Theory of Stochastic Processes
Vol. 11 (27), no. 3–4, 2005, pp. 110–120

UDC 519.21

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.

Introduction

We consider a quadratic functional errors-in-variables model

$$(1)\qquad \begin{aligned} y_i &= a_0\xi_i^2 + b_0\xi_i + c_0 + \delta_i,\\ x_i &= \xi_i + \varepsilon_i, \quad 1 \leqslant i \leqslant n, \end{aligned}$$

where (x_i, y_i), $1 \leqslant i \leqslant n$, are observed, ξ_i are unknown nonrandom parameters, ε_i, δ_i are i.i.d. normal error terms, and the vector $\beta_0 = (a_0,\ b_0,\ c_0)^{\mathrm{T}}$ 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.

1. Model assumptions and orthogonal regression estimator

Let $g(\xi, \beta) = a\xi^2 + b\xi + c$ be a regression function, where $\beta = (a;\ b;\ c)^{\mathrm{T}} \in \Theta$ 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 $g^{\beta} = \left(\xi^2;\ \xi;\ 1\right)$ is the derivative with respect to the vector variable β. The expectation of a random variable ζ is denoted by

2000 *AMS Mathematics Subject Classification.* Primary 62J02; Secondary 62F12, 62H12.
Key words and phrases. Asymptotic bias, bias correction, orthogonal regression estimator, quadratic errors-in-variables regression model, functional model.

$\mathbf{E}\,\zeta$, and its variance is denoted by $\mathbf{D}\,\zeta$. A sequence $\{\chi_n(\theta), n \geqslant 1\}$ 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 $Q(\beta) = \frac{1}{n}\sum_{i=1}^n q(x_i,y_i,\beta)$. Then the orthogonal regression estimator $\hat\beta$ is defined as a measurable solution to the optimization problem:

$$Q(\beta) \text{— min},\ \beta\in\Theta,$$

where Θ is a parameter set.

Assume that the following conditions hold:

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

(ii) $|\xi_i| \leqslant A$, $i \geqslant 1$, where A is unknown.

(iii) $\varepsilon_i, \delta_i \sim N(0,\sigma^2)$ i.i.d., $i \geqslant 1$, 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

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

Hence, h is a solution to the cubic equation $\big(y - ah^2 - bh - c\big)\,(2ah+b) + x - h = 0$. The equation can have from one to three solutions, some of them can not be a global minimum point. Note that $F\left(\xi, g(\xi,\beta),\beta,\xi\right) \equiv 0$ and $F^u\left(\xi, g(\xi,\beta),\beta,\xi\right) = -1 - \left[g^\xi(u,\beta)\right]^2 \neq 0$. Then the Implicit Functions Theorem implies the following: there exists a neighbourhood of a point $(\xi, g(\xi,\beta),\beta)$, $U_\nu(\xi,\beta) := B_\nu(\xi)\times B_\nu\left(g(\xi,\beta)\right)\times B_\nu(\beta)$, $\nu = \nu(\xi,\beta)$, such that $h: U_\nu(\xi,\beta) \to \mathbb{R}$ is a uniquely defined infinitely differentiable function. Since ξ and β 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 ε_i, δ_i is less than ν_0, then there exists only one perpendicular from (x_i,y_i) to any of the curves $y = g(\xi,\beta)$, $\xi\in\mathbb{R}$; $\beta\in U_{\nu_0}(\beta_0)$. Let ν be a fixed positive constant, $\nu\in(0,\nu_0]$, such that $U_\nu(\beta_0)\subset\Theta$. We define the index set $B_n(\nu) = \left\{i = \overline{1,n} : |\varepsilon_i| < \nu, |\delta_i| < \nu\right\}$ 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 $Q_1(\beta_0)$ and its derivatives in σ^2.

We will widely use the following statement.

Lemma 1. *Let $\{\zeta_i : i \geqslant 1\}$ be an i.i.d. sequence with $\mathbf{D}\,\zeta_1 = 1$, and let $\{a_i : i\geqslant 1\}$ be a bounded sequence of real numbers. Then*

$$\frac{1}{n}\sum_{i=1}^n a_i\zeta_i = \frac{\mathbf{E}\,\zeta_1}{n}\sum_{i=1}^n a_i + \frac{1}{\sqrt{n}}O_P(1).$$

The derivatives $g^\beta, g^{\beta\xi}, g^{\beta\xi\xi}$ are row vectors. For a couple of row vectors $\vec a, \vec b$, we define a symmetric matrix $\vec a * \vec b = \frac{1}{2}(\vec a^{\mathrm T}\vec b + \vec b^{\mathrm T}\vec a)$. For a triple $\vec a, \vec b, \vec c$, let $\vec a * \vec b * \vec c$ 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:

$$k(\xi,\beta)=\frac{g^{\xi\xi}}{(1+(g^{\xi})^2)^2}g^{\beta},\qquad V(\xi,\beta)=\frac{1}{1+(g^{\xi})^2}g^{\beta\mathrm{T}}g^{\beta},$$

$$p(\xi,\beta)=\frac{9a_0^3(g^{\xi})^2}{(1+(g^{\xi})^2)^5}g^{\beta}+\frac{3a_0^2g^{\xi}}{(1+(g^{\xi})^2)^4}g^{\beta\xi}-\frac{a_0}{2\,(1+(g^{\xi})^2)^3}g^{\beta\xi\xi},$$

$$W(\xi,\beta)=\frac{(g^{\xi\xi})^2\left(7(g^{\xi})^2-2\right)}{(1+(g^{\xi})^2)^4}g^{\beta\mathrm{T}}g^{\beta}-\frac{8g^{\xi\xi}g^{\xi}}{(1+(g^{\xi})^2)^3}g^{\beta}*g^{\beta\xi}+\frac{1}{(1+(g^{\xi})^2)}g^{\beta}*g^{\beta\xi\xi},$$

$$\mathrm{T}(\xi,\beta)=\frac{g^{\xi\xi}(g^{\xi})^2}{(1+(g^{\xi})^2)^3}g^{\beta}*g^{\beta}*g^{\beta}-\frac{2g^{\xi}}{(1+(g^{\xi})^2)^2}g^{\beta}*g^{\beta}*g^{\beta\xi}.$$

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

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

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

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*

$$Q(\beta)=Q_1(\beta)+\sigma^8 o_{\sigma P}(1), \tag{3}$$

$$Q(\beta_0)=\sigma^2-\frac{\sigma^4}{4n}\sum_{i=1}^{n}\frac{(g^{\xi\xi})^2}{(1+(g^{\xi})^2)^3}|_{(\xi_i,\beta_0)}+\frac{\sigma^2}{\sqrt{n}}O_P(1)+\sigma^4 o_{\sigma P}(1),$$

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

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

$$Q_1^{\beta\beta\beta}(\beta_0)=6\mathrm{T}_n+\sigma^2R_3+\sigma^2o_{\sigma P}(1)+\frac{\sigma}{\sqrt{n}}O_P(1),$$

$$Q_1^{\beta\beta\beta\beta}(\beta_0)=R_4+\sigma^2o_{\sigma P}(1)+\frac{\sigma}{\sqrt{n}}O_P(1),$$

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 $\hat{\beta}-\beta_0$ in powers of σ^2.

2. Asymptotic deviation

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

$$\forall\,\varepsilon>0\;\exists\,C>0:\qquad \lim_{\sigma\to 0+}\limsup_{n\to\infty}P\left(\|\eta_n(\sigma)\|>C\right)<\varepsilon.$$

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

$$\lim_{\sigma\to 0+}\limsup_{n\to\infty}P\left(\|\eta_n(\sigma)\|>C\right)\to 0,\ C\to\infty.$$

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 β, ρ be the Euclidean metrics, and $\rho(M, \Gamma_\beta)$ be the distance between a point M and the plot Γ_β.

We need the following contrast condition:

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

This condition makes it possible to estimate consistently the parameter β_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.* $\forall\, \gamma > 0\ \exists\, \sigma_\gamma > 0\ \exists\, n_\gamma = n_\gamma(\omega)\ \forall\, n \geqslant n_\gamma\, \forall\, \sigma \leqslant \sigma_\gamma : \|\hat\beta_n - \beta_0\| < \gamma$.
This implies that $\hat\beta_n - \beta_0 = \tilde o_{\sigma P}(1)$.

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

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

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

$$\hat\beta_n = \beta_0 + \sigma^2 z_n + \sigma^4 \tilde O_{\sigma P}(1),$$

$$\hat\beta_n = \beta_0 + \sigma^2 z_n + \sigma^4 a_n + \sigma^4 \tilde o_{\sigma P}(1), \tag{5}$$

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

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 $\hat k_n$, $\hat V_n$, etc. For σ^2, we have the estimator $\hat\sigma^2 := Q(\hat\beta)$. Next, we have a new estimator of the parameter β_0:

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

2) Define the more precise estimators of F_n and σ^2,

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

$$\tilde\sigma^2 = Q(\hat\beta) + \frac{Q^2(\hat\beta)}{4}\left(\hat k_n \hat V_n^{-1}\hat k_n^{\mathrm T} + \frac{1}{n}\sum_{i=1}^n \left.\frac{(g^{\xi\xi})^2}{\left(1+(g^\xi)^2\right)^3}\right|_{(x_i,\hat\beta)}\right).$$

Let $\hat z_n = -\frac{1}{2}\hat V_n^{-1}\hat k_n^{\mathrm T}$, $\quad \tilde z_n = -\frac{1}{2}\tilde V_n^{-1}\tilde k_n^{\mathrm T}$, $\quad \hat a_n = -\frac{1}{2}\hat V_n^{-1}\left(\hat p_n + 2\hat W_n \hat z_n + 3\hat{\mathrm T}_n \hat z_n^2\right)$.

A more precise estimator of β_0 is

$$\tilde{\tilde\beta}_n = \hat\beta - \tilde\sigma^2 \tilde z_n - \hat\sigma^4 \hat a_n.$$

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

$$1) \quad \tilde\beta_n - \beta_0 = \sigma^2 \tilde o_{\sigma P}(1),$$

$$2) \quad \tilde{\tilde\beta}_n - \beta_0 = \sigma^4 \tilde o_{\sigma P}(1).$$

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

$$\|F^{\beta}(\xi,\beta)\| + \|F^{\xi\xi\xi}(\xi,\beta)\| \leqslant C(1+|\xi|^k),\ \xi \in \mathbb{R},\ \beta \in \Theta, \tag{6}$$

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

$$\tilde{\varepsilon}_i = \varepsilon_i/\sigma,\ \tilde{\delta}_i = \delta_i/\sigma. \tag{7}$$

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

Proof. Theorem 1 states that $Q(\beta_0) = \sigma^2 + \sigma^4 \tilde{O}_{\sigma P}(1)$, and formula (11) from Appendix implies that $Q(\hat{\beta}) - Q(\beta_0) = \sigma^4 \tilde{O}_{\sigma P}(1)$. Hence, we obtain
$\sigma^2 - \hat{\sigma}^2 = Q(\beta_0) + \sigma^4 \tilde{O}_{\sigma P}(1) - Q(\hat{\beta}) = \sigma^4 \tilde{O}_{\sigma P}(1)$, and
$\hat{\sigma}^4 - \sigma^4 = (\hat{\sigma}^2 - \sigma^2)\left(\hat{\sigma}^2 + \sigma^2\right) = \sigma^4 \tilde{O}_{\sigma P}(1)\left(2\sigma^2 + (\hat{\sigma}^2 - \sigma^2)\right) = \sigma^6 \tilde{O}_{\sigma P}(1).$ □

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

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

1) $r_1 = \frac{1}{n}\sum_{i=1}^n F^{\beta}(x_i,\bar{\beta}_i)(\hat{\beta}_n - \beta_0) = \sigma^2 \tilde{O}_{\sigma P}(1)\frac{1}{n}\sum_{i=1}^n F^{\beta}(x_i,\bar{\beta}_i), \quad \bar{\beta}_i \in [\beta_0,\hat{\beta}_n]$, and

$$\left\|\frac{1}{n}\sum_{i=1}^n F^{\beta}(x_i,\bar{\beta}_i)\right\| \leqslant \frac{C}{n}\sum_{i=1}^n \left(|\xi_i+\varepsilon_i|^k + 1\right) \leqslant C + \frac{C\cdot 2^{k-1}}{n}\sum_{i=1}^n \left(|\xi_i|^k + \sigma^k|\tilde{\varepsilon}_i|^k\right) \leqslant$$
$$\leqslant C\left(1 + 2^{k-1}a^k + 2^{k-1}\sigma^k O_P(1)\right) = O_P(1).$$

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

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

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

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

$$\frac{\hat{\sigma}^2}{2}\hat{V}_n^{-1}\hat{k}_n^{\mathrm{T}} - \frac{\sigma^2}{2}V_n^{-1}k_n^{\mathrm{T}} = \sigma^2\tilde{o}_{\sigma P}(1). \quad □$$

Proposition 3. $\tilde{F}_n - F_n = \sigma^2 \tilde{o}_{\sigma P}(1)$.

Proof. By the Taylor expansion,

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

We have $A_1 = \frac{1}{n}\sum_{i=1}^n F^{\beta}(x_i,\bar{\beta}_i)(\tilde{\beta}_n - \beta_0) = \sigma^4\tilde{O}_{\sigma P}(1)\cdot\frac{1}{n}\sum_{i=1}^n F^{\beta}(x_i,\bar{\beta}_i)$, where $\frac{1}{n}\sum_{i=1}^n F^{\beta}(x_i,\bar{\beta}_i) = O_P(1)$ similarly to r_1 from Proposition 2. Hence, $A_1 = \sigma^2\tilde{o}_{\sigma P}(1)$. Next,

$$A_2 = \frac{1}{n}\sum_{i=1}^n F^{\xi}(\xi_i,\beta_0)\cdot\varepsilon_i + \frac{1}{2n}\sum_{i=1}^n F^{\xi\xi}(\xi_i,\beta_0)\cdot\varepsilon_i^2 +$$
$$+ \frac{1}{6n}\sum_{i=1}^n F^{\xi\xi\xi}(\bar{\xi}_i,\beta_0)\cdot\varepsilon_i^3 =: R_1 + R_2 + R_3,$$

where $R_1 = \frac{1}{n}\sum_{i=1}^n F^{\xi}(\xi_i,\beta_0)\cdot\varepsilon_i = 0 + \frac{\sigma}{\sqrt{n}}O_P(1) = \sigma^2\tilde{o}_{\sigma P}(1)$;
$R_2 = \frac{1}{2n}\sum_{i=1}^n F^{\xi\xi}(\xi_i,\beta_0)\cdot\varepsilon_i^2 = \frac{\sigma^2}{2}F_n^{\xi\xi} + \frac{\sigma^2}{\sqrt{n}}O_P(1) = \frac{\sigma^2}{2}\hat{F}_n^{\xi\xi} + \sigma^2\tilde{o}_{\sigma P}(1)$;
$\|R_3\| \leqslant \frac{C\cdot\sigma^3}{6n}\sum_{i=1}^n\left(1+|\bar{\xi}_i|^k\right)\varepsilon_i^3 \leqslant \frac{C\cdot\sigma^3}{6n}\sum_{i=1}^n\left(1+2^{k-1}(|\xi_i|^k+\sigma^k|\tilde{\varepsilon}_i|^k\right)|\tilde{\varepsilon}_i|^3 = \sigma^3 O_P(1)$.
Summarizing we have
$\frac{1}{n}\sum_{i=1}^n F(x_i,\tilde{\beta}) = F_n + \frac{\sigma^2}{2}\hat{F}_n^{\xi\xi} + \sigma^2\tilde{o}_{\sigma P}(1) = F_n + \frac{\hat{\sigma}^2}{2}\hat{F}_n^{\xi\xi} + \sigma^2\tilde{o}_{\sigma P}(1)$. □

Proposition 4. *$\tilde{\sigma}^2 - \sigma^2 = \sigma^4\tilde{o}_{\sigma P}(1)$.*

Proof. Formulae (11) and $\Delta\hat{\varphi} - z_n = \tilde{o}_{\sigma P}(1)$ from the proof of Theorem 2 (see below) imply that $Q(\hat{\beta}) - Q(\beta_0) = \sigma^4\left(k_n z_n + V_n(z_n)^2\right) + \sigma^4\tilde{o}_{\sigma P}(1)$, whence

$$Q(\beta_0) = Q(\hat{\beta}) - \sigma^4\left(-\tfrac{1}{2}k_n V_n^{-1}k_n^{\mathrm{T}} + \tfrac{1}{4}k_n V_n^{-1}k_n^{\mathrm{T}}\right) + \sigma^4\tilde{o}_{\sigma P}(1) = $$
$$= \hat{\sigma}^2 + \tfrac{\sigma^4}{4}k_n V_n^{-1}k_n^{\mathrm{T}} + \sigma^4\tilde{o}_{\sigma P}(1).$$

We replace σ^2, k_n and V_n by their estimators. Then, by Propositions 1 and 2,

$$Q(\beta_0) = \hat{\sigma}^2 + \tfrac{1}{4}\hat{\sigma}^4\cdot\hat{k}_n\hat{V}_n^{-1}\hat{k}_n^{\mathrm{T}} + \sigma^4\tilde{o}_{\sigma P},(1)$$

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

$$Q(\beta_0) = \sigma^2 - \frac{\hat{\sigma}^4}{4n}\sum_{i=1}^n \frac{(g^{\xi\xi})^2}{\left(1+(g^{\xi})^2\right)^3}\Big|_{(x_i,\hat{\beta})} + \sigma^4\tilde{o}_{\sigma P}(1).$$

From the last two expansions, we obtain

$$\sigma^2 = \hat{\sigma}^2 + \frac{\hat{\sigma}^4}{4}\left(\hat{k}_n\hat{V}_n^{-1}k_n^{\mathrm{T}} + \frac{1}{n}\sum_{i=1}^n \frac{(g^{\xi\xi})^2}{\left(1+(g^{\xi})^2\right)^3}\Big|_{(x_i,\hat{\beta})}\right) + \sigma^4\tilde{o}_{\sigma P}(1), \text{ where } \hat{\sigma}^2 = Q(\hat{\beta}). \quad \square$$

Proof of Statement 2) of Theorem 3.
Theorem 2 states that $\beta_0 = \hat{\beta} - \sigma^2 z_n - \sigma^4 a_n + \sigma^4\tilde{o}_{\sigma P}(1)$. The functions k, V, p, W, and T satisfy inequality (6). Therefore, in view of Propositions 2 and 4, $\hat{a}_n - a_n = \tilde{o}_{\sigma P}(1)$, $\tilde{z}_n - z_n = \sigma^2\tilde{o}_{\sigma P}(1)$. Hence, $\hat{\sigma}^4\hat{a}_n - \sigma^4 a_n = \sigma^4\tilde{o}_{\sigma P}(1)$, $\tilde{\sigma}^2\tilde{z}_n - \sigma^2 z_n = \sigma^4\tilde{o}_{\sigma P}(1)$. Then we obtain

$$\tilde{\tilde{\beta}}_n - \beta_0 = \tilde{\sigma}^2\tilde{z}_n - \sigma^2 z_n + \hat{\sigma}^4\hat{a}_n - \sigma^4 a_n = \sigma^4\tilde{o}_{\sigma P}(1).$$

Theorem 3 is proved. □

APPENDIX

Proof of Theorem 1.

1) *Proof of (3).* We show that $Q_2(\beta) = \sigma^8 o_{\sigma P}(1)$. Consider a component of this sum:

$$q(x,y,\beta) = \left(y - g\left(h(x,y,\beta),\beta\right)\right)^2 + \left(x - h(x,y,\beta)\right)^2 \leqslant \left(y - g(\xi,\beta)\right)^2 + (x-\xi)^2 \leqslant$$
$$\leqslant 2\left(y - g(\xi,\beta_0)\right)^2 + 2\left(g(\xi,\beta_0) - g(\xi,\beta)\right)^2 + (x-\xi)^2 \leqslant 2\delta^2 + \varepsilon^2 + const.$$

Remember that $\tilde{\varepsilon}_i$ and $\tilde{\delta}_i$ were defined in (7). We have

$$Q_2(\beta) = \frac{1}{n}\sum_{i\notin B_n(\nu_0)} q(x_i,y_i,\beta) \leqslant \frac{1}{n}\sum_{i\notin B_n(\nu_0)}\left(2\delta^2 + \varepsilon^2 + const\right) \leqslant$$
$$\leqslant \frac{1}{n}\sum_{i=1}^n\left(2\delta^2 + \varepsilon^2 + const\right)\cdot\left[I\left(|\varepsilon_i| \geqslant \nu\right) + I\left(|\delta_i| \geqslant \nu\right)\right] =$$
$$= \frac{\sigma^2}{n}\sum_{i=1}^n\left(2\tilde{\delta}_i^2 + \tilde{\varepsilon}_i^2 + const\right)\left[I\left(|\tilde{\varepsilon}_i| \geqslant \nu/\sigma\right) + I\left(|\tilde{\delta}_i| \geqslant \nu/\sigma\right)\right].$$

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

$$\sigma^2 \mathbf{E}\, I\left(|\tilde{\varepsilon}_i| \geqslant \nu/\sigma\right) \cdot \tilde{\varepsilon}_i^2 \leqslant \sigma^2 \cdot \sqrt{\frac{6\sigma}{\nu}} \cdot (2\pi)^{-1/4} e^{-\frac{\nu^2}{4\sigma^2}}.$$

Similar inequalities can be obtained for other terms, and we have finally $\mathbf{E}\, Q_2(\beta) \leqslant C\sigma e^{-\frac{\nu^2}{4\sigma^2}} = C_1\sigma^8 o(1)$, as $\sigma \to 0+$. Hence, by the Chebyshev inequality

$$P\left(\frac{Q_2(\beta)}{\sigma^8} > C\right) \leqslant \frac{\mathbf{E}\, Q_2(\beta)}{\sigma^8 C} = \frac{o(1)}{C} \to 0 \quad \text{as } \sigma \to 0+,$$

and $Q_2(\beta) = \sigma^8 o_{\sigma P}(1)$, where σ^8 can be replaced by any positive degree of σ.

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 $\Delta = O\left(|\varepsilon| + |\delta|\right)$. Indeed,

$$\Delta^2 = (\xi - h)^2 \leqslant 2\left[(\xi - x)^2 + (x - h)^2\right] \leqslant 2\varepsilon^2 + 2\left[\left(y - g(h, \beta_0)\right)^2 + (x - h)^2\right] \leqslant$$
$$\leqslant 2\varepsilon^2 + 2\left[\left(y - g(\xi, \beta_0)\right)^2 + (x - \xi)^2\right] = 4\varepsilon^2 + 2\delta^2.$$

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

$$\begin{aligned} g(h, \beta_0) &= g + g^{\xi}\Delta + \tfrac{1}{2} g^{\xi\xi}\Delta^2 = g + g^{\xi}\Delta + a_0\Delta^2, \\ g^{\xi}(h, \beta_0) &= g^{\xi} + g^{\xi\xi}\Delta = g^{\xi} + 2a_0\Delta, \\ g^{\beta}(h, \beta_0) &= g^{\beta} + g^{\beta\xi}\Delta + \tfrac{1}{2} g^{\beta\xi\xi}\Delta^2, \quad g^{\beta\xi}(h, \beta_0) = g^{\beta\xi} + g^{\beta\xi\xi}\Delta. \end{aligned} \tag{8}$$

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

$$\left(\delta - g^{\xi}\Delta - a_0\Delta^2\right)\left(g^{\xi} + 2a_0\Delta\right) + \varepsilon - \Delta = 0 \tag{9}$$

with the unknown parameters ξ and $\beta_0 = (a_0,\, b_0,\, c_0)^{\mathrm{T}}$. 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

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

Here, Δ_k is a polynomial of ε and δ with the coefficients depending of g^{ξ} and a_0. Substituting (8) in (9), we find Δ_k as

$$(\delta g^{\xi} + \varepsilon) + 2a_0\Delta\delta - 3a_0 g^{\xi}\Delta^2 - 2a_0^2\Delta^3 = \left((g^{\xi})^2 + 1\right)\Delta \quad \Leftrightarrow$$
$$\Delta = \frac{(\delta g^{\xi} + \varepsilon) + 2a_0\Delta\delta - 3a_0 g^{\xi}\Delta^2 - 2a_0^2\Delta^3}{(g^{\xi})^2 + 1}.$$

Hence,

$$\Delta_1 = \frac{\delta g^{\xi} + \varepsilon}{(g^{\xi})^2 + 1}, \qquad \Delta_2 = \frac{2a_0\Delta_1\delta - 3a_0 g^{\xi}\Delta_1^2}{(g^{\xi})^2 + 1},$$
$$\Delta_3 = \frac{2a_0\Delta_2\delta - 6a_0 g^{\xi}\Delta_1\Delta_2 - 2a_0^2\Delta_1^3}{(g^{\xi})^2 + 1},$$
$$\Delta_4 = \frac{2a_0\Delta_3\delta - 3a_0 g^{\xi}(\Delta_2^2 + 2\Delta_1\Delta_3) - 6a_0^2\Delta_1^2\Delta_2}{(g^{\xi})^2 + 1}.$$

Similarly, one can find Δ_5 and Δ_6.

3) We now find the Taylor expansions of $q, q^\beta, q^{\beta\beta}$ and $q^{\beta\beta\beta}$ at the point (x, y, β_0) as functions of ε and δ and their expectations. The expectations of odd terms are zeros, and those of even terms are certain functions of σ, ξ, β_0.

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

$$q(x,y,\beta_0) = (y - g(h,\beta_0)) + (x-h)^2 = \left(\delta - g^\xi\Delta - \tfrac{1}{2}g^{\xi\xi}\Delta^2\right) + (\varepsilon - \Delta)^2 =$$
$$= q_2(\varepsilon,\delta) + q_3(\varepsilon,\delta) + q_4(\varepsilon,\delta) + O(|\varepsilon|^5 + |\delta|^5), \qquad \text{where}$$
$$q_2(\varepsilon,\delta) = (\delta - g^\xi\Delta_1)^2 + (\varepsilon - \Delta_1)^2,$$
$$q_4(\varepsilon,\delta) = \left(g^\xi\Delta_2 + \tfrac{1}{2}g^{\xi\xi}\Delta_1^2\right)^2 - 2(\delta - g^\xi\Delta_1)(g^\xi\Delta_3 + g^{\xi\xi}\Delta_1\Delta_2).$$

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

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

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

From (8), we obtain

$$q^\beta(x,y,\beta_0) = \left(\delta - g^\xi\Delta - \tfrac{1}{2}g^{\xi\xi}\Delta^2\right)\left(g^\beta + g^{\beta\xi}\Delta + \tfrac{1}{2}g^{\beta\xi\xi}\Delta^2\right), \text{ whence}$$
$$q_2^\beta = \left(2g^\xi\Delta_2 + \tfrac{1}{2}g^{\xi\xi}\Delta_1^2\right)g^\beta - 2\left(\delta - g^\xi\Delta_1\right)\Delta_1 g^{\beta\xi};$$
$$q_4^\beta = \left(2g^\xi\Delta_4 + g^{\xi\xi}(\Delta_2^2 + 2\Delta_1\Delta_3)\right)g^\beta + \left(3g^{\xi\xi}\Delta_1^2\Delta_2 + 2g^\xi\Delta_2^2 - 2(\delta - 2g^\xi\Delta_1)\Delta_3\right)g^{\beta\xi} +$$
$$+ \left(\tfrac{1}{2}g^{\xi\xi}\Delta_1^4 - \Delta_1\Delta_2(2\delta - 3g^\xi\Delta_1)\right)g^{\beta\xi\xi}.$$

The expectations of the expansion terms are $\mathbf{E}\, q_2^\beta(\varepsilon,\delta) = \sigma^2 k(\xi,\beta_0)$, $\mathbf{E}\, q_4^\beta(\varepsilon,\delta) = \sigma^4 p(\xi,\beta_0)$.

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

$$q^{\beta\beta}(x,y,\beta) = G^{\beta\beta} + \left(G^{\beta u}\right)^{\mathrm{T}} h^\beta = G^{\beta\beta} - \left(G^{\beta u}\right)^{\mathrm{T}} G^{\beta u}\cdot(G^{uu})^{-1},\ h^\beta = -(G^{uu})^{-1}G^{\beta u}.$$

We write down the expansion terms of $G^{\beta\beta}, G^{\beta u}$, and G^{uu} at the point (x, y, β_0, h):

$$G^{\beta\beta}(x,y,\beta_0,h) = 2g^{\beta\,\mathrm{T}}(h,\beta_0)g^\beta(h,\beta_0),$$
$$G_0^{\beta\beta} = 2g^{\beta\,\mathrm{T}}g^\beta, \quad G_2^{\beta\beta} = 2\Delta_1^2 g^{\beta\xi\,\mathrm{T}}g^{\beta\xi} + 4\Delta_2\left(g^\beta * g^{\beta\xi}\right) + 2\Delta_1^2\left(g^\beta * g^{\beta\xi\xi}\right),$$
$$G^{\beta u}(x,y,\beta_0,h) = 2g^\xi(h,\beta_0)g^\beta(h,\beta_0) - 2\left(\delta - g(h,\beta_0)\right)g^{\beta\xi}(h,\beta_0),$$
$$G_0^{\beta u} = 2g^\xi g^\beta, \quad G_1^{\beta u} = 2g^{\xi\xi}\Delta_1 g^\beta - (\delta - 2g^\xi\Delta_1)g^{\beta\xi},$$
$$G_2^{\beta u} = 2g^{\xi\xi}\Delta_2 g^\beta + 2\left(2g^\xi\Delta_2 + \tfrac{3}{2}g^{\xi\xi}\Delta_1^2\right)g^{\beta\xi} - (2\delta - 3g^\xi\Delta_1)\Delta_1 g^{\beta\xi\xi},$$
$$G^{uu}(x,y,\beta_0,h) = 2\left(1 + (g^\xi(h,\beta_0))^2 - (y - g(h,\beta_0))g^{\xi\xi}(h,\beta_0)\right),$$
$$G_0^{uu} = 2\left(1 + (g^\xi)^2\right), \quad G_1^{uu} = 2g^{\xi\xi}\left(3g^\xi\Delta_1 - \delta\right), \quad G_2^{uu} = 6g^{\xi\xi}\left(g^\xi\Delta_2 + \tfrac{1}{2}g^{\xi\xi}\Delta_1^2\right).$$

Then the Taylor expansion of $(G^{uu})^{-1}$ is

$$(G^{uu})^{-1} = \frac{1}{G_0^{uu}}\left(1 - \left(\frac{G_1^{uu}}{G_0^{uu}} + \frac{G_2^{uu}}{G_0^{uu}}\right) + \left(\frac{G_1^{uu}}{G_0^{uu}} + \frac{G_2^{uu}}{G_0^{uu}}\right)^2 + O\left(|\varepsilon|^3 + |\delta|^3\right)\right),$$

whence three first terms of the expansion are

$$(G^{uu})_0^{-1} := (G_0^{uu})^{-1}, \quad (G^{uu})_1^{-1} := -\frac{G_1^{uu}}{(G_0^{uu})^2}, \quad (G^{uu})_2^{-1} := \frac{1}{G_0^{uu}}\left(\left(\frac{G_1^{uu}}{G_0^{uu}}\right)^2 - \frac{G_2^{uu}}{G_0^{uu}}\right).$$

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

$$q_0^{\beta\beta} = G_0^{\beta\beta} - \frac{1}{G_0^{uu}}(G_0^{\beta u})^{\mathrm{T}}G_0^{\beta u} = 2\left(1 - \frac{(g^\xi)^2}{1+(g^\xi)^2}\right)g^{\beta\,\mathrm{T}}g^\beta = \frac{2}{1+(g^\xi)^2}g^{\beta\,\mathrm{T}}g^\beta = 2V(\xi,\beta_0);$$
$$q_2^{\beta\beta} = G_2^{\beta\beta} - \left(G_0^{\beta u\,\mathrm{T}}G_2^{\beta u} + G_1^{\beta u\,\mathrm{T}}G_1^{\beta u} + G_2^{\beta u\,\mathrm{T}}G_0^{\beta u}\right)(G^{uu})_0^{-1} -$$
$$- \left(G_0^{\beta u\,\mathrm{T}}G_1^{\beta u} + G_1^{\beta u\,\mathrm{T}}G_0^{\beta u}\right)(G^{uu})_1^{-1} - \left(G_0^{\beta u\,\mathrm{T}}G_0^{\beta u}\right)(G^{uu})_2^{-1}.$$

The expectation of $q_2^{\beta\beta}$ is $\mathbf{E}\, q_2^{\beta\beta}(\varepsilon,\delta) = 2\sigma^2 W(\xi,\beta_0)$.

d) Consider $q^{\beta\beta\beta}(x,y,\beta_0)$. The third order derivatives of G at the point (x,y,β,u) are as follows:
$G^{\beta\beta\beta}=0$, $G^{\beta\beta u}=2\left(g^{\beta\,\mathrm{T}}g^{\beta\xi}+g^{\beta\xi\,\mathrm{T}}g^{\beta}\right)|_{(u,\beta)}$, $G^{uuu}=\left(6g^{\xi}g^{\xi\xi}\right)|_{(u,\beta)}$,
$G^{\beta uu}=\left(2g^{\xi}g^{\beta\xi}+2g^{\xi\xi}g^{\beta}+2g^{\xi}g^{\beta\xi}-(y-g)\,g^{\beta\xi\xi}\right)|_{(u,\beta)}$, and
$h^{\beta\beta}=-(G^{uu})^{-1}\left(G^{\beta\beta u}+2G^{\beta uu}h^{\beta}+G^{uuu}\cdot(h^{\beta\,\mathrm{T}}h^{\beta})\right)$, as a derivative of the implicit function. Differentiating the function $q^{\beta\beta}(x,y,\beta)=G^{\beta\beta}+G^{\beta u}\cdot h^{\beta}$, we have
$q^{\beta\beta\beta}=G^{\beta\beta\beta}+G^{\beta\beta u}*h^{\beta}+h^{\beta}*G^{\beta uu}*h^{\beta}+G^{\beta u}*h^{\beta\beta}=-3(G^{uu})^{-1}\left(G^{\beta\beta u}*G^{\beta u}\right)+$
$+\,3(G^{uu})^{-2}\left(G^{\beta uu}*G^{\beta u}*G^{\beta u}\right)-(G^{uu})^{-3}G^{uuu}\left(G^{\beta u}*G^{\beta u}*G^{\beta u}\right).$

It can be easily found from the above-written that

$$q_0^{\beta\beta\beta}=\frac{6g^{\xi\xi}(g^{\xi})^2}{\left(1+(g^{\xi})^2\right)^3}g^{\beta}*g^{\beta}*g^{\beta}-\frac{12g^{\xi}}{\left(1+(g^{\xi})^2\right)^2}g^{\beta}*g^{\beta}*g^{\beta\xi}=6\mathrm{T}(\xi,\beta_0)$$

(Remember that the notation $\vec a*\vec b*\vec c$ 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 $Q_1^{\beta}(\beta_0)$:

$$Q_1^{\beta}(\beta_0)=\frac{1}{n}\sum_{i\in B_n(\nu)}q^{\beta}(x_i,y_i,\beta_0)=\sum_{j=1}^{6}A_k+\frac{1}{n}\sum_{i\in B_n(\nu)}O\left(|\varepsilon|^7+|\delta|^7\right),\ A_k=\sum_{i\in B_n(\nu)}q_k^{\beta}(\varepsilon_i,\delta_i).$$

We use Lemma 1 several times. Start with A_1. The term $q_1^{\beta}(\varepsilon,\delta)$ is a linear form of $\tilde\varepsilon$ and $\tilde\delta$ with bounded coefficients. It follows from (ii) that, for arbitrary i, $1\leqslant i\leqslant n$,

$$q_1^{\beta}(\varepsilon_i,\delta_i)=-2\left(\delta_i-g^{\xi}\Delta_1(\varepsilon_i,\delta_i)\right)g^{\beta}=\sigma\frac{2(g^{\xi}\tilde\varepsilon_i-\tilde\delta_i)}{(g^{\xi})^2+1}g^{\beta}.$$

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

$$A_1=\frac{1}{n}\sum_{i=1}^{n}q_1^{\beta}(x_i,y_i,\beta_0)-\frac{1}{n}\sum_{i\notin B_n(\nu)}q_1^{\beta}(x_i,y_i,\beta_0)=:S_1-S_2,$$

where $S_1=\frac{\sigma}{\sqrt n}O_P(1)$, and $S_2=\sigma^8 o_{\sigma P}(1)$ like a sum in 1). Other A_k can be expanded in a similar way. Consider A_6 separately. Introduce

$$R_1=\sigma^{-6}\frac{1}{n}\sum_{i=1}^{n}E\,q_6^{\beta}(\varepsilon_i,\delta_i)=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=0}^{6}\vec c_l(\xi_i,\beta_0)\cdot E\,\tilde\varepsilon_i^{\,l}\tilde\delta_i^{6-l}.$$

It is a bounded nonrandom vector depending only on (ξ_i,β_0). Dividing A_6 into two sums similarly to A_1, we obtain

$$A_6=\frac{1}{n}\sum_{i=1}^{n}q_6^{\beta}(\varepsilon_i,\delta_i)+\sigma^6 o_{\sigma P}(1)=\sigma^6\left(R_1+\frac{O_P(1)}{\sqrt n}\right)+\sigma^6 o_{\sigma P}(1).$$

Now consider the last term. We get

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

The expansion of $Q_1^{\beta}(\beta_0)$ follows from the preceding formulas. In a similar way, we can obtain the expansions for $Q(\beta_0)$, $Q_1^{\beta\beta}$, $Q_1^{\beta\beta\beta}$, and $Q^{\beta\beta\beta\beta}$. 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 $\hat\beta_n\in U_\nu(\beta_0)$. It occurs for some $\sigma\leqslant\sigma_\varepsilon$, $n\geqslant n_\varepsilon$ with probability at least $1-\varepsilon$, where ε can be an arbitrary positive quantity.

2) Write the Taylor expansion of $Q_1(\beta)$ in $\Delta\beta = \beta - \beta_0$:

$$Q_1(\beta) = Q_1(\beta_0) + Q_1^{\beta}(\beta_0)\Delta\beta + \frac{1}{2!}Q_1^{\beta\beta}(\beta_0)\,(\Delta\beta)^2 + \frac{1}{3!}Q_1^{\beta\beta\beta}(\beta_0)\,(\Delta\beta)^3 + \\ + \frac{1}{4!}\frac{\partial^4 Q_1(\beta_0)}{\partial\beta^4}(\Delta\beta)^4 + \frac{1}{5!}\frac{\partial^5 Q_1(\bar\beta)}{\partial\beta^5}(\Delta\beta)^5, \quad \bar\beta \in [\beta_0, \beta].$$

The derivative $\frac{\partial^5 Q_1(\bar\beta)}{\partial\beta^5}$ is bounded because all the partial derivatives of $Q_1(\beta)$ are bounded. Take the expansions from Theorem 1 and denote $\Delta\varphi = \sigma^2\Delta\beta$. We obtain

$$Q(\beta) - Q(\beta_0) = Q_1(\beta) - Q_1(\beta_0) + \sigma^8 o_{\sigma P}(1) = \sigma^4\left(k_n\Delta\varphi + V_n(\Delta\varphi)^2\right) + \\ + \sigma^6\left(p_n\Delta\varphi + V_n(\Delta\varphi)^2 + \mathrm{T}_n(\Delta\varphi)^3\right) + \sigma^8 R(\varphi) + rest(\Delta\varphi), \tag{10}$$

where $R(\varphi) := R_1\Delta\varphi + R_2(\Delta\varphi)^2 + R_3(\Delta\varphi)^3 + R_4(\Delta\varphi)^4$, $rest = \sigma^8\left(\left(1 + \|\Delta\varphi\|^4\right)\tilde o_{\sigma P}(1) + \|\Delta\varphi\|^4\|\Delta\beta\|O(1)\right)$. Let $\Delta\hat\varphi = \sigma^2(\hat\beta - \beta_0) = \sigma^2\Delta\hat\beta$. Since $\Delta\hat\beta = \tilde o_{\sigma P}(1)$, relation (10) yields that

$$\frac{Q(\hat\beta) - Q(\beta_0)}{\sigma^4} = k_n\Delta\hat\varphi + V_n(\Delta\hat\varphi)^2 + \tilde o_{\sigma P}(1)\left(1 + \|\Delta\hat\varphi\|^2\right) \leqslant 0. \tag{11}$$

Let $c = \liminf_{n\to\infty}\lambda_{\min}(V_n) > 0$; $c > 0$, as follows from (v). Then, for $n \geqslant n_0$, $V_n\Delta\hat\varphi^2 \geqslant \frac{c}{2}\cdot\|\Delta\hat\varphi\|^2$ and $\forall\,\varepsilon > 0\ \exists\,\sigma_\varepsilon > 0\,\forall\,\sigma \in (0, \sigma_\varepsilon]\ \exists\,n_{\varepsilon,\sigma}\ \forall\,n \geqslant n_{\varepsilon,\sigma} : P\left(|\tilde o_{\sigma P}(1)| < c/4\right) > 1 - \varepsilon$. Then (10) implies that, for $\sigma \in (0, \sigma_\varepsilon]$, $n \geqslant \max\{n_0, n_{\varepsilon,\sigma}\}$,

$$\tfrac{c}{4}\|\Delta\hat\varphi\|^2 + k_n\Delta\hat\varphi + \tilde o_{\sigma P}(1) \leqslant 0$$

with probability at least $1 - \varepsilon$. This implies $\Delta\hat\varphi = \tilde O_{\sigma P}(1)$.

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

$$\frac{Q(\hat\beta) - Q\left(\beta_0 + \sigma^2 z_n\right)}{\sigma^4} = k_n\Delta\hat\varphi + V_n\left(\Delta\hat\varphi\right)^2 - k_n z_n - V_n\left(z_n\right)^2 + \sigma^2\tilde O_{\sigma P}(1) = \\ = V_n\left(\Delta\hat\varphi - z_n\right)^2 + \left(2V_n z_n + k_n\right)\left(\Delta\hat\varphi - z_n\right) + \sigma^2\tilde O_{\sigma P}(1) \leqslant 0. \tag{12}$$

Let $z_n = -\frac{1}{2}V_n^{-1}k_n^{\mathrm{T}}$. Then (12) changes into $V_n\left(\Delta\hat\varphi - z_n\right)^2 = \sigma^2\tilde O_{\sigma P}(1)$, and condition (v) implies that $\Delta\hat\varphi - z_n = \sigma\tilde O_{\sigma P}(1) = \tilde o_{\sigma P}(1)$.

4) Let $\Delta\varphi = z_n + t$, $\Delta\hat\varphi = z_n + \hat t$, where $t = \sigma\tilde o_{\sigma P}(1)$. Subtract expansions (10) for z_n and $\Delta\varphi = z_n + t$.

$$L(t) : = \sigma^{-4}\left[Q(\beta_0 + \sigma^2(z_n + t)) - Q(\beta_0 + \sigma^2 z_n)\right] = \\ k_n t + V_n\left((z_n + t)^2 - z_n^2\right) + \sigma^2\left[p_n t + W_n\left((z_n + t)^2 - z_n^2\right) + \mathrm{T}_n\left((z_n + t)^3 - z_n^3\right)\right] + \\ \sigma^4[\,R(z_n + t) - R(z_n)] + \sigma^{-4}[rest(z_n + t) - rest(z_n)],$$

where $j_n := p_n + 2W_n z_n + 3\mathrm{T}_n z_n^2$. Since $z_n + t = \tilde O_{\sigma P}(1)$, we have $rest(z_n + t) - rest(z) = \sigma^8\tilde o_{\sigma P}(1)$. $R(z_n + t) - R(z_n) = \tilde o_{\sigma P}(1)$ by the definition of R. Then $L(t)$ has an expansion

$$L(t) = V_n t^2 + \sigma^2 j_n t + \sigma^2\left(W_n t^2 + 3\mathrm{T}_n z_n t^2 + \mathrm{T}_n t^3\right) + \sigma^4\tilde o_{\sigma P}(1). \tag{13}$$

Prove that $\Delta\tilde\varphi = \tilde O_{\sigma P}(1)$, i.e., $\hat t = \sigma^2\hat s$, where $\hat s = \tilde O_{\sigma P}(1)$. We have

$$Q(\hat\beta) - Q(\beta_0 + \sigma^2 z_n) = L(\hat t) = \sigma^4[V_n\hat s^2 + j_n\hat s + \left(W_n\hat s\,\hat t + 3\mathrm{T}_n z_n\hat s\,\hat t + \mathrm{T}_n\hat s^2\hat t\right) + \tilde o_{\sigma P}(1)] \leqslant 0.$$

Hence,

$$V_n\hat s^2 \leqslant -j_n\hat s + \tilde o_{\sigma P}(1)\left(1 + \|\hat s\|^2\right),$$

and we have from (v) that $\hat s = \tilde O_{\sigma P}(1)$.

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

$$L(\sigma^2 a_n) = \sigma^4\left(V_n a_n^2 + j_n a_n + \tilde o_{\sigma P}(1)\right).$$

From (11) and $\hat{s} = \tilde{O}_{\sigma P}(1)$, we obtain

$$L(\sigma^2 \hat{s}) = \sigma^4 \left(V_n \hat{s}^2 + j_n \hat{s} + \tilde{o}_{\sigma P}(1) \right).$$

Subtracting these equalities, we have

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

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

$$V_n \left(\hat{s} - a_n \right)^2 \leqslant \tilde{o}_{\sigma P}(1),$$

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

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 ε_i, δ_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 $\hat{\beta}$ for sufficiently small but fixed σ 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

1. I. Fazekas, A.G. Kukush, S. Zwanzig,, *On inconsistency of the least squares estimator in nonlinear functional relations*, Preprint, Department of Statistics and Demography, Odense University, Denmark, 1998.
2. I. Fazekas, A. Kukush, S. Zwanzig, *Bias correction of nonlinear orthogonal regression*, Ukr. Math. Journ. **56** (2004), no. 8, 1101–1118.
3. A.G. Kukush, S. Zwanzig, *About the adaptive minimum contrast estimator in a model with non-linear functional relations*, Ukr. Math. Journ. **53** (2001), no. 9, 1445–1452.
4. W.A. Fuller, *Measurement errors models*, Wiley, New York, 1987.
5. G.S. Repetatska, *On inconsistency of orthogonal regression estimator in a nonlinear vector error-in-variables model*, Th. Prob. and Math. Stat. (to appear).

Kyiv 01033 Volodymyrska str. 64, Taras Shevchenko Kyiv National University
E-mail: galchonok@univ.kiev.ua