International Journal of Modern Physics B
Vol. 18, Nos. 4 & 5 (2004) 681–689

A STOCHASTIC MODEL FOR MULTIFRACTAL BEHAVIOR OF STOCK PRICES

UMBERTO L. FULCO

Departamento de Física, Universidade Federal do Piauí,
64049-550, Terezina-PI, Brazil

MARCELO L. LYRA

Departamento de Física, Universidade Federal de Alagoas,
Maceió–AL, 57072-970, Brazil

FILIPPO PETRONI and MAURIZIO SERVA

Dipartimento di Matematica and I.N.F.M.
Università dell'Aquila, I-67010 L'Aquila, Italy

GANDHI M. VISWANATHAN

Departamento de Física, Universidade Federal de Alagoas,
Maceió–AL, 57072-970, Brazil

Received 10 February 2004

We investigate the general problem of how to model the kinematics of stock prices without considering the dynamical causes of motion. We propose a Markovian stochastic process which is able to reproduce the experimentally observed volatility clustering and fat tails in the probability density functions (PDF) of financial time series. More importantly, the process also reproduces the PDF time scaling, the power law memory of volatility and the apparent multifractality of the time series up to the time scale which is experimentally observable.

1. Introduction

Remarkable progress has been made in quantitatively describing nonstationary and non-Gaussian phenomena, including those observed in economic[1] and social systems. The behavior of financial markets has recently[1–9] become a focus of interest to physicists as well as an area of active research because of its rich and complex[10–19] dynamics. The daily returns r_t for a given stock can be defined as

$$r_t = \log_{10} \frac{S_{t+1}}{S_t} ,$$

at time t (Fig. 1), where S_t is the price of a an asset or the value of an index or a currency exchange rate. In this paper we address the important yet unsolved

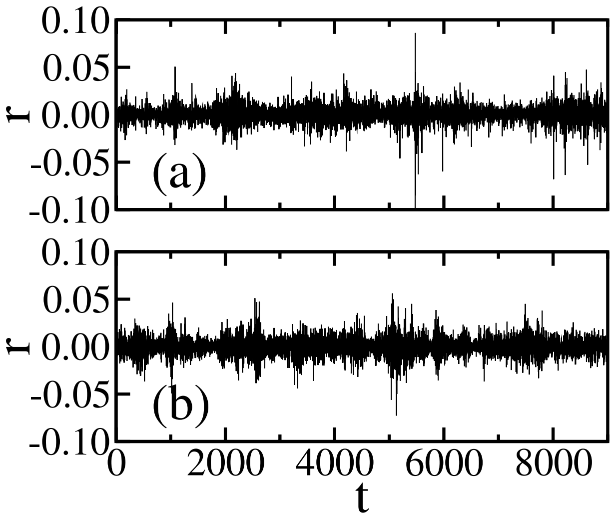


Fig. 1. Real (a) and artificial data (b) obtained from the maps in Eqs. 1 and 2 with constants $a = -0.1$, $b = 0.1$, $d = 0.98$ and $T = 10$. The striking resemblance between the time series generated by the the kinematic model and the real data is further verified by quantitative analysis, as explained in the text.

problem of how to model the kinematics of stock prices without handling the problem of the causes of motion. Our point of departure is the family of ARCH-GARCH models. These families of stochastic kinematic models neglect the causes underlying the price variations and focus only on the equations of motion governing the fluctuating returns. Specifically, the interacting buying and selling processes are not considered. ARCH models, introduced by Engle in 1982,[20] are stochastic Auto-Regressive Conditional Heteroscedasticity models, characterized by zero mean and nonconstant variances dependent on the past. Such models can simultaneously have global stationarity as well as local nonstationary behavior. The simplest ARCH model of stock returns can be defined by

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 \,,$$

with returns defined as

$$r_t = \sigma_t \omega_t \,,$$

where the random Gaussian variable ω_t has unitary variance and zero mean, while α_0 and α_1 are tunable parameters. Hence, the returns r_t are Gaussian distributed with zero mean and local variance σ_t^2. Only very recently has the physics community begun to study such models, although they are common in the economics literature. In more general ARCH processes, the local variance can depend not only on the previous value r_{t-1}, but on any finite number n of previous values $r_{t-1} \cdots r_{t-n}$. GARCH models are a further generalization in which the local variance can depend not only on previous values of the returns but also on the previous values of the locally measured variances.

Although ARCH–GARCH and generalizations are Markovian, they have succeeded in capturing important features (such as volatility clustering) of real financial data. They are, therefore, widely used, both in the financial community and in econophysics to model the kinematics of price variations. Nevertheless, several characteristic features of the real data are not well described. The three most important (and difficult to model) features not well accounted in ARCH–GARCH models are (i) the time scaling of the probability density distribution of the returns, (ii) the known long-range power law correlations in the volatility (the latter being a measure of the local standard deviation of the returns), and (iii) the apparent multifractality of the returns, i.e. volatility power law correlation exponents are non-unique and depend on the magnitude of the events considered. As an alternative model, Lévy and truncated Lévy distributions (see Ref. 21) have been proposed[22] to fit the observed fat tailed distributions. While Lévy processes can approximately reproduce the time scaling of distributions, they cannot explain one of the most relevant and characteristic phenomena: The multifractal long-range correlations in volatility which has been found to be responsible for clustering of volatility and the persistence of fat tails for long time lags.[6,23]

Here we propose a new model that does not suffer from these drawbacks. In order to test the model, we compare the model with a typical dataset: The New York Exchange (NYSE) daily composite index price closes from January 1966 to September 2001 (a total of some 9000 data points).

2. The Model

The model is inspired in the recent finding that the probability density distribution of the local variance σ_t is similar to a log-normal.[4] Other key findings include the long-range correlations found in the absolute value of the returns and the multifractal behavior of the returns.[7] Our proposal, based on these findings, is the following map for the evolution of the volatility:

$$\sigma_{t+1} = e^{[a+b\omega_{t+1}]}(\overline{\sigma_t})^d \,, \tag{1}$$

with returns defined as

$$r_t = \sigma_t \tilde{\omega}_t \,, \tag{2}$$

where $\tilde{\omega_t}$ and ω_t are independently distributed unitary Gaussian variables with zero mean that are also independent from each other. a, b and d are tunable constants. The term $\overline{\sigma_t}$ represents the average with respect to last T days, from $t-T$ to t. Since the average involves a small number T of days (from two to ten) the process, like the generalized ARCH–GARCH process, is Markovian. The interesting fact is that T is extremely small compared with the number of days in which the volatility of the model remains correlated (hundreds of days). We also remark that we have used a simple unweighted average but, in principle, we could have used any type of moving average, for instance a power law or exponentially weighted moving average.

The tunable constants a, b and d are related to the nearly log-normal form of σ_t. The value of d must be close but smaller than 1 to guaranty the stationarity of the returns; a is related to the typical size of the daily volatility, and b to the typical size of its fluctuations. Finally, note that the mutual independence of the two Gaussian variables contrasts with ARCH models, for which they coincide. We discuss why this is preferable below.

Using this map, we generate a data set of 9000 returns (Fig. 1(b)) with the following choice of constants: $a = -0.1$, $b = 0.1$, $d = 0.98$ and $T = 10$. We then compare the statistical and scaling properties of the time series generated with the real data set. Note that the above model can reproduce very well the clustering of volatility, as is evident comparing Fig. 1(a) with Fig. 1(b). Nevertheless, we proceed with a quantitative comparison of both the scaling of the distribution and the multifractal power law correlations.

3. Distribution of Returns and Time Scaling

We next study the probability density of returns for the proposed model. We estimate the probability density of returns using

$$P(r) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}\Delta} \exp\left\{ \frac{-(r - r_i)^2}{2\Delta^2} \right\}$$

with a smoothing window of $\Delta = 0.001$ (Fig. 2(a)). The r_i can be either the real or artificial data and N is their number (about 9000 for both). The dashed and dotted lines in Fig. 2(a) show the distributions of daily (one business day) returns, for both real and artificial data. Typical of financial times series is their invariance under rescaling of time. Therefore, we estimate with the same smoothing window the distribution for monthly (25 business days) returns (Fig. 2(b))

$$\frac{\sum_{i=1}^{25} r_{i+t}}{(25)^\delta} = \frac{1}{(25)^\delta} \log \frac{S_{t+25}}{S_t} ,$$

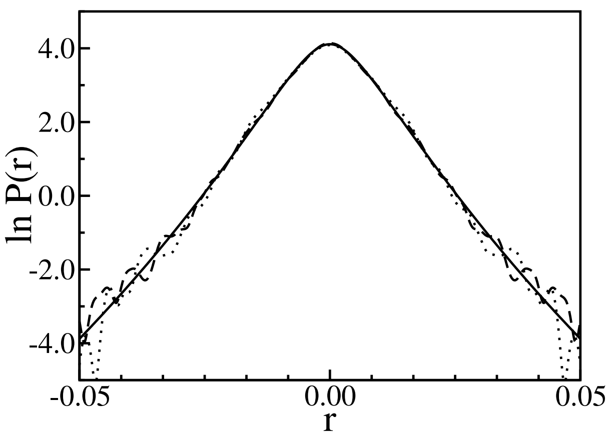
using the measured value $\delta = 0.53$ of the scaling exponent both for real and artificial data. Note that the solid line curve (see discussion below) is exactly the same in the two figures, therefore both real and artificial data exhibit almost perfect time scale invariance of the return distributions (see also Ref. 23). The lack of agreement for small monthly returns is due to short-range (1 day) correlations in the signs of the returns that are neglected in our model, but that could easily be incorporated.[23] Compared to other models of stock returns, this model shows remarkably good agreement with real data.

We next compare these probability distributions of real and artificial data with a "theoretical" distribution (solid line in Figs. 2), which we introduce better to understand the form of the real and artificial data PDF. The proposed "theoretical' distribution is a Gaussian P_{t} with log-normally distributed local variance. We can find P_{t} by convolving Gaussians of varying widths:

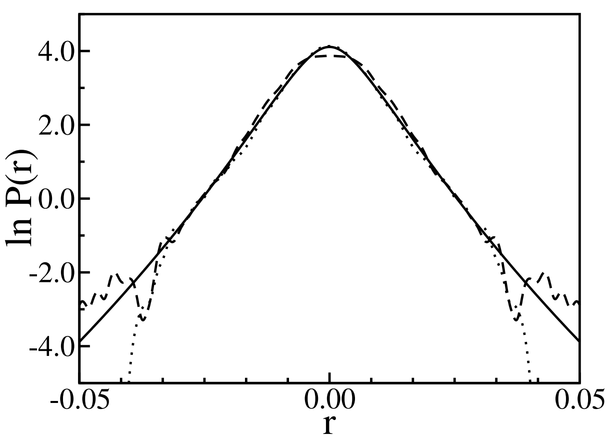
$$P_{\mathrm{t}}(r) = \int \rho(\sigma) \frac{e^{\frac{-r^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} d\sigma \tag{3}$$

$$\rho(\sigma) = \frac{1}{\sqrt{2\pi} s\sigma} e^{\frac{-1}{2}\left(\frac{\log\sigma - m}{s}\right)^2} \tag{4}$$

We have used the empirically found values $s = 0.41$ and $m = -0.34$. The solid lines in Figs. 2 leave no doubt that this distribution is one of the best candidates—if not

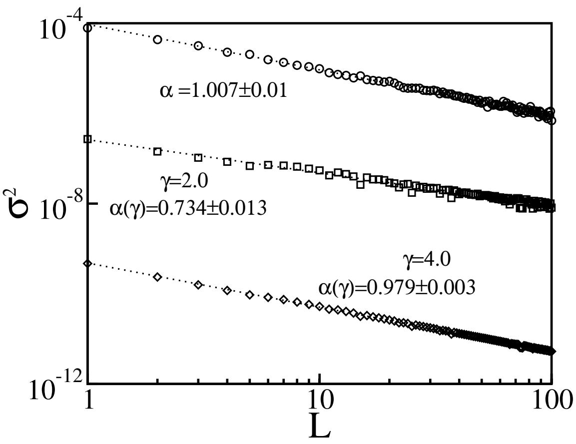


(a)

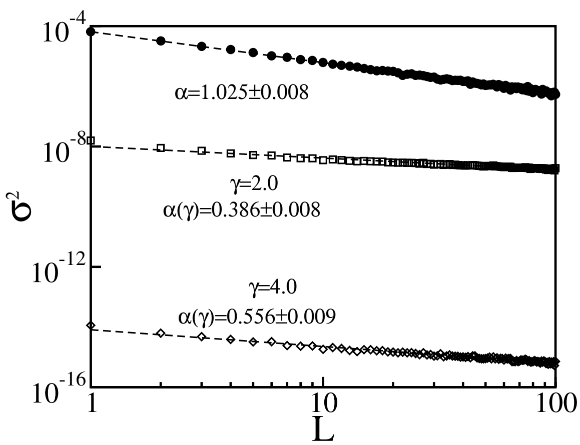


(b)

Fig. 2. Symmetrized probability density distribution $P(r)$ of the returns $r_\tau$measured over lags (a) $\tau = 1$ d and (b) $\tau = 25$ d (about one business month), shown for the artificial data (dashed line), the real data (long dashed) and compared with the "theoretical" distribution $P_t(r)$ (solid line)). For $\tau = 1$ d all three distributions are almost identical while for $\tau = 25$ a small discrepancy appears for real data for small returns. This discrepancy is due to short range correlations in real data not included in the model, as explained in the text.



(a)



(b)

Fig. 3. Variance $\sigma^2(\phi(L,\gamma)) \sim L^{-\alpha(\gamma)}$ of the generalized cumulative absolute returns as a function of L on double log scales for (a) the NYSE index and (b) the proposed kinematic model. Data for the absolute moments with $\gamma = 2$ (square) and $\gamma = 4$ (diamond) are compared with the variance $\sigma^2(\phi(L)) \sim L^{-\alpha}$ of the cumulative returns (circles). Both data and model clearly show multifractal behavior, since there is no unique scaling exponent. The exponents of the best fitting straight lines (dashed lines) are, respectively: $\alpha(2) = 0.734 \pm 0.013$, $\alpha(4) = 0.979 \pm 0.003$ and $\alpha = 1.007 \pm 0.01$ for the NYSE index; $\alpha(2) = 0.386 \pm 0.008$, $\alpha(4) = 0.556 \pm 0.009$ and $\alpha = 1.025 \pm 0.008$ for the model. The ability of this kinematic model to generate multifractal behavior distinguishes it from the well-known families of models.

the best—for describing the distribution of price variations. The agreement found is exceptionally strong.

4. Volatility Correlations and Multifractality

It is known that daily returns have no auto-correlations for lags larger than a single day, consistent with the efficient market hypothesis. This fact can be also checked by using Detrended Fluctuation Analysis (DFA)[24] and related methods. Consider the cumulative returns $\phi_t(L)$, defined as the sum of L successive returns divided by L:

$$\phi_t(L) = \frac{1}{L}\sum_{i=1}^{L} r_{t+i}\,. \tag{5}$$

One can define N/L non overlapping variables of this type, and compute the associated variance $\sigma^2(\phi(L))$, where $N \approx 9000$ is the number of data points. According to the central limit theorem, uncorrelated (or short-range correlated) r_t would lead to power-law behavior:

$$\sigma^2(\phi(L)) \sim L^{-\alpha}\,, \tag{6}$$

with exponent $\alpha = 1$ for large L. The exponent α both for the NYSE index and the model proposed here is near 1 (see Fig. 3), confirming that returns are uncorrelated.

The lack of correlations does not hold true for quantities related to the *absolute* returns. In order to perform the appropriate scaling analysis, we introduce the generalized cumulative absolute returns defined as the sum of L successive absolute return powers $|r_t|^\gamma, \ldots, |r_{t+L-1}|^\gamma$, divided by L

$$\phi_t(L,\gamma) = \frac{1}{L}\sum_{i=1}^{L} |r_{t+i}|^\gamma \tag{7}$$

where γ is a real exponent (noting that these quantities do not overlap.) Using this method, if the autocorrelation for powers of absolute returns exhibits a power-law with exponent $\alpha(\gamma) \leq 1$ for large L, then we expect

$$\sigma^2(\phi(L,\gamma)) \sim L^{-\alpha(\gamma)}\,. \tag{8}$$

(Note, however, that if the $|r_t|^\gamma$ are short-range correlated or power-law correlated with an exponent $\alpha(\gamma) > 1$, then we would not detect anomalous scaling in the analysis of variance, because it is not possible to detect $\alpha > 1$ using the method discussed here.)

We find that both the real data and the model show similar apparent multifractal behavior with non-unique scaling exponents, i.e. $\alpha(\gamma) \neq$ constant. We employ some caution and we prefer to speak of "apparent" multifractality instead of real multifractality because the involved time scales are never much longer then one or two hundred days.

Note that the function $\alpha(\gamma)$ is not universal[4,7] but depends on the particular asset considered. Therefore, it is not surprising that the values $\alpha(\gamma)$ we compute do not coincide for the real and artificial data. Indeed, this finding is somewhat expected if one takes into account the non-universality. For completeness, we note that in order to obtain full agreement for the multifractal exponents of any specific economic series one can generalize the volatility as follows:

$$x_{t+1} = e^{a+b\omega_{t+1}}(\overline{x_t^c})^d\,. \tag{9}$$

where the exponent c in the volatility average would allow for a finer control of the effects caused by extremal events. Also the resulting returns could be written as

$$r_t = (\alpha + x_t) \cdot \tilde{\omega}_t \equiv \sigma_t \tilde{\omega}_t \,. \tag{10}$$

with α representing an intrinsic constant contribution to daily volatility.[25]

5. Discussion and Conclusion

We now comment on the motivation for using two independent Gaussian variables, one for returns and one for volatility. Indeed, this choice separates the temporal evolution of volatility and returns. Besides giving reasonable agreement with experimental data, there is a deeper underlying motivation for this approach. In ARCH–GARCH models the present-day volatility depends on the previous day's absolute return. It is commonly accepted that today's sentiment about volatility does not depend on the previous day's variation of price but rather on other quantities, such as previous day intraday volatility, implicit volatility in derivative products and public (or less public) information. Therefore, the previous day's absolute return (which can be small after a fearful market day with enormous variation of prices) does not directly influence the present day volatility and so the evolution of the latter should be kept separate, exactly as we have done in the model proposed.

In conclusion, the proposed stochastic multifractal model of long-range correlated absolute returns is able to reproduce features of real financial data that are not well accounted for by existing models. Specifically, our model is able to overcome the inability of ARCH–GARCH and Lévy models to adequately explain (i) the fat tailed distribution of the returns and its time scaling, (ii) the long-range power law volatility correlations, or (iii) the apparent multifractality of the returns. We note that these results are obtained by means of the proposed Markov process and that the volatility correlations are not in the input, but are generated by the model. The model is thus unique and differs from others (see for example Ref. 26) were correlations are already present in the input ingredients, i.e. it is assumed *a priori* that increments are long-range correlated. The model presented here also represents an advance due to the exceptionally improved agreement with real data. We hope that this advance in modeling the kinematics of financial time series further contributes to the emerging study of econophysics and towards a better understanding of wider classes of complex systems.

Acknowledgment

One of us (M.S.) would like to warmly thank Francesco Guerra for his cultural and practical help in the early stage of his carrier. He especially acknowledges that his interest in stochastic processes originated from the extraordinary work of Francesco on Stochastic Mechanics and that all his research was influenced by his papers. We thank Michele Pasquini for useful discussions and for a critical reading of the manuscript. This work was partially supported by the Brazilian research agencies

CAPES and CNPq and by the Alagoas State agency FAPEAL, F.P. acknowledges the financial support of Cofin MIUR 2002 prot. 2002027798_005.

References

1. R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000).
2. J.-P. Bouchaud, A. Matacz and M. Potters, *Phys. Rev. Lett.* **87**, 228701 (2001).
3. B. Podobnik, P. Ch. Ivanov, Y. Lee, A. Chessa and H. E. Stanley, *Europhys. Lett.* **50**(6), 711 (2000).
4. M. Pasquini and M. Serva, *Euro. Phys. J.* **B16**, 195 (2000).
5. Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng and H. E. Stanley, *Phys. Rev.* **E60**, 1390 (1999).
6. P. Gopikrishnan, V. Plerou, L. A. NunesAmaral, M. Meyer and H. E. Stanley, *Phys. Rev.* **E60**, 5305 (1999).
7. M. Pasquini and M. Serva, *Econ. Lett.* **65**, 275 (1999).
8. R. N. Mantegna and H. E. Stanley, *Nature* **383**, 587 (1996).
9. R. N. Mantegna and H. E. Stanley, *Nature* **376**, 46 (1995).
10. L. Bachelier, *Ann. Sci. École Norm. Sup.* **17**, 21 (1900).
11. Z. Ding, C. W. J. Granger and R. F. Engle, *J. Empirical Finance* **1**, 83 (1993).
12. N. Crato and P. De Lima, *Economics Letters* **45**, 281 (1994).
13. J.-P. Bouchaud and D. Sornette, *J. Phys. I France* **4**, 863 (1994).
14. M. Mevy and S. Solomon, *Int. J. Mod. Phys.* **C7**, 595 (1996).
15. R. T. Baillie, *J. Econometrics* **73**, 5 (1996).
16. S. Gallucio, G. Caldarelli, M. Marsili and Y.-C. Zhang, *Physica* **A245**, 423 (1997).
17. N. Vandewalle and M. Ausloos, *Physica* **A246**, 454 (1997).
18. Y. Liu, P. Cizeau, M. Meyer, C.-K. Peng and H. E. Stanley, *Physica* **A245**, 437 (1997).
19. B. Podobnik, K. Matia, A. Chessa, P. Ch. Ivanov, Y. Lee and H. E. Stanley, *Physica* **A300**, 300 (2001).
20. R. F. Engle, *Econometrica* **50**, 987 (1982).
21. R. N. Mantegna and H. E. Stanley, *Phys. Rev. Lett.* **73**, 2946 (1994).
22. B. B. Mandelbrot, *J. Bus.* **36**, 394 (1963).
23. G. M. Viswanathan, U. L. Fulco, M. L. Lyra and M. Serva, *Physica A* (in press) (2003).
24. C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley and A. L. Goldberger, *Phys. Rev.* **E49**, 1685 (1994).
25. U. L. Fulco, M. L. Lyra, F. Petroni, M. Serva and G. M. Viswanathan, *Physica* **A329**, 273 (2003).
26. E. Bacry, J. Delour and J. F. Muzy, *Phys. Rev.* **E64**, 026103 (2001).