Record statistics for biased random walks, with an application to financial data

Gregor Wergen, Miro Bogner, and Joachim Krug

Institut für Theoretische Physik, Universität zu Köln, 50937 Köln, Germany

(Received 4 March 2011; published 9 May 2011)

We consider the occurrence of record-breaking events in random walks with asymmetric jump distributions. The statistics of records in symmetric random walks was previously analyzed by Majumdar and Ziff [Phys. Rev. Lett. **101**, 050601 (2008)] and is well understood. Unlike the case of symmetric jump distributions, in the asymmetric case the statistics of records depends on the choice of the jump distribution. We compute the record rate $P_n(c)$, defined as the probability for the nth value to be larger than all previous values, for a Gaussian jump distribution with standard deviation σ that is shifted by a constant drift c. For small drift, in the sense of $c/\sigma \ll n^{-1/2}$, the correction to $P_n(c)$ grows proportional to $\arctan(\sqrt{n})$ and saturates at the value $\frac{c}{\sqrt{2}\sigma}$. For large n the record rate approaches a constant, which is approximately given by $1 - (\sigma/\sqrt{2\pi}c)\exp(-c^2/2\sigma^2)$ for $c/\sigma \gg 1$. These asymptotic results carry over to other continuous jump distributions with finite variance. As an application, we compare our analytical results to the record statistics of 366 daily stock prices from the Standard & Poor's 500 index. The biased random walk accounts quantitatively for the increase in the number of upper records due to the overall trend in the stock prices, and after detrending the number of upper records is in good agreement with the symmetric random walk. However the number of lower records in the detrended data is significantly reduced by a mechanism that remains to be identified.

I. INTRODUCTION

The random walk is a paradigmatic model of statistical physics, which combines utmost conceptual simplicity with a surprising richness of emergent behaviors [1,2]. Among the many interesting features of random walks, recent research has focused in particular on its extremal properties, exploring quantities such as the height and position of the globally maximal excursion of a one-dimensional walk of a given number of steps, and the statistics of *records* of this process [3]. Here a record is defined as an entry in a discrete, real valued time series that is larger (upper record) or smaller (lower record) than all previous entries. While the mathematical theory of records is well developed for time series of independent, identically distributed random variables [4–6], little has been known about the record statistics of correlated processes. It is therefore remarkable that records of a large class of one-dimensional random walks can be characterized in considerable detail, as was shown in recent work by Majumdar and Ziff (MZ) [7]. Specifically, they considered the random process defined by

$$X_n = X_{n-1} + \xi_n, \qquad (1)$$

where $X_0 = 0$ (say) and the step sizes ξ_n are independent, identically distributed random variables drawn from a probability density $\phi(\xi)$ that is required to be continuous and symmetric, but is otherwise arbitrary. We say that an upper record occurs at time n if $X_n = \max\{X_0, \ldots, X_n\}$. Based on the Sparre Andersen theorem for the survival probability of the random walk [8–11], MZ show that the probability $\Pi(m,n)$ for the nth event to be the mth record is given by

$$\Pi(m,n) = \binom{2n - m + 1}{m} 2^{-2n+m-1} \qquad (2)$$

for $m \leqslant n + 1$. The first moment of this distribution with respect to m yields the mean number of records after n steps, which equals $m_n \approx \frac{2}{\sqrt{\pi}}\sqrt{n}$ for large n, and the probability P_n

for the nth event to be a record (henceforth referred to as the *record rate*) decays like $P_n \approx \frac{1}{\sqrt{\pi n}}$. In the present paper we aim to generalize these results to random walks with asymmetric jump distributions. In the first part of the paper (Sections II and III) we study records generated by random walks with a symmetric jump distribution that have an additional constant *drift* c, such that (1) generalizes to

$$X_n = X_{n-1} + \xi_n + c \qquad (3)$$

with a symmetric jump distribution $\phi(\xi)$. For the special case of a Cauchy distribution this problem was considered previously in [12]. Here, we derive approximate results for the case of a Gaussian jump distribution that apply also more generally to distributions with a finite variance.

Similar to our earlier work [13,14] on the related problem of records from independent random variables with drift [12,15], our strategy will be to analyze the limiting cases of small and large drift, respectively, as quantified by the ratio c/σ of the drift speed to the standard deviation σ of the jump distribution $\phi(\xi)$. For the Gaussian random walk we find that in the limit of $\frac{c}{\sigma} \ll \frac{1}{\sqrt{n}}$ the mean number of records and the record rate are given by

$$m_n(c) \approx \frac{2\sqrt{n}}{\sqrt{\pi}} + \frac{c}{\sigma}\frac{\sqrt{2}}{\pi}(n \arctan(\sqrt{n}) - \sqrt{n}), \qquad (4)$$

$$P_n(c) \approx \frac{1}{\sqrt{\pi n}} + \frac{c}{\sigma}\frac{\sqrt{2}}{\pi}\arctan(\sqrt{n}). \qquad (5)$$

In the limit of $\frac{c}{\sigma} \gg \frac{1}{\sqrt{n}}$ the record rate $P_n(c)$ approaches a constant value. If in addition $\frac{c}{\sigma} \gg 1$, this constant is given approximately by

$$\lim_{n \to \infty} P_n \approx 1 - \frac{c}{\sqrt{2\pi}\sigma}e^{-\frac{c^2}{2\sigma^2}}. \qquad (6)$$

In Section IV we apply our results to fluctuations in stock prices, arguably one of the most important (and ancient)

applications of random walk theory [16–18]. The basic model of a stock price is the geometric random walk $S_n = e^{X_n}$ with an upward bias reflecting long-term economic growth. Our analysis of record events in the Standard & Poor's 500 index shows a corresponding surplus of upper record events, which is consistent with the theoretical expectation. However, an asymmetry between upper and lower records remains even when the bias has been (approximately) removed [19], a feature that may be related to the gain-loss asymmetry reported in previous analyses of stock market fluctuations [20–23]. We conclude with a summary and a discussion of some open problems.

II. SURVIVAL PROBABILITIES AND FIRST PASSAGE TIMES

The record statistics of a random walk can be analyzed by considering the generating functions of the survival and first passage probabilities of the process [3,7,12]. In [7] it was shown that the generating function of $\Pi(m,n)$ is of the form

$$\sum_{n=m-1}^{\infty} \Pi(m,n)z^n = \tilde{f}_-^{m-1}(z)\tilde{q}_-(z), \qquad (7)$$

where $\tilde{q}_\pm(z)$ is the generating function of the positive (negative) survival probability $q_\pm(n)$ of the random walk. The latter is defined as the probability that the process stays above (below) the origin up to the nth step. Similarly $\tilde{f}_\pm(z)$ is the generating function of the positive (negative) first-passage probability $f_\pm(n)$ of the random walk, with $f_\pm(n) = q_\pm(n-1) - q_\pm(n)$. In the case of the symmetric random walk considered in [7] we have $q_-(n) = q_+(n) = q(n)$ and $f_-(n) = f_+(n) = f(n)$ and both $q(n)$ and $f(n)$ are completely universal for all continuous jump distributions.

Since we want to study asymmetric random walks, we need to distinguish between positive and negative survival probabilities and first passage times, and consider the functions $q_\pm(n)$ and $f_\pm(n)$. As in [7] a theorem by Sparre Andersen will play a key role in our considerations. In [8,11] it was shown that

$$\tilde{q}_\pm(z) = \sum_{n=0}^{\infty} q_\pm(n)z^n = \exp\left(\sum_{n=1}^{\infty} \frac{p_\pm(n)}{n}z^n\right), \qquad (8)$$

where $p_\pm(n)$ is the probability for the walker to be above or below the origin at the nth step. This quantity can be easily computed from $p_\pm = \int_0^\infty G(\pm x,n)\mathrm{d}x$, where $G(x,n)$ is the positional probability density of a random walk of n steps that started at the origin. Details on the computation of $G(\pm x,n)$ and $p_\pm(n)$ can be found in [3] and [10]. In the case of a symmetric random walk we simply have $p_\pm(n) = \frac{1}{2}$ independent of n and we find that in this case $\tilde{q}_\pm(z) = (1-z)^{-\frac{1}{2}}$ and $q_\pm(n) = \binom{2n}{n}2^{-2n}$ [7]. These results eventually lead to Eq. (2) [7].

In the case of an asymmetric random walk the situation gets a bit more complicated. We compute $q_\pm(n)$ and its generating function for a Gaussian random walk with drift c. Here, the jump distribution of the symmetric random variable ξ in (3) is of the form $\phi(\xi) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\xi^2}{2\sigma^2}}$ with standard deviation σ. It is easy to show that the probability density $G(x,n)$ of the random

walk after n steps is given by $G(x,n) = \frac{1}{\sqrt{2\pi}\sigma n}\exp(-\frac{(x-nc)^2}{2\sigma^2 n})$ and $p_\pm(n) = \frac{1}{2}(1 \pm \mathrm{erf}(\sqrt{\frac{n}{2}}\frac{c}{\sigma}))$. We start with the case of a small linear drift with $c \ll \frac{\sigma}{\sqrt{n}}$ such that $p_\pm(n) \approx \frac{1}{2} \pm \sqrt{\frac{n}{2\pi}}\frac{c}{\sigma}$. Now we can employ Eq. (8). Expanding up to first order in c we find

$$\tilde{q}_\pm(z) \approx \frac{1}{\sqrt{1-z}}\left(1 \pm \frac{c}{\sqrt{2\pi}\sigma}\sum_{n=1}^{\infty}\frac{z^n}{\sqrt{n}}\right). \qquad (9)$$

With $\sqrt{1-z}^{-1} = \sum_{n=0}^{\infty}\binom{2n}{n}2^{-2n}z^n$ and making use of the Cauchy formula for products of infinite sums we obtain the following expression for $\tilde{q}_\pm(z)$:

$$\tilde{q}_\pm(z) \approx \frac{1}{\sqrt{1-z}} \pm \frac{c}{\sqrt{2\pi}\sigma}\sum_{n=0}^{\infty}\sum_{k=0}^{n}\binom{2k}{k}\frac{2^{-2k}z^{n+1}}{\sqrt{n-k+1}}. \qquad (10)$$

The binomial coefficient can be approximated by $\binom{2k}{k} \approx 4^k/\sqrt{\pi k}$ and with this we can approximate the sum over k by an integral,

$$\sum_{k=0}^{n}\frac{1}{\sqrt{\pi k}}\frac{z^{n+1}}{\sqrt{n-k+1}} \approx \frac{1}{\sqrt{\pi}}\int_0^n\frac{\mathrm{d}k z^{n+1}}{\sqrt{k}\sqrt{n-k+1}}$$
$$\approx \sqrt{\pi}z^{n+1}. \qquad (11)$$

We thus obtain a simple result for the generating function of the survival probability,

$$\tilde{q}_\pm(z) \approx \frac{1}{\sqrt{1-z}} \pm \frac{c}{\sqrt{2}\sigma}\sum_{n=1}^{\infty}z^n, \qquad (12)$$

and finally the following expression for the survival probability $q_\pm(n)$ under a small linear drift:

$$q_\pm(n) \approx \frac{1}{\sqrt{\pi n}} \pm \frac{c}{\sqrt{2}\sigma}. \qquad (13)$$

The first term on the right hand side is the result for the symmetric random walk discussed in [7], which is now supplemented by a correction linear in $\frac{c}{\sigma}$. Although this particular result for $q_\pm(n)$ will not be needed in our derivation of the record statistics, we found it useful to test it against numerical simulations. The results are shown in Fig. 1. For small c, Eq. (13) is in good agreement with the simulations.

For the sake of completeness we also provide the small c expansion of $\tilde{f}_\pm(z)$ that will become important later. With $\tilde{f}_\pm(z) = 1 - (1-z)q_\pm(z)$ we find

$$\tilde{f}_\pm(z) \approx 1 - \sqrt{1-z}\left(1 \pm \frac{c}{\sqrt{2\pi}\sigma}\sum_{n=1}^{\infty}\frac{z^n}{\sqrt{n}}\right). \qquad (14)$$

From this we obtain, by methods very similar to those used above to derive $q_\pm(n)$, the result

$$f_\pm(n) \approx \frac{1}{2\sqrt{\pi}}n^{-\frac{3}{2}} \pm \frac{c}{\sqrt{2\pi}\sigma}n^{-\frac{1}{2}}. \qquad (15)$$

Next we consider the case of large drift, $\frac{c}{\sigma} \gg 1$. Here we will only discuss $q_-(n)$, as this is the quantity needed for the computation of the record rate; $q_+(n)$ has a different behavior in this regime. In the limit of $\frac{c}{\sigma} \gg 1$ we find that $p_-(n) \approx \sigma(2\pi nc^2)^{-1/2}e^{-\frac{c^2 n}{2\sigma^2}}$. Using this we find that for large n, $\tilde{q}_-(z)$
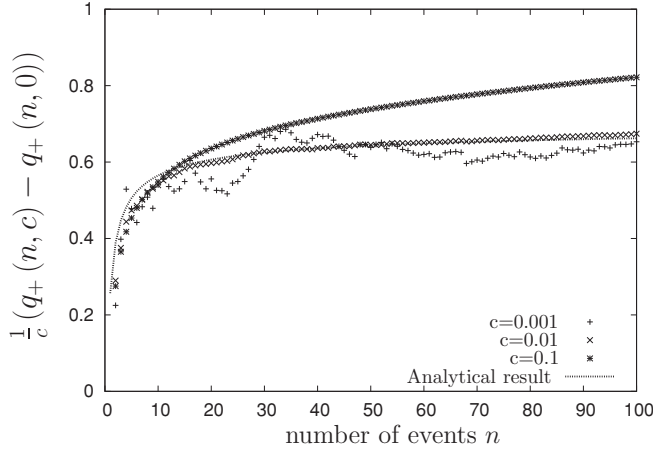
FIG. 1. Relative effect of the drift on the positive survival probability $q_+(n,c)$ of a Gaussian random walk with $\sigma = 1$. The effect of the drift is represented by $\frac{1}{c}(q_+(n,c) - q_+(n,0))$ for different drift speeds c. We simulated 10^7 realization of a random walk with $n = 100$ steps for each drift speed. The dotted line represents the analytical results obtained in Eq. (13). For small drift $c = 0.001$ and $c = 0.01$ we find good agreement with this approximation.

and $q_-(n)$ are of the form

$$\tilde{q}_-(z) \approx 1 + \sum_{n=1}^{\infty} \frac{\sigma}{c\sqrt{2\pi n^3}} e^{-\frac{c^2 n}{2\sigma^2}} z^n, \quad (16)$$

$$q_-(n) \approx \frac{\sigma}{c\sqrt{2\pi n^3}} e^{-\frac{c^2 n}{2\sigma^2}}. \quad (17)$$

These particular results were already reported in [12]. At this point it is important to notice that all results concerning the first-passage and survival probabilities in the large n limit are easily transferable to other jump distributions as long as these have a finite variance. Because of the central limit theorem, $G(\pm x, n)$ and therefore $p_{\pm}(n)$ will approach the same expressions for large n as were derived here for the Gaussian jump distribution.

III. GAUSSIAN RANDOM WALKS WITH DRIFT

A. Record rate for small c/σ and $n \ll (\sigma/c)^2$

With the small c expansions in Eqs. (9) and (14) we have all ingredients needed to derive the record statistics for a Gaussian random walk with a small linear drift. We start by computing the mean number of records m_n expected up to the nth step. For the generating function $\tilde{m}(z) = \sum_{n=0}^{\infty} m_n z^n$ of this quantity it was found in [12] that $\tilde{m}(z) = 1/((1 - z)^2 \tilde{q}_-(z))$, a result that can be obtained by computing the first moment of Eq. (7). We can now evaluate this expression making use of the generating function for $q_-(n)$ given in Eq. (9). In the limit of small $\frac{c}{\sigma}$ this yields

$$\tilde{m}(z) \approx \frac{1}{\sqrt{1 - z}^3} \left(1 + \frac{c}{\sqrt{2\pi}\sigma} \sum_{n=1}^{\infty} \frac{z^n}{\sqrt{n}} \right). \quad (18)$$

Using the series expansion of $\sqrt{1 - z}^{-3}$ and employing once again the Cauchy formula for infinite sums and the Stirling

approximation, we find

$$\tilde{m}(z) \approx \frac{1}{\sqrt{1 - z}^3} + \frac{\sqrt{2}c}{\pi\sigma} \sum_{n=1}^{\infty} z^n \sum_{k=0}^{n-1} \frac{\sqrt{k}}{\sqrt{n - k + 1}}. \quad (19)$$

If n is not too small, the sum over k can be replaced by an integral and we finally obtain an approximate expression for the generating function of m_n,

$$\tilde{m}(z) \approx \frac{1}{\sqrt{1 - z}^3} + \frac{\sqrt{2}c}{\pi\sigma} \sum_{n=1}^{\infty} z^n (n \arctan(\sqrt{n}) - \sqrt{n}). \quad (20)$$

The mean number of records of the random walk with a small linear drift c is therefore approximately given by

$$m_n \approx \binom{2n}{n} \frac{2n + 1}{2^{2n}} + \frac{\sqrt{2}c}{\pi\sigma} (n \arctan(\sqrt{n}) - \sqrt{n}). \quad (21)$$

Making use of the Stirling approximation this yields the previously announced expression (4) for $m_n(c)$ and, by taking a derivative with respect to n, the record rate $P_n(c)$ in the large n limit as given in Eq. (5). The leading order correction of the record rate due to the drift is seen to increase with $\arctan(\sqrt{n})$ and for larger n it approaches a constant value. For large n (but still in the regime $\frac{c}{\sigma} \ll \frac{1}{\sqrt{n}}$) we find the simple result

$$P_n(c) \approx \frac{1}{\sqrt{\pi n}} + \frac{c}{\sqrt{2}\sigma}. \quad (22)$$

We compared Eq. (5) to simulations and found good agreement in the regime $\frac{c}{\sigma} \ll \frac{1}{\sqrt{n}}$ (Fig. 2). We also compared this result with numerical simulations of the record rate for random walks with step sizes drawn from a uniform distribution (Fig. 3). The results for the Gaussian and the uniform distribution are very similar to each other already for small n, reflecting the convergence expected from the central limit theorem.
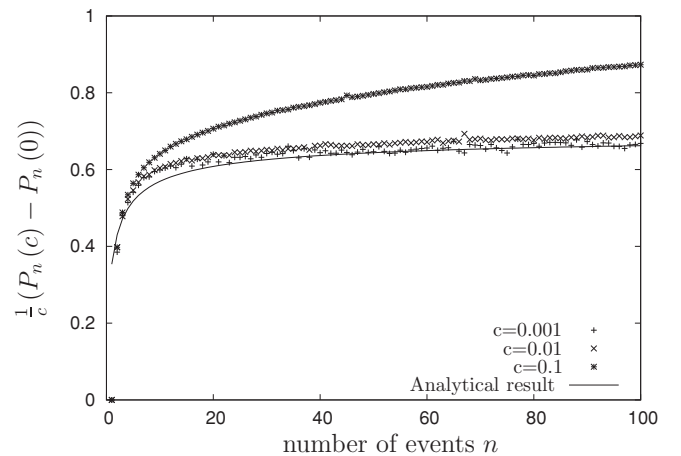

FIG. 2. Relative effect of the drift on the record rate $P_n(c)$ of a random walk with a Gaussian jump distribution ($\sigma = 1$). The effect is represented by $\frac{1}{c}(P_n(c) - P_n(0))$ for different drift speeds c. Again, we simulated 10^7 realizations of a random walk with $n = 100$ steps for each drift speed. The line represents the analytical results obtained in Eq. (5). For small drift speeds $c = 0.001$ and $c = 0.01$ we find good agreement with the approximation, but for $c = 0.1$ the approximation is no longer accurate.
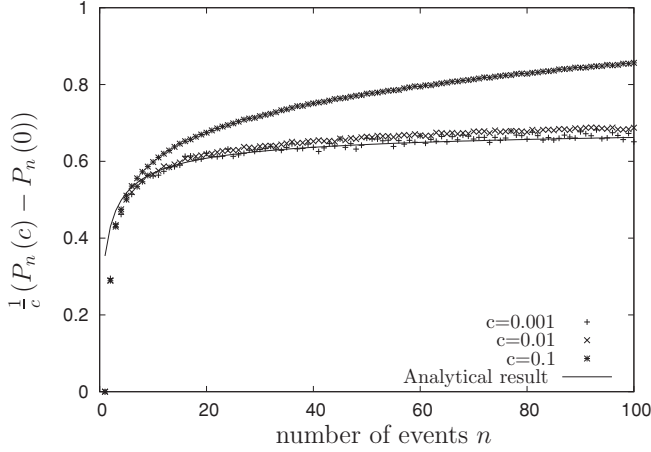
FIG. 3. Relative effect of the drift on the record rate $P_n(c)$ for a random walk with a uniform jump distribution with standard deviation $\sigma = 1$. The parameters of the simulation are the same as in Fig. 2. Even though the expression (5) was derived for a Gaussian jump distribution, it is in a good agreement with the numerical results for small c.

B. Asymptotic record rate for large n

Next we consider the limit of strong drift, $\frac{c}{\sigma} \gg 1$. Applying the same method as above and making use of our result (16) for $\tilde{q}_-(z)$ in the regime of large c/σ, we find that the number of records increases linearly with time according to

$$m_n(c) \approx n \left(1 - \frac{\sigma}{\sqrt{2\pi} c} e^{-\frac{c^2}{2\sigma^2}} \right). \tag{23}$$

Correspondingly the record rate P_n is independent of n in this case. In fact, simulations show that the record rate approaches a finite, nonzero limit $P(c) \equiv \lim_{n \to \infty} P_n(c)$ for $n \to \infty$ for any positive value of the drift (Fig. 4). This can be understood, on the basis of the general relation (7) between the distribution of record events and the negative first passage probability, to be a consequence of the fact that the negative mean first passage time of a random walk with positive drift is finite [9, 10]; roughly speaking, one expects that the asymptotic record rate $P(c)$ is proportional to the inverse of the negative mean first passage time. The result (23) implies that the asymptotic record rate behaves as

$$P(c) \approx 1 - \frac{\sigma}{\sqrt{2\pi} c} e^{-\frac{c^2}{2\sigma^2}} \tag{24}$$

for large c/σ (see inset of Fig. 4). Furthermore, since the negative mean first passage time diverges as c^{-1} for $c \to 0$ [9,10], the asymptotic record rate should behave as $P(c) \sim c$ for small c. This is confirmed by the simulations, which indicate that $P(c) \approx 1.39 \, (c/\sigma)$ for $c/\sigma \ll 1$.

The time scale $n^*(c)$ at which the saturation of the record rate occurs can be estimated by comparing the two terms in Eq. (22), which shows that

$$n^* \sim \left(\frac{\sigma}{c} \right)^2 \tag{25}$$

for small c. Not surprisingly, this is also the time scale at which the drift begins to dominate the mean square displacement of
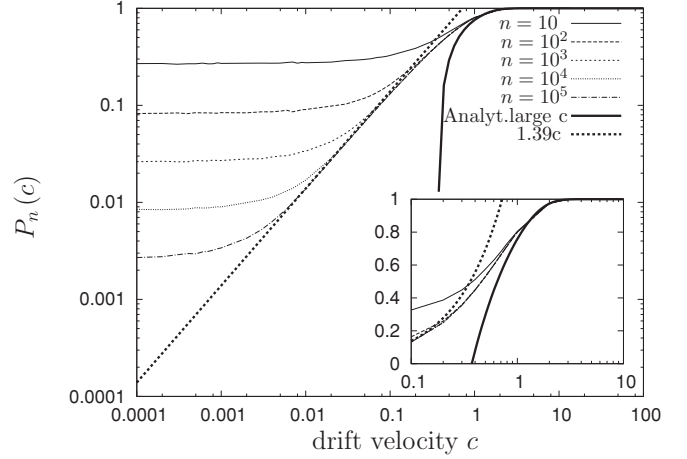

FIG. 4. Record rate for a biased Gaussian random walk with standard deviation $\sigma = 1$. The figure illustrates the convergence of $P_n(c)$ to the asymptotically constant record rate $P(c)$ for $n \to \infty$. The inset shows that the large drift result (24) becomes accurate for $c/\sigma > 1$, and the bold dotted line shows that $P(c) \approx 1.39 \frac{c}{\sigma}$ for $c \to 0$.

the random walk. Together with the linear behavior of the asymptotic record rate, this suggests the scaling form

$$P_n(c) = \frac{c}{\sigma} g((c/\sigma)^2 n) \tag{26}$$

for small c/σ and arbitrary n, where the limiting behaviors of the scaling function are $g(x \to 0) \approx \frac{1}{\sqrt{\pi x}}$ and $g(x \to \infty) \approx 1.39$. This relation is well fulfilled by the numerical data shown in Fig. 5.

IV. RECORD STATISTICS OF STOCK PRICES IN THE S&P 500

A prominent application of the random walk process can be found in the financial sciences. Originally introduced by Bachelier in 1900 [16], the geometric random walk is the standard model used to describe the evolution of stock prices. In the application of this model to actual data, trends are
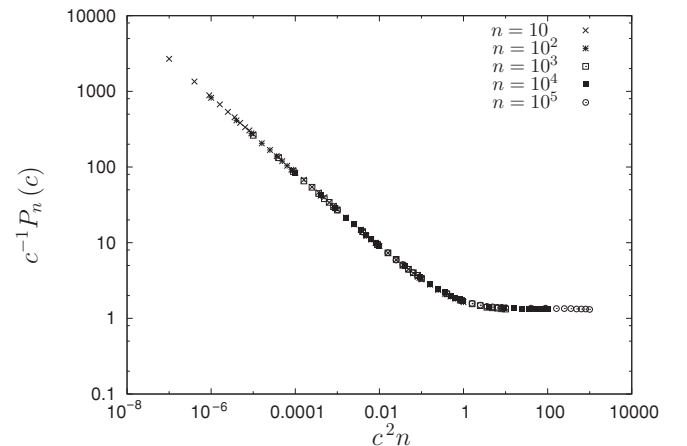

FIG. 5. Illustration of the conjectured scaling collapse (26) of the record rate $P_n(c)$ for Gaussian random walks with $\sigma = 1$ and various drift speeds $c \leqslant 0.1$.

always an issue, which in the simplest case are described by a linear drift in the logarithm of the stock price. In this section we present an empirical analysis of record events in historical stock prices taken from the Standard & Poor's 500 index, and compare the results to the theoretical predictions derived above. The observational data we used consist of daily recordings of 366 stocks that were contained in the index from January 1990 to March 2009, resulting in 366 time series of length $n = 5000$ [24]. We first analyzed the recordings without any detrending and then considered detrended data in which a fitted linear trend was subtracted from the logarithms of the stock prices.

In the raw stock data the number of upper records after $n = 5000$ trading days is considerably larger than the expected number of $2\sqrt{5000/\pi} \approx 79.79$ for a symmetric random walk. At the end of the observation period, we found an average number of 166.56 upper records in the stocks, but only 22.33 lower records. The rate of upper records was roughly constant over the entire period, whereas the rate of lower records was almost zero already after 300 days. Apparently a positive trend had a very strong effect on the record statistics of the analyzed stocks. To quantify the trend, we performed a linear regression analysis on the logarithms of the individual stock prices, determining the drift c_i and the standard deviation of increments σ_i for each stock $i = 1, \ldots, 366$. The normalized drift c_i/σ_i was then averaged over all stocks, yielding the estimate $\langle c_i/\sigma_i \rangle \approx 0.025$. At $n = 5000$ we are thus well outside the regime in which the pertubative result (4) should be valid. Still, inserting the estimated normalized drift $c/\sigma = 0.025$ into (4) we obtain a record number of 166.59, in very close agreement with the observed value. The comparison with Monte Carlo simulations of biased random walks with the same drift shows that this accuracy is actually fortuitous, but the description of the stock market data by the biased random walk model is nevertheless quite reasonable (Fig. 6).

Next we detrended the data by subtracting the fitted linear trend from the logarithmic stock prices, and counted the number of records in the detrended time series. We found an average number of 75.79 upper records after 5000 steps, in close agreement with the result for a symmetric random walk. However, the number of lower records was only 53.65, which is significantly smaller than expected. This residual asymmetry between upper and lower records persists if, instead of subtracting an overall linear trend, the data are detrended by normalizing each stock by the index [19]. To further explore this phenomenon we split the time series into 50 shorter series each lasting 100 trading days. We detrended each of the shorter time series individually by subtracting a linear trend, counted the number of upper and lower records, and then averaged the record numbers over the whole ensemble of 50×366 series of length 100. The results are shown in Fig. 7. It appears that while the number of upper records is in very good agreement with the symmetric random walk model, the number of lower records is still suppressed. This effect was found for different choices of the lengths of the time series and appears to be independent of this choice.

Qualitatively, a reduced number of lower records indicates that the positive first-passage times are increased compared to the corresponding negative first passage times. An asymmetry between first passage times to a prescribed (positive or
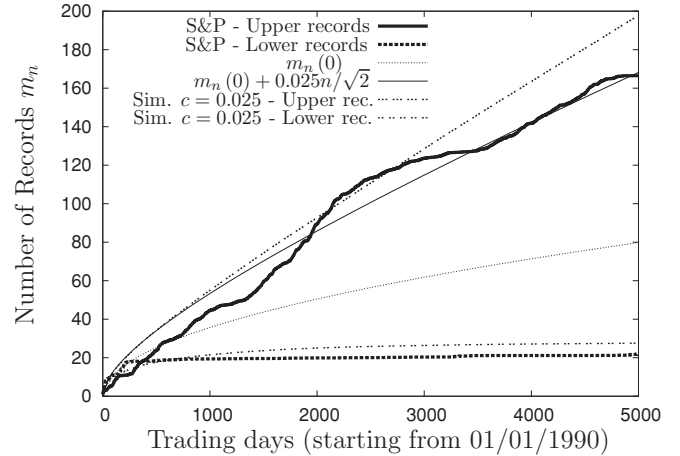

FIG. 6. Mean number of records averaged over 366 stocks from the S&P 500 index, computed from daily values for the time period 1.1.1990–31.3.2009. Full thick line shows the number of upper records, dotted thick line shows the number of lower records in the data set. The expected number of records $m_n(0) = \frac{2}{\sqrt{\pi}}\sqrt{n}$ for a symmetric random walk is shown by the thin dotted line. Also shown are the predictions of the biased random walk model with effective normalized drift $c/\sigma = 0.025$ obtained from Monte Carlo simulations as well as from the approximate expression $m_n(c) = m_n(0) + \frac{c}{\sqrt{2}\sigma}n$ (thin full line).

negative) return level has in fact been observed in previous analyses of stock market data, and is known as the *gain-loss asymmetry* [20–23]. However, this phenomenon differs in several important respects from the one reported here. First, in most (though not all [23]) cases the sign of the asymmetry is opposite to that suggested by the asymmetry
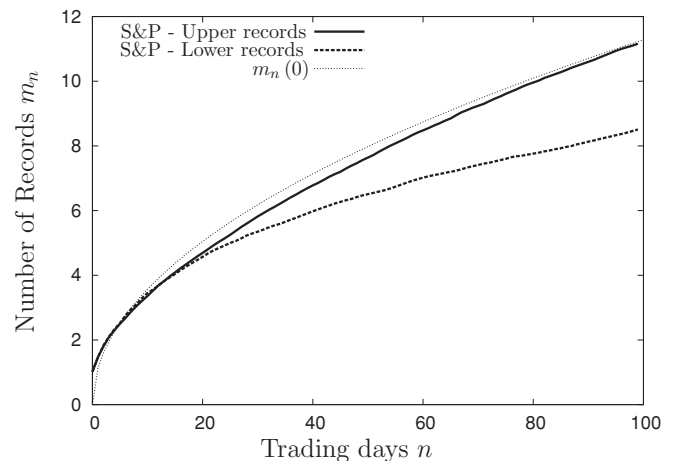

FIG. 7. Mean number of records in subsequences of the time series taken from the S&P 500 index. The entire data set of 5000 consecutive daily values was split into 50 subsequences of length 100. For each of the subsequences a linear detrending of the logarithm of the daily values was performed and the upper and lower record numbers were determined from the detrended data. The results, averaged over all stocks and all subsequences, are given by the thick black line (upper records) and the thick dashed line (lower records). The thin dashed line shows the analytical prediction for a symmetric random walk $m_n(0) = 2\sqrt{n/\pi}$. The number of upper records is in good agreement with $m_n(0)$, but the number of lower records is significantly reduced.

in the record statistics, in that first passage times for crossing a prescribed level from below are larger than for crossings from above [20,22]. Second, the asymmetry vanishes when the prescribed return level tends to zero, which is the relevant limit for the analysis of records. Finally, in contrast to the asymmetry between upper and lower records reported here, the gain-loss asymmetry is a property of entire stock indices which does not occur in individual stocks [21,22]. Indeed, a preliminary analysis of first-passage times to the origin in the detrended S&P 500 data shows an asymmetry between positive and negative excursions only when the starting point of the excursion is conditioned to be a record event [25]. An explanation of the observed residual asymmetry between upper and lower records must therefore be left to future work.

V. SUMMARY

In conclusion, using the methods introduced in [7] and a more general form of the Sparre Andersen Theorem [3,8], we were able to describe the effect of a linear drift on the record statistics of a Gaussian random walk in two regimes. For short times $n \ll (\frac{\sigma}{c})^2$ we find that the correction to the record rate $P_n(c) - P_n(0)$ increases proportional $\arctan(n)$ and then saturates at a value of $\frac{c}{\sqrt{2}\sigma}$. On the other hand, for large n the record rate saturates at a constant limiting value $P(c)$, which is linear in c for $c/\sigma \ll 1$ and approaches unity for large c/σ according to Eq. (24). The transition between the two regimes is described by the scaling form (26).

We applied our results to the statistics of records in 366 stocks contained in the S&P 500 index from 1990 to 2009. We found that, after detrending, the number of upper records in the stocks is basically identical to that predicted for the symmetric random walk. The fact that the number of lower records appears to be systematically decreased is interesting and needs to be examined more thouroughly in the future. On the theoretical side, a possible topic for future research is the record statistics of asymmetric random walks with a more complicated asymmetry than just a constant drift. The issue of asymmetric random walks with discrete jump distributions is also still open for further investigations.

[1] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A kinetic view of statistical physics* (Cambridge University Press, 2010).

[2] G. H. Weiss, *Aspects and applications of the random walk* (North-Holland, 1994).

[3] S. N. Majumdar, Physica A **389**, 4299 (2010).

[4] N. Glick, Am. Math. Mon. **85**, 2 (1978).

[5] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *Records* (John Wiley & Sons, New York, 1998).

[6] V. B. Nevzorov, *Records* (American Mathematical Society, 2004).

[7] S. N. Majumdar and R. M. Ziff, Phys. Rev. Lett. **101**, 050601 (2008).

[8] E. Sparre Andersen, Math. Scand. **2**, 195 (1954).

[9] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I and II (Wiley, New York, 1968).

[10] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, 2001).

[11] M. Bauer, C. Godrèche, and J. M. Luck, J. Stat. Phys. **96**, 963 (1999).

[12] P. Le Doussal and K. J. Wiese, Phys. Rev. E **79**, 051105 (2009).

[13] J. Franke, G. Wergen, and J. Krug, J. Stat. Mech. (2010) P10013.

[14] G. Wergen and J. Krug, Europhys. Lett. **92**, 30008 (2010).

[15] R. Ballerini and S. Resnick, Adv. Appl. Probab. **19**, 801 (1987).

[16] L. Bachelier, Annales Scientifiques de l'École Normale Supérieure Ser.3 **17**, 21 (1900).

[17] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, 2000).

[18] J. Voit, *The Statistical Mechanics of Financial Markets* (Springer, Berlin 2001).

[19] M. Bogner, *Rekordstatistik in Finanzdaten*, unpublished thesis (2009).

[20] M. H. Jensen, A. Johansen, and I. Simonsen, Physica A **324**, 338 (2003).

[21] A. Johansen, I. Simonsen, and M. H. Jensen, Physica A **370**, 64 (2006).

[22] I. Simonsen, P. T. H. Ahlgren, M. H. Jensen, R. Donangelo, and K. Sneppen, Eur. Phys. J. B **57**, 153 (2007).

[23] K. Karpio, M. A. Załuska-Kotur, and A. Orłowski, Physica A **375**, 599 (2007).

[24] *Thomson Datastream Advance 4.0 SP4*, Thomson Reuters (2003).

[25] G. Wergen (unpublished).