Physica A 400 (2014) 12–19

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Autocorrelation and cross-correlation in time series of homicide and attempted homicide

A. Machado Filho [a], M.F. da Silva [a], G.F. Zebende [a,b,*]

[a] Computational Modeling Program – SENAI CIMATEC 41650-010 Salvador, Bahia, Brazil

[b] Department of Physics – UEFS 44036-900 Feira de Santana, Bahia, Brazil

HIGHLIGHTS

- We establish the relationship between homicides and attempted homicides by DFA, DCCA, and DCCA cross-correlation coefficient.
- DCCA cross-correlation coefficient identifies a positive cross-correlation.
- The DFA analysis can be more informative depending on time scale (short or long).
- For short scale DFA did not identify auto-correlations, and for long scales DFA presents a persistent behavior.

ARTICLE INFO

Article history:
Received 20 August 2013
Received in revised form 29 November 2013
Available online 13 January 2014

Keywords:
DFA
DCCA
DCCA cross-correlation coefficient
Crime
Homicide

ABSTRACT

We propose in this paper to establish the relationship between homicides and attempted homicides by a non-stationary time-series analysis. This analysis will be carried out by Detrended Fluctuation Analysis (DFA), Detrended Cross-Correlation Analysis (DCCA), and DCCA cross-correlation coefficient, $\rho_{DCCA}(n)$. Through this analysis we can identify a positive cross-correlation between homicides and attempted homicides. At the same time, looked at from the point of view of autocorrelation (DFA), this analysis can be more informative depending on time scale. For short scale (days), we cannot identify auto-correlations, on the scale of weeks DFA presents anti-persistent behavior, and for long time scales ($n > 90$ days) DFA presents a persistent behavior. Finally, the application of this new type of statistical analysis proved to be efficient and, in this sense, this paper can contribute to a more accurate descriptive statistics of crime.

1. Introduction

Due to political, economic and social factors, crime has been studied and statistically modeled by many researchers. For example, it is possible to statistically measure the connection between unemployment and crime [1–3], the correlation between firearms and homicides [4], make a descriptive study of homicides considering author and victim [5], evaluate crime rates through probabilistic models [6], perform a temporal and spatial study of crime [7,8], analyze the flux of tourists and increase in crime [9], simulate computationally criminal activity in an urban environment [10], among others. In this way it is possible to say that crime can be modeled based on the author–victim profile, time, and geographic location, as well as, other variables. This paper aims to detect and measure the auto-correlation and the cross-correlation of homicides and attempted homicides in the city of Salvador, located in the state of Bahia (Brazil). Salvador (12°59′S, 38°29′W) is one of the largest cities in Brazil, with more than 2.7 million people, and with 3787 people per square kilometer [11]. It is worth mentioning that Salvador will host six matches of the 2014 FIFA World Cup Brazil.

* Corresponding author at: Computational Modeling Program – SENAI CIMATEC 41650-010 Salvador, Bahia, Brazil. Tel.: +55 157599922788.
E-mail addresses: gfzebende@hotmail.com, gfzebende@pq.cnpq.br (G.F. Zebende).

0378-4371/$ – see front matter
http://dx.doi.org/10.1016/j.physa.2014.01.015

The crime was studied in terms of homicides and attempted homicides because these are crimes against people and are widely used in empirical studies about the determinants of crime. In this sense Fig. 1 shows the time-series of homicides and attempted homicides per 100,000 citizens. In this figure we can see large irregularities (unpredictable), characteristic of a nonlinear system. Such systems have been studied from the point of view of complex systems. The complex systems are studied in many areas of the natural sciences, mathematics, and the social sciences [12–14]. Complex systems have nonlinear behavior, and can be studied by taking into account the properties of fractals [15], such as self-affinity in time series. If, for example, in a given time-series $\{u(i)\}$ [16] self-affinity appears, then long range power-law correlations are present [17–19]. This makes the study of complex systems very interesting, because it is possible to identify a universality in different kinds of problems [20,21]. It is known that, in the real world, data are highly non-stationary [22], and many conventional methods of analysis are not suited for non-stationary time-series [23].

For non-stationary time-series, we did our analysis in the point of view of Detrended Fluctuation Analysis, DFA [24], Detrended Cross-Correlation Analysis, DCCA [25], and DCCA cross-correlation coefficient, ρ_{DCCA} [26]. Thus, the rest of the paper is laid out as follows: Section 2 provides a brief theoretical review of these methods. Section 3 describes the data used in this paper and presents our results and, finally, Section 4 concludes the paper.

2. Brief review of DFA, DCCA, and ρ_{DCCA}

There are situations where a given observable $u(i)$ is measured at successive time intervals, forming a time-series $\{u(i)\}$ [16]. Some strategies for time-series analysis have been developed [22,23,27–38]. Today, one of the most popular methods for nonstationary time-series analysis is the Detrended Fluctuation Analysis (DFA) and will be briefly presented below.

2.1. The DFA method [24]

The DFA method was developed to analyze long-range power-law correlations in non-stationary systems like in Refs. [24, 29,33,39–47], among others. The DFA method involves the following steps: (see Fig. 2) or Ref. [48].

1. Consider a correlated signal $u(i)$ (daily homicides, attempted homicides), where $i = 1, \ldots, N_{\max}$ (the total number of points in the series). We integrate the signal $u(i)$ and obtain $y(k) = \sum_{i=1}^{k} u(i) - \langle u \rangle$, where $\langle u \rangle$ stands for the average of u;
2. The integrated signal $y(k)$ is divided into boxes of equal length n;
3. For each n-size box, we fit $y(k)$, using a polynomial function of order l, which represents the trend in the box. The y coordinate of the fitting line in each box is denoted by $y_n(k)$, since we use a polynomial fitting of order l, we denote the algorithm by DFA-l;
4. The integrated signal $y(k)$ is detrended by subtracting the local trend $y_n(k)$ in each box (of length n);
5. For a given n-size box, the root-mean-square fluctuation, $F(n)$, for this integrated and detrended signal is given by

$$F_{\text{DFA}}(n) = \sqrt{\frac{1}{N_{\max}} \sum_{k=1}^{N_{\max}} [y(k) - y_n(k)]^2}. \tag{1}$$

6. The above computation is repeated for a broad range of scales (n-sizes box) to provide a relationship between $F(n)$ and the box size n.

In accordance with Refs. [24,48], in this paper we used a polynomial fitting of order 1, with $n = 4$ for the smallest and $n = N_{\max}/4$ for the largest box width. Thus, the DFA method provides a relationship between $F_{\text{DFA}}(n)$ (root mean square fluctuation) and the time scale n, characterized by a power-law:

$$F_{\text{DFA}}(n) \propto n^{\alpha}. \tag{2}$$

In this way, α is the scaling exponent, a self-affinity parameter representing the long-range power-law correlation properties of the signal; such that if $\alpha = 0.5$, then the signal is uncorrelated; if $\alpha < 0.5$, then the correlation in the signal is anti-persistent; and if $\alpha > 0.5$, then the correlation in the signal is persistent.

However, we know that many observables can be measured and recorded simultaneously, at successive time intervals, forming time-series with the same length N [16]. For example, if we have two time-series, then the analysis of the cross-correlation between these time-series can be carried out. Naturally, in the next section, we apply a generalization of the DFA method, called detrended cross-correlation analysis (DCCA), to study the long range cross-correlations in the presence of non-stationarity [49–59].

2.2. The DCCA method [25]

Given two time-series, $\{u_1(i)\}$ and $\{u_2(i)\}$, we compute the integrated signals $R_1(k) \equiv \sum_{i=1}^{k} u_1(i)$ and $R_2(k) \equiv \sum_{i=1}^{k} u_2(i)$, where $k = 1, \ldots, N_{\max}$. Next, we divide the entire time-series into $(N-n)$ overlapping (or not) boxes, each containing $(n+1)$ values. For both time series, in each box that starts at i and ends at $i+n$, we define the local trend, $\widetilde{R}_{1,i}(k)$ and $\widetilde{R}_{2,i}(k)$ $(i \le k \le i+n)$, to be the ordinate of a linear least-squares fit. We define the detrended walk as the difference between the original walk and the local trend. Next, we calculate the covariance of the residuals in each box $f_{\text{DCCA}}^2(n, i) \equiv 1/(n+1) \sum_{k=i}^{i+n} (R_1(k) - \widetilde{R}_{1,i}(k))$