PAPER • OPEN ACCESS

Estimation of models and cycles in time series applying fractal geometry

To cite this article: H J Gallardo Pérez *et al* 2019 *J. Phys.: Conf. Ser.* **1329** 012018

View the article online for updates and enhancements.

Estimation of models and cycles in time series applying fractal geometry

H J Gallardo Pérez[1]**, O A Gallardo Pérez**[1]**, and J P Rojas Suárez**[1]
[1] Universidad Francisco de Paula Santander, San José de Cúcuta, Colombia

E-mail: henrygallardo@ufps.edu.co

Abstract. A time series, also called a time series or chronological series, consists of a set of data, coming from realizations of a random variable that are observed successively in time. Its analysis involves the use of statistical methods to adjust models to explain their behavior and make reliable forecasts. In this article integrated autoregressive models of moving average are adjusted to the studied series, complemented with specific methods of fractal geometry as support for the detection of the existence of random cycles in the series. The present investigation implies the realization of simulations, in a first phase and, later, the analysis of temporal series of economic and social variables of the country and the region.

1. Introduction
In daily activity, there are frequently random variables whose outputs are observed and collected over time, at regular periodic intervals, called time series or time series. Their analysis requires the use of statistical methods that allow information to be extracted from the observations made that do not obey any sampling plan or any experimental design, so that a model can be constructed that allows reliable forecasts to be made.

There are different methods for the analysis of time series, among which we can cite: the classical method, which explains the behavior of the series with a structural model as a function of time, and the Box and Jenkins methodology, which describes the behavior of the series as a function of the values observed in the past and their random variability. This methodology has proven to be a highly efficient technique for making predictions in situations where the pattern inherent in the series is very complex and difficult to unravel [1,2]. Both methods can be complemented with the analysis of the fractal structure of the chronological series that allows, among others, to identify the existence of random cycles in the series and estimate their expected periodicity [3].

The fundamental purpose of this research allows associating specific methods of fractal geometry [4-6] to the analysis of time series and at the same time to present a theoretical and practical reference both for the estimation of the model and in the identification of random cycles of the series. The analysis of the fractal structure of the series is given by the calculation of the Hurst coefficient (H) and the determination of the fractal dimension of the series and its associated probability space [7-9].

2. Method
This research is framed in the quantitative paradigm; part of information obtained from the free databases of different agencies of the Colombian state and based on it adjusts statistical models associated with the time series.

XIV Applied Mathematics Meeting and XI Statistics Meeting

IOP Conf. Series: Journal of Physics: Conf. Series **1329** (2019) 012018

IOP Publishing

doi:10.1088/1742-6596/1329/1/012018

The temporal series is composed of a set of realizations of a random variable Z, each one of them observed in a period of time t, this is a realization of a stochastic process in discrete time: $\{Z_t\}_T = \{Z_1, Z_2, Z_3, \dots, Z_t, \dots\}$. The observed value of the variable in period t is noted Z_t, in each period is observed a single realization of the random variable. It is assumed that there is equal spacing between the observations and that these correspond to discrete points in the time, so that the collected data can be considered as finite successions of accomplishments of stochastic variables.

Time series analysis, which consists of using sample data for inference purposes (estimation, decision-making and prediction), is complex from a functional point of view. However, it can be identified more as an art than as a science, although most of its procedures are based on results from mathematical statistics that have theoretical validity or have been empirically validated [10].

In particular, it should be noted that, even though the variable being observed is the same in each period, it has a different probability distribution in each of those periods and, when observing its realization, what is being observed is the value of a sample of size one in each of the periods. In this way, the construction of a time series model corresponds to an estimate of its parameters based on a sample size one; however, it has been demonstrated that the different time series models are estimated using methods that are highly reliable.

The observations of the phenomenon being studied by means of time series are frequently correlated, with a correlation that increases as the time interval between each pair of observations decreases.

Historically, several methods and models have been developed to estimate intertemporal behavior that describes a time series; this research addresses two of these methods: the classic or structural method and the Box-Jenkins method. Table 1 presents a summary of the most relevant models of these two methods. The structural model assumes that the observed value of the series, Z_t, in time period t is the result of the interaction of four components: trend component (T_t), cyclic component (C_t), seasonal component (S_t) and residual or random component (I_t). The autoregressive integrated moving average model (ARIMA), explains the value of the series as a function of the combination of two polynomials: the autoregressive polynomial (AR), and the moving average polynomial (MA) [2,10].

Adjustment of the ARIMA model is performed based on the exploration of the functions of autocorrelation (ACF) and partial autocorrelation (ACFP), which are obtained once adjusted to a stationary form through differentiations to stabilize the mean and transformations to stabilize variance [2,11-13]. The estimation of missing data is carried out following iterative methods with restrictions [14,15].

On the other hand, a stochastic process is said to be seasonal if its averaging function presents the behavior of a wave; in these processes, the ACF and ACFP functions reflect the correlation between consecutive periods and the correlation between seasonal periods. The proposed model is constructed from two points of view: intra- and inter-stations, obtaining a composite model ARIMA (p,d,q)x(P,D,Q)s in which p and q represent the orders of the autoregressive and moving average polynomials of the non-seasonal component and P and Q the corresponding orders of the autoregressive and moving average seasonal polynomials, respectively. The level is stabilized by means of non-seasonal differentiations and seasonal differentiations. The length of the seasonality is represented by s.

The study of the fractal series was initiated by Hurst based on Einstein's work on the Brownian movement, his work was later complemented by Mandelbrot [16]. Hurst, raise the equation $R/S = k * n^H$ to estimate the flow cycles of the Nile River and solve the problem of the construction and commissioning of a dam that would guarantee a uniform flow, since the river had periodic non-random cycles. In this equation, R/S is the standardized range (Range/standard deviation), n is the index used to denote the number of observations per time interval, k is a constant for the time series and H is the Hurst exponent.

The general Hurst equation has a characteristic of fractal geometry: its scale is in accordance with a power law. The value of H is in the closed interval $\{0,1\}$. Now, if H=0.5 the system follows a random trajectory, recovering the original scenario of a Brownian movement. Otherwise, the observations are not independent and each one carries with it a "memory" of events that have preceded it.

Three cases can be distinguished then: (i) H=0.5, independent series (Brownian noise or Brownian movement). The values observed in the series are independent of the values of the past. The series is a random walk. (ii) 0.0≤H<0.5, anti-persistent series (pink noise). The system is covering less distance than a random walk. Thus, he has the tendency to return frequently to himself. If it increases, it is very likely to decrease in the next period. If it decreases, it is very likely to be increased. (iii) 0.5<H≤1.0, persistent series (black noise). This series covers more distance than a random walk. Thus, if a series increases in a period, it is very likely that it will suffer an increase in the following period [16].

3. Results

3.1. Simulation of a white noise
The white noise process is the simplest model of time series, it consists only of a completely random series, it is usually represented $\{a_t\}_T$. One hundred and twenty simulations of white noise process are performed, with zero arithmetic mean parameters and constant variation, each composed of 300 points recorded at regular time intervals. In all cases Hurts exponent is found between 0.124 and 0.146, then its fractal dimension is in the range 2-0.146=1.854 to 2-0.124=1.876 and the fractal dimension of the probability space associated to each process is between 1/0.146=6.849 and 1/0.124=8.065.

3.2. Simulation of a chaotic process
The chaotic process is characterized by non-linearity, is highly sensitive to small changes in initial conditions and its behavior is unpredictable, even when it is deterministic. Also in this case, 120 simulations of chaotic processes were carried out, each consisting of 300 points recorded at regular time intervals. For each simulation of a chaotic process we start from the random determination of a X_o point of seed between zero and one, then we calculate the series of iterations $X_{n+1}=4*X_n*(1-X_n)$, finally we obtain the chaotic series by transformation $ln[X_n/(1 - X_n)]$. These series are also anti-persistent since the Hurts exponent is between 0.218 and 0.304, its fractal dimension is between 2-0.304=1.696 and 2-0.218=1.782 and the fractal dimension of the associated probability space is in the range 1/0.218=4.587 to 1/0.304=3.289; therefore, the series does not present random cycles.

3.3. Monthly variation of the consumer price index
The "índice del precio del consumidor (IPC)" is a statistical research conducted by the "Departamento Administrativo Nacional de Estadistica (DANE)" that allows measuring the average percentage variation of retail prices of a set of goods and services of final consumption that consumers demand. The latest methodology for calculating the IPC was developed in 2008 with updating of the assets that make up the family basket and expansion of the coverage of cities for calculation; the month of December of that same year was taken as a base, but adjustments and splices of the previous series and databases are made to have equivalent and comparable values. To estimate the model, the information corresponding to the monthly variation of the IPC for the city of San José de Cúcuta, Colombia, from January 1998 to December 2018, is taken.

The adjustment of an ARIMA model to the Z_t series implies a differentiation of one delay along with one of a twelve-month seasonal delay, in accordance with the analysis of the FAC and FACP functions obtained for the differentiated series. The random variable Wt is then defined, depending on the time delay operator B [$B^k Z_t = Z_{t-k}$], Equation (1), in such a way that it involves the delay differentiation, one and another seasonal time delay 12:

$$W_t = (1 - B)(1 - B^{12})Z_t = (1 - B - B^{12} + B^{13})Z_t = Z_t - Z_{t-1} - Z_{t-12} + Z_{t-13} \qquad (1)$$

The model proposed for the variable W_t is of type ARIMA(1,1,1)x(1,1,0)$_{12}$, which is autoregressive of moving average in time delay 1, and autoregressive in time delay 12, Equation (2) shows the model for the IPC, in the city of San José de Cúcuta, Colombia, corresponding to the differentiated series; the

series does not present atypical data; the model does not include constant, because this is significantly equal to zero:

$$\widehat{W}_t = 0.389\widehat{W}_{t-1} - 0.509\widehat{W}_{t-12} + 0{,}931\hat{a}_{t-1} + \hat{a}_t \tag{2}$$

Consequently, after replacing Equation (1) in Equation (2) and carrying out the corresponding simplifications, Equation (3) is obtained, which expresses the model associated with the GDP for the city of San José de Cúcuta, Colombia:

$$\begin{aligned}\widehat{Z}_t = {}& 1{,}389\widehat{Z}_{t-1} - 0{,}389\widehat{Z}_{t-2} + 0{,}491\widehat{Z}_{t-12} - 0{,}102\widehat{Z}_{t-13} + 0{,}389\widehat{Z}_{t-14} + 0{,}509\widehat{Z}_{t-24} \\ & - 0{,}509\widehat{Z}_{t-25} + 0{,}931\hat{a}_{t-1} + \hat{a}_t\end{aligned} \tag{3}$$

For the analysis of the fractal structure of the series, the Hurst coefficient is calculated. The procedure implies that for each value of t ($1 \le t \le n$), the range of values observed in the series from Z_1 to Z_t, respectively, is calculated. Then the regression line is estimated: $LN\left(R/S\right) = k + H * LN(n) = 0{,}6236 + 0{,}1842 * LN(n)$ of the variable $LN(R/S)$ on the variable $LN(n)$, whose slope is an estimator of the Hurst coefficient.

The following information about the fractal structure of the IPC, San José de Cúcuta, Colombia, series is estimated: this series is anti-persistent since the Hurts exponent is 0.1842, its fractal dimension is 2-0.1842=1.8158 and the fractal dimension of the associated probability space is 1/0.1842=7.245; therefore, the series does not present random cycles.

3.4. Residential consumption of electric energy
The government of Colombia, through the mining energy planning unit of the Colombian mining energy information system, presents the information corresponding to the single information system of domiciliary public services, through which it is possible to access the value of the total and invoiced residential and non-residential consumption of electrical energy, by marketing company in each city and department of the country.

$$\begin{aligned}\widehat{Z}_t = {}& 0.648\widehat{Z}_{t-1} + 0{,}392\widehat{Z}_{t-2} + 0.541\widehat{Z}_{t-12} - 0.149\widehat{Z}_{t-13} - 0{,}392\widehat{Z}_{t-14} + 0{,}459\widehat{Z}_{t-24} \\ & - 0{,}459\widehat{Z}_{t-25} + \hat{a}_t\end{aligned} \tag{4}$$

To estimate the model, the value-added series of monthly residential consumption of electric energy in the city of San José de Cúcuta, Colombia, is constructed from January 2006 to December 2018. The model adjusted in this section, Equation (4), corresponds to total consumption, but a model can also be estimated for consumption in each socioeconomic stratum. Fractal analysis of this series establishes that it is anti-persistent (H=0.1291) and does not present random cycles.

3.5. Cattle slaughter
Slaughter is the process by which an animal is killed in an appropriate manner to avoid suffering, with the aim of using its meat and parts for human consumption and use. During the first quarter of 2019, Norte de Santander, Colombia, slaughtered 16809 heads (5701 in January, 5317 in February and 5791 in March), 9777 males and 7032 females, which constitutes 2% of national production. All production was destined for domestic consumption.

The model presented in Equation (5) is estimated from the monthly production of cattle in the department of the Norte de Santander, Colombia, from January 2013 to December 2018. This series is anti-persistent and does not present random cycles.

$$\hat{Z}_t = 0.715\hat{Z}_{t-1} + 0{,}285\hat{Z}_{t-2} + 1.119\hat{Z}_{t-12} - 0.834\hat{Z}_{t-13} - 0{,}285\hat{Z}_{t-14} - 0{,}119\hat{Z}_{t-24}$$
$$+ 0{,}119\hat{Z}_{t-25} + \hat{a}_t \tag{5}$$

3.6. Movement of the urban automobile fleet

In this document reference is made to the series of the number of passengers moved quarterly by public transport vehicles in San José de Cúcuta, Colombia, from January 2005 to December 2018. The estimated model is presented in Equation (6). This series is anti-persistent (H=0.013), does not present random cycles and experiences a decreasing trend in time.

$$\hat{Z}_t = 0.824\hat{Z}_{t-1} + 0{,}176\hat{Z}_{t-2} + 0.709\hat{Z}_{t-4} - 0.533\hat{Z}_{t-5} - 0{,}176\hat{Z}_{t-6} + 0{,}291\hat{Z}_{t-8}$$
$$- 0{,}291\hat{Z}_{t-9} + \hat{a}_t \tag{6}$$

4. Discussion

Fractal geometry has allowed great advances to be made in time series analysis. Thus, from the description of the fractal structure of the series can be detected the existence of random cycles in the series, a situation that cannot be obtained by other methods. It is also possible to determine the size of the associated probability space and the "memory" of the series.

In the analyzed series, there is no evidence of long term random cycles, this is important from the point of view of adjustment of structural type models because it not only facilitates the estimation of the other components of the model, but also allows making forecasts with a greater degree of reliability.

The fundamental purpose of this research phase is the incorporation of fractal geometry to the development of time series analysis. The variables analyzed here constitute a contribution not only to the economy, but also to the line of research in this field.

The fractal geometry has allowed to make great advances in the analysis of time series. This is how, from the description of the fractal structure of the series, the existence of random cycles in the series can be detected, a situation that cannot be obtained by other methods. It is also possible to determine the dimension of the associated probability space and the "memory" of the series.

In the series analyzed, there is no evidence of long-term random cycles, this is important from the point of view of adjusting structural models, as it not only facilitates the estimation of the other components of the model, but also allows forecasts with a higher degree of reliability.

The fundamental purpose of this phase of the research is the incorporation of fractal geometry to the development of the analysis of time series, the variables analyzed here, constitute a contribution not only to the economy, but to the line of research in this field.

5. Conclusions

Fractal geometry has allowed great advances to be made in time series analysis. Thus, from the description of the fractal structure of the series can be detected the existence of random cycles in the series, a situation that cannot be obtained by other methods. It is also possible to determine the size of the associated probability space and the "memory" of the series.

In the analyzed series, there is no evidence of long term random cycles, this is important from the point of view of adjustment of structural type models because it not only facilitates the estimation of the other components of the model, but also allows making forecasts with a greater degree of reliability.

The fundamental purpose of this research phase is the incorporation of fractal geometry to the development of time series analysis. The variables analyzed here constitute a contribution not only to the economy, but also to the line of research in this field.

References

[1]   Peña D 1989 Sobre la interpretación de modelos ARIMA univariantes *Trabajos de estadística* **4(2)** 19
[2]   Box G and Jenkins G 1969 *Time series analysis, forecasting and control* (San Francisco: Holden–Day)

[3]   Plazas L, Avila M and Moncada G 2014 Estimación del exponente de Hurst y dimensión fractal para el análisis de series de tiempo de absorbancia UV-VIS *Ciencia e Ingeniería Neogranadina* **24(2)** 133

[4]   Barnsley M 1988 *Fractals everwhere* (New York: Academic Press)

[5]   Mandelbrot B 1993 *Los objetos fractales: Forma, azar, dimensión* 3ª edición (Barcelona: Tusques)

[6]   Mandelbrot B 1985 *The fractal geometry of nature* (New York: W. H. Freeman)

[7]   Villaseñor L G and Alcaraz J V 2010 Antecedentes fractales para mercados financieros *Inceptum Revista de Investigación en Ciencia de la Administración* **5(9)** 263

[8]   Gao J, Cao Y, Tung W and Hu J 2007 *Multiscale analysis of complex time series: Integration of chaos and random fractal theory, and beyond.* (New Jersey: John Wiley & Sons)

[9]   González V and Guerrero C 2001 Fractales, fundamentos y aplicaciones, parte I: Concepción geométrica en la ciencia e ingeniería *Ingenierías* **4(10)** 53

[10] Brockwell P and Davis R 2002 *Introduction to time series and forecasting* (New York: Springer) p 57

[11] Guerrero V 1991 *Análisis estadístico de series de tiempo económicas* (México: Universidad Autónoma Metropolitana) p 102

[12] Botero S and cano J 2008 Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia *Cuadernos de Economía* **27(48)** 173

[13] Moreno E and Nieto F 2014 Modelos TAR en series de tiempo financieras *Comunicaciones en estadística* **7(2)** 223

[14] Velásquez M and Martínez J 2009 Estimación de observaciones faltantes en series de tiempo usando métodos multivariados con restricciones *Comunicaciones en Estadística* **2(1)** 1

[15] Gallardo H and Nieto F 1996 Cálculo del número mínimo de observaciones para estimar el vector de datos faltantes en una serie temporal generada por un modelo AR(p) *Revista Colombiana de Estadística* **17(33-34)** 57

[16] Quintero O and Ruiz J 2011 Estimación del Exponente de Hurst y la dimensión fractal de una superficie topográfica a través de la extracción de perfiles *Revista Geomática UD.GEO* **5** 84