PHYSICAL REVIEW E **80**, 015101(R) (2009)

# Asymmetry in power-law magnitude correlations

Boris Podobnik*
*Faculty of Civil Engineering, University of Rijeka, 51000 Rijeka, Croatia;*
*Zagreb School of Economics and Management, 10000 Zagreb, Croatia;*
*and Department of Physics and Center for Polymer Studies, Boston University, Boston, Massachusetts 02215, USA*

Davor Horvatić
*Department of Physics, Faculty of Science, University of Zagreb, Bijenička c. 32, 10000 Zagreb, Croatia*

Joel N. Tenenbaum and H. Eugene Stanley
*Department of Physics and Center for Polymer Studies, Boston University, Boston, Massachusetts 02215, USA*
(Received 24 March 2009; published 17 July 2009)

Time series of increments can be created in a number of different ways from a variety of physical phenomena. For example, in the phenomenon of volatility clustering—well-known in finance—magnitudes of adjacent increments are correlated. Moreover, in some time series, magnitude correlations display asymmetry with respect to an increment's sign: the magnitude of $|x_i|$ depends on the sign of the previous increment $x_{i-1}$. Here we define a model-independent test to measure the statistical significance of any observed asymmetry. We propose a simple stochastic process characterized by a an asymmetry parameter $\lambda$ and a method for estimating $\lambda$. We illustrate both the test and process by analyzing physiological data.

DOI: 10.1103/PhysRevE.80.015101 PACS number(s): 02.50.−r, 05.40.−a, 87.19.le

The outputs of a broad class of systems ranging from physical and biological to social systems exhibit either long-range temporal or spatial correlations that can be approximated by power laws [1]. A variety of studies have also found that different complex systems spanning finance [2], physiology [3], and seismology [4,5] generate time series of increments, the magnitudes of which are power-law correlated. The correlation of these magnitudes, results in "clustering," where large increments are more likely to follow large increments and small increments are more likely to follow small increments.

Long-range magnitude correlations in increments $x_i$ are usually modeled using a time-dependent standard deviation, $\sigma_i$ [2], commonly called volatility, which essentially influences the size of subsequent increments. $\sigma_i$ is defined as a linear combination of $N$ previous values of $|x_{i-n}|$, i.e., $\sigma_i = \Sigma_{n=1}^{N} a(n)|x_{i-n}|$, where $i$ and $n$ are time scales, and $a(n)$ are statistical weights.

Magnitude correlations of many financial time series [6] are asymmetric with respect to increment sign, in that negative increments are more likely to be followed by increments of large magnitude and positive increments are more likely to be followed by increments with small magnitudes (i.e., "bad news" causes more volatility than "good news").

If we are to model this asymmetry, the standard deviation $\sigma_i$ we define must depend on both $x_{i-n}$ and $|x_{i-n}|$ to capture the dependence of both sign and magnitude. Since $\sigma_i$ must be positive, we can define $\sigma_i \equiv \Sigma_n a(n)(||x_{i-n}| + \lambda x_{i-n}|)$, where $\lambda$ is a real parameter that acts as a measure of asymmetry. For $\lambda > 0$, positive increments $x_{i-1}$ are more likely to be followed by large increments $|x_i|$ [see Fig. 1(a)], whereas for $\lambda < 0$, negative increments are more likely followed by large increments. $\lambda = 0$ reduces to the symmetric $\sigma_i$ above that has no dependence on the sign of the increment.

We ask if the concept of asymmetry in magnitude correlations is relevant to real physical data. We first create a test allowing us to determine whether an observed asymmetry is statistically significant. We then propose a stochastic process in order to (i) further test significance and (ii) model data as dependent on two parameters which characterize both the length of the power-law memory and its magnitude correlation asymmetry, parameters which we then demonstrate how to obtain. Finally, we apply our test to real-world physiological data to determine if there is statistically significant asymmetry in the magnitude correlations.

How would we know if an observed asymmetry is genuine and not due to a finite-size effect? For example, a finite-length time series generated by an independent and identically distributed (iid) (i.e., uncorrelated) process will exhibit a spurious asymmetry. To this end, we ask how large should the asymmetry be to become statistically significant? To answer this question, we generate iid series with zero mean and unit variance, and for each we calculate two sums, $S_+$ and $S_-$. The sum $S_+$ is the average of all the values $|x_i|$ preceded by positive $x_{i-1}$, while the sum $S_-$ is the average of all the values $|x_i|$ preceded by negative $x_{i-1}$. For an infinitely long iid time series, we expect $S_+ = S_-$, while finite-length time series in general have $S_+ \neq S_-$.

We therefore define a test variable:

$$S \equiv S_+ - S_-. \tag{1}$$

What is the range $(-S_c, S_c)$ such that $S$ will fall in this range 95% of the time? To answer this question, we generate a large number of finite iid time series, each with $N$ data points. For each time series we calculate $S$. We find on collecting all the $S$ values that $S$ follows a symmetrical probability distribution $P(S)$ centered at zero. By ranking the val-

*bp@phy.hr

 

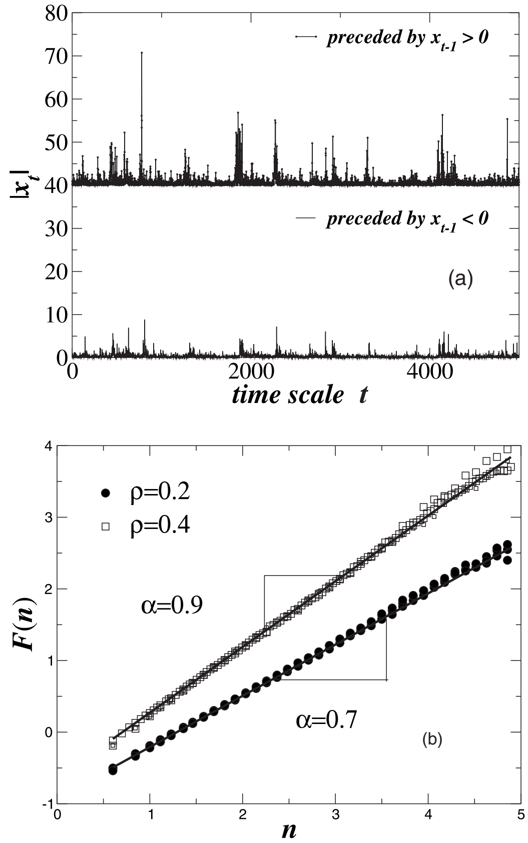


FIG. 1. Asymmetry in magnitude correlations and detrended fluctuation function $F(n)$. (a) We show that the increments are larger for $x_{i-1}>0$ (top curve, shifted upward for clarity) than for $x_{i-1}<0$ (bottom curve) as a result of positive $\lambda$. The time series is obtained from numerical simulations of the process of Eq. (2) with $\lambda=0.9$ and $\rho=0.4$. (b) Detrended fluctuation function $F(n)$, where $n$ is a time scale, obtained from numerical simulations of process of Eq. (2) with $\lambda=0.3$, 0.6, and 0.9 and $\rho=0.2$ and 0.4. For asymptotically large values of $n$, each of the $F(n)$ curves can be approximated by a power law $F(n) \propto n^{\alpha}$ with scaling exponent $\alpha \approx 0.5+\rho$ independent of the value of $\lambda$.

ues $S$ from smallest to largest, we find a critical value $S_c$ for which there is probability 0.95 that the $S$ of a random uncorrelated series is between $(-S_c, S_c)$. By repeating the same procedure for a different number of data points, in Fig. 2 and in the inset, we find an almost perfect power-law fit relating the critical value $S_c$ to the number of data points with exponent $0.5 \pm 0.006$ in agreement with the central limit theorem.

To find critical values for empirical series, we also use another approach of Ref. [5]. For a given series, we accomplish $10^4$ reshufflings, where each reshuffled time series is subtracted from the average and divided by its standard deviation. For each series, we calculate $S$ of Eq. (1). By ranking the values $S$ in ascending order, we find $S_c$ for which there is probability 0.95 that the $S$ is between $(-S_c, S_c)$. By using this approach, for subjects 02 and 08 we find $S_c=0.019$ and $S_c=0.021$, respectively.

We next argue that the interval $(-S_c, S_c)$ found for a given $N$ is a "litmus test" for significance. If the empirically calculated $S$ is found outside this interval, we consider the asymmetry statistically significant. We calculate the values of $S_c$ for various $N$ (Table I and Fig. 2).

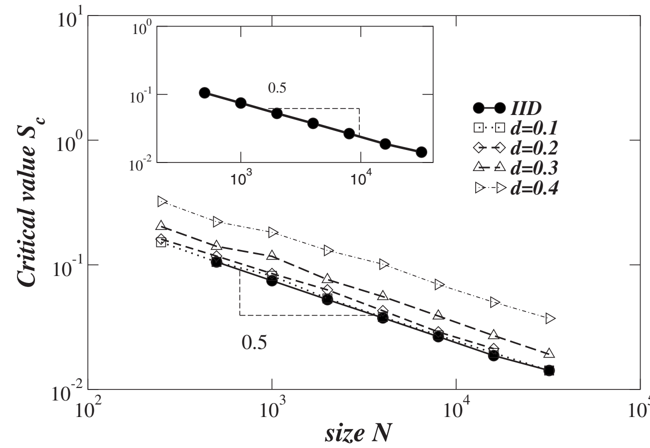


FIG. 2. To utilize the test from Eq. (1), we show that the critical value $S_c$ follows a power law with respect to sample size $N$. In order to apply the test for empirical time series, we calculate values $S_c$ for different value of $N$. We generate $10^6$ iid time series for such $N$ and for each series we calculate the test $S$. We rank all the $S$ values from smallest to largest and find the $S_c$ for which there is probability 0.95 that the $S$ of a generated series is between $(-S_c, S_c)$. We repeat the same procedure for different values of $N$. We obtain a power law (inset) $S_c \sim AN^{-\alpha}$ between $S_c$ and $N$, where $\alpha=0.5$ and $A=2.16$. We repeat the procedure for $S_c$ values for the process of Eq. (2) when $\lambda=0$ (FIARCH), a symmetric process in magnitude correlations. For four different values of $\rho$ we obtain power law relations between $S_c$ and $N$.

Note that our test is model independent—it measures asymmetry in magnitude correlations but assumes neither the memory in correlations (long or short) nor the functional form of the correlation (e.g., power law or exponential).

A concern is the possibility that in order to test significance of asymmetry in power-law magnitude correlations, we should find the intervals $(-S_c, S_c)$ not from iid but from time series generated by symmetric magnitude correlations. To address this concern, we create a stochastic process characterized by asymmetric power-law correlations in the magnitudes $|x_i|$

TABLE I. Critical values $S_c$ obtained for an iid process and for the process of Eq. (2) with $\lambda=0$. Due to finite-size effects, even these two processes may have nonzero values for $S$. In order to be considered significant asymmetry, we demand that the empirically calculated $S$ (see Table II) is outside the interval $(-S_c, S_c)$.

| $N$ | iid | $\rho=0.1$ | $\rho=0.2$ | $\rho=0.3$ | $\rho=0.4$ |
|---|---|---|---|---|---|
| 500 | 0.105 | 0.106 | 0.118 | 0.141 | 0.222 |
| 2000 | 0.053 | 0.054 | 0.063 | 0.076 | 0.131 |
| 8000 | 0.027 | 0.027 | 0.029 | 0.039 | 0.070 |
| 32000 | 0.014 | 0.014 | 0.016 | 0.019 | 0.037 |

$$x_i = \sigma_i \eta_i, \sigma_i = \sum_{n=1}^{\infty} a_n(\rho) \frac{||x_{i-n}| + \lambda x_{i-n}|}{\langle ||x_{i-n}| + \lambda x_{i-n}| \rangle}, \qquad (2)$$

where $\rho \in (0,0.5)$ and $\lambda \in (-1,1)$ are free parameters, $\sigma_i$ is a time-dependent standard deviation, $a_n(\rho)$ are power-law-distributed weights $a_n(\rho) = -\Gamma(n-\rho)/[\Gamma(-\rho)\Gamma(1+n)]$ chosen to generate power-law correlations in the magnitudes $|x_i|$. $\Gamma(x)$ denotes the Gamma function and $\eta_i$ denotes iid Gaussian random variables with mean $\langle \eta_i \rangle = 0$ and variance $\langle \eta_i^2 \rangle = 1$. The parameter $\rho$ controls the length of the power-law memory, whereas the parameter $\lambda$ controls the asymmetry in magnitude correlations. When $\lambda = 0$, the process of Eq. (2) reduces to a fractionally integrated autoregressive conditional heteroskedastic (FIARCH) process with symmetric magnitude correlations [7] for which $\alpha = 0.5 + \rho$ [8], where $\alpha$ is the exponent found from detrended fluctuation analysis (DFA) [9]. We therefore call the process of Eq. (2) asymmetric FIARCH (AFIARCH) process.

Because we include all previous increments in $\sigma_i$ of Eq. (2), our process is necessarily *long-range* correlated. We can also create a *short-range* correlated process $x_i = \sigma_i \eta_i$ by including only the most recent increment so $\sigma_i = (||x_{i-1}| + \lambda x_{i-1}|)$. In this paper, instead of $\infty$, in the sum of Eq. (2) we use the cut-off length $\ell = 500$.

By using the process of Eq. (2), we generate a number of time series and find that the magnitude correlations quantified by the DFA exponent practically do not depend on the parameter $\lambda$. To demonstrate this, Fig. 1(b) shows DFA plots for two fixed values of $\rho$ and varying values of $\lambda$. We see that the DFA plots practically overlap and that $\alpha = 0.5 + \rho$ holds as for the symmetric FIARCH process [Eq. (2) with $\lambda = 0$ [7]]. Thus, the asymmetric term in Eq. (2) ($\lambda \neq 0$) practically does not affect the correlation pattern of the magnitude time series.

We next return to our goal of determining the statistical significance of asymmetry. We use the process of Eq. (2) with $\lambda = 0$ to generate a large number of time series for various values of $\rho$ and $N$. We then determine the test variable $S$ of Eq. (1) for each of these series. Ranking the values $S$ from smallest to largest we find a critical value $S_c$ for which there is probability of 0.95 that the $S$ from a finite symmetrically defined series falls between $(-S_c, S_c)$. Varying both $\rho$ and $N$, in Fig. 2 we obtain four power-law fits relating the critical value $S_c$ and the number of data points $N$. As expected, the critical values for power-law correlated time series shown in Table I with $\rho = 0.1$ ("weak" power-law correlations) are practically the same as the critical values obtained for iid time series. However, the stronger the correlation, the larger the critical value $S_c$.

In order to estimate the parameter $\lambda$ characterizing the asymmetry of a time series, we employ the maximum likelihood estimation method [10]. One starts by deriving a likelihood function that is an expression for the probability of obtaining a given sample of $N$ known observations $(X_1, X_2, \ldots, X_N)$. We denote the probability of obtaining the $i$th observation $X_i$ as $P(X_i)$. Then the probability $L$ of obtaining our particular $N$ observations is the product of the probability $P(X_i)$ to obtain each

TABLE II. Subjects shown in column 1 are designated as in Ref. [11]. We show only a few subjects. Statistics for all subjects available on request. Column 2 shows the number $N$ of data points, while column 3 displays the results obtained for the test of Eq. (1). In columns 4 and 5 are the estimates for $\rho$ and $\lambda$ of the AFIARCH process of Eq. (2) obtained after likelihood minimization of Eq. (4) with Gaussian pdf. In columns 6 and 7 are the AFIARCH estimates for $\rho$ and $\lambda$ with the Laplace pdf. In the last column we show the AGARCH $\lambda_1$ estimate.

| Subject | $N$ | $S$ | $\rho$ | $\lambda$ | $\rho$ | $\lambda$ | AGARCH $\lambda_1$ |
|---|---|---|---|---|---|---|---|
| 02 | 28000 | −0.024 | 0.27 | 0.20 | 0.30 | 0.09 | $0.170 \pm 0.01$ |
| 06 | 30000 | 0.015 | 0.24 | −0.02 | 0.22 | −0.01 | $-0.085 \pm 0.010$ |
| 08 | 24000 | 0.070 | 0.28 | 0.02 | 0.27 | 0.03 | $-0.002 \pm 0.012$ |
| 10 | 24000 | 0.028 | 0.24 | 0.10 | 0.26 | 0.13 | $0.023 \pm 0.022$ |
| 24 | 25000 | 0.003 | 0.20 | −0.08 | 0.19 | −0.08 | $-0.062 \pm 0.012$ |
| 28 | 27000 | 0.028 | 0.24 | −0.12 | 0.25 | −0.06 | $-0.115 \pm 0.010$ |

$$L = \prod_{i=1}^{N} P(X_i). \qquad (3)$$

To make further progress, we need to posit a form for $P(X_i)$. We assume the increments $X_i$ are normally distributed $P(X_i) = (2\pi\sigma_i^2)^{-1/2} \exp(X_i^2/2\sigma_i^2)$ with a mean 0 and characterized by a time-dependent variance $\sigma_i^2$ which depends on the past values of $X_i$. In our case, all values of $\sigma_i$ and all values of $P(X_i)$ are characterized by only two adjustable parameters $(\rho, \lambda)$. Substituting the previous $P(X_i)$ into Eq. (3) and taking the logarithm we obtain the logarithmic-likelihood function for the sample [10]

$$\ln L = -\frac{1}{2} N \ln(2\pi) - \sum_{i=1}^{N} \left[ \ln(\sigma_i) + \frac{1}{2} X_i^2/\sigma_i^2 \right]. \qquad (4)$$

To illustrate the utility of the process of Eq. (2) for modeling real-world data, we next analyze a large electroencephalography (EEG) database [11] comprising records from 25 subjects randomly selected over a 6-month period at St. Vincent's University Hospital in Dublin [12]. EEG data are recorded every 0.8 s, so we obtain the number of data points $N$ between 22 000 and 30 000 (Table II). Time series of EEG magnitudes exhibit power-law long-range correlated behavior [13,14].

From each original time series we subtract the average. From Tables I and II we see that our test of Eq. (1) with probability 0.95 confirms the existence of asymmetry in magnitude correlations. The test for each subject is outside the range $(-S_c, S_c)$ for a given $N$. For example, for subject 02, characterized by $N = 28\,000$ (close to 32 000 in Table I) and $\rho = 0.27$ (close to 0.3 in Table I), we find $S = -0.024$ outside the range we obtained for iid process $(-0.014, 0.014)$, process with symmetric power-law magnitude correlations $(-0.019, 0.019)$, and the approach of Ref. [5] $(-0.019, 0.019)$.

By minimizing Eq. (4), we estimate $\rho$ and $\lambda$ for each subject (Table II) where we choose a normal probability distribution function (pdf) $P(x)$ for EEG data. Commonly one

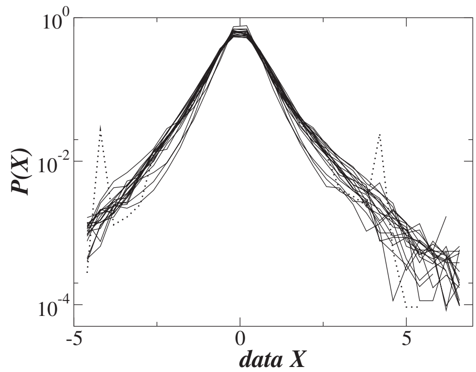


FIG. 3. The pdf in the logarithmic-linear plot for each of 20 EEG time series out of total 25 time series [11] comprising records from 20 subjects. Each pdf approximately follows the Laplace distribution. We also show the pdf of subject 10 whose tails due to bumps deviate from the Laplace pdf.

uses a normal pdf when using logarithmic-likelihood approach. In order to check if some other choice for $P(x)$ would be more appropriate, next we analyze pdf $P(x)$ of empirical data [11]. In Fig. 3 we see that for most of the empirical time series, $P(x)$ in the broad central region follows not normal, but Laplace distribution $P(X_i) = 1/(\sqrt{2}\sigma)\exp(-\sqrt{2}|x-\bar{x}|/\sigma)$, where $\sigma$ is the standard deviation. For 5 subjects, $P(x)$ exhibit some bumps in the tail parts.

Next we follow the procedure of Eq. (3) but this time with Laplace $P(x)$. We find that the parameters $\rho$ and $\lambda$ change quantitatively but not qualitatively (see Table II)—the sign of $\lambda$ does not change by replacing normal $P(x)$ by Laplace $P(x)$.

To further test the statistical significance of asymmetry in magnitude correlations found in the data, we employ another process known as asymmetric generalized autoregresive conditional heteroskedastic (AGARCH) process [6]. This process is characterized by an exponentially decaying autocorrelation function

$$\sigma_i^2 = \omega + \alpha(|x_{i-1}| + \lambda_1 x_{i-1})^2 + \beta\sigma_{i-1}^2, \tag{5}$$

where $\alpha$, $\beta$, $\omega$, and $\lambda_1$ are the free AGARCH parameters and $\lambda_1$ is an asymmetry parameter similar to the one in Eq. (2). The last column in Table II shows our estimate for $\lambda_1$ (and two standard errors).

Note that the estimates $\lambda$ and $\lambda_1$ for AFIARCH and AGARCH, respectively, calculated for different subjects are very closely related. We obtain $\lambda = -0.029 + 1.232\lambda_1$, where 0.120 is the standard error of the slope coefficient. Differences are expected since the AFIARCH is characterized by power-law magnitude correlations [see Fig. 1(b)], while the AGARCH process is characterized by exponential magnitude correlations. From the results obtained for the AGARCH process, the asymmetry parameter $\lambda_1$ is statistically insignificant (within two standard errors) only for subjects 7, 8, 10, 12, and 21. Discrepancy between the results obtained from the test of Eq. (1) and the stochastic process of Eq. (5) is likely explained by the fact that the test of Eq. (1) measures only asymmetry in magnitude correlations and does not assume either (i) the functional form of the magnitude correlations or (ii) their long-range nature, whereas our stochastic process imposes both.

From the analysis of individual time series, we conclude that magnitude correlations in observed physiological data exhibit significant asymmetry. However, universality is not confirmed. From the values of $\lambda$ in Table II obtained for different subjects, we calculate the average $\lambda - \bar{\lambda} = -0.022 \pm 0.11$. The spread of values of the asymmetry parameter $\lambda$ suggests that the asymmetry does not show universality. However, the average $\lambda$ and its standard deviation $\sigma$ should show significant differences between diseased and healthy subjects. Consequently, the present analysis and proposed test may have potential to be useful for diagnostic purposes.

We thank the NSF and the Ministry of Science of Croatia for financial support.

[1] J. B. Bassingthwaighte, L. S. Liebovitch, and B. J. West, *Fractal Physiology* (Oxford University Press, New York, 1994).
[2] R. F. Engle, Econometrica **50**, 987 (1982).
[3] Y. Ashkenazy *et al.*, Phys. Rev. Lett. **86**, 1900 (2001).
[4] A. Corral, Phys. Rev. Lett. **95**, 159801 (2005).
[5] E. Lippiello, L. de Arcangelis, and C. Godano, Phys. Rev. Lett. **100**, 038501 (2008).
[6] R. F. Engle and V. Ng, J. Financ. **48**, 1749 (1993).
[7] C. W. J. Granger and Z. Ding, J. Econometr. **73**, 61 (1996); P. Cizeau *et al.*, Physica A **245**, 441 (1997); Y. Liu *et al.*, Phys. Rev. E **60**, 1390 (1999).
[8] B. Podobnik *et al.*, Phys. Rev. E **72**, 026121 (2005).
[9] C.-K. Peng *et al.*, Phys. Rev. E **49**, 1685 (1994).
[10] J. D. Hamilton, *Time Series Analysis* (Princeton, New Jersey, 1994).
[11] www.physionet.org.pn3.ucddb/
[12] The subjects are selected for possible diagnosis of obstructive sleep apnea or primary snoring. The subject's details are available at *tt* www.physionet.org.pn3.ucddb/SubjectsDetails.xls. We use time series designated as uccdbb0*i*.eeg2, where *i* stands for an integer.
[13] C. P. Pan *et al.*, Phys. Lett. A **329**, 130 (2004).
[14] B. Podobnik and H. E. Stanley, Phys. Rev. Lett. **100**, 084102 (2008).