Journal of Applied Nonlinear Dynamics 2(2) (2013) 141–150

Journal of Applied Nonlinear Dynamics

Journal homepage: www.lhscientificpublishing.com/journals/JAND.html

Self-Similar Property of Random Signals: Solution of Inverse Problem

Raoul R. Nigmatullin[1]†and J.A. Tenreiro Machado[2]

[1]Theoretical Physics Department, Institute of Physics, Kazan(Volga Region) Federal University, Kremlevskaya str., 18, 420008, Kazan, Tatarstan, Russian Federation

[2]ISEP-Institute of Engineering, Polytechnic of Porto, Department of Electrical Engineering, Rua Dr. Antonio Bernardino de Almeida, 431 4200-072 Porto, Portugal

Submission Info

Communicated by Albert C.J. Luo
Received 28 September 2012
Accepted 9 October 2012
Available online 1 July 2013

Keywords

Scaling equation
Inverse solution
Self-similar properties of random signals

Abstract

Many random signals with clearly expressed trends can have self-similar properties. In order to see this self-similar property new presentation of signals is suggested. A novel algorithm for inverse solution of the scaling equation is developed. This original algorithm allows finding the scaling parameters, the corresponding power-law exponent and the unknown log-periodic function from the fitting procedure. The effectiveness of algorithm is tested in financial data revealing season fluctuations of annual, monthly and weekly prices. The general recommendations are given that allow the verification of this algorithm in general data series.

1 Introduction and formulation of the problem

During three decades of intensive research it became clear that our world is presumably fractal and repeats itself on different scales, both in space and time. But nowadays it is *not* sufficient to say that the object/system studied has self-similar properties. It is necessary to find the fractal dimension, to determine the limits of applicability of the scaling properties and to prove the evidence/absence of log-periodic oscillations [1] that accompany any scaling process in time or space. The same phenomena were discovered in random economical and financial activities [2]. In general, research in this field simply *supposes* or *postulates* the existence of scaling properties of the system studied. Nevertheless, if the researcher did not make this supposition initially then he must find the justification evidences that the system has really scaling/self-similar properties. How to find the convincing arguments for the skeptical scholar if a scientist has only a set of numerical data

†Corresponding author.
Email address: renigmat@gmail.com, jtm@isep.ipp.pt

ISSN 2164 – 6457, eISSN 2164 – 6473/$-see front materials
DOI : 10.5890/JAND.2013.04.003

characterizing the response of the complex system and nothing else? In this paper we demonstrate a simple criterion demonstrating the self-similarity property of the given data verified. If the data of the system under study has self-similar properties, then it is necessary to organize the adequate fit of the data over the complete solution of the simplest scaling equation given below

$$S(z\xi) = bS(z) + s_0 \tag{1}$$

$$S(z) = z^{\alpha}\,\mathrm{Pr}(\ln z) + \frac{s_0}{1-b}, \alpha = \frac{\ln b}{\ln \xi}, \tag{2}$$

$$\begin{gathered} \mathrm{Pr}(\ln z) = A_0 + \sum_{k=1}^{K}[Ac_k \cos(2\pi k \frac{\ln z}{\ln \xi}) + As_k \sin(2\pi k \frac{\ln z}{\ln \xi})], \\ \mathrm{Pr}(\ln z \pm \ln \xi) = \mathrm{Pr}(\ln z). \end{gathered} \tag{3}$$

Here $S(z)$ defines a physical value that depends on argument z which in general can accept real or complex value. The parameters ξ and b denote the scaling factors. The constant s_0 represents a possible shift and the parameter α defines in general a fractal dimension. The function Pr(ln z) in Eqs. (3) defines the unknown log-periodic function that can be decomposed to the Fourier series and represented approximately by its finite segment.

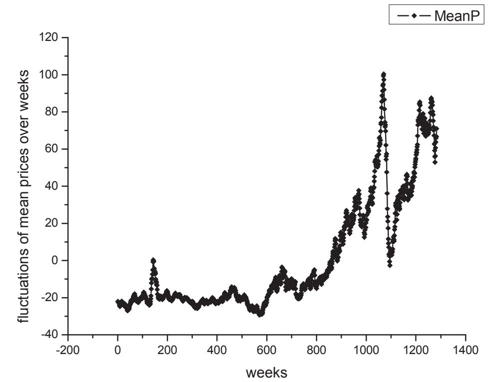
The problem can be formulated as follows. It is necessary to justify the self-similar properties of the random data characterizing the behavior of the complex system under analysis. Then it is necessary to develop a procedure for the fitting of the data to the function $S(z)$, to find the desired fitting parameters b, ξ, s_0 and to restore the scaling equation for the function $S(z)$ figuring in expression (1). Usually, the scaling equation (1) *a priori* is supposed to be known and in this case the solution (2) is restored easily. If we have only some data and any additional information is *absent* then restoration the scaling equation (1) poses some problems. That is the reason why this problem can be formulated as *inverse* problem that proves and describes the self-similar properties of random signals starting from analysis of initial data points. To the authors best knowledge, adequate algorithms for solving inverse problems involving self-similar random signals registered as a response of the complex system are *absent* in the current scientific literature. In this paper we want to show that it is possible to develop a simple and reliable algorithm that helps solving the *inverse* problem by restoring the scaling equation of the type (1).

Bearing these ideas in mind this paper is organized as follows. In Section 2 we describe the proposed algorithm. In Section 3 we consider the economical data associated with season fluctuations of oil prices. We prove that, in spite of its complex characteristics, it is possible to reveal the scaling properties and to find an accurate fit of the data to the function $S(t)$ and then to restore the desired scaling equation (1). In Section 4 we list of the basic results and general recommendations that may be helpful in the analysis of other random signals having a self-similar structure.

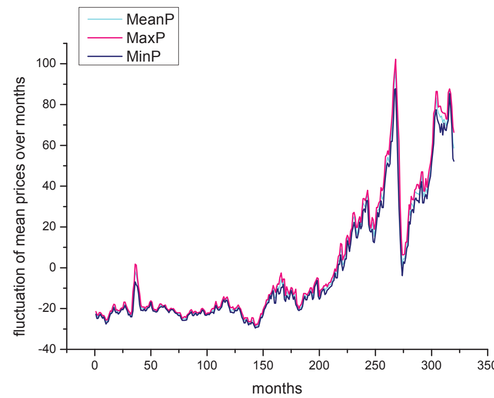
2 Description of the algorithm

We can adopt simple concepts that help showing clearly the self-similar properties of the random signal under study by having the clearly expressed trend. The idea is based on the following algorithm.

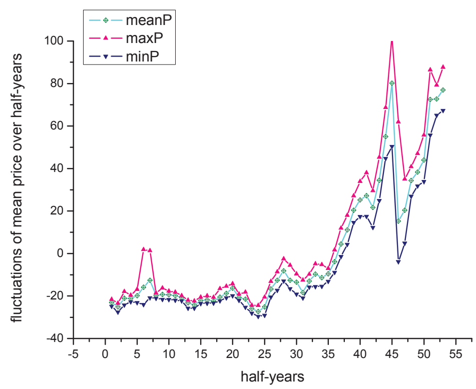
Step 1. Let us choose some interval $[x_1, x_k]$ containing a set of k data points $\{(x_1, y_1), \ldots, (x_k, y_k)\}$. One can reduce this information into *three points* if the one point is the mean value of the amplitudes



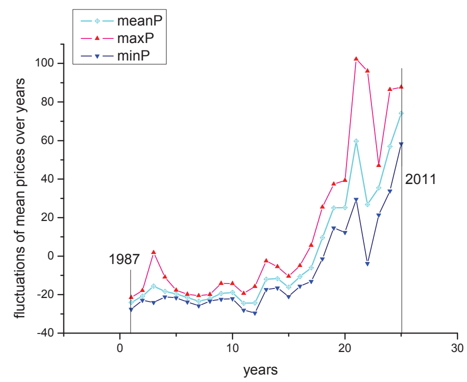
(a)



(b)



(c)



(d)

Fig. 1 (a) Fluctuations of mean prices over weeks (the trading weeks were taken over 25 years (1987—2011). (b). Fluctuations of mean prices taken over months for the same period. The upper and down curves describing the max (Pr) (red line) and min (Pr) (navy line) do not strongly differentiate from each other. These curves are similar to the one depicted on Fig1(a) that contains more detailed information. The compression data ratio is 4. (c) Fluctuations of mean prices over half-year (the trading period covers again 25 years (1987—2011). (d) shows the fluctuations over one year. As before, the limits of these season fluctuations are limited by max(Pr) and min(Pr) curves, correspondingly.

and the other two points correspond to their maximal and minimal values. So, this selection represents the simplest reduction of the given set of k random points to *three* new points $p_1 = \text{mean}\{y_1,\ldots,y_k\}$, $p_2 = \max\{y_1,\ldots,y_k\}$, $p_3=\min\{y_1, \ldots, y_k\}$. In general, these chosen intervals cannot be equaled to each other and contain different number of points. For the case of unequal intervals the number of selected points should be chosen based on some criterion that has a specific characteristic depending on the data considered. Figures 1–4 demonstrate the result of this specific reduction when adopting equal length intervals. In order to diminish the number of charts we chose for demonstration the available economical data. These data can be analyzed in details in the next section.

The initial data containing information about season fluctuations of oil prices during 25 years (1987—2011 years) is averaged over the five working days in the week (i.e., Monday to Friday) and this average value is considered as one "trading event" (the corresponds to one point on Fig.1(a)). The total number of these points covers the period 25 years and equals 1317 points. Then the data is compressed by a factor of 4 in order to produce monthly curves. If the calculation is for periods of half-year or one year fluctuations, then the initial data (depicted on Fig. 1(a)) is compressed by a factor of 24 or 51, respectively. The last 6 compressed data curves (for half-year and one year periods) are shown on Fig. 1(c) and 1(d), correspondingly.

In spite of compression of the initial data by a factor of 51 (i.e., comparing Fig.1(a) and Fig.1(d)) all these curves are similar to each other. One can notice also that the strong compression creates larger deviations of the upper (max(Pr)) and lower (min(Pr)) curves with respect to the mean(Pr) curve located always between them. These deviations are much smaller for the curves depicted on Fig.1(b). So, we can conclude that these curves keep their similarity over a wide range of scales.

Step 2. The second step is to approximate these curves by the function $S(t)$ defined by expressions (2) and (3). This fit can be realized with the help of the eigen-coordinates (ECs) method, described in papers [3, 4]. The ECs method helps to transform the curve initially containing nonlinear fitting parameters into a set of straight lines (if the chosen hypothesis is correct) and then to reduce the problem of calculation of the desired fitting parameters to the linear least square method (LLSM). We can notice from expression (1) that the proposed curve, serving as an initial hypothesis, has roughly the following form:

$$S(t) \cong c_0 + A_0 \exp(\alpha_0 \ln t) + \exp(\alpha_1 \ln t)\left[Ac_1 \cos(\omega_1 \ln t) + As_1 \sin(\omega_1 \ln t)\right]. \tag{4}$$

This function contains 7 fitting parameters and three of them $(\alpha_0, \alpha_1, \omega_1)$ enter into expression (4) by *nonlinear* way. In order to a have a reliable estimation of their values and to reduce the whole fitting procedure to the application of the well-known LLSM we notice the following relationship. Expression (4) satisfies to the differential Euler equation of the third order:

$$D^3 S(t) + a_1 D^2 S(t) + a_2 DS(t) + a_3 S(t) = K,\ D \equiv t\frac{d}{dt}. \tag{5}$$

Here the linear parameters a_k $(k = 1,2,3)$ are tightly associated with the roots of the cubic equation:

$$r^3 + a_1 r^2 + a_2 r + a_3 = (r-\alpha_0)(r-\alpha_1 - i\omega_1)(r-\alpha_1 + i\omega_1) = 0. \tag{6}$$

Triple integration of equation (5) (this procedure helps to keep the initial error in the same limit as in the function $S(t)$) allows finding the constants $a_{1,2,3}$ of the cubic equation (6) by a linear procedure using, for this purpose, the stable procedure as the LLSM. The basic linear relationship (BLR) obtained after three-fold integration of the differential equation (5) can be written as:

$$\begin{aligned}
&Y(t) = \sum_{k=1}^{6} C_k X_k(t),\\
&Y(t) = S(t) - <\cdots>,\\
&X_1(t) = \int_{t_0}^{t} S(u)\frac{du}{u} - <\cdots>,\ C_1 = -a_1,\\
&X_2(t) = \int_{t_0}^{t} (\ln t - \ln u) S(u)\frac{du}{u} - <\cdots>,\ C_2 = -a_2,
\end{aligned} \tag{7}$$

$$X_3(t) = \frac{1}{2}\int_{t_0}^{t}(\ln t - \ln u)^2 S(u)\frac{du}{u} - <\cdots>,\ C_3 = -a_3,$$

$$k = 4,5,6,$$

$$X_k(t) = \ln^{7-k}(t) - <\cdots>,\ C_k(K, DS(t_0), D^2S(t_0), D^3S(t_0)).$$

Here the pair of brackets $<\cdots> = N^{-1}\sum_{k=1}^{N}(\cdots)$ defines the arithmetic mean of the neighboring expression located on the left that should be subtracted from it. The concrete form of the constants $C_k(K, DS(t_0), D^2S(t_0), D^3S(t_0))$, depending on the initial values of the derivatives in the initial point t_0, is not essential for the calculation of the desired C_i (i=1,2,3) and can be omitted. If the initial hypothesis (2) is supposed to be correct, then we can expect that the exponents α_0 and α_1 should have at least the *same* sign and cannot be strongly deviated from each other. In the opposite case it is necessary to consider another hypothesis.

Step 3. The third step is related to optimization of the power-law exponent α in accordance with idea that was suggested in [5]. If some nonlinear fitting parameter is located in the given limits $[\alpha_{\min},\ \alpha_{\max}]$ then one can introduce the function:

$$\alpha(v) = t\cdot v + e,$$
$$t = \frac{\alpha_{\max} - \alpha_{\min}}{v_{\max} - v_{\min}},\ e = \frac{\alpha_{\min}v_{\max} - \alpha_{\max}v_{\min}}{v_{\max} - v_{\min}}, \tag{8}$$

and then one can find the optimal value v_{opt} by minimizing the fitting function:

$$S(t,v) \cong c_0A_0\exp(\alpha(v)\cdot\ln t) + \exp(\alpha(v)\cdot\ln t)[Ac_1\cos(\omega_1\ln t) + As_1\sin(\omega_1\ln t)] \tag{9}$$

with respect to the value of the minimal error:

$$\min_{v_{\text{opt}}}(\text{Re}lErr) = \left(\frac{\text{sedev}(y(t) - S(t,v))}{\text{mean}|y(t)|}\right)\cdot 100\%. \tag{10}$$

Step 4. After calculation of the optimal value of the power-law exponent $\langle\alpha\rangle = \alpha(v_{\text{opt}})$ and of the initial frequency ω_1 from (9) it is easy to find the desired values of the remaining fitting parameters:

$$\xi = \exp\left(\frac{2\pi}{\omega_1}\right),\ b = \exp\left(\langle\alpha\rangle\cdot\frac{2\pi}{\omega_1}\right) \tag{11}$$

and conclude the final fit with the help of the function:

$$y(t) \cong t^{\langle\alpha\rangle}\,\text{Pr}(\ln(t)) + c,\ \langle\alpha\rangle = \frac{\ln b}{\ln\xi},$$
$$\text{Pr}(\ln t) = A_0 + \sum_{k=1}^{k}\left(Ac_k\cos\left(2\pi k\frac{\ln t}{\ln\xi}\right) + As_k\sin\left(2\pi k\frac{\ln t}{\ln\xi}\right)\right), \tag{12}$$

using the LLSM for calculation of the set of unknown coefficients $Ac_k, As_k, k = 1,2,\ldots,K$. The accuracy of the fitting procedure is realized by choosing of the corresponding value K that minimizes the value of the given relative error. For the fitting of all season fluctuations of season prices the value K=50 is adopted in the sequel.

This algorithm can be applied for the fitting of a wide range of different self-similar data if the following conditions are satisfied:

1. The inoculating values of the power-law exponents α_0 and α_1 have the *same* sign;
2. The value v_{opt} leading to the minimal value of the relative error defined by expression (8) does exist.

If these conditions are violated then it is necessary to change the initial hypothesis (2) following from (1) and to consider more complex cases. The considerations of these cases associated with solutions of the scaling equation of the second order can constitute a subject of separate research. For example, more complex hypothesis that can be used for the solution of inverse problem should contain at least two-power law exponents. For this case it is necessary to restore the scaling equation of the second order having the following form.

$$S(x\xi^2) = bS(z\xi) + cS(z) + s_0. \tag{13}$$

The analytical solution of this scaling equation was considered recently in the paper [6] but solution of the inverse problem for equation (13) is not simple task and can constitute a subject of separate research.

3 Analysis of real data

Economical and financial data series represent an interesting topic of research due their intricate characteristics [7, 8]. In this section we use data available at the EIA - U.S. Energy Information Administration (http://www.eia.gov/petroleum/). All special or holidays where there is no data available were reconstructed, estimating the corresponding value by means of a linear interpolation. Therefore, all weeks correspond to five days of data points (i.e., going to Monday up to Friday) since there is no trade during Saturday and Sunday.

In the previous section we proved that actual economical data demonstrating the season fluctuations of the oil prices has self-similar character (see Figs. 1(a)–1(d)). Now it is necessary to consider the results of application of Steps 2, 3 and 4 of the proposed algorithm. In order to reduce its extension we present only the figures associated with fluctuation of half-year prices. The specific and important details associated with the realization of Step 4 will be given below for different season fluctuations of the oil price. We want to stress that, for simplicity, we consider initially only the fluctuations of average weekly prices taken over for the period 1283 trading weeks that were shifted with respect to their mean value. Other trading weeks belonging to 2012 year (1317–1283 = 34) are omitted.

Now everything is ready in order to realize the final fit (expression (12)) demonstrating the desired solution of the inverse problem. In order to convince the skeptically tuned reader we chose K=50 and performed the fit with very high accuracy (the value of the relative error less than 10^{-4}(%). We show only the figures for half-year fluctuations. Other fitting curves obtained for weekly, monthly and annual fluctuations of oil prices have also the best fit but they are very similar to depicted ones and so they are not shown. The *principal* feature in accomplishing the inverse problem for self-similar random functions is related to the specific distribution of coefficients Ac_k and As_k that reflects the behavior of the fitting curve on the given range of scales. Figures 2(c), (d) show the final fit of the half-year fluctuations depicted above on Figs. 2(a), (b). The final

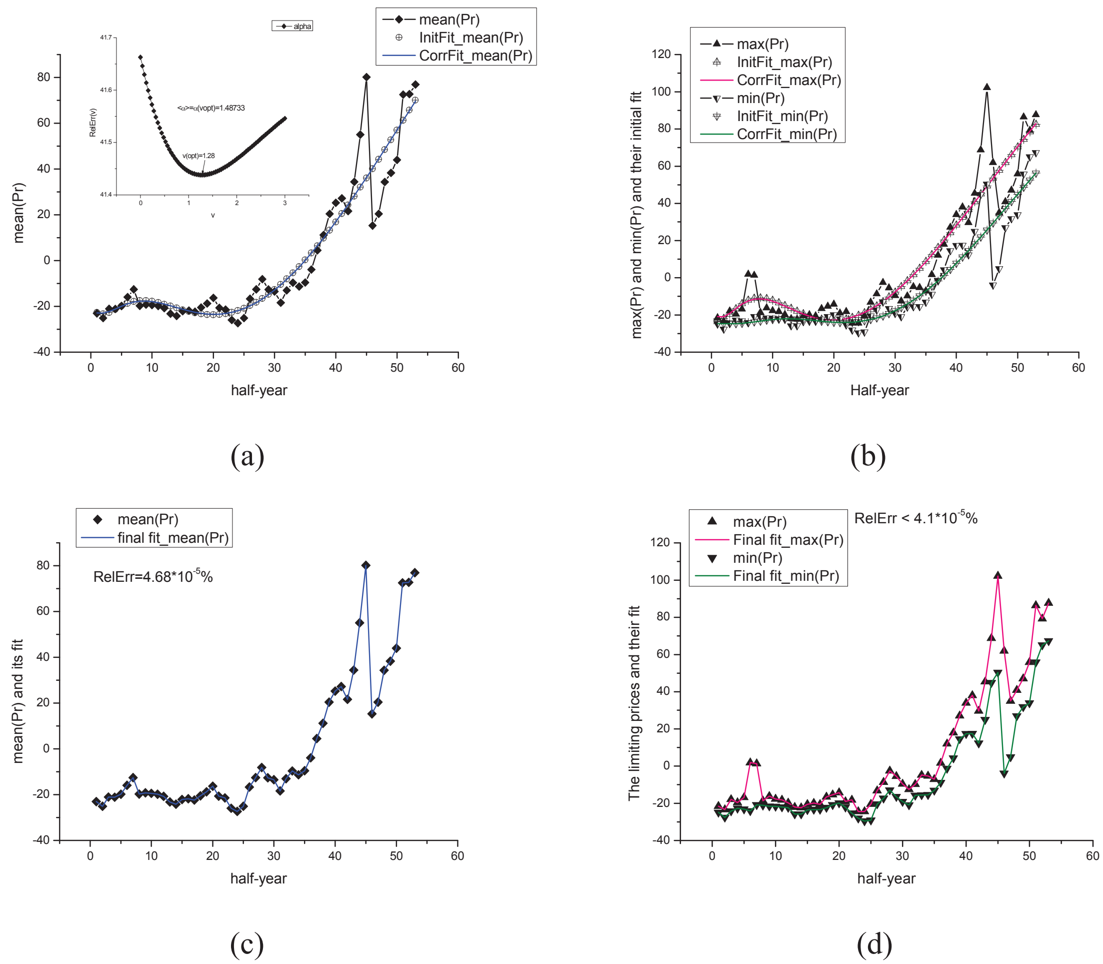
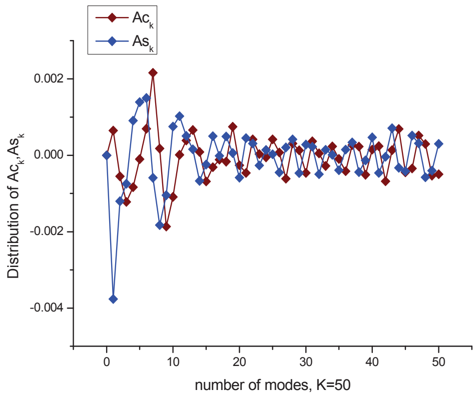
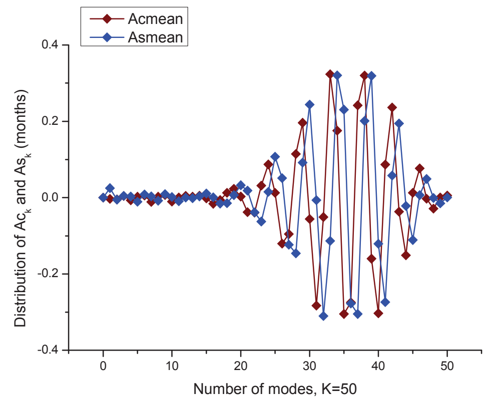


Fig. 2 (a) This figure demonstrate the half-year fluctuations of oil prices (black rhombs) and its initial fit (crossed rhombs) together with the corrected fit (navy color solid line). One can notice that the initial fit practically coincides with the corrected fit line. The upper curve demonstrates the existence of minimal value of the relative error with respect to the value v_{opt}=1.28. So, we should take the optimal value of $<\alpha>=1.4873$ for the final fitting procedure. (b) shows the same result obtained for the fluctuations of the max(Pr) (upper curve) and min(Pr) (down curve), correspondingly. The corrected fitting curves are shown by red and green solid lines.(c) Fluctuations of mean prices over half-years and its final fit by expression (10) K=50, The value of the relative error is equaled to $4.68{\cdot}10^{-5}$(%). (d) shows the final fit (realized again with the help of expression (10)) for max(Pr)-red solid line above and min(Pr) –green solid line below, correspondingly. The value of the relative error does not exceed the value $4.1{\cdot}10^{-5}$(%).

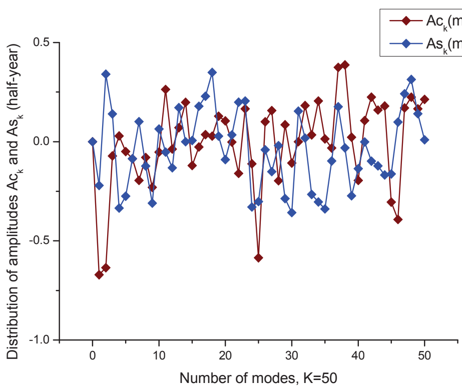
number of modes K=50 was remained the same for all treated files. The distributions of the fitting parameters $\langle\alpha\rangle, b, \xi$ associated with season fluctuations of oil prices for the given temporal period (1987–2011) are collected in the Table 1. Analyzing the parameters collected in this Table one can notice the following interesting peculiarities. The strong deviations of these parameters between the limiting season values are related to the influence of random fluctuations. The same result is valid for the parameters b and ξ. We can calculate the mean and Stdev of these parameters and find their averages. But we decided to keep these values unchanged, in order find the true reason of



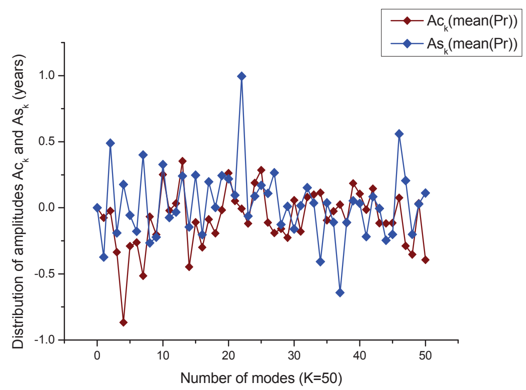
(a)



(b)



(c)



(d)

Fig. 3 (a) The distribution of amplitudes Ac_k and As_k figuring in the fitting function (10) for weekly distribution of mean prices K=50. (b) The distribution of amplitudes Ac_k and As_k figuring in the fitting function (10) for monthly distribution of mean prices. One can notice that behavior of these coefficients is different. It appears that these distributions of amplitudes reflect the specific behavior of the fitting curve corresponding to the chosen temporal interval. The similar plots calculated for max(Pr) and min(Pr) are omitted.(c) The distribution of amplitudes Ac_k and As_k figuring in the fitting function (10) for half-year distribution of mean prices $K = 50$. (d) The distribution of amplitudes Ac_k and As_k figuring in the fitting function (10) for annual distribution of mean prices. The similar plots showing the distribution of the corresponding amplitudes for max(Pr) and min(Pr) are omitted.

their strong deviation with the increasing degree of compression. One can remind that the degree of compression between the initial weekly files and final annual files is high and equaled 51. Figures 3 (a-d) depict the distribution of the coefficients Ac_k, As_k entering into decomposition (10) for all season fluctuations. Namely, this set of coefficients reflects the specific behavior of fluctuations for the calculated temporal intervals. Comparing of these plots with each other one can notice that the minimal contribution of these amplitudes takes place for the weekly fluctuations (the most detailed description of season prices) while for annual fluctuations the values of these amplitudes are increasing and their contribution to the fit are becoming more significant.

Table 1. The calculated values of some important parameters entering into the final hypothesis (10). They describe the final solution of the inverse problem. For all files that are fitted by function (10) we put K=50.

Number of file	$\langle\alpha\rangle$	b	ξ	A_0	c	RelErr.(fn)	PCC
Week Mean(Pr)	1.35274	38.4699	14.852	0.00339	−23.2729	0.36008	0.99165
Month Mean(Pr)	1.37402	39.2809	14.462	0.02167	−23.3654	0.57704	0.99466
Max(Pr)	1.33774	37.6529	15.0644	0.02857	−23.0259	0.57794	0.99462
Min(Pr)	1.4006	42.6653	14.5829	0.01719	−23.9377	0.62807	0.99365
Half-year Mean(Pr)	1.48733	40.9613	12.1359	−0.00648	−22.6892	4.0996E-5	1
Max(Pr)	1.28387	39.9092	17.663	0.30374	−23.5898	4.68858E-5	1
Min(Pr)	2.10419	52.6053	6.57516	−0.04637	−24.0724	1.55003E-5	1
Year Mean(Pr)	1.90892	50.8186	7.82904	0.04069	−19.9979	5.04593E-6	1
Max(Pr)	1.94895	46.0996	7.13898	−0.33502	−15.1736	2.05921E-6	1
Min(Pr)	2.87562	24.941	3.0604	−1.01543	−21.743	4.17339E-6	1

Comments to Table 1. The strong deviations of these parameters between the limiting season values are related to the influence of random fluctuations. The same remark for the variations of the parameters b and ξ. We can calculate the mean and Stdev of these parameters and find their mean values. But we decided to keep these values as they are in order find the true reason of their strong deviation with increasing of degree of their compression. One can remind that the degree of compression between the initial weekly files and final annual files is of 51.

4 Results and discussion

In this paper we proposed a method for solving the inverse problem associated with fitting of solutions of scaling equations for self-similar random curves. The basic results can be formulated as follows:

1. The initial fitting hypothesis (expressed by formula (2)) is suggested. It can be transformed to expression (5) that allows the application of the LLSM for calculation of the values of nonlinear power-law exponents. As an initial justification of the hypothesis (2) we expect that the calculated values of the inoculating exponents $\alpha_{1,2}$ should have the same sign and should be close to each other.
2. Optimization procedure based on the minimal value of the relative error (8) with respect to the variation parameter v was developed. This procedure helps finding the optimal value of the power-law exponent $< \alpha >$.
3. The inoculating fitting parameter ω_1 found from hypothesis (2) allows finding the desired values of b and ξ from expression (11) and actually to realize the final fit of the expression (12) into the real data (based again on the application of the LLSM).
4. It is interesting to mark also that for self-similar economic data there is a possibility to replace a simple linear hypothesis (that in many cases is used for describing the unknown trend) by more accurate curve (4) or (9) that describes the behavior of the trend and thereby to receive more reliable results associated with their prediction in the future.

In conclusion we should make also the following remark. Naturally, when the values of the calculated exponents $\alpha_{1,2}$ have the opposite signs then it is necessary to increase the limits of applicability of equation (1) and to consider more complicated solutions of the scaling equation (12). The complete solutions of this scaling equation containing two-power law exponents, or even more, are given in [6]. The consideration of this case is not trivial and can constitute the subject of the separate research. We expect that that the proposed algorithm can be explored with other classes of data series if condition $\alpha_1\alpha_2 > 0$ for the initially calculated power-law exponents is satisfied.

References

[1] Sornette, D. (1998), Discrete scale invariance and complex dimensions, *Physics Reports* **297**, 239–270.
[2] Voigt, J. (2005), *The Statistical Mechanics of the Financial Markets*, (3-rd edition) Springer-Verlag. Berlin-Heidelberg.
[3] Nigmatullin, R.R. (1998), Eigen-coordinates-new method of analytical functions identification in experimental measurements, *Journal of Applied Magnetic Resonance*, **14**, 601–633.
[4] Nigmatullin, R.R. (2000), Recognition of nonextensive statistical distributions by the eigencoordinates method, *Physica A*, **285**, 547–565.
[5] Nigmatullin, R.R. (2012), Is it possible to replace the probability distribution function by its Prony's spectrum? *Journal of Applied Nonlinear Dynamics*, **1**(2), 173–194.
[6] Nigmatullin, R.R. and Baleanu, D. (2012), The derivation of generalized functional equations describing the self-similar processes, *Fractional Calculus and Applied Analysis*, **15**(4), 718-740; DOI: 10.2478/s13540-012-0049-5.
[7] Duarte, F.B., Machado, J.T., and Duarte, G.M. (2010), Dynamics of the Dow Jones and the NASDAQ stock indexes, *Nonlinear Dynamics*, **61**(4), 691–705.
[8] Machado, J.T., Duarte, F.B., and Duarte, G.M. (2011), Analysis of stock market indices through multidimensional scaling, *Communications in Nonlinear Science and Numerical Simulation*, **16**(12), 4610–4618.