Physical applications: analysis of Colombian coffee prices using fractional Brownian motion

D A Prada\textsuperscript{1}, A Acevedo\textsuperscript{1}, H Fernandez\textsuperscript{1}, S C Prada\textsuperscript{2}, and J M Gómez\textsuperscript{3}

\textsuperscript{1} Universidad Pontificia Bolivariana, Bucaramanga, Colombia
\textsuperscript{2} Universidad Industrial de Santander, Bucaramanga, Colombia
\textsuperscript{3} Servicio Nacional de Aprendizaje, Bogotá, Colombia

E-mail: duwamg.prada@upb.edu.co

Abstract. Colombia is an exporting country of quality coffee in the world; divergent factors in the short and long term, such as inclement weather, geographic changes, and socio-political development, are some of the factors that influence the price of this product. Knowing the future behavior of this phenomenon is one of the most important studies for economists, academics, coffee growers, entrepreneurs, and exporters. Brownian motion, a physical phenomenon that describes the irregular movement of some particles suspended in a fluid, was described by the probability of finding a particle in a position at a specific time. Fractional Brownian motion describes the random fluctuation of a stochastic process continuous in time and is characterized by the Hurst coefficient to observe persistence and volatility in a time series. The percentage of volatility that changes in the price of coffee present allows generating strategies to maintain the quality of the product and, therefore, its positioning in the market. In this work it was found that the series of data on coffee prices is persistent and that its volatility is 43.77%.

1. Introduction
The continuous fluctuation suffered by financial markets within the activity of import and export of products of interest, involves the different economic sectors of a country and requires them to carry out continuous studies that allow adequate preparation regarding the changes that may affect its economy. Economists have assumed, in general, that the variation of the prices of some product has two components: a long-range one, in which prices would be governed by deep economic forces such as the opening of commercial routes, inventions that use the product, a war, some technological innovation that will modify the use of the product, a revolution, etc. The other component of the price would be to reach, the prices would vary randomly, due to a very large number of causes, many of which could not be determined with precision. These swings, called fluctuations, are transient [1]. It has been thought that there is almost no relationship between the two long and shortrange rhythms [2].

Colombia is a country recognized for cultivating and exporting high quality coffee. Factors such as climate, terrain properties, topography, culture, among others, directly affect both morphological and flavor characteristics, acidity, aftertaste, body, aroma and bitterness [1]. Particularly in some regions of Colombia there may be harvests throughout the year, which implies that Colombian coffee is permanently listed on the market due to its supply. Coffee is one of the most important export products in the country, therefore, studying the variation in its sale price internationally, is a benchmark for the economy, mainly in the coffee sector; coffee production systems in Colombia have been developed only with varieties of the Coffea arabica species [1]. The website Investing.com, is a financial information
site of worldwide reference, which evidences the variations of the prices of different products worldwide. With the coffee price data, provided by Investing.com, from 2003 to 2018, the behavior of this daily time series was analyzed, in order to observe the persistence of these changes in the future, using the rescaled rank method to find the Hurst coefficient and calculate the volatility of the coffee price.

2. Mathematical method

In the study on the processes of fertilization of flowers, Robert Brown in 1821, observed that the pollen particles presented a rapid irregular oscillatory movement suspended in a fluid. This phenomenon was defined as Brownian movement. For the year 1905, the kinetic theory is positioned which maintains that the Brownian motion is caused by a bombardment of the fluid molecules on the particle. According to Einstein, Brownian motion is described through the probability \( p(r, t) \) of finding a particle at position \( r \) at time \( t \), which satisfies the macroscopic diffusion equation \[3\].

Regarding the application of the Brownian movement in the context of fluctuations in market values, Louis Bachelier developed this theory by introducing the Chapman-Kolmogorov equation. Bachelier's work did not give direct advances in the context of Brownian motion physics, since in economics there is no place for a coefficient of friction, Stokes' law, much less for Avogadro's number; but it allows us to visualize a first application of the Brownian movement in other areas \[4\]. Today, beyond the study of the dynamics of a particle in a fluid, Brownian motion has a very wide range of different applications in areas such as hydrodynamics, in polymer dynamics, seismology for the analysis of vibrations, generation of pseudo-random sequences, among others \[3\].

In this approach for ordinary and fractional Brownian motion, the recorded time series corresponding to the long axes of the diffraction pattern of the cells when undergo shear stress were studied from nonlinear correlation to Hurst exponents. In a classical non differentiable trajectory or, more generally, ordinary Brownian motion (OBM), past increments in displacement are uncorrelated with future increments, that is, the system has no memory. In a correlated random walk, or more generally, fractional Brownian motion (FBM), past increments in displacement are correlated with future increments, at least for the first steps of the process, hence the system has memory. Hurst exponent for a time series provides a measure of whether it is a pure white noise random process or has underlying trends. Dynamic process, that might naively characterized with purely white noise, sometimes turn out to exhibit Hurst exponent statistics for long memory process, \( i.e. \), colored noise. A long memory process is a process where past events have a decaying effect on future ones. But those are forgotten as time moves forward \[5\].

The FBM is characterized by a parameter, the so called Hurst parameter \( H \). An FBM with Hurst parameter \( H > 1/2 \) is called a persistent process, \( i.e. \), the increments of this process are positively correlated. On the other hand, the increments of an FBM with \( H < 1/2 \) constitute what is called an anti-persistent process, with increments being negatively correlated. For \( H = 1/2 \), an FBM corresponds to Brownian motion which has independent increments \[6\].

The Hurst coefficient also allows to measure volatility, understood as the maximum and minimum peaks values at which data are found in a time series with respect to their average value, for the risk analysis of a time series \[7\]; the persistence of a series of time then depends on the value of the Hurst coefficient \( H \), for white noise \( H = 0 \), for Brownian noise \( H \approx 0.5 \); \( H = 1 \) indicates directed motion. Applied to financial data such as stock price, the Hurts Exponent can be interpreted as a measure for the trendingness: \( H < 0.5 \) high volatility, stock price is anti trended, \( H = 0.5 \), stock prices behaves like a Brownian process, no trend, \( H > 0.5 \) stock prices has a trend \[7-10\]. According to statistical mechanics, if \( H \) equals 0.5, the series presents a random path \[11\].

The Hurst coefficient of a time series is determined by partitioning the entire data, \( i.e. \) subgroups are processed by data aggregation. For each of these sub-series, calculate the average and the deviations of the data from the average. Using cumulative sums of the deviations, you calculate the maximum and minimum of this process to determine the rescaled range associated with the data. With this value and the standard deviation, calculated in each subgroup, a linear logarithmic regression is performed where the slope is the Hurst coefficient. Hurst found that the a-dimensional ratio R/S allows comparing the re-
sizing of several temporal series and that such a resizing can be very well described by a law of power as follows in the Equation (1).

\[
\frac{R}{S} = cN^H,
\]

where \( N \) is the time interval for the observations, \( H \) is the estimate for Hurst's Exponent as calculated from the R/S method and \( c \) is a constant \([12,13]\); using Equation (1), it is possible to perform a logarithmic linear regression in which the slope represents the Hurst coefficient.

3. Results

In order to know the behavior of the time series associated with the phenomenon of the price of Colombian coffee, the study of the fractional Brownian movement of the series will be carried out by calculating the Hurst coefficient and the logarithmic linear regression of Equation (1) , in order to determine if the time series of coffee prices present persistence or memory in the future. Investing.com recorded 5622 data on prices of Colombian export coffee between January 2, 2003 and May 27, 2018. The series of 5622 data is partitioned into 16 subgroups \([14]\).

Table 1 shows the 16 subgroups, relating each to their number of data, the rescaled range, the maximum and minimum value of the accumulated sum, and the average of said subgroup.

Table 1 shows the first 364 data included between the 2nd of January 2003 and the 31th of December 2003. With these data the standard deviation was calculated, and the rescaled range. The group 2 consists of the first 460 data recorded, whose dates vary between 2nd of January 2003 and 31th of December 2004. For the groups 3 until the 16th the same process was carried out.

Table 1. Data of the standard deviation, released number and rescaled range.

| Subgroup | Number data | Rescaled range | Max           | Min           | Average      |
|----------|-------------|----------------|---------------|---------------|--------------|
| 1        | 364         | 1228234.20     | 1080021.29    | -148212.91    | 307368.82    |
| 2        | 460         | 8185857.53     | 0             | -8185857.00   | 329420.01    |
| 3        | 1095        | 33271720.29    | 0             | -33271720.00  | 371905.35    |
| 4        | 1460        | 49290459.53    | 0             | -49290459.00  | 395409.32    |
| 5        | 1825        | 57765028.59    | 0             | -57765028.00  | 307368.82    |
| 6        | 2191        | 67989658.41    | 0             | -67989658.00  | 422120.00    |
| 7        | 2556        | 93490109.60    | 0             | -93490109.00  | 455822.11    |
| 8        | 2921        | 159834110.82   | 0             | -159834110.00 | 493249.50    |
| 9        | 3286        | 273288951.22   | 0             | -273288951.00 | 546820.55    |
| 10       | 3652        | 300020327.92   | 1729547.17    | -298290780.00 | 558182.21    |
| 11       | 4017        | 311657257.98   | 31761627.94   | -279895630.00 | 549824.95    |
| 12       | 4382        | 308015019.48   | 0             | -308015019.00 | 562590.30    |
| 13       | 4747        | 334188014.26   | 0             | -334188014.00 | 574424.31    |
| 14       | 5113        | 374641588.88   | 0             | -374641588.00 | 592701.95    |
| 15       | 5478        | 407922784.42   | 0             | -407922784.00 | 607716.61    |
| 16       | 5622        | 415444074.75   | 0             | -415444074.00 | 611109.16    |

In the Table 2 shows the data needed for the calculation of the coefficient of hurts for this time-series, in addition, the natural logarithm of the data number and the natural logarithm of the ratio between the rescaled range and the standard deviation are shown by subgroup, which are useful for logarithmic linear regression.

The logarithmic linear regression is performed, taking as data in the coordinate the natural logarithm of the number of data for each subseries, and in the ordinate, we take the natural logarithm of the quotient between the rescaled range and the standard deviation corresponding to each subseries created, these data are related in Figure 1.
Table 2. Related data from the sixteen groups depending on the range and standard deviation. Calculation of the natural logarithm of the data number and the natural logarithm of the quotient between the rescaled range and the standard deviation.

| Subgroup | Number data | Rescaled range | Standard deviation | Ln (Num) | Ln(R/S) |
|----------|-------------|----------------|--------------------|----------|---------|
| 1        | 364         | 1228234.20     | 15846.03           | 5.90     | 4.35    |
| 2        | 760         | 8185857.53     | 35111.50           | 6.63     | 5.45    |
| 3        | 1095        | 33271720.29    | 73318.00           | 7.00     | 6.12    |
| 4        | 1460        | 49290459.53    | 76967.00           | 7.29     | 6.46    |
| 5        | 1825        | 57765028.59    | 74755.54           | 7.51     | 6.56    |
| 6        | 2191        | 67989658.41    | 77412.00           | 7.69     | 6.78    |
| 7        | 2556        | 93490109.60    | 113483.02          | 7.85     | 6.71    |
| 8        | 2921        | 159834110.82   | 147700.71          | 7.98     | 6.99    |
| 9        | 3286        | 273288951.22   | 207069.75          | 8.10     | 7.19    |
| 10       | 3652        | 30002027.92    | 202081.01          | 8.20     | 7.30    |
| 11       | 4017        | 311657257.98   | 195050.26          | 8.30     | 7.38    |
| 12       | 4382        | 308015019.48   | 194113.24          | 8.39     | 7.37    |
| 13       | 4747        | 334188014.26   | 191512.65          | 8.47     | 7.46    |
| 14       | 5113        | 374641588.88   | 196937.39          | 8.54     | 7.55    |
| 15       | 5478        | 407922784.42   | 198707.70          | 8.61     | 7.63    |
| 16       | 5622        | 415444074.75   | 197292.17          | 8.63     | 7.65    |

Figure 1. Logarithmic linear regression of the data supplied in Table 2.

4. Conclusions
Ordinary and fractional Brownian motion have been helpful in explaining the behavior of random irregular motions suspended in a fluid. Particularly, fractional Brownian motion has also been studied to support hydrodynamic, seismic processes of generating random pseudo sequences, applications in stock market fluctuations. This important physical theory, in relation to the analysis in time series, has been characterized through the calculation of the Hurst coefficient, which determines correlated increment processes, under the idea of systems with persistence or memory. It also allows generating a volatility indicator associated with the time series. For the series of Colombian coffee prices provided by the Investing.com site with 5622 data between January 22, 2003 and May 27, the value of the Hurst coefficient, which was calculated using the accumulated sums, rescaling range and log arrhythmic linear regression, is equal to 1.1246, \( H > 0.5 \), which means, the time series is persistent or has long-term memory. From the data above presented, it can be concluded that the series is persistent under a volatility of 43.77%, volatility refers to the perception of the behavior of a financial asset in the timeline, with respect to coffee the risk is medium because it is directly related to the price that it presents in the Chicago mercantile exchange where Colombian coffee is the price maker world. It is important that for future studies the study of other variables such as temperature, humidity and the incidence of precipitation conditions in the place of planting and cultivation of coffee be considered.
Acknowledgments
The authors of this work thank the Universidad Pontificia Bolivariana, Bucaramanga, Colombia, for the support received during the research project entitled Financial impact of coffee crops according to rainfall in the region of Mesa de los Santos, Santander, Colombia.

References
[1] Acevedo A, Prada D, Fernández H, Chía M 2020 Gestión del Conocimiento Perspectiva Multidisciplinaria (Estado de Zulia: Fondo Editorial Universitario de la Universidad Nacional Experimental Sur del Lago Jesús María Semprúm)
[2] Braun E 2011 Caos, Fractales y Cosas Raras (Mexico: Fondo de Cultura Económica) p 92
[3] Giraldi H, Campos E 2015 ¿Es el movimiento Browniano un proceso estocástico o determinista? Boletín de la Sociedad Mexicana de Física 29(2) 103
[4] Schlesinger M, Klafter J, Zumofen G 1999 Above, below and beyond Brownian motion Amer. J. Phys. 67(12) 1253
[5] Mancilla M, Leguto A, Riquelme B, Ponce P, Bortolato S, Korol A 2017 Hurst exponent: A Brownian approach to characterize the nonlinear behavior of red blood cells deformability Physica A: Statistical Mechanics and its Applications 488 1
[6] Yerlikaya F, Vardar C, Yolcu Y, Weber G 2013 Estimation of the Hurst parameter for fractional Brownian motion using the CMARS method Journal of Computational and Applied Mathematics 259(2) 843
[7] Mandelbrot B 2006 Los Objetos Fractales. Forma, Azar y Dimensión (España: Metatemas)
[8] Racine R 2011 Estimating the Hurst Exponent (Germany: Mosaic Group)
[9] Peters E 1991 Chaos and Order in the Capital Markets (United States John Wiley & Sons, Inc.)
[10] Mandelbrot B 1997 Fractals and Scaling in Finance (United States of America: Springer)
[11] Peters E 1994 Fractal Market Analysis: Applying Chaos Theory to Investment and Economics (New York: Willey)
[12] Martin L, Aranda N, Quimbaya C 2015 Long-range correlations and trends in Colombian seismic time series Physica A: Statistical Mechanics and its Applications 421 124
[13] Henrice E, Palma D, Conçalves A, Mesquita A 2017 Support to the identification of anomalies in an external neutron source using Hurst Exponents Progress in Nuclear Energy 99 119
[14] Rodríguez S, Edgar L 2012 Hidrología de Hurst y Box Counting para el análisis de persistencia, volatilidad y riesgo en dos series de tiempo colombianas Cuadernos Latinoamericanos de Administración 8(14) 41