Autoregressive processes with anomalous scaling behavior: Applications to high-frequency variations of a stock market index

Christian Dose,[1,2] Markus Porto,[3] and H. Eduardo Roman[4]
[1]Dipartimento di Fisica, Università di Milano, Via Celoria 16, 20133 Milano, Italy
[2]Dipartimento di Ingegneria Biofisica ed Elettronica, Università di Genova, Via Opera Pia 11a, 16145 Genova, Italy
[3]Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Straße 38, 01187 Dresden, Germany
[4]Dipartimento di Fisica and INFN, Università di Milano, Via Celoria 16, 20133 Milano, Italy
(Received 25 September 2002; revised manuscript received 20 February 2003; published 27 June 2003)

We employ autoregressive conditional heteroskedasticity processes to model the probability distribution function (PDF) of high-frequency relative variations of the Standard & Poors 500 market index data, obtained at the time horizon of 1 min. The model reproduces quantitatively the shape of the PDF, characterized by a Lévy-type power-law decay around its center, followed by a crossover to a faster decay at the tails. Furthermore, it is able to reproduce accurately the anomalous decay of the central part of the PDF at larger time horizons and, by the introduction of a short-range memory, also the crossover behavior of the corresponding standard deviations and the time scale of the exponentially decaying autocorrelation function of returns displayed by the empirical data.

DOI: 10.1103/PhysRevE.67.067103 PACS number(s): 02.50.Ey, 05.40.Fb, 87.23.Ge

I. INTRODUCTION

The problem of modeling stock price or market index variations on different time scales, ranging from minutes (high-frequency) to daily and longer time horizons (low-frequency), is still open and therefore attracting a great deal of interest of both the financial and statistical physics community (see, e.g., Ref. [1]). In this paper, we approach this problem in a phenomenological manner, based on the use of autoregressive conditional heteroskedasticity (ARCH) processes [2–8]. This work complements a recent study on low-frequency (daily) variations of the Dow Jones index [9] by extending it to the high-frequency variability of the Standard & Poors 500 (S&P500) index obtained at the time horizon of 1 min [10–14]. The probability distribution function (PDF) can be studied more accurately at high frequencies than at larger time horizons due to the larger amount of empirical data available at short time scales. Although several features of the model characterizing the low-frequency behavior remain valid at high frequencies, additional aspects need to be introduced for a better description of the index dynamics at short time scales (cf. also Ref. [15]).

The paper is organized as follows. In Sec. II, we briefly review the basic definitions of the ARCH processes and their main features required for the present study. In Sec. III, we discuss a simple ARCH process aimed at describing the PDF of the S&P500 index data, displaying a crossover behavior from a Lévy-type decay, with an exponent $\alpha \approx 1.47$ around the center of the distribution, to a faster decay at the tails. In Sec. IV, we revise the simple model of Sec. III by introducing a short-range memory in the artificial time series aimed at reproducing the observed exponential decay of the autocorrelation function of returns. Finally, Sec. V contains our concluding remarks.

II. ARCH(1) PROCESSES: LÉVY-TYPE AND OTHER POWER-LAW DECAYING PDF

For completeness, we briefly review here the main features of the ARCH processes discussed in detail in previous works [9,16,17], constituting the necessary background for the present study. In the standard ARCH(1) process, one generates a sequence of random numbers x_{n+1} drawn from a Gaussian distribution

$$W(x) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{x^2}{2\sigma_n^2}\right), \tag{1}$$

where the variance σ_n^2 depends on the value of the previous outcome x_n in the form

$$\sigma_n^2 = a + b x_n^2, \tag{2}$$

with $a>0$ and $b \geqslant 0$ being the model parameters. According to Eq. (2), the average variance is then given by $\sigma^2 = a/(1-b)$. It can be shown that the corresponding PDF $P(x)$ decays as a power law

$$P(x) \sim |x|^{-(1+\alpha)} \tag{3}$$

for $|x| \to \infty$, where α is related to the parameter b by the exact relation [16,18–20]

$$b^{-\alpha/2} = \frac{2^{\alpha/2}}{\sqrt{\pi}} \Gamma\left(\frac{1+\alpha}{2}\right). \tag{4}$$

Accordingly, $P(x)$ behaves asymptotically as the Lévy distribution when $0<\alpha<2$. It is easy to show [16] that within this interval, b varies in the range $3.6377\ldots > b > 1$. As it is apparent from Eq. (4), it is more convenient to consider b to be a function of α rather than the other way around. If we do this, we can obtain b as illustrated in Fig. 1 for α varying in the range $0 \leqslant \alpha \leqslant 4$, which is relevant to this work.

III. MODELING THE HIGH-FREQUENCY VARIATIONS OF THE S&P500 INDEX

Let us start discussing the PDF of the logarithmic variations of the S&P500 index, obtained at intervals of 1 min

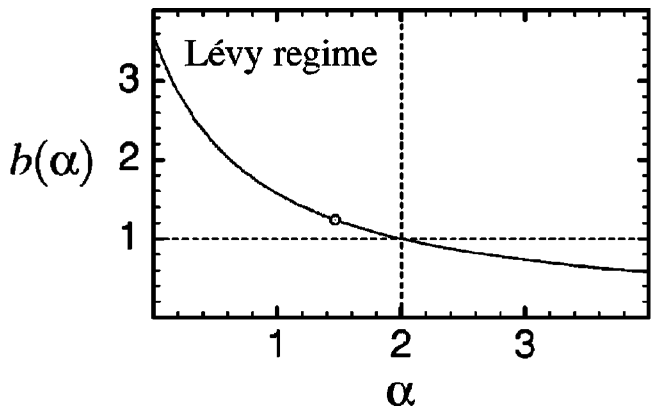

FIG. 1. The ARCH parameter b as a function of α, according to Eq. (4), in the range $0 \leqslant \alpha \leqslant 4$. The open circle denotes the values $\alpha_0 = 1.4685$ and $b_0 = 1.24$.

during the 12-year period from Jan. 1984 till Dec. 1995 [11,12]. The results for the corresponding PDF $P(x)$ are shown in Fig. 2 in a scaled form as a function of the scaled variable x/σ, where $\sigma \approx 0.07$. One clearly sees a power-law decay $P(x) \sim |x|^{-(1+\alpha)}$ at intermediate values $1 < x/\sigma < 5$, with $\alpha \approx 1.47$. For larger x, a faster decay is found, where, however, the limited number of data points does not allow for a precise determination of its actual shape.

In order to describe the empirical data in Fig. 2, we employ the ARCH model discussed in the preceding section. Clearly, we need to modify it to obtain both the power-law behavior at small $|x|$ followed by a faster decay at larger $|x|$.

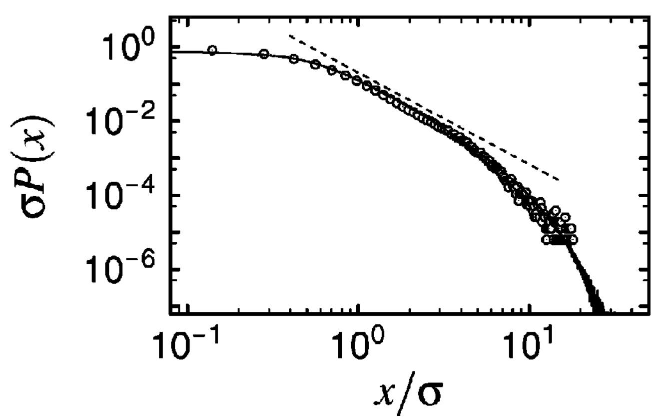

FIG. 2. Probability distribution function plotted as $\sigma P(x)$ vs x/σ for the logarithmic variations of the S&P500 index (open circles) taken every minute during the period Jan. 1984 to Dec. 1995 (from Ref. [11], where the corresponding standard deviation is $\sigma \approx 0.07$). The dashed straight line has slope -2.47 and is included as a guide. The continuous line represents the model calculations based on Eq. (5), with $a = 5.3 \times 10^{-4}$, $b = 1.24$, and $x_c = 0.7$ (resulting in a value $\sigma \approx 0.07$ as for the empirical data) obtained for a time series of 4×10^9 points. The asymptotic decay of the PDF for the present ARCH model is not known analytically, however our numerical results can be accurately fitted by the form $\sigma P(x) = 1/[1 + (y/0.45)^\gamma]$, with $\gamma = 2.365 + 0.0625y$ and $y \equiv x/\sigma$, yielding $\langle \gamma \rangle = 2.49$ in the interval $1 \leqslant y \leqslant 3$. This behavior remains to be understood. The bin width used for constructing the histogram was $0.001 \approx \sigma/70$.

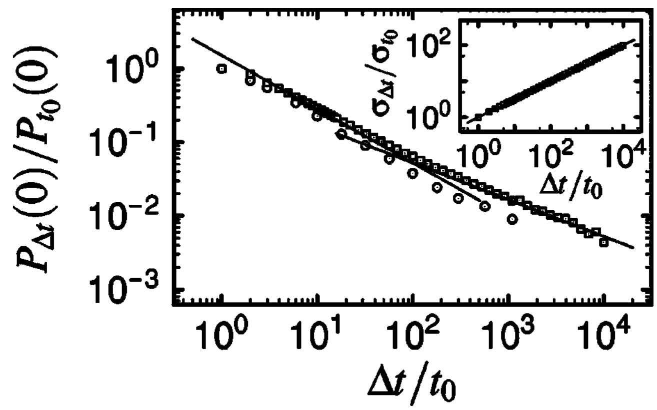

FIG. 3. Scaling behavior of the central part of the PDFs, $P_{\Delta t}(0)/P_{t_0}(0)$ (open squares) vs time horizon $\Delta t/t_0$, of the model, Eq. (5), for the parameters reported in Fig. 2. The straight lines have slopes -0.73 (top) and $-1/2$ (bottom), and are included as a guide. The open circles represent the S&P500 data [12] for $t_0 = 1$ min. The inset shows the model standard deviation $\sigma_{\Delta t}/\sigma_{t_0}$ vs $\Delta t/t_0$ (open squares), where $\sigma_{t_0} = \sigma \approx 0.07$ and the straight line has slope 1/2, indicating the absence of correlations.

To do this, we modify the definition of σ_n^2 in Eq. (2) according to

$$\sigma_n^2 = a + b e^{-|x_n|/x_c} x_n^2, \tag{5}$$

which is similar to Eq. (2) when $|x_n| \ll x_c$. In the case $|x_n| \gg x_c$, the factor $b \exp(-|x_n|/x_c) \equiv \tilde{b} \to 0$, yielding a hierarchy of increasingly faster power-law decays. This result can be understood qualitatively from Eq. (4) when $\tilde{b}$ is used in place of b, yielding $\alpha \to \infty$ when $\tilde{b} \to 0$ (cf. also Fig. 1).

The empirical data shown in Fig. 2 can be fitted accurately using the ARCH model based on Eq. (5). In particular, the parameter b can be estimated from the effective slope $\approx -(1+\alpha_0) \approx -2.47$ within the Lévy-type decay, i.e., $b = b_0 = 1.24$ (cf. also Fig. 1). The third parameter x_c can be obtained by adjusting the crossover of $P(x)$ observed around $x/\sigma \approx 5$, while a can be fixed so that the standard deviation becomes close to the empirical value $\sigma \approx 0.07$. The result of the model calculations is displayed by the continuous line in Fig. 2. Although the agreement with the S&P500 PDF is quite good, several features of the data at larger time horizons are not reproduced. To show this, we study next the *temporal aggregation* of our ARCH process in Eq. (5), defined as

$$X_n = \sum_{i=1}^{\Delta t/t_0} x_{n\Delta t/t_0 - (i-1)}, \tag{6}$$

where t_0 is the time unit used here, i.e., $t_0 = 1$ min, and calculate the corresponding PDF $P_{\Delta t}(X)$ for each time horizon Δt. To facilitate the comparison with the empirical data, we concentrate on the central part of the distributions ($X \approx 0$) by plotting the values of $P_{\Delta t}(0)$ in Fig. 3. (In random walk terminology, this corresponds to the probability of return to the origin.) The model based on Eq. (5) yields an initial

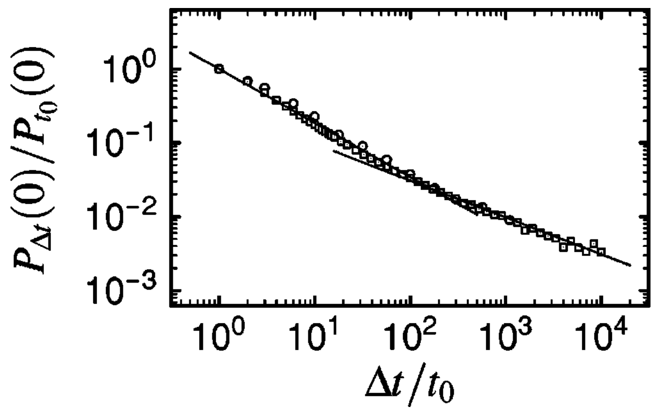

FIG. 4. Scaling behavior of the central part of the PDFs, $P_{\Delta t}(0)/P_{t_0}(0)$ (open squares) vs time horizon $\Delta t/t_0$ (as in Fig. 3) for the model based on the presence of short-range memory in the index increments, Eq. (7), with a short memory time $\tau_s = 1.68 t_0$ ($t_0 = 1$ min) and $n_{\tau_s} = 10$. The remaining parameters used were $a = 2.3\times 10^{-3}$, $b = 1.24$, and $x_c = 1.36$ (yielding again $\sigma \approx 0.07$). The straight lines have slopes -0.73 (top) and $-1/2$ (bottom), and are included as a guide. The open circles represent the S&P500 data [12] for $t_0 = 1$ min. The statistic is based on a time series of 4×10^9 points.

decay $P_{\Delta t}(0) \sim (\Delta t)^{-\beta}$, with $\beta \approx 0.7$, in qualitative agreement with the real market data, crossing over the standard decay $P_{\Delta t}(0) \sim (\Delta t)^{-1/2}$ at larger time horizons. A similar result has been obtained within a modified version of an ARCH process discussed in Ref. [17]. Note that $\beta \approx 1/\alpha_0$, as expected for a Lévy-type decay [21].

As shown in the inset of Fig. 3, however, the model standard deviation $\sigma_{\Delta t}$ behaves in the standard way as a function of the time horizon Δt, i.e., $\sigma_{\Delta t} \sim (\Delta t)^{1/2}$, in contrast to the S&P500 data [12] which shows an initial increase $\sigma_{\Delta t} \sim (\Delta t)^{\beta}$, with $\beta \approx 0.7$, for $\Delta t < 20$–30 min, and a standard behavior only at larger times (cf. Fig. 5 below). This is an indication that our simple model lacks of a short-range

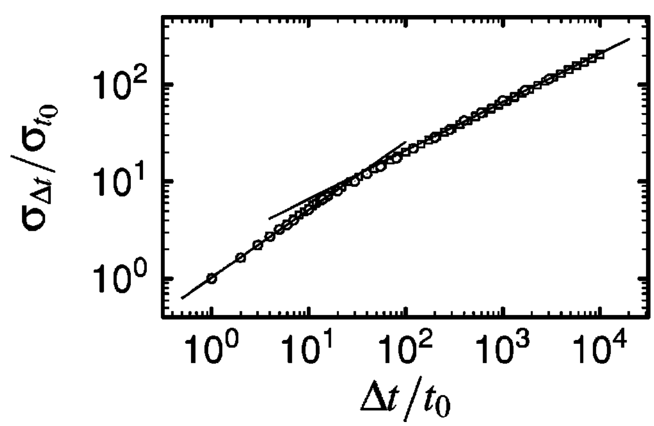

FIG. 5. Scaling behavior of the standard deviation $\sigma_{\Delta t}/\sigma_{t_0}$ (open squares) vs time horizon $\Delta t/t_0$, with $t_0 = 1$ min, for the values of the model parameters reported in Fig. 4. The straight lines have slopes 0.73 (bottom) and 1/2 (top), and are included as a guide. The crossover between the two regimes occurs at $\Delta t/t_0 \approx 25$, as for the empirical data (open circles) [12].

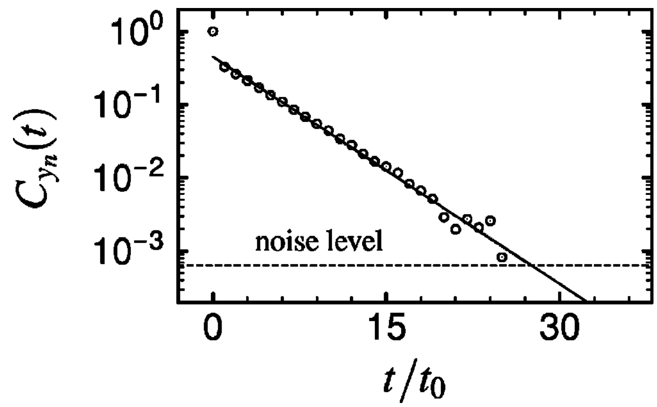

FIG. 6. Autocorrelation function of returns, $C_{y_n}(t)$ vs scaled time lag t/t_0 (open circles), for the values of the model parameters reported in Fig. 4. The continuous line is an exponential fit, $C_{y_n}(t) = 0.43\exp(-t/\tau_r)$, with $\tau_r = 4.2 t_0$ and $t_0 = 1$ min, consistent with the empirical value $\tau_r \approx 4$ min [10]. The horizontal dashed line represents the noise level.

memory. In what follows, we consider how to introduce this feature in a very simple way.

IV. ARCH PROCESS WITH MEMORY

A short-range memory is introduced by assuming that the actually observed increment, denoted next as y_n, is not just the "bare" value x_n generated by the process, Eq. (5), but rather by a linear combination of x_n and of the previous values $y_{n'}$ $(n' < n)$ of the form

$$y_n = \frac{1}{A}\left(x_n + \sum_{i=1}^{n_{\tau_s}} e^{-it_0/\tau_s} y_{n-i} \right) \quad \text{for} \quad n \geqslant 1, \tag{7}$$

where $A = \Sigma_{i=0}^{n_{\tau_s}} \exp(-it_0/\tau_s)$. We choose the initial values y_i for $i = 0, -1, -2, \ldots, 1 - n_{\tau_s}$, simply as $y_i = x_i$ for convenience. The value for n_{τ_s} is chosen sufficiently large such that the sum in Eq. (7) becomes independent of n_{τ_s}.

Using the newly generated time series y_n, we construct the corresponding temporal aggregations Y_n similarly as in Eq. (6), and obtain the associated PDF $P_{\Delta t}(Y)$. The resulting behavior of the central part of the PDF, $P_{\Delta t}(0)$, is shown in Fig. 4, where they display an initial power-law decay $P_{\Delta t}(0) \sim (\Delta t)^{-\beta}$, with $\beta \approx 0.73$, followed by a crossover to the standard decay at larger Δt, in better agreement with the real data. In addition, the crossover shape of the PDF, $P(y)$, behaves very much similar to the previous $P(x)$ shown in Fig. 2, and is therefore not shown here. The values for $\sigma_{\Delta t}$ now display the desired dependence on Δt, with a crossover at $\Delta t/t_0 \approx 25$ (cf. Fig. 5), in very good agreement with the empirical data [12]. The initial power-law exponent of the standard deviation 0.73 is consistent with the value of the short temporal decay exponent of $P_{\Delta t}(0)$.

To further assess the quality of the model predictions, we study the autocorrelation function of y_n defined as $C_{y_n}(t) = (\langle y_n y_{n+t}\rangle - \langle y_n\rangle^2)/\sigma^2$, which is displayed in Fig. 6. The results show an exponential decay with a characteristic time of about 4 min, similarly as for the empirical S&P500 data. It is interesting to observe that a rather small value for $\tau_s \approx 1.68$ min is required to get the right time decay of $C_{y_n}(t)$. (It should be noted that for the ARCH process without short-range memory, $C_{y_n}(t)=0$ for $t/t_0 \geqslant 1$.)

A similar behavior is displayed by the autocorrelation function of *absolute* returns, $C_{|y_n|}(t)$, i.e., an exponential decay with time in contrast to the slow power-law decay found for the real data [10], indicating that the model lacks a long-range memory. Modeling the slow long-range decay of the autocorrelation function of absolute returns is, however, outside the scope of the present work.

V. CONCLUSIONS

In conclusion, we have studied simple modifications of an ARCH process in order to describe the high-frequency behavior of the S&P500 market index for a time horizon of 1 min. The ARCH process we have introduced consists of two main features. The first one is an exponential cutoff within the fluctuating term of the ARCH variance, responsible for the crossover behavior observed for the PDF, and the second a short-range memory required to correctly describe the behavior of the PDF at larger time horizons. The latter reproduces the exponential decay (with a characteristic decay of ≈ 4 min) of the autocorrelation function of returns accurately. This simple scheme should be useful for modeling the high-frequency variations of stock prices, in particular, in the context of artificial many-asset markets where interesting applications can be envisaged.

ACKNOWLEDGMENTS

We would like to thank S. Cincotti, S. M. Focardi, and M. Raberto, from the University of Genoa for fruitful and stimulating discussions.

[1] *Application of Physics in Economic Modelling*, Proceedings of the NATO Advanced Research Workshop, Prague, 2001, edited by J.-P. Bouchaud, M. Marsili, B.M. Roehner, and F. Slanina [Physica A **299**, 1 (2001)].
[2] R.F. Engle, Econometrica **50**, 987 (1982).
[3] R.F. Engle and T.P. Bollerslev, Econometric Rev. **5**, 1 (1986).
[4] F.X. Diebold, *Empirical Modelling of Exchange Rate Dynamics* (Springer-Verlag, New York, 1988).
[5] T. Bollerslev, R. Engle, and D. Nelson, *ARCH Models* in *Handbook of Econometrics* (North-Holland, Amsterdam, 1993), Vol. IV.
[6] R. Engle, *ARCH, Selected Readings* (Oxford University, Oxford, 1995).
[7] C. Gouriéroux, *ARCH Models and Financial Applications*, Springer Series in Statistics (Springer, New York, 1997).
[8] T. Bollerslev, J. Econometrics **31**, 307 (1986).
[9] M. Porto and H.E. Roman, Phys. Rev. E **65**, 046149 (2002).
[10] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, and H.E. Stanley, Phys. Rev. E **60**, 1390 (1999).
[11] B. Podobnik, P.Ch. Ivanov, Y. Lee, A. Chessa, and H.E. Stanley, Europhys. Lett. **50**, 711 (2000).
[12] B. Podobnik, P.Ch. Ivanov, Y. Lee, and H.E. Stanley, Europhys. Lett. **52**, 491 (2000).
[13] H.E. Stanley, L.A.N. Amaral, X. Gabaix, P. Gopikrishnan, and V. Plerou, Physica A **299**, 1 (2001).
[14] P.Ch. Ivanov, B. Podobnik, Y. Lee, and H.E. Stanley, Physica A **299**, 154 (2001).
[15] F. Mainardi, M. Raberto, R. Gorenflo, and E. Scalas, Physica A **287**, 468 (2000).
[16] H.E. Roman and M. Porto, Phys. Rev. E **63**, 036128 (2001).
[17] H.E. Roman, M. Porto, and N. Giovanardi, Eur. Phys. J. B **21**, 155 (2001).
[18] L. de Haan, S.I. Resnick, H. Rootzen, and C.G. de Vries, Stochastic Proc. Appl. **32**, 213 (1989).
[19] D. Sornette, Physica A **250**, 295 (1998).
[20] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance* (Springer-Verlag, Heidelberg, 1999).
[21] R.N. Mantegna and H.E. Stanley, Nature (London) **376**, 46 (1995).