Теорія Ймовір. та Матем. Статист.
Вип. 75, 2006

Theor. Probability and Math. Statist.
No. 75, 2007, Pages 121–137
S 0094-9000(08)00719-9
Article electronically published on January 24, 2008

ASYMPTOTIC DISTRIBUTIONS OF LEAST SQUARES ESTIMATORS OF THE COEFFICIENTS IN THE MODEL OF LINEAR REGRESSION WITH NONLINEAR CONSTRAINTS AND LONG-MEMORY DEPENDENCE

UDC 519.21

E. M. MOLDAVS'KA

Abstract. We consider least squares estimators for linear regression models with long-memory dependence, continuous time, and nonlinear inequality constraints imposed on the parameter. We study the solution of the problem of minimization of the least squares functional in the linear regression with a given (long) radius of dependence and nonlinear inequality constraints imposed on the parameter. We prove that the solution being appropriately centered and normalized converges in distribution to the solution of the quadratic programming problem. The latter solution is non-Gaussian in contrast to known results for long-memory dependence without constraints for which an analogous transform of the solution of the minimization problem is asymptotically Gaussian in many typical cases.

1. Introduction

The asymptotic distribution of normalized least squares estimators for linear regression models with long-memory dependence (strong or long-range dependence, in other words) is studied in the paper for the case where nonlinear inequality constraints are imposed on the parameter. There are many examples of applications of such models in practice, since much of the data observed in various fields of science and technology is known to possess the long-memory dependence structure and since a number of problems in the control of industrial processes as well as problems of modelling economic processes lead to regression models with a priori known inequality constraints imposed on the parameters of the models. The presence of inequality constraints in a model allows one to obtain a better fit of the regression model to the data and to improve the quality of the estimators of the parameters of the model.

The asymptotic behavior of least squares estimators for long-memory dependence models (for both discrete and continuous cases) without constraints imposed on the parameters is studied by many authors. Taqqu [16] and Dobrushin and Major [5] proved the noncentral limit theorem that characterizes such models. The further asymptotic theory of least squares estimators for long-memory dependence models without constraints is due to Yajima [18], Künsch, Beran, and Hampel [10], Dahlhaus [4], Leonenko and Benšic [12], Leonenko [11], Ivanov and Leonenko [7]–[9].

On the other hand, regression models with independent or weakly dependent errors that involve constraints imposed on the parameters are studied in papers by Korkhin [2],

2000 *Mathematics Subject Classification.* Primary 62E20, 62F10; Secondary 60G18.
Key words and phrases. Long-memory (strong) dependence, linear regression, least squares estimators, inequality constraints, non-Gaussian distributions, asymptotic distribution, continuous time.

Knopov [1], Dupacova and Wets [6], Nagaraj and Fuller [15] for discrete cases and Wang [17] for continuous cases. Korkhin [2] studied a nonlinear regression with inequality constraints, Dupacova and Wets [6] investigate a rather general case of a nonlinear constrained regression, and the paper by Nagaraj and Fuller [15] is devoted to time series models with nonlinear inequality constraints. Wang [17] considers a very general model with nonlinear inequality constraints and equality constraints. Knopov [1] investigates a nonlinear regression with inequality constraints for the case of weakly dependent errors.

Despite a wide class of applied problems involving linear regression long-memory models with continuous time and nonlinear inequality constraints, the asymptotic properties of least squares estimators are not yet known for these models. The main aim of this paper, which is a continuation of [13], is to establish some asymptotic properties of least squares estimators for long-memory dependence models with inequality constraints.

Assumptions used in the statements of the results of this paper are introduced separately and written like **1**, **2**, ...

2. The basic model. First assumptions

We consider the estimator $\beta = (\beta_1, \dots, \beta_n)'$ in the linear regression model with continuous time

$$y(t) = \beta' g(t) + \eta(t), \qquad 0 \le t \le T, \tag{1}$$

and nonlinear constraints

$$h_j(\beta) \le 0, \qquad j = 1, \dots, r, \tag{2}$$

where β is an unknown parameter, $g(t) = [g_1(t), \dots, g_n(t)]'$, $h(\beta) = [h_1(\beta), \dots, h_r(\beta)]'$ are known functions, and $\eta(t)$, $t \in \mathbb{R}$, is a measurable mean square continuous second order stationary random process with zero mean and covariance $B_\eta(t)$, $t \in \mathbb{R}$.

In what follows we consider the case where the process $\eta(t)$ is subordinated to a long-memory Gaussian process; namely, we assume the following (see, for example, Taqqu [16]).

1. Let $\eta(t) = G(\varepsilon(t))$, $t \in \mathbb{R}$, be a random process where G is an arbitrary real-valued measurable nonrandom function and let $\varepsilon(t)$, $t \in \mathbb{R}$, be a Gaussian random process with zero mean and covariance $B_\varepsilon(t) = (1+t^2)^{-\alpha/2}$, $0 < \alpha < 1$. Assume that $\mathsf{E}\, G^2(\varepsilon(0)) < \infty$.

The model described above is rather general. Below we provide some examples. Denote by $\Phi(\cdot)$ the standard Gaussian distribution function and let U be the uniform distribution on the interval $[0;1]$. In what follows the symbol $\overset{d}{=}$ denotes the equality in distribution. According to the Smirnov transform $\Phi(\varepsilon(t)) \overset{d}{=} U$. Thus $\eta(t) \overset{d}{=} G(\varepsilon(t))$ and the marginal distribution of $G(u) = u\sigma_\eta + \mu_\eta$ is $\mathcal{N}(\mu_\eta, \sigma_\eta^2)$ if the marginal distribution of the random process $\eta(t) \overset{d}{=} \Phi(\varepsilon(t))$ is uniform on the interval $[0;1]$. Moreover every increasing distribution function F_η can be obtained in this way by considering $\eta(t) = F_\eta^{-1} \circ \Phi(\varepsilon(t))$ where "$\circ$" means the superposition of functions.

Parametric regression models (1) with discrete or with continuous time and long-memory errors without constraints (2) are considered by Dobrushin and Major [5], Yajima [18], Künsch, Beran, and Hampel [10], Dahlhaus [4], and Leonenko [11] (also see the references therein). On the other hand, regression models (1) with independent or weakly dependent errors and constraints (2) are considered by Korkhin [2] and Knopov [1].

If Assumption 1 holds, then it is well known that the function G admits the following expansion in the Hilbert space $L_2(\mathbb{R}, \phi(u)\, du)$ in terms of the Chebyshev–Hermite polynomials with the unit leading coefficient

$$G(u) = \sum_{k=0}^{\infty} \frac{C_k}{k!} H_k(u), \qquad C_k = \int_{\mathbb{R}} G(u) H_k(u) \phi(u)\, du, \tag{3}$$

where $\{H_k(u)\}_{k=0}^{\infty}$ are the Chebyshev–Hermite polynomials; that is,

$$H_k(u) = (-1)^k e^{u^2/2} \frac{d^k}{du^k} e^{-u^2/2}, \qquad k = 0, 1, \dots,$$

and $\phi(u)$ is the probability density of $\varepsilon(0)$:

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \qquad u \in \mathbb{R}.$$

Note that $C_0 = 0$, since $\mathsf{E}\,\eta(t) = 0$.

Let an integer number $m \geq 1$ be such that $C_1 = \cdots = C_{m-1} = 0$, and $C_m \neq 0$. We write $m = \operatorname{rank} G$ in this case.

2. Assume that $g_i(t) > 0$, $t > 0$, and $g_i(t)$ are bounded functions on an interval $[0; T]$, $1 \leq i \leq n$.

Let $d(T) = \operatorname{diag}(d_1(T), \dots, d_n(T))$, where

$$d_i(T) = \sqrt{\int_0^T g_i^2(t)\, dt}, \qquad \underline{\lim} T^{-1} d_{iT}^2(\theta) > 0, \qquad T \to \infty,\ i = 1, \dots, n.$$

The case where the limit equals ∞ is not excluded.

Consider the following matrices:

$$J_T = \left(J_{il,T} \right)_{i,l=1}^{n}, \qquad J_{il,T} = d_{iT}^{-1} d_{lT}^{-1} \int_0^T g_i(t) g_l(t)\, dt,$$

$$D_T^2 = \frac{1}{T} d_T^2 = \operatorname{diag}\left(\int_0^1 g_i^2(Tt)\, dt \right)_{i=1}^{n}, \qquad \Lambda_T = \left(\Lambda_T^{il} \right)_{i,l=1}^{n} = J_T^{-1},$$

$$\sigma_{T,m} = D_T^{-1} \left(\int_0^1 \int_0^1 \frac{g(Tt) g'(Ts)}{|t-s|^{\alpha m}}\, dt\, ds \right) D_T^{-1}.$$

3. $\alpha m < 1$ where $\alpha \in (0,1)$ is the parameter of the correlation function B.

We have

$$J_T = d_T^{-1} \left(\int_0^T g(t) g(t)'\, dt \right) d_T^{-1} = D_T^{-1} \left(\int_0^1 g(Tt) g(Tt)'\, dt \right) D_T^{-1}.$$

4. Let

$$\lim_{T \to \infty} \sigma_{T,m} = \sigma_m, \tag{4}$$

where σ_m is some positive definite matrix.

5. Let there exist the limit

$$\lim_{T \to \infty} J_T = J_0, \tag{5}$$

where J_0 is some positive definite matrix.

Put $R_T = J_T \sigma_m^{-1} J_T$,

$$R_0 = m!\, J_0 \sigma_m^{-1} J_0. \tag{6}$$

6. Let

1) h_j, $1 \le j \le r$, have the derivatives of the first and second order that are bounded in a neighborhood of the true value β_0;
2) $h_j(\beta_0) = 0$, $j \in \{1, \dots, q\}$, $h_j(\beta_0) < 0$, $j \in \{q+1, \dots, r\}$;
3) there exist β^* such that $h(\beta^*) < 0$;
4) the vectors $\nabla h_j(\beta_0)$, $j \in \{1, \dots, q\}$, be linearly independent;
5) $h_j(\beta)$, $j = 1, \dots, r$, be convex.

Put

$$A_j(T) = T^{-1/2} d_{jT} = \sqrt{\int_0^1 g_j^2(tT)\, dt}.$$

7. For $j = 1, \dots, n$, there are functions $\bar{g}_j(t)$, $t \in [0,1]$, that are square integrable in $t \in [0,1]$ and such that

$$\lim_{T\to\infty} \left| \frac{g_j(tT)}{A_j(T)} - \bar{g}_j(t) \right| = 0 \tag{7}$$

uniformly in $t \in [0,1]$.

Note that Assumption 4 follows from Assumptions 7, since $\sigma_m = (\sigma_{lj}))_{l,j=1}^n$ is a positive definite matrix, where

$$\sigma_{lj} = \int_0^1 \int_0^1 \bar{g}_l(t) \bar{g}_j(s) \frac{dt\, ds}{|t-s|^{m\alpha}}.$$

8. For $\alpha m < 1$ and $j = 1, \dots, n$,

$$K_{jm} = \int_{R^m} |Y_j(\lambda_1 + \dots + \lambda_m)|^2 \frac{d\lambda_1 \dots d\lambda_m}{|\lambda_1 \dots \lambda_m|^{1-\alpha}} < \infty,$$

where

$$Y_j(\lambda) = \int_0^1 \bar{g}_j(t) e^{it\lambda}\, dt, \qquad \lambda \in \mathbb{R}, \tag{8}$$

and $\bar{g}_j(t)$ is defined in Assumption 7.

Let Assumptions 1, 3, 7, and 8 hold and consider random variables

$$Q_{jm} = \frac{C_m}{m!} [c(\alpha)]^{m/2} \int_{\mathbb{R}^m}' Y_j(\lambda_1 + \dots + \lambda_m) \frac{W(d\lambda_1) \dots W(d\lambda_m)}{|\lambda_1 \dots \lambda_m|^{(1-\alpha)/2}},$$

where $c(\alpha) = (2\Gamma(\alpha) \cos(\alpha\pi/2))^{-1}$, C_m are defined by (3), Y_j is defined by (8), $\int_{\mathbb{R}^m}' \dots$ is the multiple stochastic Wiener–Itô integral, W is a complex-valued Gaussian white noise, and the hyperplanes $\lambda_i = \pm\lambda_j$, $i, j = 1, \dots, m$, $i \ne j$, are excluded from the integration domain in the above multiple stochastic integral. Note that the random variables Q_{jm} possess finite second moments.

It follows from Assumption 5 that the limits

$$\Lambda_0^{lj} = \lim_{T\to\infty} \Lambda_T^{lj}$$

exist. Let

$$Q_m = \left(\sum_{j=1}^q \Lambda_0^{lj} Q_{jm} \right)_{l=1}^n. \tag{9}$$

9. Assume that $\widehat{\beta}_T$ (a solution of the minimization problem) is a consistent estimator.

Remark 2.1. Assumptions 2–5 and 7–8 were proposed by Ivanov and Leonenko [8] for a more general setting. We use these assumptions in order to apply the central and noncentral limit theorems for processes with long-memory dependence proved in the paper [8].

Remark 2.2. Assumption 9 holds if, for example, the assumptions of Theorem 3.3 of [3] (or Theorem 3 of [14]) hold. Below we use this assumption for $J = \{\beta\colon h(\beta) \le 0\}$. According to the result mentioned above, the estimator $\widehat{\beta}_T$, being a solution of the minimization problem, is a consistent estimator of the parameter β. Moreover $\widehat{\beta}_T$ is a strong consistent estimator; that is, $\mathsf{P}\bigl\{\lim_{T\to\infty}\bigl\|\widehat{\beta}_T - \beta_0\bigr\|_p = 0\bigr\} = 1$. However, weak consistency (the convergence in probability) is enough for our purposes (recall that weak consistency is assumed in Assumption 9).

Remark 2.3. Assumptions 6, 3) and 6, 5) imply that the gradients are linearly independent, that is,

$$\nabla h_i(\widehat{\beta}_T), \qquad i \in I_\beta, \quad I_\beta = \{i\colon h_i(\widehat{\beta}_T) = 0, i \in I\}.$$

Assumptions 2–5 hold for the model of polynomial regression. Assumptions 7–8 hold, for example, in the case of $n = 1$, $g(t) = \log(e + t)$, and the quadratic function $h(\beta)$.

Now we are ready to state our main result.

3. The statement of the main result

Theorem 3.1. *Let Assumptions* 1, 2, 5, 6, *and* 9 *hold. We also assume that Assumption* 4 *holds for* $m = 1$. *Let a random vector* $\widehat{\beta}_T$ *be a solution of the following minimization problem:*

$$\begin{cases} \int_0^T [y(t) - \beta' g(t)]^2\,dt \longrightarrow \min, \\ h(\beta) \le 0. \end{cases} \tag{10}$$

Then the random vector $U_T = B(T)^{-1/2}T^{-1/2}d_T(\widehat{\beta}_T - \beta_0)$ *converges in distribution as* $T \to \infty$ *to the random vector* U *that is a solution of the following quadratic programming problem:*

$$\begin{cases} \frac{1}{2}X'R_0X - Q'X \longrightarrow \min, \\ \nabla h_j(\beta_0)X \le 0,\ j = 1, \ldots, q, \end{cases} \tag{11}$$

where R_0 *is defined in* (6) *and* Q *is a Gaussian random vector with zero mean and covariance matrix* $C_1^2 J_0 \sigma_1^{-1} J_0$. *Here* σ_1 *and* J_0 *are defined in* (4) *for* $m = 1$ *and in* (5), *respectively.*

Theorem 3.2. *Let Assumptions* 1–9 *hold for* $m > 1$. *Let a random vector* $\widehat{\beta}_T$ *be a solution of the following minimization problem:*

$$\begin{cases} \int_0^T [y(t) - \beta' g(t)]^2\,dt \longrightarrow \min, \\ h(\beta) \le 0. \end{cases} \tag{12}$$

Then the random vector $U_T = B(T)^{-m/2}T^{-1/2}d_T(\widehat{\beta}_T - \beta_0)$ *converges in distribution as* $T \to \infty$ *to the random vector* U *that is a solution of the following quadratic programming problem:*

$$\begin{cases} \frac{1}{2}X'R_0X - Q_m'X \longrightarrow \min, \\ \nabla h_j(\beta_0)X \le 0,\ j = 1, \ldots, q, \end{cases} \tag{13}$$

where R_0 *and* Q_m *are defined in* (6) *and* (9), *respectively.*

Remark 3.1. The most studied case is that of $m = 1$. Note that the random vector Q_m is Gaussian in this case.

The asymptotic behavior of the normalized least squares estimator of the regression coefficients β is studied, for example, in [4] and [8] for models with long-memory dependence and without constraints (2). Note that a more general case of the so-called M-estimators is considered in [8]. The M-estimators are asymptotically Gaussian if $m = 1$ and non-Gaussian if $m \geq 2$. The asymptotic distribution of these estimators in the problem with nonlinear constraints (2) is a solution of the quadratic programming problem (11) and almost always is non-Gaussian even in the case of $m = 1$ (despite the cases discussed below). This is a new phenomenon that did not appear in the regression models earlier.

The normalized least squares estimators are asymptotically non-Gaussian for $m \geq 2$ for both models without constraints (see, for example, Ivanov and Leonenko [8]) and in models of constrained regression.

Consider some particular cases of the function $h(\beta)$ for $m = 1$.

1) $h_j(\beta) = 0$, $j = 1, \dots, r$. Assumption 6, 3) does not hold in this case. The inequality constraints in (11) are satisfied for all β and therefore one should solve the problem (11) without constraints. The minimum is attained at $U = R_0^{-1}Q'$, which is a Gaussian vector. Therefore the normalized least squares estimators are asymptotically Gaussian in this case. Similar reasoning proves that least squares estimators are asymptotically non-Gaussian if $m \geq 2$.
2) $h_j(\beta) < 0$, $j = 1, \dots, r$. All assumptions of Theorem 3.1 hold in this case. Note that $q = 0$ in Assumption 6, 2). Constraints (11) do not make sense at all (they are not active, in other words) and, similarly to case 1), we find that $U = R_0^{-1}Q'$. Again the asymptotic distribution of the normalized least squares estimators is Gaussian. As above, the asymptotic distribution is non-Gaussian if $m \geq 2$.

We see that the models with constraints considered in the cases 1) and 2) above are asymptotically equivalent to those without constraints.

4. Auxiliary results

In what follows we need some results of the paper by Korkhin [2]. We prove some analogs of Korkhin's lemmas for continuous time and for stochastic processes with long-memory dependence (Korkhin [2] considers the case of discrete time and independent errors).

The following lemma provides some properties of random matrices. First we introduce the notion of the modal matrix.

Definition 4.1. Let A be a given matrix. The matrix whose columns are eigenvectors of A is called the *modal matrix* of A.

Lemma 4.1. *Let A_T be a symmetric random matrix of sizes $n \times n$. Assume that A_T converges to a positive definite matrix A as $T \to \infty$.*

Then

1) *the eigenvalues a_{kT}, $k = 1, \dots, n$, of A_T converge in probability to the eigenvalues a_k, $k = 1, \dots, n$, of A as $T \to \infty$;*
2) *the modal matrices C_T of A_T converge in probability to the modal matrix C of A as $T \to \infty$;*
3) $\lim_{T\to\infty} p_T = 1$ *where p_T is the probability that all eigenvalues of the matrix A_T are positive.*

The following result contains some properties of a solution of a quadratic programming problem to be studied below.

Lemma 4.2. *The quadratic programming problem*

$$\min\left\{\frac{1}{2}\beta' R\beta - Q'\beta \;\middle|\; {h_i}'\beta \le 0, i\in I\right\}$$

where $\beta, Q, h_i \in \mathbb{R}^n$ and where R is a positive definite matrix of size $n\times n$ possesses a solution $\beta(Q)$ that is continuous in Q.

Now we obtain some auxiliary results (analogs of the Korkhin lemma). First we consider the case of $m=1$.

Let

$$\Re_T = B^{m/2}(T)T^{1/2}d_T^{-1},$$

$$\widetilde{S}_T(\beta) = \frac{1}{2}\Re_T^{-2}\int_0^T (y(t)-\beta' g(t))^2\,dt.$$

The following relations are consequences of the necessary condition for the existence of an extremum of the problem (10):

$$\nabla S_T(\widehat{\beta}_T) + \sum_{i=1}^m \lambda_{iT}\nabla h_i(\widehat{\beta}_T) = 0, \tag{14}$$

$$\lambda_{iT} h_i(\widehat{\beta}_T) = 0, \qquad \lambda_{iT}\ge 0, \qquad i\in I, \tag{15}$$

where λ_{iT} are Lagrange coefficients. Relations (14)–(15) hold for $\widehat{\beta}_T$, since $h_i(\beta)$, $i\in I$, are convex and there exists β^* such that $h_i(\beta^*)\le 0$.

Let $F_T(\beta) = \widetilde{S}_T(\beta)$. Then

$$F_T(\beta) = \Re_T^{-2}\int_0^T \big(y(t)-\beta' g(t)\big)^2\,dt,$$

$$\left.\frac{\partial S_T(\beta)}{\partial \beta_k}\right|_{\beta=\beta^0} = -{\Re_T}^{-2}\int_0^T \eta(t)g(t)\,dt\Big|_{\beta=\beta^0}. \tag{16}$$

We have $p\lim_{T\to\infty}\nabla S_T(\beta^0)=0$, since the averages of a long-memory process converge to zero. This implies that

$$p\lim_{T\to\infty}\nabla S_T(\widehat{\beta}_T) = 0, \tag{17}$$

since $\widehat{\beta}$ is consistent.

We derive from (15) and Assumption 6, 1) that

$$p\lim_{T\to\infty}\lambda_{iT} = 0, \qquad i\in\{q+1,\dots,r\}, \tag{18}$$

since

$$p\lim_{T\to\infty} h_i(\widehat{\beta}_T) = h_i(\beta^0), \qquad i\in\{1,\dots,r\}, \tag{19}$$

in view of the continuity of $h_i(\beta)$ in β.

Analogously

$$p\lim_{T\to\infty}\nabla h_i(\widehat{\beta}_T) = \nabla h_i(\beta^0), \qquad i\in\{1,\dots,r\}. \tag{20}$$

The gradients $\nabla h_i(\widehat{\beta}_T)$, $i\in I_\beta$, are linearly independent by Assumptions 6, 3) and 5). Thus the system of equations (14) uniquely determines λ_{iT}, $i\in\{1,\dots,q\}\cap I_\beta$ (note that $\lambda_{iT}=0$, $i\in\{1,\dots,q\}\setminus(\{1,\dots,q\}\cap I_\beta)$ according to equalities (15)). The right hand side of system (14) converges in probability to zero in view of relations (17), (18), (20), and Assumptions 6, 1). Then we get from Assumptions 6, 4) that

$$p\lim_{T\to\infty}\lambda_{iT} = 0, \qquad i\in\{1,\dots,q\}, \tag{21}$$

since $q < n$. Thus,

$$p\lim_{T\to\infty} \lambda_{iT} = \lambda_i^0 = 0, \qquad i \in \{1, \dots, r\}. \tag{22}$$

Therefore we have proved the following result.

Lemma 4.3. *Let Assumptions* 1, 6, *and* 9 *hold. Then the Lagrange multipliers of the problem* (10),

$$\lambda_{iT}, \qquad i \in \{1, \dots, r\}, \tag{23}$$

converge in probability to zero as $T \to \infty$.

Put $L_T = \Re_T(\lambda_T - \lambda^0) = \Re_T \lambda_T$. It follows from Lemma 4.3 that

$$U_T = \Re_T(\widehat{\beta}_T - \beta^0),$$

where $\lambda_T = (\lambda_{1T}, \dots, \lambda_{rT})'$ and $\lambda^0 = (\lambda_1^0, \dots, \lambda_r^0)'$. Multiplying by $\Re_T^{-1}$ we obtain from relations (14)–(16) that

$$-\Re_T \int_0^T g(t)\eta(t)\,dt + R_T U_T + H'(\widehat{\beta}_T) L_T = 0, \tag{24}$$

$$l_{iT} h(\widehat{\beta}_T) = 0, \qquad l_{iT} \ge 0, \qquad i \in \{1, \dots, r\}, \tag{25}$$

where $H(\beta)$ is the $r \times n$ matrix whose row i is $\nabla h_i'(\beta)$, $i \in \{1, \dots, r\}$, and where

$$L_T = (l_{1T}, \dots, lr_T) \quad \text{and} \quad R_T = m!\, J_T \sigma_{T,m}^{-1} J_T.$$

Put

$$Q_T = B^{m/2}(T) T^{-1/2} d_T^{-1} \int_0^T g(t)\eta(t)\,dt. \tag{26}$$

Also let

$$L_T = (L_{1T}', L_{2T}')', \qquad H(\beta) = (H_1'(\beta), H_2'(\beta))',$$

where L_{1T} and L_{2T} are the vectors whose components are

$$l_{iT}, \quad i \in \{1, \dots, q\}, \quad \text{and} \quad l_{iT}, \quad i \in \{q+1, \dots, r\},$$

respectively; $H_1(\beta)$ is the $q \times n$ matrix whose row i is $\nabla h_i'(\beta)$, $i \in \{1, \dots, q\}$; and $H_2(\beta)$ is the $(r-q) \times n$ matrix whose row i is $\nabla h_i'(\beta)$, $i \in \{q+1, \dots, r\}$.

Taking into account this notation we obtain from (24) and (26) that

$$-Q_T + R_T U_T + H_1'(\widehat{\beta}_T) L_{1T} + H_2'(\widehat{\beta}_T) L_{2T} = 0. \tag{27}$$

According to (2),

$$h_i(\widehat{\beta}_T) + z_{iT} = 0, \qquad z_{iT} \ge 0, \qquad i \in \{1, \dots, r\},$$

whence

$$\lambda_{iT} z_{iT} = 0, \qquad i \in \{1, \dots, r\},$$

in view of the equalities in (15). Thus

$$l_{iT} w_{iT} = 0, \qquad i \in \{1, \dots, r\}, \tag{28}$$

where $w_{iT} = \Re_T z_{iT}$. Note that

$$l_{iT} \ge 0, \qquad i \in \{1, \dots, r\}, \tag{29}$$

according to the inequalities in (15), since $l_{iT} = \Re_T^{-1} \lambda_{iT}$ and $\lambda_{iT} \ge 0$, $i \in \{1, \dots, r\}$.

By definition $w_{iT} \ge 0$, $i \in \{1, \dots, r\}$. Expanding $h_i(\beta)$ in the Taylor series in a neighborhood of $\beta = \beta^0$ we get

$$\psi_i(\widehat{\beta}_T)(\widehat{\beta}_T - \beta^0) + z_{iT} = 0, \qquad i \in \{1, \dots, q\}, \tag{30}$$

where

$$\psi_i(\widehat{\beta}_T) = \nabla h_i'(\beta^0) + \frac{1}{2}(\widehat{\beta}_T - \beta^0)h_{i2}(\beta^0 + \theta_1 \Delta\widehat{\beta}_T). \tag{31}$$

Here $\theta_1 \in (0,1)$, $\Delta\widehat{\beta}_T = \widehat{\beta}_T - \beta^0$, $h_{i2}(\beta)$ is the Hessian of size $n \times n$, and $\frac{\partial^2 h_i(\beta)}{\partial\beta_j \partial\beta_k}$, $j,k = 1,\dots,n$, is its entry j,k.

Multiplying both sides of equality (30) by $\Re_T^{-1}$ we rewrite this expression in the matrix form

$$\Psi_1(\widehat{\beta}_T)U_T + W_{1T} = 0, \tag{32}$$

where $\Psi_1(\widehat{\beta}_T)$ is the matrix of size $r \times q$ whose row i is $\psi_i(\widehat{\beta}_T)$, $i \in \{1,\dots,q\}$, and W_{1T} is the q-dimensional vector with components w_{iT}, $i \in \{1,\dots,q\}$.

The above results imply the following three lemmas.

Lemma 4.4. *Let functions $g_i(t)$, $t \in [0;T]$, be bounded and let estimators $\widehat{\beta}_T$ be consistent. Then*

$$p \lim_{T\to\infty} L_{2T} = 0. \tag{33}$$

Proof. Relation (33) follows from (19) and (25), since $h_i(\beta^0) < 0$, $i \in \{q+1,\dots,r\}$, and $0 = p\lim_{T\to\infty} l_{iT} h_i(\widehat{\beta}_T) = p\lim_{T\to\infty} l_{iT} h_i(\beta^0)$. □

Lemma 4.5. *Let Assumptions* 1–6 *and* 9 *hold. Then given a number $\delta > 0$ there are numbers $\varepsilon > 0$ and $T_0 > 0$ such that*

$$\mathsf{P}\{\|L_{1T}\| \ge \varepsilon\} < \delta, \qquad T > T_0. \tag{34}$$

Proof. Equality (27) implies that $-R_T Q_T + U_T + R_T^{-1} H_1'(\widehat{\beta}_T)L_{1T} + R_T^{-1}H_2'(\widehat{\beta}_T)L_{2T} = 0$, whence

$$U_T = R_T Q_T - R_T^{-1} H_1'(\widehat{\beta}_T)L_{1T} - R_T^{-1} H_2'(\widehat{\beta}_T)L_{2T}. \tag{35}$$

Let

$$\begin{gathered} B_T = \Psi_1(\widehat{\beta}_T)R_T^{-1}H_1'(\widehat{\beta}_T), \\ q_T = \Psi_1(\widehat{\beta}_T)R_T^{-1}Q_T, \qquad e_T = \Psi_1(\widehat{\beta}_T)R_T^{-1}H_2'(\widehat{\beta}_T)L_{2T}. \end{gathered} \tag{36}$$

We have

$$L_{1T}' B_T L_{1T} = L_{1T}' A_T L_{1T}$$

where A_T is a symmetric matrix whose entries are $a_{ij}^T = \frac{1}{2}(b_{ij}^T + b_{ji}^T)$, $i,j = 1,\dots,n$, and where b_{ij}^T is an entry of B_T.

Equality (35) implies

$$\begin{aligned} &L_{1T}'\Psi_1(\widehat{\beta}_T)U_T \\ &\quad = L_{1T}\Psi_1(\widehat{\beta}_T)R_T Q_T - R_T^{-1}L_{1T}'\Psi_1(\widehat{\beta}_T)H_1'(\widehat{\beta}_T)L_{1T} - L_{1T}'\Psi_1(\widehat{\beta}_T)R_T^{-1}H_2'(\widehat{\beta}_T)L_{2T} \\ &\quad = 0. \end{aligned}$$

Using (36) and the representation of the matrices A_T in terms of the matrix B_T we have

$$L_{1T}'\Psi_1(\widehat{\beta}_T)U_T = L_{1T}'q_T - L_{1T}'B_T L_{1T} - L_{1T}'e_T,$$

whence $L_{1T}'\Psi_1(\widehat{\beta}_T)U_T = L_{1T}'q_T - L_{1T}'A_T L_{1T} - L_{1T}'e_T$. Then

$$-L_{1T}'\Psi_1(\widehat{\beta}_T)U_T = L_{1T}'A_T L_{1T} + L_{1T}'(e_T - q_T).$$

Now we derive from (28) and (32) that $L_{1T}'\Psi_1(\widehat{\beta}_T)U_T = 0$, that is,

$$L_{1T}'A_T L_{1T} + L_{1T}'(e_T - q_T) = 0. \tag{37}$$

Since A_T is a symmetric matrix, we have

$$C_T' A_T C_T = N_T, \tag{38}$$

where C is an orthogonal matrix (that is, $C_T' C_T = J$), $N_T = \operatorname{diag}(\nu_{1T}, \dots, \nu_{nT})$, and ν_{iT} is the ith eigenvalue of the matrix A_T. Since R_T and R_0 are nonsingular matrices, Assumption 4 implies that

$$p \lim_{T\to\infty} R_T^{-1} = R_0^{-1}. \tag{39}$$

Since $\widehat{\beta}_T$ is consistent, Assumptions 6, 1) and relations (20) and (31) imply that

$$p \lim_{T\to\infty} \Psi_1(\widehat{\beta}_T) = p \lim_{T\to\infty} H_1(\widehat{\beta}_T) = H_1(\beta^0). \tag{40}$$

Then we derive from (36) and (39) that

$$p \lim_{T\to\infty} B_T = p \lim_{T\to\infty} A_T = H_1(\beta^0) R_0^{-1} H_1'(\beta^0) = A.$$

This equality and Assumption 6, 4) prove that A is a positive definite matrix.

Now we establish from Lemma 4.1 that

$$p \lim_{T\to\infty} N_T = N, \qquad p \lim_{T\to\infty} C_T = C, \tag{41}$$

where $N = \operatorname{diag}(\nu_1, \dots, \nu_N)$, ν_i is the ith eigenvalue of the matrix A, and C is an orthogonal matrix such that $C'AC = N$.

Let

$$\tilde{\nu}_{iT} = \begin{cases} \nu_{iT} & \text{if } \nu_{iT} > 0, \\ 1 & \text{otherwise.} \end{cases}$$

Using the first relation in (41) one can show that

$$p \lim_{T\to\infty} \tilde{N}_T = N, \tag{42}$$

where $\tilde{N}_T = \operatorname{diag}(\tilde{\nu}_{1T}, \dots, \tilde{\nu}_{NT})$. Assertion 3) of Lemma 4.1 implies that for an arbitrary $\delta > 0$ there exists $T_1 > 0$ such that

$$\mathsf{P}\{N = \tilde{N}_T\} = \mathsf{P}\{\nu_{iT} > 0, i = 1, \dots, n\} > 1 - \frac{\delta}{3}, \qquad T > T_1.$$

Put

$$Y_T = \tilde{N}_T^{1/2} C_T^{-1} L_{1T}. \tag{43}$$

From (38) and (43) we obtain $A_T = C_T^{-1} N_T C_T^{-1}$ and $L_{1T} = C_T \tilde{N}_T^{-1/2} Y_T$, respectively. Then (37) implies that

$$\begin{aligned} &\mathsf{P}\left\{C_T' \tilde{N}'^{-1/2} Y_T' C_T^{-1} N_T C_T^{-1} C_T \tilde{N}_T^{-1/2} Y_T + C_T' \tilde{N}'^{-1/2} Y_T' (e_T - q_T) = 0\right\} \\ &\quad = \mathsf{P}\{N_T = \tilde{N}_T\} > 1 - \frac{\delta}{3}, \qquad T > T_1, \end{aligned}$$

$$\mathsf{P}\{Y_T' Y + 2Y_T' K_T = 0\} > 1 - \frac{\delta}{3}, \qquad T > T_1, \tag{44}$$

where

$$K_T = \frac{1}{2} \tilde{N}'^{-1/2} C_T' (e_T - q_T). \tag{45}$$

According to the limit theorem proved in Ivanov and Leonenko [8] (Theorem 1 therein) for long-memory processes,

$$Q_T \xrightarrow{D} Q, \qquad T \to \infty, \tag{46}$$

if $m = 1$, where Q_T and σ_1 are defined in (26) and (5), respectively, and Q is a zero mean Gaussian random vector with covariance matrix $C_1^2 J_0 \sigma_1^{-1} J_0$. Here the symbol "$\xrightarrow{D}$" stands for the convergence in distribution.

Using (36), (20), (33), (39), (40), and (46) we have $q_T \xrightarrow{D} q$ as $T \to \infty$ and

$$\lim_{T\to\infty} e_T = 0.$$

Relations (45), (41), and (42) together with the latter two equalities imply that the limit distribution of K_T coincides with the distribution of the random variable

$$K = -\frac{1}{2} N^{-1/2} C' q.$$

Put $Y_T = \tilde{Y}_T - K_T$. Then

$$\mathsf{P}\big\{\|\tilde{Y}_T\| = \|K_T\|\big\} \geq 1 - \frac{\delta}{3}, \qquad T > T_1, \tag{47}$$

by (44). Since the limit distribution of K_T exists, for an arbitrary $\delta > 0$ there are numbers $T_2 > 0$ and $\varepsilon_1 > 0$ such that

$$\begin{aligned}\frac{\delta}{6} &> \mathsf{P}\left\{\|K_T\| \geq \frac{\varepsilon_1}{2}\right\} \geq \mathsf{P}\left\{\|\tilde{Y}_T\| \geq \frac{\varepsilon_1}{2}, \|K_T\| \geq \frac{\varepsilon_1}{2}\right\} \\ &\geq \mathsf{P}\left\{\|\tilde{Y}_T\| \geq \frac{\varepsilon_1}{2}\right\} - \mathsf{P}\{\|\tilde{Y}_T\| \neq \|K_T\|\}, \qquad T > T_2.\end{aligned}$$

The latter relation and (47) yield

$$\mathsf{P}\left\{\|\tilde{Y}_T\| \geq \frac{\varepsilon_1}{2}\right\} < \frac{\delta}{2}, \qquad T > T_3 = \max(T_1, T_2),$$
$$\|Y_T\| \leq \|\tilde{Y}_T\| + \|K_T\|.$$

Then for an arbitrary $\delta > 0$ there exists a number $\varepsilon_1 > 0$ such that

$$\mathsf{P}\{\|Y_T\| < \varepsilon_1\} \geq \mathsf{P}\left\{\|\tilde{Y}_T\| < \frac{\varepsilon_1}{2}\right\} - \mathsf{P}\left\{\|K_T\| \geq \frac{\varepsilon_1}{2}\right\} > 1 - \frac{2\delta}{3}, \qquad T > T_3. \tag{48}$$

It follows from (43) that

$$\|L_{1T}\| \leq \|C_{1T}\| \cdot \|Y_T\|, \tag{49}$$

where $C_{1T} = C_T N_T^{-1/2}$. Relations (41) and (42) imply that

$$p \lim_{T\to\infty} C_{1T} = C_1 = C N^{-1/2}.$$

Thus given $\delta > 0$ there exists a number $\varepsilon > 0$ such that

$$\mathsf{P}\{\|C_{1T}\|\varepsilon_1 < \varepsilon\} \geq 1 - \frac{\delta}{3}, \qquad T > T_4. \tag{50}$$

Note that inequalities (48) and (50) hold simultaneously for $T_0 = \max(T_3, T_4)$. Consider the inequality

$$\mathsf{P}\{a \geq b\} \geq \mathsf{P}\{a \geq \zeta\} - \mathsf{P}\{b \geq \zeta\} \tag{51}$$
$$\text{for} \qquad a = \|C_{1T}\| \cdot \|Y_T\|, \qquad b = \|C_{1T}\|\varepsilon_1, \qquad \zeta = \varepsilon.$$

Then we get from (48) and (50) that

$$\frac{2}{3}\delta > \mathsf{P}\{\|C_{1T}\| \cdot \|Y_T\| \geq \|C_{1T}\|\varepsilon_1\} \geq \mathsf{P}\{\|C_{1T}\| \cdot \|Y_T\| \geq \varepsilon\} - \frac{\delta}{3}, \qquad T > T_0.$$

This together with (49) implies (34). $\square$

Lemma 4.6. *Let Assumptions* 1–6 *and* 9 *hold. Then given* $\delta > 0$ *there are numbers* $\varepsilon > 0$ *and* $T_0 > 0$ *such that*

$$\mathsf{P}\{\|U_T\| \geq \varepsilon\} < \delta, \qquad T > T_0. \tag{52}$$

Proof. Taking into account (35) we have

$$\begin{aligned} \mathsf{P}\{\|U_T\| \geq \varepsilon_3\} \leq \mathsf{P}&\left\{\left\|R_T^{-1} H_1'(\widehat{\beta}_T)\right\| \cdot \|L_{1T}\| \geq \frac{\varepsilon_3}{3}\right\} \\ &+ \mathsf{P}\left\{\left\|R_T^{-1} H_2'(\widehat{\beta}_T) L_{2T}\right\| \geq \frac{\varepsilon_3}{3}\right\} \\ &+ \mathsf{P}\left\{\left\|R_T^{-1} Q_T\right\| \geq \frac{\varepsilon_3}{3}\right\} \end{aligned} \tag{53}$$

for some number $\varepsilon_3 > 0$ and for all $T > 0$. We estimate the terms on the right hand side of (53). Since

$$p \lim_{T\to\infty} R_T^{-1} = R_0^{-1},$$

we have for given numbers $\delta > 0$ and $\varepsilon_1 > 0$ that

$$\begin{aligned} \mathsf{P}\left\{\left\|R_T^{-1} H_1'(\widehat{\beta}_T)\right\| - \left\|R_0^{-1} H_1'(\beta^0)\right\| < \varepsilon_1\right\} &\geq \mathsf{P}\left\{\left\|R_T^{-1} H_1'(\widehat{\beta}_T) - R_0^{-1} H_1'(\beta^0)\right\| < \varepsilon_1\right\} \\ &\geq 1 - \delta, \qquad T > T_1, \end{aligned}$$

by (40). Put

$$\varepsilon_2 = \varepsilon = \left\|R_0^{-1} H_1'(\beta^0)\right\| + \varepsilon\varepsilon_1$$

in the latter inequality where $\varepsilon > 0$ is arbitrary. Then

$$\mathsf{P}\left\{\varepsilon\left\|R_T^{-1} H_1'(\widehat{\beta}_T)\right\| < \varepsilon_2\right\} \geq 1 - \delta, \qquad T > T_1. \tag{54}$$

Multiply both sides of the inequality under the sign of probability by $\|R_T^{-1} H_1'(\widehat{\beta}_T)\|$. Further we put

$$a = \left\|R_T^{-1} H_1'(\widehat{\beta}_T)\right\| \cdot \|L_{1T}\|, \qquad b = \left\|R_T^{-1} H_1'(\widehat{\beta}_T)\right\|\varepsilon, \qquad \zeta = \varepsilon_2$$

in inequality (51).

It follows from (34) and (51) that

$$\delta > \mathsf{P}\left\{\|R_T^{-1} H_1'(\widehat{\beta}_T)\| \cdot \|L_{1T}\| \geq \varepsilon_2\right\} - \mathsf{P}\left\{\left\|R_T^{-1} H_1'(\widehat{\beta}_T)\right\|\varepsilon \geq \varepsilon_2\right\}, \qquad T > T_2.$$

We obtain from (54) and the latter inequality that

$$\mathsf{P}\left\{\left\|R_T^{-1} H_1'(\widehat{\beta}_T)\right\| \cdot \|L_{1T}\| \geq \varepsilon_2\right\} < \frac{\delta_1}{3}, \qquad T > \max(T_1, T_2). \tag{55}$$

Now we derive from relations (20) and (33) that

$$p \lim_{T\to\infty} R_T^{-1} H_2'(\widehat{\beta}_T) L_{2T} = 0.$$

Thus given $\varepsilon_2 > 0$ and $\delta_1 > 0$ there exists a number $T_3 > 0$ such that

$$\mathsf{P}\left\{\left\|R_T^{-1} H_2'(\widehat{\beta}_T) L_{2T}\right\| \geq \varepsilon_2\right\} < \frac{\delta_1}{3}, \qquad T > T_3. \tag{56}$$

Using (46) and (39) we prove that, for a given $\delta_1 > 0$, there are two numbers $\varepsilon_4 > 0$ and $T_4 > 0$ such that

$$\mathsf{P}\left\{\left\|R_T^{-1} Q_T\right\| \geq \varepsilon_4\right\} < \frac{\delta_1}{3}.$$

Putting $\varepsilon_3 = 3\varepsilon_4 = 3\varepsilon_2$ on the right hand side of (53) we obtain inequality (52) for $\varepsilon = \varepsilon_3$ and $T_0 = \max(T_1, T_2, T_3, T_4)$ from (53), (55), (56), and the latter inequality where $\delta = \delta_1$. $\square$

5. Proof of the main result

Now we evaluate the limit of the sequence of random variables U_T. Consider the following convex programming problem:

$$\left\{\begin{array}{l}\varphi_T(X) = \dfrac{1}{2}X' R_T(\omega)X - Q'_T(\omega)X \longrightarrow \min,\\ f_i(X) = \nabla h'_i(\beta^0)X + s_i(X) \le 0, \quad i \in \{1,\dots,q\},\\ f_i(X) = \Re_T^{-1} h_i(\beta^0) + \nabla h'_i(\beta^0)X + s_i(X) \le 0, \quad i \in \{q+1,\dots,r\},\end{array}\right. \tag{57}$$

where $X \in \mathbb{R}^n$, $Q_T(\omega) = Q_T$, $R_T(\omega) = R_T$, $\omega \in \Omega$, and Ω is the sampling space. The other symbols in (57) are

$$f_i(X) = \Re_T^{-1} h_i(\Re_T(T)X + \beta^0), \qquad i \in \{1,\dots,r\}, \tag{58}$$

$$s_i(X) = \frac{1}{2}\Re_T X' h_{i2}\big(\beta^0 + \theta_1 \Re_T^{-1} X\big)X, \qquad i \in \{1,\dots,r\}, \tag{59}$$

where the functions h_{i2} and θ_1 are defined as in (31).

A vector $U_T^*(\omega)$ considered at a fixed ω is understood as a solution of the problem (57). This means that $U_T^*(\omega)$, $\omega \in \Omega$, is a random vector.

Since $R_T(\omega)$ is a symmetric matrix, the necessary conditions for the extremum in the problem (57) are given by

$$\begin{gathered} R_T U_T^*(\omega) - R_T(\omega) + \sum_{i=1}^m \nabla f_i(U_T^*(\omega)) l_{iT}^* = 0,\\ l_{iT}^* f_i(U_T^*(\omega)) = 0, \qquad l_{iT}^* \ge 0, \qquad i \in \{1,\dots,r\}. \end{gathered} \tag{60}$$

In turn necessary conditions for the extremum in the problem (14), (15) given (25) and (27) can be rewritten as follows:

$$\begin{gathered} R_T(\omega)U_T(\omega) + \sum_{i=1}^m \nabla h_i(\widehat{\beta}_T(\omega)) l_{iT} - R_T(\omega) = 0,\\ l_{iT}(\omega) h_i(\widehat{\beta}_T(\omega)) = 0, \qquad l_{iT} \ge 0, \qquad i \in \{1,\dots,r\}, \end{gathered} \tag{61}$$

where $l_{iT}(\omega) = l_{iT}$, $U_T(\omega) + U_T$, and $\beta_T(\omega) = \beta_T$.

Put $X = U_T(\omega)$. We have

$$f_i(U_T(\omega)) = \Re_T^{-1} h_i(\widehat{\beta}_T(\omega)), \qquad \nabla f_i(U_T(\omega)) = \nabla h_i(\widehat{\beta}_T(\omega))$$

according to (58). This together with (60) and (61) implies that the vectors $U_T(\omega)$ and $L_T(\omega)$ form a solution of the system of equations (60); that is, $U_T(\omega)$ satisfies the necessary conditions for the extremum in the problem (57).

Now we are ready to prove the main result.

Proof of the main result. Consider the following quadratic programming problem:

$$\left\{\begin{array}{l}\tilde{\varphi}_T(X) = \dfrac{1}{2}X' R_0 X - Q'_T(\omega)X \to \min,\\ \nabla h'_i(\beta^0)X \le 0, \qquad i \in \{1,\dots,q\}.\end{array}\right. \tag{62}$$

Denote by $\tilde{U}_T(\omega)$ a solution of problem (62). According to Lemma 4.2, $\tilde{U}_T$ is a continuous function of $Q_T(\omega)$, namely

$$\tilde{U}_T(\omega) = \kappa \circ (\tilde{Q}_T(\omega)).$$

Then (45) implies that

$$\kappa(Q_T(\omega)) \xrightarrow{D} \kappa(Q(\omega)).$$

On the other hand, $U(\omega) = \kappa(Q(\omega))$ by (61). Thus

$$\tilde{U}_T(\omega) \xrightarrow{D} U(\omega), \qquad T \to \infty. \tag{63}$$

Put

$$o_T = \{X \colon f_i(X) \le 0, i \in \{1, \dots, r\}\},$$
$$o = \{X \colon \nabla h_i'(\beta^0) X \le 0, i \in \{1, \dots, q\}\}.$$

Since the functions $h_i(\beta)$, $i \in \{1, \dots, r\}$, are convex and $s_i(X) \ge 0$, $X \in \mathbb{R}^n$, the set O_T is convex. Therefore (57) yields

$$h_i'(\beta^0) U_T(\omega) \le 0, \qquad i \in \{1, \dots, q\},$$

since $f_i(U_T) \le 0$, $i \in \{1, \dots, r\}$ (see equality (58)). Hence $U_T(\omega) \in O$. Since the matrix R_0 is positive definite, the function $\tilde{\varphi}_T(X)$ is strongly convex in X for every fixed ω. As $U_T(\omega) \in O$ we have

$$\|U_T(\omega) - \tilde{U}_T(\omega)\|^2 \le \frac{2}{\mu} \left[\tilde{\varphi}_T(U_T(\omega)) - \tilde{\varphi}_T(\tilde{U}_T(\omega))\right]$$

for some constant $\mu > 0$ (see [2]). Thus

$$\begin{aligned} \mathsf{P}\left\{\|U_T - \tilde{U}_T\|^2 < \varepsilon^2\right\} &\ge \mathsf{P}\left\{\frac{2}{\mu}[\tilde{\varphi}_T(U_T) - \tilde{\varphi}_T(\tilde{U}_T)] < \varepsilon^2\right\} \\ &\ge \mathsf{P}\left\{|\tilde{\varphi}_T(U_T) - \varphi_T(U_T)| < \frac{\varepsilon_1}{2}\right\} \\ &\quad + \mathsf{P}\left\{|\varphi_T(U_T) - \varphi_T(\tilde{U}_T)| < \frac{\varepsilon_1}{2}\right\} - 1 \end{aligned} \tag{64}$$

for an arbitrary $\varepsilon > 0$ where $\varepsilon_1 = \mu\varepsilon^2/2$.

We estimate the probabilities on the right hand side of inequality (64). Using Lemma 4.6 we obtain from (57) and (62) that

$$\begin{aligned} \mathsf{P}\{|\tilde{\varphi}_T(U_T) - \varphi_T(U_T)| < \varepsilon_2\} &= \mathsf{P}\left\{\left\|\frac{1}{2}U_T'(R_0 - R_T)U_T\right\| < \varepsilon_2\right\} \\ &\ge 1 - \mathsf{P}\left\{\|U_T\| \ge \sqrt{b}\right\} - \mathsf{P}\left\{\|R_0 - R_T\| \ge \frac{2\varepsilon_2}{b}\right\} \\ &> 1 - \delta \end{aligned} \tag{65}$$

for $T > T_0$ and for arbitrary $\varepsilon_2 > 0$ and $\delta > 0$ where $b > 0$ is some number.

Analogously, it follows from (63) that

$$\mathsf{P}\left\{\left|\tilde{\varphi}_T(\tilde{U}_T) - \varphi_T(\tilde{U}_T)\right| < \varepsilon_2\right\} = \mathsf{P}\left\{\frac{1}{2}\tilde{U}'(R_0 - R_T)\tilde{U}_T < \varepsilon_2\right\} > 1 - \delta, \qquad T > T_0. \tag{66}$$

Now we estimate the second term in (64). We prove the inequality

$$\begin{aligned} &\mathsf{P}\left\{|\varphi_T(U_T) - \tilde{\varphi}_T(\tilde{U}_T)| < \frac{\varepsilon_1}{2}\right\} \\ &\quad \ge \mathsf{P}\left\{|\varphi_T(U_T) - \varphi_T(\tilde{U}_T)| \le 0\right\} + \mathsf{P}\left\{|\varphi_T(\tilde{U}_T) - \tilde{\varphi}_T(\tilde{U}_T)| < \frac{\varepsilon_1}{2}\right\} - 1. \end{aligned} \tag{67}$$

Inequality (67) is equivalent to the following one:

$$\begin{aligned} &\mathsf{P}\left\{|\varphi_T(U_T) - \tilde{\varphi}_T(\tilde{U}_T)| \ge \frac{\varepsilon_1}{2}\right\} \\ &\quad \le \mathsf{P}\left\{|\varphi_T(U_T) - \varphi_T(\tilde{U}_T)| > 0\right\} + \mathsf{P}\left\{|\varphi_T(\tilde{U}_T) - \tilde{\varphi}_T(\tilde{U}_T)| \ge \frac{\varepsilon_1}{2}\right\}. \end{aligned} \tag{68}$$

The latter inequality is obvious.

Now consider the first term on the right hand side of (67). If R_T is a positive definite matrix, then problem (57) has a unique solution, since O_T is a convex set and Assumption 6 holds. The above results together with (60) and (61) imply that $U_T^* = U_T$ if $\Lambda_{iT} > 0$, $i \in \{1, \ldots, n\}$. This means that

$$\begin{aligned}\mathsf{P}\left\{|\varphi_T(U_T) - \varphi_T(\tilde U_T)| \le 0\right\} &\ge \mathsf{P}\left\{\tilde U_T \in O_T, \Lambda_{iT} > 0, i \in \{1, \ldots, n\}\right\} \\ &= \mathsf{P}\left\{f_j(\tilde U_T) \le 0, j \in \{1, \ldots, r\}, \Lambda_{iT} > 0, i \in \{1, \ldots, n\}\right\},\end{aligned}$$

where $\Lambda_{iT} > 0$, $i \in \{1, \ldots, n\}$, is the i-th eigenvalue of R_T, whence we derive

$$\begin{aligned}&\mathsf{P}\left\{|\varphi_T(U_T) - \varphi_T(\tilde U_T)| \le 0\right\} \\ &\quad\ge \mathsf{P}\left\{f_j(\tilde U_T) \le 0, j \in \{1, \ldots, r\}, \Lambda_{iT} > 0, i \in \{1, \ldots, n\}\right\} \\ &\quad\ge \mathsf{P}\{\Lambda_{iT} > 0, i \in \{1, \ldots, n\}\} + \mathsf{P}\left\{\bigcap_1^r f_i(\tilde U_T) \le 0\right\} - 1 \\ &\quad\ge \mathsf{P}\{\Lambda_{iT} > 0, i \in \{1, \ldots, n\}\} + \mathsf{P}\left\{\bigcap_1^{r-1} f_i(\tilde U_T) \le 0\right\} + \mathsf{P}\left\{f_r(\tilde U_T) \le 0\right\} - 2 \\ &\quad\ge \cdots \ge \mathsf{P}\{\Lambda_{iT} > 0, i \in \{1, \ldots, n\}\} + \sum_1^r \mathsf{P}\left\{f_i(\tilde U_T) \le 0\right\} - r.\end{aligned} \tag{69}$$

In the latter chain of inequalities we use r times an obvious bound

$$\mathsf{P}(A \cap B) \ge \mathsf{P}(A) + \mathsf{P}(B) - 1.$$

Since the function h_{i2} is continuous and bounded at the point $\beta = \beta^0$, we obtain from (59) and (63) that

$$p \lim_{T\to\infty} s_i(\tilde U_T) = 0. \tag{70}$$

Using (63) and (70) together with the inequalities

$$\nabla h_i'(\beta^0) U \ge 0, \qquad i \in \{1, \ldots, q\},$$

and

$$h_i(\beta^0) < 0, \qquad i \in \{q+1, \ldots, r\},$$

one can check that, for an arbitrary $\varsigma_i > 0$, there exists a number $T_{3i} > 0$ such that

$$\mathsf{P}\left\{f_i(\tilde U_T) \le 0\right\} \ge 1 - \varsigma_i, \qquad T > T_{3i}, \qquad i \in \{1, \ldots, r\}. \tag{71}$$

Since R_T converges to R_0, we apply statement 3) of Lemma 4.1 to R_T with an arbitrary $\varsigma > 0$. As a result we get

$$\mathsf{P}\{\Lambda_{iT} > 0, i \in \{1, \ldots, n\}\} > 1 - \frac{\varsigma}{2}, \qquad T > T_5,$$

whence

$$\mathsf{P}\left\{|\varphi_T(U_T) - \varphi_T(\tilde U_T)| \le 0\right\} > 1 - \varsigma, \qquad T > T_6 = \max(T_4, T_5), \tag{72}$$

by (71) and (70) for

$$\varsigma = 2\sum_{i=1}^r \varsigma_i, \qquad T_4 = \max_{i \in \{1, \ldots, r\}} T_{3i}.$$

As an application of (67) we have

$$\mathsf{P}\left\{\varphi_T(U_T) - \tilde\varphi_T(\tilde U_T) < \varepsilon_2\right\} > 1 - \delta - \varsigma, \qquad T > \max(T_2, T_6), \tag{73}$$

by (72) and (68) for $\varepsilon_2 = \varepsilon_1/2$. Substituting estimates (73) and (66) to the right hand side of (64) for $\varepsilon_2 = \varepsilon_1/2$ we obtain $\mathsf{P}\{\|U_T - \tilde{U}_T\|^2 < \varepsilon^2\} > 1 - 2\delta - \varsigma$, $T > \max(T_1, T_2, T_6)$. Furthermore,

$$p \lim_{T\to\infty} \|U_T - \tilde{U}_T\| = 0. \tag{74}$$

Theorem 3.1 follows from (63). □

The proof for the case of $m \geq 2$ is similar. The difference is that the noncentral limit theorem (Theorem 3 of [8]) is substituted for the central limit theorem (Theorem 1 of [8]). The assumptions of Theorem 3 of [8] follow from the assumptions of our Theorem 3.1, so Theorem 3 of [3] can be used in our proof, indeed.

6. Concluding remarks

Note that the random vector (9) is Gaussian for $m = 1$ and non-Gaussian for $m \geq 2$. The solution of the quadratic programming problem (11) is non-Gaussian even in the case of $m = 1$ in contrast to the results of [8]. It is proved that the least squares estimators

$$\widehat{\beta}_T = V(T)^{-1} \int_0^T g(t) y(t)\, dt$$

of the regression coefficients β for the model without constraints (2) are asymptotically Gaussian for $m = 1$ and non-Gaussian for $m \geq 2$ if one normalizes and centers appropriately these estimators where

$$V(T) = \int_0^T g(t) g'(t)\, dt.$$

Bibliography

1. P. S. Knopov, *On a nonstationary model of M-estimators with discrete time*, Teor. Imovir. Mat. Statist. **57** (1997), 60–66; English transl. in Theory Probab. Math. Statist. **57** (1998), 61–67. MR1806885 (2003c:62061)
2. A. S. Korkhin, *Some properties of estimates for regression parameters with a priori inequality constraints*, Kibernetika **6** (1985), 106–115. (Russian) MR825195 (87c:62119)
3. E. M. Moldavs'ka, *Methods for the Identification of Parameters of Stochastic Systems with Weak and Strong Dependence*, Candidate Dissertation, Kyiv University, Kyiv, 2003. (Ukrainian)
4. R. Dahlhaus, *Efficient location and regression estimation for long-range dependence regression model*, Ann. Statist. **23** (1995), 1029–1047. MR1345213 (96h:62035)
5. R. L. Dobrushin and P. Major, *Non-central limit theorems for non-linear functionals of Gaussian fields*, Z. Wahrsch. Verw. Gebiete **50** (1979), no. 1, 27–52. MR550122 (81i:60019)
6. J. Dupacova and R. Wets, *Asymptotic behavior of statistical estimators and optimal solutions of stochastic optimization problems*, Ann. Statist. **16** (1988), no 4, 1517–1549. MR964937 (89i:62044)
7. A. V. Ivanov and N. N. Leonenko, *Asymptotic inferences for a nonlinear regression with strong dependence*, Teor. Imovir. Mat. Statist. **63** (2000), 61–79; English transl. in Theory Probab. Math. Statist. **63** (2001), 65–85. MR1870775 (2003a:62074)
8. A. V. Ivanov and N. N. Leonenko, *Asymptotic behavior of M-estimators in continuous-time non-linear regression with long-range dependent errors*, Random Oper. Stochastic Equations **10** (2002), no. 3, 201–222. MR1923424 (2003k:62080)
9. A. V. Ivanov and N. N. Leonenko, *Asymptotic theory of nonlinear regression with long-range dependence*, Math. Methods Statist. **13** (2004), no. 2, 153–178. MR2090470 (2005h:62175)
10. H. R. Künsch, J. Beran, and F. R. Hampel, *Constraints under long-range correlations*, Ann. Statist. **21** (1993), 943–964. MR1232527 (94f:62134)
11. N. N. Leonenko, *Limit Theorems for Random Fields with Singular Spectrum*, Kluwer Academic Publishers, Dordrecht, 1999. MR1687092 (2000k:60102)

12. N. N. Leonenko and M. Benšic, *On estimation of regression coefficients of long-memory random fields observed on the arrays*, Random Oper. Stochastic Equations **6** (1998), no. 3, 237–252. MR1610205 (99c:62275)
13. N. N. Leonenko and E. M. Moldavskaya, *Non-Gaussian scenarios in long-memory regression models with non-linear constraints*, Dopov. Nats. Akad. Nauk Ukr. Mat. Prirodozn. Tekh. Nauki (2002), no. 2, 44–46. MR1916769
14. E. M. Moldavskaya, *Theorem useful in proving of estimates consistency in long-memory regression models*, Visn. Kyiv. Univ. Ser. Fiz.-Mat. Nauk (2002), no. 1, 58–65. MR1934119
15. N. K. Nagaraj and W. A. Fuller, *Estimation of the parameters of linear time series models subject to nonlinear restrictions*, Ann. Statist. **19** (1991), no. 3, 1143–1154. MR1126318 (92h:62153)
16. M. S. Taqqu, *Convergence of integrated processes of arbitrary Hermite rank*, Z. Wahrsch. Verw. Gebiete **50** (1979), 53–83. MR550123 (81i:60020)
17. J. Wang, *Asymptotics of least-squares estimators for constrained nonlinear regression*, Ann. Statist. **24** (1996), no. 3, 1316–1326. MR1401852 (97i:62072)
18. Y. Yajima, *Asymptotic properties of the LSE in a regression model with long-memory stationary errors*, Ann. Statist. **19** (1991), 158–177. MR1091844 (92d:62124)

Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel
E-mail address: elinam@bgu.ac.il

Received 23/OCT/2004

Translated by O. I. KLESOV