Self-Similar Models: Relationship between the Diffusion Entropy Analysis, Detrended Fluctuation Analysis and Lévy Models

Maria C. Mariani 1,2, William Kubin 2, Peter K. Asante 2, Osei K. Tweneboah 2,*, Maria P. Beccar-Varela 1, Sebastian Jaroszewicz 3 and Hector Gonzalez-Huizar 4,5

1 Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX 79968, USA; mcmariani@utep.edu (M.C.M.); mpvarela@utep.edu (M.P.B.-V.)
2 Computational Science Program, University of Texas at El Paso, El Paso, TX 79968, USA; wkubin@miners.utep.edu (W.K.); pkasante@miners.utep.edu (P.K.A.)
3 Comisión Nacional de Energía Atómica, C1429BNP Buenos Aires, Argentina; jaroszewo@tandar.cnea.gov.ar
4 Department of Geological Sciences, University of Texas at El Paso, El Paso, TX 79968, USA; hectorg@miners.utep.edu
5 Department of Seismology, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada 22860, Mexico

* Correspondence: oktweneboh@miners.utep.edu

Received: 7 May 2020; Accepted: 25 June 2020; Published: 30 June 2020

Abstract: Financial and geophysical data, like many other low and high frequency time series, are known to exhibit some memory effects. These memory effects may be long or short, permanent or temporal depending on the event that is being modeled. The purpose of this study is to investigate the memory effects characterized by the financial market closing values and volcanic eruption time series as well as to investigate the relation between the self-similar models used and the Lévy process. This paper uses highly effective scaling methods including Lévy processes, Detrended Fluctuation Analysis (DFA) and Diffusion Entropy Analysis (DEA) to examine long-range persistence behavior in time series by estimating their respective parameters. We use the parameter of the Lévy process \(\alpha\) characterizing the data and the scaling parameters of DFA \(H\) and DEA \(\delta\) characterizing the self-similar property to generate a relationship between the three (3) aforementioned scaling methods. Findings from the numerical simulations confirm the existence of long-range persistence (long-memory behavior) in both the financial and geophysical time series. Furthermore, the numerical results from this study indicates an approximate inverse relationship between the parameter of the Lévy process and the scaling parameters of the DFA and DEA (i.e., \(H, \delta \approx \frac{1}{2}\)), which we prove analytically.

Keywords: self-similar models; Lévy flight model; Diffusion Entropy Analysis; Detrended Fluctuation Analysis; long-range correlation

1. Introduction

In recent years, researchers have developed several techniques to predict and forecast real-world events or processes that are seemingly difficult to predict due to their complex nature. Such studies have produced major improvements in predictions related to financial market crashes, volcanic activities, climates, heartbeats, etc. Mandelbrot (1982), in the article “The Fractal Geometry of Nature”, describes the surprisingly excellent resemblance between real landscapes and geometrically generated landscapes using simple fractal models [1]. He pointed out that fractals can exhibit self-similarity in enormous ways.
A small portion of snowflake depicts similar statistical properties as the whole: a simple example of a self-similar process. This is illustrated by the Koch snowflake curve shown in Figure 1.

![Koch Snowflake Curve](image)

**Figure 1.** Figure of Koch snowflake curve depicting self-similarity.

Self-similar processes are used to describe stochastic processes that exhibit the phenomenon of self-similarity. These processes occur in several areas of applied mathematics such as fractals, chaos theory, long-memory processes and spectral analysis. That a self-similar process exhibit long-memory effects implies the dependence in time series decays more slowly than an exponential decay, represented as a power-law relation. Let us now take a look at the formal definition of a self-similar process.

**Definition 1 (Self-Similarity).** *In the sense of finite dimensional distributions* $fdd$, *we define a stochastic process* $X_t$ *as self-similar if two processes* $\{X_{at}\}_{t \geq 0}$ *and* $\{a^H X_t\}_{t \geq 0}$ *have the same law if there exists* $H > 0$. *i.e.,*

$$\{X_{at}\}_{t \geq 0} fdd = \{a^H X_t\}_{t \geq 0}$$

(1)

where $a > 0$ is a scale factor and $H$ is the self-similar parameter of the stochastic process $X_t$ [2].

The variances of financial market indices and seismic energy released from volcanic eruptions are stochastic. They are time-varying and they rise during high fluctuation periods. The self-similar properties exhibited by such events are determined by estimating the trends in smaller windows of time over the entire series. This method, known as the detrended fluctuation analysis (DFA), was developed by Peng et al. in the paper “Mosaic Organization of DNA Nucleotides” [3]. The DFA converts a series into a random walk and uses least squares to estimate trends and fluctuations [4,5]. This technique is capable of detecting long-range persistence (self-similar pattern) in both stationary and non-stationary time series.

An alternative to the DFA technique is the diffusion entropy analysis (DEA) model developed by Scafetta et al. (2002), also called a pdf-scaling method. The DEA can be used to investigate if a time series is characterized by Gaussian or Lévy distribution [6,7]. It also permits the detection of long-range persistence in time series. Scafetta (2003) argues that the DEA is the only method that correctly quantifies the self-similarity parameter of time series of complex processes [6].

Researchers have applied these models to diverse fields including DNA sequences [8–10], neural oscillations [11], speech pathology detection [12], and heartbeat fluctuation in different sleep stages [13]. The suite of models offered by Peng (1994) and Scafetta (2002) prove to be some of the best techniques for detecting long-range persistence in stationary and non-stationary time series.
Unlike non-Gaussian Lévy processes that perform poorly in long-range correlation scales, Lévy flights have been used to model evolution of complex stochastic events [14–16]. The Lévy flight is a stochastic process that is based on the random succession of steps on space (random walk theory) and probability distribution of the step-sizes which are known to have heavy-tails. In particular, Lévy flights characterizes random steps on a discrete space while Lévy walk describes random steps on a continuous space. However, researchers use “Lévy flights” to describe both cases in literature sometimes [17,18]. Lévy flights have been used to model several events including financial market crashes and trajectories followed by animals when they hunt for food [19–21], etc. This is as a result of the generalized central limit theorem (GCLT). When a large collection of step-sizes of a random trajectory satisfy the power-law condition, the Lévy flight distribution is stable: the reason for choosing the Lévy flight distribution to characterize the financial and geophysical processes exhibited by the real-world time series used herein.

In the paper “detecting market crashes by analyzing long memory effects using high-frequency data” by Barany et al. (2012), the authors estimated long-memory behavior in high-frequency financial time series using the DFA, Rescaled-Range Analysis (R/S) and the Lévy flight model to detect market crashes. They go further to generate an inverse relationship between the Hurst exponent and the Lévy model parameter. This close relationship proposes a satisfactory self-similar pattern for analysis [2].

In this paper, we generate a numerical relationship between the Lévy parameter $\alpha$ characterizing the data and the resulting parameters ($H$, and $\delta$) of the DFA and DEA characterizing the self-similar property respectively. Furthermore, we prove analytically the relationship obtained. To perform these tasks, we first investigate the existence of long-range persistence in sets of financial market indices and recordings of volcanic eruptions using the DFA, DEA. We report on the scaling parameters of the DFA and DEA and conclude that the time series exhibits long-range correlation (or long-range persistence) if $0.5 < H, \delta < 1$, and anti-persistent if $0 < H, \delta < 0.5$. Useful tables of results and plots of the fitted models are reported.

The outline of this paper is as follows. Section 2 details the scaling methods and description of the real-world data used in this paper. Section 3 contains the numerical results and proposed relationships as well as proofs. Section 4 provides a detailed discussion of the results. Section 5 concludes. Some proofs, figures and plots generated from our numerical simulations are provided in the Appendix.

2. Methods

In this section, we describe three techniques for estimating self-similarity or long-range correlations in time series. There are several methods used to estimate scaling exponents of time series to investigate long-memory behavior. In this study, we limit ourselves to Lévy flight model, DFA and DEA techniques for quantifying self-similarity. Note that $\alpha, H, \delta$ are used to denote the Lévy, DFA and DEA scaling exponents respectively. An overview of the real-world data used for the purpose of evaluating and comparing these three scaling methods are also described.

2.1. Lévy Flight Model

The stable Lévy process (Lévy flight model) has independent increments but are designated as long memory processes [2]. Lévy et al. [22] discovered that the most general representation of a stable distribution is through the characteristic functions $\varphi(q)$. These characteristic functions are defined as:

$$\ln(\varphi(q)) = \mu q - \gamma \left| q \right| \left[ 1 + i \beta \frac{q}{\left| q \right|} \frac{2}{\pi} \log(\left| q \right|) \right], \text{if } \alpha = 1$$ (2)

and

$$\ln(\varphi(q)) = i \mu q - \gamma \left| q \right|^\alpha \left[ 1 - i \beta \frac{q}{\left| q \right|} \tan \left( \frac{\pi \alpha}{2} \right) \right], \text{if } \alpha \neq 1$$ (3)
where the scaling parameter of the Lévy: $0 < \alpha \leq 2$, $\gamma$ is a positive scale factor ($\gamma^2 = \sigma^2$ (variance)), $\mu$ is a real-valued constant known as the location parameter (or mean) and $-1 \leq \beta \leq 1$ is the skewness parameter.

Analytically, the stable Lévy distribution is only known for a few values of $\alpha$ and $\beta$. Consider the centered symmetric distribution ($\beta = 0$, $\mu = 0$), the characteristic function is:

$$\phi(q) = \exp \left(-\frac{q^\beta}{\sigma^\beta}\right)$$

Then, the stable Lévy distribution (Fourier transform of (4)) is defined as

$$P_L(x) = \frac{1}{\pi} \int_0^\infty \exp \left(-\frac{|q|^\alpha}{\sigma^\alpha}\right) \cos(qx) dq$$

To avoid the complications that emerge in the infinite second moment (i.e., the fact that stable Lévy processes with $\alpha < 2$ have infinite variance), consider a stochastic process with finite variance that follows scale relations called Truncated Lévy Flight (TLF) (Mantegna and Stanley (1994)).

The TLF distribution (a non-stable distribution with finite variance) is defined by

$$T(x) = cP(x)\chi_{[-l,l]}(x)$$

where $P(x)$ is a symmetric Lévy distribution. However, depending on the magnitude of the cut-off length $l$, it relaxes convergence. For a smaller cut-off length $l$, convergence is faster which generates an instantaneous cut in its tails. [14] To correct this problem, Koponen (1995) proposed a TLF with characteristic function shown in (7) below in which the cut function is a decreasing exponential function with parameter $l$.

$$\phi(q) = \exp \left[c_0 - c_1 \frac{(q^2 + 1/l)^\alpha/2}{\cos(\pi\alpha/2)} \cos(\alpha \arctan(l | q |)) \right]$$

and scale factors:

$$c_1 = \frac{2\pi \cos(\pi\alpha/2)}{a \Gamma(\alpha) \sin(\pi\alpha)} A_l, \quad c_0 = \frac{2\pi}{a \Gamma(\alpha) \sin(\pi\alpha)} A_l^{-\alpha} l$$

We obtain $T = N\Delta t$ if we discretize in time with steps $\Delta t$. Thus, we must calculate the sum of $N$ stochastic variables that are independent and identically distributed (i.i.d) at each interval.

Given a small sample of i.i.d random variables $N$, its probability distribution is very close to the stable Lévy distribution. The standardized (improved) model is given by:

$$\ln \varphi_q(q) = \ln \phi \left(\frac{q}{\sigma}\right) = c_0 - c_1 \left(\frac{(q/\sigma)^2 + 1/l^2}{\cos(\pi\alpha/2)} \right)^{\alpha/2} \cos \left(\alpha \arctan \left(\frac{l | q |}{\sigma}\right)\right)$$

$$\ln \varphi_q(q) = \frac{2\pi A_l^{-\alpha} l}{a \Gamma(\alpha) \sin(\pi\alpha)} \left[1 - \left(\frac{(q/\sigma)^2 + 1}{l^2}\right)^{\alpha/2} \cos \left(\alpha \arctan \left(\frac{q l}{\sigma}\right)\right)\right]$$

if the variance equals 1. See [16] for more details.

The Equation (10) represents the normalized Lévy model applied in this study. In the numerical simulations, we adjust the values of $A_l$, the arbitrary scale parameter $l$ and the characteristic exponent $\alpha$ to best fit the cumulative function. Please refer to [2,23–25] for more details and Appendix B for figures.

### 2.2. Detrended Fluctuation Analysis (DFA)

The DFA method is an important technique in revealing long range correlations in both stationary and non-stationary time series by estimating its scaling exponent. The numerical procedure to estimate the DFA exponent is presented below.

Let $N$ be the length of time series $(y_1, y_2, y_3, ..., y_N)$. The logarithmic ratio of the time series is obtained and the length of the new time series $M(t)$ is $N - 1$. 

Mathematics 2020, 8, 1046
\[ M(t) = \log\left( \frac{y_{t+1}}{y_t} \right), \quad t = 1, 2, ..., N - 1 \]  \hspace{1cm} (11)

The absolute value of \( M(t) \) is integrated:

\[ y(t) = \sum_{i=1}^{t} | M(i) | \]  \hspace{1cm} (12)

The integrated time series of length \( N \) is divided into \( m \) boxes of equal length \( n \) with no intersection between them. As the data is divided into equal length intervals, there may be some leftover at the end. To take account of these leftover values, the same procedure is repeated but beginning from the end, obtaining \( 2N/n \) boxes.

A least square line is fitted to each box, representing the trend in each box, thus obtaining \( y_n(t) \).

Finally the Root Mean Square Fluctuation (RMSF) is calculated using the formula

\[ F(n) = \sqrt{\frac{1}{2N} \sum_{t=1}^{2N} [y(t) - y_n(t)]^2} \]  \hspace{1cm} (13)

This computation is repeated over all box sizes to characterize a relationship between the box size \( n \) and \( F(n) \). A linear relationship between the \( F(n) \) and \( n \) (i.e., box size) in a log-log plot reveals that the fluctuations can be characterized by a scaling exponent \( H \), the slope of the line relating \( \log F(n) \) to \( \log n \). This generates the mathematical (power-law) relation:

\[ F(n) \propto n^H \]  \hspace{1cm} (14)

Next we discuss the Diffusion Entropy Analysis (DEA).

2.3. Diffusion Entropy Analysis (DEA)

The DEA can be used to analyze and detect the scaling properties of low and high frequency time series.

A function \( \Phi(r_1, r_2, ...) \) is said to be invariant to scale changes if it fulfills the property:

\[ \Phi(r_1, r_2, ...) = \gamma^a \Phi(\gamma^b r_1, \gamma^c r_2, ...) \]  \hspace{1cm} (15)

This equation shows that if we scale all coordinates \( r \) with an appropriate choice of exponents \( a, b, c, ... \), then we always recover the original function. In the case of a time series we can interpret the sequence of the numbers that form it as the generators of a diffusion process. Thus, we study the relevant probability distribution function \( \rho(x, t) \) where \( x \) denotes the variable that collects the fluctuations of the diffusion process over time span \( t \). If the time series is stationary, the scale property takes the shape

\[ \rho(x, t) = \frac{1}{\delta^t} \rho\left( \frac{x}{\delta} \right), \]  \hspace{1cm} (16)

where the \( \delta \) is the scaling exponent.

To determine the scaling exponent of the DEA, first transform the series into a diffusion process and calculate its Shannon’s entropy.

Consider the equation

\[ S(t) = - \int_{-\infty}^{\infty} \rho(x, t) \ln [\rho(x, t)] \, dx \]  \hspace{1cm} (17)

Suppose that \( \rho(x, t) \) satisfies the scale condition (16), replacing it in the (17) we obtain:
\[ S(t) = A + \delta \ln(t) \quad (18) \]

Please see proof in Appendix A.

Equation (18) describes the fact that entropy grows linearly with \( \ln(t) \) and the slope of the linear function is the scaling exponent \( \delta \).

Let us now see how we generate the diffusion process for a sequence of \( M \) numbers:

\[ \epsilon_i, i = 1, \ldots, M \quad (19) \]

The purpose of the DEA algorithm is to establish the possible existence of scalability in an efficient way without altering the series with any form of detrending. We do the following:

1. Select an integer \( l \) such that \( 1 \leq l \leq M \).
2. For each time, find \( M - l + 1 \) sub-series of length \( l \) defined such that

\[ \epsilon_i^s \equiv \epsilon_{i+s}, i = 1, \ldots, l \quad (20) \]

with \( s = 0, \ldots, M - l \).
3. For each sub-series, construct a distribution diffusion path by the position

\[ x^{(s)}(l) = \sum_{i=1}^{l} \epsilon_i^s = \sum_{i=1}^{l} \epsilon_{i+s} \quad (21) \]

4. To calculate the Shannon entropy of the diffusion process, partition the \( x - \text{axis} \) into size \( \eta(l) \) cells and tell how many parts there are in each of them at a time \( l \). Let us call this number \( N_i(l) \).
5. With \( N(l) \), we determine the probability \( p_i(l) \) that a particle is found in the \( i \)-th cell at time \( l \).

\[ p_i(l) = \frac{N_i(l)}{M - l + 1} \quad (22) \]

6. The entropy of the diffusion process at time \( l \) is given by

\[ S_d(l) = - \sum_i p_i(l) \ln[p_i(l)] \quad (23) \]

The sub-index \( d \) denotes that the entropy evaluation was generated from a discrete process. The reader should refer to [7,26–28] for more details.

2.4. Background of Data

In this subsection, we examine data from two different fields namely Financial Markets and Volcanic Eruptions and their stationary behavior.

2.4.1. Financial Market Data

The financial market data used for the analysis in this paper was obtained from YAHOO FINANCE. All data points collected are daily close values. Below are the names of countries or cities where data was collected as well as the respective start and end dates of data collection: Brazil (BOVESPA), from 27 April 1993 to 24 June 2005; USA (SPC), from 2 January 1991 to 25 October 2001; Hong Kong (HSI), from 2 January 1991 to 25 October 2001; Chile (IGPA), from 28 April 1998 to 24 October 2001; Argentina (MERVAL), from 8 October 1996 to 22 October 2001; Mexico (MXX), from 8 November 1991 to 22 October 2001; USA (NASDAQ), from 22 January 1997 to 31 December 2001; Phillipines (PSI), from 2 July 1997 to 25 October 2001; Thailand (SETI), from 2 July 1997 to 25 October 2001 and USA (S&P500) from 2 January 1991 to 25 October 2001. This data was first used for
big data analysis in [16] and recently used in [7]. The price evolution of the financial market time series are presented in Figures 2–6.

**Figure 2.** Plots of trading days closings and actual returns of BVSP and SPC.

**Figure 3.** Plots of trading days closings and actual returns of HSI and IGPA.

**Figure 4.** Plots of trading days closings and actual returns of MERV and MXX.

**Figure 5.** Plots of trading days closings and actual returns of Nasdaq and PSI.
2.4.2. Geophysical Time Series Data

The Bezymianny Volcano Campaign Seismic Network (PIRE) collected the volcanic data at different eruption times at two different seismic stations, namely BEZB and BELO. Data used in this paper were requested for 10 days before and 5 days after the published time of the volcanic eruptions. Volcanic eruptions 1 and 2 were from BEZB and Volcanic eruptions 3–8 were from BELO. Please see [24] for more details on the geophysical time series used in this study. The plots of the volcanic eruptions time series are presented in Figures 7 and 8.

2.4.3. Stationarity Test of Data

We use the Augmented Dickey-Fuller (ADF) test to examine the stationarity of both the returns of the financial market data and volcanic data. The returns of the financial market data is calculated as:

\[ \text{market returns } (r_t) = \frac{X_t - X_{t-1}}{X_{t-1}} \]
where $X_t$ is the current market closing value and $X_{t-1}$ is the previous market closing value.

2.4.4. Augmented Dickey-Fuller (ADF)

The ADF statistic tests the null hypothesis that the time series has a unit root against the alternative that it is stationary (or has no unit root) with the assumption that the data exhibits an Auto-Regressive Moving Average (ARMA) structure.

$$H_0 : \gamma = 0$$
$$H_a : \gamma < 0$$

The value of the ADF test statistic is computed as

$$ADF_t = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

As this test is asymmetrical, we are only concerned with negative values when it comes to the test statistic. If the calculated test statistic, $ADF_t$, is less (more negative) than the critical value, the null hypothesis is rejected and we conclude that no unit root is present and vice versa.

Table 1. ADF test results for returns of various financial markets.

| Financial Market | Test Statistic | Lag Order | p-Value |
|------------------|----------------|-----------|---------|
| BVSP             | -16.571        | 12        | 0.01    |
| SPC              | -19.992        | 13        | 0.01    |
| HSI              | -18.937        | 13        | 0.01    |
| IGPA             | -10.87         | 9         | 0.01    |
| MERV             | -13.298        | 10        | 0.01    |
| MXX              | -12.496        | 13        | 0.01    |
| Nasdaq           | -7.4649        | 10        | 0.01    |
| PSI              | -10.207        | 10        | 0.01    |
| SETI             | -8.4743        | 10        | 0.01    |
| SP500            | -30.856        | 24        | 0.01    |

Table 2. ADF test results for different eruption times of Volcano at the two seismic stations.

| Station | Eruption | Test Statistic | Lag Order | p-Value |
|---------|----------|----------------|-----------|---------|
| BEZB    | 1        | -3.155         | 7         | 0.10    |
|         | 2        | -1.946         | 7         | 0.60    |
|         | 3        | -3.188         | 8         | 0.09    |
|         | 4        | -3.181         | 8         | 0.09    |
|         | 5        | -3.342         | 8         | 0.07    |
|         | 6        | -3.366         | 7         | 0.06    |
|         | 7        | -2.784         | 7         | 0.25    |
|         | 8        | -2.092         | 7         | 0.54    |

Based on the values obtained in Tables 1 and 2, at 95% level, the null hypothesis of a unit root is rejected because the $ADF_t$ test statistics of most data points is more negative than the critical value of $-2.86$ from the critical values for Dickey-Fuller $t$-distribution [29]. We further confirm this conclusion using the p-values at a significant level of $\alpha = 5\%$. All $p$-values of both the returns of the financial market data and the volcanic eruption data are less than the significance level (0.05).

We reject the null hypothesis that the time series has a unit root and conclude that both the returns of financial market data and the data from different eruption times of the volcano are stationary.
3. Results

In this section, we present numerical results obtained from applying all 3 models to the data. From our numerical results we propose a mathematical relationship between the self-similar models and the Lévy model. We also present an analytical proof of the numerically derived relationship.

3.1. Numerical Results

Below are tables for both the financial markets data and volcanic eruption data. There are six (6) columns in each table where the first column indicates the particular financial market or volcanic eruption numbers, the second - fourth columns show the scaling exponents of the DFA and the DEA as well as the best Lévy parameter \( \alpha \) characterizing the data. The last three columns are basically products of the respective scaling exponents of the self-similar models \( H, \delta \) and the Lévy parameter \( \alpha \) characterizing the data.

3.1.1. Remarks on the Tables of Numerical Results

Empirical evidence presented in Tables 3 and 4 show that the product of respective scaling exponents of the DFA and Lévy as well as DEA and Lévy approximately equal 1. This suggests that there exist an inverse relationship between the DEA parameter \( \delta \) and the Lévy parameter \( \alpha \) i.e.:

\[
\delta \approx \frac{1}{\alpha} \tag{24}
\]

Also, an inverse relationship exists between the DFA parameter \( H \) and the resulting Lévy parameter \( \alpha \) i.e.:

\[
H \approx \frac{1}{\alpha} \tag{25}
\]

| Markets | DFA (\( H \)) | DEA (\( \delta \)) | Lévy (\( \alpha \)) | \( H.\alpha \) | \( \delta.\alpha \) |
|---------|---------------|---------------------|---------------------|----------------|----------------|
| BVSP    | 0.72          | 0.57                | 1.34                | 0.96           | 0.76           |
| SPC     | 0.62          | 0.60                | 1.40                | 0.87           | 0.84           |
| HSI     | 0.70          | 0.60                | 1.40                | 0.98           | 0.84           |
| IGPA    | 0.65          | 0.53                | 1.40                | 0.91           | 0.74           |
| MERV    | 0.62          | 0.56                | 1.40                | 0.87           | 0.78           |
| MXX     | 0.66          | 0.59                | 1.34                | 0.90           | 0.79           |
| Nasdaq  | 0.72          | 0.56                | 1.12                | 0.96           | 0.66           |
| PSI     | 0.71          | 0.55                | 1.40                | 0.99           | 0.77           |
| SETI    | 0.70          | 0.54                | 1.34                | 0.94           | 0.72           |
| SP500   | 0.66          | 0.65                | 1.40                | 0.92           | 0.91           |

| Seismic Station | Eruption Number | DFA (\( \alpha \)) | DEA (\( \delta \)) | Lévy (\( \alpha \)) | \( H.\alpha \) | \( \delta.\alpha \) |
|-----------------|-----------------|---------------------|---------------------|---------------------|----------------|----------------|
| BEZB            | 1               | 0.74                | 0.68                | 1.12                | 0.83           | 0.76           |
|                 | 2               | 0.92                | 0.68                | 1.34                | 1.23           | 0.91           |
|                 | 3               | 0.85                | 0.68                | 1.12                | 0.95           | 0.76           |
|                 | 4               | 0.66                | 0.68                | 1.40                | 0.92           | 0.95           |
|                 | 5               | 0.76                | 0.68                | 1.12                | 0.85           | 0.76           |
|                 | 6               | 0.67                | 0.68                | 1.34                | 0.90           | 0.91           |
|                 | 7               | 0.81                | 0.68                | 1.40                | 1.13           | 0.95           |
|                 | 8               | 0.75                | 0.68                | 1.34                | 1.01           | 0.91           |
3.2. Analytical Relations

In this section, we prove analytically the numerical relations obtained from the previous subsection.

**Property 1** (Stable distributions). Given a random variable \( X \), if we denote with \( \text{Law}(X) \) its probability density function (for example, for a Gaussian random variable, we write \( \text{Law}(X) = N(\mu, \sigma^2) \)) then we will say that the random variable \( X \) is stable, or that it has a stable distribution if for any \( n \geq 2 \) there exists a positive number \( C_n \) and a number \( D_n \) so that:

\[
\text{Law}(X_1 + X_2 + \ldots + X_n) = \text{Law}(C_n X + D_n)
\]

(26)

where \( X_1, X_2, \ldots, X_n \) are independent random copies of \( X \), this means that \( \text{Law}(X_i) = \text{Law}(X) \) for \( i = 1, 2, \ldots, n \). If \( D_n = 0 \), then \( X \) is said to be a strictly stable variable. It can be shown (see Samorodnitsky and Taqqu (1994)) that

\[
C_n = n^{1/\alpha}
\]

(27)

for some parameter \( 0 < \alpha \leq 2 \).

Equation (27) can be rewritten as

\[
\log C_n = \frac{1}{\alpha} \log n
\]

(28)

3.2.1. Detrended Fluctuation Analysis (DFA)

From (14), we obtain a mathematical relation that reveals a linear relationship between the root mean square fluctuation \( F(n) \) and the box size \( n \) in a log-log plot. i.e.,

\[
\log F(n) = \log K + H \log(n)
\]

(29)

where \( K \) is a positive constant of proportionality.

\[
\log F(n) - \log K = H \log(n)
\]

(30)

Subtracting (30) and (28) gives the relation:

\[
H = 1/\alpha
\]

since

\[
\lim_{n \to \infty} \left| \frac{\log F(n) - \log K - \log C_n}{\log n} \right| \to 0
\]

(31)

3.2.2. Diffusion Entropy Analysis (DEA)

From the probability density function (pdf) in (16), one can calculate the Shannon entropy [2,23],

\[
S(t) = -\int_{-\infty}^{\infty} \rho(x,t) \ln [\rho(x,t)] \, dx
\]

(32)

Suppose the pdf \( \rho(x,t) \) satisfies (16), then after a few simple steps (see appendix for proof) [23], we obtain the linear-log relation

\[
S(t) = A + \delta \ln(t)
\]

(33)

where Equation (33) becomes

\[
S(t) - S(1) = \delta \ln(t)
\]

(34)
From property 1, for some positive number \( C_n \), subtracting (34) and (28) gives

\[
|S(t) - S(1) - \ln(C_n)| = 0, \text{ where } t \geq 2
\]

such that

\[
\delta = \frac{1}{\alpha}
\]

4. Discussion

The three techniques described in Section 2 were used to estimate the scaling exponents of the financial and geophysical time series. Tables 3 and 4 summarizes the results of each technique applied to the real-world data. Figures A1–A18 in the appendix provide a graphical summary of results from techniques used in this paper. The results confirm an inverse relationship between the self-similar parameters \((H, \delta)\) and the Lévy flight parameter. We also observe the existence of long-range persistence in the financial and geophysical time series which confirms seasonal or irregular occurrence of real-world events such as stock market crashes and volcanoes of the past. They are helpful in the forecasting of such events using past and recent information.

Having proved the existence of an inverse relationship between the Lévy parameter and the scaling parameters of the DFA and DEA, in the event that the scaling exponent of either of the parameter estimators is not known or expensive to estimate, we are able to approximate it using the relationship generated in this study. This is because data keeps getting bigger, making the measurements and collections of big data more expensive and time-consuming.

In future work, we seek to investigate whether some of the scaling methods significantly outperform others using real-world time series with removed components such as trends, seasonality and irregularities (noise). This experiment is essential to undertake partly due to complications in measurement and collection of real-world data especially in fast-paced industries such as the financial sector.

5. Conclusions

In this study, we compare and quantify the scaling parameters depicting long-range persistence behavior of a set of financial and geophysical time series using the DFA and DEA techniques. We also characterize the respective time series with Lévy process (with stable distributions) to obtain the best Lévy flight parameter. Both the financial and the geophysical time series exhibit long-range correlations since \(0.5 < H, \delta < 1\). Empirical evidence from results suggest an inverse relationship between the DFA parameter \(H\) and the resulting Lévy parameter \(\alpha\). Similarly, an inverse relationship exists between the DEA parameter \(\delta\) and the Lévy parameter \(\alpha\).

Finally, we present proofs to verify these mathematical relationships. However, one has to note that the choice of the \(\alpha\) parameter for the Lévy processes means they are stochastic. They are influenced by the respective time variables, \(T = 1, 4, 8, 16\). The appropriate Lévy parameter is selected at appropriate time periods. Another limitation of this study is that, we adjust the values of \(A\), the arbitrary scale parameter \(l\) and the characteristic exponent \(\alpha\) of the Lévy flight model in the numerical simulations to get the best fit of the cumulative distribution function of the Lévy process. It is also worth noting that because we standardized the Lévy model we should only expect an approximate inverse relationship between the self-similar parameters \((H, \delta)\) and the Lévy parameter \((\alpha)\).

Author Contributions: Conceptualization, M.C.M.; Data curation, H.G.-H.; Formal analysis, W.K.; Methodology, M.C.M., W.K., P.K.A., O.K.T., M.P.B.-V. and S.J.; Software, W.K., M.P.B.-V. and S.J.; Supervision, M.C.M.; Validation, M.C.M. and O.K.T.; Visualization, H.G.-H.; Writing—original draft, W.K.; Writing—review & editing, O.K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.
Appendix A

This appendix contains the proof the DEA linear-log relation in Equation (33). The continuous version of the Shannon entropy reads

\[ S(t) = - \int_{-\infty}^{\infty} \rho(x, t) \ln[\rho(x, t)] \, dx \]  

(A1)

assuming that

\[ \rho(x, t) = \frac{1}{t^\vartheta} F\left(\frac{x}{t^\vartheta}\right) \]  

(A2)

and that \( F(y) \) maintains its form, namely that the statistics of the process are independent of time.

Plugging Equation (A2) into Equation (A1), we get

\[ S(t) = - \frac{1}{t^\vartheta} \int_{-\infty}^{\infty} F\left(\frac{x}{t^\vartheta}\right) \ln\left(\frac{1}{t^\vartheta} F\left(\frac{x}{t^\vartheta}\right)\right) \, dx \]  

(A3)

\[ = - \frac{1}{t^\vartheta} \ln\left(\frac{1}{t^\vartheta}\right) \int_{-\infty}^{\infty} F\left(\frac{x}{t^\vartheta}\right) \, dx - \frac{1}{t^\vartheta} \int_{-\infty}^{\infty} F\left(\frac{x}{t^\vartheta}\right) \ln\left(\frac{x}{t^\vartheta}\right) \, dx \]  

(A4)

Let \( y = \frac{x}{t^\vartheta} \Rightarrow dx = t^\vartheta dy \). Substituting for \( dx \), we get

\[ = - \frac{1}{t^\vartheta} \ln\left(\frac{1}{t^\vartheta}\right) \int_{-\infty}^{\infty} F\left(y t^\vartheta\right) \, dy - \frac{1}{t^\vartheta} \int_{-\infty}^{\infty} F\left(y\right) \ln\left(F\left(y\right)\right) t^\vartheta \, dy \]  

After canceling out some common terms, we obtain the following

\[ = - \ln\left(\frac{1}{t^\vartheta}\right) \int_{-\infty}^{\infty} F\left(y\right) \, dy - \int_{-\infty}^{\infty} F\left(y\right) \ln\left(F\left(y\right)\right) \, dy \]  

(A5)

\[ = - \left[ \ln(1) - \ln\left(t^\vartheta\right) \right] \int_{-\infty}^{\infty} F\left(y\right) \, dy - \int_{-\infty}^{\infty} F\left(y\right) \ln\left(F\left(y\right)\right) \, dy \]  

\[ = \delta \ln(t) \int_{-\infty}^{\infty} F\left(y\right) \, dy - \int_{-\infty}^{\infty} F\left(y\right) \ln\left(F\left(y\right)\right) \, dy \]  

(A6)

Equation (A6) produces the linear-log relation

\[ S(t) = \delta \ln(t) + A \]

where

\[ A = - \int_{-\infty}^{\infty} F(y) \ln[F(y)] \, dy \]  

(A7)

and the Cumulative Density Function (CDF) equals 1.

Appendix B

Plots obtained from the numerical simulation of the various financial and geophysical time series data are shown in Figures A1–A18 in this appendix.

• BLUE points depict the Normalized Data
• RED curve depicts the Cumulative Distribution Function
• GREEN curve depicts the Gaussian (Normal) distribution fit.

Figure A1. Lévy model fits for PSI and SETI.

Figure A2. Lévy model fits for SP500 and SPC USA.

Figure A3. DFA for PSI and SETI.
Figure A4. DFA for SP500 and SPC USA.

Figure A5. DEA for PSI and SETI.

Figure A6. DEA for SP500 and SPC USA.
Figure A7. Lévy model fit for Volcanic Eruptions 1 and 2.

Figure A8. Lévy model fit for Volcanic Eruptions 3 and 4.

Figure A9. Lévy model fit for Volcanic Eruptions 5 and 6.
Figure A10. Lévy model fit for Volcanic Eruptions 7 and 8.

Figure A11. DFA for Volcanic Eruption 1 and 2.

Figure A12. DFA for Volcanic Eruption 3 and 4.
Figure A13. DFA for Volcanic Eruption 5 and 6.

Figure A14. DFA for Volcanic Eruption 7 and 8.

Figure A15. DEA for Volcanic Eruption 1 and 2.
(a) Volcanic Eruption 3

(b) Volcanic Eruption 4

Figure A16. DEA for Volcanic Eruption 3 and 4.

Figure A17. DEA for Volcanic Eruption 5 and 6.

Figure A18. DEA for Volcanic Eruption 7 and 8.

References
1. Mandelbrot, B.B. *The Fractal Geometry of Nature*; Times Books: New York, NY, USA, 1982; p. 44.
2. Barany, E.; Beccar Varela, M.P.; Florescu, I.; Sengupta, I. Detecting market crashes by analyzing long-memory effects using high-frequency data. *Quant. Financ.* 2012, 12, 623–634, doi:10.1080/14697688.2012.664937.
3. Peng, C.K.; Buldyrev, S.V.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L. Mosaic organization of DNA nucleotides. *Phys. Rev. E* 1994, 49, 1685–1689, doi:10.1103/physreve.49.1685.
4. Shlesinger, M.F.; West, B.J.; Klafter, J. Lévy dynamics of enhanced diffusion: Application to turbulence. *Phys. Rev. Lett.* 1987, 58, 11–16.
5. Ben-Avraham, D.; Havlin, S. *Diffusion and Reactions in Fractals and Disordered Systems*; Cambridge University Press: Cambridge, UK, 2000.
6. Scafetta, N. *An Entropic Approach to the Analysis of Time Series*; University of North Texas: Denton, TX, USA, 2003; p. 5906.

7. Mariani, M.C.; Asante, P.K.; Bhuiyan, M.A.M.; Beccar-Varela, M.P.; Jaroszewicz, S.; Tweneboah, O.K. Long-Range Correlations and Characterization of Financial and Volcanic Time Series. *Mathematics* 2020, 8, 441.

8. Buldyrev, S.V.; Goldberger, A.L.; Havlin, S.; Mantegna, R.N.; Matsa, M.E.; Peng, C.K.; Simons, M.; Stanley, H.E. Long-Range Correlation-Properties of Coding And Noncoding DNA-Sequences—Genbank Analysis. *Phys. Rev. E* 1995, 51, 508–5091.

9. Bunde, A.; Havlin, S. *Fractals and Disordered Systems*; Springer: Berlin/Heidelberg, Germany, 1996.

10. Kantelhardt, J.W.; Koscielny-Bunde, E.; Rego, H.H.A.; Havlin, S.; Bunde, A. Detecting Long-range Correlations with Detrended Fluctuation Analysis. *Physica A* 2001, 295, 441–454.

11. Hardstone, R.; Poil, S.S.; Schiavone, G.; Jansen, R.; Nikulin, V.V.; Mansvelder, H.D.; Linkenkaer-Hansen, K. Detrended Fluctuation Analysis: A Scale-Free View on Neuronal Oscillations. *Front. Physiol.* 2012, 3, 450.

12. Little, M.; McSharry, P.; Moroz, I.; Roberts, S. Nonlinear, Biophysically-Informed Speech Pathology Detection. *IEEE Int. Conf. Acoust. Speed Signal Process. Proc.* 2006, 2, 1080–1083.

13. Bunde, A. Correlated and uncorrelated regions in heart-rate fluctuations during sleep. *Phys. Rev. E* 2000, 85, 373–3739.

14. Mantegna, R.N; Stanley, H.E. Stochastic process with ultra-slow convergence to a Gaussian: the truncated Lévy flight. *Phys. Rev. Lett.* 1994, 73, 2946–2949.

15. Koponen, I. Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Phys. Rev. E* 1995, 52, 1197–1199.

16. Mariani, M.C.; Liu, Y. Normalized truncated Lévy walks applied to the study of financial indices. *Physica A* 2007, 377, 590–598.

17. Kleinberg, J.M. Navigation in a small world. *Nature* 2000, 406, 845.

18. Li, G.; Reis, S.D.S.; Moreira, A.A.; Havlin, S.; Stanley, H.E.; Andrade, J.S., Jr. Towards Design Principles for Optimal Transport Networks. *Phys. Rev. Lett.* 2010, 104, 018701.

19. Viswanathan, G.M.; Buldyrev, S.V.; Havlin, S.; da Luz, M.G.E.; Raposo, E.P.; Stanley, H.E. Optimizing the success of random searches. *Nature* 1999, 401, 910–914.

20. Reynolds, G. Navigating Our World Like Birds and some authors have claimed that the motion of bees. *The New York Times*, 2014.

21. Sims, D.W.; Reynolds, A.M.; Humphries, N.E.; Southall, E.J.; Wearmouth, V.J.; Metcalfe, B.; Twitchett, R.J. Hierarchical random walks in trace fossils and the origin of optimal search behavior. *Proc. Natl. Acad. Sci. USA* 2014, 111, 1107–1108.

22. Khintchine, A.Y.; Lévy, P. Sur les lois stables. *CR Acad. Sci. Paris* 1936, 202, 374–376.

23. Mariani, M.C.; Florescu, I; Beccar Varela, M.P.; Ncheuguim, E. Long correlations and Lévy models applied to the study of memory effects in high frequency (tick) data. *Phys. A Stat. Mech. Its Appl.* 2009, 388, 1659–1664.

24. Beccar-Varela, M.P.; Gonzalez-Huizar, H.; Mariani, M.C.; Tweneboah, O.K. Lévy Flights and Wavelets Analysis of Volcano-Seismic Data. *Pure Appl. Geophys.* 2020, 177, 723–736.

25. Beccar-Varela, M.; Gonzalez-Huizar, H.; Mariani, M.C.; Serpa, L.F.; Tweneboah, O.K. Chile2015: Lévy Flight and Long-Range Correlation Analysis of Earthquake Magnitudes in Chile. In *The Chile-2015 (Illapel) Earthquake and Tsunami*; Birkhäuser: Cham, Switzerland, 2017; pp. 113–122.

26. Scafetta, N.; Grigolini, P. Scaling detection in time series: Diffusion entropy analysis. *Phys. Rev. E* 2002, 66, 036130.

27. Clausius, R. Available online: http://www.ueltschi.org/teaching/chapShannon.pdf (accessed on 15 October 2019).

28. Clausius, R. Available online: http://galileo.phys.virginia.edu/classes/152.mf1i.spring02/Entropy.pdf (accessed on 10 October 2019).

29. Fuller, W.A. *Introduction to Statistical Time Series*; John Wiley and Sons: New York, NY, USA, 1976.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).