PHYSICAL REVIEW E **75**, 060102(R) (2007)

Markov analysis and Kramers-Moyal expansion of nonstationary stochastic processes with application to the fluctuations in the oil price

Fatemeh Ghasemi,[1] Muhammad Sahimi,[2,*] J. Peinke,[3] R. Friedrich,[4] G. Reza Jafari,[5] and M. Reza Rahimi Tabar[3,6,7,†]
[1]*The Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, D-01187 Dresden, Germany*
[2]*Mork Family Department of Chemical Engineering & Materials Science, University of Southern California, Los Angeles, California 90089-1211, USA*
[3]*Carl von Ossietzky University, Institute of Physics, D-26111 Oldenburg, Germany*
[4]*Institute for Theoretical Physics, University of Münster, D-48149 Münster, Germany*
[5]*Department of Physics, Shahid Beheshti University, Evin, Tehran 19839, Iran*
[6]*Department of Physics, Sharif University of Technology, Tehran 11365, Iran*
[7]*CNRS UMR 6529, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France*
(Received 10 January 2007; revised manuscript received 9 May 2007; published 18 June 2007)

We describe a general method for analyzing a nonstationary stochastic process $X(t)$ which, unlike many of the previous analysis methods, does not require $X(t)$ to have any scaling feature. The method is used to study the fluctuations in the daily price of oil. It is shown that the *returns* time series, $y(t)=\ln[X(t+1)/X(t)]$, is a stationary and Markov process, characterized by a Markov time scale t_M. The coefficients of the Kramers-Moyal expansion for the probability density function $P(y,t|y_0,t_0)$ are computed. $P(y,t|,y_0,t_0)$ satisfies a Fokker-Planck equation, which is equivalent to a Langevin equation for $y(t)$ that provides *quantitative* predictions for the oil price over times that are of the order of t_M. Also studied is the average frequency of positive-slope crossings, $\nu_\alpha^+=P(y_i>\alpha, y_{i-1}<\alpha)$, for the returns, where $T(\alpha)=1/\nu_\alpha^+$ is the average waiting time for observing $y(t)=\alpha$ again.

DOI: 10.1103/PhysRevE.75.060102 PACS number(s): 05.40.−a, 05.10.Gg, 05.45.Tp

Characterizing nonstationary stochastic processes has been a problem of fundamental interest for a long time. Examples of such processes include various indicators of economic activity [1], fluctuations in the porosity and permeability of porous media [2], velocity fluctuations in turbulent flows, and heartbeat dynamics [3]. We propose in this Rapid Communication a general method for (i) generating a stationary process $y(t)$, given a nonstationary one, $X(t)$; (ii) analyzing the statistical properties of $y(t)$; and (iii) constructing stochastic continuum equations that not only reconstruct $y(t)$ [and hence $X(t)$], but also provide *quantitative* predictions for it over a certain time scale that we identify below.

Given $X(t)$, one may be able to construct a stationary process $y(t)$ by at least one of the two following methods. (i) Constructing the *algebraic increments*, $y(t)=X(t+1)-X(t)$. The best-known example of such processes is the fractional Brownian motion (FBM) with a power spectrum, $S(f) \propto 1/f^{2H+1}$, where H is the Hurst exponent. It is well known that the FBM's increments [with $S(f)\propto 1/f^{2H-1}$] are stationary. Moreover, when $H=1/2$, the increments are uncorrelated, while for $H=-1/2$ $X(t)$ itself becomes random. (ii) Let $Z=\ln X(t)$. Then, one may construct $y(t)$, by $y(t)=Z(t+1)-Z(t)=\ln[X(t+1)/X(t)]$, so that $y(t)$ represents the *logarithmic increments*.

We then analyze $y(t)$ based on the application of Markov processes and development of a Langevin equation for it. As a concrete example, we analyze the fluctuations in oil's daily price, a most notorious nonstationary process, and show that they fall in the class of nonstationary processes, the logarithmic increments of which are stationary. The method is, however, general and applicable to a large class of nonstationary processes.

Figure 1 presents the fluctuations in oil's daily price, $X(t)$ [4]. It is not difficult to show that $X(t)$ is not stationary by showing, for example, that its variance computed in a window is not stable if we increase the windows size or move it. Hence we construct the logarithmic increments, or the *log-returns*, of $X(t)$ by $y(t)=\ln[X(t+1)/X(t)]$; see Fig. 2. It is now straightforward to show that $y(t)$ is stationary using three different methods. We computed its average and variance in moving windows of increasing sizes to check that they are stable. We then computed the spectral density $S(f)$ of $y(t)$. The result, $S(f)\propto f^\beta$ with $\beta\simeq 0$, indicated the absence of long-range correlations in $y(t)$. We also analyzed $y(t)$ us-

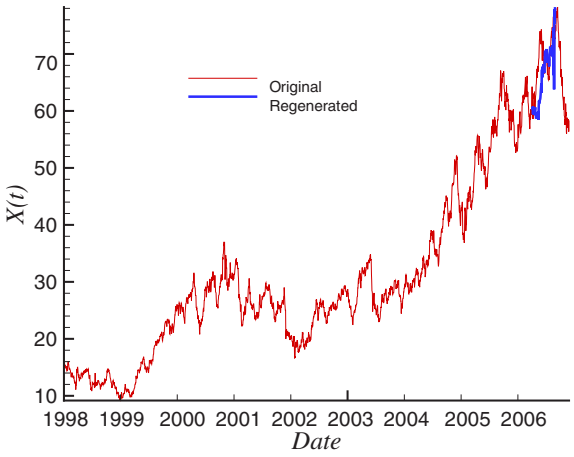

FIG. 1. (Color online) The daily oil price [4] (in $). The time lag is 1 day. Shown is a sample of the actual daily oil prices (red) and the reconstructed data (blue), using Eq. (8).

*Electronic address: moe@iran.usc.edu
†Electronic address: rahimitabar@gmail.com

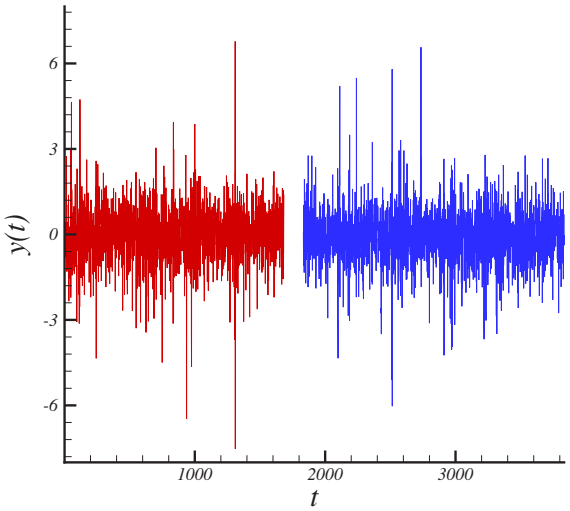

FIG. 2. (Color online) Comparison of the actual return data (red, left) and the reconstructed ones using the Langevin equation (blue, right). For clarity, the time series have been shifted on the t axis.

ing the detrended fluctuation analysis and the rescaled-range method, to further check that $y(t)$ is stationary [5–7]. They both yielded $\beta \simeq 0$ and thus $y(t)$ is, at least to a good degree of approximation, stationary.

Since long-range correlations are absent in $y(t)$, but short-range correlations may exist, we check whether $y(t)$ follows a Markov chain [8–11], in which case we estimate its Markov time scale t_M—the minimum time interval over which $y(t)$ can be approximated by a Markov process. Characterizing the statistical properties of $y(t)$ requires evaluation of the joint probability density function (PDF) $P_n(y_1,t_1;\ldots;y_n,t_n)$ for an arbitrary n, the number of the data points. If, however, $y(t)$ is a Markov process, the n-point joint PDF P_n is the product of the conditional probabilities $P(y_{i+1},t_{i+1}|y_i,t_i)$, for $i=1,\ldots,n-1$. A necessary condition for $y(t)$ to be a Markov process is that the Chapman-Kolmogorov (CK) equation [12],

$$P(y_2,t_2|y_1,t_1) = \int dy_3\, P(y_2,t_2|y_3,t_3)P(y_3,t_3|y_1,t_1), \quad (1)$$

should hold for any t_3 in $t_1<t_3<t_2$. The validity of the CK equation for different values of y_1 is checked by comparing the directly evaluated $P(y_2,t_2|y_1,t_1)$ with those calculated according to right side of Eq. (1). Note that the opposite is not necessarily true, namely, that if a stochastic process satisfies the CK equation, it is not necessarily Markovian [13].

To estimate t_M, we used the least-squares method. If $y(t)$ is a Markov process, one has

$$P(y_3,t_3|y_2,t_2;y_1,t_1) = P(y_3,t_3|y_2,t_2). \quad (2)$$

Thus we compare the PDF, $P(y_3,t_3;y_2,t_2;y_1,t_1) = P(y_3,t_3|y_2,t_2;y_1,t_1)P(y_2,t_2;y_1,t_1)$, with that obtained based on the Markov process. Using the properties of the Markov process and substituting in Eq. (2), we obtain

$$P_{\mathrm{M}}(y_3,t_3;y_2,t_2;y_1,t_1) = P(y_3,t_3|y_2,t_2)P(y_2,t_2;y_1,t_1). \quad (3)$$

[Note that the stationarity of a stochastic process is not necessary for using Eqs. (2) and (3)]. To check whether $y(t)$ is a Markov process, we must compute the three-point joint PDF through Eq. (2) and compare the result with that obtained through Eq. (3). To do so, we first determine the quality of the fit through computing the least-squares fitting quantity χ^2, defined by

$$\chi^2 = \int dy_3\, dy_2\, dy_1 [P(y_3,t_3;y_2,t_2;y_1,t_1) - P_{\mathrm{M}}(y_3,t_3;y_2,t_2;y_1,t_1)]^2/(\sigma_{3j}^2+\sigma_{\mathrm{M}}^2), \quad (4)$$

where σ_{3j}^2 and σ_{M}^2 are the variances of $P(y_3,t_3;y_2,t_2;y_1,t_1)$ and $P_{\mathrm{M}}(y_3,t_3;y_2,t_2;y_1,t_1)$, respectively. To estimate t_M, we used the likelihood statistical analysis [14]. In the absence of a prior constraint, the probability of the set of three-point joint PDFs is given by

$$P(t_3-t_1) = \Pi_{y_3,y_2,y_1} \frac{1}{\sqrt{2\pi(\sigma_{3j}^2+\sigma_{\mathrm{M}}^2)}} \exp\left\{\frac{[P(y_3,t_3;y_2,t_2;y_1,t_1)-P_{\mathrm{M}}(y_3,t_3;y_2,t_2;y_1,t_1)]^2}{2(\sigma_{3j}^2+\sigma_{\mathrm{M}}^2)}\right\}. \quad (5)$$

$P(x)$ must be normalized. Evidently, when, for a set of the parameters, $\chi_\nu^2=\chi^2/N$ is minimum (N is the degree of freedom), the probability is maximum. Figure 3 presents χ_ν^2; its minimum is ≈ 0.6, corresponding to $t_M=t_3-t_1\simeq 1$ day. Figure 4 shows the likelihood function of t_M. Here, we used the χ^2 test to estimate t_M, and also used the method proposed in Refs. [6–11,15–18] which enables us to estimate t_M via a direct check of the CK equation. The result is, again, $t_M \simeq 1$ day.

For a Markov process, knowledge of $P(y_2,t_2|y_1,t_1)$ is sufficient for generating the entire statistics of $y(t)$, encoded in the n-point PDF which satisfies a master equation which, in turn, is reformulated by a Kramers-Moyal (KM) expansion:

$$\frac{\partial}{\partial t}P(y,t|y_0,t_0) = \sum_k \left(-\frac{\partial}{\partial y}\right)^k [D^{(k)}(y,t)P(y,t|y_0,t_0)]. \quad (6)$$

The KM coefficients $D^{(k)}(y,t)$ are given by

$$D^{(k)}(y,t) = \frac{1}{k!}\lim_{\Delta t\to 0} M^{(k)},$$

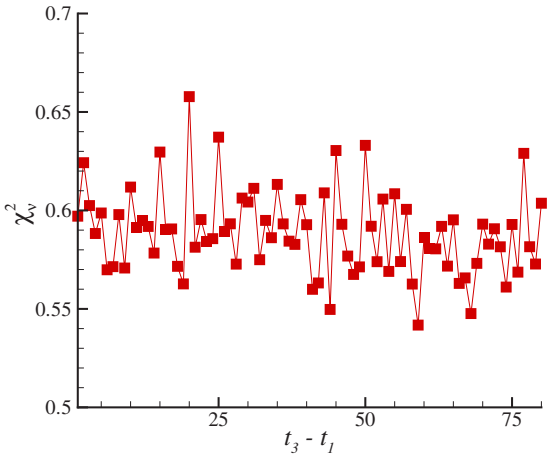

FIG. 3. (Color online) The χ^2 test for estimation of t_M.

$$M^{(k)} = \frac{1}{\Delta t}\int dy'(y'-y)^k P(y',t+\Delta t|y,t). \tag{7}$$

For a general stochastic process, all the KM coefficients may be nonzero. However, provided that $D^{(4)}$ vanishes or is small compared to the first two coefficients [12], truncation of the KM expansion after the second term is meaningful in the statistical sense. For the oil data, $D^{(4)} \simeq 10^{-2} D^{(2)}$, where $y(t)$ is measured in units of its maximum, $y_{\max}$. Thus we truncate the KM expansion after the second term, reducing it to a Fokker-Planck (FP) equation. According to the Ito calculus [12,19], the FP equation is equivalent to a Langevin equation,

$$\frac{d}{dt}y(t) = D^{(1)}(y) + \sqrt{D^{(2)}(y)}f(t), \tag{8}$$

where $f(t)$ is a random "force" with zero mean and Gaussian statistics, δ-correlated in t, i.e., $\langle f(t)f(t')\rangle = 2\delta(t-t')$.

Furthermore, Eq. (8) enables us to reconstruct a time series for $y(t)$ which is similar to the original one *in the statistical sense*. In Fig. 2 the original and reconstructed $y(t)$ are both shown. We find that $D^{(1)}$ and $D^{(2)}$, estimated directly from the data, are well represented by the approximants,

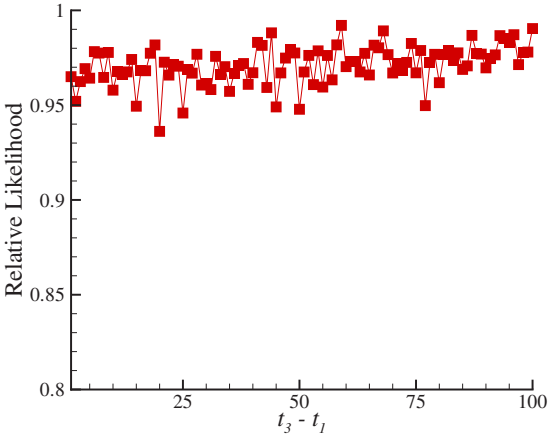

FIG. 4. (Color online) Relative likelihood function for t_M.

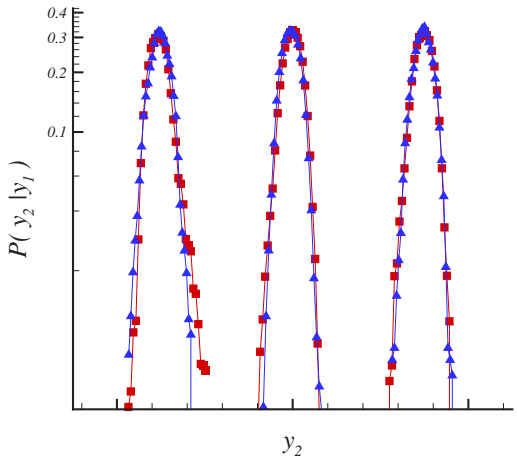

FIG. 5. (Color online) Comparison of the directly evaluated PDFs using the actual data (squares), and the PDFs obtained from Eq. (10) (triangles). Values for y_1, from left to right, are −0.1, 0.0, and 0.1 [measured in units of $y_{\max}(t)$]. For better presentation, the PDFs have been shifted on the horizontal axis.

$$D^{(1)}(y) = -1.09y,$$

$$D^{(2)}(y) = 0.0033 - 0.003y + 0.716y^2, \tag{9}$$

but the estimates become relatively inaccurate for large y and thus the uncertainty in them increases.

We now evaluate the precision of the reconstructed $y(t)$, by computing the conditional PDF through the numerical solution of the FP equation, which is very sensitive to the numerical errors in $D^{(1)}$ and $D^{(2)}$ [6–11,19–22]. The solution of the FP equation for small Δt is given by

$$P(y_2,t+\Delta t|y_1,t) = \frac{1}{2\sqrt{\pi D^{(2)}(y_2)\Delta t}} \times \exp\left\{\frac{-[y_2-y_1-D^{(1)}(y_2)\Delta t]^2}{4D^{(2)}(y_2)\Delta t}\right\}. \tag{10}$$

Equation (10) enables us to predict the probability of an "observation" y_2 at time $t+\Delta t$, if we know y_1 at t. In Fig. 5 we show the computed conditional PDFs using the data, and those using Eq. (10), for three values of y_1 with $\Delta t = 1$. To further check the accuracy of the reconstructed $y(t)$, we used the Kolmogorov-Smirnov test to compare the cumulative distribution function for the original and reconstructed [i.e., Eq. (10)] $y(t)$. With 1682 data points, we find the maximum difference between the two cumulative PDFs to be about 0.030. For the α levels 10%, 5%, and 1%, we find the critical values to be 0.042, 0.046, and 0.056, respectively.

To make predictions for the *future*, we write $x(t+1)$ in terms of $x(t)$,

$$x(t+1) = x(t)\exp\{\sigma_y[y(t)+\bar{y}]\}, \tag{11}$$

where $\bar{y}$ and σ_y are the mean and standard deviations of $y(t)$. To use Eq. (11) to predict $x(t+1)$, we need $[x(t),y(t)]$. We select three consecutive points in the series $y(t)$ and search for three consecutive points in the reconstructed $y(t)$ with the smallest difference with the selected points. The difference is

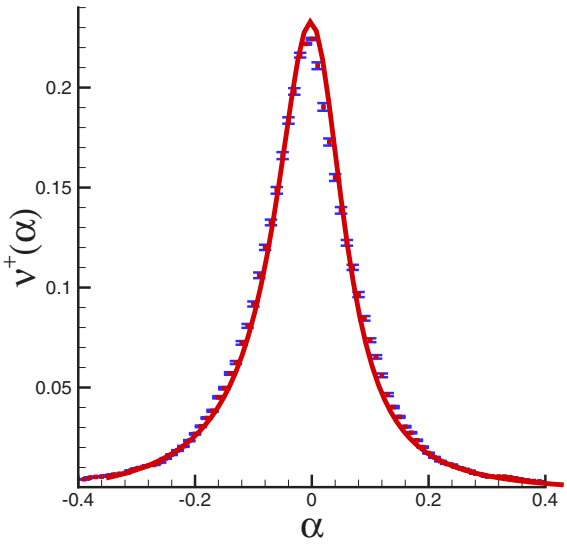

FIG. 6. (Color online) The level crossing $\nu^+\alpha$ for the returns time series. Actual data (symbols) and reconstructed ones (curve).

considered minimum if it is less than $0.05y_{\max}$. Wherever this happens is taken to be the time t which fixes $[x(t),y(t)]$. Shown in Fig. 1 are the actual data and the predictions for some interval in the oil price $x(t)$, beginning with $t \simeq 2006$. Our computations indicate that the predictions are accurate for up to 8 days (recall that t_M is on the order of 1 day), but the uncertainties increase beyond this time.

Finally, we computed the frequency of the level crossings at a given level α [23–25], given by $\nu_\alpha^+ = P(y_i > \alpha, y_{i-1} < \alpha)$, where ν_α^+ is the number of positive-difference crossings of $y(t)$, $y(t)-\bar{y}=\alpha$, in the interval T. The quantity $T(\alpha)=1/\nu_\alpha^+$ is the average time interval that one should wait in order to observe $y=\alpha$ again. The frequency ν_α^+ is given by

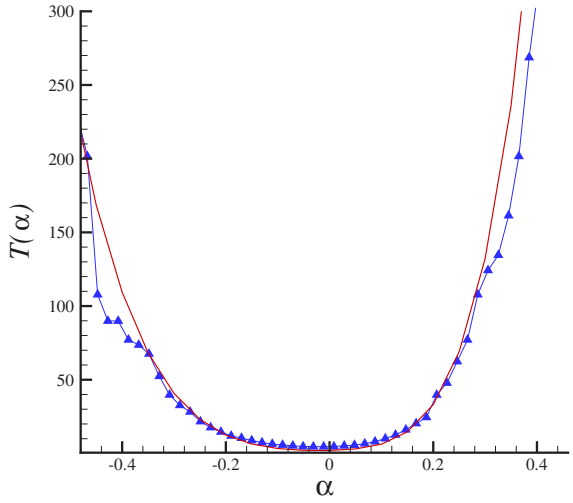

FIG. 7. (Color online) The average waiting time $T(\alpha)$ (in days) for observing $y(t)=\alpha$ again. Actual data (symbols) and reconstructed ones (curve).

$$\nu_\alpha^+ = \int_{-\infty}^{\alpha}\int_{\alpha}^{\infty} P(y_i,y_{i-1})dy_i\,dy_{i-1}$$
$$= \int_{-\infty}^{\alpha}\int_{\alpha}^{\infty} P(y_i|y_{i-1})P(y_{i-1})dy_i\,dy_{i-1}, \qquad (12)$$

where $P(y_{i-1}=y)=[C/D^{(2)}]\exp[\int_0^y dy' D^{(1)}(y')/D^{(2)}(y')]$, and $P(y_i|y_{i-1})$ is given by Eq. (10) with $\Delta t=1$, with C being a normalization constant. In Figs. 6 and 7, we present the computed level-crossing frequency and $T(\alpha)$, in units of days, over a time interval, for both the actual data set and the reconstructed one obtained through Eq. (11). The maximum and minimum of y are 0.4 and −0.4, respectively.

We would like to thank M. S. Movahed and D. Sornette for useful comments and discussions.

[1] R. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexities in Finance* (Cambridge University Press, New York, 2000).
[2] M. Sahimi, *Flow and Transport in Porous Media and Fractured Rock* (VCH, Weinheim, 1995).
[3] P. Ch. Ivanov *et al.*, Nature (London) **399**, 461 (1999); Y. Ashkenazy *et al.*, Phys. Rev. Lett. **86**, 1900 (2001).
[4] The data were taken from the Energy Information Agency of the United States Department of Energy; see http://www.eia.doe.gov/emeu/international/crude1.html
[5] C. K. Peng *et al.*, Phys. Rev. E **49**, 1685 (1994); S. M. Ossadnik *et al.*, Biophys. J. **67**, 64 (1994).
[6] M. S. Taqqu *et al.*, Fractals **3**, 785 (1995); A. R. Mehrabi *et al.*, Phys. Rev. E **56**, 712 (1997).
[7] J. Feder, *Fractals* (Plenum, New York, 1988).
[8] R. Friedrich and J. Peinke, Phys. Rev. Lett. **78**, 863 (1997).
[9] G. R. Jafari *et al.*, Phys. Rev. Lett. **91**, 226101 (2003).
[10] F. Ghasemi *et al.*, Eur. Phys. J. B **47**, 411 (2005); F. Ghasemi *et al.*, J. Biol. Phys. **32**, 117 (2006).
[11] M. R. Rahimi Tabar *et al.*, Comput. Sci. Eng. **8**, 86 (2006).
[12] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1984).
[13] W. Feller, Ann. Math. Stat. **30**, 1252 (1959).
[14] R. Colistete *et al.*, Int. J. Mod. Phys. D **13**, 669 (2004).
[15] K. E. Bassler *et al.*, Physica A **369**, 343 (2006).
[16] P. Billingsley, *Probability and Measure* (Wiley, New York, 1995).
[17] Y. Ait-Sahalia, L. P. Larsen, and J. A. Scheinkman, in *Handbook of Financial Econometrics*, edited by Y. Ait-Sahalia and L. P. Hansen, home.uchicago.edu/lhansen/handbook.htm
[18] F. Ghasemi *et al.*, J. Stat. Mech.: Theory Exp. 2006, P11008 .
[19] J. P. Bouchaud and R. Cont, Eur. Phys. J. B **6**, 543 (1998).
[20] C. Renner *et al.*, J. Fluid Mech. **433**, 383 (2001).
[21] J. Davoudi and M. R. Tabar, Phys. Rev. Lett. **82**, 1680 (1999).
[22] M. Waechter *et al.*, Europhys. Lett. **64**, 579 (2003).
[23] F. Shahbazi *et al.*, J. Phys. A **36**, 2517 (2003).
[24] G. R. Jafari *et al.*, J. Stat. Mech.: Theory Exp. 2006, P06008.
[25] A. Bahraminasab *et al.*, J. Stat. Phys. **124**, 1471 (2006).