UDC 519.234 DOI: 10.18698/1812-3368-2020-4-42-57

ROBUST IDENTIFICATION OF AN EXPONENTIAL AUTOREGRESSIVE MODEL

A.V. Goryainov[1] agoryainov@gmail.com
V.B. Goryainov[2] vb-goryainov@bmstu.ru
W.M. Khing[1] waimyokhing@gmail.com

[1] Moscow Aviation Institute (National Research University), Moscow, Russian Federation
[2] Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

One of the most common nonlinear time series (random processes with discrete time) models is the exponential autoregressive model. In particular, it describes such nonlinear effects as limit cycles, resonant jumps, and dependence of the oscillation frequency on amplitude. When identifying this model, the problem arises of estimating its parameters — the coefficients of the corresponding autoregressive equation. The most common methods for estimating the parameters of an exponential model are the least squares method and the least absolute deviation method. Both of these methods have a number of disadvantages, to eliminate which the paper proposes an estimation method based on the robust Huber approach. The obtained estimates occupy an intermediate position between the least squares and least absolute deviation estimates. It is assumed that the stochastic sequence is described by the autoregressive equation of the first order, is stationary and ergodic, and the probability distribution of the innovations process of the model is unknown. Unbiased, consistency and asymptotic normality of the proposed estimate are established by computer simulation. Its asymptotic variance was found, which allows to obtain an explicit expression for the relative efficiency of the proposed estimate with respect to the least squares estimate and the least absolute deviation estimate and to calculate this efficiency for the most widespread probability distributions of the innovations sequence of the equation of the autoregressive model

Keywords

Exponential autoregression, robust estimate, consistency, asymptotic normality, asymptotic relative efficiency

Received 09.10.2019
Accepted 13.12.2019
© Author(s), 2020

Introduction. The exponential autoregression model is one of the most popular of time series models (random processes with discrete time) [1]. The advantage of this model lies in the possibility of obtaining, with its use, a description of the nonlinear effects of a number, in particular, limit cycles, jumps resonance, and the dependence of the oscillation frequency on the amplitude, which is impossible in the framework of the linear autoregressive model. The exponential autoregressive model has proven itself in technology [1], economics [2], climatology [3], oceanology [4], biology [1].

The main task in identifying an exponential autoregressive model is to evaluate its parameters — the coefficients of the autoregressive equation that describes this model. The most common methods for estimating coefficients are the least squares method and the least absolute deviation method. These methods have more than two hundred years of history and are well studied for linear models. In particular, if time series observations are Gaussian (normal) random variables, then the least squares method gives the best results. If observations of the time series were made with large measurement errors, then the least absolute deviation method is more effective. In the second half of the 20th century, the M-estimates method was developed, which includes the advantages of both the least squares method and the least modulus method. For linear models, the M-estimates almost not inferior to the least squares estimate in the Gaussian case and surpasses it even with a small deviation of the probability distribution of observations from the Gaussian. The M-estimate is almost always better than the least absolute deviation estimates, second only to probability distributions with the so-called heavy tails, which is typical for observations obtained with gross measurement errors.

For nonlinear models, a comparison of the above three methods is poorly understood. Separate results were obtained for threshold autoregression and autoregression with random coefficients [5, 6]. In the present work, a comparative study of these methods is carried out by computer simulation by evaluating the parameters of exponential first-order autoregression.

Problem statement. The time series X_t, $t = 0, \pm 1, \pm 2, \ldots,$ described by the first-order model of exponential autoregression satisfies the recurrence equation

$$X_t = \left(a_0 + b_0 e^{-c_0 X_{t-1}^2} \right) X_{t-1} + \varepsilon_t. \tag{1}$$

The coefficients a_0, b_0, c_0 of equation (1) are real numbers and are model parameters. The updating process ε_t, $t = 1, 2, \ldots,$ is a sequence of independent identically distributed random variables with zero expectation function

$E\varepsilon_t = 0$ and finite variance $D\varepsilon_t = E\varepsilon_t^2 = \sigma^2 < \infty$. Suppose that model (1) is stationary and ergodic. A sufficient condition for this is, for example, the simultaneous fulfillment of conditions $|a| < 1$, $c > 0$ and $\sigma^2 < \infty$ [7].

Model (1) is an example of smoothing another popular nonlinear model — a threshold autoregressive model [8] of the form

$$X_t = \begin{cases} a_0 X_{t-1} + \varepsilon_t, & \text{if } |X_{t-1}| > C; \\ (a_0 + b_0) X_{t-1} + \varepsilon_t, & \text{if } |X_{t-1}| \le C, \end{cases}$$

where $C > 0$ is some threshold constant. Indeed, if $|X_{t-1}|$ it is large, then the value $e^{-c_0 X_{t-1}^2}$ is close to zero; therefore, the right-hand side of (1) is practically indistinguishable from $a_0 X_{t-1} + \varepsilon_t$. As the $|X_{t-1}|$ decreases, the role of the coefficient b_0 increases.

Consider the problem of estimating the parameters (a_0, b_0, c_0) of equation (1) from observations $X_1, X_2, \ldots, X_n$ of a process X_t, satisfying this equation. We define the M-estimates of the coefficients (a_0, b_0, c_0) as the minimum point $(\hat{a}, \hat{b}, \hat{c})$ of the function

$$g(a,b,c) = \sum_{t=2}^{n} \rho\left(X_t - \left(a + b e^{-c X_{t-1}^2} \right) X_{t-1} \right)^2, \tag{2}$$

where ρ is the so-called ρ-function. It is usually assumed that ρ is an even and downward-convex function. The least squares and the least absolute deviation estimates are a special case of the M-estimates for $\rho(x) = x^2$ and $\rho(x) = |x|$, respectively.

The most common ρ-functions are [9] ρ-Huber function

$$\rho_H(x) = \begin{cases} x^2, & \text{if } |x| \le k; \\ 2k|x| - k^2, & \text{if } |x| > k, \end{cases} \tag{3}$$

and the ρ-Tukey function

$$\rho_H(x) = \begin{cases} 1 - \left(1 - \left(\dfrac{x}{k} \right)^2 \right)^3, & \text{if } |x| \le k; \\ 1, & \text{if } |x| > k. \end{cases} \tag{4}$$

Here $k \in (0, \infty)$ is the tuning parameter, the change of which allows to achieve the maximum efficiency of the M-estimates, depending on the specific type of probability distribution density function f of the members ε_t of the updating process.

According to (2), (3) and the curves shown in Fig. 1, the M-estimates with the ρ-Huber function is a compromise between the least squares estimate and the least absolute deviation estimate. Since $\rho_H(x)$ it coincides with x^2 in the neighborhood $(-k, k)$ of the origin and behaves linearly outside this neighborhood, similar to $|x|$, the large residuals $X_t - \left(a + be^{-cX_{t-1}^2}\right)X_{t-1}$, generated by large errors in the observations (large perturbations ε_t), they affect the target function $g(a,b,c)$ in a linear rather than quadratic manner, thereby reducing the effect of these large errors on the minimum $g(a,b,c)$, and on the accuracy of parameter estimation. The ρ-Tukey function, which ignores large residuals (larger in magnitude than k), ignores it even more, reduces the contribution of sharply distinguished observations, replacing them with unity.

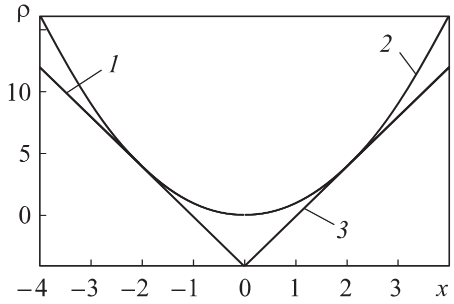

Fig. 1. ρ-Huber function:
1 — $\rho(x)$; *2* — x^2; *3* — $2k|x| - k^2$

The purpose of the work is to study the accuracy of the M-estimates with the ρ-Huber function depending on the probability distribution of the members of the update sequence ε_t.

Simulation studies of the properties of the M-estimates. At the first stage, the *non-bias and consistency* of the M-estimates were studied by computer simulation. For definiteness, it was assumed that $a_0 = -0.3$, $b_0 = -0.8$, $c_0 = 1$, the sample size n varied from 100 to 800 in increments of 100. Random values ε_t, $t = 1, \ldots, n$, were modeled $N = 1{,}000$ once using *MATLAB* random number generator simulating the normal, logistic, and double exponential probability distributions. Based on the generated values $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in}$, using the recurrence relation (1) with a zero initial condition $X_0 = 0$, the realizations $x_{i1}, x_{i2}, \ldots, x_{in}$, $i = 1, 2, \ldots, N$, of the time series $x_{i1}, x_{i2}, \ldots, x_{in}$, $i = 1, 2, \ldots, N$, observations were calculated. For each realization of the observations $x_{i1}, x_{i2}, \ldots, x_{in}$, $i = 1, 2, \ldots, N$, the realization was found $\left(\hat{a}_i, \hat{b}_i, \hat{c}_i\right)$, an M-estimates $\left(\hat{a}, \hat{b}, \hat{c}\right)$ of the parameters (a_0, b_0, c_0), was determined, which was determined as the minimum point of the objective function $g(a,b,c)$ of the form (2). As ρ-functions in (2), ρ-functions (3) and (4) were used. The function $g(a,b,c)$ was minimized using the Levenberg — Marquardt optimization algorithm [10], the essence of which is a combination of the Newton method with the gradient descent method.

To study the probabilistic properties of the M-estimates, its expectation function $\left(\mathrm{E}\hat{a}, \mathrm{E}\hat{b}, \mathrm{E}\hat{c}\right)$ was approximated by averaging the realizations $\left(\hat{a}_i, \hat{b}_i, \hat{c}_i\right)$

by $i=1,2,\ldots,N,$ over the vector $(\overline{\hat{a}},\overline{\hat{b}},\overline{\hat{c}})=\left(n^{-1}\sum_{i=1}^{N}\hat{a}_i,\ n^{-1}\sum_{i=1}^{N}\hat{b}_i,\ n^{-1}\sum_{i=1}^{N}\hat{c}_i\right)$, since it follows from the law of large numbers that $(\overline{\hat{a}},\overline{\hat{b}},\overline{\hat{c}})\to(\mathrm{E}\hat{a},\mathrm{E}\hat{b},\mathrm{E}\hat{c})$ at $N\to\infty$, by definition, the non-bias of the estimate means that $(\mathrm{E}\hat{a},\mathrm{E}\hat{b},\mathrm{E}\hat{c})=(a_0,b_0,c_0)$. Thus, under the condition of non-bias, the difference $(\overline{\hat{a}},\overline{\hat{b}},\overline{\hat{c}})-(a_0,b_0,c_0)$ in the simulation should take values close to zero.

The errors $\delta=\overline{\hat{b}}-b_0$ of estimation of the second coordinate b_0 for $n=100,200,\ldots,800$ the double exponential probability distribution of random variables ε_t and ρ-function (3) are given in the table. Small values of the δ allow us to conclude that the estimate $\hat{b}$ is not biased. For estimates of the other two coordinates obtained including with the ρ-function (4) and for other probability distributions of the updating sequence ε_t, the simulation results are similar. This makes it possible to draw a similar conclusion about the non-bias of M-estimates $\hat{a}$ and $\hat{c}$ of parameters, a_0 and c_0, respectively.

The bias δ and approximation $\overline{\Delta}$ of the second moment of parameter *b* estimation for various values of *n* of the sample size

n	100	200	300	400	500	600	700	800
δ	−0.00262	−0.0203	0.00247	−0.00431	0.00828	−0.001832	−0.000646	−0.00227
$\overline{\Delta}$	0.0802	0.0399	0.0272	0.0198	0.0157	0.0132	0.0112	0.0105

The validity of the M-estimates $(\hat{a},\hat{b},\hat{c})$ was verified by approximating the second moments $(\mathrm{E}(\hat{a}-a_0)^2,\mathrm{E}(\hat{b}-b_0)^2,\mathrm{E}(\hat{c}-c_0)^2)$ and variances $(\mathrm{D}\hat{a},\mathrm{D}\hat{b},\mathrm{D}\hat{c})$ of its coordinates. Consistency by definition means that $(\hat{a},\hat{b},\hat{c})\to(a_0,b_0,c_0)$, in probability $n\to\infty$. From the Chebyshev inequality it follows that for consistency, for example $\hat{b}$, it is enough that if a $\Delta=\mathrm{E}(\hat{b}-b_0)^2\to 0\ n\to\infty$, taking into account the established nonbias of $\hat{b}$, it is enough if $n\to\infty$ the condition $\mathrm{D}\hat{b}\to 0$ is fulfilled. According to the law of large numbers, $\overline{\Delta}=n^{-1}\sum_{i=1}^{N}(\hat{b}_i-b_0)^2$ tends to Δ when $N\to\infty$, therefore, by the behavior of $\overline{\Delta}$, one can judge the validity of the $\hat{b}$ estimation. The table also shows the values $\overline{\Delta}$ for $n=100,200,\ldots,800$. With increasing n values $\overline{\Delta}$ decrease, approaching zero, which indicates the validity of the assessment $\hat{b}$ estimation.

Usually, in mathematical statistics, the rate of convergence of a consistent estimate is proportional $n^{-1/2}$. Such estimates are called $\sqrt{n}$-consistent. To test the hypothesis that the rate of convergence of the M-estimates to the estimated parameter is proportional to $n^{-1/2}$ the expression $n^{-1/2}$ as a function of n was approximated according to the table by a polynomial of the first degree. Least squares obtained

$$\sqrt{n}\Delta \approx 1.0699 - 0.0032683n + 4.9245 \cdot 10^{-6} n^2 - 2.5734 \cdot 10^{-9} n^3,$$

which confirms the hypothesis of $\mathrm{D}\hat{b} \sim \mathrm{const}\, n^{-1/2}$. Similarly established $\sqrt{n}$-consistency of estimates $\hat{a}$ and $\hat{c}$.

At the second stage, the *asymptotic normality of the M-estimates* was studied by computer simulation. For definiteness, it was assumed that $a_0 = -0.3$, $b_0 = -0.8$, $c_0 = 1$, are random variables of ε_t have a double exponential distribution of probability, the ρ-function is described (3). Figure 2 shows the histogram constructed from the results of 10,000 simulated implementations $\hat{b}$. The symmetric and bell-shaped histograms suggest that the probability distribution of a random variable $\hat{b}$ is close to normal.

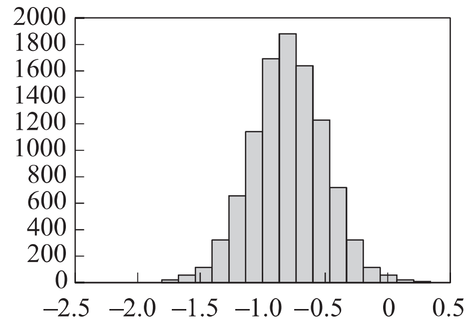

Fig. 2. Histogram of M-estimate of $\hat{b}$

Testing the hypothesis of normality of the $\hat{b}$ estimate for 10,000 implemen-tations using the χ^2-test allowed us to accept it at a significance level of 0.001. A similar result was obtained for estimates $\hat{a}$ and $\hat{c}$.

Asymptotic properties of the M-estimates. The exact probability distribution of estimates is very difficult to find and is possible only in the simplest cases. It is usually possible to establish the asymptotic distribution of the estimate, i.e., the distribution to which the probability distribution of the estimate converges (weakly, by distribution) with an unlimited increase in the number of observations. As a rule, such a limiting distribution turns out to be normal, and the estimates in this case are called asymptotically normal. The results obtained above suggest that the M-estimates $\left(\hat{a}, \hat{b}, \hat{c}\right)$ is asymptotically normal. Below we will justify the asymptotic normality of the M-estimates $\left(\hat{a}, \hat{b}, \hat{c}\right)$ with ρ-functions (3), (4).

Having expanded the objective function $g(a,b,c)$ of the form (2) according to the Taylor formula at a point (a_0, b_0, c_0) up to the second order inclusive, we obtain

$$g(a,b,c)=g(a_0,b_0,c_0)+A^T\theta+\frac{1}{2}\theta^T B\theta+\delta(a,b,c),$$

where

$$A=\frac{1}{\sqrt{n}}\left(\frac{\partial g(a_0,b_0,c_0)}{\partial a},\ \frac{\partial g(a_0,b_0,c_0)}{\partial b},\ \frac{\partial g(a_0,b_0,c_0)}{\partial c}\right)^T=$$

$$=\left(-\sum_{t=2}^{n}\rho'(\varepsilon_t)X_{t-1},\ -\sum_{t=2}^{n}\rho'(\varepsilon_t)X_{t-1}e^{-c_0X_{t-1}^2},\ \sum_{t=2}^{n}\rho'(\varepsilon_t)X_{t-1}^3 b_0 e^{-c_0X_{t-1}^2}\right);$$

$$\theta=\left(\sqrt{n}(a-a_0),\sqrt{n}(b-b_0),\sqrt{n}(c-c_0)\right);$$

$$B=\frac{1}{n}\begin{pmatrix}\dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial a^2} & \dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial a\partial b} & \dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial a\partial c}\\ \dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial a\partial b} & \dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial b^2} & \dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial b\partial c}\\ \dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial a\partial c} & \dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial b\partial c} & \dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial c^2}\end{pmatrix};$$

$\delta(a,b,c)$ is an infinitesimal function of a higher order when $(a,b,c)\to$ $\to(a_0,b_0,c_0)$ compared with $(a-a_0)^2+(b-b_0)^2+(c-c_0)^2$. The matrix B elements have the form

$$\frac{\partial^2 g(a_0,b_0,c_0)}{\partial a^2}=\sum_{t=2}^{n}\rho''(\varepsilon_t)X_{t-1}^2,\qquad \frac{\partial^2 g(a_0,b_0,c_0)}{\partial a\partial b}=\sum_{t=2}^{n}\rho''(\varepsilon_t)X_{t-1}^2e^{-c_0X_{t-1}^2},$$

$$\frac{\partial^2 g(a_0,b_0,c_0)}{\partial a\partial c}=-\sum_{t=2}^{n}\rho''(\varepsilon_t)X_{t-1}^4 b_0 e^{-c_0X_{t-1}^2},$$

$$\frac{\partial^2 g(a_0,b_0,c_0)}{\partial b^2}=\sum_{t=2}^{n}\rho''(\varepsilon_t)X_{t-1}^2e^{-2c_0X_{t-1}^2},$$

$$\frac{\partial^2 g(a_0,b_0,c_0)}{\partial b\partial c}=-\sum_{t=2}^{n}\rho''(\varepsilon_t)X_{t-1}^4 b_0 e^{-2c_0X_{t-1}^2}+\sum_{t=2}^{n}\rho'(\varepsilon_t)X_{t-1}^3e^{-c_0X_{t-1}^2},$$

$$\frac{\partial^2 g(a_0,b_0,c_0)}{\partial c^2}=\sum_{t=2}^{n}\rho''(\varepsilon_t)X_{t-1}^6 b_0^2 e^{-2c_0X_{t-1}^2}-\sum_{t=2}^{n}\rho'(\varepsilon_t)X_{t-1}^5 b_0 e^{-c_0X_{t-1}^2}.$$

We show that there is $\lim\limits_{n\to\infty}B$. For example, find $\lim\limits_{n\to\infty}\dfrac{1}{n}\dfrac{\partial^2 g(a_0,b_0,c_0)}{\partial c^2}$. By the assumption, random sequences ε_t and X_t are both stationary and ergodic. Therefore, the sequences $\rho''(\varepsilon_t)X_{t-1}^6e^{-2c_0X_{t-1}^2}$ and $\rho'(\varepsilon_t)X_{t-1}^5e^{-c_0X_{t-1}^2}$

will also be stationary and ergodic [11]. Under the law of large numbers for stationary and ergodic sequences (see [11]), we obtain that with $n \to \infty$ probability

$$\frac{1}{n}\sum_{t=2}^{n} \rho''(\varepsilon_t) X_{t-1}^6 b_0^2 e^{-2c_0 X_{t-1}^2} \to \mathrm{E}[\rho''(\varepsilon_1)]\,\mathrm{E}\left(X_0^6 b_0^2 e^{-2c_0 X_0^2} \right);$$

$$\frac{1}{n}\sum_{t=2}^{n} \rho'(\varepsilon_t) X_{t-1}^5 b_0 e^{-c_0 X_{t-1}^2} \to \mathrm{E}[\rho'(\varepsilon_t)]\,\mathrm{E}\left(X_0^5 b_0 e^{-c_0 X_0^2} \right).$$

From the form of the ρ-functions defined by formulas (3), (4), it follows that $\mathrm{E}\rho'(\varepsilon_t) = 0.$ So from independence ε_t from X_{t-1} it follows: $\mathrm{E}[\rho'(\varepsilon_t)]\,\mathrm{E}\left[X_0^5 b_0 e^{-c_0 X_0^2} \right] = 0.$ Thus, with $n \to \infty$ probability

$$\frac{1}{n}\frac{\partial^2 g(a_0, b_0, c_0)}{\partial c^2} \to \mathrm{E}[\rho''(\varepsilon_1)]\,\mathrm{E}\left(X_0^6 b_0^2 e^{-2c_0 X_0^2} \right).$$

Calculating similarly the remaining elements of the matrix $\lim\limits_{n\to\infty} B,$ we obtain that in probability $\lim\limits_{n\to\infty} B = \mathrm{E}[\rho''(\varepsilon_1)]K,$ where

$$K = \begin{pmatrix} \mathrm{E}(X_0^2) & \mathrm{E}\left(X_0^2 e^{-c_0 X_0^2} \right) & -\mathrm{E}\left(X_0^4 b_0 e^{-c_0 X_0^2} \right) \\ \mathrm{E}\left(X_0^2 e^{-c_0 X_0^2} \right) & \mathrm{E}\left(X_0^2 e^{-2c_0 X_0^2} \right) & -\mathrm{E}\left(X_0^4 b_0 e^{-2c_0 X_0^2} \right) \\ -\mathrm{E}\left(X_0^4 b_0 e^{-c_0 X_0^2} \right) & -\mathrm{E}\left(X_0^4 b_0 e^{-2c_0 X_0^2} \right) & \mathrm{E}\left(X_0^6 b_0^2 e^{-2c_0 X_0^2} \right) \end{pmatrix}.$$

Consequently

$$g(a,b,c) = g(a_0, b_0, c_0) + A^T\theta + \frac{1}{2}\mathrm{E}[\rho''(\varepsilon_1)]\theta^T K\theta + \delta(a,b,c) + \tilde{\delta}(a,b,c),$$

where with $n \to \infty$ $\tilde{\delta}(a,b,c) \to 0$ in probability.

Reasoning as in [12], we find the following: the asymptotic distribution of the M-estimates $\left(\hat{a}, \hat{b}, \hat{c}\right),$ which is the minimum point of the objective function $g(a,b,c),$ coincides with the asymptotic distribution of the minimum point $\left(\tilde{a}, \tilde{b}, \tilde{c}\right)$ of the function

$$\tilde{g}(a,b,c) = g(a_0, b_0, c_0) + A^T\theta + \frac{1}{2}\mathrm{E}[\rho''(\varepsilon_1)]\theta^T K\theta,$$

representing a quadratic form. It is easy to show that

$$(\tilde{a}, \tilde{b}, \tilde{c}) = -(\mathrm{E}[\rho''(\varepsilon_1)]K)^{-1} A.$$

Let us prove the asymptotic normality of the random vector A, from which the asymptotic normality of the random vector $\left(\tilde{a}, \tilde{b}, \tilde{c}\right)$ will follow, and therefore the asymptotic normality of the M-estimates $\left(\hat{a}, \hat{b}, \hat{c}\right)$. Denote by the $\mathcal{A}_{t-1}$ σ-algebra of events generated by the set of random variables $X_0, \ldots, X_{t-1}$. As was noted $\mathrm{E}\rho'(\varepsilon_t) = 0$, then it follows from (1) that ε_t does not depend on X_{t-1}. By the definition of σ-algebra $\mathcal{A}_{t-1}$, a random variable X_{t-1} is measurable with respect to $\mathcal{A}_{t-1}$. Therefore, from the properties of conditional mathematical expectations [13] we have

$$\mathrm{E}[\rho'(\varepsilon_t) X_{t-1} \,|\, \mathcal{A}_{t-1}] = X_{t-1} \mathrm{E}[\rho'(\varepsilon_t) \,|\, \mathcal{A}_{t-1}] = X_{t-1} \mathrm{E}[\rho'(\varepsilon_t)] = 0.$$

Given the form of (3), (4) ρ-functions, we obtain: $0 < \mathrm{E}[(\rho'(\varepsilon_t) X_{t-1})^2] < \infty$. Therefore, by the central limit theorem for martingales (for example, [14]) a random sequence

$$\frac{1}{\sqrt{n}} \frac{\partial g(a_0, b_0, c_0)}{\partial a} = -\frac{1}{\sqrt{n}} \sum_{t=2}^{n} \rho'(\varepsilon_t) X_{t-1}$$

is asymptotically normal with a expectation function

$$\lim_{n \to \infty} \mathrm{E}\left(\frac{1}{\sqrt{n}} \sum_{t=2}^{n} \rho'(\varepsilon_t) X_{t-1} \right)$$

and variance

$$\lim_{n \to \infty} \mathrm{D}\left(\frac{1}{\sqrt{n}} \sum_{t=2}^{n} \rho'(\varepsilon_t) X_{t-1} \right).$$

Since the random variables ε_t and X_{t-1} are both independent and $\mathrm{E}\rho'(\varepsilon_t) = 0$, so

$$\mathrm{E}\left(\frac{1}{\sqrt{n}} \sum_{t=2}^{n} \rho'(\varepsilon_t) X_{t-1} \right) = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \mathrm{E}[\rho'(\varepsilon_t) X_{t-1}] = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \mathrm{E}[\rho'(\varepsilon_t)] \mathrm{E}[X_{t-1}] = 0.$$

Since for $s < t$ the random variables bb and dd are independent, and $a = b$, then for all $s < t$.

The random variables $\rho'(\varepsilon_t)$ and $\rho'(\varepsilon_s) X_{s-1} X_{t-1}$ are independent and $\mathrm{E}\rho'(\varepsilon_t) = 0$. Then for all $s < t$

$$\mathrm{E}[\rho'(\varepsilon_s) X_{s-1} \rho'(\varepsilon_t) X_{t-1}] = \mathrm{E}[\rho'(\varepsilon_t)] \mathrm{E}[\rho'(\varepsilon_s) X_{s-1} X_{t-1}] = 0.$$

Further, independence of ε_t and X_{t-1} implies

$$\mathrm{E}[(\rho'(\varepsilon_t)X_{t-1})^2] = \mathrm{E}[\rho'(\varepsilon_t)^2]\mathrm{E}[X_{t-1}^2] = \mathrm{E}[\rho'(\varepsilon_1)^2]\mathrm{E}[X_0^2].$$

So at $n \to \infty$

$$\mathrm{D}\left(\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\rho'(\varepsilon_t)X_{t-1}\right) = E\left(\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\rho'(\varepsilon_t)X_{t-1}\right)^2 =$$

$$= \frac{1}{n}\sum_{t=2}^{n}\sum_{s=2}^{n}\mathrm{E}[\rho'(\varepsilon_s)X_{s-1}\rho'(\varepsilon_t)X_{t-1}] \to \frac{1}{n}\sum_{t=2}^{n}\mathrm{E}[(\rho'(\varepsilon_t)X_{t-1})^2] = \mathrm{E}[\rho'(\varepsilon_1)^2]\mathrm{E}[X_0^2].$$

Thus, the random sequence $\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial a}$ is asymptotically normal with zero mean and variance $\mathrm{E}[\rho'(\varepsilon_1)^2]\mathrm{E}[X_0^2]$.

It is similarly proved that the random sequence $\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial b}$ is asymptotically normal with zero mean and variance $\mathrm{E}[\rho'(\varepsilon_1)^2]\mathrm{E}\left[X_0^2 e^{-2c_0X_0^2}\right]$, and the random sequence $\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial c}$ is asymptotically normal with zero mean and variance $\mathrm{E}[\rho'(\varepsilon_1)^2]\mathrm{E}\left[X_0^6 b^2 e^{-2c_0X_0^2}\right]$.

To find the asymptotic covariance matrix of the vector A we get:

$$\lim_{n\to\infty}\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial a}\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial b}\right) = \mathrm{E}[\rho'(\varepsilon_1)^2]\mathrm{E}\left(X_0^2 e^{-c_0X_0^2}\right);$$

$$\lim_{n\to\infty}\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial a}\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial c}\right) = -\mathrm{E}[\rho'(\varepsilon_1)^2]\mathrm{E}\left(X_0^4 b_0 e^{-c_0X_0^2}\right);$$

$$\lim_{n\to\infty}\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial b}\frac{1}{\sqrt{n}}\frac{\partial g(a_0,b_0,c_0)}{\partial c}\right) = -\mathrm{E}[\rho'(\varepsilon_1)^2]\mathrm{E}\left(X_0^4 b_0 e^{-2c_0X_0^2}\right).$$

Thus, the random vector A is asymptotically normal with zero mean and the covariance matrix $\mathrm{E}[\rho'(\varepsilon_1)^2]K$. Since $\left(\tilde{a},\tilde{b},\tilde{c}\right) = -(\mathrm{E}[\rho''(\varepsilon_1)]K)^{-1}A$, then the covariance matrix of the vector $\left(\tilde{a},\tilde{b},\tilde{c}\right)$, and, therefore, the covariance matrix of M-estimates $\left(\hat{a},\hat{b},\hat{c}\right)$ is equals (see for example [15])

$$\frac{\mathrm{E}[\rho'(\varepsilon_1)^2]}{(\mathrm{E}[\rho''(\varepsilon_1)])^2}K^{-1}.$$

Substituting into this expression $\rho(x)=x^2$, we obtain the following: since $(x^2)'=2x$ and $(x^2)''=2$, the least squares estimate $\left(a^*, b^*, c^*\right)$ is asymptotically normal with zero mean and the covariance matrix

$$\frac{\mathrm{E}[2(\varepsilon_1)^2]}{(\mathrm{E}[2])^2}K^{-1}=\mathrm{E}\varepsilon_1^2K^{-1}=\sigma^2K^{-1}.$$

Let us compare the quality of the M-estimates $\left(\hat{a}, \hat{b}, \hat{c}\right)$ and the least-squares estimate $\left(a^*, b^*, c^*\right)$. Among the two scalar estimates, we will consider the best one that deviates less from the estimated parameter. Estimates are random; therefore, their deviations from the estimated parameter are also random. It is logical to measure the accuracy of the estimate by expectation of the square of the difference between the estimate and the parameter being estimated, which for an asymptotically unbiased estimate coincides with its asymptotic variance. Thus, it is advisable to compare the accuracy of two scalar asymptotically unbiased estimates by comparing their asymptotic variances. If the estimated parameter is a vector, and its estimates are asymptotically unbiased and have asymptotic covariance matrices proportional to each other, then it is natural to compare the accuracy of these estimates with the ratio of the proportionality coefficients of their asymptotic variances.

Estimates $\left(\hat{a}, \hat{b}, \hat{c}\right)$ and $\left(a^*, b^*, c^*\right)$ are asymptotically normal with proportional covariance matrices; therefore, we will compare the accuracy of the estimates $\left(\hat{a}, \hat{b}, \hat{c}\right)$ and $\left(a^*, b^*, c^*\right)$ with each other by comparing the quantities $\dfrac{\mathrm{E}[\rho'(\varepsilon_1)^2]}{(\mathrm{E}[\rho''(\varepsilon_1)])^2}$ and σ^2. The ratio

$$e(f,k)=\frac{\sigma^2}{\dfrac{\mathrm{E}[\rho'(\varepsilon_1)^2]}{(\mathrm{E}[\rho''(\varepsilon_1)])^2}}=\frac{\sigma^2(\mathrm{E}[\rho''(\varepsilon_1)])^2}{\mathrm{E}[\rho'(\varepsilon_1)^2]}$$

will be called the asymptotic relative efficiency of the M-estimate with respect to the least squares estimate. Asymptotic relative efficiency shows how many times more observations are needed to the least squares estimate compared with the number of observations required by the M-estimate to achieve the same accuracy. For example, $e=2$ means that to achieve the same accuracy, the M-estimate requires 2 times less observations n, than the least squares estimate.

Asymptotic relative efficiency depends on the type of probability distribution density $f(x)$ of the updating process ε_t. Let us calculate the value e for various distributions of the updating process ε_t.

If ε_t are Gaussian random variables with $E\varepsilon_t = 0$ and $D\varepsilon_t = 1$, so $f(x) = f_g(x)$, where

$$f_g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

So

$$E[(\rho'(\varepsilon_1))^2] = \int_{-\infty}^{\infty} (\rho'(x))^2 f_g(x)dx = 4\left(k^2 - \frac{k\sqrt{2}}{\sqrt{\pi}} e^{-\frac{k^2}{2}} + (1-k^2)\operatorname{erf}\left(\frac{k}{\sqrt{2}}\right)\right) =$$

$$= 4\left(k^2 - \frac{k\sqrt{2}}{\sqrt{\pi}} e^{-\frac{k^2}{2}} + 2(1-k^2)\Phi_0(k)\right);$$

$$E[\rho''(\varepsilon_1)] = \int_{-\infty}^{\infty} \rho''(x) f_g(x)dx = 2\operatorname{erf}\left(\frac{k}{\sqrt{2}}\right) = 4\Phi_0(k);$$

$$e(f_g, k) = \frac{\operatorname{erf}^2\left(\frac{k}{\sqrt{2}}\right)}{k^2 - \frac{k\sqrt{2}}{\sqrt{\pi}} e^{-\frac{k^2}{2}} + (1-k^2)\operatorname{erf}\left(\frac{k}{\sqrt{2}}\right)} = \frac{4\Phi_0^2(k)}{k^2 - \frac{k\sqrt{2}}{\sqrt{\pi}} e^{-\frac{k^2}{2}} + 2(1-k^2)\Phi_0(k)}.$$

Here $\operatorname{erf}(x) = \int_0^x \frac{2}{\sqrt{\pi}} e^{-t^2} dt$ is error function; $\Phi_0(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ is Laplace's function.

Figure 3 shows the relationship between $e(f_g, k)$ and k with a Gaussian distribution ε_t. The asymptotic relative efficiency of the M-estimate with increases of k monotonically increases, that is, the effectiveness of the M-estimate increases, at $k \to \infty$ approaching the least-squares efficiency. We note that $\lim_{k\to\infty} e(f_g, k) = 1$, i.e., $k \to \infty$, when the M-estimate goes over to the least-squares estimate, the quality of both estimates becomes the same. Also, when $k = 0$ the M-estimate becomes the least absolute deviation estimate. Since when $x \to 0$

$$\Phi_0(x) = \frac{1}{2}\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) = \frac{x}{\sqrt{2\pi}} + O(x^3), \quad e^{-\frac{x^2}{2}} = 1 - x^2 + O(x^3), \tag{5}$$

then, $\lim_{k\to 0} e(f_g, k) = 2/\pi$. Therefore, with a Gaussian distribution of random variables ε_t, the least squares estimate is approximately 1.5 times more effective

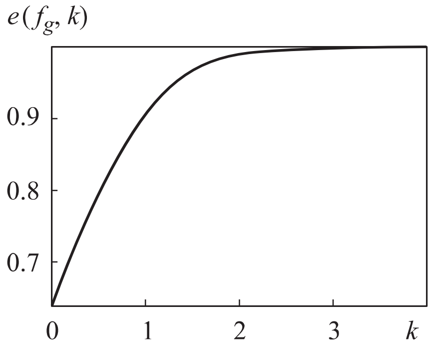

Fig. 3. Relationship between $e(f_g, k)$ and k on normal ε_t distribution

than the least absolute deviation estimate. In other words, the accuracy of the least squares estimate constructed from $n = 100$ the observations will be achieved using the least absolute deviation estimates only based on approximately $n = 150$ observations.

The assumption that ε_t are Gaussian random variables is usually justified by the central limit theorem of probability theory. However, this theorem is limiting in nature and the assumption that the probability distribution of ε_t deviates slightly from the Gaussian one looks more realistic. A typical model for violating the assumption of ε_t Gaussianity is the assumption that it has a contaminated (clogged) Gaussian distribution, or Tukey distribution, with a density [16]

$$f_T(x) = (1-\gamma)\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \gamma \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{x^2}{2\tau^2}},\ 0 \le \gamma \le 1,\ \tau > 1.$$

In this case, the least squares method usually sharply loses its effectiveness [17].

Estimate $e(f_T, k)$. In this case the values $\sigma^2 = \mathrm{E}\varepsilon_t$, $\mathrm{E}[\rho'(\varepsilon_1)^2]$ and $\mathrm{E}[\rho''(\varepsilon_1)]$ are the following

$$\sigma^2 = 1 + \tau^2\gamma - \gamma;$$

$$\mathrm{E}[\rho'(\varepsilon_1)^2] = 4k^2\left(1 + (\gamma-1)\operatorname{erf}\left(\frac{k}{\sqrt{2}}\right) - \gamma \operatorname{erf}\left(\frac{k}{\sqrt{2}\tau}\right)\right) + 4\tau^2\gamma \operatorname{erf}\left(\frac{k}{\tau\sqrt{2}}\right) +$$

$$+ 4(1-\gamma)\left(\operatorname{erf}\left(\frac{k}{\sqrt{2}}\right) - \frac{k\sqrt{2}}{\sqrt{\pi}} e^{-\frac{k^2}{2}}\right) - \frac{4\tau\gamma k\sqrt{2}}{\sqrt{\pi}} e^{-\frac{k^2}{2\tau^2}} =$$

$$= 4\left(k^2 + 2(1-\gamma)(1-k^2)\Phi_0(k) + 2\gamma(\tau^2 - k^2)\Phi_0\left(\frac{k}{\tau}\right) +\right.$$

$$\left. + \frac{k\sqrt{2}(\gamma-1)}{\sqrt{\pi}} e^{-\frac{k^2}{2}} - \frac{k\tau\gamma\sqrt{2}}{\sqrt{\pi}} e^{-\frac{k^2}{2\tau^2}}\right);$$

$$\mathrm{E}[\rho''(\varepsilon_1)] = 2(1-\gamma)\operatorname{erf}\left(\frac{k}{\sqrt{2}}\right) + 2\gamma \operatorname{erf}\left(\frac{k}{\tau\sqrt{2}}\right) = 4\left((1-\gamma)\Phi_0(k) + \gamma\Phi_0\left(\frac{k}{\tau}\right)\right).$$

Consequently,

$$e(f_T,k)=$$
$$=\frac{(1+\tau^2\gamma-\gamma)\left((1-\gamma)\Phi_0(k)+\gamma\Phi_0\left(\frac{k}{\tau}\right)\right)^2}{k^2+2(1-\gamma)(1-k^2)\Phi_0(k)+2\gamma(\tau^2-k^2)\Phi_0\left(\frac{k}{\tau}\right)+\frac{k\sqrt{2}}{\sqrt{\pi}}\left((\gamma-1)e^{-\frac{k^2}{2}}-\tau\gamma e^{-\frac{k^2}{2\tau^2}}\right)}.$$

Using (5), at $\tau\to\infty$ get $e(f_T,k)=C\tau^2+O(\tau)$, where $C>0$ is some constant depending on k and γ.

Therefore, with increasing τ value $e(f_T,k)$ the value increases indefinitely for any $k>0$ and $\gamma\in(0,1)$. Thus, the asymptotic relative efficiency of the M-estimates with respect to the least squares estimate for $\tau\to\infty$ can be arbitrarily large.

When $k\to 0$ we obtain the asymptotic relative efficiency of the least absolute deviation estimate with respect to the estimate of least squares

$$e(f_T,0)=\lim_{k\to 0} e(f_T,k)=\frac{2(1+\tau^2\gamma-\gamma)(\tau^2-2\tau^2\gamma+2\gamma\tau+\gamma^2\tau^2-2\gamma^2\tau+\gamma^2)}{\tau^2\pi}=$$
$$=\frac{2(1-\gamma+\tau^2\gamma)(\tau\gamma-\tau-\gamma)^2}{\tau^2\pi},$$

which, as is easily seen, coincides with $2/\pi$ at $\tau=1$ and $\gamma=0$.

Conclusion. Using computer simulation, it was found that the M-estimates of the coefficients of the equation of the exponential autoregression is unbiased, consistent and asymptotically normal. The relationship betweem the asymptotic variance of the M-estimates and the form of the probability distribution density of the updating process of the autoregressive equation is found. The values of the asymptotic variance of the M-estimates are calculated for the main types of density. It is shown that in conditions close to practical, the M-estimates is more efficient than the least squares estimate and the least absolute deviation estimate.

Translated by K. Zykova

REFERENCES

[1] Ozaki T. Time series modeling of neuroscience data. CRC Press, 2012.

[2] Olugbode M., El-Masry A., Pointon J. Exchange rate and interest rate exposure of UK industries using first-order autoregressive exponential GARCH-in-mean (EGARCH-M) approach. *Manch. Sch.*, 2014, vol. 82, iss. 4, pp. 409–464.
DOI: https://doi.org/10.1111/manc.12029

[3] Gurung B. An exponential autoregressive (EXPAR) model for the forecasting of all India annual rainfall. *Mausam*, 2015, vol. 66, no. 4, pp. 847–849.

[4] Ghosh H., Gurung B., Gupta P. Fitting EXPAR models through the extended Kalman filter. *Sankhya B*, 2015, vol. 77, no. 1, pp. 27–44.
DOI: https://doi.org/10.1007/s13571-014-0085-8

[5] Goryainov A.V., Goryainov V.B. M-Estimates of autoregression with random coefficients. *Autom. Remote Control*, 2018, vol. 79, pp. 1409–1421.
DOI: https://doi.org/10.1134/S0005117918080040

[6] Goryainov V.B., Goryainova E.R. Comparative analysis of robust and classical methods for estimating the parameters of a threshold autoregression equation. *Autom. Remote Control*, 2019, vol. 80, no. 4, pp. 666–675.
DOI: https://doi.org/10.1134/S0005117919040052

[7] Chan K.S., Tong H. On the use of deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.*, 1985, vol. 17, iss. 3, pp. 666–678. DOI: https://doi.org/10.2307/1427125

[8] Tong H. Threshold models in time series analysis 30 years on. *Stat. Interface*, 2011, vol. 4, no. 2, pp. 107–118. DOI: https://dx.doi.org/10.4310/SII.2011.v4.n2.a1

[9] Huber P., Ronchetti E.M. Robust statistics. Wiley, 2009.

[10] Rhinehart R.R. Nonlinear regression modeling for engineering applications: modeling, model validation, and enabling design of experiments. Wiley, 2016.

[11] White H. Asymptotic theory for econometricians. Academic Press, 2000.

[12] Goryainov V.B. M-estimates of the spatial autoregression coefficients. *Autom. Remote Control*, 2012, vol. 73, no. 8, pp. 1371–1379.
DOI: https://doi.org/10.1134/S0005117912080103

[13] Shiryaev A.N. Veroyatnost' [Probability]. Moscow, Nauka Publ., 2011.

[14] Billingsley P. Convergence of probability measures. Wiley, 1999.

[15] Magnus J.R., Neudecker H. Matrix differential calculus with applications in statistics and econometrics. Wiley, 1999.

[16] Hampel F.R., Ronchetti E.M., Rousseeuw P.J., et al. Robust statistics. Wiley, 2005.

[17] Goryainov V.B., Goryainova E.R. The influence of anomalous observations on the least squares estimate of the parameter of the autoregressive equation with random coefficient. *Herald of the Bauman Moscow State Technical University. Series Natural Sciences*, 2016, no. 2 (65), pp. 16–24 (in Russ.).
DOI: http://doi.org/10.18698/1812-3368-2016-2-16-24

Goryainov A.V. — Cand. Sc. (Phys.-Math.), Assoc. Professor, Department of Probability Theory and Computer Modeling, Moscow Aviation Institute (National Research University) (Volokolamskoe shosse 4, Moscow, 125993 Russian Federation).

Goryainov V.B. — Dr. Sc. (Phys.-Math.), Professor, Department of Mathematical Simulation, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Khing W.M. — Post-Graduate Student, Department of Mathematical Simulation, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Please cite this article as:

Goryainov A.V., Goryainov V.B., Khing W.M. Robust identification of an exponential autoregressive model. *Herald of the Bauman Moscow State Technical University, Series Natural Sciences*, 2020, no. 4 (91), pp. 42–57.
DOI: https://doi.org/10.18698/1812-3368-2020-4-42-57

В Издательстве МГТУ им. Н.Э. Баумана
вышла в свет монография авторов
И.В. Фомина, С.В. Червона, А.Н. Морозова

«Гравитационные волны ранней Вселенной»

Рассмотрены применение скалярных полей в космологии и методы построения моделей ранней Вселенной на основе их динамики. Выполнен анализ динамики Вселенной на различных стадиях ее эволюции. Проведен расчет параметров космологических возмущений. Представлены методы верификации инфляционных моделей и новые методы детектирования гравитационных волн.

По вопросам приобретения обращайтесь:
105005, Москва, 2-я Бауманская ул., д. 5, стр. 1
+7 (499) 263-60-45
press@bmstu.ru
https://bmstu.press