Analysis of Nonstationary Stochastic Processes with Application to the Fluctuations in the Oil Price

Fatemeh Ghasemi,[a] Muhammad Sahimi,[b] J. Peinke,[c] R. Friedrich,[d]

G. Reza Jafari,[e] and M. Reza Rahimi Tabar[c,f,g]

[a] *The Max Planck Institute for the Physics of Complex Systems, Nthnitzer Strasse 38, 01187 Dresden, Germany*

[b] *Mork Family Department of Chemical Engineering & Materials Science, University of Southern California, Los Angeles,*

CA 90089-1211

[c] *Carl von Ossietzky University, Institute of Physics, D-26111 Oldenburg, Germany*

[d] *Institute for Theoretical Physics, University of Münster, D-48149 Münster, Germany*

[e] *Department of Physics, Shahid Beheshti University, Evin, Tehran 19839, Iran*

[f] *Dep. of Physics, Sharif University of Technology, P.O. Box 11365-9161, Tehran 11365, Iran*

[g] *CNRS UMR 6529, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France*

We describe a method for analyzing a nonstationary stochastic process $x(t)$, and utilize it to study the fluctuations in the oil price. Evidence is presented that the fluctuations in the returns $y(t)$, defined as, $y(t) = \ln\{x(t+1)/x(t)\}$, where $x(t)$ is the datum at time t, constitute a Markov process, characterized by a Markov time scale t_M. We compute the coefficients of the Kramers-Moyal expansion for the probability distribution function $P(y, t|y_0, t_0)$, and show that $P(y, t|, y_0, t_0)$ satisfies a Fokker-Planck equation, which is equivalent to a Langevin equation for $y(t)$. The Langevin equation provides quantitative predictions for the oil price over Markov time scale 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 $y(t)$, where $T(\alpha) = 1/\nu_\alpha^+$ is the average waiting time for observing $y(t) = \alpha$ again. The method described is applicable to a wide variety of nonstationary stochastic processes which, unlike many of the previous methods, does not require the data to have any scaling feature.

PACS numbers(s): 05.10.Gg, 05.40.-a, 05.45.Tp

Analyzing and characterizing nonstationary stochastic processes which fluctuate widely and contain extended correlations has been a problem of fundamental interest for a long time. Examples of such processes that occur in many natural and man-made phenomena include various indicators of economic activities [1], such as the stock market, to fluctuations in the porosity and permeability of porous materials [2], velocity fluctuations in turbulent flows, and heartbeat dynamics [3]. We propose in this Rapid Communication a method for, (1) generating a stationary process out of a nonstationary process; (2) analyzing the extent and nature of the correlations in the resulting stationary process, and (3) constructing effective stochastic continuum equations that not only reconstruct the process, but also provide quantitative predictions for it over a period of time which is on the order of the Markov time scale.

The method is based on application of Markov processes and development of a Langevin equation for a time series $y(t)$ which is constructed based on the original nonstationary process $x(t)$. The method, which is quite general, is used to analyze the fluctuations in the oil price, although the method is applicable to other similar nonstationary stochastic processes. Although the oil price is one of the most notorious nonstationary stochastic processes, we show that the method that we describe can analyze its fluctuations in a meaningful manner, and provide quantitative predictions for the price.

Figure 1 presents the daily fluctuations in the oil price $x(t)$ in the period 1998-2006 [4]. It is not difficult to show that the fluctuations do not constitute a stationary process, for instance one can show that the variance of the signal in some window are not stable under increasing the its window‘s size. Let us introduce the *returns* $y(t)$ defined by, $y(t) = \ln[x(t+1)/x(t)]$. The resulting series for $y(t)$ is shown in Fig. 2. It can be shown straightforwardly, by measuring the average and variance of $y(t)$ in a moving window, that $y(t)$ is stationary. We then computed the spectral density $S(f)$ of $y(t)$ in order to check whether there are long-range, and in particular nondecaying, correlations in $y(t)$. The result, $S(f) \propto f^{\beta}$ with $\beta \simeq 0$, indicated the absence of long-range, nondecaying correlation in $y(t)$. Also according to our calculation for Hurst exponent by Deterended fluctuation analysis (DFA) and rescaled range analysis (R/S), the return can be considered as a stationary series [5-7]. Indeed Hurst exponent of the return time series is $H = 0.51 \pm 0.2$, therefore one concludes the return time series is stationary signal.

The proposed method for analyzing the fluctuations in $y(t)$ is then based on constructing a Langevin equation for $y(t)$. The construction of the Langevin equation is carried out in two steps. Since long-range, nondecaying correlations are absent in $y(t)$, but short-range decaying correlations do exist, we first check whether the time series follows a Markov chain [8-11], in which case we estimate its Markov time scale t_M - the minimum time interval over which the series can be considered as a Markov process. In general, complete characterization of the statistical properties of stochastic fluctuations of $y(t)$ requires evaluation of the joint probability distribution function (PDF) $P_n(y_1, t_1; \cdots; y_n, t_n)$ 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 can be generated by the product of the conditional probabilities $P(y_{i+1}, t_{i+1}|y_i, t_i)$, for $i = 1, \cdots, n-1$. A necessary condition for a stochastic phenomenon to be a Markov process is that the Chapman-Kolmogorov (CK) equation [12],

$$P(y_2, t_2|y_1, t_1) = \int dy_3 \; P(y_2, t_2|y_3, t_3)P(y_3, t_3|y_1, t_1) \; , \tag{1}$$

should hold for any t_3 in the interval $t_1 < t_3 < t_2$. One should check the validity of the CK equation for different values of y_1 by comparing the directly-evaluated conditional probability distributions $P(y_2, t_2|y_1, t_1)$ with those calculated according to right side of Eq. (1).

In what follows, we use the least squares method to determine the Markov time scale of the $y(t)$. If $y(t)$ be a Markov process then one finds

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

In order to determine the Markov time scale, we compare the three-point PDF with that obtained on the basis of the Markov process. The joint three-point PDF, in terms of conditional probability functions, is given by $P(y_3,t_3;y_2,t_2;y_1,t_1) = P(y_3,t_3|y_2,t_2;y_1,t_1)P(y_2,t_2;y_1,t_1)$. Using the properties of the Markov process and substituting in equation (2), we obtain

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

In order to check the condition for the data being a Markov process, we must compute the three-point jiont PDF through equation (2) and compare the result with equation (3). The first step in this direction is to determine the quality of the fit through the least squares fitting quantity χ^2 defined by

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

where $\sigma_{3.joint}$ and σ_{Mar} are the variances of $P(y_3,t_3;y_2,t_2;y_1,t_1)$ and $P_{Mar}(y_3,t_3;y_2,t_2;y_1,t_1)$, respectively. To compute the Markov time scale, we use the likelihood statistical analysis [13]. In the absence of a prior constraint, the probability of the set of three-point joint PDFs is given by a product of Gaussian functions:

$$p(t_3 - t_1) =$$

$$\Pi_{y_3,y_2,y_1} \frac{1}{\sqrt{(\sigma^2_{3.joint} + \sigma^2_{Mar})^2}} \exp\left[\frac{[P(y_3,t_3;y_2,t_2;y_1,t_1) - P_{Mar}(y_3,t_3;y_2,t_2;y_1,t_1)]^2}{2(\sigma^2_{3.joint} + \sigma^2_{Mar})}\right] \tag{5}$$

This probability distribution must be normalized. Evidently, when, for a set of values of the parameters, the $\chi^2\nu$ is minimum, the probability is maximum. Figure 3 shows the normalized $\chi^2\nu$ as a function of the $t_3 - t_1$, where $\chi^2\nu = \chi^2/N$ and N being the number of degrees of freedom. The minimum value of $\chi^2\nu$ is ≈ 0.6, corresponding to $t_M = t_3 - t_1 \simeq 1$ day. Figure 4 shows the likelihood function of the Markov time scale of the $y(t)$. Here we have used the χ^2 test to determine the Markov time scale. Using the method proposed in [6-11,14-16], one can estimate the Markov time scale via direct check of CK equation. The result is again $t_M \simeq 1$ day. However in χ^2 method we can estimate the likelihood function of the Markov time scale (see [17] for more details).

We next derive a stochastic equation that describes the fluctuations of $y(t)$. Knowledge of $P(y_2,t_2|y_1,t_1)$ (for Markov process) is sufficient for generating the entire statistics of the returns encoded in the $n-$point probability density, which satisfies a master equation. The master equation is then put in the form of a Kramers-Moyal (KM) expansion:

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

The KM coefficients $D^{(k)}(y,t)$ are defined in terms of the conditional moments $M^{(k)}(y,t)$:

$$\begin{aligned} D^{(k)}(y,t) &= \frac{1}{k!}\lim_{\Delta t\to 0} M^{(k)} \ , \\ M^{(k)} &= \frac{1}{\Delta t}\int dy'(y'-y)^k P(y',t+\Delta t|y,t) \ . \end{aligned} \tag{7}$$

For a general stochastic process, all the KM coefficients may be nonzero. According to Pawula's theorem, however, the KM expansion may be truncated after the second term, provided that the fourth-order coefficient $D^{(4)}$ vanishes [12]. For our data, $D^{(4)} \simeq 10^{-2}D^{(2)}$, where values of $y(t)$ are measured in units of the maximum of the time series, $y_{\max}$. Then we can truncation the KM expansion. In that case, the KM expansion reduces to a Fokker-Planck (FP) equation (i.e. the equation (4) with $D^{(k)}(y,t) = 0$ for $k \geq 3$). We note that the FP equation is equivalent to the following Langevin equation (using the Ito interpretation)[12,18]:

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

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) separates the deterministic and stochastic components of the returns' fluctuations in terms of the coefficients $D^{(1)}$ and $D^{(2)}$. It turns out that the resulting drift and diffusion coefficients, estimated directly from the data, are, respectively, linear and quadratic functions of y, and are well-represented by the approximates,

$$\begin{aligned} D^{(1)}(y) &= -1.09y \ , \\ D^{(2)}(y) &= 0.0033 - 0.003y + 0.716y^2 \ . \end{aligned} \tag{9}$$

Estimates of $D^{(1)}(y)$ and $D^{(2)}(y)$ become poor for large y and, thus, the uncertainty in the estimates increases. Equation (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 time series for $y(t)$ are shown.

Next, we evaluate the precision of the reconstructed $y(t)$. This is done by reconstructing the conditional PDF through the numerical solution of the FP equation for the conditional PDF, which is much sensitive to the errors in $D^{(1)}$ and $D^{(2)}$ [6-11, 18-21]. The solution of the FP equation for small Δt is given by,

$$P(y_2, t+\triangle t|y_1, t) = \frac{1}{2\sqrt{\pi D^{(2)}(y_2)\Delta t}}$$

$$\times \exp\left[\frac{-(y_2 - y_1 - D^{(1)}(y_2)\Delta t)^2}{4D^{(2)}(y_2)\Delta t}\right] . \qquad (10)$$

Equation (8) enables us to predict the probability of an "observation" y_2 at time $t+\triangle t$, if we know y_1 at time 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 do more check we used Kolmogorov-Smirnov test to compare the cumulative distribution function for the original and reconstructed (i.e. eq.(10)) time series. With 1682 data we find the maximum difference between two cumulative PDFs to be about 0.030. For Alpha levels %10, %5 and %1, we find the critical values are 0.042, 0.046 and 0.056, respectively.

To make predictions, we use the definition of $y(t)$ to write $x(t+1)$ in terms of $x(t)$. But, since the reconstructed data have unit variance and zero mean, we have

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

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 series of $y(t)$ that have the smallest difference with the selected points. We consider the difference to be minimum if it is less than $0.05y_{\max}$ (stricter rules are clearly possible). 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 $t \simeq 2006$ (blue).

Finally, using the calculated drift and diffusion coefficients, we can compute the frequency of the level-crossings at a given level α [22-24]. This quantity is 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 time scale $T(\alpha) = 1/\nu_\alpha^+$ is then the average time interval that one should wait in order to observe the given $y = \alpha$ again. The frequency ν_α^+ has the following expression in terms of joint and conditional PDF's,

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

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

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

In summary, a method has been described for analyzing a nonstationary stochastic process in terms of a stationary one. In particular, we used the method to analyze the fluctuations in the oil price. The method is based on a Kramers-Moyal expansion for the probability $P(y, t)$ of the returns $y(t)$, and construction of a Fokker-Planck equation for the conditional probability $P(y, t|y_0, t_0)$ and a Langevin equation for $y(t)$. The method is applicable to any nonstationary stochastic process that contains no long-range, nondecaying correlations. The method can predict the oil price over a period time on the order of an optimal time that, as we described above, can be estimated from the data.

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

REFERENCES

[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] See, for example, 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] Peng C K, Buldyrev S V, Havlin S, Simons M, Stanley H E, and Goldberger A L, 1994 Phys. Rev. E **49**, 1685; Ossadnik S M, Buldyrev S B, Goldberger A L, Havlin S, Mantegna R N, Peng C K, Simons M and Stanley H E, 1994 Biophys. J. **67**, 64

[6] Taqqu M S, Teverovsky V and Willinger W, 1995 Fractals **3**, 785

[7] Feder J, 1988 *Fractals* (Plenum Press, New York)

[8] R. Friedrich and J. Peinke, Phys. Rev. Lett. **78**, 863 (1997).

[9] G. R. Jafari, S. M. Fazlei, F. Ghasemi, S. M. Vaez Allaei, M. R. Rahimi Tabar, A. Iraji Zad, and G. Kavei, Phys. Rev. Lett. **91**, 226101 (2003).

[10] F. Ghasemi, J. Peinke, M. Sahimi, and M. R. Rahimi Tabar, Eur. Phys. J. B **47**, 411 (2005); F. Ghasemi, Muhammad Sahimi, J. Peinke and M. Reza Rahimi Tabar, Journal of Biological Physics 32,117-128 (2006)

[11] M. R. Rahimi Tabar, F. Ghasemi, J. Peinke, R. Friedrich, K. Kaviani, F. Taghavi, S. Sadeghi, G. Bijani, and M. Sahimi, Comput. Sci. Eng. **8**, 86 (2006).

[12] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1984).

[13] R. Jr. Colistete, J. C. Fabris, S. V. B. Goncalves and P. E. de Souza, Int. J. Mod. Phys. D 13 669 (2004)

[14] K. E. Bassler, G. H. Gunaratne, and J. L. McCauley, Physica A, 369, 343-353 (2006).

[15] P. Billingsley, Probability and Measure, Wiley, 1995.

[16] Y. Ait-Sahalia, L.P. Larsen, J.A. Scheinkman, Operator methods for continuous-time Markov processes, in Handbook of Financial Econometrics, Edited by Yacine Ait-Sahalia and Lars Peter Hansen, http://home.uchicago.edu/ lhansen/handbook.htm

[17] F Ghasemi, A Bahraminasab, M Sadegh Movahed, Sohrab Rahvar, K R Sreenivasan, M Reza Rahimi Tabar, J. Stat. Mech. P11008 (2006).

[18] J. P. Bouchaud and R. Cont, Eur. Phys. J. B **6**, 543 (1998),

[19] C. Renner, J. Peinke, and R. Friedrich, J. Fluid Mech. **433**, 383 (2001).

[20] J. Davoudi and M. R. Rahimi Tabar, Phys. Rev. Lett. **82**, 1680 (1999).

[21] M. Waechter, F. Riess, H. Kantz, and J. Peinke, Europhys. Lett. **64**, 579 (2003).

[22] F. Shahbazi, S. Sobhanian, M. R. Rahimi Tabar, S. Khorram, G. R. Frootan, and H. Zahed, J. Phys. A **36**, 2517 (2003).

[23] G. R. Jafari, M. S. Movahed, S. M. Fazeli, M. R. Rahimi Tabar, and S. F. Masoudi, J. Stat. Mech., P06008 (2006).

[24] Bahraminasab A ., Movahed , M. Sadegh, Nasiri , S. D., Masoudi A. A. , Sahimi M, Journal of Statistical Physics 124 (6): 1471-1490, (2006)

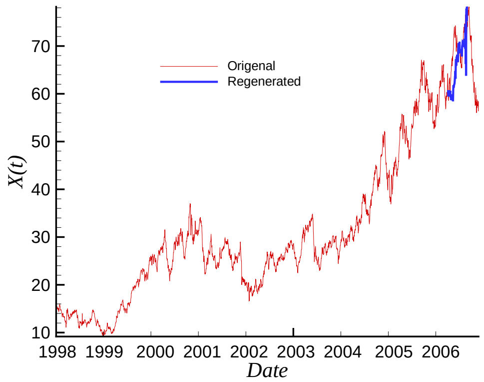

FIG. 1. Daily fluctuations in the oil price [4]. The time lag is one day. Shown are a sample of the actual daily oil prices (red) and the reconstructed data (blue), using the Langevin equation. The inset shows the predicted range.

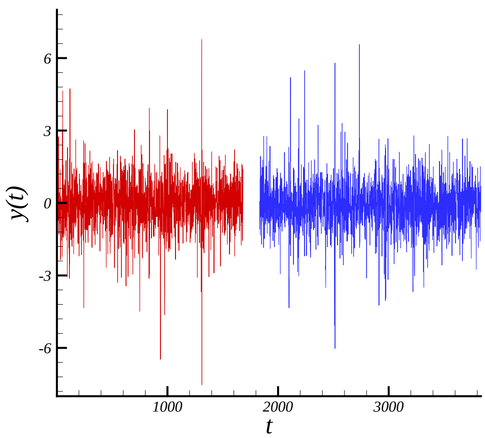

FIG. 2. Comparison of the actual return data (red) and the reconstructed ones using the corresponding Langevin equation (blue). For clarity of presentation the time series have been shifted on the $t-$axis.

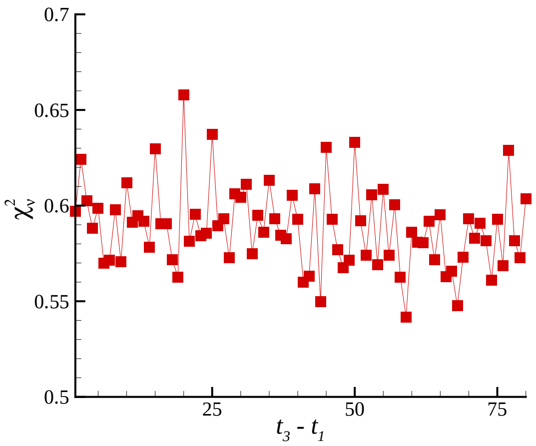

FIG. 3. The χ^2 test for estimation of Markov time scale of the time series of the returns, indicating that the Markov time scale t_M is 1 day.

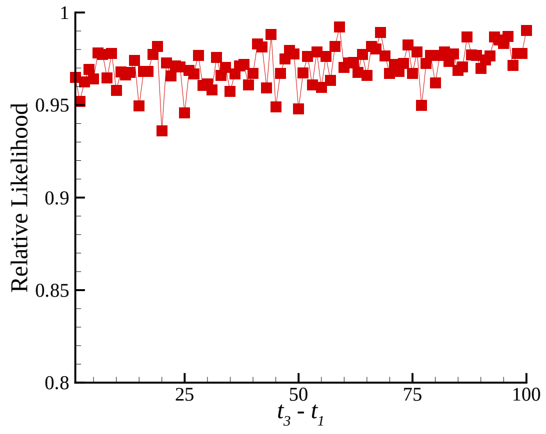

FIG. 4. Relative likelihood function for the Markov time scale of the $y(t)$ fluctuations, as a function of $t_3 - t_1$.

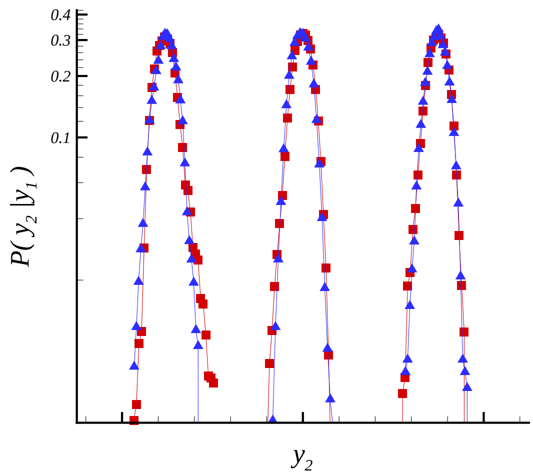

FIG. 5. Comparison of the directly-evaluated PDFs using the actual data, and the PDFs obtained from Eq. (10). Values for y_1, from left to right, are -0.1, 0.0, and 0.1 [measured in units of $y_{\max}(t)$]. For better presentation, the PDFs have been shifted on the horizontal axis.

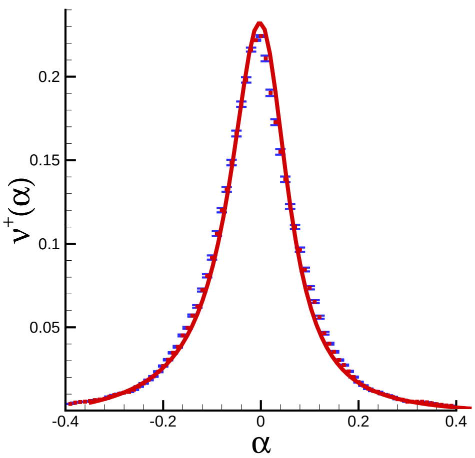

FIG. 6. The level crossing $\nu^+\alpha$ for the returns time series.

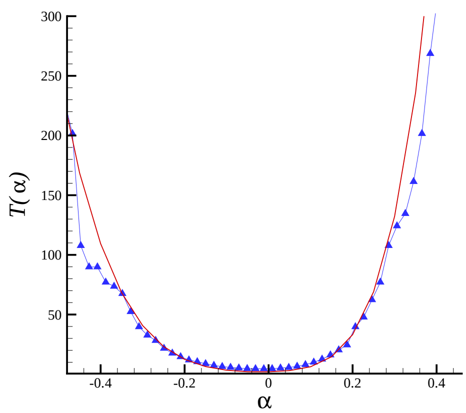

FIG. 7. The average waiting time for observing $y(t) = \alpha$ again.