ISSN 0005-1179, Automation and Remote Control, 2009, Vol. 70, No. 2, pp. 233–243. © Pleiades Publishing, Ltd., 2009.
Original Russian Text © A.V. Timofeev, 2009, published in Avtomatika i Telemekhanika, 2009, No. 2, pp. 68–79.

Nonasymptotic Confidence Sets of Prescribed Dimensions for Parameters of Nonlinear Regressions

A. V. Timofeev

"Center of Vocal Technologies," St. Petersburg, Russia

Received July 23, 2007

Abstract—The article suggests the sequential plan for confidence estimation of multidimensional parameters that nonlinearly enter into the regression equation. The solution is obtained in the nonasymptotic statement under the condition of the incomplete prior definiteness relative to the distribution of heteroscedastic observations.

PACS numbers: 02.30.Yy, 02.50.Cw, 02.50.Sk

DOI: 10.1134/S0005117909020052

1. INTRODUCTION

In the last years, interest significantly increased to the practical use of nonlinear models in various applied branches, such as, for example, econometrics and processing of stochastic signals [1]. In particular, in econometrics this interest stems from the growing level of dissatisfaction of the results of using linear models for the solution of the complex of econometric problems. The number of these problems includes the prediction of stochastic dynamics of stock indices [2], investigation of oscillations of business activity [3, 4]. It is customary to subject to the investigation only asymptotic properties of the estimates of nonlinear parameters [5–11]. At the same time, in view of the boundedness of the volume accessible in practice of the analyzable sample, nonasymptotic indices of the accuracy of the obtained estimates take on particular significance. For the case of linear models, interesting nonasymptotic results are presented in [12]. Nonasymptotic confidence intervals foe estimating the scalar parameter of the nonlinear regression, which are obtained from the position of the sequential analysis, were considered in [13]. In this work, the sequential plan is suggested for the nonasymptotic solution of the problem for confidence estimation of the nonlinear regression parameter, which is serviceable in the incomplete prior definiteness relative to the distribution of heteroscedactic observations. In contrast to [14], in the solution suggested in this article, explicit expressions for gradients of the loss function are not used, which make the algorithm more feasible.

2. STATEMENT OF THE PROBLEM

The observations $\mathbf{X} = \{X(k) \,|\, k \geq 0\}$ are described by the following equation:

$$X(k) = A(k,\, \theta^*) + \xi(k), \quad k \geq 1,$$

where $X(k)$, $A(k, \theta^*)$, $\xi(k) \in R^m$. The nonlinear functions relative to the unknown parameter θ^*, namely, the functions $\{A(k, \theta^*) \mid k \geq 1\}$ are prescribed. The sequence of random, interindependent vectors $\{\xi(k) \mid k \geq 1\}$ with unknown distributions is such that

$$\forall\, k \geq 0 : \Big(\mathbf{E}\left(\xi(k)\right) = 0, \quad \mathbf{E}\left(\xi(k)\xi(k)^{\mathrm{T}}\right) = L(k)\Big).$$

Here, $\{L(k) \mid k \geq 0\}$ is the preset sequence of nonrandom diagonal $m \times m$ matrices, in which case

$$\forall\, k \geq 0 : L(k) = \operatorname{diag}\left(l_1^2(k), \ldots, l_m^2(k)\right), \quad \exists K : \overset{m}{\underset{j=1}{\forall}}\, l_j^2(k) < K < \infty.$$

Here the real parameter K is defined a priori. The unknown parameter θ^* belongs to the closed sphere in R^m. Let $\Delta = \{\delta_1, \delta_2, \ldots, \delta_m\} \subseteq R^1$, $\forall \delta_j \in \Delta : \delta_j \in]0, \infty[$ be the set of required dimensions of the confidence parallelepiped, while the quantity $P_c \in]0, 1[$ represents the required value of the confidence factor.

It is necessary to draw up a sequential plan for confidence estimation of the parameter θ^*, which is defined in the following way:

- a random moment of ceasing of the observations $\tau \geq 1$;
- the rule of construction of the confidence parallelepiped $\Xi(\tau)$ in the compact Θ, which satisfies the following conditions:

$$\mathbf{P}_{\theta^*}\left(\theta^* \in \Xi(\tau)\right) \geq P_c, \quad \mathbf{P}_{\theta^*}(\tau < \infty) = 1,$$
$$\forall \theta_1, \theta_2 \in \Xi(\tau) : \overset{m}{\underset{j=1}{\forall}} \left|\langle \theta_1 \rangle_j - \langle \theta_2 \rangle_j\right| \leq \delta_j, \quad \delta_j \in \Delta.$$

From here on, we will denote by the symbol $\|\cdot\|$ the norm of the space in which the compact Θ is embedded. $\langle a \rangle_i$ is the ith component of the vector a. For simplicity, if the sense ambiguity does not arise, instead of $\mathbf{P}_\theta$ we will write the $\mathbf{P}$ and instead of $\mathbf{E}_\theta$ we will write the $\mathbf{E}$.

3. METHOD OF THE SOLUTION

Let for random vector functions $\{A(k, \theta) \mid k \geq 0\}$ in Θ the following condition be fulfilled:

$$\forall\left(\theta_1, \theta_2 \in \Theta, \theta_1 \neq \theta_2, n \geq 1\right) : \sum_{k=0}^{n-1} R_k\left(\theta_1, \theta_2\right)(nm)^{-1} > 0,$$
$$R_k\left(\theta_1, \theta_2\right) = \left[A(k, \theta_1) - A(k, \theta_2)\right] L^{-1}(k) \left[A(k, \theta_1) - A(k, \theta_2)\right]^{\mathrm{T}}. \tag{1}$$

Let $\Theta^* = \{\theta_n \mid n \geq 1\}$ be the sequence of the least-squares method (LSM) estimates of the parameter θ^*, which admits the notation:

$$\forall n \geq 1 : \theta_n = \operatorname{Arg} \inf_{\theta \in \Theta} I(n, \theta),$$

where

$$I(n, \theta) = \sum_{k=1}^{t(n)} \left[X(k) - A(k, \theta)\right] L^{-1}(k) \left[X(k) - A(k, \theta)\right]^{\mathrm{T}} (t(n)m)^{-1}.$$

Here, $\{t(n)\} \subseteq N$, $\lim\limits_{n \to \infty} t(n) \to \infty$. For each $n \geq 1$ we will define the functional $\Phi\left(n, \theta, \theta_n\right)$ and the sequence of closed sets $\{\Xi(n) \mid n \geq 1\} \subseteq \Theta$, such that

$$\Phi\left(n, \theta, \theta_n\right) = I(n, \theta) - I(n, \theta_n), \quad \forall\left(n \geq 1, \theta \in \Xi(n)\right) : \Phi\left(n, \theta, \theta_n\right) \leq c(n).$$

Here, $\{c(n) \mid n \geq 1\}$ is the known sequence of nonrandom functions. For each $n \geq 1$ we will define elements $\theta_L(n)$, $\theta_U(n) \in \Theta$, such that

$$\forall\left(n \geq 1, \theta \in \Xi(n)\right) : \overset{m}{\underset{j=1}{\forall}} \left[\langle \theta_L(n) \rangle_j \leq \langle \theta \rangle_j \leq \langle \theta_U(n) \rangle_j\right].$$

The sequence of sets $\{\Xi^*(n) \mid n \geq 1\} \subseteq \Theta$ admits the following notation:

$$\forall n \geq 1 : \Xi^*(n) = \left\{\theta \mid \theta \in \Theta, \overset{m}{\underset{j=1}{\forall}} \left[\langle\theta_L(n)\rangle_j \leq \langle\theta\rangle_j \leq \langle\theta_U(n)\rangle_j\right]\right\}.$$

We will consider the sequential plan for the confidence estimation of the parameter θ^* in the form of the pair $(\gamma(n), \tau)$, such that

$$\forall n \geq 1 : \gamma(n) = \theta_U(n) - \theta_L(n), \quad \tau = \inf\left\{n \geq 1 \middle| \overset{m}{\underset{j=1}{\forall}} \left(\langle\gamma(n)\rangle_j \leq \delta_j\right)\right\}.$$

Properties of the plan $(\gamma(n), \tau)$ are defined by the following theorem.

Theorem 1. *Let the following conditions be fulfilled*:

(1) $\overset{m}{\underset{j=1}{\forall}} \delta_j > 0,\ \delta_j \in \Delta$;

(2) *for the sequence* $\{A(k, \theta) \mid k \geq 1\}$ *the conditions* (1) *are fulfilled;*

(3) $\forall\, k \geq 0\, \exists\, L_0 \in]0, \infty[\,, \quad \forall\, \theta_1,\, \theta_2 \in \Theta\colon \overset{m}{\underset{j=1}{\forall}} \left|\langle A\,(k,\, \theta_1)\rangle_j - \langle A\,(k,\, \theta_2)\rangle_j\right| l_j^{-2}(k) \leq L_0 \,\|\theta_1 - \theta_2\|$,

where $\|\cdot\|$ *is the norm in* R^m;

(4) $t(n) = \left[n^{2+r}\right] + 1,\ n \geq 1, [a]$ *is the integral part of the parameter* $a,\ r > 0$;

(5) *for some* $s, z \in]0, \infty[$ *there exist* $\rho,\ g \in]0, \infty[$, *such that*

$$\forall\,(\theta, \theta' \in \Theta, n \geq s) : \mathbf{G}\,(n, \theta, \theta') \geq g \left\|\theta - \theta'\right\|^2,$$
$$\rho = \inf\left\{n \geq s \mid g(\delta^*/2)^2 - 2c(n) \geq z\right\},$$

where $\delta^* = \mathrm{Inf}\ \{\delta_i \mid \delta_i \in \Delta\}$, $\mathbf{G}\,(n, \theta_1, \theta_2) = \sum\limits_{k=1}^{t(n)} R_k\,(\theta_1, \theta_2)\,(t(n)m)^{-1}, n \geq 1$;

(6) *for the preset* $\Pi, W \in]0, \infty[$: $\mathbf{E}_{\theta^*}\left(\sup\limits_{\theta_1,\theta_2\in\Theta} (\eta^4(n, \theta_1, \theta_2))\right) \leq \Pi n^{-p}$, $\sum\limits_{n\geq 1} n^{-p} \leq W$, *where* $\eta(n, \theta_1, \theta_2) = \sum\limits_{k=1}^{t(n)} 2\xi(k+1)L^{-1}(k)\,(A(k, \theta_2) - A(k, \theta_1))^{\mathrm{T}} / (t(n)m)$.

Further, if the sequence $\{c(n) \mid n \geq 1\}$ *is such that*

$$\forall n \geq 1 : c(n) = \frac{4L_0qK^{0.5}}{(m(1-P_c))^{0.5}} n^{-r/2} \left(\frac{\pi^2}{6}\right)^{0.5},$$

where $q = \sup\limits_{\theta_1,\theta_2\in\Theta} \|\theta_1 - \theta_2\|$, *then the following assertions will be true:*

(1) $\forall\theta^* \in \Theta : \mathbf{P}_{\theta^*}\,(\theta^* \in \Xi^*(\tau)) \geq P_c$;

(2) $\forall\theta^* \in \Theta : \mathbf{P}_{\theta^*}\,(\tau < \infty) = 1$;

(3) $\mathbf{E}_{\theta^*}\tau \leq \rho + 2W\Pi z^{-4}$.

Here, $\mathbf{E}_{\theta^*}\tau$ is the mean time of observation in the suggested sequential plane $(\gamma(n), \tau)$. As follows from the statement of the Theorem 1, in the implementation of the sequential plan $(\gamma(n), \tau)$, with probability 1, in the final moment of ceasing of the observations τ, the confidence parallelepiped $\Xi^*(\tau)$ will be constructed, which will contain, with the required confidence factor P_c, the estimable parameter $\theta^* \in \Theta$. In this case, the dimensions of the confidence parallelepiped $\Xi^*(\tau)$ will have the required size. In view of the nonlinearity of the functions $\{A(k, \theta) \mid k \geq 1\}$, the quantities

$\theta_U(n), \theta_L(n)$, in the general case, cannot be fixed analytically. One of the possible methods of estimation of these quantities is the approach based on the solution of a group of extremal problems. In this case, the quantities $\theta_U(n), \theta_L(n)$ can be defined in the following way:

$$\forall(n \geq 1, j \in \{1, \dots, m\}) : \left\{ \langle\theta_U(n)\rangle_j = \operatorname{Arg} \sup_{\theta \in \Xi^*(n)} (\langle\theta\rangle_j), \langle\theta_L(n)\rangle_j = \operatorname{Arg} \inf_{\theta \in \Xi^*(n)} (\langle\theta\rangle_j) \right\}.$$

4. INFORMAL DISCUSSION AND THE EXAMPLE

The principal idea of the method is based on the investigation of the behavior of the loss function $I(n, \theta)$ in the neighborhood of the minimum point θ_n. For each $t(n)$, the size of the confidence set for the nonlinear parameter θ^* is defined by the quantity $c(n)$ and the sensitivity of the function $I(n, \theta)$ to variations of the parameter θ in the neighborhood of the point θ_n. The higher the quantities q (sphere diameter Θ), K (upper bound for noise dispersion), P_c (confidence factor), and L_0 (Lipschitz constant), the higher the value of the quantity $c(n)$, and hence, the larger the size of the confidence set $\Xi^*(n)$ defined for a specific value of the quantity $t(n)$. As the power of the observed sample increases, the quantity $t(n)$ grows, which is the cause of a decrease of a value of the quantity $c(n)$ corresponding to it. In this case, the size of the confidence set $\Xi^*(n)$ steadily decreases tending to zero at $t(n) \to \infty$. The assumption of the existence for functions $\{A(k, \theta)\}_{k>0}$ of the Lipschitz constant in the compact Θ is common and is used in a considerable number of works devoted to the investigation of nonlinear models of the regression analysis. For example, similar assumptions were used in works [9, 11] in the proof of the strong consistency of point estimates by the least-squares method for nonlinear regression models. Following the highly convex version of the Weistrass theorem, it is easy to see that for the fulfilment of the condition (5), it is sufficient that the functions $\sqrt{\mathbf{G}(n, \theta, \theta')}$, $n \geq 1$ be heavily convex in the compact Θ with the coefficient $2(gq)^{0.5}$. The higher the value of the parameter g by the absolute value, the more sensitive the function $\mathbf{G}(n, \theta, \theta')$ to variations of the parameter θ, the larger amount of the information that is contained in the observed sample, and the smaller the volume of the sample $t(n)$ that will be necessary for the construction of the confidence set with the required sizes at the fixed confidence factor.

We will consider a simple example. In work [11] the model was examined for the generation of observations of the form $y(t) = \exp(-\theta x(t)) + \xi(t)$, where $\theta \in [\alpha, \beta] = \Theta, \alpha > 0, \{x(t)\}$ is the sequence of bounded positive regressors. We will assume in addition that $\forall t \geq 0 : (\mathbf{E}\xi(t) = 0, \mathbf{E}\xi^2(t) = 1$, $\mathbf{E}(\xi_j^4(t)) = \mu_4)$, $\{x(t)\}$ are random independent quantities and for the preset quantities A and B: $P(x(t) \in [\mathrm{A}, \mathrm{B}]) = 1$. In this case,

$$R_k(\theta_1, \theta_2) = (\exp(-\theta_2 x(k)) - \exp(-\theta_1 x(k)))^2,$$

$$\eta(n, \theta_1, \theta_2) = \sum_{k=1}^{t(n)} 2\xi(k) (\exp(-\theta_2 x(t)) - \exp(-\theta_1 x(t))) t^{-1}(n).$$

Following [11], for each θ_2, $\theta_1 \in [\alpha, \beta]$ there exist constants c_2, $c_1 > 0$, such that

$$c_1|\theta_2 - \theta_1|x(t) \leq |\exp(-\theta_2 x(t)) - \exp(-\theta_1 x(t))| \leq c_2|\theta_2 - \theta_1|x(t) \leq c_2\mathrm{B}|\theta_2 - \theta_1|,$$

and the conditions (2) and (3) of the theorem are fulfilled.

Further, the following inequalities are evident:

$$\mathbf{G}(n, \theta, \theta') = \sum_{k=1}^{t(n)} R_k(\theta, \theta') t(n)^{-1} \geq c_1^2 |\theta_2 - \theta_1|^2 x^2(t) \geq c_1^2 \mathrm{A}^2 |\theta_2 - \theta_1|^2,$$

and the condition (5) is met of the Theorem 1 for the process $y(t)$. It is evident that $\forall\theta_1,\theta_2 \in [\alpha,\beta]$: $\mathbf{E}\eta(n,\theta_1,\theta_2) = 0$, and using the Dharmadhikari–Jogdeo theorem [15], we obtain the following upper bound:

$$\mathbf{E}\left(\eta(n,\theta_1,\theta_2)\right)^4 \leq R(4)t^{-3}(n)\sum_{k=1}^{t(n)}\mu_4 c_2^4 \left|\theta_2-\theta_1\right|^4 \mathrm{B}^4 = U(4)t^{-2}(n)\mu_4 c_2^4 \left|\theta_2-\theta_1\right|^4 \mathrm{B}^4.$$

Here $U(p) = \dfrac{1}{2}p(p-1)\max(1,2^{p-3})\left(1+\dfrac{2}{p}K_{2m}^{(p-2)/2m}\right)$, $K_{2m} = \sum\limits_{r=1}^{m}\dfrac{r^{2m-1}}{(r-1)!}$, the integer m is such that $2m \leq p < 2m+2$. In this case we have $\Pi = U(4)\mu_4 c_2^4 \left|\theta_2-\theta_1\right|^4 \mathrm{B}^4$ and $W = \sum\limits_{n\geq 1} n^{-2} = \pi^2/6$. Thus, the constants Π, W from the condition (6) of the Theorem 1 are defined, and so for the model of the considered example, all conditions of this theorem are fulfilled.

5. RESULTS OF NUMERICAL MODELING

For the numerical modeling use was made of the following model of observations:

$$y(t) = H\exp(-\theta x(t)) + \xi(t).$$

Here $\{\xi(t)\}$ is the sequence of independent random quantities that have the distribution $N(0,1)$, $\{x(t)\}$ is the sequence of random quantities that are uniformly distributed in the interval $[0,1]$, the parameter is $\theta^* \in [0,1]$. The confidence factor P_c was chosen equal to 0.95. In this case, $\forall n \geq 1 : c(n) = 4\pi H\,(n\,0.3)^{-0.5}$. The required size of the confidence interval δ was taken equal to 0.05.

The observations ceased at a random instant of time $\tau = \inf\{n \geq 1 \mid \theta_U(n) - \theta_L(n) \leq \delta\}$, where $\theta_U(n)$ is the upper bound of the confidence interval, $\theta_L(n)$ is the lower bound of the confidence interval. Thus, at the instant of ceasing of observations, the required size of the confidence interval for the parameter θ^* was reached.

The estimate $\overline{\tau}(\theta^*)$ of the quantity of the mean time of observation in the sequential plan was defined as a mean value of the times of observation in implementations of a group of the numerical experiments performed for a fixed value of the pair of model parameters H and θ^*. Each group of experiments that corresponds to the next pair of the parameters (H,θ^*), consisted of 20 series. In addition, for each group we define the following: the mean lower bound of the confidence interval $(\overline{\theta}_L(\tau))$ and the mean upper bound of the confidence interval $(\overline{\theta}_U(\tau))$.

The numerical modeling results are collected in the table. As follows from the obtained results, the quantity $\overline{\tau}(\theta^*)$ is defined to a large extent not only by a value of the parameter H, but also

Confidence intervals for the parameter θ^*

H	θ^*	$\overline{\tau}(\theta^*)$	Mean lower bound	Mean upper bound
20	0	9	−0.03	0.01
	0.1	9	0.08	0.12
	0.15	9	0.11	0.16
	0.2	28	0.18	0.23
	0.25	65	0.22	0.27
10	0	65	−0.03	0.02
	0.1	65	0.07	0.12
	0.15	128	0.12	0.16
	0.2	128	0.18	0.23
	0.25	217	0.23	0.28

by a value of the estimable parameter θ^*. The parameter H defines the signal-to-noise ratio in the observable sample, hence large values of H at the fixed θ^* correspond to lower values of the quantity $\overline{\tau}(\theta^*)$ and conversely: lower values of the parameter correspond to large values of the mean time of observation $\overline{\tau}(\theta^*)$. Thus, the supposition that is obvious from the viewpoint of the common sense is confirmed: to reach the preset size of the confidence interval at the fixed P_c and θ^* in the case of a higher value of the signal-to-noise ratio, it is necessary to have a lower mean volume of the sample in comparison with the situation corresponding to a lower value of the signal-to-noise ratio.

On the other hand, the relationship between a value of the parameter θ^* and the quantity $\overline{\tau}(\theta^*)$ at the fixed P_c and H is defined by the parametric sensitivity of the functional $I(n,\theta)$ in the neighborhood of the point θ^*. In conformity with [16], the parametric sensitivity of the functional $I(n,\theta)$ is specified by the sensitivity coefficient $k(n,\theta)$, which is defined as a derivative of the function $I(n,\theta)$ with respect to the θ.

Reasoning in the informal way, the character of the dependence of the quantity $\overline{\tau}(\theta^*)$ on a value of the parameter θ^* is rather evident: the higher the value of the function $k(n,\theta)$ in the neighborhood of the point θ^*, the lower is the value of the mean time of observation $\overline{\tau}(\theta^*)$. In other words, the higher the sensitivity of the functional to variations of the parameter θ in the neighborhood of the point θ^*, the lower the mean volume of the sample will be taken to construct the confidence interval of the prescribed size at the fixed confidence factor P_c.

6. CONCLUSIONS

In this article, we suggested the method of developing confidence sets with prescribed dimensions for the parameter that nonlinearly enters into observation equations. The solution is obtained in the nonasymptotic statement at the incomplete prior definiteness relative to the distribution of heteroscedastic observations. The suggested algorithm is synthesized in the context of the sequential analysis, is based on estimates of the least-squares method, and can be generalized in the event of dependent observations. Upper bounds are obtained for a value of the mean time of observation in the suggested sequential plan. The results of numerical modeling demonstrate the performance of the suggested method.

APPENDIX

Proof of Theorem 1. In the case when the expression X implicates the expression Y, we will write X $\Rightarrow$ Y, but in the situation when the event $\omega(1)$ entails the event $\omega(2)$, we will write $\omega(1) \subset \omega(2)$. For any θ_1, $\theta_2 \in \Theta$, $n \geq 1$ we will consider the quantity

$$\eta(n,\theta_1,\theta_2) = \sum_{k=1}^{t(n)} 2\xi(k)L^{-1}(k)\left(A(k,\theta_2) - A(k,\theta_1)\right)^{\mathrm{T}} / \left(t(n)m\right).$$

It is evident that

$$\forall\theta_1,\theta_2 \in \Theta : \mathbf{E}\left(\eta^2(n,\theta_1,\theta_2)\right) = 4\mathbf{E}\left(\sum_{k=1}^{t(n)}\sum_{j=1}^{m}\left(\frac{\langle\xi(k)\rangle_j^2}{l_j^2(k)}\right)\left(\frac{\langle\mathbf{A}(k,\theta_1)\rangle_j - \langle\mathbf{A}(k,\theta_2)\rangle_j}{l_j(k)t(n)m}\right)^2\right)$$

$$= 4\left(\sum_{k=1}^{t(n)}\sum_{j=1}^{m}\left(\frac{\langle\mathbf{A}(k,\theta_1)\rangle_j - \langle\mathbf{A}(k,\theta_2)\rangle_j}{l_j(k)t(n)m}\right)^2\mathbf{E}\left(\frac{\langle\xi(k)\rangle_j^2}{l_j^2(k)}\right)\right) = 4\sum_{k=1}^{t(n)} R_k(\theta_1,\theta_2)(t(n)m)^{-2}.$$

It is easy to see that

$$\begin{aligned}\sup_{\theta_1,\theta_2\in\Theta} R_k(\theta_1,\theta_2) &= \sup_{\theta_1,\theta_2\in\Theta}\left(\left(\mathbf{A}(k,\theta_2)-\mathbf{A}(k,\theta_1)\right)L^{-1}(k)\left(\mathbf{A}(k,\theta_2)-\mathbf{A}(k,\theta_1)\right)^{\mathrm{T}}\right)\\ &= \sup_{\theta_1,\theta_2\in\Theta}\left(\sum_{j=1}^{m}\frac{\left(\langle\mathbf{A}(k,\theta_1)\rangle_j-\langle\mathbf{A}(k,\theta_2)\rangle_j\right)^2}{l_j^2(k)}\right)\\ &\le \sup_{\theta_1,\theta_2\in\Theta}\left(L_0\left\|\theta_1-\theta_2\right\|\sum_{j=1}^{m}\left|\langle\mathbf{A}(k,\theta_1)\rangle_j-\langle\mathbf{A}(k,\theta_2)\rangle_j\right|\right)\\ &\le \sup_{\theta_1,\theta_2\in\Theta}\left(mKL_0^2\left\|\theta_1-\theta_2\right\|^2\right)\le mKL_0^2q^2.\end{aligned}$$

Thus, the following notation is admissible:

$$\forall\theta_1,\theta_2\in\Theta:\mathbf{E}\left(\eta^2(n,\theta_1,\theta_2)\right)=4\sum_{k=1}^{t(n)}\left(R_k(\theta_1,\theta_2)\right)(t(n)m)^{-2}\le 4KL_0^2q^2\left(mt(n)\right)^{-1}.$$

We will denote for brevity

$$\varepsilon(n)=\mathbf{E}_{\theta^*}\left(\eta(n,\theta_n,\theta)\right),\quad \chi(n)=\eta(n,\theta_n,\theta)-\varepsilon(n),\quad \theta\in\Theta.$$

Further, for any $\theta\in\Theta$

$$\begin{aligned}&P_{\theta^*}\left(|\eta(n,\theta_n,\theta)|\le c(n)\right)\\ &=P_{\theta^*}\left(|\chi(n)+\varepsilon(n)|\le c(n)\right)\ge P_{\theta^*}\left(|\chi(n)|+\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\le c(n)\right)\\ &=P_{\theta^*}\left(|\chi(n)|\le c(n)-\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\right).\end{aligned}$$

Next, $\forall\theta\in\Theta:\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|^2\le 4KL_0^2q^2\left(mt(n)\right)^{-1}$ and in view of the Lyapunov inequality, the following notation is admissible:

$$\forall\theta\in\Theta:\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\le\left(\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|^2\right)^{-0.5}\quad\text{and therefore}$$
$$\forall\theta\in\Theta:\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\le 2K^{0.5}L_0q\left(mt(n)\right)^{-0.5}.$$

Further, $(t(n))^{-0.5}=\left(\left[n^{2+r}\right]+1\right)^{-0.5}\le\left[n^{2+r}\right]^{-0.5}\le n^{-r/2}$, $n>0$ and therefore

$$\forall\theta\in\Theta:\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\le 2K^{0.5}L_0q\left(mt(n)\right)^{-0.5}\le c(n)\left(1-P_c\right)^{0.5}\left(\pi^2/6\right)^{-0.5}/2.$$

It is evident that $(1-P_c)^{0.5}<1$ and $\left(\pi^2/6\right)^{-0.5}<1$, therefore $\forall\theta\in\Theta:\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\le c(n)/2$ and for a certain $t=c(n)/2-\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|>0$ the following notation is admissible:

$$\begin{aligned}&P_{\theta^*}\left(|\eta(n,\theta_n,\theta)|\le c(n)\right)\ge P_{\theta^*}\left(|\chi(n)\le c(n)|-\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\right)\\ &=P_{\theta^*}\left(|\chi(n)|\le c(n)/2+t\right)\ge P_{\theta^*}\left(|\chi(n)|\le c(n)/2\right).\end{aligned}$$

Further,

$$\forall\theta\in\Theta:P_{\theta^*}\left(|\eta(n,\theta_n,\theta)|>c(n)\right)<P_{\theta^*}\left(|\chi(n)|>c(n)/2\right).$$

Using now the Chebyshev inequality, we write

$$\begin{aligned}&P_{\theta^*}\left(|\chi(n)|>c(n)/2\right)\le 4c^{-2}(n)\mathbf{E}_{\theta^*}\left(\chi(n)\right)^2\\ &\le 4c^{-2}(n)\mathbf{E}_{\theta^*}\sup_{\theta_1,\theta_2\in\Theta}\left(\eta^2(n,\theta_1,\theta_2)\right)\le 16c^{-2}(n)KL_0^2q^2\left(mt(n)\right)^{-1},\end{aligned}$$

and the inequality takes place:

$$\forall\theta \in \Theta : P_{\theta^*}\left(|\eta(n, \theta_n, \theta)| > c(n)\right) \leq 16c^{-2}(n)KL_0^2q^2\left(mt(n)\right)^{-1}. \tag{A.1}$$

The following notation is admissible:

$$\Phi(n, \theta, \theta_n) = \mathbf{G}\left(n, \theta^*, \theta\right) - \mathbf{G}\left(n, \theta^*, \theta_n\right) + \eta(n, \theta_n, \theta) \geq 0, \quad n \geq 1, \quad \theta \in \Theta.$$

If $\theta = \theta^*$, then $\mathbf{G}\left(n, \theta^*, \theta\right) - \mathbf{G}\left(n, \theta^*, \theta_n\right) \leq 0$ and the parameter θ^* will belong to the set $\Xi(n)$ only in the case when the quantity $\eta(n, \theta_n, \theta^*)$ is such that

$$I(n, \theta^*) - I(n, \theta_n) = \mathbf{G}\left(n, \theta^*, \theta^*\right) - \mathbf{G}\left(n, \theta^*, \theta_n\right) + \eta(n, \theta_n, \theta^*) > c(n).$$

However, we will consider the situation when $|\eta(n, \theta_n, \theta^*)| \leq c(n)$. In this case, it is evident that $I(n, \theta^*) - I(n, \theta_n) \leq c(n)$ because $\operatorname{sgn}\left(I(n, \theta^*) - I(n, \theta_n)\right) = \operatorname{sgn}(\mathbf{G}(n, \theta^*, \theta^*) - \mathbf{G}(n, \theta^*, \theta_n)) = -1$. In other words, if $|\eta(n, \theta_n, \theta^*)| \leq c(n)$, then $\theta^* \in \Xi(n)$ and in the set $\Xi(n)$, the following implications are valid:

$$\forall n \geq 1 : \left(|\eta(n, \theta_n, \theta^*)| \leq c(n)\right) \Rightarrow \left(\theta^* \in \Xi(n)\right) \Rightarrow \left(\theta^* \in \Xi^*(n)\right). \tag{A.2}$$

Using (A.2), we have

$$\forall n \geq 1 : P\left(|\eta(n, \theta_n, \theta^*)| \leq c(n)\right) \leq P\left(\theta^* \in \Xi^*(n)\right).$$

Using (A.1), it is easy to see that

$$\forall\left(n \geq 1, \theta^* \in \Theta\right) : P\left(\theta^* \notin \Xi^*(n)\right) \leq P\left(|\eta(n, \theta_n, \theta^*)| > c(n)\right) \leq 16c^{-2}(n)KL_0^2q^2\left(mt(n)\right)^{-1}.$$

We will now estimate the probability of unoccurrence of the event $\omega_\tau : \theta^* \notin \Xi^*(\tau)$.

$$\begin{aligned} P\left(\theta^* \notin \Xi^*(\tau)\right) &= \sum_{n\geq 1} P\left(\theta^* \notin \Xi^*(n), \tau = n\right) \\ &\leq \sum_{n\geq 1} P\left(\theta^* \notin \Xi^*(n)\right) \leq \sum_{n\geq 1} 16c^{-2}(n)KL_0^2q^2\left(mt(n)\right)^{-1} \\ &\leq \left(16KL_0^2q^2m^{-1}\right)\sum_{n\geq 1} c^{-2}(n)t^{-1}(n) = (1 - P_c)\left(\frac{\pi^2}{6}\right)^{-1}\sum_{n\geq 1} n^r t^{-1}(n) \\ &= (1 - P_c)\left(\frac{\pi^2}{6}\right)^{-1}\sum_{n\geq 1} n^r\left(\left[n^{2+r}\right] + 1\right)^{-1} \\ &\leq (1 - P_c)\left(\frac{\pi^2}{6}\right)^{-1}\sum_{n\geq 1} n^r n^{-2-r} = 1 - P_c. \end{aligned}$$

Here, account was taken of the fact that $\sum\limits_{n\geq 1} n^{-2} = \dfrac{\pi^2}{6}$. Hence, we have

$$P\left(\theta^* \in \Xi^*(\tau)\right) \geq P_c.$$

It is evident that the rectangular parallelepiped $\Xi^*(\tau)$ is confidential for the parameter θ^*, satisfies the requirements for the statement of the problem, and is completely defined by the vectors $\theta_L(n)$, $\theta_U(n)$, the search for which mainly reduces to the implementation of the suggested sequential plan $(\gamma(n), \tau)$. Thus, the first assertion of the theorem is proved.

Further, taking into account the fact that $\mathbf{G}(n,\theta^*,\theta^*)=0$, $n\geq 1$, we have

$$n\geq 1: I(n,\theta_U(n))-I(n,\theta^*)=\mathbf{G}(n,\theta^*,\theta_U(n))+\mathbf{G}(n,\theta^*,\theta^*)+\eta(n,\theta^*,\theta_U(n))$$
$$=\mathbf{G}(n,\theta^*,\theta_U(n))+\eta(n,\theta^*,\theta_U(n))\geq 0.$$

In a similar way,

$$\forall n\geq 1: I(n,\theta_L(n))-I(n,\theta^*)=\mathbf{G}(n,\theta^*,\theta_L(n))+\eta(n,\theta^*,\theta_U(n))\geq 0.$$

We denote: $H=2\|\theta^*-\theta_U(n)\|\,\|\theta^*-\theta_L(n)\|$ and consider the following events:

$$\omega_U(0,n):\left\{I(n,\theta^*)-I(n,\theta_U(n))\leq g(\delta^*/2)^2\right\},$$
$$\omega_U(1,n):\left\{\mathbf{G}(n,\theta^*,\theta_U(n))=c(n)+\eta(n,\theta^*,\theta_n)+s(n,\theta^*,\theta_U(n))\leq g(\delta^*/2)^2\right\},$$
$$\omega_L(0,n):\left\{I(n,\theta^*)-I(n,\theta_L(n))\leq g(\delta^*/2)^2\right\},$$
$$\omega_L(1,n):\left\{\mathbf{G}(n,\theta^*,\theta_L(n))=c(n)+\eta(n,\theta_L(n),\theta^*)+s(n,\theta^*,\theta_L(n))\leq g(\delta^*/2)^2\right\},$$
$$\omega_U(4,n):\left\{\|\theta^*-\theta_U(n)\|^2\leq(\delta^*/2)^2\right\},$$
$$\omega_L(4,n):\left\{\|\theta^*-\theta_L(n)\|^2\leq(\delta^*/2)^2\right\},$$
$$\omega(*,n):\left\{H\leq(\delta^*)^2/2\right\}.$$

Here,

$$s(n,\theta^*,\theta_U(n))=c(n)-\left(\mathbf{G}(n,\theta^*,\theta_U(n))+\eta(n,\theta^*,\theta_U(n))\right),$$
$$s(n,\theta^*,\theta_L(n))=c(n)-\left(\mathbf{G}(n,\theta^*,\theta_L(n))+\eta(n,\theta^*,\theta_L(n))\right).$$

Because $(\mathbf{G}(n,\theta^*,\theta_U(n))+\eta(n,\theta^*,\theta_U(n)))>0$ and $(\mathbf{G}(n,\theta^*,\theta_L(n))+\eta(n,\theta^*,\theta_L(n)))>0$, we have

$$s(n,\theta^*,\theta_U(n))<c(n),\quad s(n,\theta^*,\theta_L(n))<c(n). \tag{A.3}$$

Because $\|\theta^*-\theta_U(n)\|^2+\|\theta^*-\theta_L(n)\|^2\geq\|\theta_U(n)-\theta_L(n)\|^2-H$, the following implications are valid:

$$\begin{gathered}\left(\|\theta^*-\theta_L(n)\|^2\leq(\delta^*/2)^2\right)\&\left(\|\theta^*-\theta_U(n)\|^2\leq(\delta^*/2)^2\right)\\ \Rightarrow\left(\|\theta^*-\theta_U(n)\|^2+\|\theta^*-\theta_L(n)\|^2\leq(\delta^*)^2/2\right),\\ (H\leq(\delta^*)^2/2)\Rightarrow\left(\|\theta^*-\theta_U(n)\|^2+\|\theta^*-\theta_L(n)\|^2\geq\|\theta_U(n)-\theta_L(n)\|^2-H\right.\\ \left.\geq\|\theta_U(n)-\theta_L(n)\|^2-(\delta^*)^2/2\right)\\ \Rightarrow\left\{\left(\|\theta^*-\theta_U(n)\|^2+\|\theta^*-\theta_L(n)\|^2\leq(\delta^*)^2/2\right)\Rightarrow\left(\|\theta_U(n)-\theta_L(n)\|^2-(\delta^*)^2/2\leq(\delta^*)^2/2\right)\right\}\\ \Rightarrow\left(\|\theta_U(n)-\theta_L(n)\|^2\leq(\delta^*)^2\right)\Rightarrow\left(\|\theta_U(n)-\theta_L(n)\|\leq\delta^*\right).\end{gathered} \tag{A.4}$$

Further, with due regard for the condition (5) of the provable theorem,

$$\omega_U(1,n)\subset\omega_U(4,n),\quad \omega_L(1,n)\subset\omega_L(4,n). \tag{A.5}$$

Using the triangle inequality, in view of (A.4) and (A.5), we can write

$$\begin{gathered}\omega_U(1,n)\omega_L(1,n)\subset\omega_U(4,n)\omega_L(4,n)\subset\omega(*,n)\\ \subset\omega(5,n):\left\{\|\theta^*-\theta_U(n)\|^2+\|\theta^*-\theta_L(n)\|^2\leq(\delta^*)^2/2\right\}\\ \subset\omega(6,n):\left\{\|\theta_U(n)-\theta_L(n)\|\leq\delta^*\right\}\subset\omega(7,n):\{\tau\leq n\}.\end{gathered} \tag{A.6}$$

We will denote the following events:

$$\omega_U(8,n) : \left\{\eta(n,\theta^*,\theta_U(n)) + s(n,\theta^*,\theta_U(n)) \le g(\delta^*/2)^2 - c(n)\right\},$$
$$\omega_L(8,n) : \left\{\eta(n,\theta^*,\theta_L(n)) + s(n,\theta^*,\theta_L(n)) \le g(\delta^*/2)^2 - c(n)\right\},$$
$$\omega_U(9,n) : \left\{|\eta(n,\theta^*,\theta_U(n))| + c(n) \le g(\delta^*/2)^2 - c(n)\right\},$$
$$\omega_L(9,n) : \left\{|\eta(n,\theta^*,\theta_L(n))| + c(n) \le g(\delta^*/2)^2 - c(n)\right\},$$
$$\overline{\omega}_U(9,n) : \left\{|\eta(n,\theta^*,\theta_U(n))| > g(\delta^*/2)^2 - 2c(n)\right\},$$
$$\overline{\omega}_L(9,n) : \left\{|\eta(n,\theta^*,\theta_L(n))| > g(\delta^*/2)^2 - 2c(n)\right\}.$$

Using the Boole inequality and (A.3),

$$\begin{aligned} &\forall n > \rho : P_{\theta^*}\left(\omega_U(1,n) \times \omega_L(1,n)\right) \\ &= P_{\theta^*}\left(\omega_U(8,n) \times \omega_L(8,n)\right) \ge P_{\theta^*}\left(\omega_U(9,n) \times \omega_L(9,n)\right) \\ &\ge 1 - \left(P_{\theta^*}\left(\overline{\omega}_U(9,n)\right) + P_{\theta^*}\left(\overline{\omega}_L(9,n)\right)\right) \\ &\ge 1 - \left(P_{\theta^*}\left(|\eta(n,\theta^*,\theta_U(n))| > z\right) + P_{\theta^*}\left(|\eta(n,\theta^*,\theta_L(n))| > z\right)\right). \end{aligned} \tag{A.7}$$

In view of the condition (6) of the provable theorem, we have

$$\forall\left(n \ge 1; \theta,\theta' \in \Theta\right) : \mathbf{E}_{\theta^*}\left(\eta^4(n,\theta,\theta')\right) \le \mathbf{E}_{\theta^*}\left(\sup_{\theta_1,\theta_2 \in \Theta}(\eta^4(n,\theta,\theta'))\right) \le \Pi n^{-p}.$$

Further, using the Chebyshev inequality, we write

$$\forall\left(n \ge 1; \theta,\theta' \in \Theta\right) : P_{\theta^*}\left(|\eta(n,\theta,\theta')| > z\right) \le \mathbf{E}_{\theta^*}\left(\eta^4(n,\theta,\theta')\right) z^{-4} \le \Pi n^{-p} z^{-4}.$$

Considering (A.7), we can write

$$\begin{aligned} &\forall n > \rho : P_{\theta^*}\left(\omega_U(1,n)\omega_L(1,n)\right) \\ &\ge 1 - \left(P_{\theta^*}\left(|\eta(n,\theta^*,\theta_U(n))| > z\right) + P_{\theta^*}\left(|\eta(n,\theta^*,\theta_L(n))| > z\right)\right) \\ &\ge 1 - 2\Pi n^{-p} z^{-4}. \end{aligned} \tag{A.8}$$

It follows from (A.6) that for any $\alpha \in]0,1[$, the following implications are valid:

$$\left(P_{\theta^*}\left(\omega_U(1,n) \times \omega_L(1,n)\right) \ge \alpha\right) \Rightarrow \left(P_{\theta^*}\left(\tau \le n\right) \ge \alpha\right) \Rightarrow \left(P_{\theta^*}\left(\tau > n\right) < 1 - \alpha\right), \quad n \ge 1.$$

In this case, using (A.8), at $n > \rho$ we have

$$P_{\theta^*}\left(\tau > n\right) < 2\Pi n^{-p} z^{-4}.$$

Therefore, with due regard for the theorem condition, we can write

$$\mathbf{E}_{\theta^*}\tau = \sum_{n \ge 1} P_{\theta^*}\left(\tau > n\right) \le \rho + \sum_{n \ge 1} 2\Pi n^{-p} z^{-4} \le \rho + 2\Pi W z^{-4}.$$

Thus, the third assertion of the theorem is proved. It follows from here, too, that

$$\sum_{n \ge 1} P_{\theta^*}\left(\tau > n\right) < \infty.$$

Resorting now to the Borel–Cantelli lemma, we conclude that

$$\mathbf{P}_{\theta^*}\left(\tau < \infty\right) = 1$$

and so the second assertion of the Theorem 1 is proved. The theorem is proved completely.

REFERENCES

1. Arce, G.R., Nonlinear Signal Processing: A Statistical Approach, Hoboken: Wiley, 2004.
2. McMillan, D.G., Non-linear Forecasting of Stock Returns: Does Volume Help?, *Int. J. Forecasting*, 2007, vol. 23, pp. 115–126.
3. Mills, T.C., Business Cycle Asymmetries and Non-linearities in UK Macroeconomic Time Series, *Ricerche Econom.*, 1995, vol. 49, pp. 97–124.
4. Cross, R., *The Natural Rate of Unemployment: Reflections on 25 Years of the Hypothesis*, London: Cambridge Univ. Press, 1995.
5. Ibragimov, I.A. and Khas'minskii, R.Z., *Asimptoticheskaya teoriya otsenivaniya* (Asumptotic Theory of Estimation), Moscow: Nauka, 1979.
6. Ljung, L., *System Identification—Theory for User*, Upper Saddle River: Prentice Hall, 1999.
7. Lai, T.L., Asymptotic Properties of Nonlinear Least Squares Estimates in Stochastic Regression, *Models. Ann. Statist.*, 1994, vol. 22, no. 4, pp. 1917–1930.
8. Anderson, T.W. and Taylor, J., Strong Consistency of Least Squares Estimation in Dynamic Models, *Ann. Statist.*, 1979, vol. 7, pp. 484–489.
9. Wu, C.F., Asymptotic Theory of Nonlinear Least Squares Estimation, *Ann. Statist.*, 1981, vol. 9, pp. 501–513.
10. Hu, I., Strong Consistency of Bayes Estimates in Stochastic Regression Models, *J. Multivariate Anal.*, 1996, vol. 57, pp. 215–227.
11. Skouras, K., Strong Consistency in Nonlinear Stochastic Regression Models, *Ann. Statist.*, 2000, vol. 28, no. 3, pp. 871–879.
12. Weyer, E. and Campi, M.C., Non-asymptotic Confidence Ellipsoids for the Least Squares Estimate, *Automatica*, 2002, vol. 38, no. 9, pp. 1539–1547.
13. Timofeev, A.V., Non-asymptotic Solution of Confidence-Estimation Parameter Task of a Non-linear Regression by Means of Sequential Analysis, *Probl. Control Inform. Theory*, 1991, vol. 20, no. 5, pp. 341–351.
14. Timofeev, A.V., Nonasymptotic Confidence Estimation of Nonlinear Regression Parameters in the Context of Sequential Analysis, *Avtom. Telemekh.*, 1997, no. 10, pp. 70–76.
15. Dharmadhikari, S.W. and Jogdeo, K., Bounds on Moments of Sums of Random Variables, *Math. Statist.*, 1969, vol. 40, pp. 1506–1509.
16. Rosenvasser, E.N. and Yusupov, R.M., *Chuvstvitel'nost' sistem upravleniya* (Sensitivity of Control Systems), Moscow: Nauka, 1981.

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