

# Record statistics of financial time series and geometric random walks

Behloul Sabir\* and M. S. Santhanam

*Indian Institute of Science Education and Research, Dr. Homi Bhabha Road, Pune 411 008, India*

(Received 1 July 2014; published 19 September 2014)

The study of record statistics of correlated series in physics, such as random walks, is gaining momentum, and several analytical results have been obtained in the past few years. In this work, we study the record statistics of correlated empirical data for which random walk models have relevance. We obtain results for the records statistics of select stock market data and the geometric random walk, primarily through simulations. We show that the distribution of the age of records is a power law with the exponent  $\alpha$  lying in the range  $1.5 \leq \alpha \leq 1.8$ . Further, the longest record ages follow the Fréchet distribution of extreme value theory. The records statistics of geometric random walk series is in good agreement with that obtained from empirical stock data.

DOI: [10.1103/PhysRevE.90.032126](https://doi.org/10.1103/PhysRevE.90.032126)

PACS number(s): 05.40.-a, 89.65.Gh, 02.50.Ey

## I. INTRODUCTION

In popular parlance, records are associated with record-breaking events. Common examples include extreme weather events such as the occurrences of lowest or highest temperature [1], unparalleled sport performances in Olympic and other events [2], and financial downturns like the major stock market crashes. In recent years, there is an increasing interest in the study of record statistics in the context of global warming and climate change [3], occurrences of cyclones and floods [4], and stock markets. In physics, records statistics is useful in understanding the behavior of stochastic motion of a domain wall in metallic ferromagnetic materials [5], as a model for the growth of networks based on record events [6], in understanding magnetization of superconductors and spin-glasses [7] and as an alternative indicator of quantum chaos in kicked rotor model [8]. Even as the record-breaking events continue to enjoy media attention, there is also an increased research interest in the theoretical study of record events [9–14].

For a discretely sampled stochastic time series  $x_t, t = 1, 2, 3, \dots, N$ , the record events are those that are larger (smaller) than all the preceding events. An event at  $t = T$  would be an upper record if  $x_T > \max(x_1, x_2, \dots, x_{T-1})$ . Then, some of the relevant questions of interest are the probability for the occurrence of record at any given time, mean number of records in a certain time window, and record age, i.e., how long a record is expected to survive. Results for most of these questions is known for uncorrelated random variables [15]. However, it is known that most of the physically observed time series, e.g., temperature, stock market volatility, earthquake magnitudes, are strongly correlated [16]. The record statistics for such correlated systems is beginning to receive research attention.

Recently, the record statistics of correlated series such as the positions of random walker was studied [9,12,13]. Random walk is a fundamental model in physics and has applications in many areas, including the dynamics of stock markets. If the increments of the random walker are drawn from a continuous and symmetric function  $\phi(\xi)$ , it was shown that the mean number of records, for large  $N$ , is proportional to  $\sqrt{N}$  and the mean record age  $\langle r \rangle \propto N$  [9]. These results have further been

generalized to the case of random walk with a constant drift with application to stock market data [17] and also to multiple random walkers [13].

In spite of such growing interest in correlated series, very few works have focused on the record statistics in empirical stock data [13,14,18]. In this paper, we report on the record statistics of empirical stock data to understand two quantities of interest not studied earlier, namely: (i) the distribution of record age and (ii) the distribution of longest record ages. We present our analysis of stock data in the context of geometric random walk (GRW) model, which is considered as one of the suitable models for the dynamics of stock data [19]. In addition, it must be pointed out that GRW has other applications as well, including as a model for interacting neurons [20].

In this paper, we analyze the upper record statistics for 18 stocks, for which longest data is available in the public domain. The data used in this work is described in the Appendix. Most aspects of record statistics, especially quantities such as the mean number of records, record age distribution, longest record age, etc., depend only on the position of record-breaking event on time axis and *not* on its magnitude. We study these quantities using geometric random walk as the benchmark model. We show that both for the records in stock data and geometric random walk series the distribution of record age  $r$  is consistent with  $P(r) \sim r^{-\alpha}$ , with the exponent  $1.5 \leq \alpha \leq 1.8$ , and the longest record age  $r_{\max}$  falls in the class of type-II generalized extreme value (Fréchet) distribution.

## II. DISTRIBUTION OF RECORD AGES

GRW has not attracted as much attention as the random walk model except in the context of financial applications [19]. GRW model is given by

$$y_{i+1} = y_i \exp(\xi_i), \quad i = 1, 2, 3 \dots N. \quad (1)$$

In this,  $\xi_i$  is Gaussian distributed  $G(\mu, \sigma)$  with mean  $\mu$  and standard deviation  $\sigma$ . This implies that the “log returns”  $R_i = \log(y_{i+1}/y_i)$  are also Gaussian. The log-returns from the empirical stock data is known to be approximately Gaussian distributed over a wide range of timescales [19].

Record age is the time duration  $r$  between successive occurrences of record events, i.e., the time for which a record survives. Record age distribution will provide insights into how long a record can be expected to survive and is useful in

---

\*Present address: InvenZone, Sakinaka, Mumbai 400072, India.

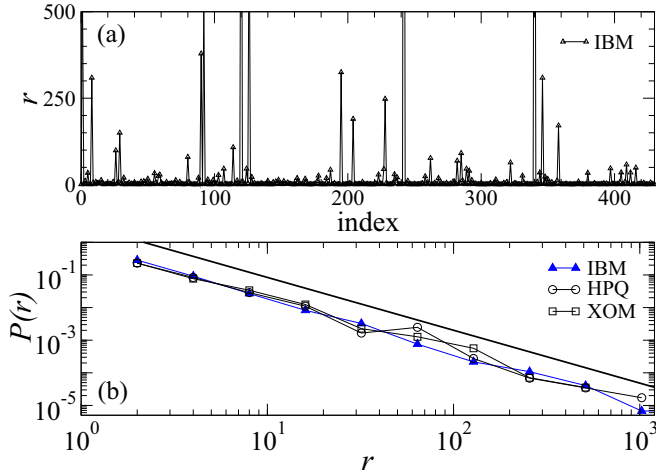


FIG. 1. (Color online) (a) Record ages (in days) calculated from IBM stock data. (b) The (logarithmically binned) distribution of record ages for three stocks. The solid line in (b) has slope  $-1.623 \pm 0.081$ .

hazard estimation problems. Though the mean record age  $\langle r \rangle$  has been analytically determined for random walk problems in the earlier works [9,17], no results have been obtained for the distribution of record age.

In Fig. 1(a), we show the record ages obtained from IBM stock data. In this, record ages longer than 500 are not shown since they mask the details near  $r = 1$ . The longest record age [not visible in Fig. 1(a)] is 2313 days and the shortest is 1 day. Thus, in this case, the record ages vary over 3 orders of magnitude. Clearly, they depend on the length  $N$  of data being considered since the longest observed record age cannot exceed the length of data. Fig. 1(b) displays the distribution of record age computed from the stock prices of three stocks (HPQ, XOM, and IBM) with the longest available time series. In the plot shown in this figure, the distribution, for the most part, is consistent with a power law of the form

$$P(r) \sim Ar^{-\alpha}, \quad (2)$$

with the exponent  $\alpha \approx 1.623$  and  $A$  being the normalization constant that can be written in terms of generalized  $\zeta$  function  $\zeta(\alpha, r_{\min})$ , where the power law regime begins at  $r = r_{\min}$ . However, the tail of computed distribution flattens out due to effects arising from the finite size of the data. In order to improve the statistics, we use GRW simulations [Eq. (1)] with  $\xi_i$  drawn from normal distribution with parameters values  $\mu = \langle \mu_{\text{emp}} \rangle = 0.00031$  and  $\sigma = \langle \sigma_{\text{emp}} \rangle = 0.015$ . These parameter values  $\langle \mu_{\text{emp}} \rangle$  and  $\langle \sigma_{\text{emp}} \rangle$  were computed from the empirical stock data by averaging over the individual values of  $\mu$  and  $\sigma$  obtained for each stock. The record age distribution for each value of  $N$ , shown in Fig. 2(a), is averaged over  $10^5$  GRW realizations. Clearly, the distribution in Fig. 2(a) can be described by a power law as in Eq. (2), with the exponent  $\alpha = 1.652 \pm 0.006$ . Significantly, it is independent of the value of  $N$ . In contrast to quantities like the mean number of records that depend on  $N$  [9,14,17], the distribution of record ages is characteristic statistical property of record breaking events independent of the length of data. Further, as  $N$  increases, the range over which the power law is valid

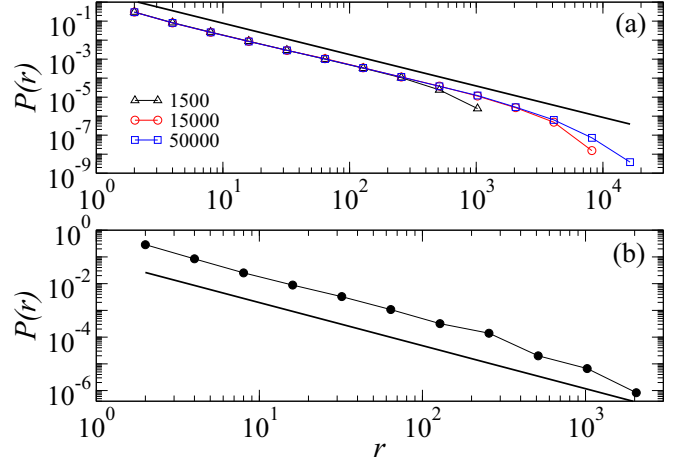


FIG. 2. (Color online) The (logarithmically binned) distribution of record ages obtained from (a) GRW simulations for three values of  $N$  and (b) stock data other than those shown in Fig. 1(b). The solid line in (a) has slope  $-1.652 \pm 0.006$  and in (b) has slope  $-1.611 \pm 0.051$ . See text for details.

also increases, implying that the tail behavior is a finite size effect. Within the parametric regime relevant for the stocks listed in the Appendix, namely,  $0.0001 \leq \mu \leq 0.0005$  and  $0.01 \leq \sigma \leq 0.05$ , we did not find any systematic relation between these parameters and the exponent  $\alpha$ .

Based on the results displayed in Fig. 1(b), we might expect that all the individual stocks will display nearly the same value of  $\alpha$  even if  $N$  is different for each one of them. As shown in Fig. 3, the value of the exponent lies in the range  $1.5 \leq \alpha \leq 1.8$  for the stocks listed in the Appendix. These values are maximum likelihood estimates obtained using the method in Ref. [21]. Hence, we combined the record ages computed from the rest of stock data in the Appendix (other than HPQ, XOM, and IBM) and the resulting distribution is displayed in Fig. 2(b). The power law form [Eq. (2)] is seen in the figure and the maximum likelihood estimate of the exponent is  $\alpha \approx 1.611 \pm 0.051$ .

### III. LONGEST RECORD AGE

Given that the record age is distributed as a power law, it is of interest to understand the distribution of longest record

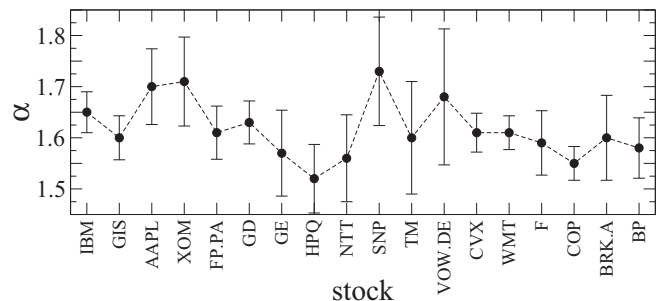


FIG. 3. Maximum likelihood estimate for the power-law exponent  $\alpha$  for the stocks studied in this work. The error bars represent the uncertainty in the estimated value of  $\alpha$  [21]. Full name of the stocks in  $x$  axis is given in the Appendix.

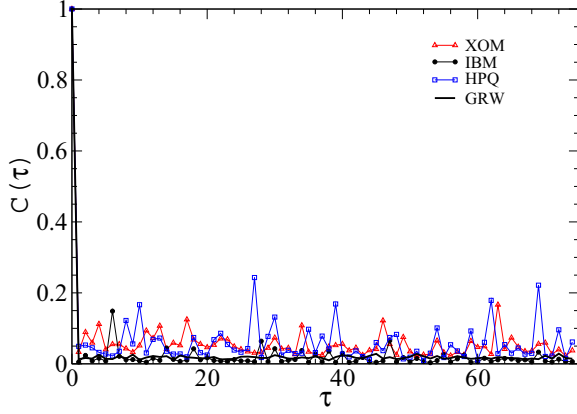


FIG. 4. (Color online) The autocorrelation function of the record ages obtained from three different stock data and GRW simulations. For GRW results, we have used the same value of  $N$  as for IBM stock,  $\langle \mu_{\text{emp}} \rangle = 0.00031$  and  $\langle \sigma_{\text{emp}} \rangle = 0.015$ . See text for details.

age. Clearly, shortest record age cannot be less than unity, a restriction arising from the resolution of the data measurement. Similarly, any record age longer than the length of the time series  $N$  cannot be resolved. In Refs. [9,14], it was pointed out that for a symmetric random walk process, the longest record age is proportional to  $N$ . However, the distribution of longest record age has not been discussed earlier.

In this section, we show that the longest record age falls in the class of type-II generalized extreme value distribution, namely, the Fréchet distribution [22]. First clue for this result arises from the record ages that are uncorrelated, to a good approximation. Figure 4 shows the autocorrelation function  $C(\tau) = \langle x_t x_{t+\tau} \rangle$  for the stock data. It reveals that the record ages are, at best, weakly correlated. The record ages obtained from GRW simulations [with parameters  $\mu = \langle \mu_{\text{emp}} \rangle, \sigma = \langle \sigma_{\text{emp}} \rangle$  chosen as done for Fig. 2(a)] also show a similar behavior. For such fast decay of correlations, extreme value theory for independent variables holds good [23]. Hence, we can expect the longest record age to follow the generalized extreme value distributions.

Figures 5(a) and 5(b) show the distribution of longest record age  $r_{\text{max}}$ , in terms of the scaled variable

$$z = 1 + k(r_{\text{max}} - a_N)/b_N,$$

for the GRW simulations with parameters the same as described in the legend of Fig. 2(a). In this,  $a_N$  and  $b_N$  are, respectively, the location and scale parameters dependent on  $N$ . This figure reveals a good agreement with the Fréchet distribution [22]

$$F(z) = \frac{1}{b_N} z^{-1-1/k} e^{-z^{-1/k}}, \quad (z > 0), \quad (3)$$

and  $F(z) = 0$  for  $z \leq 0$  and  $k > 0$  is the shape parameter. This is the extreme value distribution consistent with results shown in Figs. 1 and 2, i.e., the distribution of record ages  $P(r)$  has a lower-end cutoff and its tail decays as a power law. The agreement with Fréchet distribution gets better for  $N \gg 1$ . The dependence of the location parameter  $a_N$  and the scale parameter  $b_N$  on  $N$  shown in Figs. 5(c) and 5(d) reveals that  $\ln N$  function provides a good representation of

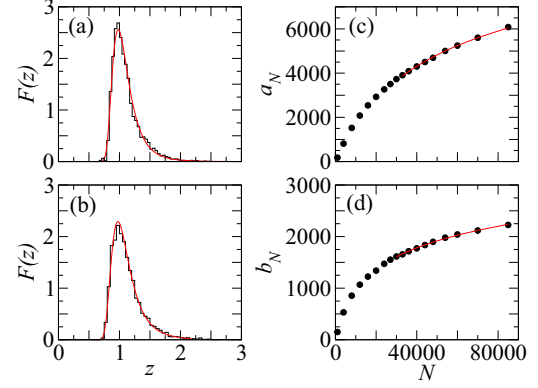


FIG. 5. (Color online) Scaled distribution of longest record age obtained from GRW simulations with (a)  $N = 15\,000$  and (b)  $N = 85\,000$ . The solid curve is the Fréchet distribution with shape parameter  $k > 0$ . (c) The location parameter  $a_N$  and (d) the scale parameter  $b_N$  of the Fréchet distribution shown as a function of  $N$ . The solid line in (c, d) is the logarithmic fit for  $N > 30\,000$ .

the data for  $N > 30\,000$ . Using this fit and the mean of Fréchet distribution  $\langle z \rangle = a_N + (b_N/k)[\Gamma(1-k) - 1]$ , we get the asymptotic mean of the longest record ages as  $\langle r_{\text{max}} \rangle \propto \ln N$ . This is the result obtained analytically in Ref. [14] without using extreme value theory.

Finally, we compute the distribution of the longest record age from stock data. To circumvent the shortage of data, we divided the empirical stock data into windows of length  $N = 1\,000$ . The longest record age from each of these windows was tabulated for each stock. All such data of extreme record ages from all the stocks were combined together to compute the (scaled) distribution shown in Fig. 6 as solid circles. The histogram in the figure is obtained from  $10^5$  ensemble GRW simulations with  $N = 1\,000$  and other parameters chosen as described in the legend of Fig. 2. The solid curve is the Fréchet distribution with  $k > 0$ . The distribution  $F(z)$  computed from stock data displays a reasonable agreement with Fréchet distribution. The deviations could partly be attributed to the

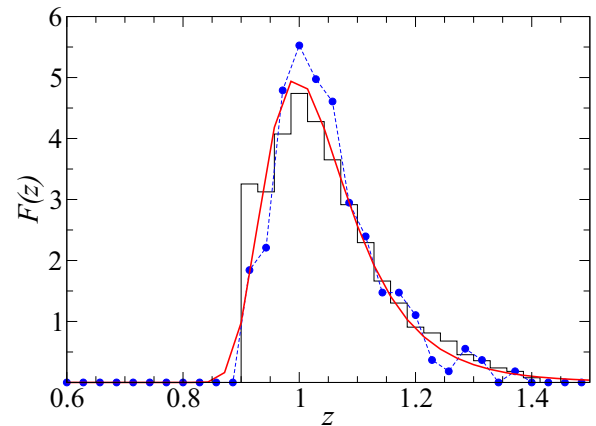


FIG. 6. (Color online) The distribution of scaled longest record ages computed from stock data (solid circles), GRW simulations (histogram) with  $N = 1\,000$ . The solid curve is the Fréchet distribution with parameters  $a_N$  and  $b_N$  corresponding to  $N = 1\,000$ .

insufficient stock market data to compute extreme record ages. We must also point out that both GRW simulations and stock data display pronounced deviation from Fréchet distribution for  $z < 1$ .

#### IV. SUMMARY AND DISCUSSIONS

In summary, we have analyzed the stock data for two quantities of interest in the study of record statistics, namely, the distribution of record age and the longest record age. The results have been obtained based on the analysis of 18 stocks for which the data is available in the public domain. We also study the geometric random walk series as a suitable reference model in the context of the time series of stocks. For the stock data and the GRW simulations, the record ages are distributed as a power law with exponent in the range  $1.5 \leq \alpha \leq 1.8$ . The record ages are uncorrelated, to a good approximation. The distribution of the longest record age is well described by the Fréchet distribution of the extreme value theory.

The results presented in this work also apply to the records statistics of the positions of a standard random walker. This is possible because the random walk and GRW are related through a simple time-independent transformation. The record age distribution  $P(r)$  is independent of  $N$  to within the numerical errors and it does not preclude the mean record age from being dependent on  $N$  [14]. Record ages being nearly uncorrelated implies that predicting the length of time before the occurrence of next record event based on historical data is unlikely to be easy, even though the mean record age can be determined [9]. The longest record age is Fréchet distributed for  $N \gg 1$  and pronounced deviations exist for small  $N$ . While an analysis of longer and bigger portfolio of stock data will yield better estimate for the power law exponent  $\alpha$  and also

for the longest record age distribution, it would be interesting to analytically obtain these results.

#### APPENDIX: DATA USED IN THE ANALYSIS

In this work, we use the daily closing values, corrected for splits and dividends, of the following stocks. These are publicly accessible from [finance.yahoo.com](http://finance.yahoo.com). Standard stock symbols used in the respective stock exchanges are used to indicate stock names.

Stock	Years	Length of data	Stock exchange
IBM (IBM)	1962–2012	12 764	NYSE
GIS (General Mills Inc.)	1983–2012	7 358	NYSE
AAPL(Apple Inc.)	1984–2012	7 067	NASDAQ
XOM (Exxon Mobil Inc.)	1970–2012	10 777	NYSE
FP.PA (Total SA)	2000–2012	3 435	PARIS
GD (General Dynamics Co.)	1977–2012	8 600	NYSE
GE (General Electric Co.)	1962–2012	12 764	NYSE
HPQ (Hewlett-Packard Co.)	1962–2012	12 747	NYSE
NTT (Nippon Telegraph...)	1994–2013	4 656	NYSE
SNP (China Petroleum and...)	2000–2013	3 124	NYSE
TM (Toyoto Motor Co.)	1993–2013	5 021	NYSE
VOW.DE (Volkswagen AG)	2000–2013	3 423	XETRA
CVX (Chevron Co.)	1970–2013	10 910	NYSE
WMT (Walmart Stores Inc.)	1972–2013	10 238	NYSE
F (Ford Motors)	1977–2013	9 141	NYSE
COP (ConocoPhillips)	1982–2013	7 878	NYSE
BRK.A (Berkshire Hathaway)	1990–2013	5 825	NYSE
BP (BP plc)	1988–2013	6 445	NYSE

- [1] C. DeConcini and C. F. Tompkins, World Resources Institute Fact Sheet (2013), [www.wri.org/publication/fact-sheet-2012-year-record-breaking-extreme-weather-and-climate](http://www.wri.org/publication/fact-sheet-2012-year-record-breaking-extreme-weather-and-climate).
- [2] J. J. Hernandez Gmez, V. Marquina, and R. W. Gomez, *Eur. J. Phys.* **34**, 1227 (2013); see also [www.wired.com/playbook/2012/08/physics-long-jump-linear-regression](http://www.wired.com/playbook/2012/08/physics-long-jump-linear-regression); D. Gembris, J. G. Taylor, and D. Suter, *J. Appl. Stat.* **34**, 529 (2007).
- [3] S. Rahmstorf and D. Coumou, *Proc. Natl. Acad. Sci. USA* **108**, 17905 (2011).
- [4] T. C. Peterson, M. P. Hoerling, P. A. Stott, and S. C. Herring, *Bull. Am. Meteorol. Soc. (Special Supplement)* **94**, S1 (2013).
- [5] B. Alessandro, C. Beatrice, G. Bertotti, and A. Montorsi, *J. Appl. Phys.* **68**, 2901 (1990).
- [6] C. Godrèche and J. M. Luck, *J. Stat. Mech.* (2008) P11006.
- [7] L. P. Oliveira, H. J. Jensen, M. Nicodemi, and P. Sibani, *Phys. Rev. B* **71**, 104526 (2005); P. Sibani, G. F. Rodriguez, and G. G. Kenning, *ibid.* **74**, 224407 (2006).
- [8] Shashi C. L. Srivastava, A. Lakshminarayan, and Sudhir R. Jain, *Europhys. Lett.* **101**, 10003 (2013).
- [9] S. N. Majumdar and R. M. Ziff, *Phys. Rev. Lett.* **101**, 050601 (2008).
- [10] G. Wergen, *J. Phys. A* **46**, 223001 (2013); J. Krug, *J. Stat. Mech. Theory Exp.* (2007) P07001; I. Eliazar and J. Klafter, *Phys. Rev. E* **80**, 061117 (2009); J. Franke, G. Wergen, and J. Krug, *J. Stat. Mech. Theory Exp.* (2010) P10013; E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **88**, 022145 (2013).
- [11] C. M. Rowe and L. E. Derry, *Geophys. Res. Lett.* **39**, L16703 (2012); W. I. Newman, B. D. Malamud, and D. L. Turcotte, *Phys. Rev. E* **82**, 066111 (2010); G. Wergen and J. Krug, *Europhys. Lett.* **92**, 30008 (2010).
- [12] Y. Edery, A. B. Kostinski, S. N. Majumdar, and B. Berkowitz, *Phys. Rev. Lett.* **110**, 180602 (2013); S. Sabhapandit, *Europhys. Lett.* **94**, 20003 (2011); S. N. Majumdar, *Physica A* **389**, 4299 (2010); C. Godreche, S. N. Majumdar, and G. Schehr, *J. Phys. A* **47**, 255001 (2014).
- [13] G. Wergen, S. N. Majumdar, and G. Schehr, *Phys. Rev. E* **86**, 011119 (2012).
- [14] S. N. Majumdar, G. Schehr, and G. Wergen, *J. Phys. A* **45**, 355002 (2012).
- [15] V. B. Nevzorov, *Records: Mathematical Theory* (American Mathematical Society, Providence, RI, 2001); B. Schmittmann and R. K. P. Zia, *Am. J. Phys.* **67**, 1269 (1999).

- [16] P. Doukhan, G. Oppenheim, and M. S. Taqqu, *Theory and Applications of Long-Range Dependence* (Springer, Berlin, 2003).
- [17] G. Wergen, M. Bogner, and J. Krug, *Phys. Rev. E* **83**, 051109 (2011).
- [18] G. Wergen, *Physica A* **396**, 114 (2014).
- [19] David Ruppert, *Statistics and Data Analysis for Financial Engineering* (Springer, New York, 2011); Frank J. Fabozzi, *Encyclopedia of Financial Models*, 1st ed. (Wiley, New York, 2012).
- [20] R. Kuhn and P. Neu, *J. Phys. A* **41**, 324015 (2008).
- [21] A. Clauset, C. R. Shalizi, and M. E. J. Newman, *SIAM Rev.* **51**, 661 (2009).
- [22] S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer, Berlin, 2001).
- [23] M. R. Leadbetter and H. Rootzen, *Ann. Probab.* **16**, 431 (1988).