combinations that capture the common information in the set most successfully. The first principal component (PC) identifies the linear combination of the variables with maximum variance, the second principal component yields a second linear combination of the variables, orthogonal to the first, with maximal remaining variance, and so on [9]. For our objective, i.e. defining an index of genocide intensity, we are interested in the first PC, which will be an appropriate summary of genocide intensity if it captures a relatively high percentage of the total variance present in the genocide proxies set and the "loadings" of that PC have roughly equal values.

PCA relies on maximizing the classical sample variance. Therefore, it is sensitive to outliers. Since we are dealing with highly skewed data that includes a non-negligible number of outlier values, we revert to a recently proposed PCA that is robust to outliers in skewed distributions (Hubert, Rousseeuw, and Verdonck (2009)), referred to as ROBPCA. ROBPCA reduces the effect of outliers by replacing the classical sample covariance matrix used in classical PCA with a robust covariance matrix that is calculated for a subset of data points for which outlyingness is below a predefined threshold value. ROBPCA for skewed data uses the skewness-adjusted whiskers as a benchmark for defining outlyingness (see above).

---

[9] Formally, suppose that x is a vector of p random variables and x^* is a vector of the standardized p variables, having zero mean and unit variance, then the first principal component $PC1$ is the linear function $\alpha_1' x^*$ having maximum variance, where α_1 is a vector of p constants $\alpha_{11}, a_{12}, ..., \alpha_{1p}$ and $\prime$ denotes transpose.

$PC1 = \alpha_1' x^* = \alpha_{11} x_1^* + \alpha_{12} x_2^* + ... + \alpha_{1p} x_p^*$,

Mathematically, the vector α_1 maximizes $var[\alpha_1' x^*] = \alpha_1' \Sigma \alpha_1$, with Σ the covariance matrix of x^*,which corresponds to the correlation matrix of the vector x of the original, unstandardized variables. For the purpose of finding a closed form solution for this maximization problem, a normalization constraint, $\alpha_1' \alpha_1 = 1$, is imposed. To maximize $\alpha_1' \Sigma \alpha_1$ subject to $\alpha_1' \alpha_1 = 1$, the standard approach is to use the technique of Lagrange multipliers. It can be shown that this maximization problem leads to choosing α_1 as the eigenvector of Σ corresponding to the largest eigenvalue of Σ, λ_1 and $var[\alpha_1' x^*] = \alpha_1' \Sigma \alpha_1 = \lambda_1$.

To interpret the PC in terms of the original variables, each coefficient α_{1l} must be divided by the standard deviation, s_l, of the corresponding variable x_l. For example, a one unit increase in x_l, leads to a change in the $1st$ PC equal to α_{1l}/s_l.

For a detailed exposition of principal component analysis we refer to Jolliffe (2002) and Dunteman (2001).

Given the relatively recent introduction of ROBPCA for skewed data, it has not yet been applied for constructing indices of conflict intensity. In contrast, a number of studies have used classical PCA for the purpose of summarizing conflict indicators by a conflict index. Pioneering work by Hibbs (1973) derives indices of "collective protest" and "internal war" from a 108-nation cross-sectional analysis of six event variables on mass political violence. Following Hibbs (1973) a large number of cross-country studies have used an index of sociopolitical instability as an explanatory variable in regressions in which the dependent variable is growth, savings or investment (e.g. Venieris and Gupta (1986), Barro (1991), Alesina and Perotti (1996)). To the best of our knowledge, only one micro-economic study, González and Lopez (2007), uses PCA to summarize variables into a micro level index of violent conflict. This study looks at the effect of political violence in Columbia on farm household efficiency. Five indicators of violence are defined: homicides, the number of attacks by FARC guerrillas, the number of attacks by ELN guerrillas, kidnappings, and displaced population. The first PC accounts for 43% of the joint variance of the five indicators, and is retained as an index of political violence.

4.2 Results

We subject four different subsets of the nine genocide intensity proxies to ROBPCA. Scholars looking into the participation to the genocide will be interested in using the suspect categories. Therefore we start with the sets $[GE1, GE2]$ and $[GE1, GE2, GE3]$. The resulting first PCs explain respectively up to 81% and 76% of the total variation in the underlying set of variables. The corresponding linear combinations are:

$$\begin{aligned} GEI12 &= 0.58\,(GE1) + 0.81\,(GE2) \\ GEI13 &= 0.52\,(GE1) + 0.63\,(GE2) + 0.57\,(GE3) \end{aligned}$$

Studies focusing on the causes and consequences of genocide intensity may benefit from a genocide index augmented with information on excess mortality. Furthermore, enlarging the set of variables subjected to ROBPCA reduces the effect of measurement error and outliers in each of the proxies separately. We subject the augmented variable sets $[GE1, GE2, GE3, GE4, GE5, GE6]$ and $[GE1, GE2, GE3, GE7, GE8, GE9]$ to ROBPCA. The resulting first PCs, explaining respectively up to 58% and 39% of the total variation in the underlying set of variables, are:

$$GEI16 = 0.42\,(GE1) + 0.50\,(GE2) + 0.42\,(GE3) + 0.44\,(GE4) + 0.42\,(GE5) + 0.19\,(GE6)$$

$$GEI1379 = 0.53\,(GE1) + 0.60\,(GE2) + 0.54\,(GE3) + 0.19\,(GE7) + 0.14\,(GE8) + 0.12\,(GE9)$$

The province level averages and rankings of these indices are listed in Table 4. Butare receives the highest rank for each of the four indices, while Kibuye and Gitarama always feature in the top 4 (put in bold). Kibungo and Gikongoro take a place among the top 4 in half of the cases.

In order to reduce the effect of survival bias, we increase the weight of sectors that are close to sites of large-scale massacres. In these sites, the probability that entire families were exterminated is likely to be higher. We take the proximity to a large-scale massacre into account by adding the distance to the nearest mass grave to the set of variables subjected to ROBPCA[10]. The massacre-adjusted $GEIs$ are given by the following linear combinations, with MG "log distance to nearest mass grave" (see Table 3)[11]:

$$GEI12mg = 0.42\,(GE1) + 0.49\,(GE2) - 0.76\,(MG)$$

$$GEI13mg = 0.42\,(GE1) + 0.50\,(GE2) + 0.41\,(GE3) - 0.64\,(MG)$$

$$GEI16mg = 0.36\,(GE1) + 0.43\,(GE2) + 0.33\,(GE3) + 0.44\,(GE4) + 0.40\,(GE5) + 0.19\,(GE6) - 0.43\,(MG)$$

$$GEI1379mg = 0.45\,(GE1) + 0.50\,(GE2) + 0.43\,(GE3) + 0.19\,(GE7) + 0.16\,(GE8) + 0.13\,(GE9) - 0.53\,(MG)$$

The resulting province level values and rankings for the massacre-adjusted $GEIs$ are reported in Table 5. The most noteworthy change is that Gitarama now only features once among the top 4, whereas Gikongoro is more predominant with three appearances in the top 4.

[10] The distance to the nearest massgrave is calculated in km at the sector level by overlaying a geo-referenced administrative map with the location of 71 massgraves in Rwanda taken from the Yale Genocide Studies website.

[11] The vectors of loadings stemming from classical PCA are [0.62, 0.64, -0.45], [0.52, 0.58, 0.55,-0.31], [0.39, 0.44, 0.38, 0.44, 0.42, 0.27, -0.26], and [0.48, 0.54, 0.50, 0.26, 0.22, 0.11, -0.32] for gei12mg, gei13mg, gei16mg and gei1379mg respectively. Correlation coefficients between the first PCs of classical PCA and ROBPCA are 0.96, 0.98, 0.99 and 0.99.

This change of order using the massacre-adjusted $GEIs$ is an indication that we succeed in attenuating survival bias.

The province level averages of the genocide indices may hide important differences within provincial borders. Figure 3 illustrates this with a plot of quintiles of the genocide index $GEI1379mg$. We observe a large number of top quintile sectors in Butare, the eastern part of Gikongoro province and Kibuye province, as well as in the northwestern corner of Kibungo. In addition, we find smaller local clusters in and around Kigali City and in the western province Cyangugu.

5 Genocide dummies

5.1 Method

The objective is to assign sectors into categories that are distinct with respect to genocide intensity. A multitude of possibilities exist, including using percentiles (e.g. assigning "1" to the top 10% or top 20% values), or identifying a cut-off value (e.g. assigning "1" to values of the standardized indices that exceed 0.5). Both choices involve a degree of arbitrariness. In addition, they run the risk of wrongly classifying erroneous outlier values. To avoid these caveats, we turn to Local Indicators of Spatial Association (LISA). LISA allow us to identify areas with high values of a variable that are surrounded by high values on the neighboring areas, i.e. high-high clusters. Concomitantly, the low-low clusters are also identified from this analysis (Anselin (1995)).

More formally, LISA provide a measure of the extent to which the arrangement of values around a specific location deviates from spatial randomness. A general expression of a LISA statistic for a variable y_i, observed at location i, is:

$$L_i = f(y_i, y_{J_i}),$$

where f is a function expressing the correlation between y_i and y_{J_i}, and the y_{J_i} are the values observed in the neighborhood J_i of location i. The LISA statistic we look at is the local Moran statistic for an observation i:

$$I_i = \left(y_i - \overline{y}\right) \sum_{j=1}^{n} w_{ij} \left(y_j - \overline{y}\right),$$

with w_{ij} a spatial weighting matrix indicating the relevant neighbors for the LISA analysis.

The weighting matrix w_{ij} can be defined in different ways, although contiguity-based definitions are by far mostly used. We use a first order rook-contiguity based weighting matrix for neighbors, where w_{ij} equals 1 for sectors with a common boundary.

By looking explicitly at areas instead of individual sectors, we can to a large extent avoid wrong classification of erroneous outliers. Arbitrariness in identifying "high" is avoided by assessing the significance of high-high clusters. The procedure employed to assess statistical significance relies on a Monte Carlo simulation of different arrangements of the data and the construction of an empirical distribution of simulated statistics. Afterwards the value obtained originally is compared to the distribution of simulated values and, if the value exceeds the 95^{th} percentile, it is said that the relation found is significant at 5%.

LISA have been used in Anselin (1995) for analyzing spatial patterns of conflict in Africa. In addition, an umber of micro-level studies have used LISA for detecting hot spots in crime (e.g. Murray, McGuffog, Western, and Mullins (2001)). Several other cluster detection methods have been proposed and used for analyzing the location of armed conflict across countries (e.g. Ward and Gleditsch (2002)). A recent micro-level application uses the SaTScan program for detecting space-time clusters in DR Congo (Raleigh, Witmer, and Loughlin (2009)).

5.2 Results

Table 6 gives the share of sectors within a province belonging to high-high and low-low clusters for the massacre-adjusted $GEIs$. Much in line with the results of Table 5, Butare and Kibuye lead the ranks, closely followed by Gikongoro. Figure 4 shows the locations with significant high-high (dark grey) and significant low-low clusters (light grey) for the genocide index $GEI1379mg$. We note a very large low-low cluster in the North, corresponding to low shares of Tutsi in the northern provinces. The significant high-high clusters confirm the pattern detected before: Butare clearly stands out with close to half its territory belonging to high-high clusters, Kibuye comes in second with several high-high clusters on a relatively small area; Gikongoro, Kibungo and Gitarama follow closely. Finally, a few small high-high clusters turn up in Rural Kigali and Cyangugu.

6 Conclusion

This article facilitates the use of the sector level data released by the Rwandan transitional justice system charged with judging genocide suspects. After discussing the general reliability of the data, we detected a number of outlying observations using a recently proposed method for skewed data. Importantly, we found that these outliers are not spatially correlated.

We identified nine genocide intensity proxies from the data, three indicating participation and six indicating excess mortality. Subjecting subsets of these proxies to ROBPCA (Skewness-Adjusted Robust PCA), we proposed four different indices of genocide intensity, We corrected these indices for survival bias. Finally, we used Local Indicators of Spatial Association (LISA) to transform the continuous indices into categorical variables in a non-arbitrary way that is robust to spatial outliers.

The gacaca data can be matched with several nationally representative Rwandan household surveys in which sectors are used as sample units. This means that the scope for using the proposed genocide indices and genocide dummies in empirical applications is large. Enlarging the scope for using fine spatial measures of violent conflict intensity is one way forward in the micro-economic literature on the causes and consequences of armed conflict.

Appendix: legal definition of suspect categories

Originally, four categories of genocide suspects were created in 1996 by the Act on the Organisation and Pursuits of Crimes against Humanity. However, the Organic Law 16/2004 of 19.06.2004 reorganizes the Gacaca process and reduced the categories to three: the former categories 2 and 3 were combined to make category 2 and the 4th category became the 3rd one.

- Category 1:

(a) The person whose criminal acts or criminal participation place him or her among the planners, organizers, incitators, supervisors and ringleaders of the genocide or crimes against humanity, together with his or her accomplices;

(b) The person who, at that time, was in the organs of leadership, at the national level, at the level of Prefecture, Sub-prefecture, Commune, in political parties, army, gendarmerie, communal police, religious denominations or in militia, has committed these offences or encouraged other people to commit them, together with his or her accomplices;

(c) The well known murderer who distinguished himself or herself in the location where he or she lived or wherever he or she passed, because of the zeal which characterized him or her in killings or excessive wickedness with which they were carried out, together with his or her accomplices;

(d) The person who committed acts of torture against others, even though they did not result into death, together with his or her accomplices;

(e) The person who committed acts of rape or acts of torture against sexual organs, together with his or her accomplices;

(f) The person who committed dehumanizing acts on the dead body, together with his or her accomplices.

- Category 2:

(a) The person whose criminal acts or criminal participation place him or her among the killers or who committed acts of serious attacks against others, causing death, together with his or her accomplices;

(b) The person who injured or committed other acts of serious attacks with the intention to kill, but who did not attain his or her objective, together with his or her accomplices;

(c) The person who committed or aided to commit other offences against persons, without the intention to kill, together with his or her accomplices.

- Category 3:

(a) The person who only committed offences against property.

References

Alesina, A., and R. Perotti (1996): "Income distribution, political instability, and investment," European Economic Review, 40(6), 1203–1228.

Anselin, L. (1995): "Local indicators of spatial association," Geographical Analysis, 27, 93–115.

Barnett, V., and T. Lewis (1993): Outliers in Statistical Data (3rd edn). John Wiley & Sons Ltd.: Chichester, U.K.

Barro, R. J. (1991): "Economic growth in a cross-section of countries," Quarterly Journal of Economics, 106, 407–444.

Des Forges, A. (1999): Leave None to Tell the Story: Genocide in Rwanda. New York: Human Rights Watch.

Dunteman, G. H. (2001): Principal Components Analysis, Sage university papers, Quantitative applictions in the social sciences. Sage.

Filmer, D., and L. H. Pritchett (2001): "Estimating wealth effects without expenditure data–or tears: an application to educational enrollments in states of India," Demography, 38(1), 115 – 132.

González, M. A., and R. A. Lopez (2007): "Political Violence and Farm Household Efficiency in Colombia.," Economic Development and Cultural Change, 55(2), 367 – 392.

Government of Rwanda, . (2005): "Report on Data Collection: Annexes," Discussion paper, National service of Gacaca Jurisdiction.

——— (2008): "Recencement des Rescapés du Génocide de 1994: rapport final," Discussion paper, Service National de Recensement.

Hibbs, D. (1973): Mass Political Violence: A Cross-national Causal Analysis. John Wiley & Sons.

Hill, K., and T. Trussell (1977): "Further Developments in Indirect Mortality Estimation," Population Studies, 31, 75–84.

Hubert, M., and S. V. der Veeken (2008): "Outlier detection for skewed data," Journal of Chemometrics, 22, 235–246.

Hubert, M., P. Rousseeuw, and T. Verdonck (2009): "Robust PCA for skewed data and its outlier map," Computational Statistics and Data Analysis, 53, 2264–2274.

Jolliffe, I. (2002): Principal Component Analysis, no. XXIX in Springer Series in Statistics. Springer, NY, 2nd edn.

Longman, T. (2009): "An Assessment of Rwanda's Gacaca Courts," Peace Review, 21(3), 304–312.

Murray, A., I. McGuffog, J. Western, and P. Mullins (2001): "Exploratory Spatial Data Analysis Techniques for Examining Urban Crime," The British Journal of Criminology, 41, 309–329.

Office of the Prosecutor, . (2002): "Abantu Bafungiye mu Magereza Kasho na Burigade," Discussion paper, Rwandan Ministry of Justice.

Pitsch, A. (2002): "Rhe gacaca law of Rwanda: possibilities and problems in adjudicating genocide suspects," Discussion paper, Working Paper NUR-UMD Partnership.

Prunier, G. (1998): The Rwanda Crisis: History of a Genocide. London: Hurst & Company.

Raleigh, C., and H. Hegre (2005): "Introducing ACLED: An Armed Conflict Location and Event Dataset," Discussion paper, 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, March.

Raleigh, C., F. Witmer, and J. O. Loughlin (2009): A Review and Assessment of Spatial Analysis and Conflictchap. in "The Geography of War in Geographic Contributions to International Studies" (ed. C. Flint). Oxford: Basil Blackwell.

Restrepo, J., M. Spagat, and J. Vargas (2006): "The Severity of the Colombian Conflict: Cross-Country Datasets Versus New Micro-Data," Journal of Peace Research, 43(1), 99–115.

Straus, S. (2004): "How many perpetrators were there in the Rwandan genocide? An estimate," Journal of Genocide Research, 6(1), 85–98.

Venieris, Y., and D. Gupta (1986): "Income distribution and socio-political instability as determinants of savings: A cross-sectional model," Journal of Political Economy, 96, 873–883.

Verpoorten, M. (2005): "The death toll of the Rwandan genocide: a detailed analysis for Gikongoro Province," Population, 60(4), 331–368.

Ward, M. D., and K. S. Gleditsch (2002): "Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace," Political Analysis, 10, 244–260.

Wolters, S. (2005): "The Gacaca Process: Eradicating the culture of impunity in Rwanda," Situation report, Institute for Security Studies.

Yanagizawa, D. (mimeo): "Propaganda and Conflict: Theory and Evidence from the Rwandan Genocide," Discussion paper, Mimeo.

Table 1. Genocide proxies

	Nationwide total (1)	% 1994 population Mean (2)	St. Dev. (3)	Skewness-adjusted whisker (4)	Number of outliers (5)	% 1994 population, outliers excluded Mean (6)	St. Dev. (7)
Panel A. Sector level information collected by the gacaca courts (N=1484)							
(G1) Category 1 suspects	76,650	1.1%	1.5%	7.0%	14	1.0%	1.2%
(G2) Category 2 suspects	432,670	6.0%	6.4%	29.4%	12	5.6%	5.1%
(G3) Category 3 suspects	309,500	4.4%	5.7%	25.5%	15	4.0%	4.4%
(G4) Widowed genocide survivors	28,061	0.4%	0.5%	3.5%	5	0.4%	0.5%
(G5) Orphaned genocide survivors	75,078	1.1%	1.5%	7.8%	10	1.0%	1.2%
(G6) Disabled genocide survivors	12,191	0.2%	0.5%	1.4%	21	0.1%	0.2%
Male survivors	103,342	1.4%	2.1%	9.9%	16	1.3%	1.6%
Female survivors	138,207	1.9%	2.6%	13.5%	10	1.8%	2.2%
Panel B: Three additional genocide proxies derived from the sector level gacaca data							
(G7) Widowed / total survivors (%)		11.7%	17.0%	88.8%	8	10.9%	10.1%
(G8) Orphaned / total survivors (%)		36.4%	48.3%	252.7%	7	34.2%	31.3%
(G9) Disabled / total survivors (%)		4.6%	9.0%	51.3%	12	4.0%	5.9%

Notes: 1994 population is projected forward from the 1991 population census using commune level 1978-1991 population growth rates; outliers are identified by means of skewness-adjusted whiskers. Total survivors is defined as the sum of male and female survivors (see last two rows of Table 1)

Table 2: Province level averages of G1 - G9

Panel A	G1	G2	G3	G4	G5	G6	G7	G8	G9
Butare	**0.019**	**0.095**	**0.071**	**0.008**	**0.020**	**0.003**	**0.189**	**0.474**	**0.049**
Byumba	0.002	0.013	0.013	0.001	0.002	0.001	0.057	0.200	0.041
Cyangugu	0.010	0.066	0.028	**0.006**	**0.017**	**0.002**	0.109	0.365	0.034
Gikongoro	0.014	0.076	**0.078**	0.005	**0.013**	0.002	**0.138**	0.402	0.037
Gisenyi	0.005	0.032	0.029	0.001	0.005	0.000	0.100	**0.440**	0.030
Gitarama	**0.015**	**0.080**	**0.066**	**0.006**	**0.014**	**0.002**	0.130	0.308	0.044
Kibungo	**0.015**	**0.083**	0.046	**0.005**	0.012	0.002	**0.132**	0.375	0.049
Kibuye	**0.018**	**0.091**	**0.070**	0.003	0.011	**0.002**	**0.133**	**0.468**	**0.085**
Kigalirural	0.010	0.075	0.043	0.003	0.010	0.002	0.110	0.368	0.045
Kigaliville	0.007	0.043	0.017	0.003	0.011	0.002	0.131	**0.500**	**0.061**
Ruhengeri	0.002	0.012	0.012	0.001	0.002	0.000	0.069	0.309	0.032
Umutara	0.006	0.025	0.019	0.002	0.005	0.001	0.087	0.190	**0.075**
Panel B: excluding outliers									
Butare	**0.016**	**0.088**	**0.059**	**0.007**	**0.017**	**0.002**	**0.161**	**0.391**	0.045
Byumba	0.002	0.013	0.013	0.001	0.002	0.001	0.043	0.143	0.018
Cyangugu	0.010	0.066	0.028	**0.006**	**0.016**	0.002	0.109	0.365	0.034
Gikongoro	0.013	0.072	**0.067**	**0.005**	0.012	0.001	0.118	0.351	0.037
Gisenyi	0.005	0.029	0.027	0.001	0.005	0.000	0.100	**0.440**	0.030
Gitarama	**0.014**	**0.080**	**0.066**	**0.006**	**0.014**	**0.002**	0.130	0.308	0.044
Kibungo	**0.014**	**0.076**	0.046	0.005	**0.012**	**0.002**	**0.132**	0.375	**0.045**
Kibuye	**0.015**	**0.085**	**0.068**	0.003	0.011	0.001	**0.133**	**0.468**	**0.075**
Kigalirural	0.010	0.068	0.037	0.003	0.009	0.001	0.110	0.368	0.041
Kigaliville	0.007	0.043	0.017	0.003	0.011	**0.002**	**0.131**	**0.500**	**0.061**
Ruhengeri	0.002	0.012	0.012	0.001	0.002	0.000	0.063	0.289	0.027
Umutara	0.006	0.025	0.019	0.002	0.005	0.001	0.070	0.190	**0.050**

Notes: For definitions of G1-G9 see Table 1

Table 3: 1991 % Tutsi and distance to Mass graves across provinces

Province	Proportion of Tutsi in a sector	Distance from a sector to the nearest mass grave (km)	log(Distance from a sector to the nearest mass grave)
Butare	17.3	6.1	1.7
Byumba	1.5	25.2	3.2
Cyangugu	10.5	8.3	2.0
Gikongoro	12.8	7.5	1.9
Gisenyi	2.9	9.9	2.2
Gitarama	9.2	10.2	2.2
Kibungo	7.7	9.1	2.0
Kibuye	14.8	7.1	1.8
Rural Kigali	8.8	12.3	2.4
Kigali City	17.9	3.4	1.1
Ruhengeri	0.5	14.2	2.5
Umutara	NA	37.1	3.3
Total	8.9	12.2	2.2

Notes: Proportion of Tutsi taken from the 1991 population census; Distance to mass grave stems from own calculations by overlaying a geo-referenced administrative map with the location of mass graves taken from Yale genocide studies

Table 4: Values and ranking of genocide indices across provinces

Resulting from applying ROBCA to the following sets of variables:	GI12 G1, G2		GI13 G1, G2, G3		GI16 G1, G2, G3, G4, G5, G6		GI1379 G1, G2, G3, G7, G8, G9	
	(1) Values	(2) Ranking	(3) Values	(4) Ranking	(5) Values	(6) Ranking	(7) Values	(8) Ranking
Butare	0.77	**12**	0.91	**12**	1.39	**12**	1.01	**12**
Byumba	-0.94	1	-1.08	2	-1.38	2	-1.14	1
Cyangugu	0.04	6	-0.13	6	0.29	7	-0.13	6
Gikongoro	0.32	8	0.61	**10**	0.61	**10**	0.62	8
Gisenyi	-0.57	4	-0.62	4	-0.91	3	-0.60	4
Gitarama	0.43	**9**	0.58	**9**	0.75	**11**	0.58	**10**
Kibungo	0.48	**10**	0.41	8	0.45	8	0.45	**9**
Kibuye	0.68	**11**	0.83	**11**	0.60	**9**	0.92	**11**
Rural Kigali	0.18	7	0.13	7	0.02	6	0.14	7
Kigali City	-0.33	5	-0.54	5	-0.45	5	-0.43	5
Ruhengeri	-0.94	2	-1.08	1	-1.42	1	-1.11	2
Umutara	-0.59	3	-0.73	3	-0.91	4	-0.72	3

Notes: For definitions of G1-G9 see Table 1

Table 5: Values and ranking of massacre-adjusted genocide indices across provinces

Resulting from applying ROBCA to the following sets of variables:	GI12mg G1, G2, distance to mass grave		GI13mg G1, G2, G3, distance to mass grave		GI16mg G1, G2, G3, G4, G5, G6, distance to mass grave		GI1379mg G1, G2, G3, G7, G8, G9, distance to mass grave	
	(1) Values	(2) Ranking	(3) Values	(4) Ranking	(5) Values	(6) Ranking	(7) Values	(8) Ranking
Butare	1.06	**12**	1.19	**12**	1.60	**12**	1.27	**12**
Byumba	-1.53	1	-1.58	1	-1.75	1	-1.62	1
Cyangugu	0.31	6	0.17	6	0.49	7	0.11	6
Gikongoro	0.57	8	0.78	**10**	0.75	**11**	0.77	**10**
Gisenyi	-0.28	4	-0.39	4	-0.77	4	-0.42	4
Gitarama	0.34	7	0.51	7	0.72	**9**	0.51	8
Kibungo	0.60	**9**	0.59	**9**	0.58	8	0.59	**9**
Kibuye	0.88	**10**	1.02	**11**	0.76	**10**	1.10	**11**
Rural Kigali	0.00	5	0.03	5	-0.04	5	0.04	5
Kigali City	0.92	**11**	0.54	8	0.27	6	0.51	7
Ruhengeri	-0.87	3	-1.05	3	-1.42	2	-1.12	3
Umutara	-1.42	2	-1.41	2	-1.38	3	-1.35	2

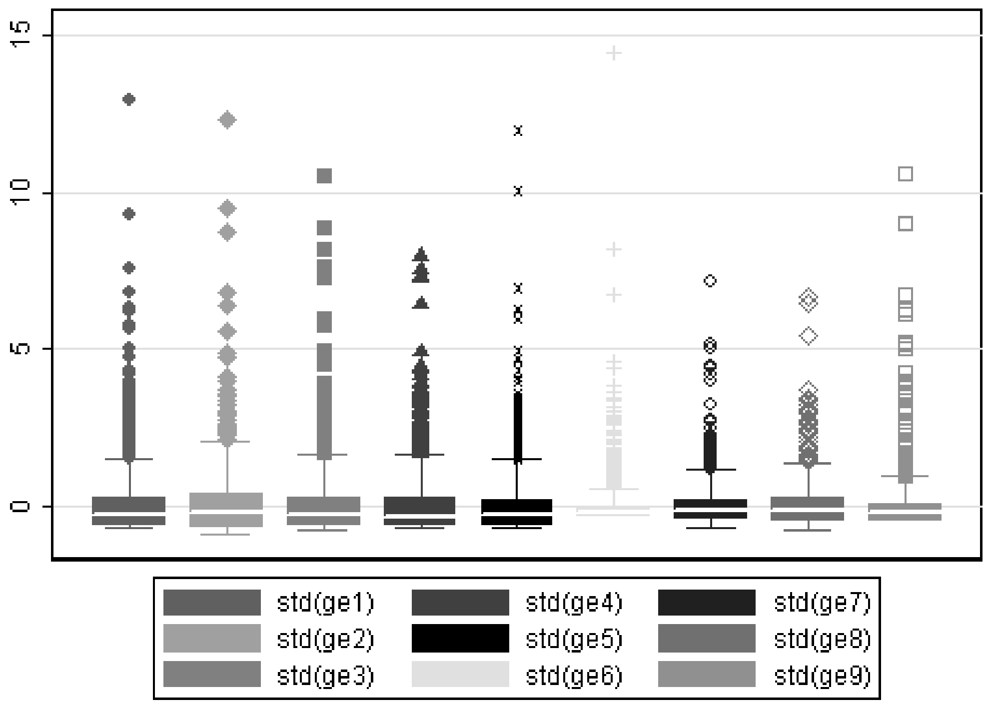
Notes: For definitions of G1-G9 see Table 1

Table 6: Proportion of sectors in a province belonging to high-high and low-low clusters

Resulting from applying ROBCA to the following sets of variables:	GI12mg G1, G2, distance to mass grave		GI13mg G1, G2, G3, distance to mass grave		GI16mg G1, G2, G3, G4, G5, G6, distance to mass grave		GI1379mg G1, G2, G3, G7, G8, G9, distance to mass grave	
	(1) high-high	(2) low-low	(3) high-high	(4) low-low	(5) high-high	(6) low-low	(7) high-high	(8) low-low
Butare	**0.49**	0.00	**0.42**	0.00	**0.48**	0.00	**0.45**	0.00
Byumba	0.00	0.90	0.00	0.88	0.00	0.86	0.00	0.86
Cyangugu	0.20	0.01	0.09	0.01	**0.21**	0.01	0.07	0.02
Gikongoro	0.30	0.02	**0.29**	0.00	**0.24**	0.01	**0.25**	0.01
Gisenyi	0.01	0.12	0.00	0.12	0.00	0.24	0.00	0.15
Gitarama	0.17	0.00	0.16	0.00	0.17	0.00	0.15	0.00
Kibungo	**0.39**	0.08	**0.30**	0.06	**0.22**	0.05	**0.26**	0.06
Kibuye	**0.32**	0.01	**0.33**	0.01	0.11	0.01	**0.37**	0.00
Rural Kigali	0.09	0.14	0.04	0.09	0.04	0.13	0.04	0.09
Kigali City	**0.44**	0.00	0.03	0.00	0.03	0.00	0.03	0.00
Ruhengeri	0.00	0.57	0.00	0.61	0.00	0.73	0.00	0.60
Umutara	0.07	0.62	0.03	0.61	0.03	0.61	0.03	0.59

Notes: For definitions of G1-G9 see Table 1; the clusters are identified in GeoDa using LISA-analysis

Figure 1: standard box plot of standardized (std) genocide proxies GE1-GE9



Notes: Two far outlying observations (std. Values >20) were removed from this figure

Figure 2. Administrative map of Rwanda

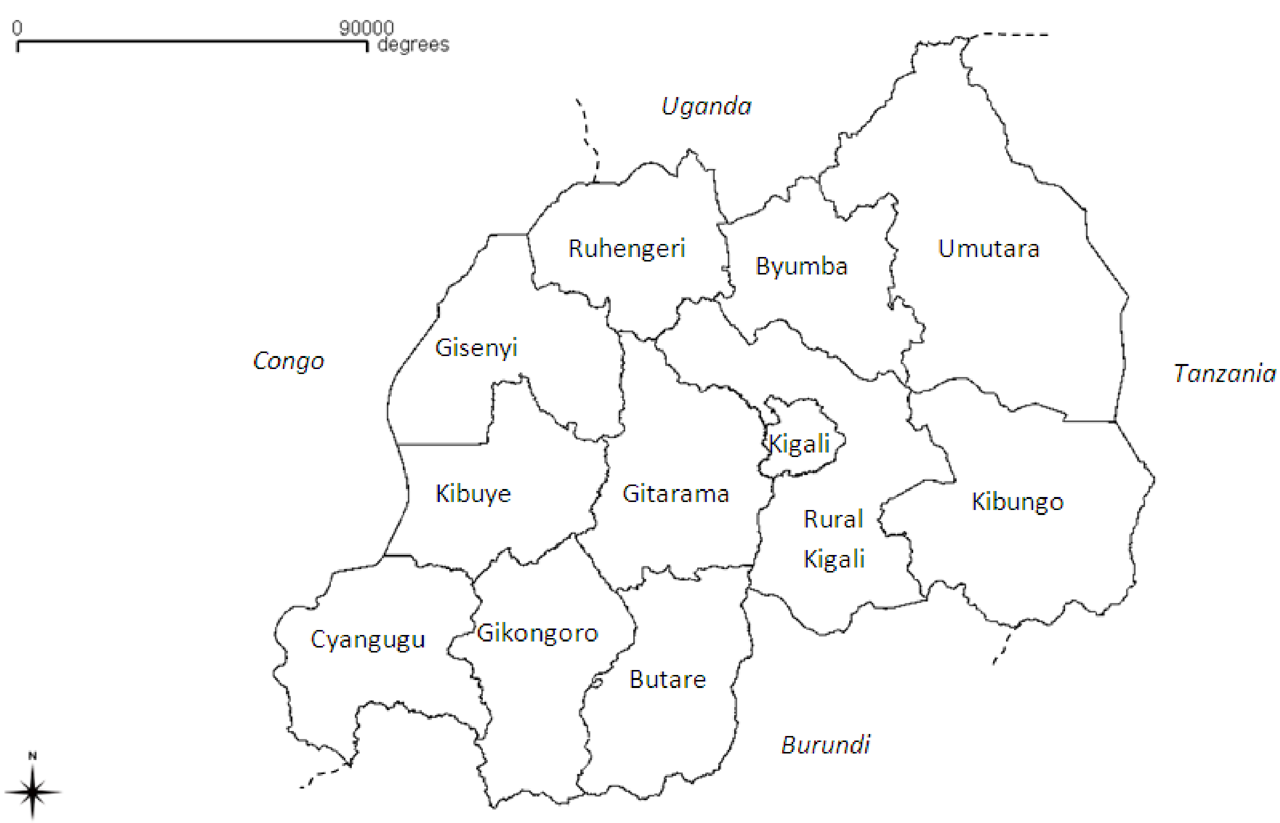


Figure 3. Quintiles of the genocide index GEI1379mg (darkest grey = upper quintile)

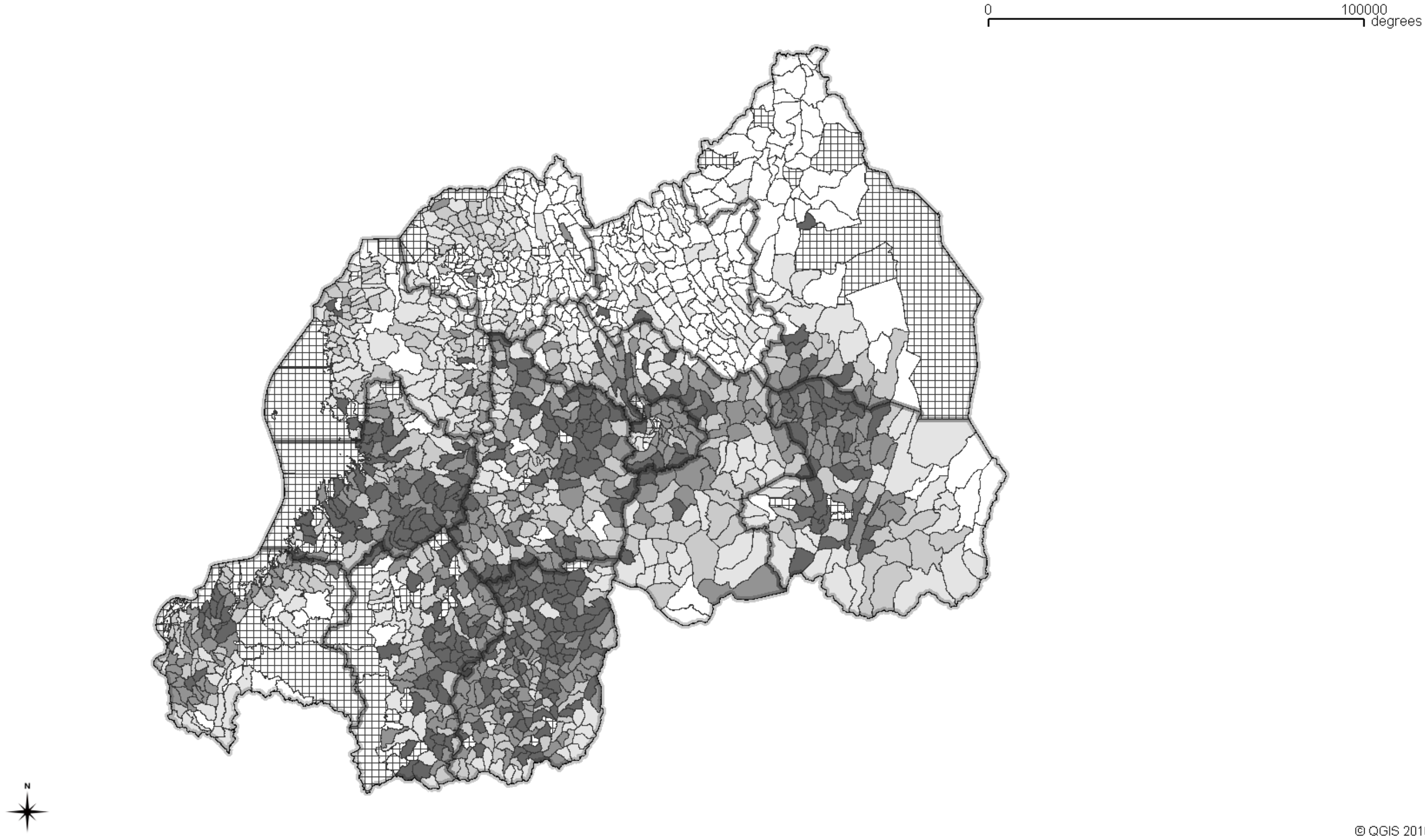


Figure 4. significant high-high (dark grey) and low-low clusters (light grey) for the genocide index GEI1379mg