ISSN 0005-1179, Automation and Remote Control, 2017, Vol. 78, No. 10, pp. 1803–1818. © Pleiades Publishing, Ltd., 2017.
Original Russian Text © S.E. Vorobeichikov, V.V. Konev, 2017, published in Avtomatika i Telemekhanika, 2017, No. 10, pp. 90–108.

On Sequential Confidence Estimation of Parameters of Stochastic Dynamical Systems with Conditionally Gaussian Noises

S. E. Vorobeichikov[*,a] **and V. V. Konev**[*,b]

[*] *Tomsk State University, Tomsk, Russia*
e-mail: [a]*sev@mail.tsu.ru,* [b]*vvkonev@mail.tsu.ru*

Received March 30, 2016

Abstract—We consider the problem of non-asymptotical confidence estimation of linear parameters in multidimensional dynamical systems defined by general regression models with discrete time and conditionally Gaussian noises under the assumption that the number of unknown parameters does not exceed the dimension of the observed process. We develop a non-asymptotical sequential procedure for constructing a confidence region for the vector of unknown parameters with a given diameter and given confidence coefficient that uses a special rule for stopping the observations. A key role in the procedure is played by a novel property established for sequential least squares point estimates earlier proposed by the authors. With a numerical modeling example of a two-dimensional first order autoregression process with random parameters, we illustrate the possibilities for applying confidence estimates to construct adaptive predictions.

Keywords: confidence estimation, sequential methods for dependent observations, conditionally Gaussian noises, multidimensional dynamical systems, nonlinear time series.

DOI: 10.1134/S0005117917100058

1. INTRODUCTION

Many applied problems of control and identification of dynamical systems, signal processing, and data analysis in information and measurement complexes successfully employ discrete time models defined by stochastic difference equations with conditionally Gaussian noises. These models are interesting because they lead to a sufficiently adequate description for a wide class of real world phenomena in technics, physics, finances, economics, and so on. In filtering theory, for instance, they lead to Kalman–Bucy equations in a closed form with no stationarity assumptions, and coefficients in the equations of the dynamical system may depend on all past values [1]. Various efficient approaches have been developed for identification of unknown parameters in dynamical systems subject to random noises: least squares (LS), maximal likelihood (ML), stochastic approximation, and others (see, e.g., [2]). All identification methods aim to construct point estimates or confidence regions for unknown parameters with a given quality. Here the quality of point estimates is usually measured as the mean squared deviation of estimates of parameters from their true values, while confidence estimates are evaluated by the size of confidence regions and probability of the fact that the true value of the parameter will lie inside the confidence region.

It is known that even in case when unknown parameters occur in the stochastic equation in a linear way, and the distribution of noises that define process dynamics is known, the study of properties of estimates becomes radically more complicated when we pass from a deterministic regression model to a stochastic one. Linear analysis of deterministic regression has been comprehensively presented in [3] and has direct applications to real world problems; in particular, it lets one find the necessary sample size that would ensure a given quality of point or confidence estimates for unknown parameters.

Complications in the studies of stochastic regression stem from the fact that classical estimates of parameters computed with a fixed size sample are nonlinear functions of observations, which is easy to see by considering, e.g., an LS estimate of the parameter of a first order autoregression. This is due to the fact that a central place in identification theory is occupied by asymptotic methods of analysis that study convergence conditions for the estimates and their distributions under an unbounded growth of the number of observations (see, e.g., [4]). Asymptotic identification theory has led to strict justifications of many models used in time series analysis.

In practical use of identification algorithms, one often assumes that properties of estimates for small and moderate sample sizes are close to asymptotical. However, limit and sublimit properties of estimates may differ significantly, which may be reflected in the quality of decisions made based on these estimates. For example, in the construction of confidence estimates one should be cautious when using the asymptotic normality of classical least squares estimates because the rate of convergence to a normal law may depend on unknown parameters, and the normal approximation itself may be unsatisfactory for every fixed number of observations (see, e.g., [5]). Therefore, while asymptotical results obtained by classical methods are no doubt important, and they establish a potential possibility to construct models of dynamical systems, in the non-asymptotical identification theory many problems related to a practical use of these estimates remain open.

These problems include, first, constructing procedures for point and confidence estimation with guaranteed quality that would use finite sample sizes. Second, search for additional ways to improve the asymptotical properties of estimates. A large number of works (see [6, 7] and references therein) have been devoted to developing methods of guaranteed point estimation based on the sequential analysis approach. An important characteristic feature of these methods is that they use in their decision procedures samples of variable (random) size. The need to apply sequential analysis for a scheme with independent observations was first justified by Wald [8]. For processes with dependent values, in the optimal stopping problems sequential analysis was used by A.N. Shiryaev [9]. Novikov [10] and Liptser and Shiryaev [1] proved that the maximal likelihood estimate for the shift coefficient in the diffusion process computed at a special (random) time moment has better properties than the classical estimate: it is unbiased and has predefined mean squared accuracy. Note that for stochastic dynamical systems with discrete time and for more complex models with continuous time sequential estimates of unknown parameters obtained by substituting a certain stopping moment in a classical least squares or maximal likelihood estimate cannot be studied for a finite number of observations. Nevertheless, it has been established in [7, 11–13] and other works that constructing sequential point estimates for parameters of stochastic systems with guaranteed accuracy based on classical LS and ML estimates becomes possible under sufficiently general conditions on the model if together with introducing special rules for stopping the observations one also makes certain additional structural changes in classical basic estimates. The works [11, 13] have developed a method for constructing unbiased sequential point estimates for the parameters of stochastic dynamical systems with discrete time that guarantee a given mean squared accuracy under the assumption that the number of unknown parameters does not exceed the dimension of the observed process.

Together with point estimates for parameters of stochastic systems with discrete time, both theoretical and applied studies pay special attention to developing efficient procedures for confidence estimation of the parameters. The problem of constructing confidence estimates in regression models has been considered in different settings in a number of works. The work [14, 15] proposed methods for constructing confidence intervals for survival quantiles in the Cox regression model, including one based on sequential analysis. An approach to constructing confidence regions for the parameters of a threshold autoregression with non-Gaussian noises has been developed in [16, 17]. The confidence estimation problem for the parameters of multidimensional regression with constraints on predicted variables has been studies in [18]. The work [19] proposed a confidence estimate for the slope of a linear regression with Gaussian noises. The problem of simultaneous confidence estimation of the parameters of a linear regression with high dimension and hypothesis testing has been studied in [20].

The main objective of this work is to develop, based on sequential analysis, algorithms for constructing confidence intervals and sets of a fixed size with a given confidence level for the estimation of unknown parameters in stochastic dynamical systems with conditionally Gaussian noises. To refine the probabilities of deviations of the estimates and reduce the size of the confidence region we propose to improve the sequential point estimation procedure from [13] and obtain exact probabilities for the deviations of estimates that take into account the specifics of noises acting on the system. Accuracy of the proposed confidence estimation algorithm is supported by modeling results in Section 4.

2. PROBLEM SETTING. CONSTRUCTING CONFIDENCE ESTIMATION ALGORITHMS

Suppose that an observable n-dimensional process $x_t = (x_1(t), \ldots, x_n(t))'$, $t = 0, 1, \ldots$, satisfies equations

$$x_{t+1} = A_0(t,x) + A_1(t,x)\theta + B(t,x)\varepsilon_{t+1}, \quad t = 0, 1, \ldots, \tag{1}$$

where $\theta = (\theta_1, \ldots, \theta_p)'$ is the vector of unknown parameters, $p \leqslant n$; the prime denotes transposition; $\varepsilon_t = (\varepsilon_1(t), \ldots, \varepsilon_n(t))'$, $t = 0, 1, \ldots$, is the unobserved sequence of independent Gaussian vectors with parameters $\mathbf{E}\varepsilon_t = 0$, $cov(\varepsilon_t, \varepsilon_t) = I_n$, I_n is the unit matrix of order n. We assume that the vector of initial values x_0 does not depend on the noise (ε_t), and vector functions $A_0(t,x)$ and matrix functions $A_1(t,x)$ and $B(t,x)$ of sizes $n \times 1$, $n \times p$, and $n \times n$ respectively depend on the process $\{x_t,\ t = 0, 1, \ldots\}$ in a nonanticipatory way, i.e., for a given t their elements depend only on $x_0, \ldots, x_t$. The problem is to construct a confidence region for the unknown vector of parameters of a given diameter that contains the true value of the parameter with at least a given probability.

Remark 1. It is known (see, e.g. [6, 21, 22]), that in a number of applied problems models where unknown parameters are subject to disturbances and remain constant only on average are more adequate. In this version, the dynamics of process $(x_t)_{t\geqslant 0}$ is defined by equations

$$x_{t+1} = A_0(t,x) + A_1(t,x)\left(\theta + \eta_t\right) + B(t,x)\varepsilon_{t+1}, \tag{2}$$

where η_t is an unobserved multiplicative noise. If process $(\eta_t)_{t\geqslant 1}$ is a sequence of independent identically distributed Gaussian random vectors with parameters 0 and Σ independent of the noise $(\varepsilon_t)_{t\geqslant 1}$, then it is easy to see that Eq. (2) reduces to an equation of type (1),

$$x_{t+1} = A_0(t,x) + A_1(t,x)\theta + \tilde{B}(t,x)\tilde{\varepsilon}_{t+1}, \tag{3}$$

where

$$\tilde{B}(t,x) = D^{1/2}(t,x), \quad D(t,x) = A_1(t,x)\Sigma A_1'(t,x)(t,x) + B(t,x)B'(t,x).$$

Thus, the confidence estimation problem for parameter θ in Eq. (2) reduces to estimating parameter θ in Eq. (1), with the difference that noises $(\tilde{\varepsilon}_t)_{t\geqslant 1}$ are only conditionally Gaussian. However, as we will see below, this difference does not complicate the analysis.

In the choice of estimation method for unknown parameters $\theta_1, \ldots, \theta_p$ one should take into account that structural matrices $A_i(t, x)$ and $B(t, x)$ are, generally speaking, random. Besides, it is natural to require that the estimates would let us control the size of the confidence region for unknown parameters with a finite sample. For a practical application, the confidence region is better when it is smaller. Unfortunately, immediate use of least squares, maximal likelihood estimates, and other estimates based on a fixed number of observations of the process $\{x_t\}$, although they are relatively easy to compute, does not allow for confidence estimates with desired properties.

A more suitable version for our purposes are sequential variants of least squares (maximal likelihood) estimates that include a special choice of weight matrices in the LS method and rules for stopping the observations. To estimate the vector parameter θ in process (1), the authors proposed a sequential LS estimate [13] that guarantees estimation of unknown parameters with given mean squared accuracy.

Since we propose to use these point estimates in the considered confidence estimation problem, we show the necessary formulas, omitting derivations and justifications.

Case A. Each parameter θ_i in the vector $\theta = (\theta_1, \ldots, \theta_p)'$ is estimated by the observations of process (1) separately.

In this case, for every positive number h we introduce a system of sequential plans $(\tau_i(h), \hat{\theta}_i(h))$, $1 \leqslant i \leqslant p$, where $\tau_i(h)$ is the number of observations of process (1) used to estimate the scalar parameter θ_i, and $\hat{\theta}_i(h)$ is the sequential point estimate of parameter θ_i, where

$$\tau_i(h) = \inf\left\{n \geqslant 0 : \sum_{t=0}^{n} c_i(t) \geqslant h\right\}, \quad \inf\{\emptyset\} = \infty; \tag{4}$$

$$\hat{\theta}_i(h) = \frac{1}{h} \sum_{t=0}^{\tau_i(h)} \beta_i(t) c_i(t) y_i(t). \tag{5}$$

Here

$$c_i(t) = 1/ < \left[A_1' \left(B(t)B'(t)\right)^+ A_1(t)\right]^{-1} >_{ii}; \quad y_i(t) = < W(t)\left(x_{t+1} - A_0(t)\right) >_i;$$

$$W(t) = \left(A_1'(t)\left(B(t)B'(t)\right)^{-1} A_1(t)\right)^+ A_1'(t)\left(B(t)B'(t)\right)^+;$$

$$\beta_i(t) = \begin{cases} 1, & 0 \leqslant t < \tau_i(h) \\ \alpha_i(h), & t = \tau_i(h), \end{cases}$$

where $\alpha_i(h)$ is a correction factor, $0 < \alpha_i(h) \leqslant 1$, that satisfies equation

$$\sum_{t=0}^{\tau_i(h)-1} c_i(t) \; + \; \alpha_i(h) c_i(\tau_i(h)) = h;$$

$< A >_{ij}$ denotes the (i, j)th element of matrix A, $< a >_i$ is the ith coordinate of vector $(a_1, \ldots, a_p)'$; A^+ is the pseudoinverse matrix for matrix A (see, e.g., [23]).

Case B. All coordinates of vector $\theta = (\theta_1, \ldots, \theta_p)'$ are estimated simultaneously.

Here for every $h > 0$ we introduce a sequential plan

$$\left(\tau(h),\ \hat{\theta}(h) = (\hat{\theta}_1(h), \ldots, \hat{\theta}_p(h))'\right)$$

for the estimation of vector $\theta = (\theta_1, \dots, \theta_p)'$ by formulas

$$\tau(h) = \inf\left\{n \geqslant 0: \sum_{t=0}^{n} c(t) \geqslant h\right\}, \tag{6}$$

$$\hat{\theta}(h) = \frac{1}{h}\sum_{t=0}^{\tau(h)} \beta(t) A_1'(t,x) V(t) \left(x_{t+1} - A_0(t,x)\right), \tag{7}$$

where

$$\beta(t) = \begin{cases} 1, & 0 \leqslant t < \tau(h) \\ \alpha(h), & t = \tau(h), \end{cases}$$

$\alpha(h)$ is a correction factor, $0 < \alpha(h) \leqslant 1$, that satisfies equation

$$\sum_{t=0}^{\tau(h)-1} c(t) + \alpha(h) c(\tau(h)) = h,$$

$$V(t) = c(t)\left[A_1^+(t,x)\right]' A_1^+(t,x); \tag{8}$$

$$c(t) = \begin{cases} 1/\nu^*(t), & \text{if } A_1^+(t,x) A_1(t,x) = I_p \\ 0, & \text{otherwise,} \end{cases}$$

$\nu^*(t)$ is the maximal eigenvalue of matrix

$$U(t) = A_1^+(t,x) B(t,x) B'(t,x) \left(A_1^+(t,x)\right)'.$$

Durations $\tau_i(h)$ and $\tau(h)$ of the said identification procedures are random and depend on the value of the parameter h. For these durations to be finite for an arbitrary $h > 0$, the set Θ of admissible values of parameter θ in Eq. (1) must satisfy

$$P_\theta\left(\sum_{t \geqslant 0} c_i(t) = +\infty\right) = 1, \quad i = \overline{1,p}; \quad P_\theta\left(\sum_{t \geqslant 0} c(t) = +\infty\right) = 1 \tag{9}$$

for all $\theta \in \Theta$.

The works [11, 13] have established that sequential identification procedures (4), (5) and (6), (7), if they can be implemented, have a number of advantages over classical LS estimates based on a fixed number of observations in point estimation problems for the parameters of stochastic dynamical systems defined by Eq. (1).

First, the sequential point estimates (4), (5) and (6), (7) are unbiased, i.e.,

$$\mathbf{E}_\theta \hat{\theta}_i(h) = \theta_i, \quad i = \overline{1,p}, \qquad \mathbf{E}_\theta \hat{\theta}(h) = \theta, \quad \theta \in \Theta.$$

Second, for every $h > 0$ they satisfy the following inequalities:

$$\sup_{\theta \in \Theta} \mathbf{E}_\theta (\hat{\theta}_i(h) - \theta_i)^2 \leqslant \frac{1}{h}, \quad i = \overline{1,p}, \tag{10}$$

$$\sup_{\theta \in \Theta} \mathbf{E}_\theta (\hat{\theta}(h) - \theta)(\hat{\theta}(h) - \theta)' \leqslant \frac{1}{h} I_p, \tag{11}$$

where I_p is the unit matrix of order p; matrix inequality $A \leqslant B$ means that for all $x \in \mathbf{R}^p$ the inequality holds for the corresponding quadratic forms $x'Ax \leqslant x'Bx$. Since the value of parameter h

can be chosen in the procedure, these inequalities can be regarded as the guarantee property in the mean squared sense for sequential point estimates (4), (5) and (6), (7).

Note that the said properties of sequential point estimates have been proven in [11, 13] under the assumption that the distribution of noises (ε_t) in Eq. (1) is unknown, and $\{\varepsilon_t\}$ is either a sequence of independent vectors with $\mathbf{E}\varepsilon_t = 0$ and $cov(\varepsilon_t, \varepsilon_t) = I_n$ or the process (ε_t) is a quadratically integrable martingale difference with respect to a certain filtration $(\mathcal{F}_t)_{t\geqslant 0}$ (compatible with the structural functions $A_i(t,x)$ and $B(t,x)$) of the process, i.e.,

$$\mathbf{E}\left(\varepsilon_t|\mathcal{F}_{t-1}\right) = 0, \quad \mathbf{E}\left(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right) \leqslant I_n.$$

The goal of this work is to show that sequential point estimates (4), (5) and (6), (7) can be used as a foundation for constructing confidence intervals and sets for unknown parameters of the process (1).

According to confidence estimation theory [3], the confidence estimation problem for the vector of unknown parameters θ by observations $Y_n = (X_0, \ldots, X_n)$ of the process (1) is to construct a subset $S\left(Y_n\right)$ in the parametric region $\Theta \subset \mathbf{R}^p$ that covers the true value of the parameter θ with a given confidence coefficient $1-\Delta > 0$, i.e.,

$$P_\theta\left(\theta \in S\left(Y_n\right)\right) \geqslant 1-\Delta, \quad \theta \in \Theta. \tag{12}$$

Sequential confidence estimation theory (see, e.g., [8]) admits the possibility to construct confidence sets with samples of random size τ, i.e., $Y_\tau = (X_0, \ldots, X_\tau)$. Here the confidence set can be constructed with some sequential point estimate of the unknown parameter. For the considered dynamical system (1) with structural functions $A_i(t,x)$ and $B(t,x)$, we will use sequential point estimates (4), (5) and (6), (7) to construct confidence estimates.

If we need to estimate only one of the parameters $\theta_1, \ldots, \theta_p$, it is natural to take as the confidence set for θ_i the confidence interval

$$S_i\left(Y_{\tau_i}, z, h\right) = \left(\hat{\theta}_i(h) - z,\ \hat{\theta}_i(h) + z\right), \quad z > 0, \quad h > 0, \tag{13}$$

where $\hat{\theta}_i(h)$ is computed with (4) and (5).

If we are estimating the entire vector θ, we define the confidence set by equality

$$S\left(Y_\tau, z, h\right) = \left\{\theta \in \Theta:\ ||\hat{\theta}(h) - \theta|| < z\right\}, \quad z > 0, \tag{14}$$

where $\hat{\theta}(h)$ is computed with (6) and (7); $||\theta|| = \left(\sum_{i=1}^p \theta_i^2\right)^{1/2}$. The value z defines the radius of the confidence ball set.

3. PROPERTIES OF SEQUENTIAL CONFIDENCE ESTIMATES

Sequential confidence estimates (4), (5) and (6), (7) depend on two parameters z and h, where z defines the size of the confidence set and h controls the duration for the identification procedure. For a fixed $z > 0$ the value of h should be chosen in such a way that confidence sets (13), (14) cover the corresponding true values of unknown parameters θ_i and θ with probability no less than the predefined confidence coefficient $1-\Delta > 0$; for all $\theta \in \Theta$ the following inequalities must hold:

$$\begin{gathered} P_\theta\left(\theta_i \in S_i\left(Y_{\tau_i}, z, h\right)\right) \geqslant 1-\Delta, \quad i = \overline{1,p}, \\ P_\theta\left(\theta \in S\left(Y_\tau, z, h\right)\right) \geqslant 1-\Delta. \end{gathered} \tag{15}$$

It is problematic to find the optimal value of h that minimizes sample sizes in sequential confidence estimates (13), (14), because to compute exact upper bounds with respect to $\theta \in \Theta$ for the probabilities in the left-hand sides of inequalities (13), (14) we need to know non-asymptotical (for finite h) distributions for estimates $\hat{\theta}_i(h)$ and $\hat{\theta}(h)$. Since we do not know these distributions, to choose h it is natural to use upper bounds of the said probabilities. Note that rough estimates of probabilities in the left-hand side of inequalities (13), (14) obtained with the guarantee properties (10), (11) and Chebyshev inequalities should be excluded from consideration since this choice for the value of h would lead to very long confidence estimation procedures (4), (5) and (6), (7). To finish constructing the non-asymptotical confidence estimation procedure for the unknown parameters of process (1) we will need new properties of point estimates proposed in [13] that we formulate as Theorem 1.

Theorem 1. *Suppose that an observable process* $(x_t)_{t\geqslant 0}$ *satisfies Eqs.* (1), *where* $\{\varepsilon_t\}_{t\geqslant 1}$ *is a sequence of independent n-dimensional standard Gaussian vectors, and structural functions* $A_0(t,x)$, $A_1(t,x)$, *and* $B(t,x)$ *are such that for some parametric set* $\Theta \subset \mathbf{R}^p$ *conditions* (9) *hold.*

Then point estimate of the coordinates $\hat{\theta}_i(h)$, $i=\overline{1,p}$, *and vector* $\hat{\theta}(h)$ *defined by equalities* (5) *and* (7) *satisfy for all* $h>0$ *and* $z>0$ *the following inequalities:*

1) $P_\theta\left(|\hat{\theta}_i(h)-\theta_i| > z\right) \leqslant 2e^{-\frac{z^2h}{2}}, \quad i=\overline{1,p} \quad \forall\, \theta\in\Theta;$

2) $P_\theta\left(||\hat{\theta}(h)-\theta|| > z\right) \leqslant 2p\exp\left(-\dfrac{z^2h}{2p}\right).$

Proof of Theorem 1, which is postponed to the Appendix, is based on a generalization of the well-known Azuma–Hoeffding inequality for martingales (see, e.g., [24]), which is also proven in the Appendix.

Using exponential upper bounds for the probabilities of estimates deviating from unknown parameters obtained in Theorem 1, we can find the parameters of the confidence estimation procedure that would guarantee a given estimation quality.

Theorem 2. *Suppose that conditions of Theorem 1 hold, and we are given* $z>0$ *and* $0<\Delta<1$.

Then confidence intervals for θ_i, $i=\overline{1,p}$, *defined in* (13), *for* $h=(2/z^2)ln(2/\Delta)$ *have confidence level at least* Δ, *i.e.,*

$$P_\theta\left(\theta_i\in S_i\left(Y_{\tau_i},z,h\right)\right) \geqslant 1-\Delta, \quad \forall\,\theta\in\Theta, \quad i=\overline{1,p}. \tag{16}$$

Theorem 3. *Under the assumptions of Theorem 2, confidence set for vector* θ *defined by equality* (14), *with* $h=(2p/z^2)ln(2p/\Delta)$ *covers the true value of* θ *with probability at least* $1-\Delta$, *i.e.,*

$$P_\theta\left(\theta\in S\left(Y_\tau,z,h\right)\right) \geqslant 1-\Delta, \quad \forall\,\theta\in\Theta.$$

Statements of Theorems 2 and 3 that finish the construction of non-asymptotical confidence estimates for unknown parameters of the process (1) immediately follow from Theorem 1. Let us check inequality (16). Finding h by given z and Δ from equations

$$\exp\left(-\frac{z^2h}{2}\right)=\Delta,$$

we get that $h=\frac{2}{z^2}ln\frac{2}{\Delta}$. This together with Theorem 1 implies that

$$P_\theta\left(\theta_i\in S_i\left(Y_{\tau_i},z,h\right)\right) = 1-P_\theta\left(|\hat{\theta}_i-\theta_i|>z\right) \geqslant 1-2\exp\left(-\frac{z^2h}{2}\right)=1-\Delta.$$

This completes the proof of Theorem 2.

4. NUMERICAL MODELING RESULTS

To test the quality of constructed procedures for guaranteed confidence estimation, we have conducted numerical modeling (10 000 realizations for every process). We have modeled a two-dimensional process x_t with conditionally Gaussian noises defined by equation

$$x_{t+1} = A_1(t,x)\theta + B(t,x)\varepsilon_{t+1}, \quad t = 0, 1, \ldots, \tag{17}$$

where

$$A_1(t,x) = \begin{pmatrix} x_1(t) & x_2(t) \\ -x_1(t) & x_2(t) \end{pmatrix}, \tag{18}$$

$\theta = (\theta_1, \theta_2)'$ is the vector of unknown parameters, $\{\varepsilon_t\}_{t\geqslant 1}$ is the sequence of independent standard two-dimensional Gaussian vectors. Here the structural matrix $B(t,x)$ of size 2×2, which defines the level of noises, has been used in two variations. In the first case, $B(t,x)$ was assumed to be the unit matrix I_2; in the second, its elements are ARCH processes

$$B(t,x) = \begin{pmatrix} \sqrt{0.3 + 0.1x_1^2(t) + 0.2x_2^2(t)} & \sqrt{0.2 + 0.3x_1^2(t)} \\ \sqrt{0.4 + 0.2x_2^2(t)} & \sqrt{0.4 + 0.2x_1^2(t) + 0.1x_2^2(t)} \end{pmatrix}. \tag{19}$$

In case $B(t,x) = I_2$, we get from (5) that

$$c_1(t) = 2x_1^2(t), \quad c_2(t) = 2x_2^2(t).$$

It can be checked immediately that sequences $\{c_1(t)\}$ and $\{c_2(t)\}$ satisfy conditions (9). One can test condition (9) for system (17) with matrix $B(t,x)$ defined by (19) in a similar way.

Modeling results are shown in Tables 1 and 2, which show sample averages of the duration of estimating $\mathbf{E}_\theta\tau_1$ and $\mathbf{E}_\theta\tau_2$ for parameters θ_1 and θ_2. Values of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ show the sample probability of the confidence interval (of semi-width z) not covering the true value of the parameters, and the procedure's parameter h was chosen so as to get the value of the upper bound Δ for the probability of not covering the true value.

In the modeling of the confidence estimation procedure, we used a modification of the sequential procedure (4) with the previous stopping moment (4) and estimate

$$\theta_i^*(h) = \frac{\sum\limits_{t=0}^{\tau_i(h)} \sqrt{\beta_i(t)}c_i(t)y_i(t)}{\sum\limits_{t=0}^{\tau_i(h)} \sqrt{\beta_i(t)}c_i(t)},$$

which differs from estimate (5) by a weight coefficient in the last terms of the sums. This estimate satisfies the following proposition.

Proposition 1. *Under the assumptions of Theorem 1, estimate $\theta_i^*(h)$, $i = \overline{1,p}$, satisfies the following inequalities*

$$P_\theta\left(|\theta_i^*(h) - \theta_i| \geqslant c\right) \leqslant 2\left(1 - \Phi(c\sqrt{h})\right) = \Delta.$$

Proof of these inequalities is similar to the proof of the first statement in Theorem 1 and is not shown here.

Table 1. Modeling the process with $B(t, x) = I_2$

$h = 270.55$		$z = 0.1$		$\Delta = 0.1$	
θ_1	θ_2	$\mathbf{E}_\theta\tau_1$	$\mathbf{E}_\theta\tau_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
0.4	–0.2	108.2	112.3	0.1011	0.0930
0.5	0.5	71.3	70.9	0.0979	0.1004
0.2	–0.9	25.3	21.9	0.0877	0.0927
0.2	–0.7	68.2	55.6	0.0994	0.0938
$h = 385.15$		$z = 0.1$		$\Delta = 0.05$	
θ_1	θ_2	$\mathbf{E}_\theta\tau_1$	$\mathbf{E}_\theta\tau_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
0.4	–0.2	152.6	158.2	0.0509	0.0505
0.5	0.5	99.9	99.5	0.0452	0.0494
0.2	–0.9	29.0	24.9	0.0456	0.0445
0.2	–0.7	94.5	75.6	0.0454	0.0515

Table 2. Modeling the process with ARCH noises

$h = 270.55$		$z = 0.1$		$\Delta = 0.1$	
θ_1	θ_2	$\mathbf{E}_\theta\tau_1$	$\mathbf{E}_\theta\tau_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
0.4	–0.2	2.6	299.0	0.0487	0.1008
0.5	0.5	2.7	261.9	0.0453	0.0996
0.2	–0.9	2.3	172.3	0.0445	0.0924
0.2	–0.7	2.4	235.7	0.0466	0.1029
$h = 385.15$		$z = 0.1$		$\Delta = 0.05$	
θ_1	θ_2	$\mathbf{E}_\theta\tau_1$	$\mathbf{E}_\theta\tau_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
0.4	–0.2	3.4	423.5	0.0207	0.0563
0.5	0.5	3.5	370.8	0.0150	0.0520
0.2	–0.9	2.9	242.0	0.0149	0.0513
0.2	–0.7	3.1	333.5	0.0184	0.0513

Table 3. Modeling the process with drifting parameters

$h = 270.55$		$z = 0.1$		$\Delta = 0.1$	
θ_1	θ_2	$\mathbf{E}_\theta\tau_1$	$\mathbf{E}_\theta\tau_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
0.4	–0.2	109.8	132.9	0.0953	0.1015
0.5	0.5	72.2	92.0	0.0969	0.0975
0.2	–0.9	28.7	32.7	0.0964	0.0964
0.2	–0.7	69.4	74.6	0.0976	0.0926
$h = 384.15$		$z = 0.1$		$\Delta = 0.05$	
θ_1	θ_2	$\mathbf{E}_\theta\tau_1$	$\mathbf{E}_\theta\tau_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
0.4	–0.2	155.1	187.9	0.0485	0.0488
0.5	0.5	101.1	129.4	0.0474	0.0481
0.2	–0.9	34.6	40.7	0.0448	0.0478
0.2	–0.7	95.7	102.9	0.0514	0.0494

We have also modeled a process with drifting parameters

$$x_{t+1} = A_1(t, x)\,(\theta + \eta_t) + B(t, x)\varepsilon_{t+1}, \tag{20}$$

where $B(t, x) = I_2$, $(\eta_t)_{t\geqslant 0}$ is a sequence of independent identically distributed two-dimensional Gaussian random vectors with zero mean and diagonal covariance matrix $\Sigma = \mathrm{diag}(0.01,\ 0.04)$. Sample characteristics of confidence estimates (13) are shown in Table 3.

Besides, using the resulting confidence estimates we have constructed one-step predictions for the two-dimensional process (17) with drifting parameters (20) $\theta_1 = 0.2$, $\theta_2 = -0.7$. On the first

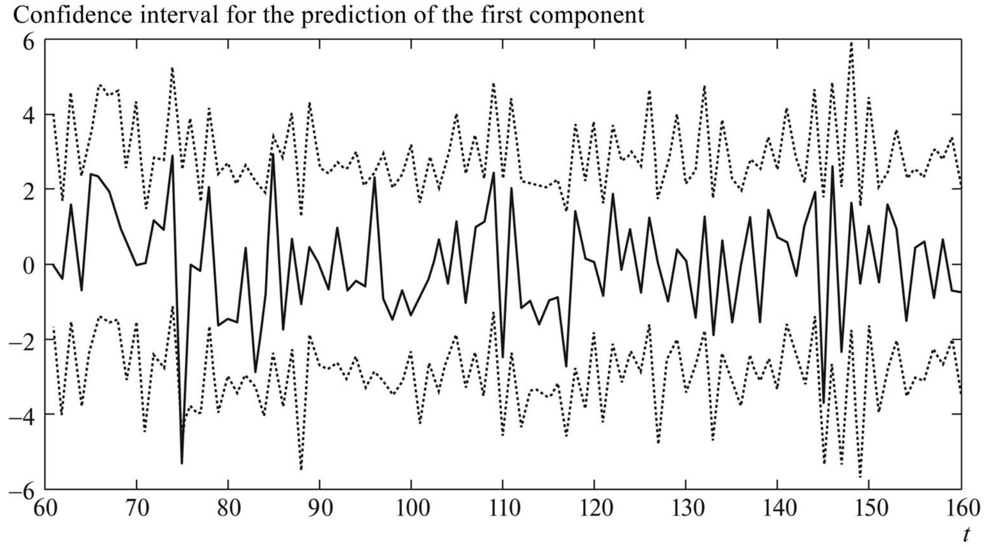

Figure.

step, we computed an estimate for the unknown parameters θ_1 and θ_2 by formulas (4), (5). Next, for $t = \max(\tau_1(h), \tau_2(h)) + 1, \ldots$ we have constructed predictions for the values of process x_{t+1} and confidence intervals for these predictions. The prediction of value x_{t+1} is defined as

$$\hat{x}_{t+1} = A_1(t, x)\hat{\theta}.$$

To construct a confidence interval for the prediction we compute the value

$$z_i(t) = c \times (|(A_1(t,x))_{i,1}| + (A_1(t,x))_{i,2}|) + F^{-1}(0.995) \left[< \tilde{B}(t,x) >_{i,i}\right]^{0.5},$$
$$\tilde{B} = A_1 \Sigma A_1' + I,$$

where $F^{-1}(0.995)$ is the quantile of level 0.995 for the standard normal distribution; the value c defines the semi-width of the confidence interval for parameter estimates, which was chosen to equal 0.1 in our modeling. The procedure's parameter h was chosen to be such that the probability of not covering the true value of the parameter θ_i would not exceed 0.02. Confidence intervals for the coordinates of vector $x_{t+1} = (x_1(t+1), x_2(t+1))$ are

$$\left(\hat{x}_i(t+1) - z_i(t),\ \hat{x}_i(t+1) + z_i(t)\right), \quad i = 1, 2.$$

The probability of one of the coordinates of the next value of the process x_{t+1} to not fall into the corresponding interval does not exceed 0.05. Therefore, true values of the process x_{t+1} for $t \geqslant 61$ will fall into the confidence region shown on the figure with probability 0.95. Prediction results for the second component are similar. To estimate parameter θ_i with given accuracy, we needed 60 observations.

5. CONCLUSION

In Sections 2 and 3 of this work, we have proposed and studied identification algorithms for the parameters of stochastic dynamical systems defined by Eq. (1) with confidence intervals and sets. A characteristic feature of the algorithms for estimating both individual coordinates and the entire vector of unknown parameters is that they are constructed by samples of random size, using special rules for stopping the observations, as it is often done in sequential analysis. We have established that such sequential confidence estimates yield a solution for the identification problem in the non-asymptotical setting with guaranteed accuracy and define the required number of observations for a given confidence set and given probability that this set does not cover the true value of the parameter. Note that this conclusion holds for a wide class of structural functions in the observable process (1) with no additional assumptions such as stationarity.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of the Russian Federation (state contracts no. 2.3208.2017/PCh) and the Russian Foundation for Basic Research, project no. 16-01-00121A.

APPENDIX

1. To prove Theorem 1 we will need a generalization of the well-known Azuma–Hoeffding inequalities (see, e.g., [24], Theorem 6.3.3), which we formulate as a theorem since it can also be of separate interest.

Theorem A.1. *Let* $(M_k, \mathcal{F}_k)_{k\geqslant 0}$ *be a quadratically integrable martingale such that*

(a) *its quadratic characteristic* [9] *satisfies condition*

$$P(<M>_\infty = +\infty) = 1;$$

(b) $Law(\Delta M_k|\mathcal{F}_{k-1}) = \mathcal{N}\left(0, \sigma^2_{k-1}\right)$, *i.e.,* $\mathcal{F}_{k-1}$, *the conditional distribution of* $\Delta M_k = M_k - M_{k-1}$, *is Gaussian with parameters* 0 *and* $\sigma^2_{k-1} = \mathbf{E}((\Delta M_k)^2|\mathcal{F}_{k-1})$.

For every $h > 0$ *we introduce the stopping moment*

$$\tau = \tau(h) = \inf\left\{n \geqslant 1 : \sum_{k=1}^{n} \sigma^2_{k-1} \geqslant h\right\} \tag{A.1}$$

and random value

$$M^*(h) = \frac{1}{h}\sum_{k=1}^{\tau(h)} \alpha_k(h)\Delta M_k, \tag{A.2}$$

where

$$\alpha_k(h) = \begin{cases} 1, & 1 \leqslant k < \tau(h) \\ \beta, & k = \tau(h), \end{cases}$$

β *is a correction factor* $0 < \beta \leqslant 1$ *defined by equations*

$$\sum_{k=1}^{\tau(h)-1} \sigma^2_{k-1} + \beta\sigma^2_{\tau(h)-1} = h.$$

Then for all $h > 0$ *and* $z > 0$

1) $\mathbf{E}e^{\gamma M^*(h)} \leqslant e^{\frac{\gamma^2}{2h}}$;

2) $P(|M^*(h))| > z) \leqslant 2e^{\frac{-z^2 h}{2}}$.

Remark 2. Random value $M^*(h)$ differs from the value of martingale $M_{\tau(h)}$ stopped at time moment $\tau(h)$ by the presence of the correction factor β at the last increment $\Delta M_{\tau(h)}$.

Remark 3. A transformation of type (A.2) was first used in [11].

Proof of Theorem A.1. We represent $\gamma M^*(h)$ as

$$\begin{aligned}\gamma M^*(h) = \sum_{k=1}^{\tau(h)} \left(\frac{\gamma}{h}\alpha_k(h)\Delta M_k - \frac{\gamma^2\alpha_k^2}{2h^2}\mathbf{E}\left((\Delta M_k^2)|\mathcal{F}_{k-1}\right)\right) \\ + \frac{\gamma^2}{2h^2}\sum_{k=1}^{\tau(h)} \alpha_k^2(h)\sigma_{k-1}^2 = \zeta_{\tau(h)} \ + \ \frac{\gamma^2}{2h^2}\sum_{k=1}^{\tau(h)} \alpha_k^2(h)\sigma_{k-1}^2,\end{aligned} \tag{A.3}$$

where

$$\zeta_n = \sum_{k=1}^{n} \left(\frac{\gamma}{h}\alpha_k(h)\Delta M_k - \frac{\gamma^2\alpha_k^2}{2h^2}\sigma_{k-1}^2\right).$$

The definition of $\tau(h)$ in (A.1) implies that

$$\frac{\gamma^2}{2h^2}\sum_{k=1}^{\tau(h)} \alpha_k^2(h)\sigma_{k-1}^2 \leqslant \frac{\gamma^2}{2h^2}\sum_{k=1}^{\tau(h)} \alpha_k(h)\sigma_{k-1}^2 = \frac{\gamma^2}{2h^2}.$$

This together with (A.3) implies that

$$\exp(\gamma M^*(h)) \leqslant e^{\frac{\gamma^2}{2h}} \exp\left(\zeta_{\tau(h)}\right). \tag{A.4}$$

Let us show that

$$\mathbf{E}\exp(\zeta_{\tau(h)}) \leqslant 1. \tag{A.5}$$

We introduce the sequence of truncated moments $\overline{\tau}(h,N) = \tau(h) \wedge N,\ N = 1, 2, \ldots,$ and denote $\overline{\zeta}_N = \zeta_{\overline{\tau}(h,N)}$. Noting that

$$\overline{\zeta}_N = \sum_{k=1}^{N} \left(\frac{\gamma}{h}\alpha_k(h)\Delta M_k \chi_{(k\leqslant\tau(h))} - \frac{\gamma^2\alpha_k^2}{2h^2}\sigma_{k-1}^2\chi_{(k\leqslant\tau(h))}\right),$$

and computing again the conditional expectations, we get

$$\begin{aligned}\mathbf{E}\exp(\overline{\zeta_N}) = \mathbf{E}\mathbf{E}\exp(\overline{\zeta}_N|\mathcal{F}_{N-1}) \\ = \mathbf{E}e^{\overline{\zeta}_{N-1}}\mathbf{E}\left(\exp\left[\frac{\gamma\alpha_N(h)}{h}\chi_{(N\leqslant\tau(h))}(\Delta M_k) - \frac{\gamma^2\alpha_N^2(h)}{2h^2}\chi_{(N\leqslant\tau(h))}\right]\Big|\mathcal{F}_{N-1}\right) \\ = \mathbf{E}e^{\overline{\zeta}_{N-1}} = \ldots = 1.\end{aligned}$$

By Fatou's lemma, we have inequalities

$$\mathbf{E}\exp(\zeta_{\tau(h)}) \leqslant \underline{\lim}_{N\to\infty} \mathbf{E}\exp(\overline{\zeta_N}) = 1,$$

which together with (6) imply the first statement of Theorem A.1.

Further, for every $\gamma > 0$ and $z > 0$ Chebyshev inequalities imply that

$$P\left(M^*(h) > z\right) = P\left(\gamma M^*(h) > \gamma z\right) \leqslant e^{-\gamma z}\mathbf{E}e^{\gamma M^*(h)} \leqslant e^{-\gamma z}e^{\frac{\gamma^2}{2h}}.$$

Minimizing the right-hand side with respect to γ, we get the estimate

$$P\left(M^*(h) > z\right) \leqslant e^{-\frac{z^2 h}{2}}.$$

We can also similarly find that

$$P\left(M^*(h) < -z\right) \leqslant e^{-\frac{z^2 h}{2}}.$$

This completes the proof of Theorem A.1.

2. **Proof of Theorem 1.** Substituting (1) into (5), we have

$$\hat{\theta}_i(h) = \frac{1}{h}\sum_{t=0}^{\tau_i(h)} \beta_i(t)c_i(t) < W(t)\left(A_1(t,x)\theta + B(t,x)\varepsilon_{t+1}\right) >_i$$

$$= \frac{1}{h}\sum_{t=0}^{\tau_i(h)} \beta_i(t)c_i(t) < \left(A_1'(t)(B(t)B'(t))^+ A_1(t)\right)^+ A_1'(t)(B(t)B'(t))^+ A_1(t,x)\theta >_i + \eta_i(h)$$

$$= \frac{1}{h}\sum_{t=0}^{\tau_i(h)} \beta_i(t)c_i(t)\theta_i + \eta_i(h) = \theta_i + \eta_i(h),$$

where

$$\eta_i(h) = \frac{1}{h}\sum_{t=0}^{\tau_i(h)} \beta_i(t)c_i(t) < W(t)B(t,x)\varepsilon_{t+1} >_i .$$

Next we introduce the process $(M_k)_{k\geqslant 0}$ defined by equalities

$$M_0 = 0, \quad M_k = \sum_{t=0}^{k-1} c_i(t) < W(t)B(t,x)\varepsilon_{t+1} >_i, \quad k \geqslant 1,$$

and filtration $\{\mathcal{F}\}_{k\geqslant 0}$ with $\mathcal{F} = \sigma(X_0)$, $\mathcal{F}_k = \sigma(x_0, \varepsilon_1, \ldots, \varepsilon_k)$, $k \geqslant 1$. Due to the conditions on $\{\varepsilon_k\}_{k\geqslant 1}$ in (1) the sequence $(M_k, \mathcal{F}_k)_{k\geqslant 0}$ is a martingale with conditionally Gaussian increments. Using the definition of matrix $W(t)$ and functions $c_i(t)$ in (5), we compute the conditional variance of the increments

$$\sigma_{k-1}^2 = \mathbf{E}\left[(\Delta M_k)^2 | \mathcal{F}_{k-1}\right] = c_i^2(k-1)\mathbf{E} < W(k-1)B(k-1,x)\varepsilon_k >_{ii}^2$$
$$= c_i^2(k-1) < W(k-1)B(k-1,x)B'(k-1,x)W'(k-1,x) >_{ii}$$
$$= c_i^2(k-1) < A_1'(k-1,x)(B(k-1,x)B'(k-1,x))^+ A_1^+(k-1,x) >_{ii} = c_i(k-1).$$

Hence, comparing the definitions of stopping moments (4) and (A.1) and taking into account (A.2), we get

$$M^*(h) = \eta_i(h).$$

Using Theorem 2, we get the first statement of Theorem 1.

Let us now check the second statement. Using (1), we write estimate (7) as

$$\hat{\theta}(h) = \frac{1}{h}\sum_{t=0}^{\tau(h)} \beta(t)A_1'(t,x)V(t,x)A_1(t,x)\theta + \eta(h), \tag{A.6}$$

where

$$\eta(h) = \frac{1}{h}\sum_{t=0}^{\tau(h)} \beta(t)A_1'(t,x)V(t,x)B(t,x)\varepsilon_{t+1}. \tag{A.7}$$

Taking into account (8), we find the deviation of estimate

$$\hat{\theta}(h) = \frac{1}{h}\sum_{t=0}^{\tau(h)} \beta(t)c(t)A_1'(t,x)\left(A_1^+(t,x)A_1^+(t,x)A_1(t,x)\theta\right) + \eta(h)$$
$$= \frac{1}{h}\sum_{t=0}^{\tau(h)} \beta(t)c(t)I_p\theta + \eta(h) = \theta + \eta(h). \tag{A.8}$$

Let us show that every nonzero vector $\lambda = (\lambda_1, \ldots, \lambda_p)' \in \mathbf{R}^p$ and number $z > 0$ satisfy the following inequality:

$$P_\theta\left(|\lambda'\eta(h)| > z\right) \leqslant 2\exp\left(-\frac{z^2 h}{2\lambda'\lambda}\right), \quad \theta \in \Theta. \tag{A.9}$$

For every $\gamma > 0$ we have that

$$P_\theta\left(\lambda'\eta(h) > z\right) \leqslant e^{-\gamma z}\mathbf{E}_\theta e^{\gamma\lambda'\eta(h)}. \tag{A.10}$$

We represent the value $\gamma\lambda'\eta(h)$ as

$$\gamma\lambda'\eta(h) = \sum_{t=0}^{\tau(h)} b(t)\varepsilon_{t+1} = \zeta_{\tau(h)} + \frac{1}{2}\sum_{t=0}^{\tau(h)} ||b||^2(t), \tag{A.11}$$

where

$$b(t) = \frac{\gamma}{h}\beta(t)\lambda' A_1'(t,x)V(t,x)B(t,x), \quad \zeta_n = \sum_{t=0}^{n}\left(b(t)\varepsilon_{t+1} - \frac{||b(t)||^2}{2}\right).$$

Since matrix $V(t)$ in (8) satisfies inequality $V(t) \geqslant V(t)B(t,x)B'(t,x)V'(t)$, the definition of $\tau(h)$ in (6) implies that

$$\frac{1}{2}\sum_{t=0}^{\tau(h)} ||b(t)||^2 = \frac{\gamma^2}{2h^2}\sum_{t=0}^{\tau(h)} \beta^2(t)||\lambda' A_1'(t,x)V(t,x)B(t,x)||^2$$
$$= \frac{\gamma^2}{2h^2}\sum_{t=0}^{\tau(h)} \beta^2(t)\lambda' A_1'(t,x)V(t,x)B(t,x)B'(t,x)V(t,x)A_1(t,x)\lambda$$
$$\leqslant \frac{\gamma^2}{2h^2}\lambda'\sum_{t=0}^{\tau(h)} \beta^2(t)A_1'(t,x)V(t,x)A_1(t,x)\lambda = \frac{\gamma^2}{2h}\lambda'\lambda.$$

This together with (A.11) implies the inequality

$$\mathbf{E}_\theta \exp\left(\gamma\lambda'\eta(h)\right) \leqslant e^{\frac{\gamma^2}{2h}\lambda'\lambda}\mathbf{E}_\theta e^{\zeta_{\tau(h)}}. \tag{A.12}$$

Let us show that

$$\mathbf{E}_\theta \exp(\zeta_{\tau(h)}) \leqslant 1. \tag{A.13}$$

We introduce the sequence of truncated stopping moments $\overline{\tau}(h,N) = \tau(h) \wedge N$, $N = 1, 2, \ldots$ and denote $\overline{\zeta}_N = \zeta_{\overline{\tau}(h,N)}$. Since for every $\theta \in \Theta$

$$\overline{\zeta}_N \to \zeta_{\tau(h)} \ P_\theta\text{—almost surely for } N \to \infty,$$

by Fatou's lemma it suffices to check that

$$\mathbf{E}_\theta \exp(\overline{\zeta}_N) \leqslant 1, \quad \theta \in \Theta.$$

Writing $\overline{\zeta}_N$ as

$$\overline{\zeta}_N = \sum_{t=0}^{N} \left(b(t)\chi_{(t\leqslant\tau(h))}\varepsilon_{t+1} - \frac{||b(t)||^2}{2}\chi_{(t\leqslant\tau(h))} \right),$$

taking into account that sequence $(\varepsilon(t))_{t\geqslant 1}$ is Gaussian, and computing conditional expectations repeatedly, we get the necessary estimate

$$\mathbf{E}_\theta \exp(\overline{\zeta}_N) \ = \mathbf{E}_\theta \mathbf{E}_\theta \left(\exp(\overline{\zeta}_N) | \mathcal{F}_N \right)$$
$$= \mathbf{E}_\theta \exp(\overline{\zeta}_{N-1}) \mathbf{E} \left(\exp\left[b(N)\chi_{N\leqslant\tau(h)}\varepsilon_{N+1} - \frac{||b(N)||^2}{2}\chi_{N\leqslant\tau(h)} \right] \Big| \mathcal{F}_N \right)$$
$$= \mathbf{E}_\theta \exp(\overline{\zeta}_{N-1}) = \ldots = 1.$$

Combining inequalities (A.10), (A.12), and (A.13), we get the inequality

$$P_\theta \left(\lambda' \eta(h) > z \right) \leqslant \ e^{-\gamma z} e^{\frac{\gamma^2}{2h}\lambda'\lambda}, \quad \gamma > 0, \quad z > 0.$$

Minimizing the right-hand side of this inequality with respect to γ, we get

$$P_\theta \left(\lambda' \eta(h) > z \right) \leqslant \exp\left(-\frac{z^2}{2h\lambda'\lambda} \right).$$

We can similarly show that

$$P_\theta \left(\lambda' \eta(h) < -z \right) \leqslant \exp\left(-\frac{z^2}{2h\lambda'\lambda} \right).$$

Consequently, inequality (A.9) holds. Using formula (8) for the deviation of point estimate (A.8) and inequality (A.9) with $\lambda = (\delta_{1i}, \ldots, \delta_{pi})'$, where δ_{ki} is Kroneker's symbol, we get

$$P_\theta(|\hat{\theta}_i(h) - \theta_i| > z) \leqslant 2\exp\left(-\frac{z^2 h}{2} \right), \quad z > 0. \tag{A.14}$$

It remains to estimate the probability of estimate $\hat{\theta}(h)$ falling outside the confidence region. Next we use the inclusions

$$\left\{ |\hat{\theta}(h) - \theta| > z \right\} = \left\{ \sum_{i=1}^{p} (\hat{\theta}_i(h) - \theta_i)^2 > z^2 \right\}$$
$$\subset \bigcup_{i=1}^{p} \left\{ (\hat{\theta}_i(h) - \theta_i)^2 > \frac{z^2}{p} \right\} = \bigcup_{i=1}^{p} \left\{ |\hat{\theta}_i(h) - \theta_i| > \frac{z}{\sqrt{p}} \right\}.$$

This together with (A.14) yields the second statement of Theorem 1:

$$P_\theta \left\{ |\hat{\theta}(h) - \theta| > z \right\} \leqslant \sum_{i=1}^{p} P_\theta \left\{ |\hat{\theta}_i(h) - \theta_i| > \frac{z}{\sqrt{p}} \right\} \leqslant 2p\exp\left(-\frac{z^2 h}{2p} \right).$$

This completes the proof of Theorem 1.

REFERENCES

1. Liptser, R.Sh. and Shiryaev, A.N., *Statistika sluchainykh protsessov* (Statistics of Random Processes), Moscow: Nauka, 1974.
2. Ljung, L. and Söderstrom, T., *Theory and Practice of Recursive Identification*, Cambridge: MIT Press, 1986.
3. Seber, G., *Linear Regression Analysis*, New York: Wiley, 1977. Translated under the title *Lineinyi regressionnyi analiz*, Moscow: Mir, 1980.
4. Anderson, T.W., *The Statistical Analysis of the Time Series*, New York: Wiley, 1971. Translated under the title *Statisticheskii analiz vremennykh ryadov*, Moscow: Mir, 1976.
5. Lai, T.L. and Siegmund, D., Fixed Accuracy Estimation of an Autoregressive Parameter, *Ann. Statist.*, 1983, vol. 11, pp. 478–485.
6. Kashkovskii, D.V. and Konev, V.V., Successive Identification of the Random-parameter Linear Dynamic System, *Autom. Remote Control*, 2008, vol. 69, no. 8, pp. 1344–1356.
7. Konev, V.V. and Pergamenshchikov, S.M., Sequential Plans of Parameter Identification in Dynamic Systems, *Autom. Remote Control*, 1981, vol. 42, no. 7, part 1, pp. 917–924.
8. Wald, A., *Sequential Analysis*, New York: Wiley, 1947. Translated under the title *Posledovatel'nyi analiz*, Moscow: Fizmatgiz, 1960.
9. Shiryaev, A.N., *Statisticheskii posledovatel'nyi analiz* (Statistical Sequential Analysis), Moscow: Nauka, 1976.
10. Novikov, A.A., Sequential Estimation of Parameters of Diffusion Processes, *Teor. Veroyat. Primen.*, 1972, vol. 36, no. 2, pp. 394–396.
11. Borisov, V.Z. and Konev, V.V., On Sequential Parameter Estimation in Discrete-time Processes, *Autom. Remote Control*, 1977, vol. 38, no. 10, pp. 1475–1480.
12. Vasil'ev, V.A. and Konev, V.V., Sequential Estimation of Parameters of Dynamical Systems under Incomplete Observation, *Izv. Akad. Nauk SSSR, Tekh. Kibern.*, 1982, no. 6, pp. 145–154.
13. Vorobeichikov, S.E. and Konev, V.V., On Sequential Identification of Stochastic Systems, *Izv. Akad. Nauk SSSR, Tekh. Kibern.*, 1980, no. 4, pp. 91–98.
14. Lai, T.L. and Li, W., Confidence Intervals in Group Sequential Trials with Random Group Sizes and Applications to Survival Analysis, *Biometrika*, 2006. vol. 93 (3), pp. 641–654.
15. Lai, T.L. and Siegmund, Z., Confidence Intervals for Qurvival Quantiles in the Cox Regression Model, *Lifetime Data Anal.*, 2006, vol. 12 (4), pp. 407–419.
16. Sriram, T.N., Fixed Size Confidence Region for Parameters of Threshold AR(1) Models, *J. Statist. Plann. Inference*, 2001, vol. 97, pp. 293–304.
17. Sriram, T.N. and Iaci, R., Editor's Special Invited Paper: Sequential Estimation for Time Series Models, *Sequential Anal.*, 2014, vol. 33, pp. 136–157.
18. Liu, W. and Lin, S., Construction of Exact Simultaneous Confidence Bands in Multiple Linear Regression with Predictor Variables Constrained in an Ellipsoidal Region, *Statist. Sinica*, 2009, vol. 19, pp. 213–232.
19. Hayter, A.J., Wynn, H.P., and Liu, W., Slope Modified Confidence Bands for a Simple Linear Regression Model, *Statist. Methodology*, 2006, vol. 3, pp. 186–192.
20. Javanmard, A. and Montanari, A., Confidence Intervals and Hypothesis Testing for High-dimensional Regression, *J. Mach. Learning Res.*, 2015, vol. 15, pp. 2869–2909.
21. Feigin, P.D. and Tweedie, R.D., Random Coefficient Autoregressive Process: A Markov Chain Analysis of Stationarity and Finiteness of Moments, *Time Ser. Anal.*, 1985, vol. 6, pp. 1–14.
22. Singpurwalla, N.D. and Soyer, R., Assessing Software Reliability Grow Using a Random Coefficient Autoregressive Process and Its Ramifications, *IEEE Trans. Software Eng.*, 1985, vol. SE-1, pp. 1456–1464.
23. Albert, A., *Regressiya, psevdoinversiya i rekurrentnoe otsenivanie* (Regression, Pseudoinversion, and Recurrent Estimation), Moscow: Nauka, 1977.
24. Ross, S.M., *Stochastic Processes*, New York: Wiley, 1996, 2nd ed.

This paper was recommended for publication by A.V. Nazin, a member of the Editorial Board