Effect of Nonlinear Correlations on the Statistics of Return Intervals in Multifractal Data Sets

Mikhail I. Bogachev, Jan F. Eichner, and Armin Bunde

Institut für Theoretische Physik III, Justus-Liebig-Universität Giessen, 35392 Giessen, Germany

(Received 8 August 2007; published 10 December 2007)

We study the statistics of return intervals between events above a certain threshold in multifractal data sets without linear correlations. We find that nonlinear correlations in the record lead to a power-law (i) decay of the autocorrelation function of the return intervals, (ii) increase in the conditional return period, and (iii) decay in the probability density function of the return intervals. We show explicitly that all the observed quantities depend both on the threshold value and system size, and hence there is no simple scaling observed. We also demonstrate that this type of behavior can be observed in real economic records and can be used to improve considerably risk estimation.

DOI: 10.1103/PhysRevLett.99.240601 PACS numbers: 05.40.−a, 02.50.−r, 05.45.Df, 89.75.Da

The understanding of the occurrence of extreme events is of great importance in various fields of science [1–4]. One of the central quantities here is the time interval between successive events above (or below) some threshold Q, sometimes referred to as reoccurrence times or return intervals. By studying the statistics of the return intervals for increasing threshold heights Q one aims to find out the laws governing the occurrence of extreme events.

It is well known that for uncorrelated data sets the probability density function (PDF) of the return intervals is a simple exponential, $\ln P_Q(r) \sim -r/R_Q$, where R_Q is the mean return interval ("return period"), and the return intervals are uncorrelated. For long-term correlated records (x_i), where the autocorrelation function $C_x(s)$ between two events separated by the time s decays by a power law, $C_x(s) \sim s^{-\gamma}$, $0 < \gamma < 1$, the statistics of the return intervals is changed considerably. The PDF decays by a stretched exponential [5–7], $\ln P_Q(r) \sim -(r/R_Q)^{\gamma}$, and the sequence of return intervals is also long-term correlated with the same exponent γ as the original record [6,7]. It was shown that this leads to a clustering of extreme events, which can be observed both in climate records [8,9] and in volatility records in the financial markets [10]. For studies of long-term persistence in condensed matter physics, we refer to [11].

In this Letter, we are interested in the statistics of the return intervals in multifractal data sets (x_i). Unlike long-term correlated ("monofractal") data sets, the correlation structure of multifractal data sets cannot be quantified by a simple exponent, but an infinite hierarchy of exponents is needed [12–14]. To study how the nonlinear correlations affect the statistics of the return intervals, we focus on data sets where the linear correlations [described by the autocorrelation function $C_x(s)$] vanish. Perhaps the most prominent examples for these kinds of data sets are precipitation [15,16] and financial records [17–20]. We find that the nonlinear correlations lead to linear long-term correlations among the return intervals, with an autocorrelation function that decays by a power law and an exponent that increases monotonically with increasing threshold Q. Also, the PDF of the return intervals shows pronounced power-law behavior with an exponent that depends on Q. At very large values of r/R_Q, the PDF shows a stronger decay which may be interpreted as a finite-size effect. Finally, we show explicitly that all these features can be observed in financial data sets.

For generating multifractal data sets, we consider a variant of the multiplicative random cascade process, described, e.g., in [13,14,21,22]. In this process, the data set is obtained in an iterative way, where the length of the record doubles in each iteration. We start with the zeroth iteration $n = 0$, where the data set (x_i) consists of one value, $x_1^{(n=0)} = 1$. In the nth iteration, the data $x_i^{(n)}$, $i = 1, 2, \ldots, 2^n$, is obtained from

$$x_{2l-1}^{(n)} = x_l^{(n-1)} m_{2l-1}^{(n)} \quad \text{and} \quad x_{2l}^{(n)} = x_l^{(n-1)} m_{2l}^{(n)}, \tag{1}$$

where the multipliers m are independent and identically distributed random numbers with zero mean and unit variance. The resulting PDF is symmetric with log-normal tails. As we show below, there are no linear correlations in these data sets, i.e., $C_x(s) = 0$ for $s > 0$.

There are several ways to characterize multifractal data sets. Here we chose the multifractal detrended fluctuation analysis (MFDFA), introduced by Kantelhardt *et al.* in [23]. In the MFDFA one considers the profile, i.e., the cumulated data series $Y_j = \sum_{i=1}^{j}(x_i - \langle x \rangle)$, and splits the record into N_s (nonoverlapping) segments of size s. In each segment k a local polynomial fit $y_k(j)$ of, e.g., second order is estimated. Then one determines the variance $F_k^2(s) = (1/s)\sum_{j=1}^{s}[Y_{[(k-1)s+j]} - y_k(j)]^2$ between the local trend and the profile in each segment k and calculates a generalized fluctuation function $F_q(s)$,

$$F_q(s) \equiv \left\{\frac{1}{N_s}\sum_{k=1}^{N_s}[F_k^2(s)]^{q/2}\right\}^{1/q}. \tag{2}$$

In general, $F_q(s)$ scales with s as $F_q(s) \sim s^{h(q)}$. The generalized Hurst exponent $h(q)$ is directly related to the scaling exponent $\tau(q)$ defined by the standard partition function-based multifractal formalism, via $\tau(q) = qh(q) - 1$. For a monofractal record, $h(q)$ is independent of q. For stationary records, $h(2)$ is related to the autocorrelation function $C_x(s)$. In the absence of linear correlations [where $C_x(s) = 0$ for $s > 0$], $h(2) = 1/2$.

In general, $h(q)$ depends on both the distribution of the data and their correlation structure [23]. To eliminate the dependence on the log-normal tailed distribution, we have first ranked the N numbers in the multifractal data, and then exchanged them rankwise by a set of N numbers from a Gaussian distribution. Now the deviations of the resulting $h(q)$ from $h(2)$ depend only on the nonlinear correlations and therefore can be used to characterize them. It can be easily seen that this procedure conserves, for a fixed return period R_Q, the arrangement of the return intervals as well as their statistics, since the temporal arrangement of the values of the original data remains unchanged. Figure 1(a) displays $F_q(s)/s^{1/2}$ for $q = 1$, 2, and 5. The double-logarithmic plot shows that $F_q(s)$ follows the anticipated power-law scaling, with different exponents for the different values of q. For $q = 2$, $F_q(s)/s^{1/2}$ has reached a plateau, i.e., $h(2) = 1/2$, indicating the absence of linear correlations in the data. To show this feature explicitly, we also calculated directly the autocorrelation function $C_x(s)$. The inset in Fig. 1(a) shows that (as expected) $C_x(s)$ fluctuates around zero for all $s \geq 1$.

In the following, we focus on the return intervals between subsequent events above some threshold Q, which we obtain directly from the original data. Figure 1(b) shows the PDF of the return intervals for $R_Q = 10$, 70, and 500. For all return periods, we find a pronounced power-law behavior

$$P_Q(r) \sim (r/R_Q)^{-\delta(Q)}, \tag{3}$$

in marked contrast to the uncorrelated or long-term correlated monofractal data sets. The exponent δ [shown in Fig. 1(c)] depends explicitly on R_Q and seems to converge to a limiting curve for large data sets. When shuffling the multifractal data set, the nonlinear correlations are destroyed and the scaled PDFs collapse (as expected) to a single exponential [also shown in Fig. 1(b)].

Next, we study the way the return intervals are arranged in time. Figure 1(d) shows the autocorrelation function $C_Q(s)$ [24] of the return intervals for $R_Q = 10$, 70, and 500. While $C_x(s) \equiv 0$ for $s \geq 1$, $C_Q(s)$ decays by a power law,

$$C_Q(s) \sim s^{-\beta(Q)}, \tag{4}$$

indicating long-term correlations among the return intervals. Figure 1(d) shows that the exponent β increases monotonically with R_Q, $\beta = 0.47$, 0.56, and 0.7 for $R_Q = 10$, 70, and 500, respectively. Obviously, these long-term correlations have been induced by the nonlinear correlations in the multifractal data set. Extracting the return interval sequence from a data set is a nonlinear operation, and thus the return intervals are influenced by the nonlinear correlations in the original data set. Accordingly, the return intervals in data sets without linear correlations are sensitive indicators for nonlinear correlations in the data sets.

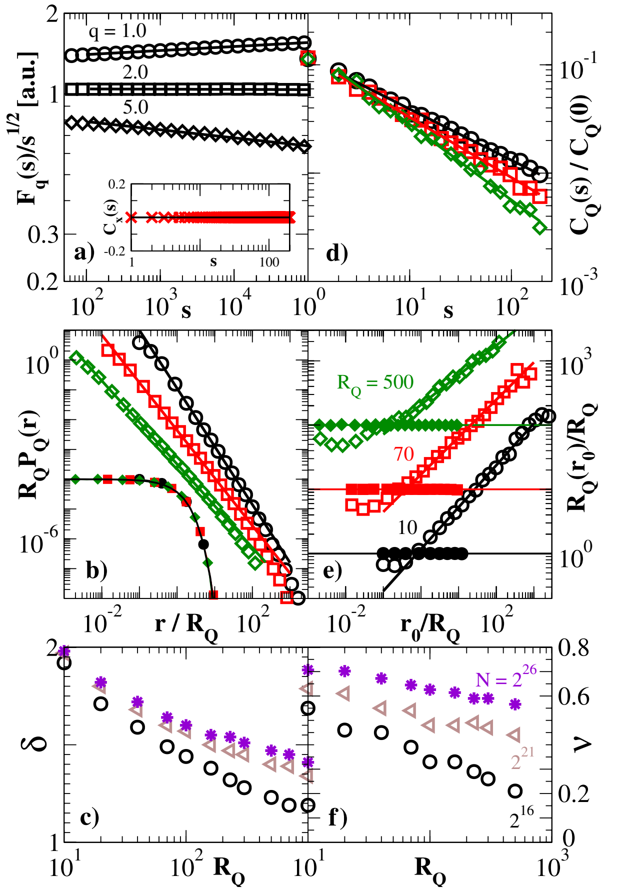

FIG. 1 (color online). Analysis of the multifractal cascade model: (a) MFDFA fluctuation function for $q = 1$ (○), 2 (□), and 5 (◇). The autocorrelation function $C_x(s)$ of the data is shown in the inset. (b) Scaled PDFs $P_Q(r)$ of the return intervals for return periods $R_Q = 10$ (○), 70 (□), and 500 (◇). To avoid overlapping, symbols were shifted downwards by a factor of 10 (□) and 100 (◇). The relevant filled symbols show the corresponding PDFs for the shuffled data, shifted downwards by a factor of 10^4. (c) Exponents $\delta(Q)$ vs R_Q, for different systems sizes $N = 2^{16}$ (○), 2^{21} (◁), and 2^{26} (*). (d) Return intervals autocorrelation function $C_Q(s)$ for the same R_Q values as in (b). (e) Conditional return periods $R_Q(r_0)$ in units of R_Q vs r_0/R_Q for the same R_Q values as in (b) (○), The filled symbols are for the shuffled data. The curves for $R_Q = 70$ and 500 were raised by a factor of 10 and 100, respectively, to avoid overlapping symbols. (f) Exponents $\nu(Q)$, for the same systems sizes as in (c). The results in (a), (b), (d), and (e) are based on data sets of length $N = 2^{21}$ and averaged over 150 configurations.

To further quantify the memory among the return intervals, we consider the conditional return intervals; i.e., we regard only those intervals whose preceding interval is of a fixed size r_0. In Fig. 1(e) the conditional return period $R_Q(r_0)$, which is the average of all conditional return intervals for a fixed threshold Q, is plotted versus r_0/R_Q (in units of R_Q). The figure demonstrates that, as a consequence of the memory, large return intervals are rather followed by large ones, and small intervals by small ones. In particular, for r_0 values exceeding the return period R_Q, $R_Q(r_0)$ increases by a power law,

$$R_Q(r_0) \sim r_0^{\nu(Q)} \quad \text{for} \quad r_0 > R_Q, \tag{5}$$

where the exponent ν approximately decreases logarithmically with increasing value of R_Q [Fig. 1(f)]. Note that only for an infinite record the value of $R_Q(r_0)$ can increase infinitely with r_0. For real (finite) records, there exists a maximum return interval which limits the values of r_0, and therefore $R_Q(r_0)$. As well as for $P_Q(r)$ and $C_Q(s)$, there is no scaling, and accordingly, the occurrence of extreme events cannot be deducted straightforwardly from the occurrence of smaller events. When shuffling the original data, the memory vanishes and $R_Q(r_0) \equiv R_Q$, indicated by the filled symbols.

A central quantity in risk estimation is the probability $W_Q(x; \Delta x)$ that, after an elapsed time x from the last event above Q, the next event above Q will occur within a (short) time interval $\Delta x \ll R_Q$. By definition, W_Q is given by

$$W_Q(x; \Delta x) = \frac{\int_x^{x+\Delta x} P_Q(r)dr}{\int_x^{\infty} P_Q(r)dr}. \tag{6}$$

For uncorrelated records, $W_Q = 1 - \exp(\Delta x/R_Q) \simeq \Delta x/R_Q$, independent of x. Since due to the nonlinear correlations in the multifractal data set $P_Q(r)$ decays approximately by a power law, it can be shown straightforwardly that $W_Q(x; \Delta x)$ decays as

$$W_Q(x; \Delta x) \simeq [\delta(Q) - 1]\frac{\Delta x}{x} = [\delta(Q) - 1]\frac{\Delta x/R_Q}{x/R_Q}, \tag{7}$$

which we also verified numerically. As expected, due to the multifractality, W_Q does not scale with R_Q, but the explicit Q dependence occurs solely in the prefactor $[\delta(Q) - 1]$ and can be estimated. Accordingly, in the absence of linear correlations in the original data set, the nonlinear correlation allows for a considerably improved risk estimation. The estimation can be further improved by considering the conditional distribution function $P_Q(r|r_0)$ instead of $P_Q(r)$.

We like to note that we obtained similar results also for the multifractal random walk model [25] which constitutes another important class of multifractal models. A detailed comparison of the results will be presented elsewhere.

To show that the effects found here can be observed in real world data sets, we have considered financial records. It is well known that the arithmetic returns $(P_i - P_{i-1})/P_{i-1}$ of daily stock closing prices P_i form multifractal data sets, with vanishing autocorrelation function [17,26,27]. We have analyzed several stocks (IBM, GM, GE, Boeing, etc.), exchange rates versus U.S. dollar (e.g., DM, AUS dollar, CAN dollar), oil crude prices (Brent and WTI), and integral market indices (e.g., Dow Jones and S&P 500), with qualitatively identical results. As a representative example we focus on the IBM record from January, 1962, until February, 2007 [28]. First, to compare the MFDFA results with the simulated data, we exchange the returns by Gaussian data by conserving the rank ordering, as we did for the model data. The result is displayed in Fig. 2(a). The dashed lines correspond to the simulated data, with a record length comparable to the length of the IBM data set. The inset shows the (vanishing) autocorrelation function for the IBM returns.

The figure shows that the nonlinear correlations measured by $F_q(s)$ are modeled quite well by (1). Thus we expect that the return intervals will show a similar behavior, too. This is shown in Figs. 2(b)–2(d), where for several values of R_Q the PDF $P_Q(r)$, the autocorrelation function

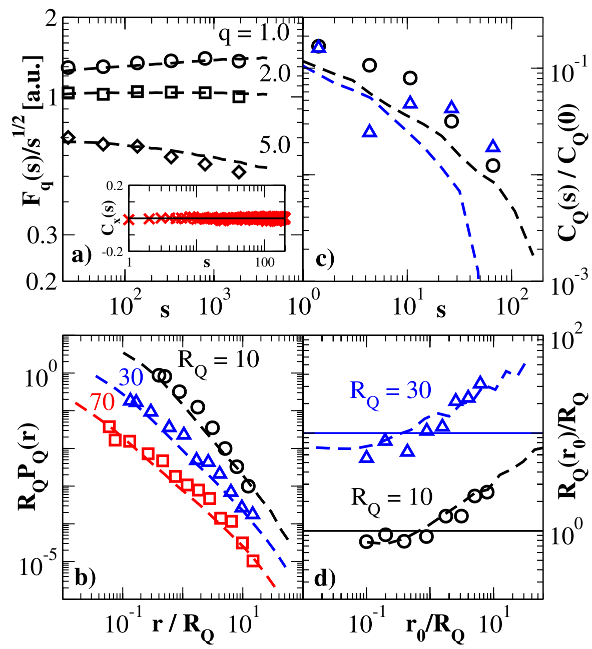

FIG. 2 (color online). Analysis of the arithmetic returns of IBM daily stock closing prices (open symbols), and simulated multifractal data of similar length $N = 2^{14}$ (dashed lines), averaged over 500 configurations: (a) MFDFA fluctuation function for three moments $q = 1$ (○), 2 (□), and 5 (◇). The autocorrelation function $C_x(s)$ of the IBM returns is shown in the inset. (b) Scaled PDFs $P_Q(r)$ of the return intervals versus r/R_Q for $R_Q = 10$ (○), 30 (△), and 70 (□). To avoid overlapping, symbols were shifted downwards by a factor of 10 (△) and 100 (□). (c) Autocorrelation function $C_Q(s)$ of return intervals (r_j) for $R_Q = 10$ and 30 [symbols correspond to those in (b)]. (d) Conditional return periods $R_Q(r_0)$ in units of R_Q vs r_0/R_Q for the same R_Q values as in (c). The curve for $R_Q = 30$ was raised by a factor of 10 to avoid overlapping symbols.

$C_Q(r)$, and the conditional return periods $R_Q(r_0)$ of the return intervals are shown and compared with the model data of similar length (shown as dashed lines) averaged over 500 configurations each. The agreement between observed and model data is striking. As in Fig. 1(b) the PDFs decay approximately by a power law, with different exponents for different values of R_Q. Because of the comparably short data set of the IBM data set, finite-size effects are considerably stronger than for the model data, leading to deviations from the power law at smaller return intervals than in Fig. 1(b). Also the autocorrelation function $C_Q(s)$ behaves qualitatively the same as the model data [see Fig. 1(c)] but due to less statistics finite-size effects are more pronounced. The conditional return periods shown in Fig. 2(d) agree very well with the model data, but due to less statistics very large values of r_0/R_Q cannot be tested. We like to note that $P_Q(r)$ and $R_Q(r_0)$ for several financial data sets for negative thresholds Q have been studied by Yamasaki *et al.* [29], but different conclusions regarding scaling and functional forms have been drawn.

In summary, we have studied the return intervals in multifractal data sets without linear correlations. We found that due to the inherent nonlinear correlations in the data, the relevant quantities characterizing the return intervals (autocorrelation function, PDF, and conditional return period) exhibit (nonuniversal) power-law behavior with exponents depending explicitly on the height of the threshold. We have demonstrated that these features, which can be observed in economic records, allow for an improved risk estimation.

Finally, we like to emphasize that similar consequences of nonlinear correlations should be looked after also in the other data sets that are known to be multifractal, e.g., turbulence [30], rainfall [15,16], river flows [31], teletraffic in large networks [32], physiology [33], and also in persistent data sets in condensed matter physics [11].

We would like to acknowledge DAPHNET for financial support.

[1] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics* (John Wiley & Sons, New York, 1978).
[2] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes* (Springer, New York, 1983).
[3] H. v. Storch and F. W. Zwiers, *Statistical Analysis in Climate Research* (Cambridge University Press, Cambridge, England, 2001).
[4] A. Bunde, J. Kropp, and H. J. Schellnhuber, *The Science of Disasters* (Springer, Berlin, 2002).
[5] When considering zero-level crossings in long-term correlated data with Gaussian noise, it has been proven by G. F. Newell and M. Rosenblatt, Ann. Math. Stat. **33**, 1306 (1962) that the probability of having no zero-level crossing after t time steps is bounded from above by a stretched exponential.
[6] A. Bunde, J. F. Eichner, S. Havlin, and J. W. Kantelhardt, Physica (Amsterdam) **A342**, 308 (2004).
[7] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, Phys. Rev. E **75**, 011128 (2007).
[8] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, Phys. Rev. Lett. **94**, 048701 (2005).
[9] M. S. Santhanam and H. Kantz, Physica (Amsterdam) **A345**, 713 (2005).
[10] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **102**, 9424 (2005).
[11] S. N. Majumdar, Curr. Sci. **77**, 370 (1999).
[12] B. B. Mandelbrot, *Fractal Geometry of Nature* (W. H. Freeman, New York, 1982).
[13] J. Feder, *Fractals* (Plenum, New York, 1988).
[14] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals: New Frontiers of Science* (Springer, New York, 1992).
[15] S. Lovejoy and D. Schertzer, *Multifractals and Rain, in New Uncertainty Concepts in Hydrology and Water Resources*, edited by Z. W. Kunzewicz (Cambridge University Press, New York, 1995).
[16] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin, J. Geophys. Res. **111**, D01106 (2006).
[17] Z. Ding, C. W. J. Granger, and R. F. Engle, J. Empir. Finance **1**, 83 (1993).
[18] B. B. Mandelbrot, *Gaussian Self-Affinity and Fractals* (Springer, New York, 2001).
[19] T. Lux and M. Ausloos, in The Science of Disasters (Ref. [4]).
[20] T. Lux, Int. J. Mod. Phys. C **15**, 481 (2004).
[21] C. Meneveau and K. R. Sreenivasan, Phys. Rev. Lett. **59**, 1424 (1987).
[22] M. Greiner, H. C. Eggers, and P. Lipa, Phys. Rev. Lett. **80**, 5333 (1998).
[23] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, Physica (Amsterdam) **A316**, 87 (2002).
[24] For the record r_i, $i = 1, 2, \ldots, N_Q$, of the return intervals between events above the threshold Q, $C_Q(s) \equiv \langle (r_i - R_Q)(r_{i+s} - R_Q)\rangle / \langle (r_i - R_Q)^2 \rangle$, $s = 1, 2, 3, \ldots$ [8].
[25] E. Bacry, J. Delour, and J. F. Muzy, Phys. Rev. E **64**, 026103 (2001).
[26] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, Nature (London) **381**, 767 (1996).
[27] K. Matia, Y. Ashkenazy, and H. E. Stanley, Europhys. Lett. **61**, 422 (2003).
[28] http://finance.yahoo.com/q/hp?s=IBM.
[29] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H. E. Stanley, in *Practical Fruits of Econophysics* (Springer, Tokyo, 2006), p. 43.
[30] J. F. Muzy, E. Bacry, and A. Arneodo, Phys. Rev. Lett. **67**, 3515 (1991).
[31] G. Pandey, S. Lovejoy, and D. Schertzer, J. Hydrol. **208**, 62 (1998).
[32] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk, IEEE Trans. Inf. Theory **45**, 992 (1999).
[33] P. Ch. Ivanov, A. L. Goldberger, and H. E. Stanley, in The Science of Disasters (Ref. [4]).