Теорія Ймовір. та Матем. Статист.
Вип. 73, 2005

Theor. Probability and Math. Statist.
No. 73, 2006, Pages 163–179
S 0094-9000(07)00690-4
Article electronically published on January 17, 2007

INCONSISTENCY OF THE ORTHOGONAL REGRESSION ESTIMATOR FOR THE VECTOR NONLINEAR ERRORS-IN-VARIABLES MODEL

UDC 519.21

G. S. REPETATS'KA

Abstract. We prove that the orthogonal regression estimator for the vector errors-in-variables model is inconsistent and study its asymptotic deviation from the true value of the parameter. We also propose another estimator whose deviation from the true value is smaller than that for the orthogonal regression estimator.

1. Introduction

Consider a nonlinear errors-in-variables model

$$y_i = g(\xi_i, \beta_0) + \delta_i,$$
$$x_i = \xi_i + \varepsilon_i, \qquad i = 1, \dots, n,$$

where $(x_i, y_i) \in \mathbb{R}^q \times \mathbb{R}^s$ are known observations, ξ_i are unknown nonrandom constants, $\{(\delta_i^{\mathrm{T}}, \varepsilon_i^{\mathrm{T}})^{\mathrm{T}}, i \geq 1\}$ are independent identically distributed errors, and β_0 is the parameter to be estimated. All the parameters are vector-valued.

We study the orthogonal regression estimator and its asymptotic properties as $n \to \infty$. This estimator, widely used in the literature and in practice, is consistent for the linear model. Moreover it coincides with the maximum likelihood estimator if the errors have the normal distribution. The linear model is analyzed in the book [1]. The estimator is consistent for the nonlinear model under additional restrictions if ξ_i are known to have some specific properties. This holds, for example, in the case of repeated observations [2] or in the case where the errors tend to zero [1].

In the general case, if the regression function is nonlinear in ξ, the estimator is inconsistent and is isolated from the true value even if the errors are small (and fixed). Section 2 contains the proof of the inconsistency of the estimator. We obtain the principal term of the asymptotic deviation between the estimator and the true value in Section 3. We propose a new estimator in Section 4; the order of the asymptotic deviation between this estimator and the true value is smaller if the errors are small.

The current paper continues the investigations of [4] where the asymptotic deviation is studied for the scalar model, as well as those of [3] where the inconsistency is proved for the model with scalar variables y_i and vector variables ξ_i. We extend the latter result to the general vector model and study the asymptotic deviation for this model. We also obtain a new estimate for the variance constructed with the help of a goal function.

2000 *Mathematics Subject Classification.* Primary 62J02; Secondary 62F12, 62H12.
Key words and phrases. Orthogonal regression estimators, nonlinear errors-in-variables model.

2. The inconsistency of the orthogonal regression estimator

By I_n and O we denote the unit $n \times n$ matrix and the zero matrix of an arbitrary order, respectively. The symbol $\|\cdot\|$ stands for the Euclidean norm, $U_r(x)$ means the r-neighborhood of a point x. Vectors and the regression function are understood as column vectors throughout the paper. On the other hand, the derivatives are understood as row vectors. The superscript denotes the (vector) variable used to evaluate the derivative. For example, g^β is the Jacobi matrix for the vector function g with respect to the vector argument β. In what follows we also use the derivatives $g^{\xi\xi}$, $g^{\beta\xi}$, $g^{\xi\xi\xi}$ of vector-valued functions g with respect to vector arguments, where $g^{\xi\xi}$, $g^{\beta\xi}$, and $g^{\xi\xi\xi}$ are a symmetric bilinear operator assuming values in $\mathbb{R}^s$, a bilinear operator assuming values in $\mathbb{R}^s$, and a symmetric trilinear operator, respectively. The mathematical expectation and the variance of a random variable ζ are denoted by $\mathsf{E}\,\zeta$ and $\operatorname{Var}\zeta$, respectively.

Now we introduce some notation for sequences of random vectors depending on parameters $\beta \in \Theta$ and $\sigma > 0$.

Definition 2.1. We write $\eta_n(\beta,\sigma) = O_P(1)$ if

$$\lim_{c\to\infty} \sup_{n\ge 1} \sup_{\beta\in\Theta} \sup_{\sigma>0} \mathsf{P}\left(\|\eta_n(\beta,\sigma)\| > c\right) = 0,$$

where $O_P(1)$ is a stochastically and uniformly bounded sequence of random vectors.

Definition 2.2. We write $\eta_n(\beta) = o_P(1)$ if $\eta_n(\beta) \to 0$ in probability for all $\beta \in \Theta$.

Definition 2.3. We write $\eta_n(\beta,\sigma) = o_{\sigma P}(1)$ if

$$\lim_{\sigma\to 0+} \sup_{n\ge 1} \mathsf{P}\left(\sup_{\beta\in\Theta} \|\eta_n(\beta,\sigma)\| > c\right) = 0$$

for all $c > 0$.

Below we list the assumptions imposed on the regression model.

(i) $\beta_0 \in \operatorname{int}\Theta$ where $\Theta \subset \mathbb{R}^p$ is a compact set.

(ii) $\|\xi_i\| \le a$, $i \ge 1$, where a is an unknown positive number.

(iii) $g\colon \mathbb{R}^q \times U \to \mathbb{R}^s$ is a three times continuously differentiable function, where U is an open set, $U \supset \Theta$.

(iv) $[\delta_i^{\mathrm T}; \varepsilon_i^{\mathrm T}]^{\mathrm T} \sim N(\vec{0}; \sigma^2\Gamma)$ where Γ is a known positive definite matrix, and $\sigma > 0$ is an unknown number. We allow the errors ε_i and δ_i to be correlated. Without loss of generality assume that $\Gamma = \begin{pmatrix} I_s & S \\ S^{\mathrm T} & I_q \end{pmatrix}$, where $S \in \mathbb{R}^{s\times q}$.

(v) $\underline{\lim}_{n\to\infty} \|k_n\| > 0$, where

$$k_n = \frac{1}{n}\sum_{i=1}^n \overrightarrow{\operatorname{tr}}^{\mathrm T}\left(g^{\xi\xi}H^{-1}\right)\left[I_s - SS^{\mathrm T} + (g^\xi - S)(g^\xi - S)^{\mathrm T}\right]^{-1} g^\beta\Big|_{(\xi_i,\beta_0)},$$

$$H = \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm T} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}, \qquad \overrightarrow{\operatorname{tr}}\left(g^{\xi\xi}H^{-1}\right) = \left(\operatorname{tr}(g_1^{\xi\xi}H^{-1}), \dots, \operatorname{tr}(g_s^{\xi\xi}H^{-1})\right)^{\mathrm T}.$$

The orthogonal regression estimator is defined by

$$\hat\beta \in \arg\min_{\beta\in\Theta} \frac{1}{n}\sum_{i=1}^n \min_{u\in\mathbb{R}^q} \begin{pmatrix} y_i - g(u,\beta) \\ x_i - u \end{pmatrix}^{\mathrm T} \Gamma^{-1} \begin{pmatrix} y_i - g(u,\beta) \\ x_i - u \end{pmatrix}. \tag{1}$$

The matrix Γ^{-1} has the same block structure as Γ. Its blocks are denoted by

$$\Gamma^{-1} =: \begin{pmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{pmatrix},$$

where $V^{11} \in \mathbb{R}^{s\times s}$, $V^{22} \in \mathbb{R}^{q\times q}$, and $(V^{12})^{\mathrm{T}} = V^{21}$. We use these superscripts throughout the paper to denote the corresponding blocks of $(s+q)\times(s+q)$ block matrices. It is easy to check that

$$\Gamma^{-1} = \begin{pmatrix} I_s & S \\ S^{\mathrm{T}} & I_q \end{pmatrix}^{-1} = \begin{pmatrix} V & -VS \\ -S^{\mathrm{T}}V & I_q + S^{\mathrm{T}}VS \end{pmatrix},$$

where $V := \left(I_s - SS^{\mathrm{T}}\right)^{-1}$. Note that the matrix Γ is positive definite if and only if V is positive definite.

Consider the function

$$G(x,y,\beta,u) := \begin{pmatrix} y - g(u,\beta) \\ x-u \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} y - g(u,\beta) \\ x-u \end{pmatrix},$$

where $x, u \in \mathbb{R}^q$, $y \in \mathbb{R}^s$, and $\beta \in \Theta$.

We prove the existence and uniqueness of the minimum of the function G in u under conditions (i)–(iii).

1) *Existence*. The minimum exists since the function G is continuous in u and therefore

$$\lim_{\|u\|\to\infty} G(x,y,\beta,u) = +\infty.$$

Let $h(x,y,\beta)$ be one of the minimum points of the function $G(x,y,\beta,u)$ in u.

2) *Uniqueness*. The function $G(x,y,\beta,u)$ is differentiable with respect to u at the minimum point, whence $G^u\big|_{u=h(x,y,\beta)} = \vec{0}$. This means that $h(x,y,\beta)$ is defined implicitly by

$$F(x,y,\beta,h) := -\frac{1}{2}G^{u\mathrm{T}}(x,y,\beta,u)\big|_{u=h} = \begin{pmatrix} g^{\xi}(h,\beta) \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} y - g(h,\beta) \\ x-h \end{pmatrix} = \vec{0}. \tag{2}$$

Lemma 2.1. *The function $h(x,y,\beta)$ is uniquely defined by equation (2) and is twice continuously differentiable in a neighborhood of the points $(\xi, g(\xi,\beta),\beta)$. Moreover one can choose a common diameter of the neighborhood for all ξ and β such that $\|\xi\| \le a$ and $\beta \in \Theta$, that is,*

$$h\colon U_{\nu_0}(\xi)\times U_{\nu_0}(g(\xi,\beta))\times U_{\nu_0}(\beta) \to \mathbb{R}^q.$$

Proof. We apply the implicit function theorem. It is known that

$$F(\xi, g(\xi,\beta),\beta,\xi) = \vec{0}.$$

Moreover

$$\begin{aligned} F^u(x,y,\beta,u) = {} & \begin{pmatrix} g^{\xi}(u,\beta) \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^{\xi}(u,\beta) \\ I_q \end{pmatrix} \\ & + \left(g^{\xi\xi}(u,\beta)\right)^{\mathrm{T}} \left(V^{11}(y - g(u,\beta)) + V^{12}(x-u)\right) \end{aligned}$$

and

$$F^u(\xi, g(\xi,\beta),\beta,\xi) = \begin{pmatrix} g^{\xi}(\xi,\beta) \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^{\xi}(\xi,\beta) \\ I_q \end{pmatrix} = H(\xi,\beta)$$

is a positive definite matrix. According to the implicit function theorem, there is a neighborhood $U_\nu(\xi,\beta) := U_\nu(\xi)\times U_\nu(g(\xi,\beta))\times U_\nu(\beta)$, $\nu = \nu(\xi,\beta)$, of the point $(\xi, g(\xi,\beta),\beta)$ such that $h\colon U_\nu(\xi,\beta)\to\mathbb{R}^q$ is a single-valued continuously differentiable function. Since the arguments ξ and β vary in compact sets, one can choose $\nu = \nu_0 > 0$, common for all points $\beta\in\Theta$ and ξ if $\|\xi\|\le a$. For any point of the graph

$$A := \left\{(\xi, g(\xi,\beta)) : \|\xi\| \le a, \beta\in\Theta\right\},$$

there exists a neighborhood $U_\nu(\xi,\beta)$ where h is single-valued. Note that

$$\{U_\nu(\xi,\beta)\colon \|\xi\|\le a, \beta\in\Theta\}$$

is an open covering of the compact set A and that it contains a finite subcovering. Thus there is a common ν_0 for all points of the set. The lemma is proved. □

The following result is useful when constructing the asymptotic expansions.

Lemma 2.2. *Let $\{a_i\colon i\ge 1\}$ be a bounded sequence of numbers, and let ζ_i be independent identically distributed random variables having finite second moments. If ζ has the same distribution as ζ_i, then*

$$\frac{1}{n}\sum_{i=1}^{n} a_i\zeta_i = \frac{\mathsf{E}\,\zeta}{n}\cdot\sum_{i=1}^{n} a_i + \frac{\sqrt{\operatorname{Var}\zeta}}{\sqrt{n}} O_P(1),$$

where the random variable $O_P(1)$ is stochastically and uniformly bounded in the sense of Definition 2.1 *with either* $\operatorname{Var}\zeta=\sigma^2$ *or* $\operatorname{Var}\zeta=\sigma^4$.

We consider the model for small σ^2. The errors are bounded stochastically; thus $h(x,y,\beta)$ is not always uniquely determined for large errors. To account for this phenomenon we consider the set of indices $B_n(\nu)=\{i=1,\dots,n\colon \|\varepsilon_i\|\le\nu, \|\delta_i\|\le\nu\}$, where the constant $\nu\in(0,\nu_0]$ is such that $U_\nu(\beta_0)\in\operatorname{int}\Theta$ and ν_0 is defined in Lemma 2.1. Consider the following goal function:

$$Q(\beta)=Q_n(\beta):=\frac{1}{n}\sum_{i=1}^{n} q(x_i,y_i,\beta),$$

where $q(x,y,\beta):=G(x,y,\beta,h(x,y,\beta))$. Then estimator (1) can be rewritten as follows:

$$\hat\beta\in\arg\min_{\beta\in\Theta} Q(\beta).$$

Hence

$$Q(\beta)=Q_1(\beta)+Q_2(\beta):=\frac{1}{n}\sum_{i\in B_n(\nu)} q(x_i,y_i,\beta)+\frac{1}{n}\sum_{i\notin B_n(\nu)} q(x_i,y_i,\beta),$$

where $Q_1(\beta)$ is the principal part, while $Q_2(\beta)$ is treated as a remainder.

Theorem 2.1. *Let conditions* (i)–(iv) *hold for model* (1). *Then*

$$Q(\beta)=Q_1(\beta)+\sigma^4 o_{\sigma P}(1), \tag{3}$$

$$Q_1^\beta(\beta_0)=\sigma^2 k_n+\left(\sigma^3+\frac{\sigma}{\sqrt{n}}\right)O_P(1)+\sigma^4 o_{\sigma P}(1), \tag{4}$$

where k_n is defined by assumption (v).

Proof. We will omit the subscript i for the variables x_i, y_i, ξ_i, ε_i, δ_i, and $h_i:=h(x_i,y_i,\beta_0)$, $i=1,\dots,n$.

1) We prove that $Q_2(\beta)=\sigma^4 o_{\sigma P}(1)$. Consider $q(x_i,y_i,\beta)$:

$$\begin{aligned}
q(x,y,\beta)&=\begin{pmatrix} y-g(h(x,y,\beta),\beta)\\ x-h(x,y,\beta)\end{pmatrix}^{\mathrm T}\Gamma^{-1}\begin{pmatrix} y-g(h(x,y,\beta),\beta)\\ x-h(x,y,\beta)\end{pmatrix}\\
&\le\begin{pmatrix} y-g(\xi,\beta)\\ x-\xi\end{pmatrix}^{\mathrm T}\Gamma^{-1}\begin{pmatrix} y-g(\xi,\beta)\\ x-\xi\end{pmatrix}\\
&=\begin{pmatrix} \delta+(g(\xi,\beta_0)-g(\xi,\beta))\\ \varepsilon\end{pmatrix}^{\mathrm T}\Gamma^{-1}\begin{pmatrix} \delta+(g(\xi,\beta_0)-g(\xi,\beta))\\ \varepsilon\end{pmatrix}\\
&\le C\left(\|\delta\|^2+\|\varepsilon\|^2\right)+D,
\end{aligned}$$

where C and D are universal constants. Now consider $Q_2(\beta)$:

$$Q_2(\beta) = \frac{1}{n} \sum_{i \notin B_n(\nu)} q(x_i, y_i, \beta) \le \frac{1}{n} \sum_{i \notin B_n(\nu)} \left(C \left(\|\varepsilon_i\|^2 + \|\delta_i\|^2 \right) + D \right)$$

(5)

$$\le \frac{1}{n} \sum_{i=1}^{n} \left(C \left(\|\varepsilon_i\|^2 + \|\delta_i\|^2 \right) + D \right) \cdot \left(I(\|\varepsilon_i\| \ge \nu) + I(\|\delta_i\| \ge \nu) \right).$$

Let $\tilde{\varepsilon}_i := \varepsilon_i/\sigma$, $\tilde{\delta}_i := \delta_i/\sigma$, and $[\tilde{\delta}_i^{\mathrm{T}}, \tilde{\varepsilon}_i^{\mathrm{T}}]^{\mathrm{T}} \sim N(\vec{0}, \Gamma)$. We estimate one of the terms on the right hand side of (5):

$$\mathsf{E}\, \|\varepsilon_i\|^2 I(\|\varepsilon_i\| \ge \nu) = \sigma^2\, \mathsf{E}\, \|\tilde{\varepsilon}_i\|^2 I\left(\|\tilde{\varepsilon}_i\|^2 \ge \nu^2/\sigma^2\right) = \frac{\sigma^4}{\nu^2}\, \mathsf{E}\, \frac{\nu^2}{\sigma^2} \|\tilde{\varepsilon}_i\|^2 I\left(\|\tilde{\varepsilon}_i\|^2 \ge \nu^2/\sigma^2\right)$$
$$\le \frac{\sigma^4}{\nu^2}\, \mathsf{E}\, \|\tilde{\varepsilon}_i\|^4 I\left(\|\tilde{\varepsilon}_i\|^2 \ge \nu^2/\sigma^2\right) = \sigma^4 o(1), \qquad \sigma^2 \to 0+.$$

The other terms in (5) are considered similarly. Thus

$$\mathsf{E}\, Q_2(\beta) = \sigma^4 o(1), \qquad \sigma \to 0+.$$

It remains to show that $\sigma^{-4} Q_2(\beta) = o_{\sigma P}(1)$ to complete the proof of (3).

By the Chebyshev inequality,

$$\mathsf{P}\left(\sigma^{-4} Q_2(\beta) > C\right) \le \frac{\mathsf{E}\, \sigma^{-4} Q_2(\beta)}{C} = \frac{o\,(1)}{C} \to 0, \qquad \sigma \to 0+.$$

Thus $Q_2(\beta) = \sigma^4 o_{\sigma P}(1)$.

2) We prove equality (4). We express the necessary terms by using errors ε_i, δ_i and values of the function and its derivatives at points (ξ_i, β_0). For the sake of brevity we do not write the subscript i assuming that $i \in B_n(\nu)$. Recall that Lemma 2.1 holds for $i \in B_n(\nu)$. Define Δ from the equality

$$h = h(x, y, \beta_0) = \xi + \Delta$$

and note that $\Delta = O(\|\varepsilon\| + \|\delta\|)$. Indeed,

$$\|\Delta\|^2 = \|h - \xi\|^2 \le 2\left(\|h - x\|^2 + \|x - \xi\|^2\right) \le 2\left(\|\varepsilon\|^2 + G(x, y, \beta_0, h)\right)$$
$$\le 2\left(\|\varepsilon\|^2 + G(x, y, \beta_0, \xi)\right) \le 2\left(\|\varepsilon\|^2 + \|\Gamma^{-1}\| \left(\|\varepsilon\|^2 + \|\delta\|^2\right)\right) = O\left(\|\varepsilon\|^2 + \|\delta\|^2\right).$$

Now expand the functions into a Taylor series in a neighborhood of the points (ξ, β_0):

$$g(h, \beta_0) = g(\xi, \beta_0) + g^{\xi}(\xi, \beta_0)\Delta + \frac{1}{2}\left(\Delta^{\mathrm{T}} g^{\xi\xi}(\xi, \beta_0)\Delta\right) + O(\|\Delta\|^3),$$
$$g^{\xi}(h, \beta_0) = g^{\xi}(\xi, \beta_0) + \Delta^{\mathrm{T}} g^{\xi\xi}(\xi, \beta_0) + O(\|\Delta\|^2),$$
$$g^{\beta}(h, \beta_0) = g^{\beta}(\xi, \beta_0) + \Delta^{\mathrm{T}} g^{\beta\xi}(\xi, \beta_0) + O(\|\Delta\|^2).$$

Here and in what follows the symbol $O(\cdot)$ stands for a uniformly bounded variable, that is,

$$\|O(t)\| \le C\|t\|$$

for all β, ξ, σ, i, and n. We treat the derivatives that cannot be represented in the matrix form as operators applied to every component of the vector function $g = (g_1, \dots, g_s)^{\mathrm{T}}$. For example,

$$\Delta^{\mathrm{T}} g^{\xi\xi} \Delta = \left(\Delta^{\mathrm{T}} g_1^{\xi\xi} \Delta, \dots, \Delta^{\mathrm{T}} g_s^{\xi\xi} \Delta\right)^{\mathrm{T}}.$$

We omit the argument of a function if it is evaluated at the point (ξ, β_0). Substituting the above expansions in the equation for the implicit function we get

$$\begin{pmatrix} g^{\xi} + \Delta^{\mathrm{T}} g^{\xi\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \delta - g^{\xi}\Delta - \frac{1}{2}\Delta^{\mathrm{T}} g^{\xi\xi}\Delta \\ \varepsilon - \Delta \end{pmatrix} = O(\|\varepsilon\|^3 + \|\delta\|^3). \tag{6}$$

Now we express Δ in terms of the errors ε and δ:

$$\Delta = \Delta_1 + \Delta_2 + O(\|\varepsilon\|^3 + \|\delta\|^3), \tag{7}$$

where Δ_1 is the linear part with respect to ε and δ, while Δ_2 is the quadratic part. Then

$$\Delta_1 = O(\|\varepsilon\| + \|\delta\|), \qquad \Delta_2 = O(\|\varepsilon\|^2 + \|\delta\|^2).$$

To determine Δ_1 we substitute (7) into (6), find the linear part, and equate it to zero:

$$\begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix} = 0, \qquad \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix} = \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} \Delta_1 = H\Delta_1,$$
$$\Delta_1 = H^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}.$$

To determine Δ_2, we again substitute (7) into (6), find the quadratic part, and equate it to zero:

$$\begin{pmatrix} g^\xi + \Delta_1^{\mathrm{T}} g^{\xi\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} (\delta - g^\xi \Delta_1) - \frac{1}{2}\Delta_1^{\mathrm{T}} g^{\xi\xi} \Delta_1 - g^\xi \Delta_2 \\ (\varepsilon - \Delta_1) - \Delta_2 \end{pmatrix} = 0,$$
$$\begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} -\frac{1}{2}\Delta_1^{\mathrm{T}} g^{\xi\xi} \Delta_1 - g^\xi \Delta_2 \\ -\Delta_2 \end{pmatrix} + \begin{pmatrix} \Delta_1^{\mathrm{T}} g^{\xi\xi} \\ O_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix} = 0,$$
$$\Delta_2 = H^{-1} \left[\begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} -\frac{1}{2}\Delta_1^{\mathrm{T}} g^{\xi\xi} \Delta_1 \\ O \end{pmatrix} + \begin{pmatrix} \Delta_1^{\mathrm{T}} g^{\xi\xi} \\ O_q \end{pmatrix} \Gamma^{-1} \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix} \right].$$

Having obtained Δ with an essential precision we pass to the evaluation of $q^\beta(x, y, \beta_0)$. By definition,

$$q(x, y, \beta) = G(x, y, \beta, h(x, y, \beta)),$$

whence we get that

$$\begin{aligned} q^\beta(x, y, \beta_0) &= G^\beta(x, y, \beta_0, u)\big|_{u=h} + G^u(x, y, \beta_0, u)\big|_{u=h} \cdot h^\beta(x, y, \beta_0) \\ &= G^\beta(x, y, \beta_0, u)\big|_{u=h} = -2 \begin{pmatrix} y - g(h, \beta_0) \\ x - h \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\beta(h, \beta_0) \\ O \end{pmatrix} \end{aligned}$$

by (2). Now we substitute the above expansions into the latter expression:

$$q^\beta(x, y, \beta_0) = -2 \begin{pmatrix} \delta - g^\xi \Delta - \frac{1}{2}\Delta^{\mathrm{T}} g^{\xi\xi} \Delta \\ \varepsilon - \Delta \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\beta + \Delta^{\mathrm{T}} g^{\beta\xi} \\ O \end{pmatrix} + O(\|\varepsilon\|^3 + \|\delta\|^3).$$

Using the expansion of Δ in terms of ε and δ instead of Δ, we obtain the linear part L, quadratic part W, and the remainder $O\left(\|\varepsilon\|^3 + \|\delta\|^3\right)$. We treat the quadratic part W and evaluate its mathematical expectation:

$$\begin{aligned} W &= -2 \begin{pmatrix} -g^\xi \Delta_2 - \frac{1}{2}\Delta_1^{\mathrm{T}} g^{\xi\xi} \Delta_1 \\ -\Delta_2 \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\beta \\ O \end{pmatrix} + \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \Delta_1^{\mathrm{T}} g^{\beta\xi} \\ O \end{pmatrix} \\ &= 2\Delta_2^{\mathrm{T}} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\beta \\ O \end{pmatrix} + \begin{pmatrix} \Delta_1^{\mathrm{T}} g^{\xi\xi} \Delta_1 \\ O \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\beta \\ O \end{pmatrix} \\ &\quad + \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \Delta_1^{\mathrm{T}} g^{\beta\xi} \\ O \end{pmatrix}. \end{aligned}$$

Below are some technical calculations needed to find $\mathsf{E}\,W$:

$$\mathsf{E}\,\Delta_1\Delta_1^{\mathrm{T}} = \mathsf{E}\,H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}\begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix} H^{-1}$$

$$= \sigma^2 H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix} H^{-1} = \sigma^2 H^{-1} H H^{-1} = \sigma^2 H^{-1},$$

$$\mathsf{E}\,\Delta_1\varepsilon^{\mathrm{T}} = \mathsf{E}\,H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}\varepsilon^{\mathrm{T}} = \sigma^2 H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} S \\ I_q \end{pmatrix}$$

$$= \sigma^2 H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} O \\ I_q \end{pmatrix} = \sigma^2 H^{-1},$$

$$\mathsf{E}\,\Delta_1\delta^{\mathrm{T}} = \mathsf{E}\,H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}\delta^{\mathrm{T}} = \sigma^2 H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} I_s \\ S^{\mathrm{T}} \end{pmatrix}$$

$$= \sigma^2 H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} I_s \\ O \end{pmatrix} = \sigma^2 H^{-1} g^{\xi\mathrm{T}}.$$

Next we determine $\mathsf{E}(H\Delta_2) = H\,\mathsf{E}\,\Delta_2$:

$$\mathsf{E}(H\Delta_2) = \mathsf{E}\left[\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} -\frac{1}{2}\Delta_1^{\mathrm{T}} g^{\xi\xi}\Delta_1 \\ O \end{pmatrix} + \begin{pmatrix} \Delta_1^{\mathrm{T}} g^{\xi\xi} \\ O \end{pmatrix}\Gamma^{-1}\begin{pmatrix} \delta - g^{\xi}\Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix}\right].$$

Each of the two terms is considered separately.

a)

$$\mathsf{E}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} -\frac{1}{2}\Delta_1^{\mathrm{T}} g^{\xi\xi}\Delta_1 \\ O \end{pmatrix} = -\frac{1}{2}\,\mathsf{E}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} \overrightarrow{\mathrm{tr}}(g^{\xi\xi}\Delta_1\Delta_1^{\mathrm{T}}) \\ O \end{pmatrix}$$

$$= -\frac{\sigma^2}{2}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} \overrightarrow{\mathrm{tr}}(g^{\xi\xi}H^{-1}) \\ O \end{pmatrix}.$$

b) If the column i of a matrix X is denoted by $X_{(i)}$, then $(AB)_{(i)} = A \cdot B_{(i)}$ for all matrices A and B with appropriate sizes. Thus

$$\mathsf{E}\begin{pmatrix} \Delta_1^{\mathrm{T}} g^{\xi\xi} \\ O_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} \delta - g^{\xi}\Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix} = \mathsf{E}\left(g^{\xi\xi}\Delta_1\right)^{\mathrm{T}}\left(V^{11}(\delta - g^{\xi}\Delta_1) + V^{12}(\varepsilon - \Delta_1)\right)$$

$$= \mathsf{E}\sum_{\substack{1\le i\le s \\ 1\le j\le s}} \left(g_i^{\xi\xi}\Delta_1\delta_j - \Delta_1\Delta_1^{\mathrm{T}} g_j^{\xi\mathrm{T}}\right) V_{ij}^{11} + \mathsf{E}\sum_{\substack{1\le i\le s \\ 1\le j\le q}} g_i^{\xi\xi}(\Delta_1\varepsilon_j - \Delta_1\Delta_{1j})V_{ij}^{12}$$

$$= \sigma^2\sum_{\substack{1\le i\le s \\ 1\le j\le s}} \left(g_i^{\xi\xi}\left(H^{-1}g^{\xi}\right)_{(j)} - g_i^{\xi\xi}\left(H^{-1}g^{\xi}\right)_{(j)}\right) V_{ij}^{11}$$

$$+ \sigma^2\sum_{\substack{1\le i\le s \\ 1\le j\le q}} g_i^{\xi\xi}\left(\left(H^{-1}\right)_{(j)} - \left(H^{-1}\right)_{(j)}\right) V_{ij}^{12} = \vec{0}.$$

The mathematical expectation of the second term in the expression for W can be found similarly to a), while the expectation of the third term is zero (the reasoning is similar to b)). Therefore

$$\mathsf{E}\,W = \mathsf{E}\begin{pmatrix} \Delta_1^{\mathrm{T}} g^{\xi\xi}\Delta_1 \\ O \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} g^{\beta} \\ O \end{pmatrix} + 2\,\mathsf{E}\,\Delta_2^{\mathrm{T}}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1}\begin{pmatrix} g^{\beta} \\ O \end{pmatrix}$$

$$= \sigma^2\begin{pmatrix} \overrightarrow{\mathrm{tr}}(g^{\xi\xi}H^{-1}) \\ O \end{pmatrix}^{\mathrm{T}}\left\{\Gamma^{-1} - \Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix} H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\right\}\begin{pmatrix} g^{\beta} \\ O \end{pmatrix}.$$

Further, we find the left $s \times s$ block of the matrix in the braces:

$$\Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix} = \begin{pmatrix} V & -VS \\ -S^{\mathrm{T}}V & I_q + S^{\mathrm{T}}VS \end{pmatrix}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix} = \begin{pmatrix} V\left(g^{\xi} - S\right) \\ I_q - S^{\mathrm{T}}\left(g^{\xi} - S\right) \end{pmatrix},$$

$$\Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix} H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}}\Gamma^{-1} = \begin{pmatrix} V\left(g^{\xi} - S\right) \\ I_q - S^{\mathrm{T}}\left(g^{\xi} - S\right) \end{pmatrix} H^{-1}\begin{pmatrix} V\left(g^{\xi} - S\right) \\ I_q - S^{\mathrm{T}}\left(g^{\xi} - S\right) \end{pmatrix}^{\mathrm{T}}.$$

The left upper block of the same matrix can be found as follows. If A, B, C, D, and N are matrices with agreed dimensions, then

$$\begin{bmatrix} A \\ B \end{bmatrix} N \begin{bmatrix} C & D \end{bmatrix} = \begin{bmatrix} ANC & AND \\ BNC & BND \end{bmatrix}.$$

This implies that

$$\left(\Gamma^{-1} - \Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix} H^{-1}\begin{pmatrix} g^{\xi} \\ I_q \end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\right)^{11} = V - V\left(g^{\xi} - S\right)H^{-1}\left(g^{\xi} - S\right)^{\mathrm{T}}V.$$

This matrix is such that

$$V - V\left(g^{\xi} - S\right)H^{-1}\left(g^{\xi} - S\right)^{\mathrm{T}}V = \left(I_s - SS^{\mathrm{T}} + (g^{\xi} - S)(g^{\xi} - S)^{\mathrm{T}}\right)^{-1}. \tag{8}$$

Indeed, let $A = V^{1/2}(g^{\xi} - S)$. Then $H = I_q + A^{\mathrm{T}}A$ and it is easy to check that

$$I_s - A(I_q + A^{\mathrm{T}}A)^{-1}A = (I_s + AA^{\mathrm{T}})^{-1}.$$

Then the left hand side of (8) equals

$$\begin{aligned} &V^{1/2}\left(I_s - A(I_q + A^{\mathrm{T}}A)^{-1}A\right)V^{1/2} = V^{1/2}\left(I_s + AA^{\mathrm{T}}\right)^{-1}V^{1/2} \\ &\quad = \left(V^{-1} + V^{-1/2}AA^{\mathrm{T}}V^{-1/2}\right)^{-1} = \left(I_s - SS^{\mathrm{T}} + (g^{\xi} - S)(g^{\xi} - S)^{\mathrm{T}}\right)^{-1}. \end{aligned}$$

Hence $\mathsf{E}\,W = \sigma^2 \cdot \overrightarrow{\mathrm{tr}}^{\mathrm{T}}\left(g^{\xi\xi}H^{-1}\right) \cdot \left[I_s - SS^{\mathrm{T}} + (g^{\xi} - S)(g^{\xi} - S)^{\mathrm{T}}\right]^{-1} g^{\beta}$. Now we consider the term

$$Q_1^{\beta}(\beta_0) = \frac{1}{n}\sum_{i\in B_n(\nu)} q(x_i, y_i, \beta_0) = \frac{1}{n}\sum_{i\in B_n(\nu)} (L_i + W_i) + O\left(\|\varepsilon_i\|^3 + \|\delta_i\|^3\right) = S_1 - S_2 + R,$$

where

$$S_1 = \frac{1}{n}\sum_{i=1}^{n} L_i + \frac{1}{n}\sum_{i=1}^{n} W_i, \qquad S_2 = \frac{1}{n}\sum_{i\notin B_n(\nu)} (L_i + W_i),$$

$$R = \frac{1}{n}\sum_{i\in B_n(\nu)} O\left(\|\varepsilon_i\|^3 + \|\delta_i\|^3\right).$$

First we treat S_1. Since $\mathsf{E}\,L_i = 0$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} L_i = \frac{\sigma}{\sqrt{n}}O_P(1), \qquad \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\,W_i + \frac{\sigma^2}{\sqrt{n}}O_P(1) = \sigma^2 k_n + \frac{\sigma}{\sqrt{n}}O_P(1),$$

$$S_1 = \sigma^2 k_n + \frac{\sigma}{\sqrt{n}}O_P(1)$$

by Lemma 2.2.

Similarly to 1) we prove that $S_2 = \sigma^4 o_{\sigma P}(1)$. Next we consider the remainder R:

$$\|R\| = \frac{1}{n}\sum_{i\in B_n(\nu)} O\left(\|\varepsilon_i\|^3 + \|\delta_i\|^3\right) \le \text{const}\,\frac{1}{n}\sum_{i\in B_n(\nu)} \left(\|\varepsilon_i\|^3 + \|\delta_i\|^3\right)$$
$$\le \text{const}\,\frac{1}{n}\sum_{i=1}^{n} \left(\|\varepsilon_i\|^3 + \|\delta_i\|^3\right) = \text{const}\,\sigma^3\frac{1}{n}\sum_{i=1}^{n}\left(\|\tilde\varepsilon_i\|^3 + \|\tilde\delta_i\|^3\right) = \sigma^3 O_P(1).$$

Combining the latter relations we complete the proof of (4). □

Theorem 2.2. *Let conditions* (i)–(v) *hold for model* (1)*. Then for an arbitrary* $\varepsilon > 0$ *there are* $\tau > 0$ *and* $\sigma_\varepsilon > 0$ *such that*

$$\varliminf_{n\to\infty} \mathsf{P}\left(\|\hat\beta_n - \beta_0\| > \sigma^2\tau\right) > 1 - \varepsilon$$

for all $\sigma \in (0, \sigma_\varepsilon]$.

Proof. 1) The function $Q_1^{\beta\beta}(\beta)$, $\beta \in U_\nu(\beta_0)$, is bounded, since Lemma 2.1 holds for $\beta \in U_\nu(\beta)$ and all functions involved in the expression for $q(x_i, y_i, \beta)$, $i \in B_n(\nu)$, are continuous and bounded.

2) *Representation for* $Q(\beta)$*.* For $\beta \in U_\nu(\beta_0)$ and $\Delta\beta = \beta - \beta_0$, we write the Taylor expansion of the function $Q_1(\beta)$:

$$Q_1(\beta) = Q_1(\beta_0) + Q_1^\beta(\beta_0)\Delta\beta + \frac{1}{2}\Delta\beta^{\mathrm T} Q_1^{\beta\beta}(\bar\beta)\Delta\beta, \quad \text{where } \bar\beta \in [\beta_0, \beta].$$

By Theorem 2.1,

$$Q(\beta) = Q(\beta_0) + \left(\sigma^2 k_n + \sigma^3 O_P(1) + \sigma n^{-1/2} O_P(1)\right)\Delta\beta + O(1)\|\Delta\beta\|^2 + \sigma^4 o_{\sigma P}(1).$$

Let $\Delta\beta = \sigma^2\Delta\varphi$. The latter equality implies that

$$\frac{Q(\beta) - Q(\beta_0)}{\sigma^4} = \left(k_n + \sigma O_P(1) + \frac{1}{\sigma\sqrt{n}} O_P(1)\right)\Delta\varphi + O(1)\|\Delta\varphi\|^2 + o_{\sigma P}(1). \tag{9}$$

Let $\hat\beta - \beta_0 = \Delta\hat\beta = \sigma^2\Delta\hat\varphi$. Then

$$\frac{Q(\hat\beta) - Q(\beta_0)}{\sigma^4} = \left(k_n + \sigma O_P(1) + \frac{1}{\sigma\sqrt{n}} O_P(1)\right)\Delta\hat\varphi + O(1)\|\Delta\hat\varphi\|^2 + o_{\sigma P}(1).$$

Put $\Delta\varphi = -tk_n$, where $t > 0$ is a sufficiently small number. Since k_n is bounded, $\beta_t = \beta_0 - tk_n \in U_\nu(\beta_0) \subset \Theta$ for small t. Thus (9) implies

$$\frac{Q(\beta_t) - Q(\beta_0)}{\sigma^4} = -\|k_n\|^2 t + \frac{1}{\sigma\sqrt{n}} O_P(1)\|k_n\| t + O(1)\|k_n\|^2 t^2 + o_{\sigma P}(1).$$

By the definition of $\hat\beta$,

$$0 \ge \frac{Q(\hat\beta) - Q(\beta_t)}{\sigma^4} = p\left(\Delta\hat\varphi\right) + R_1(t) + R_2(n) + R_3(\sigma),$$

where $p\left(\Delta\hat\varphi\right) = k_n\Delta\hat\varphi + \sigma O_P(1)\|\Delta\hat\varphi\| + O(1)\|\Delta\hat\varphi\|^2$,

$$R_1(t) = \|k_n\|^2 t + O(1)\|k_n\|^2 t^2, \qquad R_2(n) = \frac{1}{\sigma\sqrt{n}} O_P(1)\left(\|\Delta\hat\varphi\| + \|k_n\| t\right),$$
$$R_3(\sigma) = o_{\sigma P}(1).$$

Fix $\varepsilon > 0$. Assumption (v) allows one to choose a sufficiently small number t such that

$$R_1(t) \ge t_0 > 0$$

for $n \geq n_0$. For an arbitrary σ, there exists $n_1 = n_1(\varepsilon, \sigma)$ such that

$$\mathsf{P}\left(|R_2(n)| < t_0/4\right) > 1 - \varepsilon/2$$

for all $n \geq n_1$. There exists $\sigma_\varepsilon > 0$ such that, for an arbitrary $\sigma \in (0, \sigma_\varepsilon]$, we have

$$\mathsf{P}\left(|R_3(\sigma)| < t_0/4\right) > 1 - \varepsilon/2$$

starting with some number $n_1 = n_1(\varepsilon, \sigma)$. Therefore the probability that

$$p\left(\Delta\hat{\varphi}\right) \leq -t_0/2$$

is not smaller than $1 - \varepsilon$ for all $\sigma \in (0, \sigma_\varepsilon]$ and all $n \geq N(\varepsilon, \sigma) := \max(n_0, n_1)$. This means that the probability that $\Delta\hat{\varphi}$ is isolated from zero is not smaller than $1 - \varepsilon$. In other words,

$$\mathsf{P}\left(\|\Delta\hat{\varphi}\| > \tau\right) \geq 1 - \varepsilon$$

for some $\tau > 0$. This completes the proof of Theorem 2.2. □

3. Asymptotic deviation

By Theorem 2.2, the estimator $\hat{\beta}$ is inconsistent and it is isolated from the true value β_0 if errors are small. Below we study the order of deviation between the estimator and the true value of the parameter with respect to σ and construct an estimator with a smaller deviation than $\hat{\beta}$ for small σ.

In what follows we need the following notation used for a sequence of random variables.

Definition 3.1. We write $\eta_n(\sigma) = \tilde{O}_{\sigma P}(1)$ if for an arbitrary $\varepsilon > 0$ there exists $C > 0$ such that

$$\lim_{\sigma \to 0+} \varlimsup_{n \to \infty} \mathsf{P}\left(\|\eta_n(\sigma)\| > C\right) < \varepsilon.$$

Definition 3.2. We write $\eta_n(\sigma) = \tilde{o}_{\sigma P}(1)$ if

$$\lim_{\sigma \to 0+} \varlimsup_{n \to \infty} \mathsf{P}\left(\|\eta_n(\sigma)\| > C\right) = 0$$

for all $C > 0$.

Note that $\eta_n(\sigma) = \tilde{o}_{\sigma P}(1)$ if $\eta_n(\sigma) = o_{\sigma P}(1)$. The converse is not always true.

Let $M_i(x_i, y_i)$ and $M_i^0\left(\xi_i, g(\xi_i, \beta_0)\right)$ be points of the space $\mathbb{R}^q \times \mathbb{R}^s$,

$$\Gamma_\beta = \{(\xi, g(\xi, \beta)) \colon \xi \in \mathbb{R}^q\}$$

be the graph of the regression function depending on a parameter β, ρ be the Euclidean distance, and $\rho(M, \Gamma_\beta)$ be the distance between the point M and the graph of the function Γ_β.

In what follows we also need the following contrast condition (in other words, the condition of the asymptotic separation of the point β_0 from the other points β):

$$\text{(con)} \qquad \text{for all } \delta > 0, \qquad \varliminf_{n \to \infty} \inf_{\|\beta - \beta_0\| > \delta} \frac{1}{n} \sum_{i=1}^{n} \rho^2\left(M_i^0, \Gamma_\beta\right) > 0.$$

The contrast condition guarantees that one can construct a consistent estimator of the parameter β as the variance of errors tends to zero; namely, the following result holds.

Lemma 3.1. *Let the contrast condition hold and let $g \in C(\mathbb{R}^q \times \Theta)$. Then*

$$\hat{\beta}_n - \beta_0 = \tilde{o}_{\sigma P}(1).$$

We considered $Q_1^\beta(\beta_0)$ in the proof of the inconsistence of the estimator, while $Q_1^{\beta\beta}(\beta_0)$ is used to estimate the deviation.

Theorem 3.1. *Let assumptions* (i)–(iv) *hold. Then*

$$Q_1^{\beta\beta}(\beta_0) = 2V_n + \sigma O_P(1) + \sigma^4 o_{\sigma P}(1),$$

where

$$V_n := \frac{1}{n}\sum_{i=1}^{n} g^{\beta\mathrm{T}}\left[I_s - SS^{\mathrm{T}} + (g^{\xi} - S)(g^{\xi} - S)^{\mathrm{T}}\right]^{-1} g^{\beta}\Big|_{(\xi_i,\beta_0)}.$$

Proof. Consider $q^{\beta\beta}(x_i, y_i, \beta_0)$ for $i \in B_n(\upsilon)$ (we do not write the subscript i):

$$q^{\beta\beta}(x,y,\beta) = \frac{\partial}{\partial\beta}\left(G^{\beta}(x,y,\beta,h(x,y,\beta))\right) = G^{\beta\beta}(x,y,\beta,u)\Big|_{u=h} + G^{\beta u}(x,y,\beta,u)\Big|_{u=h} h^{\beta}$$
$$= \left(G^{\beta\beta} - G^{\beta u}(G^{uu})^{-1}G^{u\beta}\right)\Big|_{u=h(x,y,\beta)}.$$

Further, we find the second derivatives of the function G (its first derivatives were already evaluated in the preceding section):

$$G^{\beta\beta} = 2\begin{pmatrix} g^{\beta}(u,\beta) \\ O\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\beta}(u,\beta) \\ O\end{pmatrix} - 2\begin{pmatrix} y - g(u,\beta) \\ x-u\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\beta\beta}(u,\beta) \\ O\end{pmatrix},$$

$$G^{\beta u} = \frac{\partial}{\partial u}G^{\beta\mathrm{T}} = 2\begin{pmatrix} g^{\beta}(u,\beta) \\ O\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\xi}(u,\beta) \\ I_q\end{pmatrix} - 2\begin{pmatrix} g^{\beta\xi}(u,\beta) \\ O\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} y - g(u,\beta) \\ x-u\end{pmatrix},$$

$$G^{uu} = 2\begin{pmatrix} g^{\xi}(u,\beta) \\ I_q\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\xi}(u,\beta) \\ I_q\end{pmatrix} - 2\begin{pmatrix} g^{\xi\xi}(u,\beta) \\ O\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} y - g(u,\beta) \\ x-u\end{pmatrix}.$$

Since Γ is a positive definite matrix, there exist positive numbers C and C_1 such that $C\left(\|y - g(h,\beta_0)\|^2 + \|x-h\|^2\right) \le G(x,y,\beta_0,h(x,y,\beta_0)) \le G(x,y,\beta_0,\xi) \le C_1\left(\|\varepsilon\|^2 + \|\delta\|^2\right)$. Thus

$$\|y - g(h(x,y,\beta_0),\beta_0)\| + \|x - h(x,y,\beta_0)\| = O(\|\varepsilon\| + \|\delta\|).$$

The latter result helps us to obtain the principal part of $q(x,y,\beta_0)$ by observing that all x, y, and $h(x,y,\beta_0)$ are bounded.

If $f(t)$, $t \in U$, is a function differentiable at the point $t_0 \in U$, then

$$f(t) - f(t_0) = O\left(\|t - t_0\|\right), \qquad t \to t_0.$$

Putting $h = h(x,y,\beta_0)$ we get

$$G^{\beta\beta}(x,y,\beta_0,h) = 2\begin{pmatrix} g^{\beta} \\ O\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\beta} \\ O\end{pmatrix} + O\left(\|\varepsilon\| + \|\delta\|\right),$$

$$G^{\beta u}(x,y,\beta_0,h) = 2\begin{pmatrix} g^{\beta} \\ O\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q\end{pmatrix} + O\left(\|\varepsilon\| + \|\delta\|\right),$$

$$G^{uu}(x,y,\beta_0,h) = 2\begin{pmatrix} g^{\xi} \\ I_q\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q\end{pmatrix} + O\left(\|\varepsilon\| + \|\delta\|\right) = 2H + O\left(\|\varepsilon\| + \|\delta\|\right),$$

where all functions are considered at the argument (ξ, β_0). Next,

$$q^{\beta}(x,y,\beta_0) = \left(G^{\beta\beta} - G^{\beta u}(G^{uu})^{-1}G^{u\beta}\right)\Big|_{(x,y,\beta_0,h(x,y,\beta_0))}$$
$$= 2\begin{pmatrix} g^{\beta} \\ O\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\beta} \\ O\end{pmatrix} - 2\begin{pmatrix} g^{\beta} \\ O\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\xi} \\ I_q\end{pmatrix} H^{-1}\begin{pmatrix} g^{\xi} \\ I_q\end{pmatrix}^{\mathrm{T}}\Gamma^{-1}\begin{pmatrix} g^{\beta} \\ O\end{pmatrix}$$
$$+ O\left(\|\varepsilon\| + \|\delta\|\right)$$
$$= 2g^{\beta\mathrm{T}}\left(V - V(g^{\xi} - S)^{\mathrm{T}}H^{-1}(g^{\xi} - S)V\right)g^{\beta} + O\left(\|\varepsilon\| + \|\delta\|\right).$$

Hence

$$Q_1^{\beta\beta}(\beta_0) = \frac{2}{n} \sum_{i\in B_n(\upsilon)} g^{\beta\mathrm{T}} \left(V - V(g^\xi - S)^\mathrm{T} H^{-1}(g^\xi - S)V\right) g^\beta\big|_{(\xi_i,\beta_0)} + R$$
$$= \frac{2}{n} \sum_{i=1}^{n} g^{\beta\mathrm{T}} \left(V - V(g^\xi - S)^\mathrm{T} H^{-1}(g^\xi - S)V\right) g^\beta\big|_{(\xi_i,\beta_0)} + \sigma^4 o_{\sigma P}(1) + R,$$

where

$$\|R\| = \frac{1}{n} \sum_{i\in B_n(\upsilon)} O\left(\|\varepsilon_i\| + \|\delta_i\|\right) \le \frac{C}{n} \sum_{i=1}^{n} \left(\|\varepsilon_i\| + \|\delta_i\|\right) = \frac{C\cdot\sigma}{n} \sum_{i=1}^{n} (\|\tilde\varepsilon_i\| + \|\tilde\delta_i\|)$$
$$= \sigma O_P(1)$$

and $C > 0$ is a constant. It remains to apply (8). □

The following condition plays an important role when studying the asymptotic deviation between the estimator and the true value.

(vi) $\underline{\lim}_{n\to\infty} \lambda_{\min}(V_n) > 0$, where $\lambda_{\min}$ is the minimal eigenvalue of the matrix V_n.

Theorem 3.2. *Let conditions* (i)–(vi) *and the contrast condition* (con) *hold. Then*

$$\hat\beta_n = \beta_0 - \frac{\sigma^2}{2} V_n^{-1} k_n^\mathrm{T} + \sigma^2 \tilde o_{\sigma P}(1). \tag{10}$$

Remark 3.1. If the conditions of Theorem 3.2 hold for the estimator $\hat\beta_n$, then

$$\hat\beta_n - \beta_0 = \sigma^2 \tilde O_{\sigma P}(1).$$

On the other hand, $\hat\beta_n - \beta_0 \ne \sigma^2 \tilde o_{\sigma P}(1)$.

Proof. 1) Lemma 3.1 implies that, given arbitrary $\varepsilon > 0$ and $\nu > 0$, there are numbers $\sigma_{\varepsilon\nu} > 0$ and $n_{\varepsilon\nu}$ such that

$$\mathsf{P}\left(\|\hat\beta - \beta_0\| < \nu\right) > 1 - \varepsilon$$

for all $\sigma \in (0, \sigma_{\varepsilon\nu}]$ and $n \ge n_{\varepsilon\nu}$. In the proof below we assume that $\sigma \in (0, \sigma_{\varepsilon\nu}]$ and $n \ge n_{\varepsilon\nu}$. Then the probability that $\hat\beta \in U_\nu(\beta_0)$ is not less than $1 - \varepsilon$.

2) The function $q(x_i, y_i, \beta)$ is three times continuously differentiable in β for $i \in B_n(\nu)$; thus $\|Q_1^{\beta\beta\beta}(\beta)\| <$ const for $\beta \in U_\nu(\beta_0)$, where $Q_1^{\beta\beta\beta}(\beta)$ is a symmetric trilinear form.

3) We expand $Q_1(\beta)$ by the Taylor formula for $\beta \in U_\nu(\beta_0)$:

$$Q_1(\beta) = Q_1(\beta_0) + Q_1^{\beta\mathrm{T}}(\beta_0)\Delta\beta + \frac{1}{2}\Delta\beta^\mathrm{T} Q_1^{\beta\beta}(\beta_0)\Delta\beta + \frac{1}{6}\sum_{i,j,k=1}^{p} Q_1^{\beta_i\beta_j\beta_k}(\bar\beta)\Delta\beta_i\Delta\beta_j\Delta\beta_k,$$

where $\Delta\beta = \beta - \beta_0$ and $\bar\beta \in [\beta_0, \beta]$. We use the expressions obtained in Theorems 2.1 and 3.1 instead of $Q_1^\beta(\beta_0)$ and $Q_1^{\beta\beta}(\beta_0)$:

$$Q(\beta) = Q_1(\beta) + Q_2(\beta)$$
$$= Q_1(\beta_0) + \left(\sigma^2 k_n + \sigma^3 O_P(1) + \frac{\sigma}{\sqrt{n}} O_P(1) + \sigma^4 o_{\sigma P}(1)\right)\Delta\beta$$
$$+ \Delta\beta^\mathrm{T}\left(V_n + \sigma O_P(1) + \sigma^4 o_{\sigma P}(1)\right)\Delta\beta + O(1)\|\Delta\beta\|^3 + \sigma^4 o_{\sigma P}(1) + Q_2(\beta_0)$$
$$+ \left(Q_2(\beta) - Q_2(\beta_0)\right),$$

$$Q(\beta) = Q(\beta_0) + \left(\sigma^2 k_n + \sigma^3 O_P(1) + \frac{\sigma}{\sqrt{n}} O_P(1)\right)\Delta\beta + \Delta\beta^\mathrm{T}\left(V_n + \sigma O_P(1)\right)\Delta\beta$$
$$+ O(1)\|\Delta\beta\|^3 + \sigma^4 o_{\sigma P}(1).$$

Put $\Delta\beta = \sigma^2\Delta\varphi$ in the latter formula:

$$Q(\beta) = Q(\beta_0) + \sigma^4\Big(k_n\Delta\varphi + \Delta\varphi^{\mathrm{T}}V_n\Delta\varphi + \sigma O_P(1)\left(\|\Delta\varphi\| + \|\Delta\varphi\|^2\right) + \frac{1}{\sigma^3\sqrt{n}}O_P(1)\|\Delta\beta\| + O(1)\|\Delta\varphi\|^2\|\Delta\beta\| + o_{\sigma P}(1)\Big),$$

$$\begin{aligned}\frac{Q(\beta)-Q(\beta_0)}{\sigma^4} &= k_n\Delta\varphi + \Delta\varphi^{\mathrm{T}}V_n\Delta\varphi + \sigma O_P(1)\left(\|\Delta\varphi\| + \|\Delta\varphi\|^2\right)\\ &\quad + \frac{1}{\sigma^3\sqrt{n}}O_P(1)\|\Delta\beta\| + O(1)\|\Delta\varphi\|^2\|\Delta\beta\| + o_{\sigma P}(1).\end{aligned} \tag{11}$$

Since $\hat\beta - \beta_0 = \tilde o_{\sigma P}(1)$ by Lemma 3.1, we rewrite (11) for $\Delta\hat\varphi = \frac{\hat\beta-\beta_0}{\sigma^2}$:

$$\begin{aligned}\frac{Q(\hat\beta)-Q(\beta_0)}{\sigma^4} &= k_n\Delta\hat\varphi + \Delta\hat\varphi^{\mathrm{T}}V_n\Delta\hat\varphi + \sigma O_P(1)\|\Delta\hat\varphi\| + \|\Delta\hat\varphi\|^2\tilde o_{\sigma P}(1) + o_{\sigma P}(1)\\ &\quad + \frac{1}{\sigma^3\sqrt{n}}O_P(1)\\ &\le 0.\end{aligned} \tag{12}$$

Given arbitrary $\varepsilon > 0$ one can choose σ_0 and $n_{\sigma,\varepsilon}$ such that

$$\mathsf{P}\left(|\tilde o_{\sigma P}(1)| < \frac{1}{2}\lambda_{\min}(V_n)\right) > 1-\varepsilon$$

for $\sigma \in (0,\sigma_0]$ and $n \ge n_{\sigma,\varepsilon}$. Then (12) implies that $\Delta\hat\varphi = \tilde O_{\sigma P}(1)$. Hence

$$\frac{Q(\hat\beta)-Q(\beta_0)}{\sigma^4} = k_n\Delta\hat\varphi + \Delta\hat\varphi^{\mathrm{T}}V_n\Delta\hat\varphi + \tilde o_{\sigma P}(1). \tag{13}$$

Let $z_n := -\frac{1}{2}V_n^{-1}k_n^{\mathrm{T}}$. By the definitions of V_n and k_n and by the assumptions of the theorem, $\|z_n\|$ is bounded and asymptotically isolated from zero. Now we rewrite (11) for $\Delta\varphi = z_n$:

$$\begin{aligned}\frac{Q(\beta_0+\sigma^2 z_n)-Q(\beta_0)}{\sigma^4} &= k_n z_n + z_n^{\mathrm{T}}V_n z_n + \sigma O_P(1) + \tilde o_{\sigma P}(1) + \frac{1}{\sigma^3\sqrt{n}}O_P(1)\\ &= k_n z_n + z_n^{\mathrm{T}}V_n z_n + \tilde o_{\sigma P}(1).\end{aligned} \tag{14}$$

Using the definition of $\hat\beta$ we obtain

$$Q(\hat\beta) \le Q(\beta_0 + \sigma^2 z_n),$$

whence

$$\begin{aligned}\frac{Q(\hat\beta)-Q(\beta_0+\sigma^2 z_n)}{\sigma^4} &= k_n\Delta\hat\varphi + \Delta\hat\varphi^{\mathrm{T}}V_n\Delta\hat\varphi - k_n z_n - z_n^{\mathrm{T}}V_n z_n + \tilde o_{\sigma P}(1)\\ &= (\Delta\hat\varphi - z_n)^{\mathrm{T}}V_n(\Delta\hat\varphi - z_n) + 2z_n^{\mathrm{T}}V_n\Delta\hat\varphi - 2z_n^{\mathrm{T}}V_n z_n + k_n\Delta\hat\varphi - k_n z_n + \tilde o_{\sigma P}(1)\\ &= (\Delta\hat\varphi - z_n)^{\mathrm{T}}V_n(\Delta\hat\varphi - z_n) + (2z_n^{\mathrm{T}}V_n - k_n)(\Delta\hat\varphi - z_n) + \tilde o_{\sigma P}(1)\\ &= (\Delta\hat\varphi - z_n)^{\mathrm{T}}V_n(\Delta\hat\varphi - z_n) + \tilde o_{\sigma P}(1) \le 0\end{aligned}$$

by (13) and (14).

Using condition (vi) we get $\underline{\lim}_{n\to\infty}\lambda_{\min}(V_n) = \tau > 0$. Therefore there exists an integer n_0 such that

$$0 \le \|\Delta\hat\varphi - z_n\|^2\frac{\tau}{2} \le (\Delta\hat\varphi - z_n)^{\mathrm{T}}\,V_n\,(\Delta\hat\varphi - z_n) \le \tilde o_{\sigma P}(1)$$

for all $n \ge n_0$.

We have proved that $\Delta\hat\varphi - z_n = \tilde o_{\sigma P}(1)$, so that $\hat\beta = \beta_0 - \frac{\sigma^2}{2}V_n^{-1}k_n^{\mathrm{T}} + \sigma^2\tilde o_{\sigma P}(1)$. □

4. An improved estimator

Theorem 3.1 allows one to estimate the deviation between the orthogonal regression estimator and the true value if σ^2, k_n, and V_n are known with sufficient precision. Below we introduce another estimator $\tilde{\beta}_n$ as follows:

$$\tilde{\beta}_n := \hat{\beta}_n + \frac{\hat{\sigma}^2}{2}\hat{V}_n^{-1}\hat{k}_n^{\mathrm{T}},$$

where

$$\hat{k}_n := \frac{1}{n}\sum_{i=1}^{n} \overrightarrow{\mathrm{tr}}^{\mathrm{T}}\left(g^{\xi\xi}H^{-1}\right)\left[I_s - SS^{\mathrm{T}} + (g^{\xi}-S)(g^{\xi}-S)^{\mathrm{T}}\right]^{-1} g^{\beta}\Big|_{(x_i,\hat{\beta}_n)},$$

$$\hat{V}_n := \frac{1}{n}\sum_{i=1}^{n} g^{\beta\mathrm{T}}\left[I_s - SS^{\mathrm{T}} + (g^{\xi}-S)(g^{\xi}-S)^{\mathrm{T}}\right]^{-1} g^{\beta}\Big|_{(x_i,\hat{\beta}_n)},$$

$$\hat{\sigma}^2 := Q(\hat{\beta})/s.$$

We show that $\tilde{\beta} - \beta_0 = \sigma^2\tilde{o}_{\sigma P}(1)$.

Lemma 4.1. *Let $F \in C^1(\mathbb{R}^q \times U)$, $U \supset \Theta$, and let U be an open set. Assume that the contrast condition* (con) *holds and that*

$$\|F^{\xi}(\xi,\beta)\| \le Ce^{A\|\xi\|}, \qquad \xi \in \mathbb{R}^q,\ \beta \in \Theta,$$

for some constants C and A. Then

$$\frac{1}{n}\sum_{i=1}^{n} F(x_i,\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} F(\xi_i,\beta_0) + \tilde{o}_{\sigma P}(1).$$

Proof. Consider the difference

$$\frac{1}{n}\sum_{i=1}^{n}\Big(F(x_i,\hat{\beta}) - F(\xi_i,\beta_0)\Big) = \frac{1}{n}\sum_{i=1}^{n}\Big(F(x_i,\hat{\beta}) - F(\xi_i,\hat{\beta})\Big) + \frac{1}{n}\sum_{i=1}^{n}\Big(F(\xi_i,\hat{\beta}) - F(\xi_i,\beta_0)\Big).$$

Every term is estimated separately. First,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\Big(F(x_i,\hat{\beta}) - F(\xi_i,\hat{\beta})\Big)\right| = \left|\frac{1}{n}\sum_{i=1}^{n} F^{\xi}(\bar{\xi}_i,\hat{\beta})\sigma\tilde{\varepsilon}_i\right| \le \frac{\sigma}{n}\sum_{i=1}^{n}\left|F^{\xi}(\bar{\xi},\hat{\beta})\right|\cdot\|\tilde{\varepsilon}_i\|$$
$$\le \frac{C\sigma}{n}\sum_{i=1}^{n}\exp\big(A\|\xi_i\| + \sigma\|\tilde{\varepsilon}_i\|\big)\|\tilde{\varepsilon}_i\| \le \mathrm{const}\,\frac{\sigma}{n}\sum_{i=1}^{n} e^{A\sigma\|\tilde{\varepsilon}_i\|}\|\tilde{\varepsilon}_i\| = \sigma O_P(1) = \tilde{o}_{\sigma P}(1).$$

For $\beta \in U_\nu(\beta_0)$,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\Big(F(\xi_i,\hat{\beta}) - F(\xi_i,\beta_0)\Big)\right| \le \sup_{\substack{\|\xi\|\le a\\ \beta\in U_\nu(\beta_0)}} \|F^{\beta}(\xi,\beta)\|\cdot\|\hat{\beta}-\beta_0\| = \tilde{o}_{\sigma P}(1),$$

since $\|\hat{\beta}-\beta_0\| = \tilde{o}_{\sigma P}(1)$ by Lemma 3.1. □

Lemma 4.2. *Let conditions* (i)–(vi) *and the contrast condition* (con) *hold. Then*

$$\sigma^2 = \frac{Q(\hat{\beta})}{s} + \sigma^3\tilde{O}_{\sigma P}(1).$$

Proof. Since
$$\Delta\hat{\varphi} = \tilde{O}_{\sigma P}(1),$$
it follows from (13) that $Q(\hat{\beta}) - Q(\beta_0) = \sigma^4 \tilde{O}_{\sigma P}(1)$. Thus $Q(\hat{\beta}) = Q(\beta_0) + \sigma^4 \tilde{O}_{\sigma P}(1)$. It remains to prove that
$$Q(\beta_0) = \sigma^2 s + \frac{\sigma^2}{\sqrt{n}} O_P(1) + \sigma^3 O_P(1) + \sigma^4 o_{\sigma P}(1) = \sigma^2 s + \sigma^3 \tilde{O}_{\sigma P}(1).$$

The term $Q(\beta_0)$ is expanded in the same way as $Q_1^\beta(\beta_0)$ in the proof of Theorem 2.1. By (3),
$$Q(\beta_0) = Q_1(\beta_0) + \sigma^4 o_{\sigma P}(1).$$
Considering $Q_1(\beta_0) = n^{-1} \sum_{i \in B_n(\nu)} q(x_i, y_i, \beta_0)$, we get
$$\begin{aligned} q(x, y, \beta_0) &= \begin{pmatrix} y - g(h, \beta_0) \\ x - h \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} y - g(h, \beta_0) \\ x - h \end{pmatrix} \\ &= \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix} + O\left(\|\varepsilon\|^3 + \|\delta\|^3\right). \end{aligned}$$

To evaluate the mathematical expectation of the first term, we apply the explicit expressions for the mathematical expectations obtained in the proof of Theorem 2.1:
$$\begin{aligned} &\mathsf{E} \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \delta - g^\xi \Delta_1 \\ \varepsilon - \Delta_1 \end{pmatrix} \\ &\quad = \mathsf{E} \left[\begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix} - 2 \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} \Delta_1 + \Delta_1^{\mathrm{T}} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} \Delta_1 \right] \\ &\quad = \mathsf{E} \left[\operatorname{tr} \Gamma^{-1} \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix} \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}^{\mathrm{T}} - 2 \operatorname{tr} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} \Delta_1 \begin{pmatrix} \delta \\ \varepsilon \end{pmatrix}^{\mathrm{T}} + \operatorname{tr} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} \Delta_1 \Delta_1^{\mathrm{T}} \right] \\ &\quad = \sigma^2 \operatorname{tr}(\Gamma^{-1}\Gamma) - 2\sigma^2 \operatorname{tr} \left\{ \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} H^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \right\} + \sigma^2 \operatorname{tr} \left\{ \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} H^{-1} \right\} \\ &\quad = \sigma^2 \operatorname{tr}(I_{s+q}) - 2\,\sigma^2 \operatorname{tr} \left\{ \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} H^{-1} \right\} + \sigma^2 \operatorname{tr} \left\{ \begin{pmatrix} g^\xi \\ I_q \end{pmatrix}^{\mathrm{T}} \Gamma^{-1} \begin{pmatrix} g^\xi \\ I_q \end{pmatrix} H^{-1} \right\} \\ &\quad = \sigma^2 \operatorname{tr}(I_{s+q}) - \sigma^2 \operatorname{tr}\left\{ H H^{-1} \right\} = \sigma^2 s. \end{aligned}$$

Similarly to the proof of Theorem 2.1, $Q(\beta_0) = \sigma^2 s + (\sigma^2/\sqrt{n}) O_P(1) + \sigma^4 o_{\sigma P}(1) + R$, where
$$\begin{aligned} |R| &= \frac{1}{n} \sum_{i \in B_n(\nu)} O\left(\|\varepsilon_i\|^3 + \|\delta_i\|^3\right) \le \frac{\text{const}}{n} \sum_{i \in B_n(\nu)} \left(\|\varepsilon_i\|^3 + \|\delta_i\|^3\right) \\ &\le \frac{\text{const}}{n} \sum_{i=1}^n \left(\|\varepsilon_i\|^3 + \|\delta_i\|^3\right) = \sigma^3 \frac{\text{const}}{n} \sum_{i=1}^n \left(\|\tilde{\varepsilon}_i\|^3 + \|\tilde{\delta}_i\|^3\right) = \sigma^3 O_P(1). \quad \square \end{aligned}$$

Below we state the main result on the improved estimator.

Theorem 4.1. *Let conditions* (i)–(vi) *and the contrast condition* (con) *hold. Assume that*
$$\left\|g^{\xi\beta}(\xi, \beta)\right\| + \left\|g^{\xi\xi\xi}(\xi, \beta)\right\| \le C e^{A\|\xi\|}, \qquad \xi \in \mathbb{R}^q,\ \beta \in \Theta,$$
for some constants C and A. Then
$$\tilde{\beta}_n = \beta_0 + \sigma^2 \tilde{o}_{\sigma P}(1).$$

Proof. If

$$\|g^{\beta}(\xi,\beta)\| + \|g^{\xi}(\xi,\beta)\| + \|g^{\xi\beta}(\xi,\beta)\| + \|g^{\xi\xi}(\xi,\beta)\| + \|g^{\xi\xi\xi}(\xi,\beta)\| \le Ce^{A\|\xi\|}$$

for $\xi \in \mathbb{R}^q$ and $\beta \in \Theta$, then $k_n = \hat{k}_n + \tilde{o}_{\sigma P}(1)$ and $V_n = \hat{V}_n + \tilde{o}_{\sigma P}(1)$ by Lemma 4.1. We show that the latter condition follows from the hypotheses of the theorem (however, the constants C and A are different in the hypotheses of the theorem and in the condition above). Let a function f be such that $\|f^{\xi}(\xi,\beta)\| \le Ce^{A\|\xi\|}$, $\xi \in \mathbb{R}^q$, $\beta \in \Theta$. Then

$$\int_{\vec{0}}^{\xi} (f^{\xi}(t,\beta), dt) = f(\xi,\beta) - f(\vec{0},\beta)$$

and $f(\vec{0},\beta)$, as a function of β, is bounded on the compact set Θ. Hence

$$|f(\xi,\beta)| \le C_1 e^{A_1\|\xi\|}, \qquad \xi \in \mathbb{R}^q,\ \beta \in \Theta,$$

for some constants C_1 and A_1.

Hence $\sigma^2 V_n^{-1} k_n^{\mathrm{T}} = \hat{\sigma}^2 \hat{V}_n^{-1} \hat{k}_n^{\mathrm{T}} + \sigma^2 \tilde{o}_{\sigma P}(1)$, whence the theorem follows by (10). □

Example. Consider the following errors-in-variables model with the regression function

$$g(\xi,\beta) = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} e^{\beta_1\xi_1+\beta_2\xi_2} \\ \beta_3\xi_1 \end{pmatrix},$$

where $\beta = (\beta_1,\beta_2,\beta_3)^{\mathrm{T}} \in \mathbb{R}^3$, $\xi = (\xi_1,\xi_2)^{\mathrm{T}} \in \mathbb{R}^2$, and the covariance matrix of the vector $(\delta_i^{\mathrm{T}}, \varepsilon_i^{\mathrm{T}})^{\mathrm{T}}$ is $\sigma^2 I_4$. Let the true values of the regressors be $\xi(i) = \bigl(\xi_1(i), \xi_2(i)\bigr)^{\mathrm{T}}$. Then $h = (h_1,h_2)^{\mathrm{T}}$ is determined by the system of equations

$$\begin{gathered}(x_1 - h_1) + \beta_1(y_1 - e^{\beta_1 h_1+\beta_2 h_2})e^{\beta_1 h_1+\beta_2 h_2} + \beta_3(y_2 - \beta_3 h_1) = 0,\\ (x_2 - h_2) + \beta_2(y_1 - e^{\beta_1 h_1+\beta_2 h_2})e^{\beta_1 h_1+\beta_2 h_2} = 0.\end{gathered}$$

Put $L(\xi,\beta) := 1 + \left(\beta_1^2+\beta_2^2\right) g_1^2 + \beta_3^2\left(1+\beta_2^2 g_1^2\right)^2$. Then

$$k_n^{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\beta_1^2+\beta_2^2(1+\beta_3^2)g_1^2}{L} \begin{pmatrix} (1+\beta_3^2)\xi_1 \\ (1+\beta_3^2)\xi_2 \\ -\beta_1\beta_3\xi_1 \end{pmatrix}\Bigg|_{\left(\xi(i),\beta_0\right)},$$

$$V_n = \frac{1}{n}\sum_{i=1}^{n} \frac{g_1^2}{L} \begin{pmatrix} (1+\beta_3^2)\xi_1^2 & (1+\beta_3^2)\xi_1\xi_2 & -\beta_1\beta_3\xi_1^2 \\ (1+\beta_3^2)\xi_1\xi_2 & (1+\beta_3^2)\xi_2^2 & -\beta_1\beta_3\xi_1\xi_2 \\ -\beta_1\beta_3\xi_1^2 & -\beta_1\beta_3\xi_1\xi_2 & (\beta_1^2+\beta_2^2+g_1^{-2})\xi_1^2 \end{pmatrix}\Bigg|_{\left(\xi(i),\beta_0\right)}.$$

The estimator $\hat{\beta}_n$ is inconsistent if condition (v) holds. For example, this is the case if all components of $\xi(i)$, $i \ge 1$, are positive and isolated from zero. Let $\hat{k}_n$ and $\hat{V}_n$ be the corresponding estimators of k_n and V_n, respectively, and $\hat{\sigma}^2 = Q(\hat{\beta}_n)/2$. Then, for small σ, the improved estimator is given by

$$\tilde{\beta}_n = \hat{\beta}_n + \frac{1}{4} Q(\hat{\beta}_n)\hat{V}_n^{-1}\hat{k}_n^{\mathrm{T}}.$$

5. Concluding remarks

We considered the orthogonal regression estimator $\hat{\beta}_n$ for the vector errors-in-variables model and proved that it is inconsistent for small but fixed σ. We also proved that the estimator belongs with a large probability to a neighborhood of the point $\beta_0 - \sigma^2 V_n^{-1} k_n^{\mathrm{T}}/2$. We introduced a new estimator $\tilde{\beta}_n$ whose asymptotic deviation from the true value has a smaller order with respect to σ:

$$\tilde{\beta}_n - \beta_0 = \tilde{o}_{\sigma P}(1), \qquad \hat{\beta}_n - \beta_0 = \tilde{O}_{\sigma P}(1).$$

The estimator $\tilde{\beta}_n$ approximates β_0 better than $\hat{\beta}_n$ does for small fixed σ and large n. One can further improve the estimator by finding additional terms of the asymptotic deviation with respect to σ^2.

Another question is that condition (v), basic for the proof of the inconsistency of the estimator, does not hold for some nonlinear models. For example, if $g = \xi_1 \tan(\beta_1 \xi_2 + \beta_2)$ and $S = O$, then $\operatorname{tr}\left(g^{\xi\xi} H^{-1}\right) \equiv 0$, whence $k_n = \vec{0}$. It is interesting to find out whether the estimator is consistent in this model and, if this is not the case, to obtain the deviation between the true value and the estimator.

Bibliography

1. W. A. Fuller, *Measurement Errors Models*, Wiley, New York, 1987. MR898653 (89a:62160)
2. K. M. Wolter and W. A. Fuller, *Estimation of nonlinear errors-in-variables models*, Ann. Statist. **10** (1982), no. 2, 539–548. MR653528 (83j:62099)
3. I. Fazekas, A. G. Kukush, and S. Zwanzig, *On Inconsistency of the Least Squares Estimator in Nonlinear Functional Relations*, Preprint, vol. 1, Department of Statistics and Demography, Odense University, Denmark, 1998.
4. ———, *Bias correction of nonlinear orthogonal regression*, Ukrain. Mat. Zh. **56** (2004), no. 8, 1101–1118. MR2136312 (2006g:62065)

Department of Probability Theory and Mathematical Statistics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue 6, Kiev 03127, Ukraine

E-mail address: galchonok@univ.kiev.ua

Received 17/SEP/2004

Translated by S. KVASKO