GALYNA REPETATSKA

MODIFIED ORTHOGONAL REGRESSION ESTIMATOR
IN THE QUADRATIC ERRORS-IN-VARIABLES MODEL

The quadratic functional measurement error model with equal error variances is considered. The asymptotic bias of an orthogonal regression estimator is derived. A modified estimator which has smaller asymptotic bias for small measurement errors is presented.

1. Model assumptions and orthogonal regression estimator

Let \( g(\xi, \beta) = a\xi^2 + b\xi + c \) be a regression function, where \( \beta = (a; b; c)^T \in \Theta \) is the vector of the unknown parameters. In the paper, all the vector values are column vectors. The derivatives are denoted by superscripts, and the vector derivatives are row vectors. For example, \( g^\xi(\xi, \beta) = 2a\xi + b \), and \( g^\beta = (\xi^2; \xi; 1) \) is the derivative with respect to the vector variable \( \beta \). The expectation of a random variable \( \zeta \) is denoted by

---

**Introduction**

We consider a quadratic functional errors-in-variables model

\[
y_i = a_0 \xi_i^2 + b_0 \xi_i + c_0 + \delta_i,
\]

\[
x_i = \xi_i + \epsilon_i, \quad 1 \leq i \leq n,
\]

where \((x_i, y_i), 1 \leq i \leq n, \) are observed, \( \xi_i \) are unknown nonrandom parameters, \( \epsilon_i, \delta_i \) are i.i.d. normal error terms, and the vector \( \beta_0 = (a_0, b_0, c_0)^T \) consists of the parameters to be estimated. Noise variances are unknown.

The general discussion of the linear error-in-variables model is given in [4]. Concerning the orthogonal regression estimator, it is proved in [1] that, for nonlinear errors-in-variables models including (1), this estimator is inconsistent. In [2], a new corrected estimator is presented which has smaller asymptotic bias. In [5], this estimator with some changes was extended for a model, where all variables are vectors.

In this paper, the next term of the asymptotic bias is derived, and a new estimator is proposed. A similar estimator can be used for other nonlinear regression models, but we consider, for simplicity, only the quadratic regression function. In Section 1, the model assumptions and an orthogonal regression estimator are presented. In Section 2, two leading terms of the asymptotic bias of the estimator are derived. In Section 3, two corrected estimators are proposed. The first estimator has been proposed in [2], another one is original. It has less asymptotic deviation than the first one.

Some calculations were performed with the Mathematica 3.0 program. The proofs of Theorems 1 and 2 are put in Appendix.
$E\zeta$, and its variance is denoted by $D\zeta$. A sequence $\{X_n(\theta), n \geq 1\}$ of random functions is denoted by $O_P(1)$ if it is uniformly stochastically bounded.

Let $G(x, y, \beta, u) = (y - g(u, \beta))^2 + (x - u)^2$. Then $q(x, y, \beta) := \min_{u \in \mathbb{R}} G(x, y, u, \beta)$ is the squared distance between a point $(x, y)$ and a parabola $y = g(u, \beta), u \in \mathbb{R}$.

Introduce the objective function $Q(\beta) = \frac{1}{n} \sum_{i=1}^{n} q(x, y, \beta)$. Then the orthogonal regression estimator $\hat{\beta}$ is defined as a measurable solution to the optimization problem:

$$Q(\beta) = \min, \beta \in \Theta,$$

where $\Theta$ is a parameter set.

Assume that the following conditions hold:

(i) $\beta_0 \in \text{int } \Theta, \Theta$ is a compact set in $\mathbb{R}^3$.

(ii) $|\xi_i| \leq A, i \geq 1$, where $A$ is unknown.

(iii) $\varepsilon_i, \delta_i \sim N(0, \sigma^2)$ i.i.d., $i \geq 1$, where $\sigma > 0$ is the unknown parameter.

(iv) $a_0 \neq 0$, i.e., the true regression function is nonlinear.

Consider the problem of existence and uniqueness of a minimum point of the function $G(x, y, \beta, u), u \in \mathbb{R}$.

1. Existence. The function $G(x, y, \beta, u)$ is continuous and tends to $+\infty$ as $u \to \infty$. Therefore, there exists at least one minimum point. Denote one of such points by $h(x, y, \beta)$. Note that, for any minimum point $h$,

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

2. Uniqueness. For a minimum point, $G^u(x, y, \beta, u)|_{u=h(x,y,\beta)} = 0$ holds. Hence, $h(x, y, \beta)$ is implicitly defined by the normal equation

$$(2) \quad F(x, y, \beta, h) := -\frac{1}{2} G^u|_{u=h} = (y - g(h, \beta)) g^\xi(h, \beta) + x - h = 0.$$

Hence, $h$ is a solution to the cubic equation $(y - ah^2 - bh - c)(2ah + b) + x - h = 0$. The equation can have from one to three solutions, some of them can not be a global minimum point. Note that $F(\xi, g(\xi, \beta), \beta, \xi) \equiv 0$ and $F^u(\xi, g(\xi, \beta), \beta, \xi) = -1 - [g^\xi(u, \beta)]^2 \neq 0$.

Then the Implicit Functions Theorem implies the following: there exists a neighbourhood of a point $(\xi, g(\xi, \beta), \beta, U_\nu(\xi, \beta) := B_\nu(\xi) \times B_\nu(g(\xi, \beta)) \times B_\nu(\beta), \nu = \nu(\xi, \beta)$, such that $\nu : U_\nu(\xi, \beta) \to \mathbb{R}$ is a uniquely defined infinitely differentiable function. Since $\xi$ and $\beta$ belong to compact sets, it is possible to find a common value $\nu_0 > 0$ for all $\beta \in \Theta, \xi \in [-A - 1, A + 1]$.

If the absolute value of both error terms $\varepsilon_i, \delta_i$ is less than $\nu_0$, then there exists only one perpendicular from $(x_i, y_i)$ to any of the curves $y = g(\xi, \beta), \xi \in \mathbb{R}, \beta \in U_\nu(\beta_0)$. Let $\nu$ be a fixed positive constant, $\nu \in (0, \nu_0]$, such that $U_\nu(\beta_0) \subset \Theta$. We define the index set $B_n(\nu) = \{i = 1, n : |\varepsilon_i| < \nu, |\delta_i| < \nu\}$ and divide the objective function into two parts:

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

Here, $Q_1(\beta)$ is the leading term and $Q_2(\beta)$ is the remainder one. Now we find an asymptotic expansion of $Q_1(\beta)$ and its derivatives in $\sigma^2$.

We will widely use the following statement.

**Lemma 1.** Let $\{\xi_i : i \geq 1\}$ be an i.i.d. sequence with $D\zeta = 1$, and let $\{a_i : i \geq 1\}$ be a bounded sequence of real numbers. Then

$$\frac{1}{n} \sum_{i=1}^{n} a_i \xi_i = \frac{E\zeta}{n} \sum_{i=1}^{n} a_i + \frac{1}{\sqrt{n}} O_P(1).$$

The derivatives $g^\beta, g^\xi, g^\xi \xi$ are row vectors. For a couple of row vectors $\vec{a}, \vec{b}$, we define a symmetric matrix $\vec{a} \ast \vec{b} = \frac{1}{2}(\vec{a}^T \vec{b} + \vec{b}^T \vec{a})$. For a triple $\vec{a}, \vec{b}, \vec{c}$, let $\vec{a} \ast \vec{b} \ast \vec{c}$ be a
cubic matrix corresponding to a symmetric trilinear form which acts on a vector $\vec{x}$ as $(\vec{a}, \vec{x}) \cdot (\vec{b}, \vec{x}) \cdot (\vec{c}, \vec{x})$.

Define the following functions:

\[
k(\xi, \beta) = \frac{g^{2\xi}}{(1 + (g^{2})^2)^2} g^\beta, \quad V(\xi, \beta) = \frac{1}{1 + (g^{2})^2} g^{\beta} g^\beta.
\]

\[
p(\xi, \beta) = \frac{9\alpha^2(\xi, \beta)^2}{(1 + (g^{2})^2)^2} g^\beta + \frac{3\alpha^2g^\xi}{(1 + (g^{2})^2)^4} g^\beta + \frac{a_0}{2(1 + (g^{2})^2)^2} g^{\beta^{\beta \xi}},
\]

\[
W(\xi, \beta) = \frac{(g^{2\xi})(7(\xi, \beta)^2 - 2)}{(1 + (g^{2})^2)^4} g^{\beta} g^\beta - \frac{8g^{2\xi}g^\xi}{(1 + (g^{2})^2)^4} g^\beta g^\beta + \frac{1}{(1 + (g^{2})^2)^2} g^\beta g^\beta,
\]

\[
T(\xi, \beta) = \frac{g^{2\xi}g^\xi}{(1 + (g^{2})^2)^3} g^\beta g^\beta - \frac{2g^\xi}{(1 + (g^{2})^2)^2} g^\beta g^\beta.
\]

For an arbitrary function $F(\xi, \beta)$, let $F_n = \frac{1}{n} \sum_{i=1}^{n} F(\xi, \beta_0)$. In this way, we can define the quantities $k_n, V_n, p_n, W_n, T_n$.

**Definition 1.** A sequence of random vectors $\eta_n(\beta, \sigma) = o_\sigma P(1)$, if, for each $c > 0$,

\[
\lim_{\sigma \to 0^+} \sup_{n \geq 1} P \left( \sup_{\beta \in \Theta} ||\eta_n(\beta, \nu, \sigma)|| > c \right) = 0.
\]

The following theorem gives the asymptotic expansions of the function $Q_1(\beta)$ and its derivatives.

**Theorem 1.** Suppose that, for model (1), assumptions (i)–(iii) are satisfied. Then

\[
Q(\beta) = Q_1(\beta) + \sigma^2 o_\sigma P(1),
\]

\[
Q_1(\beta_0) = \sigma^2 - \frac{\sigma^4}{4n} \sum_{i=1}^{n} \frac{(g^{2\xi})^2}{(1 + (g^{2})^2)^2} ||\xi, \beta_0|| + \frac{\sigma^2}{\sqrt{n}} O_P(1) + \sigma^4 o_\sigma P(1),
\]

\[
Q_1^3(\beta_0) = 2\sigma^2 k_n + \sigma^4 p_n + \sigma^6 R_1 + \sigma^4 o_\sigma P(1) + \frac{\sigma}{\sqrt{n}} O_P(1),
\]

\[
Q_1^{3\beta}(\beta_0) = 2V_n + 2\sigma^2 W_n + \sigma^4 R_2 + \sigma^4 o_\sigma P(1) + \frac{\sigma}{\sqrt{n}} O_P(1),
\]

\[
Q_1^{3\beta^3}(\beta_0) = 6T_n + \sigma^2 R_3 + \sigma^2 o_\sigma P(1) + \frac{\sigma}{\sqrt{n}} O_P(1),
\]

where $R_1, R_2, R_3, R_4$ are bounded nonrandom terms.

The inconsistency of an orthogonal regression estimator was proved in [1] in the case where $k_n$ is separated from zero. Theorem 1 helps us to find two leading terms of the asymptotic expansion of $\beta - \beta_0$ in powers of $\sigma^2$.

2. **Asymptotic deviation**

**Definition 2.** A sequence of random vectors $\eta_n(\sigma) = \tilde{o}_\sigma P(1)$, if

\[
\forall \varepsilon > 0 \exists C > 0 : \lim_{\sigma \to 0^+} \limsup_{n \to \infty} P(\|\eta_n(\sigma)\| > C) < \varepsilon.
\]

**Definition 3.** A sequence of random vectors $\eta_n = \tilde{o}_\sigma P(1)$, if

\[
\lim_{\sigma \to 0^+} \limsup_{n \to \infty} P(\|\eta_n(\sigma)\| > C) \to 0, \quad C \to \infty.
\]
Let $M_i(x_i, y_i), M_i^0(\xi_i, g(\xi_i, \beta_0))$ be the points on the plane, $\Gamma_\beta := \{(\xi, g(\xi, \beta)) : \xi \in \mathbb{R}\}$ be a plot of the regression function with a parameter $\beta, \rho$ be the Euclidean metrics, and $\rho(M, \Gamma_\beta)$ be the distance between a point $M$ and the plot $\Gamma_\beta$.

We need the following contrast condition:

$$(\text{con}) \quad \forall \delta > 0 : \liminf_{n \to \infty} \inf_{\|\beta - \beta_0\| > \delta} \frac{1}{n} \sum_{i=1}^{n} \rho^2(M^0_i, \Gamma_\beta) > 0. $$

This condition makes it possible to estimate consistently the parameter $\beta_0$ by $\hat{\beta}$, as $n \to \infty$ and $\sigma \to 0$.

**Lemma 2** [2]. Suppose that, for model (1), the contrast condition (con) is satisfied. Then a.s. $\forall \gamma > 0$ $\exists \sigma_\gamma > 0$ $\exists n_\gamma = n_\gamma(\omega)$ $\forall n \geq n_\gamma \forall \sigma_\gamma \leq \sigma_\gamma : \|\hat{\beta}_n - \beta_0\| < \gamma$.

This implies that $\hat{\beta}_n - \beta_0 = \hat{\alpha}_o P(1)$.

Denote the minimal eigenvalue of a matrix $A$ by $\lambda_{\text{min}}(A)$. To find the asymptotic deviation of the estimate, we need the following assumption:

$$(v) \quad \liminf_{n \to \infty} \lambda_{\text{min}}(V_n) > 0. $$

**Theorem 2.** Suppose that, for model (1), conditions $(i)$–$(v)$ and (con) are satisfied. Then

$$\hat{\beta}_n = \beta_0 + \sigma^2 z_n + \sigma^4 \hat{\alpha}_o P(1),$$

$$\hat{\beta}_n = \beta_0 + \sigma^2 z_n + \sigma^4 a_n + \sigma^4 \hat{\alpha}_o P(1),$$

where $z_n := -\frac{1}{2} V_n^{-1} k_n^T$, $a_n := -\frac{1}{2} V_n^{-1} \left( p_n + 2 z_n^T W_n + 3 T_n (z_n)^2 \right)$.

3. Modified estimators

We will construct consequently two estimators which have smaller asymptotic bias than $\hat{\beta}$. We have to estimate the terms $z_n$ and $a_n$ of the asymptotic expansion in (3).

Let $F(\xi, \beta), \xi \in \mathbb{R}, \beta \in \Theta$ be an arbitrary twice differentiable function.

1) For $F_n = \frac{1}{n} \sum_{i=1}^{n} F(\xi_i, \beta_0)$, we introduce the following estimator:

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^{n} F(x_i, \hat{\beta}).$$

Thus, we have the estimators of the terms $k_n, V_n, p_n, W_n, T_n$ in the form $\hat{k}_n, \hat{V}_n, \hat{\beta}$, etc.

For $\sigma^2$, we have the estimator $\hat{\sigma}^2 := Q(\hat{\beta})$.

Next, we have a new estimator of the parameter $\beta_0$:

$$\tilde{\beta}_n = \hat{\beta}_n + \frac{\hat{\sigma}^2}{2} V_n^{-1} k_n.$$

2) Define the more precise estimators of $F_n$ and $\sigma^2$,

$$\tilde{F}_n = \frac{1}{n} \sum_{i=1}^{n} F(x_i, \hat{\beta}) - \frac{\hat{\sigma}^2}{2 n} \sum_{i=1}^{n} F^{(2)}(x_i, \hat{\beta}),$$

$$\tilde{\sigma}^2 = Q(\hat{\beta}) + \frac{Q(\hat{\beta})^2}{4} \left( \hat{k}_n \hat{V}_n^{-1} k_n^T + \frac{1}{n} \sum_{i=1}^{n} \left( \frac{g^{(\xi)^2}}{(1 + (g^{(\xi)})^2)^2} \right)_{(x_i, \hat{\beta})} \right).$$

Let $\tilde{z}_n = -\frac{1}{2} \tilde{V}_n^{-1} k_n^T, \quad \tilde{z}_n = -\frac{1}{2} \tilde{V}_n^{-1} k_n^T, \quad \tilde{a}_n = -\frac{1}{2} \tilde{V}_n^{-1} \left( \tilde{p}_n + 2 \tilde{W}_n \tilde{z}_n + 3 \tilde{T}_n \tilde{z}_n^2 \right).$

A more precise estimator of $\beta_0$ is

$$\tilde{\beta}_n = \hat{\beta} - \tilde{\sigma}^2 \tilde{z}_n - \tilde{\sigma}^4 \tilde{a}_n.$$

**Theorem 3.** Suppose that conditions $(i)$–$(v)$ and (con) hold for model (1). Then

1) $\tilde{\beta}_n - \beta_0 = \sigma^2 \hat{\alpha}_o P(1),$ 
2) $\tilde{\beta}_n - \beta_0 = \sigma^4 \hat{\alpha}_o P(1).$
Proof. We start with some auxiliary statements. In these statements, we suppose that the conditions of Theorem 3 hold, the function $F$ is three times differentiable, and, for some positive $C, k$, the inequality

$$
\|F^\beta(\xi, \beta)\| + \|F^{\beta \xi \xi}(\xi, \beta)\| \leq C(1 + |\xi|^k), \ \xi \in \mathbb{R}, \ \beta \in \Theta,
$$

holds. We normalize the error terms to obtain standard normal variables:

$$
\tilde{\varepsilon}_i = \varepsilon_i / \sigma, \ \tilde{\delta}_i = \delta_i / \sigma.
$$

Proposition 1. $\hat{\sigma}^2 - \sigma^2 = \sigma^4 \hat{O}_{\sigma P}(1), \ \hat{\sigma}^4 - \sigma^4 = \sigma^6 \hat{O}_{\sigma P}(1)$.

Proof. Theorem 1 states that $Q(\beta_0) = \sigma^2 + \sigma^4 \hat{O}_{\sigma P}(1)$, and formula (11) from Appendix implies that $Q(\hat{\beta}) - Q(\beta_0) = \sigma^4 \hat{O}_{\sigma P}(1)$. Hence, we obtain

$$
\sigma^2 - \hat{\sigma}^2 = Q(\beta_0) + \sigma^4 \hat{O}_{\sigma P}(1) - Q(\hat{\beta}) = \sigma^4 \hat{O}_{\sigma P}(1),
$$

and

$$
\hat{\sigma}^4 - \sigma^4 = (\hat{\sigma}^2 - \sigma^2)(\hat{\sigma}^2 + \sigma^2) = \sigma^4 \hat{O}_{\sigma P}(1)(2\sigma^2 + (\hat{\sigma}^2 - \sigma^2)) = \sigma^6 \hat{O}_{\sigma P}(1). \ \square
$$

Proposition 2. $\hat{F}_n = F_n + \delta_{\sigma P}(1)$.

Proof. $\hat{F}_n - F_n = \frac{1}{n} \sum_{i=1}^{n} F(x_i, \hat{\beta}_n) - F(x_i, \beta_0) + \frac{1}{n} \sum_{i=1}^{n} F(x_i, \beta_0) - F(\xi, \beta_0) =: r_1 + r_2.$

1) $r_1 = \frac{1}{n} \sum_{i=1}^{n} F^\beta(x_i, \hat{\beta}_n)(\hat{\beta}_n - \beta_0) = \sigma^2 \delta_{\sigma P}(1) + \frac{1}{n} \sum_{i=1}^{n} F^\beta(x_i, \hat{\beta}_n), \ \hat{\beta}_n \in [\beta_0, \hat{\beta}_n]$, and

$$
\left\| \frac{1}{n} \sum_{i=1}^{n} F^\beta(x_i, \hat{\beta}_n) \right\| \leq \frac{C}{n} \sum_{i=1}^{n} (|\xi_i|^k + 1) \leq \frac{C}{n} \sum_{i=1}^{n} (|\xi_i|^k + \sigma^k |\tilde{\varepsilon}_i|^k) \leq C(1 + 2^{k-1} \sigma^k + 2^{k-1} \sigma^k O_{\sigma P}(1)) = O_{\sigma P}(1).
$$

2) $r_2 = \frac{1}{n} \sum_{i=1}^{n} F^\xi(\tilde{\xi}_i, \beta_0) \cdot \tilde{\varepsilon}_i = \sigma \cdot \frac{1}{n} \sum_{i=1}^{n} F^\xi(\tilde{\xi}_i, \beta_0) \tilde{\varepsilon}_i = \sigma O_{\sigma P}(1)$, similarly to $r_1$.

Proof of Statement 1) of Theorem 3. Theorem 2 states that

$$
\hat{\beta} - \beta_0 = \frac{\hat{\sigma}^2}{2} \hat{V}_n^{-1} \hat{k}_n^T - \frac{\sigma^2}{2} V_n^{-1} k_n^T + \sigma^2 \delta_{\sigma P}(1).
$$

The functions $k(\xi, \beta)$ and $V(\xi, \beta)$ satisfy inequality (4). Hence,

$$
\frac{\hat{\sigma}^2}{2} \hat{V}_n^{-1} \hat{k}_n^T - \frac{\sigma^2}{2} V_n^{-1} k_n^T = \sigma^2 \delta_{\sigma P}(1). \ \square
$$

Proposition 3. $\hat{F}_n - F_n = \sigma^2 \delta_{\sigma P}(1)$.

Proof. By the Taylor expansion,

$$
\frac{1}{n} \sum_{i=1}^{n} F(x_i, \hat{\beta}) - \frac{1}{n} \sum_{i=1}^{n} F(\xi, \beta_0) = \frac{1}{n} \sum_{i=1}^{n} \left( F(x_i, \hat{\beta}) - F(x_i, \beta_0) \right) + \frac{1}{n} \sum_{i=1}^{n} \left( F(x_i, \beta_0) - F(\xi, \beta_0) \right) =: A_1 + A_2.
$$

We have $A_1 = \frac{1}{n} \sum_{i=1}^{n} F^\beta(x_i, \hat{\beta}_n)(\hat{\beta}_n - \beta_0) = \sigma^4 \delta_{\sigma P}(1) \cdot \frac{1}{n} \sum_{i=1}^{n} F^\beta(x_i, \hat{\beta}_n)$, where

$$
\frac{1}{n} \sum_{i=1}^{n} F^\beta(x_i, \hat{\beta}_n) = O_{\sigma P}(1) \text{ similarly to } r_1 \text{ from Proposition 2}.
$$

Next,

$$
A_2 = \frac{1}{n} \sum_{i=1}^{n} F^\xi(\xi_i, \beta_0) \cdot \varepsilon_i + \frac{1}{2n} \sum_{i=1}^{n} F^{\xi \xi}(\xi_i, \beta_0) \cdot \varepsilon_i^2 + \frac{1}{6n} \sum_{i=1}^{n} F^{\xi \xi \xi}(\xi_i, \beta_0) \cdot \varepsilon_i^3 =: R_1 + R_2 + R_3.
$$
where $R_1 = \frac{1}{n} \sum_{i=1}^{n} F^\xi (\xi_i, \beta_0) \cdot \varepsilon_i = 0 + \frac{\sigma_P}{\sqrt{\nu}} \cdot O_P(1) = \sigma^2 \sigma_\delta P(1)$;

$R_2 = \frac{1}{n} \sum_{i=1}^{n} F^\xi (\xi_i, \beta_0) \cdot \varepsilon_i = \frac{\sigma^2}{\sqrt{n}} O_P(1) = \frac{\sigma^2}{\sqrt{n}} F^\xi (\xi_i, \beta_0) + \sigma^2 \sigma_\delta P(1)$;

$s(\beta) = \frac{\sigma^2}{\sqrt{n}} F^\xi (\xi_i, \beta_0) + \sigma^2 \sigma_\delta P(1)$.

Summarizing we have

$\frac{1}{n} \sum_{i=1}^{n} F(x_i, \beta) = F_n + \frac{\sigma^2}{\sqrt{n}} F^\xi (\xi_i, \beta_0) + \sigma^2 \sigma_\delta P(1) = F_n + \frac{\sigma^2}{\sqrt{n}} F^\xi (\xi_i, \beta_0) + \sigma^2 \sigma_\delta P(1).$  □

**Proposition 4.** $\sigma^2 = \sigma^2 = \sigma^2 \sigma_\delta P(1)$.

**Proof.** Formulae (11) and $\Delta \hat{\phi} - z_n = \sigma_\delta P(1)$ from the proof of Theorem 2 (see below) imply that $Q(\hat{\beta}) - Q(\beta_0) = \sigma^4 \sum_{i=1}^{n} (1 + \frac{\sigma^2}{\sqrt{n}} F^\xi (\xi_i, \beta_0)) \varepsilon_i^2 = \frac{\sigma^2}{\sqrt{n}} F^\xi (\xi_i, \beta_0) + \sigma^2 \sigma_\delta P(1)$, whence

$Q(\hat{\beta}) = Q(\hat{\beta}) - \sigma^4 \sum_{i=1}^{n} (1 + \frac{\sigma^2}{\sqrt{n}} F^\xi (\xi_i, \beta_0)) \varepsilon_i^2 = \frac{\sigma^2}{\sqrt{n}} F^\xi (\xi_i, \beta_0) + \sigma^2 \sigma_\delta P(1)$.

We replace $\sigma^2, k_n$ and $V_n$ by their estimators. Then, by Propositions 1 and 2,

$Q(\hat{\beta}) = \hat{\sigma}^2 + \frac{\sigma^2}{\sqrt{n}} k_n \cdot \hat{V}_n^{-1} k_n^T + \sigma^2 \sigma_\delta P(1)$.

and the second formula from the condition of Theorem 1 takes the form

$Q(\hat{\beta}) = 2 \hat{\sigma}^2 + \frac{\sigma^2}{\sqrt{n}} k_n \cdot \hat{V}_n^{-1} k_n^T + \sigma^2 \sigma_\delta P(1)$.

From the last two expansions, we obtain

$\hat{\sigma}^2 = \frac{\sigma^2}{\sqrt{n}} k_n \cdot \hat{V}_n^{-1} k_n^T + \frac{\sigma^2}{\sqrt{n}} \sum_{i=1}^{n} \frac{(g^\xi)^2}{(1 + (g^\xi)^2)^{\frac{3}{2}}}(x_i, \beta_0) + \sigma^2 \sigma_\delta P(1)$.

□

**Proof of Statement 2) of Theorem 3.**

Theorem 2 states that $\beta_0 = \hat{\beta} = \sigma^2 z_n - \sigma^4 a_n + \sigma^4 \sigma_\delta P(1)$. The functions $k, V, p, W,$ and $T$ satisfy inequality (6). Therefore, in view of Propositions 2 and 4, $\hat{a}_n - a_n = \sigma_\delta P(1)$, $\hat{z}_n - z_n = \sigma^2 \sigma_\delta P(1)$, and $\hat{\sigma}^2 \hat{z}_n - \sigma^2 \hat{z}_n = \sigma^2 \sigma_\delta P(1)$. Then we obtain

$\hat{\beta}_n - \beta_0 = \hat{\sigma}^2 \hat{z}_n - \sigma^2 \hat{z}_n - \sigma^4 \hat{a}_n - \sigma^4 a_n = \sigma^4 \sigma_\delta P(1)$.

Theorem 3 is proved. □

**APPENDIX**

**Proof of Theorem 1.**

1) **Proof of (3).** We show that $\sigma^2 = \sigma^2 \sigma_\delta P(1)$. Consider a component of this sum:

$q(x, y, \beta) = (y - g(h(x, y, \beta)))^2 + (x - h(x, y, \beta))^2 \leq \frac{1}{n} \sum_{i=1}^{n} (I(\varepsilon_i \geq \nu) + I(\delta_i \geq \nu)] = \frac{\sigma^2}{\sqrt{n}} \sum_{i=1}^{n} (2\delta^2 + \epsilon^2 + \text{const}) \cdot I(\varepsilon_i \geq \nu) + I(\delta_i \geq \nu/\sigma)$.

Let us consider the expectations of the terms in the former expression by using the following inequality: $1 - F_X(x) \leq \frac{1}{x} f_X(x)$, $x > 0$, where $f_X$ and $F_X$ are, respectively, the
standard normal density and the normal distribution function. Hence, \( P(|\xi| \geq \nu / \sigma) = 2 \left( 1 - F_N(\nu / \sigma) \right) \leq \frac{2 \alpha}{\nu} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{\nu^2}{2\sigma^2}} \) and then

\[
\sigma^2 E I (|\xi| \geq \nu / \sigma) \cdot \varepsilon^2 \leq \sigma^2 \cdot \frac{6 \sigma}{\nu} \cdot (2\pi)^{-1/4} e^{-\frac{\nu^2}{4\sigma^2}}.
\]

Similar inequalities can be obtained for other terms, and we have finally

\[
E Q_2(\beta) \leq C \sigma e^{-\frac{\nu^2}{2\sigma^2}} = C_1 \sigma^8 o(1), \text{ as } \sigma \to 0+.
\]

Hence, by the Chebyshev inequality

\[
P \left( \frac{Q_2(\beta)}{\sigma^8} > C \right) \leq \frac{E Q_2(\beta)}{\sigma^8 C} = o(1) \to 0 \text{ as } \sigma \to 0+,
\]

and \( Q_2(\beta) = \sigma^8 o \sigma^P(1) \), where \( \sigma^8 \) can be replaced by any positive degree of \( \sigma \).

2) Now consider the case \( i \in B_n(\nu) \). We denote \( h_i = h(x_i, y_i, \beta_0) \). We omit the index \( i \) for the terms \( x_i, y_i, \varepsilon_i, \delta_i, h_i \). All these terms belong to a compact set for all \( i \in B_n(\nu) \). Introduce \( \Delta = h - \xi \). Note that \( \Delta = O (|\varepsilon| + |\delta|) \). Indeed,

\[
\Delta^2 = (\xi - h)^2 \leq 2 \big((\xi - x)^2 + (x - h)^2\big) \leq 2 \varepsilon^2 + 2 \big((y - g(h, \beta_0))^2 + (x - h)^2\big) \leq
\]

\[
\leq 2 \varepsilon^2 + 2 \big((y - g(\xi, \beta_0))^2 + (x - \xi)^2\big) = 4 \varepsilon^2 + 2 \delta^2.
\]

We write down the Taylor expansion for the regression function \( g \). When some function is taken at the point \((\xi, \beta_0)\), we write it without the argument. Then

\[
g(h, \beta_0) = g + g^\xi \Delta + \frac{1}{2} g^{\xi \xi} \Delta^2 = g + g^\xi \Delta + a_0 \Delta^2,
\]

\[
g^\xi(h, \beta_0) = g^\xi + g^{\xi \xi} \Delta = g^\xi + 2a_0 \Delta,
\]

\[
g^{\xi \xi}(h, \beta_0) = g^{\xi \xi} + g^{\xi \xi \xi} \Delta^2, \quad g^{\xi \xi}(h, \beta_0) = g^{\xi \xi} + g^{\xi \xi \xi} \Delta.
\]

(8)

We substitute it into (8) and obtain the equation for \( \Delta \):

\[
(\delta - g^\xi \Delta - a_0 \Delta^2) \left(g^{\xi} + 2a_0 \Delta\right) + \varepsilon - \Delta = 0
\]

(9)

with the unknown parameters \( \xi \) and \( \beta_0 = (a_0, b_0, c_0)^T \). The equation has a unique solution \( \Delta = \Delta(\varepsilon, \delta) \), for any \( i \in B_n(\nu) \). The function \( \Delta = \Delta(\varepsilon, \delta) \) is infinitely differentiable for \( |\varepsilon| < \nu, |\delta| < \nu \), and we can find its Taylor expansion. For an arbitrary function \( s(\varepsilon, \delta) \), we denote the \( k \)-th term of the expansion by \( s_k \). Then

\[
\Delta = \Delta_1 + \ldots + \Delta_6 + O \left(|\varepsilon|^7 + |\delta|^7\right), \quad \Delta_k = \sum_{i+j=k} c^{(k)}_{ij} \varepsilon^i \delta^j.
\]

Here, \( \Delta_k \) is a polynomial of \( \varepsilon \) and \( \delta \) with the coefficients depending of \( g^\xi \) and \( a_0 \). Substituting (8) in (9), we find \( \Delta_k \) as

\[
(\delta g^\xi + \varepsilon) + 2a_0 \Delta \delta - 3a_0 g^\xi \Delta^2 - 2a_0^2 \Delta^3 = (g^\xi)^2 + 1 \Delta \implies
\]

\[
\Delta = \frac{(\delta g^\xi + \varepsilon) + 2a_0 \Delta \delta - 3a_0 g^\xi \Delta^2 - 2a_0^2 \Delta^3}{(g^\xi)^2 + 1}.
\]

Hence,

\[
\Delta_1 = \frac{\delta g^\xi + \varepsilon}{(g^\xi)^2 + 1}, \quad \Delta_2 = \frac{2a_0 \Delta_1 \delta - 3a_0 g^\xi \Delta_1^2}{(g^\xi)^2 + 1},
\]

\[
\Delta_3 = \frac{2a_0 \Delta_2 \delta - 6a_0 g^\xi \Delta_1 \Delta_2 - 2a_0^2 \Delta_2^2}{(g^\xi)^2 + 1},
\]

\[
\Delta_4 = \frac{2a_0 \Delta_3 \delta - 3a_0 g^\xi (\Delta_2^2 + 2\Delta_1 \Delta_3) - 6a_0^2 \Delta_2 \Delta_3}{(g^\xi)^2 + 1}.
\]

Similarly, one can find \( \Delta_5 \) and \( \Delta_6 \).
3) We now find the Taylor expansions of \( q, q^2, q^{3}\beta \) and \( q^{3}\beta \xi \) at the point \((x, y, \beta_0)\) as functions of \( \varepsilon \) and \( \delta \) and their expectations. The expectations of odd terms are zeros, and those of even terms are certain functions of \( \sigma, \xi, \beta_0 \).

a) Consider \( q(x, y, \beta_0) \):

\[
q(x, y, \beta_0) = (y - g(h, \beta_0)) + (x - h)^2 = (\delta - g^\varepsilon \Delta - \frac{1}{2}g^\varepsilon \Delta^2) + (\varepsilon - \Delta)^2 = q_2(\varepsilon, \delta) + q_3(\varepsilon, \delta) + O(|\varepsilon|^{5} + |\delta|^{5}),
\]

where

\[
q_2(\varepsilon, \delta) = (\delta - g^\varepsilon \Delta)^2 + (\varepsilon - \Delta)^2,
\]

\[
q_4(\varepsilon, \delta) = (g^\varepsilon \Delta_2 + \frac{1}{2}g^\varepsilon \Delta^2)^2 - 2(\delta - g^\varepsilon \Delta_1)(g^\varepsilon \Delta_3 + g^\varepsilon \Delta_1 \Delta_2).
\]

The expectations of these terms are \( E \ q_2(\varepsilon, \delta) = \sigma^2, E \ q_4(\varepsilon, \delta) = - \frac{\sigma^2}{4}(g^\varepsilon)^2 (1 + (g^\varepsilon)^2)^{-3} \).

b) Consider \( q^3(x, y, \beta_0) \).

\[
q^3(x, y, \beta) = G^3(x, y, \beta, u)|_{u=h} + G^u(x, y, \beta, u)|_{u=h} \cdot h^3(x, y, \beta) = -2(y - g(h, \beta)) g^\beta(h, \beta), \quad \text{because } G^u|_{u=h} = 0.
\]

From (8), we obtain

\[
q^3(x, y, \beta_0) = (\delta - g^\varepsilon \Delta - \frac{1}{2}g^\varepsilon \Delta^2) \left(g^\beta + g^\varepsilon \Delta + \frac{1}{2}g^\varepsilon \Delta^2\right), \quad \text{whence}
\]

\[
q^3_2 = (2g^\varepsilon \Delta_2 + \frac{1}{2}g^\varepsilon \Delta^2) g^\beta - 2(\delta - g^\varepsilon \Delta_1) g^\varepsilon \Delta^3;
\]

\[
q^3_4 = (2g^\varepsilon \Delta_2 + g^\varepsilon \Delta_3 + \frac{1}{2}g^\varepsilon \Delta^2_2 + 2\Delta_1 \Delta_3) g^\beta + (3g^\varepsilon \Delta^2_1 \Delta_2 + 2g^\varepsilon \Delta^3 - 2(\delta - g^\varepsilon \Delta_1) g^\varepsilon \Delta^3 + (\frac{1}{2}g^\varepsilon \Delta^2_1 - \Delta_1 \Delta_2) (2\delta - 3g^\varepsilon \Delta_1) g^\varepsilon \Delta^5).
\]

The expectations of the expansion terms are \( E \ q^3_2(\varepsilon, \delta) = \sigma^2 k(\xi, \beta_0), E \ q^3_4(\varepsilon, \delta) = \sigma^4 p(\xi, \beta_0). \)

c) Consider \( q^{3}\beta(x, y, \beta_0) \).

\[
q^{3}\beta(x, y, \beta) = G^{3}\beta + (G^{3}\beta u)^T T h^3 = G^{3}\beta - (G^{3}\beta u)^T T G^{3}\beta u \cdot (G^{3}\beta u)^{-1} h^3 = -(G^{3}\beta u)^{-1} G^{3}\beta u.
\]

We write down the expansion terms of \( G^{3}\beta(x, y, \beta_0, h) \):

\[
G^{3}\beta_0 = 2g^\varepsilon T g^\beta, \quad G^{3}\beta^2 = 2g^\varepsilon T g^\beta T g^\varepsilon + 4\Delta_2 \left( g^\beta * g^\varepsilon \right) + 2\Delta_1 \left( g^\beta * g^\varepsilon \right),
\]

\[
G^{3}\beta^3 = 2g^\varepsilon \Delta_2 g^\beta + (g^\beta g^\varepsilon \Delta_2 + \frac{1}{2}g^\varepsilon \Delta^2) g^\beta - (2\delta - 3g^\varepsilon \Delta_1) g^\varepsilon \Delta^3,
\]

\[
G^{3}\beta^4 = 2(1 + (g^\beta)^2), \quad G^{3}\beta^5 = 2g^\varepsilon \Delta_2 (3g^\varepsilon \Delta_1 - \delta), \quad G^{3}\beta^6 = 6g^\varepsilon \Delta_2 + \frac{1}{2}g^\varepsilon \Delta^2.
\]

Then the Taylor expansion of \( (G^{3}\beta u)^{-1} \) is

\[
\left((G^{3}\beta u)^{-1}\right)^{-1} = \frac{1}{\sigma^2_0} \left(1 - \left(G^{3}\beta u / \sigma^2_0 + G^{3}\beta u / \sigma^2_0\right)^2 + O(|\varepsilon|^3 + |\delta|^3)\right),
\]

whence three first terms of the expansion are

\[
(G^{3}\beta u)^{-1} = -(G^{3}\beta u)^{-1}, \quad (G^{3}\beta u)^{-1} = -\frac{1}{\sigma^2_0} \left((G^{3}\beta u / \sigma^2_0)^2 - G^{3}\beta u / \sigma^2_0\right).
\]

The terms of degrees 0 and 2 for \( q^{3}\beta(x, y, \beta_0) \) are as follows:

\[
q^{3}\beta_0 = G^{3}\beta_0 - \frac{1}{\sigma^2_0} (G^{3}\beta u)^T G^{3}\beta u = 2 \left(1 - \frac{(g^\beta)^2}{T \sigma^2_0}\right) g^\beta T g^\beta - \frac{2}{T \sigma^2_0} g^\beta T g^\beta = 2V(\xi, \beta_0);
\]

\[
q^{3}\beta_2 = G^{3}\beta^2 - (G^{3}\beta u)^T G^{3}\beta u + G^{3}\beta^2 T G^{3}\beta u + G^{3}\beta^2 u T G^{3}\beta u \cdot (G^{3}\beta u)^{-1} - (G^{3}\beta u)^{-1} G^{3}\beta u T G^{3}\beta u \cdot (G^{3}\beta u)^{-1}.
\]

The expectation of \( q^{3}\beta_2 \) is \( E \ q^{3}\beta_2(\varepsilon, \delta) = 2\sigma^2 W(\xi, \beta_0) \).
d) Consider $q^{333}(x, y, \beta, \nu)$. The third order derivatives of $G$ at the point $(x, y, \beta, u)$ are as follows:

$$
G^{333} = 0, \quad G^{33u} = 2 \left( g^3 y g^3 + g^3 T g^3 \right) |_{(u, \beta)}, \quad G^{uuu} = (6 g^3 y g^3) |_{(u, \beta)},
$$

$$
G^{3au} = (2 g^3 y g^3 + 2 y g^3 g^3 - (y - g) g^3 g^3) |_{(u, \beta)}, \quad h^{33} = -(G^{uuu})^{-1} (G^{333} + 2 G^{3au} h^3 + G^{uuu} (h^3 T h^3)),
$$

as a derivative of the implicit function. Differentiating the function $q^{333}(x, y, \beta) = G^{333} + G^{33u} h^3$, we have

$$
q^{333} = G^{333} + G^{33u} * h^3 + h^3 * G^{33u} * h^3 + G^{3u} * h^{33} = -3(G^{uuu})^{-1} (G^{333} * G^{3u}) +
$$

$$
+ 3(G^{uuu})^{-2} (G^{3uu} * G^{3u} * G^{3u}) - (G^{uuu})^{-3} G^{uuu} (G^{3u} + G^{3u} * G^{3u}).
$$

It can be easily found from the above-written that

$$
q^{333}_0 = \frac{6 g^3 y (g^3)^2}{(1 + (g^3)^2)^2} g^3 * g^3 * g^3 - \frac{12 g^6}{(1 + (g^3)^2)^2} g^3 * g^3 * g^3 = 6 T(\xi, \beta_0).
$$

(Remember that the notation $\tilde{a} * \tilde{b} * \tilde{c}$ was given just after Lemma 1).

4) Proof of the statements of Theorem 1. The first of them has been already proved, and the rest ones are easily inferred from the formulas stated above. We derive expansion (4) for $Q_1^3(\beta_0)$:

$$
Q_1^3(\beta_0) = \sum_{i \in B_n(\nu)} q^{0}(x_i, y_i, \beta_0) = \sum_{i \in B_n(\nu)} A_k + \sum_{i \in B_n(\nu)} O((\varepsilon_1^2 + |\delta_1|^2), A_k = \sum_{i \in B_n(\nu)} q^{0}(\varepsilon_i, \delta_i).
$$

We use Lemma 1 several times. Start with $A_1$. The term $q^{0}_1(\varepsilon, \delta)$ is a linear form of $\varepsilon$ and $\delta$ with bounded coefficients. It follows from (ii) that, for arbitrary $i, 1 \leq i \leq n, q^{0}(\varepsilon_i, \delta_i) = -2 (\delta_i - g^3 \Delta_1(\varepsilon_i, \delta_i)) g^3 = \sigma^6 \frac{2(g^3 \varepsilon_i - \delta_i)}{(g^3)^2 + 1} g^3.
$$

To apply Lemma 1, we divide $A_1$ into two sums:

$$
A_1 = \frac{1}{n} \sum_{i=1}^{n} q^{0}_1(x_i, y_i, \beta_0) - \frac{1}{n} \sum_{i \in B_n(\nu)} q^{0}_1(x_i, y_i, \beta_0) =: S_1 - S_2,
$$

where $S_1 = \frac{1}{n} \sum_{i=1}^{n} O_P(1)$, and $S_2 = \sigma^6 \sigma_P(1)$ like a sum in 1). Other $A_k$ can be expanded in a similar way. Consider $A_6$ separately. Introduce

$$
R_1 = \sigma^{-6} \frac{1}{n} \sum_{i=1}^{n} E q^{0}_6(\varepsilon_i, \delta_i) = \frac{1}{n} \sum_{i=1}^{n} \sum_{l=0}^{6} E \varepsilon_i^l v_i^6 \delta_i^{6-l}.
$$

It is a bounded nonrandom vector depending only on $(\xi, \beta_0)$. Dividing $A_6$ into two sums similarly to $A_1$, we obtain

$$
A_6 = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^6(\varepsilon_i, \delta_i) + \sigma^6 \sigma_P(1) = \sigma^6 \left( R_1 + \frac{O_P(1)}{\sqrt{n}} \right) + \sigma^6 \sigma_P(1).
$$

Now consider the last term. We get

$$
\left\| \frac{1}{n} \sum_{i \in B_n(\nu)} O(|\varepsilon_i|^2 + |\delta_i|^2) \right\| \leq const \frac{\sigma^7}{n} \sum_{i=1}^{n} (|\varepsilon_i|^2 + |\delta_i|^2) = \sigma^7 O_P(1).
$$

The expansion of $Q_1^3(\beta_0)$ follows from the preceding formulas. In a similar way, we can obtain the expansions for $Q(\beta_0), Q_1^{33}, Q_1^{333},$ and $Q^{3333}$. The last two derivatives are considered as the matrices corresponding to tri- and four-linear forms.□

Proof of Theorem 2.

1) We find an expansion of $Q(\beta)$. Consider the case where $\hat{\beta}_n \in U_n(\beta_0)$. It occurs for some $\sigma \leq \sigma_c, n \geq n_c$ with probability at least $1 - \varepsilon$, where $\varepsilon$ can be an arbitrary positive quantity.
2) Write the Taylor expansion of $Q_1(\beta)$ in $\Delta \beta = \beta - \beta_0$:

$$Q_1(\beta) = Q_1(\beta_0) + \frac{\partial Q_1(\beta_0)}{\partial \beta} (\Delta \beta) + \frac{1}{2!} \frac{\partial^2 Q_1(\beta_0)}{\partial \beta^2} (\Delta \beta)^2 + \frac{1}{3!} \frac{\partial^3 Q_1(\beta_0)}{\partial \beta^3} (\Delta \beta)^3 + \cdots$$

where

$$\frac{\partial Q_1(\beta)}{\partial \beta} = \sigma^2 \frac{\partial^2 Q_1(\beta)}{\partial \beta^2}, \quad \tilde{\beta} \in [\beta_0, \beta].$$

The derivative $\frac{\partial Q_1(\beta)}{\partial \beta}$ is bounded because all the partial derivatives of $Q_1(\beta)$ are bounded. Take the expansions from Theorem 1 and denote $\Delta \varphi = \sigma^2 \Delta \hat{\beta}$. We obtain

$$Q(\beta) - Q(\beta_0) = Q_1(\beta) - Q_1(\beta_0) + \sigma^2 o_{\sigma P}(1) = \sigma^4 (k_n \Delta \varphi + V_n(\Delta \varphi)^2) +$$

$$+ \sigma^6 (p_n \Delta \varphi + V_n(\Delta \varphi)^2 + T_n(\Delta \varphi)^3) + \sigma^8 R(\varphi) + rest(\Delta \varphi),$$

where $R(\varphi) = R_1(\Delta \varphi) + R_2(\Delta \varphi)^2 + R_3(\Delta \varphi)^3 + R_4(\Delta \varphi)^4$,

$$rest = \sigma^8 (1 + \|\Delta \varphi\|^4) \sigma_{\sigma P}(1) + \|\Delta \varphi\|^2 \|\Delta \beta\|O(1)).$$

Let $\Delta \hat{\beta} = \sigma^2 (\hat{\beta} - \beta_0) = \sigma^2 \Delta \hat{\beta}$. Since $\Delta \beta = \sigma_{\sigma P}(1)$, relation (10) yields that

$$Q(\hat{\beta}) - Q(\beta_0) = k_n \Delta \hat{\beta} + V_n(\Delta \hat{\beta})^2 + \sigma^2 \sigma_{\sigma P}(1) \leq 0. \quad (11)$$

Let $c = \lim \inf_{n \to \infty} \lambda_{\min}(V_n) > 0, \ c > 0$, as follows from (v). Then, for $n \geq n_0, V_n \Delta \varphi^2 \geq \frac{\sigma^4}{c} \|\Delta \varphi\|^2$ and $\forall \epsilon > 0 \exists \sigma > 0 \forall \varphi \in (0, \sigma) \exists m_{n, \sigma} \forall n \geq m_{n, \sigma}, P(|\sigma_{\sigma P}(1)| < c/4) > 1 - \epsilon.$

Then (10) implies that, for $\sigma \in (0, \sigma_{\sigma P}(1)], n \geq m_{n, \sigma}$,

$$\frac{\sigma^4}{c} \|\Delta \varphi\|^2 + k_n \Delta \hat{\beta} + \sigma^2 \sigma_{\sigma P}(1) \leq 0$$

with probability at least $1 - \epsilon$. This implies $\Delta \hat{\beta} = \sigma_{\sigma P}(1)$.

3) Write expansions (10) for $\Delta \hat{\beta}$ and $\Delta \varphi = \epsilon$ and subtract them. We recall that $\Delta \hat{\beta} = \sigma_{\sigma P}(1)$ and $\epsilon$ are some nonrandom bounded vectors. We obtain

$$Q(\epsilon) - Q(\beta_0 + \sigma^2 \epsilon) = k_n \Delta \hat{\beta} + V_n(\Delta \hat{\beta})^2 - k_n \epsilon_n - V_n(\epsilon_n)^2 + \sigma^2 \sigma_{\sigma P}(1) =$$

$$= V_n(\Delta \hat{\beta} - \epsilon_n)^2 + (2V_n \epsilon_n + k_n) (\Delta \hat{\beta} - \epsilon_n) + \sigma^2 \sigma_{\sigma P}(1) \leq 0. \quad (12)$$

Let $\epsilon_n = -\frac{1}{2} V_n^{-1} k_n T_n$. Then (12) changes into $V_n(\Delta \hat{\beta} - \epsilon_n)^2 = \sigma^2 \sigma_{\sigma P}(1)$, and condition (v) implies that $\Delta \hat{\beta} - \epsilon_n = \sigma_{\sigma P}(1) = \sigma_{\sigma P}(1)$.

4) Let $\Delta \varphi = \epsilon + t, \Delta \hat{\beta} = \epsilon + t$, where $t = \sigma \sigma_{\sigma P}(1)$. Subtract expansions (10) for $\epsilon_n$ and $\Delta \varphi = \epsilon + t$.

$L(t) := \sigma^{-4} \left[ Q(\beta_0 + \sigma^2 \epsilon_n + t) - Q(\beta_0 + \sigma^2 \epsilon) \right] = \sigma^{-4} \left[ k_n + V_n(\epsilon_n + t)^2 - z_n^2 + \sigma^2 \left[ p_n t + W_n(\epsilon_n + t)^2 - z_n^2 \right] + T_n(\epsilon_n + t)^3 - 3z_n^3 + \sigma^4(R(t) \epsilon_n + t) - R(\epsilon_n) \right] + \sigma^{-4} \left[ rest(t) + \sigma^2 rest(z) \right]$$

where $j_n := p_n + 2W_n \epsilon_n + 3T_n \epsilon_n^2 + T_n \epsilon_n^3$. Since $t = \sigma_{\sigma P}(1)$, we have $rest(t) + \sigma^2 rest(z) = \sigma^4 \sigma_{\sigma P}(1)$, $R(t) - R(\epsilon_n) = \sigma_{\sigma P}(1)$ by the definition of $R$. Then $L(t)$ has an expansion

$$L(t) = V_n t^2 + \sigma^2 j_n t + \sigma^2 (W_n t^2 + 3T_n \epsilon_n \epsilon_n^2 + T_n t^3) + \sigma^4 \sigma_{\sigma P}(1). \quad (13)$$

Prove that $\Delta \hat{\beta} = \sigma_{\sigma P}(1)$, i.e., $t = \sigma^2 \epsilon$, where $\epsilon = \sigma_{\sigma P}(1)$. We have

$$Q(\epsilon) - Q(\beta_0 + \sigma^2 \epsilon) = L(t) = \sigma^4 [V_n \epsilon^2 + j_n \epsilon + W_n \epsilon \epsilon t + 3T_n \epsilon_n \epsilon_n t + T_n \epsilon_n^2 t] + \sigma^4 \sigma_{\sigma P}(1) \leq 0. \quad (14)$$

Hence,

$$V_n \epsilon^2 \leq -j_n \epsilon + \sigma_{\sigma P}(1) \quad (1 + \|\epsilon\|^2) \leq 0,$$

and we have from (v) that $\epsilon = \sigma_{\sigma P}(1)$.

5) The first leading term of the asymptotic deviation is $\sigma^2 \epsilon_n$. Show that the second leading term is $\sigma^4 a_n$. Note that $a_n$ is a bounded nonrandom vector. Then (13) implies

$$L(\sigma^2 a_n) = \sigma^4 \left( V_n a_n^2 + j_n a_n + \sigma_{\sigma P}(1) \right) \leq 0.$$
From (11) and \( \hat{s} = \tilde{O}_{\sigma P}(1) \), we obtain

\[
L(\sigma^2 \hat{s}) = \sigma^2 \left( V_n \hat{s}^2 + j_n \hat{s} + \tilde{o}_{\sigma P}(1) \right).
\]

Subtracting these equalities, we have

\[
\frac{L(\sigma^2 \hat{s}) - L(\sigma^2 a_n)}{\sigma^4} = V_n (\hat{s}^2 - \tilde{a}_n^2) + j_n (\hat{s} - a_n) + \tilde{o}_{\sigma P}(1) = V_n (\hat{s} - a_n)^2 + (2V_n a_n + j_n) (\hat{s} - a_n) + \tilde{o}_{\sigma P}(1) \leq 0.
\]

Remember that \( a_n = -\frac{1}{2}V_n^{-1}j_n \), then we have the inequality

\[
V_n (\hat{s} - a_n)^2 \leq \tilde{o}_{\sigma P}(1),
\]

whence \( \hat{s} = a_n + \tilde{o}_{\sigma P}(1) \). Theorem 2 is proved. \( \square \)

**Conclusion**

We have found the second term of the asymptotic bias of the orthogonal regression estimator. It is possible to find the subsequent terms in a similar way and then, with sufficiently precise estimates, to construct more accurate estimators. The corrected estimators can be found for any nonlinear regression function in the way like that used in the proof of Theorem 1.

The condition of normality of the error terms \( \varepsilon_i, \delta_i \) is important for calculations. The results can be extended to the non-normal case where the error terms have a symmetric distribution with finite fourth-order moments. The deviation of the proposed estimators is less than the deviation of \( \tilde{\beta} \) for sufficiently small but fixed \( \sigma \) and \( n \to \infty \).

We intend to test the quality of the proposed estimators by simulations and to consider an implicit regression model. In such models, there are no dependent and independent variables, and \( x_i \) and \( y_i \) appear in a symmetric way, see [3].

**Bibliography**


E-mail: galchonok@univ.kiev.ua