The Model of Real Data Constructing Using Fractional Brownian Motion

Valeria Bondarenko and Victor Bondarenko

Abstract

In this paper we investigate the properties of the fractional Brownian motion as a basic process of stochastic time series models. New method of estimating the Hurst exponent is substantiated. Stochastic model, which is representing a time series analysis in the form of increments converted to fractional Brownian motion. The method of checking the adequacy of the proposed models. The research results are implemented in software for the simulation and analysis temporal data.

Keywords: Stochastic model; Fractional Brownian motion; Estimation of parameters

Introduction

Let \( x(t), t \in [0, T] \) is an observed trajectory, which is describing the stochastic evolution of some dynamic object. Mathematical model of this trajectory is defined as a random process, \( \xi(t) \). Where: \( x(t) = X(t), X(\cdot) \) is realization of the process \( \xi \).

As a rule, we chose as a model random process with known characteristics. Direct use of this definition requires broad classes of these processes. On the other hand, this class includes Gauss and Markov processes. Let’s introduce another definition of continuous mathematical models for the observed trajectory \( x(\cdot) \in C(0; T) \) using nonlinear conversion.

Definition

Mathematical model of observed trajectory \( x(t) \) is a pair \((\Phi, \xi)\), where \( x(t) = \Phi(X(t)) \), \( \xi(t) \) is a random process with known characteristics, \( \Phi \) is a reversible conversion in \( c(0, T) \).

Let’s assume \( \xi(t) = \xi_{t_1} \ldots \xi_{t_n} = \xi_{t_1, t_2, \ldots, t_n} = (t_2 - t_1) \ldots (T - 0) n, \) \( x(t) = \Phi(X(t)) \), \( t \) is a model of observed time series \( x_1, x_2, \ldots, x_n \).

Let’s call \( \xi \) as a basic process of the model. Levy processes with independent stationary increments have been considered as basic for models of time series (particularly financial) [1-4]. The next step in the development of the models is transition to diffusion processes. For example, diffusion model of stock price \( S(t) \) is obtained from the following considerations:

\[
S(t) = S(0) \exp(\sigma \sqrt{t} + \mu t), \quad X(t) = \ln S(t),
\]

Where \( \omega(t) \) is a standard Wiener process, \( \sigma \) is volatility and interest rate, \( \mu \) is a constant. Then let’s propose the equation:

\[
dS(t) = \sigma S(t) \omega(t) + \mu S(t) dt
\]

Which can be interpreted as a stochastic equation Ito and its solution could be written as a geometric (economic) Brownian motion:

\[
S(t) = S(0) \exp\left\{ \sigma \omega(t) + \left( \mu - \frac{\sigma^2}{2} \right) t \right\}
\]

For the model \( (1) \) of a stock price have been obtained a number of known results, including the Black-Scholes formula for a rational option pricing [5-8]. The main drawback of Levy processes (and diffusion) is a priori satisfies only the simplest physical phenomena. The absence of impact on processes in biology, economics, climate, etc. looks unconvincing. In this paper we propose a non-Markovian model of the time series.

Selection of the Base Process and its Properties

One of the most popular Markov models of time series is Gaussian random process, and fractional Brownian motion [9-11]. The demand of this process is caused by “convenient” properties, which are described below.

Fractional Brownian motion is defined as a Gaussian random process with characteristics:

\[
B_H(t), \quad E B_H(t) = 0, \quad E B_H(0) = 0, \quad E B_H(t) B_H(s) = \frac{1}{2} (t^{2H} + s^{2H} - |t-s|^{2H})
\]

Note that with \( H = 0.5 \) we get a standard Wiener process.

Smoothness of the trajectories of the process \( B_H(t) \) is defined by the parameter \( H \): almost all the trajectories satisfy the Holder condition:

\[
[X(t) - X(s)] \leq c |t-s|^\alpha, \quad \alpha < H,
\]

This generalizes known Levy’s result for the Wiener process.

The increments of fBM \( B_H(t_2) - B_H(t_1), \quad B_H(t_3) - B_H(t_1), \quad t_1 < t_2 < t_3 < t_4 \) form a Gaussian random vector with a correlation between the coordinates:

\[
\frac{1}{2} (t_4 - t_1)^{2H} + (t_3 - t_2)^{2H} - (t_4 - t_2)^{2H} - (t_3 - t_1)^{2H}
\]

For discrete time:

\[
\xi_k = B_H \left( \frac{k}{n} \right) - B_H \left( \frac{k-1}{n} \right)
\]

We obtain the correlation coefficient:

\[
P \{ X(t_n) \in \Delta | X(t_{n-1}) = a_{n-1} \} =
\]
Let’s introduce the notation:

\[ \hat{H}_n = \frac{\ln \left( \frac{\sigma}{\sqrt{\pi n}} \right)}{\ln n} \]

With known \( \sigma \):

\[ \hat{\sigma}_n = n^H \sqrt{\frac{n}{2}} R_n = 1.25n^H R_n \]

(6)

Let’s propose a new method of estimation Hurst exponent [16-18].

Let’s introduce the notation:

\[ \hat{Q}(H) = 0.8 \frac{R_n}{\sqrt{n}} (S^{-1}y,y) \]

Where matrix \( S = S_H \) is defined.

**Statement**

Statistic

\[ \hat{H} = \arg \min_{H} \| \hat{Q}(H) - H \| \]

Is a consistent estimator of the parameter H.

**Proof**

\( E \) is the canonical Gaussian vector with the following characteristics:

\[ E \varepsilon = 0, E(\varepsilon,u)(\varepsilon,v) = (u,v) \dim \varepsilon = n. \]

Then

\[ y = V^2 \varepsilon, \] therefore

\[ n = E(\varepsilon, \varepsilon) = E(V^{-1}y, y) = \frac{n^{2H}}{\sigma^2} E \left( (S^{-1}y, y) \right) \]

And consequently the statistic

\[ \hat{\sigma}_2 = (n)^{2H-1}(S^{-1}y, y) \]

And here statistics \( (n)^{2H-1}(S^{-1}y, y) \) is an unbiased estimate of the parameter \( \sigma^2 \). Let’s introduce:

\[ \hat{\sigma}_n^2 = n^{2H-1} \left( (S^{-1}y, y) \right) \]

(7)

With calculating the dispersion

\[ D\hat{\sigma}_2 = n^{4H-2}E(S^{-1}y, y)^2 - \sigma^4 \]

Use the formula for integration by parts for the Gaussian measure [19] and get

\[ D\hat{\sigma}_2 = 2\sigma^4 \rightarrow 0, \quad n \rightarrow \infty \]

\[ \hat{\sigma}_2 \]

Is a consistent parameter estimation of \( \sigma^2 \).

The equalities (6,7) are form the system, from which follows the relation:

\[ \frac{\hat{\sigma}_n}{\sigma} = 0.8 \frac{R_n}{\sqrt{n}} (S^{-1}y,y) \approx 1, \quad \left| \frac{\hat{\sigma}_n}{\sigma} - 1 \right| \rightarrow \min \]

This proves the statement. The efficiency of proposed estimation method has been tested by numerical experiment [16].

**The limit theorems for some statistics**

The limit theorems for statistics from increments of fractional Brownian motion have been proved in works of Nourdin 1 and others [20-23].
The general idea of approximation is an one-dimensional functional transformation \( g \) of each increment \( y_k \), where \( B \) is an increasing odd function, 

\[
z_k = g(y_k)
\]

Let’s assume

\[
\lim_{n \to \infty} \frac{1}{n-1} \sum_{i=1}^{n} [g(y_i)]^2 = \frac{2}{\pi}
\]

Where \( z_i = g(y_i) \) is assumed as a Gaussian random value. Let’s demonstrate proposed algorithm with \( g \) as a power function. Assume

\[
z_k = \text{sgn} y_k |y_k|^\lambda = |y_k|^\lambda
\]

\[
y_k = \text{sgn} y_k |y_k|^\lambda
\]

Then

\[
d_n = \frac{1}{n-1} \sum_{i=1}^{n} |y_i|^\lambda
\]

\[
d = \lim d_n
\]

Is equal to ratio of the corresponding mathematical expectations. For \( \xi \sim N(0, \sigma^2) \),

\[
E|\xi|^{\lambda} = \sqrt{\frac{\sigma^2}{\Gamma(\frac{\lambda+1}{2})}}
\]

\[
\lambda = \frac{\Gamma(\frac{\lambda+1}{2})}{\frac{\Gamma(\lambda/2)}{\sqrt{\pi}}}
\]

Where the parameter \( \lambda \) is defined from the equation. Thus, the proposed approximation leads to the following model of original time series:

\[
x_k = \sum_{j=1}^{n} \text{sgn} y_j |z_j|^\lambda
\]

If we’ll assume that values of sequence \( |z_1, z_n| \) are increments of \( \xi_{\text{Bm}} \), let’s calculate Hurst exponent by the following algorithm, which shows proposed method:

1) Construct the statistic:

\[
R_1 = \frac{1}{n} \sum_{k=1}^{n} |y_k|
\]

2) Calculate the matrix \( S_H^{-1} \), where \( S_H \) is a correlation matrix of increments:

\[
s_{jk} = E(B_n(\frac{k}{n}) − B_n(\frac{j}{n})) / (B_n(\frac{k}{n}) - B_n(\frac{j}{n})) = \varrho (\xi_k, \xi_j)
\]

\[
Q = \frac{0.8}{R_1} \sum_{\xi=1}^{n} |\xi|^\lambda
\]

The statistics \( Q \) is calculating for difference values of Hurst exponent with step (0.0, 0.5-1) and:

\[
Q(H - |\hat{H} - arg min \{Q(H) - |\}|)
\]

3) The testing of hypothesis \( \text{H} = (\text{statistics} \ z_1,\ldots, z_n) \) which obtained by transformation (11) of real data are simulated by increments from fractional Brownian motion. The algorithm with known \( H \) is the following. Denote

\[
c = \frac{1}{n} \sum_{k=1}^{n} z_k^2
\]

And assume that hypothesis \( \text{H} \) is done

\[
z_k = \sqrt{c} \xi_k = \sqrt{c} n^H \left( B_n \left( \frac{k+1}{n} \right) - B_n \left( \frac{k}{n} \right) \right)
\]

Assume \( y_k = \frac{z_j}{\sqrt{c}} \) and construct the statistics

\[
A_n = \frac{1}{n} \sum_{k=1}^{n} \frac{y_k^3}{\sqrt{c}^3}, \ H \in \left( \frac{0}{2}, \frac{1}{2} \right)
\]
The Model of Real Data Constructing Using Fractional Brownian Motion

The third example: The oscillation of waves in the North Atlantic, 10.1980-10.2014 409 data (Figures 3 and 4).

The Comparative Analysis of Used Models

Let’s compare the time series model (approximation of fractional Brownian motion) with known models and estimate the quality of modeling. Note that the choice of the quality criterion is dependent from the type of model.

The values of exchange rate, Banque de France. Let’s compare the effectiveness of approximation method with other models for real 1020 data [21].

For modeling of selected values are used these models:

- Autoregression,
- Autoregression with moving average (ARMA) \((p,q)\),
- Autoregressive with integrated moving average \((p,d,q)\),
- Autoregressive moving average (ARMA) \((p,q)\).

These methods have been selected, because after using the special tests for statistical data, we’ve revealed high autocorrelation value and existence of a trend.

Based on analysis of values of the constructed partial autocorrelation and autocorrelation function of data series, the order AR (1) model may be in the range from 1 to 5. The model AR (1) is given:

\[ y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) = 0.101 + 0.908 y(k-1) \]  

Where \( y(k) \) is a basic variable; \( \varepsilon(k) \) is a random process. Characteristics of the adequacy and quality for short-term forecasts for the training sample had the following values:

\[ R^2 = 0.816, \sum c_i^2 = 52.591, DW = 1.957, Mard = 10.71, U = 0.069 \]

Some deterioration of forecasting is obtained by expansion of the order of autoregression for two:

\[ y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) = 1.176 + 1.0047 y(k-1) - 0.111 y(k-2) \]  

The Real Examples

Let’s consider examples of real nature:

The first example: the monthly data of market rate of the Bundesbank (Germany) (http://www.bundesbank.de) for 2003-2012 (120 data) (Figure 1).

The second example: 1020 data of exchange rate EUR / USD for 2011-2014 (http://www.banque-france.fr) (Figure 2).

\[ B_n = \frac{1}{n^{1/2}} \sum v_k^2 z_k^3, \quad H \in \left( 0, \frac{1}{2} \right); \]
\[ D_n = \frac{1}{n^{1/2}} \sum v_k z_k^3, \quad H \in \left( \frac{1}{2}; 1 \right). \]  

If hypothesis \( T \) is true, there is convergence:

\[ A_n \rightarrow -\frac{3}{2} \sigma^2; \quad B_n \rightarrow 3c^2 \eta; \quad D_n \rightarrow \frac{3}{2} c^2 B^2. \]  

Let’s consider examples of real nature:

The first example: the monthly data of market rate of the Bundesbank (Germany) (http://www.bundesbank.de) for 2003-2012 (120 data) (Figure 1).

The second example: 1020 data of exchange rate EUR / USD for 2011-2014 (http://www.banque-france.fr) (Figure 2).
\[ R^2 = 0.820, \sum \varepsilon^2 = 50.771, DE=1.957, MARD=11.197, U = 0.071. \]
\[ y(k) = a_0 + a_1y(k - 1) + a_2y(k - 5) + a_kk + a_3k^2 + a_4k^3 + \varepsilon(k) = \]
\[ = 3.957 + 0.743y(k - 1) + 0.005y(k - 5) - 0.035k + 0.00036k^3 - 1.24E - 0.6k^4 \]  
(17)
\[ R^2 = 0.844, \sum \varepsilon^2 = 43.842, DW = 1.720, \hat{\text{MARD}} = 5.783, \]
\[ \hat{\text{MARD}} = 5.783, U = 0.036 \]

Thus, mean absolute relative difference has been reduced from 11.20% to 5.78% after introducing the trend in model. Let’s construct a model with autoregressive moving average:
\[ y(k) = a_0 + a_1y(k - 1) + a_2y(k - 5) \]
\[ b_1 \varepsilon(k - 1) + b_2 \varepsilon(k - 2) + \varepsilon(k) = 10.82 + 0.757y(k - 1) + 0.134y(k - 5) + 0.209\varepsilon(k - 1) + 0.271\varepsilon(k - 2) \]

(18)

However, its characteristics aren’t better than in the previous case (the model with the trend), except for Durbin-Watson statistic (Table 1):
\[ R^2 = 0.832, \sum \varepsilon^2 = 4.6931, DW = 1.943, MARD=10.611, U = 0.068 \]

Thus, the best is a structural model of the process from all constructed mathematical models, which takes into account explicitly the trend of process and vibrations (MARD=5.78%) [17-19]. This is quite a logical result, because the structural models are describing these processes with a higher degree of adequacy than others. As expected, the introduction of a moving average model didn’t improve the quality as compared with a simple model AR (1). The value of the Durbin-Watson statistic for the approximation model is more closer to 2 than...
For all examples the approximation has antipersistent character (H<0.5) and it's adequate, if the conditions are satisfied [14].

**Figure 4:** The block diagram.

<table>
<thead>
<tr>
<th>Real data</th>
<th>$\hat{H}$</th>
<th>$A_0$</th>
<th>$B_0$</th>
<th>$A$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>The monthly data of Bundesbank</td>
<td>0.4</td>
<td>-19.1</td>
<td>-0.8</td>
<td>-16.3</td>
<td>58.3</td>
<td>—</td>
</tr>
<tr>
<td>Exchange rate €/$, 2011-2014, Banque de France</td>
<td>0.4</td>
<td>-1.72</td>
<td>-1.61</td>
<td>-1.51</td>
<td>5.28</td>
<td>—</td>
</tr>
<tr>
<td>The oscillation of waves in the North Atlantic</td>
<td>0.1</td>
<td>-3.44</td>
<td>-6.38</td>
<td>-3.575</td>
<td>9.883</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 2: The values of control statistics and parameters of limit distributions for Examples 1-3.

**References**