Siberian Mathematical Journal, Vol.42, No. 2, 2001.
Translated from:
*Sibirskiĭ Matematicheskiĭ Zhurnal*, Vol. 42, No. 2, pp. 372–388.

# ASYMPTOTICALLY NORMAL ESTIMATION OF A MULTIDIMENSIONAL PARAMETER IN THE LINEAR-FRACTIONAL REGRESSION PROBLEM

YU. YU. LINKE AND A. I. SAKHANENKO

## 1. Statement of the Problem

Suppose that in a series of $N$ trials, $N \to \infty$, we observe a sequence of random variables $Z_1, \dots, Z_N$ and assume that they are representable as

$$Z_i = g_i(\theta) + \xi_i \equiv \frac{\alpha_i(\theta)}{\beta_i(\theta)} + \xi_i, \quad i = 1, \dots, N, \tag{1.1}$$

where

$$\alpha_i(\theta) \equiv a_{0i} + \sum_{j=1}^{m} a_{ji}\theta_j, \quad \beta_i(\theta) \equiv 1 + \sum_{j=1}^{m} b_{ji}\theta_j \tag{1.2}$$

are linear functions depending on an unknown $m$-dimensional parameter $\theta$ with coordinates $\theta_1, \dots, \theta_m$, while

$$b_{ji} \geq 0,\ a_{0i},\ a_{ji}, \quad i = 1, \dots, N,\ j = 1, \dots, m, \tag{1.3}$$

are known numbers. The random variables $\xi_i$, $i = 1, \dots, N$, in (1.1) are nonobservable measurement errors. Below we impose some constraints on the limit behavior of the distributions of some linear combinations of these random variables.

In this article we consider the problem of estimating the unknown vector $\theta$ with coordinates $\theta_j > 0$, $j = 1, \dots, m$, through the random variables $Z_1, \dots, Z_N$. We propose some rather simple method for obtaining asymptotically normal estimators of unknown parameters for the linear-fractional regression model (1.1)–(1.3). Unlike the method of least squares which is usually used for solving such nonlinear regression problems, implementation of the proposed method does not require iterational procedures which in turn create difficulties in selecting initial approximation, convergence rate of the process, etc. and necessitate employment of computers due to a huge number of iterations.

The main goal of this article is to describe the method for constructing estimators in its general form together with the scheme of studying these estimators, as well as to demonstrate application of some ideas that can be used in studying the estimators. A mathematically-rigorous complete justification of the method was given in [1] in the simplest one-dimensional case of the linear-fractional regression problem. In a forthcoming article, the authors thoroughly study the case

*Date*: July 24, 2000. Revision submitted November 30, 2000.

of two unknown parameters which covers the Michaelis–Menten equation playing an important role in biochemistry.

Our method bases on an analogy that we revealed between the linear-fractional regression problem (1.1) and the classical linear regression problem (see (2.2) and (2.3)). For this reason, there is a certain analogy between the standard method for obtaining estimators of parameters in the classical linear regression problem (see, for instance, [2]) and the method for obtaining estimators in a more complicated linear-fractional regression problem (1.1) which is proposed in (2.5). In both problems, the estimators of parameters are sought as solutions to systems of linear equations in which the number of equations coincides with the number of the unknown parameters. In both cases, obtaining the estimators, we do not use laborious iterative procedures for approximate search of estimators.

**Structure of the article.** In § 2 we describe the method for constructing estimators in the problem under consideration in the general case and in § 3 we present a condition for consistency and asymptotic normality of the estimators. In § 4 we describe a broader class of estimators and find conditions for their asymptotic normality. In § 5 we obtain necessary conditions for optimality of the estimators and thereby indicate a way of finding such estimators. In § 6 and § 7 we consider several important particular cases. Some recommendations on practical application of the estimators can be found in § 6–§ 8.

**Notation.** The expression $\mathbf{A} = A_{m\times n}$ means that $\mathbf{A}$ is a matrix with $m$ rows and $n$ columns. The entry at the intersection of the $p$th row and the $q$th column is denoted by $(\mathbf{A})_{pq}$. The symbol $^\top$ stands for transposition of a vector or matrix. If $t$ is a vector with coordinates $t_1, \dots, t_N$ then $t$ is a column vector and $t^\top = (t_1, \dots, t_N)$ is a row vector. The symbols $\mathbf{I} = I_{n\times n}$ and $\mathbf{0} = 0_{n\times n}$ denote the identity and zero matrices of the corresponding size and $\operatorname{diag}\{h_1, \dots, h_N\}$ denotes the diagonal $(N \times N)$-matrix with entries $h_1, \dots, h_N$ on the principal diagonal. Given a nonnegative definite symmetric matrix $\mathbf{B}$, we denote by $\mathbf{B}^{1/2}$ the unique nonnegative definite symmetric matrix that satisfies the equality

$$(\mathbf{B}^{1/2})^2 = \mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{1/2\top}.$$

All limits are calculated as $N \to \infty$. If $\mathbf{A}$ is a random matrix or vector (whose entries may depend on $N$) then the notation $\mathbf{A} \xrightarrow{p} \mathbf{A}_0$ means coordinatewise convergence in probability of the entries of the matrix or vector (as $N \to \infty$). In this case the convergence

$$\mathbf{A} \Longrightarrow \Phi_m(0, \mathbf{I})$$

means that the distribution of the vector $\mathbf{A}$ may depend on $N$ and converges weakly (as $N \to \infty$) to the $m$-dimensional standard normal distribution $\Phi_m(0, \mathbf{I})$.

We say that some $m$-dimensional statistic $\tilde{\theta}^*$ is an *asymptotically normal estimator* for an $m$-dimensional parameter $\theta$ with asymptotic covariance matrix $\mathbf{K}\mathbf{K}^\top$ if the distribution of the vector $\mathbf{K}^{-1}(\tilde{\theta}^* - \theta)$ converges weakly to the $m$-dimensional standard normal distribution $\Phi_m(0, \mathbf{I})$.

## 2. Construction of Estimators for an Unknown Parameter

**2.1.** First of all, we reduce (1.1) to a form more convenient for finding estimators. To this end, we multiply both sides of (1.1) by the denominator $\beta_i(\theta)$ in (1.2) to obtain

$$Z_i + \sum_{j=1}^{m} b_{ji} Z_i \theta_j = a_{0i} + \sum_{j=1}^{m} a_{ji}\theta_j + \beta_i(\theta)\xi_i. \tag{2.1}$$

Put

$$X_{ji} = a_{ji} - b_{ji} Z_i, \quad Y_i = Z_i - a_{0i}, \quad \eta_i = \beta_i(\theta)\xi_i. \tag{2.2}$$

Inserting (2.2) in (2.1), we rewrite (1.1) in the following equivalent form:

$$Y_i = \sum_{j=1}^{m} X_{ji}\theta_j + \eta_i, \quad i = 1, \ldots, N. \tag{2.3}$$

**2.2.** It is easy to see that the form of equations (2.3) is analogous to that of the equations in the classical linear regression problem (see, for instance, [2]). Therefore, it is natural to try to find an estimator for the unknown parameters $\theta_1, \ldots, \theta_m$ in the same way as in the linear regression problem. We multiply (2.3) by some constants $c_{ki}$, where $k = 1, \ldots, m$, $i = 1, \ldots, N$, and sum up the results over $i$ to obtain

$$\sum_{i=1}^{N} c_{ki} Y_i - \sum_{i=1}^{N}\sum_{j=1}^{m} c_{ki} X_{ji}\theta_j = \sum_{i=1}^{N} c_{ki}\eta_i. \tag{2.4}$$

We may suppose that the right-hand side of (2.4), representing a weighted sum of measurement errors, is small as compared with the other summands; therefore, it is natural to discard the right-hand side of (2.4), replacing the unknown parameter $\theta$ in the so-obtained identity with the estimator $\theta^*$. Therefore, at the first step we always find the required estimators $\theta_1^*, \ldots, \theta_m^*$ as solutions to the following system of linear equations:

$$\sum_{i=1}^{N}\sum_{j=1}^{m} c_{ki} X_{ji}\theta_j^* = \sum_{i=1}^{N} c_{ki} Y_i, \quad k = 1, \ldots, m, \tag{2.5}$$

with appropriately chosen constants $\{c_{ki}\}$. Below in § 6 we give recommendations concerning the choice of these constants.

Observe that the main and most serious difference of equations (2.3) from the similar equations of linear regression consists in the fact that $\{X_{ji}\}$ in (2.3) are random variables of a special form (see (2.2)) rather than constants. This circumstance makes the problem of studying the properties of the estimators $\theta_1^*, \ldots, \theta_m^*$ much more laborious than in the case of linear regression.

**2.3.** To study the properties of the constructed estimators, it is more convenient to pass to matrix notation. We introduce the matrices $\mathbf{C} = \mathbf{C}_{m\times N}$ and $\mathbf{X} = \mathbf{X}_{m\times N}$ with entries

$$(\mathbf{C})_{ki} = c_{ki}, \quad (\mathbf{X})_{ji} = X_{ji} = a_{ji} - b_{ji} Z_i$$

and introduce the vectors

$$Y = (Y_1, \ldots, Y_N)^\top, \quad \eta = (\eta_1, \ldots, \eta_N)^\top, \quad \xi = (\xi_1, \ldots, \xi_N)^\top,$$

$$\theta = (\theta_1, \ldots, \theta_m)^\top, \quad \theta^* = (\theta_1^*, \ldots, \theta_m^*)^\top.$$

With these notations, we can rewrite (2.3) and (2.5) in the following more compact form:

$$Y = \mathbf{X}^\top\theta + \eta, \quad \mathbf{C}\mathbf{X}^\top\theta^* = \mathbf{C}Y. \tag{2.6}$$

In particular, from (2.6) we derive the equality

$$\mathbf{C}\mathbf{X}^\top(\boldsymbol{\theta}^* - \boldsymbol{\theta}) = \mathbf{C}\eta. \tag{2.7}$$

Now, we define the matrices $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\theta})_{m\times N}$ and $\boldsymbol{\Psi} = \boldsymbol{\Psi}_{m\times N}$ by putting

$$(\boldsymbol{\Lambda})_{ji} = \lambda_{ji}(\theta) = a_{ji} - b_{ji}g_i(\theta), \quad (\boldsymbol{\Psi})_{ji} = \psi_{ji} = b_{ji}\xi_i.$$

In this case

$$\mathbf{X} = \boldsymbol{\Lambda} - \boldsymbol{\Psi} \tag{2.8}$$

and we can rewrite (2.7) as follows:

$$(\mathbf{C}\boldsymbol{\Lambda}^\top - \mathbf{C}\boldsymbol{\Psi}^\top)(\theta^* - \theta) = \mathbf{C}\eta. \tag{2.9}$$

Equalities (2.7) and (2.9) will be useful in the next section while studying the behavior of the difference $\theta^* - \theta$ determining the accuracy of estimation.

## 3. Consistency and Asymptotic Normality

**3.1.** Henceforth we suppose that the entries of the matrix $\mathbf{C}$ are chosen so that the matrix $\mathbf{C}\boldsymbol{\Lambda}^\top = (\mathbf{C}\boldsymbol{\Lambda}(\theta)^\top)_{m\times m}$ is nondegenerate. Multiplying both sides of (2.9) by $(\mathbf{C}\boldsymbol{\Lambda}^\top)^{-1}$, we rewrite (2.9) as

$$(I - \mathbf{G}\boldsymbol{\Psi}^\top)(\boldsymbol{\theta}^* - \boldsymbol{\theta}) = \mathbf{G}\eta \quad \text{for} \quad \mathbf{G} = (\mathbf{C}\boldsymbol{\Lambda}^\top)^{-1}\mathbf{C}. \tag{3.1}$$

Using (3.1), we can easily "guess" the form of Theorem 1 stated below.

**Theorem 1**. *Suppose that the following conditions are satisfied:*

$$(\mathbf{C}\boldsymbol{\Lambda}^\top)^{-1}\mathbf{C}\boldsymbol{\Psi}^\top \xrightarrow{p} \mathbf{0}, \tag{3.2}$$

$$(\mathbf{C}\boldsymbol{\Lambda}^\top)^{-1}\mathbf{C}\eta \xrightarrow{p} \mathbf{0}. \tag{3.3}$$

*Then the estimator $\theta^*$ is consistent.*

**Remark 1** It is clear that

$$\|\mathbf{G}\eta\|^2 = \eta^\top\mathbf{G}^\top\mathbf{G}\eta = \sum_{i,l=1}^{N}(\mathbf{G}^\top\mathbf{G})_{il}\beta_i(\theta)\beta_l(\theta)\xi_i\xi_l.$$

Thus, if $\mathbf{E}\xi_i^2 < \infty$ for all $i = 1, 2, \ldots$, then for validity of (3.3), it is sufficient that

$$\mathbf{E}\|\mathbf{G}\eta\|^2 = \sum_{i,l=1}^{N}(\mathbf{G}^\top\mathbf{G})_{il}\beta_i(\theta)\beta_l(\theta)\mathbf{E}\xi_i\xi_l \xrightarrow{p} 0.$$

Similarly, in this case we can easily write down the conditions sufficient for convergence (3.2):

$$\sum_{j,k=1}^{m}\mathbf{E}(\mathbf{G}\boldsymbol{\Psi}^\top)_{jk}^2 = \sum_{k=1}^{m}\sum_{i,l=1}^{N}(\mathbf{G}^\top\mathbf{G})_{il}b_{ki}b_{kl}\mathbf{E}\xi_i\xi_l \to 0.$$

**3.2.** Now, suppose that the random vector $\mathbf{C}\eta$ is asymptotically normal in the sense that there is a random or nonrandom nondegenerate matrix $\mathbf{U}$ such that

$$\mathbf{U}\mathbf{C}\eta \Longrightarrow \Phi_m(0, \mathbf{I}). \tag{3.4}$$

In this case from (2.7) we obtain the equality

$$\mathbf{U}\mathbf{C}\mathbf{X}^\top\mathbf{A}^{-1}\mathbf{A}(\theta^* - \theta) = \mathbf{U}\mathbf{C}\eta \tag{3.5}$$

for every nondegenerate matrix $\mathbf{A}$. Hence, we can easily "guess" the form of Theorem 2 below.

**Theorem 2**. *Suppose that* (3.4) *is satisfied and*

$$\mathbf{UCX}^\top\mathbf{A}^{-1} \xrightarrow{p} \mathbf{I} \tag{3.6}$$

*for some nondegenerate random or nonrandom matrix* $\mathbf{A}$. *Then*

$$\mathbf{A}(\theta^* - \theta) \Longrightarrow \Phi_m(0, \mathbf{I}). \tag{3.7}$$

**Corollary 1**. *Suppose that* (3.4) *is satisfied and*

$$\mathbf{UC\Lambda}^\top\mathbf{A}^{-1} \xrightarrow{p} \mathbf{I}, \tag{3.8}$$

$$\mathbf{UC\Psi}^\top\mathbf{A}^{-1} \xrightarrow{p} \mathbf{0}. \tag{3.9}$$

*Then assertion* (3.7) *of Theorem* 2 *is valid.*

**Corollary 2**. *Suppose that assertion* (3.7) *of Theorem* 2 *is valid for some nonrandom matrix* $\mathbf{A}$. *Then*

$$(\mathbf{A}^\top\mathbf{A})^{1/2}(\theta^* - \theta) \Longrightarrow \Phi_m(0, \mathbf{I}).$$

**Remark 2**. In the above statements, it is natural to put $\mathbf{A} = \mathbf{UC\Lambda}^\top$. In this case condition (3.8) of Corollary 1 is fulfiled automatically. It is in this way that we proceed below in Corollary 5 and Theorem 6. In Theorem 2, the simplest choice is $\mathbf{A} = \mathbf{UCX}^\top$, since in this case condition (3.6) is checked elementarily (see also Remark 13).

**Remark 3**. Observe that the estimator $\theta^*$ may satisfy the conditions of Theorem 2 but fail to be consistent. The corresponding example is constructed in [1, Remark 6].

In § 7, we present simple sufficient conditions guaranteeing validity of all above conditions in the case of independent observations.

**3.3.** We turn to the proofs of the claims. We need the following

**Lemma 1**. *Suppose that an m-dimensional random vector* $\zeta_N$ *satisfies the condition*

$$\zeta_N \Longrightarrow \Phi_m(0, \mathbf{I}) \tag{3.10}$$

*and a random* $(m \times m)$*-matrix* $\mathbf{J}_N$ *and an m-dimensional random vector* $\delta_N$ *are such that*

$$\mathbf{J}_N \xrightarrow{p} \mathbf{I}, \quad \delta_N \xrightarrow{p} \mathbf{0}.$$

*Then*

$$\mathbf{J}_N(\zeta_N + \delta_N) \Longrightarrow \Phi_m(0, \mathbf{I}) \quad \text{and} \quad \mathbf{J}_N^{-1}(\zeta_N + \delta_N) \Longrightarrow \Phi_m(0, \mathbf{I}).$$

This assertion is easy on using the Cramér–Wald method; therefore, we omit the proof.

*Proof of Theorem 1.* Rewrite (3.1) as

$$\theta^* - \theta = (\mathbf{I} - \mathbf{G\Psi}^\top)^{-1}\mathbf{G}\eta,$$

where $\mathbf{G\Psi}^\top \xrightarrow{p} 0$ by (3.2). Now, the claim ensues from (3.3) and Lemma 1 for $\mathbf{J}_N = \mathbf{I} - \mathbf{G\Psi}^\top$.

*Proof of Theorem 2.* From (3.5) we find that

$$\mathbf{A}(\theta^* - \theta) = (\mathbf{UCX}^\top\mathbf{A}^{-1})^{-1}\mathbf{UC}\eta.$$

Now, we easily derive the assertion of the theorem from (3.4), (3.6), and Lemma 1 for $\mathbf{J}_N = \mathbf{UCX}^\top\mathbf{A}^{-1}$.

*Proof of Corollary 1.* Note that by (2.8)

$$\mathbf{UCXA}^{-1} = \mathbf{UC\Lambda}^\top\mathbf{A}^{-1} - \mathbf{UC\Psi}^\top\mathbf{A}^{-1}. \tag{3.11}$$

The above representation, together with (3.8) and (3.9), immediate yields assumption (3.6) of Theorem 2.

**3.4.** Before proving Corollary 2, we state one useful assertion.

**Lemma 2**. *Suppose that* (3.10) *is satisfied and* $\mathbf{Q}_N$ *is some orthogonal matrix; i.e.,* $\mathbf{Q}_N\mathbf{Q}_N^\top = \mathbf{I}$. *Then*

$$\mathbf{Q}_N\zeta_N \Longrightarrow \Phi_m(0,\mathbf{I}).$$

*Proof.* Using a generalization of the theorem about uniform convergence of characteristic functions for multidimensional random variables (see [3, Section 13.3]), we conclude that

$$(\forall K<\infty)\ \sup_{\|t\|\le K}|\mathbf{E}e^{i\langle\zeta_N,t\rangle} - e^{-\|t\|^2/2}| \to 0.$$

From orthogonality of $\mathbf{Q}_N$ we obtain $\|t\| = \|\mathbf{Q}_N t\|$ and therefore

$$\sup_{\|t\|\le K}|\mathbf{E}e^{i\langle\mathbf{Q}_N\zeta_N,t\rangle} - e^{-\|t\|^2/2}| = \sup_{\|\mathbf{Q}_N^\top t\|\le K}|\mathbf{E}e^{i\langle\zeta_N,\mathbf{Q}_N^\top t\rangle} - e^{-\|\mathbf{Q}_N^\top t\|^2/2}| \to 0.$$

The so-obtained convergence of the characteristic functions leads to the sought assertion of Lemma 2.

*Proof of Corollary 2.* Put $\mathbf{A}_0 = (\mathbf{A}^\top\mathbf{A})^{1/2}$. Then

$$(\mathbf{A}_0\mathbf{A}^{-1})(\mathbf{A}_0\mathbf{A}^{-1})^\top = \mathbf{A}_0\mathbf{A}^{-1}\mathbf{A}^{-1\top}\mathbf{A}_0{}^\top = \mathbf{A}_0(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}_0{}^\top = \mathbf{I}.$$

Hence, $\mathbf{Q}_N = \mathbf{A}_0\mathbf{A}^{-1}$ is an orthogonal matrix. Thus, from Theorem 2 and Lemma 2 we immediately obtain the claim:

$$(\mathbf{A}\mathbf{A}^\top)^{1/2}(\theta^*-\theta) = \mathbf{Q}_N\mathbf{A}(\theta^*-\theta) \Longrightarrow \Phi_m(0,\mathbf{I}).$$

## 4. Improvement of Estimators

**4.1.** Suppose that at the first step of estimation we have constructed some estimator $\theta^*$ for $\theta$ as a solution to (2.5). At the second step we can introduce the estimators $\theta_1^{**},\dots,\theta_N^{**}$ as solutions to the system of the equations

$$\sum_{i=1}^N\sum_{j=1}^m \gamma_{ki}(\theta^*)X_{ji}\theta_j^{**} = \sum_{i=1}^N \gamma_{ki}(\theta^*)Y_i, \quad k=1,\dots,m, \tag{4.1}$$

where $\gamma_{ki}(\theta)$ are some specially chosen functions depending only on the unknown parameter $\theta$. Practical recommendations concerning the choice of the functions $\gamma_{ki}(\theta)$ are given in § 6.

Observe that (4.1) differs from (2.5) in replacing the numbers $c_{ki}$ by the statistics $\gamma_{ki}(\theta^*)$. We emphasize that the employment of the statistics $\gamma_{ki}(\theta^*)$ instead of the numbers $c_{ki}$ essentially expands the class of estimators. Thereby, for a good choice of the functions $\gamma_{ki}(\theta)$, we acquire the possibility of using the “improved” estimators $\theta_k^{**}$ instead of $\theta_k^*$.

In particular, in § 5 and § 6 we show that, under rather general conditions, we can choose the functions $\gamma_{ki}$ so that the covariance matrix of the improved estimator $\theta^{**}$ be minimal.

**4.2.** Now, we turn to studying these estimators. By analogy with (2.6), it is convenient to rewrite (4.1) in the following equivalent matrix form:

$$\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top\theta^{**} = \mathbf{\Gamma}(\theta^*)Y, \tag{4.2}$$

where $\mathbf{\Gamma}(\boldsymbol{\theta}) = \mathbf{\Gamma}(\boldsymbol{\theta})_{m\times N}$ with $(\mathbf{\Gamma}(\theta))_{ki} = \gamma_{ki}(\theta)$. In particular, from (2.6) and (4.2) we obtain the following analog of (2.7):

$$\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top(\theta^{**} - \theta) = \mathbf{\Gamma}(\theta^*)\eta. \tag{4.3}$$

**4.3.** Now, suppose that the random vector $\mathbf{\Gamma}(\theta)\eta$ is asymptotically normal in the sense that there is a random or nonrandom nondegenerate matrix $\mathbf{U}_\Gamma$ such that

$$\mathbf{U}_\Gamma\mathbf{\Gamma}(\theta)\eta \Longrightarrow \Phi_m(0, \mathbf{I}). \tag{4.4}$$

In this case we have the following analogs of assertions of the preceding section:

**Theorem 3**. *Suppose that* (4.4) *is satisfied and*

$$\delta_{0,N} \equiv \mathbf{U}_\Gamma(\mathbf{\Gamma}(\theta^*) - \mathbf{\Gamma}(\theta))\eta \xrightarrow{p} \mathbf{0}, \tag{4.5}$$

$$\mathbf{U}_\Gamma\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top\mathbf{A}_\Gamma^{-1} \xrightarrow{p} \mathbf{I} \tag{4.6}$$

*for some random or nonrandom nondegenerate matrix* $\mathbf{A}_\Gamma$. *Then*

$$\mathbf{A}_\Gamma(\theta^{**} - \theta) \Longrightarrow \Phi_m(0, \mathbf{I}). \tag{4.7}$$

**Corollary 3**. *Suppose that* (4.4) *and* (4.5) *are satisfied and*

$$\mathbf{U}_\Gamma\mathbf{\Gamma}(\theta)\mathbf{\Lambda}^\top\mathbf{A}_\Gamma^{-1} \xrightarrow{p} \mathbf{I}, \tag{4.8}$$

$$\mathbf{\Delta}_{1,N} \equiv \mathbf{U}_\Gamma\mathbf{\Gamma}(\theta)\mathbf{\Psi}^\top\mathbf{A}_\Gamma^{-1} \xrightarrow{p} \mathbf{0}, \tag{4.9}$$

$$\mathbf{\Delta}_{2,N} \equiv \mathbf{U}_\Gamma(\mathbf{\Gamma}(\theta^*) - \mathbf{\Gamma}(\theta))\mathbf{\Lambda}^\top\mathbf{A}_\Gamma^{-1} \xrightarrow{p} \mathbf{0}, \tag{4.10}$$

$$\mathbf{\Delta}_{3,N} \equiv \mathbf{U}_\Gamma(\mathbf{\Gamma}(\theta^*) - \mathbf{\Gamma}(\theta))\mathbf{\Psi}^\top\mathbf{A}_\Gamma^{-1} \xrightarrow{p} \mathbf{0}. \tag{4.11}$$

*Then assertion* (4.7) *of Theorem* 3 *is valid.*

**Corollary 4**. *Suppose that assertion* (4.7) *of Theorem* 3 *is valid for some nonrandom matrix* $\mathbf{A}_\Gamma$. *Then*

$$(\mathbf{A}_\Gamma^\top\mathbf{A}_\Gamma)^{1/2}(\theta^* - \theta) \Longrightarrow \Phi_m(0, \mathbf{I}).$$

**Remark 4** In the above assertions the simplest choice is $\mathbf{A}_\Gamma = \mathbf{U}_\Gamma\mathbf{\Gamma}(\boldsymbol{\theta})\mathbf{\Lambda}^\top$. In this case, condition (4.8) of Corollary 3 is satisfied automatically. It is in this way that we proceed in Corollary 6. If we put $\mathbf{A}_\Gamma = \mathbf{U}_\Gamma\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top$ in Theorem 3 then (4.6) is verified elementarily. We propose a similar idea below in Remark 13.

**4.4.** The rest of the section is devoted to proving the claims.

*Proof of Theorem 3.* Multiplying equality (4.3) by the matrix $\mathbf{U}_\Gamma$ from the left, we obtain

$$\mathbf{U}_\Gamma\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top\mathbf{A}_\Gamma^{-1}\mathbf{A}_\Gamma(\theta^{**} - \theta) = \mathbf{U}_\Gamma\mathbf{\Gamma}(\theta^*)\eta;$$

whence, using the notations introduced in the statement of the theorem, we can easily derive the following important identity:

$$\mathbf{A}_\Gamma(\theta^{**} - \theta) = \big(\mathbf{U}_\Gamma\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top\mathbf{A}_\Gamma^{-1}\big)^{-1}(\mathbf{U}_\Gamma\mathbf{\Gamma}(\theta)\eta + \delta_{0,N}).$$

To complete proof of the theorem, it remains to use Lemma 1, the above identity, and the conditions of the theorem.

*Proof of Corollary 3.* By (2.8),

$$\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top = \mathbf{\Gamma}(\theta)(\mathbf{\Lambda} - \mathbf{\Psi}) + (\mathbf{\Gamma}(\theta^*) - \mathbf{\Gamma}(\theta))(\mathbf{\Lambda} - \mathbf{\Psi}).$$

Involving (4.9)–(4.11), we thus obtain

$$\mathbf{U}_\Gamma\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top\mathbf{A}_\Gamma^{-1} = \mathbf{U}_\Gamma\mathbf{\Gamma}(\boldsymbol{\theta})\mathbf{\Lambda}^\top\mathbf{A}_\Gamma^{-1} - \mathbf{\Delta}_{1,N} + \mathbf{\Delta}_{2,N} - \mathbf{\Delta}_{3,N}.$$

Using this identity, we can easily verify that conditions (4.8)–(4.11) of Corollary 3 are sufficient for validity of assumption (4.6) of Theorem 3.

*Proof of Corollary 4.* Repeating the argument of the proof of Corollary 2, we easily establish that $\mathbf{Q}_N = \left(\mathbf{A}_\Gamma^\top\mathbf{A}_\Gamma\right)^{1/2}\mathbf{A}_\Gamma^{-1}$ is an orthogonal matrix. Consequently,

$$\left(\mathbf{A}_\Gamma^\top\mathbf{A}_\Gamma\right)^{1/2}(\theta^{**} - \theta) = \mathbf{Q}_N\mathbf{A}_\Gamma(\theta^{**} - \theta) \Longrightarrow \Phi_m(0, \mathbf{I})$$

which is immediate from Lemma 2 and Theorem 3.

## 5. Optimization of Estimators

**5.1.** Henceforth we suppose that the measurement errors satisfy the following natural assumptions:

$$(\forall i)\ \mathbf{E}\eta_i = \beta_i(\theta)\mathbf{E}\xi_i = 0,\ 0 < \mathbf{D}\eta_i = \beta_i^2(\theta)\mathbf{D}\xi_i < \infty.$$

In this case we denote by $\mathbf{V}$ the covariance matrix of the random vector $\eta$. In other words, we suppose below that

$$\exists\mathbf{E}\eta = 0, \quad \exists\mathbf{V} = \mathbf{E}\eta\eta^\top, \quad \exists\mathbf{V}^{-1}.$$

Introduce the following class of matrices:

$$\mathcal{M}(\theta, \mathbf{V}) = \{\mathbf{\Gamma} = \mathbf{\Gamma}_{m\times N} : \exists(\mathbf{\Gamma}\mathbf{\Lambda}^\top)^{-1},\ \exists(\mathbf{\Gamma}\mathbf{V}\mathbf{\Gamma}^\top)^{-1}\}.$$

For $\mathbf{\Gamma} \in \mathcal{M}(\boldsymbol{\theta}, \mathbf{V})$ we use the notation

$$\mathbf{B}(\mathbf{\Gamma}, \theta, \mathbf{V}) \equiv (\mathbf{\Gamma}\mathbf{\Lambda}^\top)^{-1}\mathbf{\Gamma}\mathbf{V}\mathbf{\Gamma}^\top(\mathbf{\Gamma}\mathbf{\Lambda}^\top)^{-1\top}. \tag{5.1}$$

Henceforth we also assume that $\mathbf{C}, \mathbf{\Gamma}(\theta) \in \mathcal{M}(\theta, \mathbf{V})$.

**5.2.** Take $\mathbf{U}$ and $\mathbf{U}_\Gamma$ to be arbitrary nonrandom matrices satisfying the conditions

$$\mathbf{U}^\top\mathbf{U} = (\mathbf{C}\mathbf{V}\mathbf{C}^\top)^{-1}, \quad \mathbf{U}_\Gamma^\top\mathbf{U}_\Gamma = (\mathbf{\Gamma}(\theta)\mathbf{V}\mathbf{\Gamma}^\top(\theta))^{-1}. \tag{5.2}$$

Now, put

$$\mathbf{A} = \mathbf{U}\mathbf{C}\mathbf{\Lambda}^\top, \quad \mathbf{A}_\Gamma = \mathbf{U}_\Gamma\mathbf{\Gamma}(\theta)\mathbf{\Lambda}^\top. \tag{5.3}$$

For the so-defined matrices, (3.8) and (4.8) hold automatically; (3.4) and (4.4) mean validity of the classical central limit theorem; and Corollaries 2 and 4 take the following simple form:

**Corollary 5**. *Suppose that the matrices* $\mathbf{U}$ *and* $\mathbf{A}$ *in* (5.2) *and* (5.3) *satisfy* (3.4) *and* (3.9)*. Then*

$$\mathbf{B}_C^{-1/2}(\theta^* - \theta) \Longrightarrow \Phi_m(0, \mathbf{I}) \quad \textit{for} \quad \mathbf{B}_C = (\mathbf{A}^\top\mathbf{A})^{-1}.$$

**Corollary 6**. *Suppose that the matrices* $\mathbf{U}_\Gamma$ *and* $\mathbf{A}_\Gamma$ *in* (5.2) *and* (5.3) *satisfy* (4.4)*,* (4.5)*, and* (4.9)–(4.11)*. Then the following convergence holds:*

$$\mathbf{B}_\Gamma^{-1/2}(\theta^{**} - \theta) \Longrightarrow \Phi_m(0, \mathbf{I}) \quad \textit{for} \quad \mathbf{B}_\Gamma = ({\mathbf{A}_\Gamma}^\top\mathbf{A}_\Gamma)^{-1}.$$

**Remark 5** It is easy to see that the estimators $\theta^*$ and $\theta^{**}$ remain the same if we multiply their defining equations (2.6) and (4.2) from the left by some nondegenerate matrices. In other words, we can find these estimators as solutions to the equations

$$\mathbf{R}_C\mathbf{C}\mathbf{X}^\top\theta^* = \mathbf{R}_C\mathbf{C}Y, \quad \mathbf{R}_\Gamma\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top\theta^{**} = \mathbf{R}_\Gamma\mathbf{\Gamma}(\theta^*)Y$$

for arbitrary nondegenerate matrices $\mathbf{R}_C = (\mathbf{R}_C)_{m\times N}$ and $\mathbf{R}_\Gamma = (\mathbf{R}_\Gamma)_{m\times N}$.

It is easy to see that in this case the asymptotic covariance matrices of the estimators $\theta^*$ and $\theta^{**}$ remain the same.

**5.3.** Thus, under the conditions of Corollaries 5 and 6, the estimators $\theta^*$ and $\theta^{**}$ are asymptotically normal with the respective asymptotic covariance matrices $\mathbf{B}_C$ and $\mathbf{B}_\Gamma$. Clearly, the less the asymptotic covariance matrices of estimators are, the more exact are the estimators. However, observe that

$$\mathbf{B}_C = \mathbf{B}(\mathbf{C}, \theta, \mathbf{V}), \quad \mathbf{B}_\Gamma = \mathbf{B}(\mathbf{\Gamma}(\theta), \theta, \mathbf{V}). \tag{5.4}$$

Therefore, it is natural to try to find matrices that are in a sense minimal among the matrices of the form $\mathbf{B}(\mathbf{\Gamma}, \boldsymbol{\theta}, \mathbf{V})$.

Furthermore, by [4] the inequality $\mathbf{B}_1 \geq \mathbf{B}_2$ for two nonnegative definite symmetric matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ means that the matrix $\mathbf{B}_1 - \mathbf{B}_2$ is nonnegative definite; i.e., we have the inequality $t^\top\mathbf{B}_1 t \geq t^\top\mathbf{B}_2 t$ between the quadratic forms for every column vector $t = (t_1, \ldots, t_N)^\top$.

We introduce the notation

$$\mathbf{B}^{opt}(\theta, \mathbf{V}) = (\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{\Lambda}^\top)^{-1}, \quad \mathbf{\Gamma}^o(\theta, \mathbf{V}) = \mathbf{\Lambda}\mathbf{V}^{-1}. \tag{5.5}$$

**Theorem 4**. *There is a nonnegative definite symmetric matrix* $\mathbf{B}^o$ *such that the inequality*

$$(\mathbf{\Gamma}\mathbf{\Lambda}^\top)(\mathbf{B}(\mathbf{\Gamma}, \theta, \mathbf{V})) - \mathbf{B}^{opt}(\theta, \mathbf{V}))(\mathbf{\Gamma}\mathbf{\Lambda}^\top)^\top = (\mathbf{\Gamma} - \mathbf{R}\mathbf{\Gamma}^o)\mathbf{B}^o(\mathbf{\Gamma} - \mathbf{R}\mathbf{\Gamma}^o)^\top \tag{5.6}$$

*holds for all matrices* $\mathbf{\Gamma} \in \mathcal{M}(\theta, \mathbf{V})$ *and all nondegenerate matrices* $\mathbf{R} = \mathbf{R}_{m\times m}$.

**Corollary 7**. *The following relation is valid for all matrices* $\mathbf{\Gamma} \in \mathcal{M}(\theta, \mathbf{V})$*:*

$$\mathbf{B}(\mathbf{\Gamma}, \theta, \mathbf{V}) \geq \mathbf{B}^{opt}(\theta, \mathbf{V}). \tag{5.7}$$

*Moreover, the equality*

$$\mathbf{B}(\mathbf{\Gamma}^{opt}, \theta, \mathbf{V}) = \mathbf{B}^{opt}(\theta, \mathbf{V}) \tag{5.8}$$

*holds if*

$$\mathbf{\Gamma}^{opt} = \mathbf{R}\mathbf{\Lambda}\mathbf{V}^{-1} \equiv \mathbf{R}\mathbf{\Gamma}^o(\theta, \mathbf{V}), \tag{5.9}$$

*where* $\mathbf{R} = R_{m\times m}$ *is an arbitrary nondegenerate matrix.*

We emphasize that in the general case the matrix $\mathbf{\Gamma}^{opt}$ always depends on the unknown parameter $\theta$ and the covariance matrix $\mathbf{V}$. It is easy to see that the entries of $\mathbf{\Gamma}^{opt}$ are constant only under very special additional constraints on $\mathbf{\Lambda}$ and $\mathbf{V}$ (see Examples 3 and 5 in § 6).

**5.4. Remark 6**. We can interpret identity (5.6) as follows: the better the approximation of the matrix $\mathbf{C}$ at the first step or the matrix $\mathbf{\Gamma}(\theta)$ at the second step is by some matrix $\mathbf{\Gamma}^{opt} = \mathbf{R}\mathbf{\Gamma}^o(\theta, \mathbf{V})$, the closer the asymptotic covariance matrices $\mathbf{B}_C$ and $\mathbf{B}_\Gamma$ of the estimators $\theta^*$ and $\theta^{**}$ are to the optimal matrices $\mathbf{B}^{opt}(\theta, \mathbf{V})$. Moreover, the matrix on the right-hand side of (5.6) characterizes the degree of proximity rather clearly.

**Remark 7**. Suppose that the errors $\xi_i$ are independent and have standard normal distribution and that the variances $\sigma_i^2$ are independent of the parameter $\theta$. Then

$$\mathbf{B}^{opt}(\theta, \mathbf{V}) = \mathbf{I}_N^{-1}(\theta), \tag{5.10}$$

where $\mathbf{I}_N(\theta)$ is the Fisher information for the sample $Z_1, \dots, Z_N$. By analogy with the Rao–Cramér inequality, we should naturally expect a certain unimprovability of the estimators $\theta^*$ if we choose $\mathbf{C} = \mathbf{\Gamma}^{opt}$ and the estimators $\theta^{**}$ if we choose $\mathbf{\Gamma}(\theta) = \mathbf{\Gamma}^{opt}$.

Assertion (5.10) follows from representation (5.7) for $\mathbf{B}^{opt}(\theta, \mathbf{V})$ and the equality

$$(\mathbf{I}_N(\theta))_{jk} = \sum \frac{\lambda_{ji}(\theta)\lambda_{ki}(\theta)}{\beta_i^2(\theta)\sigma_i^2} = (\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{\Lambda}^\top)_{jk}$$

whose simple derivation is omitted.

**5.5.** To prove the claims, we need the following lemma improving Theorem 4.

**Lemma 3**. *Equality* (5.6) *holds for the matrix*

$$\mathbf{B}^o \equiv \mathbf{V} - \mathbf{\Lambda}^\top \mathbf{B}^{opt}\mathbf{\Lambda} \tag{5.11}$$

*which satisfies the relation*

$$\mathbf{B}^{o\top} = \mathbf{B}^o = \mathbf{B}^o\mathbf{V}^{-1}\mathbf{B}^{o\top} \geq 0. \tag{5.12}$$

*Proof.* Using (5.1) and (5.7) for $\mathbf{B}(\mathbf{\Gamma}, \theta, \mathbf{V}))$ and $\mathbf{B}^{opt}(\theta, \mathbf{V})$, we immediately infer that the left-hand side of (5.6) coincides with $\mathbf{\Gamma}^\top\mathbf{B}^o\mathbf{\Gamma}$ for $\mathbf{B}^o$ in (5.11). Using the identity

$$\mathbf{\Gamma}^o(\theta, \mathbf{V})\mathbf{B}^o = \mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{V} - \mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{\Lambda}^\top\mathbf{B}^{opt}\mathbf{\Lambda} = \mathbf{\Lambda} - (\mathbf{B}^{opt})^{-1}\mathbf{B}^{opt}\mathbf{\Lambda} = \mathbf{0},$$

we easily see that the right-hand side of (5.6) coincides with $\mathbf{\Gamma}^\top\mathbf{B}^o\mathbf{\Gamma}$ as well.

We have thus proven the first assertion of Lemma 3. Applying definitions (5.5) and (5.11) of $\mathbf{B}^{opt}(\theta, \mathbf{V})$ and $\mathbf{B}^o$, we readily validate (5.12).

Thus, Lemma 3 and Theorem 4 are proven completely. Now, we turn to derivation of Corollary 7. Relation (5.7) ensues from the fact that $\mathbf{B}(\mathbf{\Gamma}, \theta, \mathbf{V})) - \mathbf{B}^{opt}(\theta, \mathbf{V})$ is a nonnegative definite matrix by (5.6). If (5.9) is satisfied then (5.8) is an obvious consequence of the same identity (5.6).

## 6. On Choosing Constants and Functions in the Definition of Estimators

**6.1. Example 1** Suppose that the covariance matrix $\mathbf{V}$ is representable as

$$\mathbf{V} = \mathbf{W}(\theta)\sigma^2,$$

where $\mathbf{W}(\theta)$ is a matrix whose entries are known functions of $\theta$ and $\sigma^2$ is some unknown parameter. Put

$$\mathbf{\Gamma}(\theta) = \mathbf{\Lambda}(\theta)\mathbf{W}^{-1}(\theta).$$

If the matrix $\mathbf{\Gamma}(\theta)$ satisfies the conditions of Theorem 3 then the estimator $\theta^{**}$ is asymptotically normal with the optimal covariance matrix $\mathbf{B}^{opt}(\theta, \mathbf{V})$.

This is immediate from Corollary 8, since in this case $\mathbf{\Gamma}(\theta) = \sigma^2\mathbf{\Lambda}\mathbf{V}^{-1} \equiv \sigma^2\mathbf{\Gamma}^o(\theta, \mathbf{V})$.

**Remark 8**. In Example 1, we could recommend to take the numbers $c_{ki}$ at the first step to be $c_{ki} = \gamma_{ki}(\theta_0) \equiv (\mathbf{\Gamma}(\theta_0))_{ki}$ for some $\theta_0$ chosen in advance. Clearly, the closer the chosen value $\theta_0$ is to the unknown true value of the parameter $\theta$, the more exact is the estimator $\theta^*$.

Observe that, for $\mathbf{W}(\theta) \equiv \mathbf{I}$, the general case in Example 1 covers the situation of the classical regression analysis when the measurement errors $\xi_1, \dots, \xi_N$ are independent, identically distributed, and have mean zero, while the variances $\mathbf{D}\xi_i = \sigma^2$ coincide and are unknown.

**6.2.** In Examples 2–5, we assume that the random errors $\xi_1, \dots, \xi_N$ in (1.1) are not correlated and satisfy the conditions

$$\forall i \ \mathbf{E}\xi_i = 0, \ 0 < \beta_i^2(\theta)\mathbf{D}\xi_i \equiv \mathbf{D}\eta_i = \sigma^2/w_i(\theta) < \infty, \tag{6.1}$$

where $w_i(\theta)$ are known functions and $\sigma^2$ is an unknown parameter.

**Example 2** If (6.1) is satisfied then, as observed in Example 1, we can always put

$$(\mathbf{\Gamma}^{opt})_{ki} = \lambda_{ki}(\theta)w_i(\theta) \equiv (a_{ki} - b_{ki}g_i(\theta))w_i(\theta). \tag{6.2}$$

**Example 3** Suppose that

$$g_i(\theta) = a_{0i}/\beta_i(\theta) \tag{6.3}$$

and (6.1) is satisfied for

$$w_i(\theta) = w_{0i}/g_i(\theta), \tag{6.4}$$

where $w_{0i}$ are known constants. It is easy to see that in this case we can choose the optimal constants $c_{ki}$ already at the first step of construction of the estimator $\theta^*$ as follows:

$$c_{ki} = b_{ki}w_{0i} \equiv (\mathbf{\Gamma}^{opt})_{ki}. \tag{6.5}$$

**Example 4** Suppose that (6.1) is satisfied and

$$g_i(\theta) = (c_0 a_{mi} + a_{mi}\theta_m)/\beta_i(\theta). \tag{6.6}$$

As we show below in Lemma 4, in this case we can define the functions $(\mathbf{\Gamma}^{opt})_{ki}$ in Corollary 8 by the following simpler formula than (6.2):

$$(\mathbf{\Gamma}^{opt})_{mi} = (1 - c_0 b_{mi})g_i(\theta)w_i(\theta) \quad \text{and} \quad (\mathbf{\Gamma}^{opt})_{ki} = b_{ki}g_i(\theta)w_i(\theta) \text{ for } k < m. \tag{6.7}$$

**Example 5** Suppose that (6.1), (6.4), and (6.6) are satisfied. Under these assumptions, the functions $(\mathbf{\Gamma}^{opt})_{ki}$ in (6.7) are independent of $\theta$. Hence, in this case we can choose the optimal constants $c_{ki}$ already at the first step by putting

$$c_{mi} = (1 - c_0 b_{mi})w_{0i} \equiv (\mathbf{\Gamma}^{opt})_{mi} \quad \text{and} \quad c_{ki} = b_{ki}w_{0i} \equiv (\mathbf{\Gamma}^{opt})_{ki} \text{ for } k < m. \tag{6.8}$$

**6.3. Remark 9**. If in Example 3 we only know that (6.3) is satisfied then we can recommend using $c_{ki} = b_{ki}$ as the simplest $c_{ki}$ in (6.5) even when we do not know whether (6.4) holds or not. However, this choice of $c_{ki}$ is not optimal.

Similarly, if in Example 5 we only know that (6.4) is satisfied then we can recommend using $c_{ki}$ in (6.8) for $w_{0i} \equiv 1$ even when (6.3) fails or there is no information of its validity.

**Remark 10**. If the exact form of the covariance matrix $\mathbf{V}$ is unknown then we cannot find the matrix $\mathbf{\Gamma}^{opt}$ and construct the estimator $\theta^{**}$ for $\mathbf{\Gamma}(\theta) = \Gamma^{opt}$. Then we can recommend to take the entries of $\mathbf{\Gamma}(\theta)$ to be functions of which we may assume that they "differ slightly" from the unknown entries of the matrix $\mathbf{\Gamma}^{opt}$. Therefore, as mentioned in Remark 6, the "closer" the entries of the matrix $\mathbf{\Gamma}(\theta)$ are to the entries of the optimal matrix $\mathbf{\Gamma}^{opt}$, the less the asymptotic variance of the so-obtained estimator differs from $\mathbf{B}^{opt}$.

**Remark 11**. If there is no information of the behavior of the covariance matrix $\mathbf{V}$ then we can recommend to take at the first step $c_{ki} = \lambda_{ki}(\theta_0)$ for some $\theta_0$. In this case the closer the chosen value of $\theta_0$ is to the unknown true value of the parameter $\theta$ and the closer the unknown matrix $\mathbf{V}$ is to the matrix of the form $\sigma^2 I$ for some $\sigma$, the more exact is the estimator $\theta^*$.

**6.4.** Now, we prove an auxiliary assertion which is essentially used in Examples 4 and 5.

**Lemma 4**. *Suppose that* (6.1) *and* (6.6) *are satisfied and* $\mathbf{\Gamma}^{opt}$ *is the matrix with entries defined by* (6.7). *Then there is a matrix* $\mathbf{R}$ *such that*

$$\mathbf{\Gamma}^{opt} = \mathbf{R}\mathbf{\Lambda}\mathbf{V}^{-1}.$$

*Proof.* Let

$$(\mathbf{R})_{mj} = \begin{cases} \theta_j \sigma^2, & j < m, \\ (c_0 + \theta_m)\sigma^2, & j = m, \end{cases} \quad \text{and} \quad (\mathbf{R})_{kj} = \begin{cases} -\sigma^2, & k = j < m, \\ 0, & m \neq k \neq j. \end{cases}$$

Observe that the so-defined matrix $\mathbf{R}$ is nondegenerate. Since

$$\lambda_{ji} = \begin{cases} -b_{ji} g_i(\theta), & j < m, \\ a_{mi} - b_{mi} g_i(\theta), & j = m; \end{cases}$$

for $k < m$ we have

$$(\mathbf{R}\mathbf{\Lambda})_{ki} = \sum_{j=1}^{m} (\mathbf{R})_{kj} (\mathbf{\Lambda})_{ji} = b_{ki} g_i(\theta) \sigma^2$$

and for $k = m$ we have

$$\begin{aligned} (\mathbf{R}\mathbf{\Lambda})_{mi} &= -\sum_{j=1}^{m-1} b_{ji}\theta_j g_i(\theta)\sigma^2 + (c_0 + \theta_m)(a_{mi} - b_{mi} g_i(\theta))\sigma^2 \\ &= g_i(\theta) \left( -\sum_{j=1}^{m} b_{ji}\theta_j + \beta_i(\theta) - c_0 b_{mi} \right) \sigma^2 = (1 - c_0 b_{mi}) g_i(\theta) \sigma^2. \end{aligned}$$

Hence, we obtain the sought assertion:

$$(\mathbf{R}\mathbf{\Lambda}\mathbf{V}^{-1})_{ki} = \begin{cases} b_{ki} g_i(\theta) w_i(\theta), & k < m, \\ (1 - c_0 b_{mi}) g_i(\theta) w_i(\theta), & k = m. \end{cases}$$

## 7. Consequences for Independent Observations

**7.1.** Suppose that $\{\xi_i\}$ are independent errors and

$$\forall i \ \mathbf{E}\xi_i = 0, \ 0 < \mathbf{D}\xi_i = \sigma_i^2 < \infty. \tag{7.1}$$

In this case

$$\forall i \ \eta_i = \beta_i(\theta)\xi_i, \ \mathbf{D}\eta_i = \beta_i^2(\theta)\sigma^2, \ \mathbf{V} = \operatorname{diag}\{\mathbf{D}\eta_1, \ldots, \mathbf{D}\eta_N\}. \tag{7.2}$$

We also emphasize that throughout this section we assume that all assumptions of (5.2) and (5.3) are satisfied, and we use the notation $\mathbf{B}_C$ of Corollary 5.

The following assertion is useful in checking consistency:

**Theorem 5**. *If the measurement errors are independent and satisfy* (7.1) *and if the condition*

$$\max_{j\le m}(\mathbf{B}_C)_{jj}\to 0 \tag{7.3}$$

*is fulfiled then the estimator* $\theta^*$ *is consistent.*

**7.2.** Now, suppose that the errors $\{\xi_i\}$ are representable as

$$\forall i\ \xi_i=\sigma_i\varepsilon_i,\ \mathbf{E}\varepsilon_i=0,\ \mathbf{D}\varepsilon_i=1, \tag{7.4}$$

where $\varepsilon_1,\dots,\varepsilon_N$ is a sequence of independent identically distributed random variables.

The following theorem is useful in checking asymptotic normality:

**Theorem 6**. *Suppose that the independent measurement errors are representable in the form* (7.4) *and satisfy* (7.3)*; moreover,*

$$\max_{i\le N}(\mathbf{V}^{1/2}\mathbf{C}^\top(\mathbf{C}\mathbf{V}\mathbf{C}^\top)^{-1}\mathbf{C}\mathbf{V}^{1/2})_{ii}\to 0. \tag{7.5}$$

*Then the convergence*

$$\mathbf{U}(\theta^*-\theta)\Longrightarrow\Phi_m(0,\mathbf{I})$$

*holds for every nonrandom matrix* $\mathbf{U}$ *satisfying the condition*

$$\mathbf{U}^\top\mathbf{U}=\mathbf{B}_C^{-1}. \tag{7.6}$$

Thus, under rather simple conditions (7.3) and (7.5), the estimator $\theta^*$ is consistent and asymptotically normal.

**Remark 12**. At first sight, the simplest choice in (7.6) is $\mathbf{U}=\mathbf{B}_C^{-1/2}$. However, it is sometimes easier to use the standard orthogonalization procedure and find a triangle matrix $\mathbf{U}$ satisfying (7.6).

**7.3. Remark 13**. We emphasize that the difference $(\theta^*-\theta)$ in Theorems 2 and 3 and their corollaries is normalized by the matrices $\mathbf{A}$, $\mathbf{A}_\Gamma$ and $\mathbf{B}_C$, $\mathbf{B}_\Gamma$, depending essentially on the unknown parameter $\theta$, and the matrix $\mathbf{V}$ which may be unknown as well. For this reason, the assertions in which these matrices are replaced with known ones may be useful in construction of confidence intervals and test of hypotheses.

We indicate one possible approach to solution of these problems in the case of independent observations. Put

$$\eta_i^*=\beta_i(\theta^*)Z_i-\alpha_i(\theta^*),\quad \mathbf{V}^*=\mathrm{diag}\{\eta_1^{*2},\dots,\eta_N^{*2}\},$$

$$\mathbf{A}^*=(\mathbf{C}\mathbf{V}^*\mathbf{C}^\top)^{-1/2}\mathbf{C}\mathbf{X}^\top\quad\text{and}\quad\mathbf{A}^{**}=(\mathbf{\Gamma}(\theta^*)\mathbf{V}^*\mathbf{\Gamma}^\top(\theta^*))^{-1/2}(\mathbf{\Gamma}(\theta^*)\mathbf{X}^\top.$$

By analogy with [1], we should expect the following convergences under rather stringent additional assumptions:

$$\mathbf{A}^*(\theta^*-\theta)\Longrightarrow\Phi_m(0,\mathbf{I})\quad\text{and}\quad\mathbf{A}^{**}(\theta^{**}-\theta)\Longrightarrow\Phi_m(0,\mathbf{I}).$$

**7.4.** We turn to proving the claims. First we prove several auxiliary lemmas.

**Lemma 5**. *Suppose that the independent measurement errors satisfy* (7.1) *and* (7.3)*. Then* (3.2)*,* (3.3)*, and* (3.9) *hold.*

*Proof.* Comparing the notations of (3.1), (5.1), and (5.4), we conclude that $\mathbf{B}_C = \mathbf{A}^{-1}\mathbf{A}^{-1\top} = \mathbf{G}\mathbf{V}\mathbf{G}^\top$. Using this and (7.2), we find that (7.3) is equivalent to the following:

$$(\mathbf{B}_C)_{jj} = \sum_{l=1}^{m} (\mathbf{A}^{-1})^2_{jl} = \sum_{i=1}^{N} (\mathbf{G})^2_{ji} \mathbf{D}\eta_i^2 \to 0, \quad j = 1, \dots, m. \tag{7.7}$$

Thus,

$$\mathbf{D}(\mathbf{G}\eta)_j = \mathbf{D}\left(\sum_{i=1}^{N} (\mathbf{G})_{ji}\eta_i\right) = \sum_{i=1}^{N} (\mathbf{G})^2_{ji}\mathbf{D}\eta_i = (\mathbf{B}_C)_{jj} \to 0. \tag{7.8}$$

Now, by (1.3) the numbers $b_{ji}$ are nonnegative and so

$$\mathbf{D}b_{qi}\xi_i = b^2_{qi}\sigma_i^2 \le \beta_i^2(\theta)\sigma_i^2/\theta_q^2 = \mathbf{D}\eta_i/\theta_q^2. \tag{7.9}$$

In particular, from (7.9) we obtain

$$\mathbf{D}(\mathbf{G}\mathbf{\Psi}^\top)_{jk} = \mathbf{D}\left(\sum_{i=1}^{N} (\mathbf{G})_{ji} b_{ki}\xi_i\right) \le \sum_{i=1}^{N} (\mathbf{G})^2_{ji}\mathbf{D}\eta_i/\theta_k^2 \to 0. \tag{7.10}$$

Applying (7.9) again, we find that

$$\begin{aligned} \mathbf{D}(\mathbf{U}\mathbf{C}\mathbf{\Psi}^\top\mathbf{A}^{-1})_{jk} &= \mathbf{D}\left(\sum_{i=1}^{N}\sum_{q=1}^{m} (\mathbf{U}\mathbf{C})_{ji} b_{qi}\xi_i (\mathbf{A}^{-1})_{qk}\right) \\ = \sum_{i=1}^{N}\left(\sum_{q=1}^{m} (\mathbf{U}\mathbf{C})_{ji} b_{qi} (\mathbf{A}^{-1})_{qk}\right)^2 \sigma_i^2 &\le a_j \sum_{q=1}^{m} (\mathbf{A}^{-1})^2_{qk}/\theta_q^2, \end{aligned} \tag{7.11}$$

where

$$a_j = \sum_{i=1}^{N} (\mathbf{U}\mathbf{C})^2_{ji}\mathbf{D}\eta_i.$$

But the last sum equals the diagonal entry of the matrix $(\mathbf{U}\mathbf{C})\mathbf{V}(\mathbf{U}\mathbf{C})^\top$. However,

$$(\mathbf{U}\mathbf{C})\mathbf{V}(\mathbf{U}\mathbf{C})^\top = \mathbf{U}\mathbf{C}\mathbf{V}\mathbf{C}^\top\mathbf{U}^\top = \mathbf{U}(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top = \mathbf{I}$$

in view of (5.2). Consequently, $a_j = 1$. From this fact, (7.7), and (7.11) we obtain

$$\mathbf{D}(\mathbf{U}\mathbf{C}\mathbf{\Psi}^\top\mathbf{A}^{-1})_{jk} \le \sum_{q=1}^{m} (\mathbf{A}^{-1})^2_{qk}/\theta_q^2 \le \sum_{q=1}^{m} (\mathbf{B}_C)_{qq}/\theta_q^2 \to 0. \tag{7.12}$$

But

$$\mathbf{E}(\mathbf{G}\eta)_j = \mathbf{E}(\mathbf{G}\mathbf{\Psi}^\top)_{jk} = \mathbf{E}(\mathbf{U}\mathbf{C}\mathbf{\Psi}^\top\mathbf{A}^{-1})_{jk} = 0.$$

Therefore, by Chebyshev's inequality, from (7.8), (7.10), and (7.12) we obtain the sought convergences (3.3), (3.2), and (3.9).

Theorem 5 is an obvious particular case of Lemma 5. To validate Theorem 6, we only have to verify condition (3.4) of Theorem 2. We do this in the following lemma:

**Lemma 6**. *Suppose that the independent measurement errors are representable in the form* (7.4) *and satisfy* (7.5). *Then* (3.4) *is satisfied.*

*Proof.* Put

$$\widetilde{\mathbf{C}} = \mathbf{U}\mathbf{C}\mathbf{V}^{1/2}, \quad (\widetilde{\mathbf{C}})_{ij} = \tilde{c}_{ij}, \quad \varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^\top. \tag{7.13}$$

In this case it follows from (5.2) that

$$\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}^\top = \mathbf{I}, \quad \text{i.e.,} \quad \sum_{i=1}^N \tilde{c}_{ui}\tilde{c}_{vi} = \begin{cases} 0, & u \neq v, \\ 1, & u = v. \end{cases} \tag{7.14}$$

By (7.13), we can rewrite (3.4) as

$$\mathbf{U}^{-1}\mathbf{C}\eta = \widetilde{\mathbf{C}}\varepsilon \Longrightarrow \Phi_m(0, \mathbf{I}). \tag{7.15}$$

Now, according to the Cramér–Wald theorem (see, for instance, [5]), the random vectors $\zeta_n = (\zeta_{n1}, \dots, \zeta_{nm})$ in $\mathbf{R}^m$ converge in distribution to $\zeta = (\zeta_1, \dots, \zeta_m)$ if and only if each linear combination of the components $\zeta_n$ converges in distribution to the corresponding linear combination of the components of $\zeta$. Therefore, (7.15) is equivalent to convergence of the distributions of the sums $\sum_{j=1}^m t_j \sum_{i=1}^N \widetilde{\mathbf{C}}_{ji}\varepsilon_i$ to $\Phi_1(0, \sum_{j=1}^m t_j^2)$ at each point $t = (t_1, \dots, t_m)$ of $\mathbf{R}^m$. By (7.14), we can use Lemma 1 of [1]. It suffices to demonstrate that

$$\max_{i \le N} \left( \sum_{j=1}^m t_j (\widetilde{\mathbf{C}})_{ji} \right)^2 \le \sum_{j=1}^m t_j^2 \max_{i \le N} \sum_{j=1}^m (\widetilde{\mathbf{C}})_{ji}^2 \to 0. \tag{7.16}$$

However, recalling (5.2), we easily see that the last sum in (7.16) is representable as

$$(\widetilde{\mathbf{C}}^\top \widetilde{\mathbf{C}})_{ii} = (\mathbf{V}^{1/2}\mathbf{C}^\top \mathbf{U}^\top \mathbf{U}\mathbf{C}\mathbf{V}^{1/2})_{ii} = (\mathbf{V}^{1/2}\mathbf{C}^\top (\mathbf{C}\mathbf{V}\mathbf{C}^\top)^{-1}\mathbf{C}\mathbf{V}^{1/2})_{ii}.$$

Consequently, the sought convergence (7.16) ensues from (7.5). □

## 8. Concluding Remarks

**Remark 14**. It is worth to observe that many conditions in the statements of this article are mostly of a methodological and conceptual character and, in the authors' opinion, such manner of exposing the results seems clearest, since superfluous details may obscure the basic ideas. For instance, the conditions of Theorems 1 and 2 are the multidimensional LNL and CLT for special schemes of series constructed for a sequence of independently identically distributed random variables. Using the Cramér–Wald method, in each concrete case we can reduce the verification of conditions (3.2)–(3.4) and (3.6) to validation of the one-dimensional LNL and CLT (in the case of independent observations we do this in § 7). At first sight, it seems most difficult to verify the conditions of Theorem 3 which, even on assuming (6.1), take the form

$$\sum_{i=1}^N (f_{i,N}(\theta^*) - f_{i,N}(\theta))\xi_i \xrightarrow{p} 0$$

if we use the Cramér–Wald method. However, the technics of verification of such conditions was developed by the authors (see [1, Lemmas 5–12]).

**Remark 15**. We emphasize that all assertions of the article remain valid in the situation when the observable random variables constitute a scheme of series, i.e., the following representation is valid for the $i$th observation:

$$Z_i^{(N)} = \frac{a_{0i}^{(N)} + \sum\limits_{j=1}^m a_{ji}^{(N)}\theta_j}{1 + \sum\limits_{j=1}^m b_{ji}^{(N)}\theta_j} + \xi_i^{(N)},$$

where the upper index indicates the dependence of $\{Z_i\}$, $\{a_{ji}\}$, $\{b_{ji}\}$, and $\{\xi_i\}$ and the entries of the covariance matrix $\mathbf{V}$ on the number $N$ of observations. (The form of some conditions was specially chosen for this remark to hold.) There is one exception: in Theorem 6 we have to require that the distributions of the random variables $\varepsilon_i^{(N)}$ be independent of $i$ and $N$.

**Remark 16**. Throughout the article, $\theta$ is an unknown parameter. Moreover, the entries of the covariance matrix of errors may be unknown too. Thus, most conditions in all assertions of the article represent constraints on quantities which contain unknown parameters. It is clear that for practical application of these assertions we must verify all such conditions *for all values of all unknown parameters* (as, for instance, in [4]).

## References


[1] Linke Yu. Yu. and Sakhanenko A. I. *Asymptotically normal estimation of a parameter in a linear-fractional regression problem.* Sibirsk. Mat. Zh., 41, is.1 (2000), 150–163.

[2] Demidenko E. Z. *Linear and Nonlinear Regression.* Finansy i Statistika, Moscow, (in Russian), 1981.

[3] Loève M. *Probability Theory.* Izdat. Inostr. Lit., Moscow, (Russian translation), 1962.

[4] Borovkov A. A. *Mathematical Statistics. Estimation of Parameters. Test of Hypotheses.* Nauka, Moscow, (in Russian), 1984.

[5] Billingsley P. P. *Convergence of Probability Measures.* Nauka, Moscow, (Russian translation), 1977.