CORRECTION OF NONLINEAR ORTHOGONAL REGRESSION ESTIMATOR

I. Fazekas,[1] **A. Kukush,**[2] **and S. Zwanzig**[3]   UDC 519.21

For any nonlinear regression function, it is shown that the orthogonal regression procedure delivers an inconsistent estimator. A new technical approach to the proof of inconsistency based on the implicit-function theorem is presented. For small measurement errors, the leading term of the asymptotic expansion of the estimator is derived. We construct a corrected estimator, which has a smaller asymptotic deviation for small measurement errors.

Introduction

We consider the nonlinear errors-in-variables model

$$y_i = g(\xi_i, \beta^0) + \varepsilon_{1i}, \tag{1}$$

$$x_i = \xi_i + \varepsilon_{2i}, \tag{2}$$

where $i = 1, \ldots, n$. The design points or variables $\{\xi_1, \ldots, \xi_n\} \subset \mathbb{R}$ are unknown and fixed. In this model, the application of the least-squares method is often called orthogonal regression because the sum of orthogonal distances between the observations and the regression curve has to be minimized.

This method is also known in numerical literature under the name of total least squares (see the works of Boggs, Byrd and Schnabel [1] and Schwetlick and Tiller [2] and the references therein). The numerical algorithms are globally and locally convergent and already implemented in ODRPACK, FUNKE, and GaussFit software packages, as discussed by Boggs and Rogers in [3] (ODRPACK) and by Strebel, Sourlier and Gander in [4] (FUNKE). The application of the nonlinear orthogonal distance estimator and the use of these packages are recommended in meteorology by Strebel, Sourlier and Gander [4], in astronomy by Branham [5] and Jefferys [6] (GaussFit), in biology by Van Huffel [7], and in robotics by Mallick [8].

For linear errors-in-variables models, this estimation procedure is consistent. In the case of normally distributed errors, the least-squares estimator is the maximum-likelihood one and is also efficient. An excellent and thorough survey of linear errors-in-variables models was given by Fuller in [9].

In the nonlinear case, the consistency of the least-squares estimator is only given under additional conditions that guarantee that the unknown design points are consistently estimable. This is the case, e.g., under the entropy condition on the set of design points [10], or in the case of repeated observations [11], or in an asymptotic inference with respect to a vanishing error variance [9, p. 240].

In statistical literature, the inconsistency of the unrestricted nonlinear orthogonal distance estimator has been known for a long time, and several adjusting proposals were given by Wolter and Fuller [11], Stefanski

---

[1] Institute of Informatics, University of Debrecen, Debrecen, Hungary.

[2] Shevchenko Kyiv University, Kyiv.

[3] Uppsala University, Uppsala, Sweden.

0041–5995/04/5608–1308

[12], Stefanski and Carroll [13], Nagelkerke [14], Armstrong [15], Schafer [16], Hillegers [17], Amemiya [18], Gleser [19], and Kukush and Zwanzig [20].

Nevertheless, inconsistency results are proved only for special cases. Carroll et al. [21] assumed, instead of (1), that y_i is a Bernoulli variable with expected value

$$G\left(\xi_i^T \beta\right) \tag{3}$$

and that, in (2), the error term is normally distributed with known covariance matrix. They argued that the maximum likelihood estimator for β is not consistent and advised to consult the authors in this point. Stefanski [22] gave the proof of inconsistency for the above binary regression model with logistic link function $G(t) = (1 + \exp(-t))^{-1}$ in (3). Stefanski [12] proposed M-estimators $\overline{\beta}$ defined as a measurable solution of the estimating equation

$$\sum_{i=1}^{n} \psi_i\left(x_i, y_i, \overline{\beta}\right) = 0. \tag{4}$$

The main point is that the estimating functions ψ_i in (4) have to be unbiased, i.e.,

$$E_{\xi, \beta}\left(\sum_{i=1}^{n} \psi_i(x_i, y_i, \beta) \right) = o(1), \tag{5}$$

to obtain the consistency of the M-estimator $\overline{\beta}$. Stefanski [22] argued that if (5) fails, then the M-estimator for β in inconsistent. The fact that (5) is violated is established only in special cases, like for the exponential regression function.

In this paper, we give a general proof of the inconsistency of the orthogonal regression procedure for arbitrary nonlinear smooth regression functions. The main idea is to use the technique of implicitly defined functions and to derive an expansion of the corresponding score functions

$$\sum_{i=1}^{n} \psi_i(x_i, y_i, \beta).$$

This expansion includes terms that do not vanish in the nonlinear case with fixed error variances. This is also a new technical approach for such inconsistency proof in statistics.

Under mild additional assumptions, we consider the asymptotic deviation of the orthogonal distance estimator. We derive the leading term of the asymptotic expansion for small measurement errors and present a corrected estimator, which has a smaller asymptotic deviation. Our new estimator is different from the adjusted estimator proposed by Amemiya and Fuller in [23], where an asymptotic expansion of the estimator is given in a replication-type model. In particular, they required that the variances decrease faster than the sample sizes increase and obtained another nonvanishing leading term within the framework of their asymptotic approach.

The paper is organized as follows: In Sec. 2, model assumptions and the orthogonal regression estimator are given. In Sec. 3, the inconsistency of the orthogonal regression estimator and related results are formulated. In Sec. 4, a leading term of the asymptotic expansion is presented, and, in Sec. 5, the corrected estimator is constructed. Section 6 contains conclusions. The proofs are given in Appendices 1 and 2.

2. Model

Assume that we have observations $(y_1, x_1), \dots, (y_n, x_n)$ that are independently and, in general, not identically distributed and are generated by (1) and (2). The errors $\{\varepsilon_{ji}\}$ are

$$\varepsilon_{ji} \sim \mathcal{N}(0, \sigma^2) \quad \text{i.i.d.,} \quad i = 1, \dots, n, \quad j = 1, 2. \tag{6}$$

This assumption is made for convenience. In [24], the proof of inconsistency is given for arbitrary error distributions with moment condition and weak dependence between the variables ε_{ji}.

The regression function $g(\cdot, \cdot)$ is known. The unknown parameters are β^0, ξ_i, $i = 1, \dots, n$, and σ^2. The parameter of interest is $\beta^0 \in \Theta \subset \mathbb{R}^p$. The variables $\xi_1, \dots, \xi_n$ are the nuisance parameters, whose number grows up with the sample size n.

We assume that the variables $\xi_1, \dots, \xi_n$ come from a product set

$$[-a, a]^n, \tag{7}$$

where a is fixed but unknown and β^0 lies in the interior of a compact set:

$$\beta^0 \in \operatorname{int}\Theta, \quad \Theta \subset \mathbb{R}^p \text{ is compact.} \tag{8}$$

We also assume the smoothness condition

$$g \in C^3(\mathbb{R} \times U) \quad \text{for some open} \quad U \supset \Theta. \tag{9}$$

Derivatives are denoted by superscripts, e.g.,

$$g^\xi(\xi, \beta) = \frac{\partial}{\partial \xi} g(\xi, \beta), \quad g^{\xi\xi}(\xi, \beta) = \frac{\partial^2}{\partial \xi^2} g(\xi, \beta).$$

The orthogonal regression estimator $\hat{\beta}$ of β^0 is defined as a measurable solution of the optimization problem:

$$\hat{\beta} \in \arg\min_{\beta \in \Theta} \frac{1}{n} \sum_{i=1}^n \min_{\xi \in \mathbb{R}} \left[(y_i - g(\xi, \beta))^2 + (x_i - \xi)^2 \right].$$

3. Inconsistency Results

In this section, we use an asymptotic approach for an increasing sample size $n \to \infty$ and arbitrary small fixed variances. We show that, under this setup, $\hat{\beta}$ is inconsistent.

The sum of the projected squares is denoted by

$$Q_{\text{Proj}}(\beta) \;=\; \frac{1}{n}\sum_{i=1}^{n}\;\min_{\xi\in\mathbb{R}}\Big[\,(y_i - g(\xi,\beta))^2 + (x_i - \xi)^2\,\Big] \qquad \text{for all}\quad \beta\in\Theta. \tag{10}$$

The function $Q_{\text{Proj}}(\beta)$ is our estimating criterion for the parameter of interest β, where the nuisance parameters are eliminated. Note that, under (6), the orthogonal regression estimator coincides with the maximum likelihood one.

We have

$$Q_{\text{Proj}}(\beta) \;=\; \frac{1}{n}\sum_{i=1}^{n} q(x_i, y_i, \beta),$$

where

$$q(x, y, \beta) \;=\; [\,y - g(h(x,y,\beta),\beta)\,]^2 \;+\; [\,x - h(x,y,\beta)\,]^2 \tag{11}$$

and $h(x, y, \beta)$ is the minimum point of the function

$$f(\xi) \;=\; (\,y - g(\xi,\beta)\,)^2 + (x-\xi)^2.$$

Then the function $h(x, y, \beta)$ is implicitly defined by the normal equation

$$F(x, y, \beta, h) \;=\; (\,y - g(h,\beta)\,)\,g^{\xi}(h,\beta) + x - h \;=\; 0 \qquad \text{for all}\quad x,\; y,\; \text{and}\; \beta \tag{12}$$

with the initial condition

$$h(\xi, g(\xi,\beta),\beta) \;=\; \xi \qquad \text{for all}\quad \beta\in\Theta.$$

Since

$$F^{h}(x, y, \beta, h) \;=\; -1 - (g^{\xi}(h,\beta))^2 + (\,y - g(h,\beta)\,)\,g^{\xi\xi}(h,\beta)$$

with

$$F^{h}(\xi, g(\xi,\beta),\beta,\xi) \;=\; -1 - (g^{\xi}(\xi,\beta))^2 \;\neq\; 0,$$

the implicit-function theorem implies the following: Under the smoothness condition (9), there exist a constant v_0 and an ε-neighborhood $U_{\varepsilon}(\beta^0)$ of β^0 such that

$$h(\cdot,\cdot,\cdot)\colon\; [\xi - v_0, \xi + v_0] \times [\,g(\xi,\beta^0) - v_0,\, g(\xi,\beta^0) + v_0\,] \times U_{\varepsilon}(\beta^0) \;\to\; \mathbb{R} \tag{13}$$

and $h(\cdot,\cdot,\cdot)$ is a uniquely defined twice differentiable function. For the derivative $\dfrac{\partial h(x,y,\beta)}{\partial \beta} = h^{\beta}(x, y, \beta)$, we have

$$h^\beta(x, y, \beta) = \frac{1}{1 + (g^\xi)^2 - (y - g)g^{\xi\xi}}\left((y - g)g^{\xi\beta} - g^\xi g^\beta\right),\tag{14}$$

where the regression function g and its derivatives are taken at the point (h, β).

For illustration, consider the simple linear model

$$g(\xi, \beta) = \beta\xi.$$

In this case, we know that $h(x, y, \beta) = (y\beta + x)/(\beta^2 + 1)$ and $h^\beta(x, y, \beta) = (y - y\beta^2 - 2x\beta)/(\beta^2 + 1)^2$.

In the theorem below, we derive a stochastic expansion of the first derivatives of the leading term $Q_{\text{Lead}}(\beta)$ of the estimation criterion $Q_{\text{Proj}}(\beta)$ defined in (10).

Theorem 1. *Suppose that, for model (1), (2), assumptions (6)–(9) are satisfied. Then, for every positive constant $\nu \le \nu_0$, where ν_0 is from (13), one has*

$$Q_{\text{Proj}}(\beta) = Q_{\text{Lead}}(\beta) + \sigma^4 \text{rest}_{(1)}(n, \beta, \nu, \sigma^2),\tag{15}$$

$$Q_{\text{Lead}}^\beta(\beta^0) = \sigma^2\kappa_n + \left(\nu\sigma^2 + \frac{\sigma}{\sqrt{n}}\right)O_P(1) + \sigma^4\text{rest}_{(2)}(n, \nu, \sigma^2),\tag{16}$$

where

$$\kappa_n = \frac{1}{n}\sum_{i=1}^n \frac{g^{\xi\xi}(\xi_i, \beta^0)}{\left(1 + \left(g^\xi(\xi_i, \beta^0)\right)^2\right)^2}g^\beta(\xi_i, \beta^0),\tag{17}$$

and, for all constants $c > 0$,

$$\lim_{\sigma\to 0}\sup_{n\ge 1}P_{\xi_1,\dots,\xi_n,\beta^0}\left(\sup_{\beta\in\Theta}\left|\text{rest}_{(1)}(n, \beta, \nu, \sigma^2)\right| > c\right) = 0,\tag{18}$$

$$\lim_{\sigma\to 0}\sup_{n\ge 1}P_{\xi_1,\dots,\xi_n,\beta^0}\left(\left|\text{rest}_{(2)}(n, \nu, \sigma^2)\right| > c\right) = 0,\tag{19}$$

where $O_P(1)$ denotes the remainder, which is uniformly bounded in probability $P_{\xi_1,\dots,\xi_n,\beta^0}$ with respect to all n, $\nu \le \nu_0$, and $\sigma > 0$.

The leading term κ_n is related to the curvature of the regression function. Recall that the curvature of the graph $\Gamma_\beta = \{(\xi, g(\xi, \beta)), \xi \in \mathbb{R}\}$ at the point $(\xi^0, g(\xi^0, \beta))$ is given by $(g^{\xi\xi}(\xi^0, \beta))\left(1 + \left(g^\xi(\xi^0, \beta)\right)^2\right)^{-3/2}$.

Theorem 1 implies the main result of this paper, which states that the orthogonal regression estimator is inconsistent if the leading term in expansion (16) is nonvanishing. Actually, Theorem 2 below states much more than inconsistency.

Theorem 2. *Suppose that, for model (1), (2), the conditions of Theorem 1 are satisfied. Also assume that*

$$\liminf_{n\to\infty} \|\kappa_n\| > 0, \tag{20}$$

where κ_n is given in (17). Then, for every $\varepsilon > 0$, there exist $\tau > 0$ and $\sigma_\varepsilon > 0$ such that, for any $\sigma \in (0, \sigma_\varepsilon]$,

$$\liminf_{n\to\infty} P_{\xi_1,\dots,\xi_n,\beta^0}\left(\left\|\hat{\beta} - \beta^0\right\| > \sigma^2\tau\right) \geq 1 - \varepsilon.$$

Corollary 1. *Suppose that condition (20) in Theorem 2 is replaced by the condition*

$$\limsup_{n\to\infty} \|\kappa_n\| > 0,$$

where κ_n is given in (17). Then, for every $\varepsilon > 0$, there exist $\tau > 0$ and $\sigma_\varepsilon > 0$ such that, for any $\sigma \in (0, \sigma_\varepsilon]$,

$$\limsup_{n\to\infty} P_{\xi_1,\dots,\xi_n,\beta^0}\left(\left\|\hat{\beta} - \beta^0\right\| > \sigma^2\tau\right) \geq 1 - \varepsilon.$$

Remark 1. Theorem 2 states inconsistency for small enough but fixed variances σ^2. The case $\sigma^2 \to 0$ is excluded in Theorem 2.

Remark 2. We have no inconsistency in the case where the regression function is linear in the design points because $g^{\xi\xi} \equiv 0$ and, hence, $\kappa_n \equiv 0$. We also have $\kappa_n \equiv 0$ if the regression function is independent of β. But then the necessary contrast condition for the consistency of the orthogonal regression estimator is not satisfied.

Example 1. Consider model (1), (2) with $g(\xi, \beta) = \exp(\beta\xi)$, $\xi \in \mathbb{R}$, $\beta \in \mathbb{R}$. For the exponential model, we have

$$\kappa_n = \frac{(\beta^0)^2}{n} \sum_{i=1}^{n} \frac{\xi_i e^{2\beta^0\xi_i}}{\left(1 + (\beta^0)^2 e^{2\beta^0\xi_i}\right)^2}.$$

If $\beta^0 \neq 0$ and the design points ξ_i are positive, bounded, and separated from zero, then (20) holds, and, under assumption (6), the orthogonal regression estimator is inconsistent for small enough but fixed variances σ^2.

4. Asymptotic Deviation

Definition 1. *Let $\eta_n = \eta_n(\sigma^2)$ be a sequence of random vectors depending on σ^2, $\sigma > 0$. Then we write $\eta_n = o_{P_\sigma}(1)$ if, for every $\varepsilon > 0$ and $\gamma > 0$, there exists $\sigma_{\varepsilon\gamma} > 0$ such that, for all $\sigma \in (0, \sigma_{\varepsilon\gamma}]$,*

$$\liminf_{n\to\infty} P\left(\left\|\eta_n(\sigma^2)\right\| \leq \gamma\right) \geq 1 - \varepsilon.$$

Definition 2. *Let* $\eta_n = \eta_n(\sigma^2)$ *be a sequence of random vectors depending on* σ^2, $\sigma > 0$. *Then we write* $\eta_n = O_{P_\sigma}(1)$ *if, for every* $\varepsilon > 0$, *there exist* C_ε *and* $\sigma_\varepsilon > 0$ *such that, for any* $\sigma \in (0, \sigma_\varepsilon]$,

$$\liminf_{n \to \infty} P(\|\eta_n\| \le C_\varepsilon) \ge 1 - \varepsilon.$$

Further, we need the following contrast condition: For every $\delta > 0$,

$$\liminf_{n \to \infty} \inf_{\|\beta - \beta^0\| \ge \delta} \frac{1}{n} \sum_{i=1}^{n} \rho^2\left(P_i^0, \Gamma_\beta\right) > 0, \tag{21}$$

where $\rho^2\left(P_i^0, \Gamma_\beta\right)$ is the distance between the point $P_i^0 = (\xi_i, g(\xi_i, \beta^0))$ and the graph $\Gamma_\beta = \{(\xi, g(\xi, \beta)):$ $\xi \in \mathbb{R}\}$.

The result presented below is very close to Lemma 1 in [23]. We give it without proof. Recall that the estimator $\hat{\beta}$ is a random vector depending on the sample size n and the error variance σ^2.

Lemma 1. *Suppose that, for model (1), (2), assumptions (6), (7), (8), and (21) are satisfied and* $g \in C(\mathbb{R} \times \Theta)$. *Then*

$$\left\|\hat{\beta} - \beta^0\right\| = o_{P_\sigma}(1).$$

We introduce the matrix

$$V_n = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + \left(g^\xi(\xi_i, \beta^0)\right)^2} g^\beta(\xi_i, \beta^0) g^\beta(\xi_i, \beta^0)^T.$$

Note that V_n^{-1} corresponds to the asymptotic covariance matrix of $\hat{\beta} - \beta^0$ in the setup of Amemiya and Fuller [23].

Then we can show that the total least-squares estimator $\hat{\beta}$ is with high probability near the point $\beta^0 - \sigma^2 V_n^{-1} \kappa_n / 2$.

Theorem 3. *Suppose that, for model (1), (2), the conditions of Theorem 2 are satisfied. Also assume that condition (21) is satisfied and*

$$\liminf_{n \to \infty} \lambda_{\min}(V_n) > 0, \tag{22}$$

where $\lambda_{\min}$ *denotes the smallest eigenvalue. Then*

$$\hat{\beta} = \beta^0 - \frac{\sigma^2}{2} V_n^{-1} \kappa_n + \sigma^2 o_{P_\sigma}(1). \tag{23}$$

Definition 2, the fact that κ_n is bounded, and Theorem 3 yield

$$\hat{\beta} = \beta^0 + \sigma^2 O_{P_\sigma}(1). \tag{24}$$

5. Corrected Estimator

Relation (23) enables us to define a corrected estimator $\tilde{\beta}$ as follows:

$$\tilde{\beta} = \hat{\beta} + \frac{\hat{\sigma}^2}{2} \hat{V}_n^{-1} \hat{\kappa}_n, \tag{25}$$

where $\hat{\sigma}^2$ is the corrected variance estimator given by

$$\hat{\sigma}^2 = \left(\frac{1}{n} \sum_{i=1}^{n} \left(y_i - g(x_i, \hat{\beta}) \right)^2 \right) \left(1 + \frac{1}{n} \sum_{i=1}^{n} \left(g^\xi(x_i, \hat{\beta}) \right)^2 \right)^{-1}, \tag{26}$$

$\hat{V}_n$ is the estimate of the matrix V_n determined as

$$\hat{V}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + \left(g^\xi(x_i, \hat{\beta}) \right)^2} g^\beta(x_i, \hat{\beta}) g^\beta(x_i, \hat{\beta})^T,$$

and $\hat{\kappa}_n$ is the estimate of κ_n occurring in Theorem 1:

$$\hat{\kappa}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{g^{\xi\xi}(x_i, \hat{\beta})}{\left[1 + \left(g^\xi(x_i, \hat{\beta}) \right)^2 \right]^2} g^\beta(x_i, \hat{\beta}).$$

Lemma 2. *Suppose that the conditions of Lemma 1 are satisfied. Consider a function $F \in C^1(\mathbb{R} \times U)$ for some open $U \supset \Theta$. Assume that, for some fixed $C > 0$ and $A > 0$,*

$$\left| F^\xi(\xi, \beta) \right| \leq C \exp(A|\xi|), \quad \xi \in \mathbb{R}, \quad \beta \in U. \tag{27}$$

Then, for model (1), (2),

$$\frac{1}{n} \sum_{i=1}^{n} F(x_i, \hat{\beta}) = \frac{1}{n} \sum_{i=1}^{n} F(\xi_i, \beta^0) + o_{P_\sigma}(1).$$

Using Lemma 2, we obtain

$$\hat{V}_n = V_n + o_{P_\sigma}(1), \tag{28}$$

$$\hat{\kappa}_n = \kappa_n + o_{P_\sigma}(1) \tag{29}$$

if

$$\sum_{i=1}^{3} \left| \frac{\partial^i}{\partial \xi^i} g(\xi, \beta) \right| + \| g^{\beta}(\xi, \beta) \| + \| g^{\xi\beta}(\xi, \beta) \| \leq C \exp(A|\xi|) \tag{30}$$

for some fixed $C > 0$ and $A > 0$ and for all $\xi \in \mathbb{R}$ and $\beta \in U$.

Lemma 3. *Suppose that the conditions of Theorem 3 hold. Also assume that inequality (30) is satisfied with omitting the terms $\partial^3 g(\xi, \beta)/\partial \xi^3$ and $g^{\xi\beta}(\xi, \beta)$. Then*

$$\hat{\sigma}^2 = \sigma^2 + \sigma^2 o_{P_\sigma}(1). \tag{31}$$

Summarizing (28), (29), and (31), we obtain the following result:

Theorem 4. *Suppose that condition (30) and the assumptions of Theorem 3 are satisfied. Then the corrected estimator $\tilde{\beta}$ in (25) has the representation*

$$\tilde{\beta} = \beta^0 + \sigma^2 o_{P_\sigma}(1).$$

Remark 3. If the variances are different, i.e., $D^2\varepsilon_{1i} \neq D^2\varepsilon_{2i}$, but their ratio is known, then one can transform (1) and (2) to obtain equal variances.

Remark 4. It is also possible to find a correction of the naive estimator of β^0 defined by

$$\hat{\beta}_{\text{naive}} \in \arg\min_{\beta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} (y_i - g(x_i, \beta))^2.$$

The naive estimator is also inconsistent, and its asymptotic expansion has the leading term of order σ^2 involving $g^{\beta\beta}$. The correction demands stronger restrictions than (30) on the derivatives of g. For example, a bound for the third derivative with respect to β is needed. Recall that, in a linear model, $\hat{\beta}_{\text{naive}}$ is inconsistent, whereas $\hat{\beta}$ is consistent.

6. Conclusions

We considered an orthogonal regression estimator $\hat{\beta}$ in a nonlinear functional errors-in-variables model. In the situation where the model is strictly separated from a linear model, we gave a mathematical proof of the inconsistency of $\hat{\beta}$. The proof relies on the implicit-function theorem.

Moreover, we derived an expansion of the asymptotic deviation for small measurement errors and constructed a new corrected estimator $\tilde{\beta}$, which has smaller asymptotic deviation for small errors.

It would be interesting to derive the next term of order σ^4 in the expansion of $\hat{\beta} - \beta^0$ and to construct a correction of higher order.

7. Appendix 1: Proof of Inconsistency

7.1. Proof of Theorem 1. The proof is divided into several steps.

Truncation. Let ν_0 be the constant introduced in (13). For an arbitrary positive constant ν, $\nu \le \nu_0$, we define the index set

$$B_n(\nu) = \{i: 1 \le i \le n, \ |\varepsilon_{1i}| \le \nu, \ |\varepsilon_{2i}| \le \nu\}. \tag{32}$$

We decompose the projected sum of squares $Q_{\text{Proj}}(\beta)$ into two parts

$$Q_{\text{Proj}}(\beta) = \frac{1}{n} \sum_{i \in B_n(\nu)} q(x_i, y_i, \beta) + \frac{1}{n} \sum_{i \notin B_n(\nu)} q(x_i, y_i, \beta)$$

and define the leading term

$$Q_{\text{Lead}}(\beta) = \frac{1}{n} \sum_{i \in B_n(\nu)} q(x_i, y_i, \beta).$$

We now show that

$$\frac{1}{n} \sum_{i \notin B_n(\nu)} q(x_i, y_i, \beta) = \sigma^2 \text{rest}_{(1)}(n, \beta, \nu, \sigma^2), \tag{33}$$

where the remainder $\text{rest}_{(1)}$ satisfies (18). By virtue of (11), we have

$$q(x_i, y_i, \beta) \le (y_i - g(\xi_i, \beta))^2 + (x_i - \xi_i)^2$$

$$= (\varepsilon_{1i} + g(\xi_i, \beta^0) - g(\xi_i, \beta))^2 + \varepsilon_{2i}^2 \le 2\varepsilon_{1i}^2 + \varepsilon_{2i}^2 + \text{const}$$

for some constant const because $g(\cdot, \cdot)$ is continuous and, therefore, bounded on the compact set $[-a, a] \times \Theta$. Hence,

$$\frac{1}{n} \sum_{i \notin B_n(\nu)} q(x_i, y_i, \beta) \le \frac{2}{n} \sum_{i \notin B_n(\nu)} \left(\varepsilon_{1i}^2 + \varepsilon_{2i}^2 + \text{const}\right) \le \frac{2}{n} \sum_{i=1}^{n} \left[\varepsilon_{1i}^2 + \varepsilon_{2i}^2 + \text{const}\right] \left[I(|\varepsilon_{1i}| \ge \nu) + I(|\varepsilon_{2i}| \ge \nu)\right],$$

where $I(A)$ is the indicator function of the set A. The typical terms of the expectation of the above expression are $E\left(\varepsilon_{1i}^2 I(|\varepsilon_{1i}| \ge \nu)\right)$, $E(I(|\varepsilon_{1i}| \ge \nu))$, and $E\left(\varepsilon_{1i}^2 I(|\varepsilon_{2i}| \ge \nu)\right)$. We can now use inequalities of the form

$$E\left(\varepsilon_{1i}^2 I(|\varepsilon_{1i}| \ge \nu)\right) \le \frac{\sigma^4}{\nu^2} E\left(\frac{\varepsilon_{1i}^2}{\sigma^4} I\left(\frac{|\varepsilon_{1i}|}{\sigma} \ge \frac{\nu}{\sigma}\right)\right),$$

where ε_{1i}/σ is standard normally distributed. Therefore, by using the Chebyshev inequality, we obtain

$$P\left(\frac{1}{n}\sum_{i\notin B_n(\nu)} q(x_i, y_i, \beta) > \varepsilon\right) \leq \text{const}\,\varepsilon^{-1}(\sigma^4(\nu^{-2} + \nu^{-4})o(1)) \tag{34}$$

as $\sigma^2 \to 0$. Inequality (34) yields (33) with $\text{rest}_{(1)}$ satisfying (18).

Taylor Expansions. Now consider the case $i \in B_n(\nu)$. Then, under the assumptions made above, all observations y_i, x_i belong to a compact set. Let us omit the index i and set $\varepsilon_1 =: \delta$ and $\varepsilon_2 =: \varepsilon$. We have

$$x = \xi + \varepsilon \tag{35}$$

and

$$y = g(\xi, \beta^0) + \delta, \tag{36}$$

where

$$|\varepsilon| \leq \nu, \quad |\delta| \leq \nu. \tag{37}$$

We introduce Δ by the equality

$$h(x, y, \beta^0) = \xi + \Delta,$$

where $h(x, y, \beta^0)$ is defined in (12). Under assumptions (7)–(9), the expansions of the regression function and of its derivatives at the point $h = h(x, y, \beta^0)$ are as follows:

$$g(h, \beta^0) = g(\xi, \beta^0) + \Delta g^\xi(\xi, \beta^0) + \frac{1}{2}\Delta^2 g^{\xi\xi}(\xi, \beta^0) + O(\Delta^3), \tag{38}$$

$$g^\xi(h, \beta^0) = g^\xi(\xi, \beta^0) + \Delta g^{\xi\xi}(\xi, \beta^0) + O(\Delta^2), \tag{39}$$

$$g^\beta(h, \beta^0) = g^\beta(\xi, \beta^0) + \Delta g^{\beta\xi}(\xi, \beta^0) + O(\Delta^2). \tag{40}$$

By virtue of (37), all variables in (38)−(40) belong to some compact set. Thus, relation (9) implies that, for $k = 2, 3$, the remainders satisfy the inequality

$$\sup_{x,y,\xi} \frac{\left\|O(\Delta^k)\right\|}{\left|\Delta^k\right|} \leq \text{const.}$$

We insert (35)–(39) into (12) and obtain

$$\Delta^2 A + \Delta\delta B + \Delta C - \delta g^\xi - \varepsilon = O\left(|\Delta|^3 + |\delta|\Delta^2\right) \tag{41}$$

with

$$A = \frac{3}{2}(g^\xi g^{\xi\xi}), \quad B = -g^{\xi\xi}, \quad C = 1 + (g^\xi)^2, \tag{42}$$

where the regression function g and its derivatives are taken at the point (ξ, β^0). Further, let

$$\Delta_1 = \frac{\delta g^\xi + \varepsilon}{C}.$$

Note that

$$\Delta_1 = O(|\varepsilon| + |\delta|). \tag{43}$$

Using the definition of Δ and $h(x, y, \beta)$, we obtain

$$\Delta = O(|\varepsilon| + |\delta|). \tag{44}$$

Relation (41) now yields

$$\Delta = \Delta_1 - \Delta\frac{\delta B}{C} - \Delta^2\frac{A}{C} + O(|\varepsilon|^3 + |\delta|^3).$$

Thus,

$$\Delta = \Delta_1 + \frac{1}{2}\Delta_2 + O(|\varepsilon|^3 + |\delta|^3), \tag{45}$$

where Δ_2 is of order $O(|\varepsilon|^2 + |\delta|^2)$. Substituting this into (41), we obtain

$$\Delta_2 = -\frac{2(\Delta_1^2 A + \Delta_1 \delta B)}{C},$$

or, more explicitly,

$$\Delta_2 = \frac{g^{\xi\xi}}{C^3}\left(\delta^2\left(2g^\xi - (g^\xi)^3\right) - 3\varepsilon^2 g^\xi + \varepsilon\delta\left(2 - 4(g^\xi)^2\right)\right). \tag{46}$$

Proof of (16). We now consider

$$Q_{\text{Lead}}^\beta(\beta^0) = \frac{1}{n}\sum_{i\in B_n(v)} q^\beta(x_i, y_i, \beta^0). \tag{47}$$

Using (11), we get

$$q^\beta(x_i, y_i, \beta) = -2\left(h^\beta\left[(x - h) + (y - g)g^\xi\right] + (y - g)g^\beta\right),$$

where the regression function g and its derivatives are taken at the point $(h(x, y, \beta), \beta)$. Since $h(x, y, \beta)$ satisfies the normal equation (12), we have

$$q^\beta(x, y, \beta) = -2(y - g(h(x, y, \beta), \beta))g^\beta(h(x, y, \beta), \beta). \tag{48}$$

We insert (38) and (40) into (48) and use (44). Thus, we get

$$q^\beta(x, y, \beta^0) = -2\delta g^\beta + 2\Delta(g^\beta g^\xi - \delta g^{\xi\beta}) + \Delta^2(2g^{\xi\beta}g^\xi + g^{\xi\xi}g^\beta) + O\left(|\varepsilon|^3 + |\delta|^3\right), \tag{49}$$

where all derivatives are taken at the point (ξ, β^0). Using (45) with (43) and (46), we obtain

$$\sum_{i \in B_n(v)} q^\beta(x_i, y_i, \beta^0) = L + V + R. \tag{50}$$

Here, L is the linear term, which has the form

$$L = \sum_{i \in B_n(v)} (a_i \varepsilon_{1i} + b_i \varepsilon_{2i})$$

and V is the quadratic term, which has the form

$$V = \sum_{i \in B_n(v)} \left(c_i \varepsilon_{1i}^2 + d_i \varepsilon_{2i}^2 + m_i \varepsilon_{1i} \varepsilon_{2i}\right),$$

where

$$d_i = \frac{-3}{C^3} g^{\xi\xi}(g^\xi)^2 g^\beta + \frac{2}{C^2} g^\xi g^{\xi\beta} + \frac{1}{C^2} g^{\xi\xi}g^\beta$$

and

$$c_i = \frac{3}{C^3} g^{\xi\xi}(g^\xi)^2 g^\beta - \frac{2}{C^2} g^\xi g^{\xi\beta}.$$

The coefficients a_i, b_i, c_i, d_i, and m_i depend only on bounded partial derivatives of the regression function.

In (50), R is the remainder consisting of terms with orders of ε_{1i} and ε_{2i} higher than 2, and

$$R \leq \frac{1}{n} \sum_{i \in B_n(v)} r_i\left(|\varepsilon_{1i}|^3 + |\varepsilon_{2i}|^3\right),$$

where r_i depends only on bounds of partial derivatives of the regression function and $\max_{i=1,\ldots,n} r_i \leq \text{const.}$ Then

$$E|R| \leq \text{const } E\left(|\varepsilon_{11}|^3 I(|\varepsilon_{11}| \leq v)\right) \leq \text{const } v\sigma^2.$$

Therefore,

$$R = \nu \sigma^2 O_P(1).$$

Here and in what follows, $O_P(1)$ is uniformly bounded in probability $P_{\xi_1, \dots, \xi_n, \beta^0}$ with respect to all n, $\nu \leq \nu_0$, and $\sigma > 0$.

Using (47) and (50), we get

$$Q_{\text{Lead}}^\beta(\beta^0) = \frac{1}{n} \sum_{i \in B_n(\nu)} q^\beta(x_i, y_i, \beta^0) = S_1 - S_2 + \nu \sigma^2 O_P(1), \tag{51}$$

where

$$S_1 = \frac{1}{n} \sum_{i=1}^n \left(a_i \varepsilon_{1i} + b_i \varepsilon_{2i} + c_i \varepsilon_{1i}^2 + d_i \varepsilon_{2i}^2 + m_i \varepsilon_{1i} \varepsilon_{2i} \right)$$

and

$$S_2 = \frac{1}{n} \sum_{i \notin B_n(\nu)} \left(a_i \varepsilon_{1i} + b_i \varepsilon_{2i} + c_i \varepsilon_{1i}^2 + d_i \varepsilon_{2i}^2 + m_i \varepsilon_{1i} \varepsilon_{2i} \right).$$

Similarly to (34), we have

$$S_2 = \sigma^4 \text{rest}_{(2)}(n, \nu, \sigma^2), \tag{52}$$

where $\text{rest}_{(2)}$ satisfies (19). According to (6), we get

$$S_1 = \frac{1}{n} \sum_{i=1}^n \sigma^2(c_i + d_i) + \frac{\sigma}{\sqrt{n}} O_P(1) \tag{53}$$

and, furthermore,

$$\frac{1}{n} \sum_{i=1}^n \sigma^2(c_i + d_i) = \frac{1}{n} \sum_{i=1}^n \sigma^2 \frac{1}{\left(1 + (g^\xi)^2 \right)^2} g^{\xi\xi} g^\beta = \sigma^2 \kappa_n,$$

where the derivatives are taken at (ξ_i, β^0) and κ_n is introduced in (17). Using (51)–(53), we obtain expansion (16) for $Q_{\text{Lead}}^\beta(\beta^0)$.

Theorem 1 is proved.

7.2. Proof of Theorem 2. $Q_{\text{Lead}}^{\beta\beta}(\beta)$ *is Bounded.* Let

$$G(x, y, \beta, u) = (y - g(u, \beta))^2 + (x - u)^2,$$

where $x, y, u \in \mathbb{R}$ and $\beta \in \Theta$. Then, by virtue of (11),

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

Since $h(x, y, \beta)$ is the minimum point, we have $G^u\big|_{u=h(x,y,\beta)} = 0$ for x, y, and β in the neighborhood of $(\xi, g(\xi, \beta^0), \beta^0)$. Using $\xi^\beta(x, y, \beta) = G^\beta + G^u h^\beta$, we get

$$\xi^\beta(x, y, \beta) = G^\beta(x, y, \beta, u)\big|_{u=h(x,y,\beta)}.$$

The second derivative is

$$\xi^{\beta\beta}(x, y, \beta) = G^{\beta\beta}(x, y, \beta, u)\big|_{u=h(x,y,\beta)} + G^{\beta u}(x, y, \beta, u)\big|_{u=h(x,y,\beta)} h^\beta(x, y, \beta). \tag{54}$$

Using formulas (54) and (14), we establish that, under condition (9), for γ and ν small enough but positive,

$$\sup_{\|\beta-\beta^0\|<\gamma} \left\| Q_{\text{Lead}}^{\beta\beta}(\beta) \right\| \leq \Lambda < \infty, \tag{55}$$

where Λ is a deterministic constant depending only on γ and ν.

Representation of $Q_{\text{Proj}}(\beta)$. Denote

$$U_\gamma(\beta^0) = \left\{ \beta \colon \|\beta - \beta^0\| < \gamma \right\}.$$

Taking (8) and (9) into account, for $\beta \in U_\gamma(\beta^0)$ and $\Delta\beta = \beta - \beta^0$ we get

$$Q_{\text{Lead}}(\beta) = Q_{\text{Lead}}(\beta^0) + Q_{\text{Lead}}^\beta(\beta^0)\Delta\beta + \frac{1}{2}\Delta\beta^T Q_{\text{Lead}}^{\beta\beta}(\bar{\beta})\Delta\beta, \tag{56}$$

where $\bar{\beta}$ is an intermediate point between β and β^0. It follows from Theorem 1 that, for $\beta \in U_\gamma(\beta^0)$,

$$Q_{\text{Proj}}(\beta) - Q_{\text{Proj}}(\beta^0) = Q_{\text{Lead}}(\beta) - Q_{\text{Lead}}(\beta^0) + \sigma^4 \text{rest}_{(3)},$$

where

$$\text{rest}_{(3)} = \text{rest}_{(3)}(n, \beta, \nu, \sigma^2) = \text{rest}_{(1)}(n, \beta, \nu, \sigma^2) - \text{rest}_{(1)}(n, \beta^0, \nu, \sigma^2).$$

Relation (56) with $\Delta\varphi = \sigma^{-2}\Delta\beta$, assertion (16) of Theorem 1, and the boundedness of $Q_{\text{Lead}}^{\beta\beta}(\beta)$ yield

$$Q_{\text{Proj}}(\beta) - Q_{\text{Proj}}(\beta^0) = \sigma^4\left[\kappa_n + \left(\nu + \frac{1}{\sigma\sqrt{n}} \right)O_P(1) + \sigma^2 \text{rest}_{(2)} \right]\Delta\varphi + O(1)\sigma^4\|\Delta\varphi\|^2 + \sigma^4\text{rest}_{(3)}. \tag{57}$$

Inconsistency. Let us show that $\Delta\hat{\varphi} = \sigma^{-2}\left(\hat{\beta} - \beta^0\right)$ is separated from zero with large probability. We fix $\sigma_0 > 0$ and consider $0 < \sigma \leq \sigma_0$. Since κ_n is bounded, one can find $t > 0$ such that, for all $n \geq 1$,

$$\beta_t = \beta^0 + \sigma^2(-t\kappa_n) \in U_\gamma(\beta^0) \subset \Theta.$$

We insert both $\Delta\varphi = -t\kappa_n$ and $\Delta\hat{\varphi}$ into (57) and recall that $Q_{\text{Proj}}\left(\hat{\beta}\right) \leq Q_{\text{Proj}}(\beta_t)$. We obtain

$$0 \leq \sigma^{-4}\left(Q_{\text{Proj}}(\beta_t) - Q_{\text{Proj}}\left(\hat{\beta}\right)\right) = p(\Delta\hat{\varphi}) + R_1(t) + R_2(n) + R_3(\sigma), \tag{58}$$

where

$$p(\Delta\hat{\varphi}) = -\kappa_n\Delta\hat{\varphi} + \nu O_P(1)\|\Delta\hat{\varphi}\| + O(1)\|\Delta\hat{\varphi}\|^2$$

is a polynomial in $\Delta\hat{\varphi}$ and

$$R_1(t) = -\|\kappa_n\|^2 t + \nu O_P(1)\|\kappa_n\| t + O(1)\|\kappa_n\|^2 t^2,$$

$$R_2(n) = \frac{1}{\sigma\sqrt{n}}O_P(1)(\|\Delta\hat{\varphi}\| + \|\kappa_n\| t),$$

$$R_3(\sigma) = \text{rest}_{(3)}(n, \beta_t, \nu, \sigma^2) - \text{rest}_{(3)}(n, \hat{\beta}, \nu, \sigma^2) - \sigma^2\text{rest}_{(2)}(n, \nu, \sigma^2)\Delta\hat{\varphi} - \sigma^2\text{rest}_{(2)}(n, \nu, \sigma^2)\kappa_n t.$$

Now let $\kappa \in (0, 1)$. (In what follows, κ can be different in different statements, but it can be chosen to be arbitrary close to 1.) By using (20), one can choose $\nu > 0$ and n_0 such that, for $n > n_0$,

$$P\left(\nu O_P(1) < \frac{\|\kappa_n\|}{2}\right) > \kappa. \tag{59}$$

At the same time, one can find $t_0 > 0$ such that, for suitable small positive t, for ν chosen above, and for $n > n_0$, we have

$$R_1(t) \leq -t_0$$

with probability greater than κ. There is an $n_\sigma \geq n_0$ such that, for $n \geq n_\sigma$,

$$R_2(n) \leq \frac{t_0}{4}$$

with probability greater than κ. Moreover, we can find and fix a suitable small positive σ_0 such that, for all $\sigma \in (0, \sigma_0]$ and $n \geq 1$,

$$R_3(\sigma) \leq \frac{t_0}{4}$$

with probability greater than κ. Therefore, relation (58) implies that, for $n \geq n_\sigma$ and $\sigma < \sigma_0$,

$$\frac{t_0}{2} \leq p(\Delta\hat{\varphi})$$

with probability greater than κ. Since the coefficients of the polynomial p are stochastically bounded, $\|\Delta\hat{\varphi}\|^2$ cannot be arbitrarily close in probability to 0. This implies Theorem 2.

7.3. Proof of Corollary 1. In the proof of Theorem 2, we have used (20) to prove (59). If

$$\limsup_{n\to\infty} \|\kappa_n\| > 0,$$

then we can choose a subsequence $n(m)$ such that

$$\lim_{m\to\infty} \|\kappa_{m(n)}\| > 0.$$

For this subsequence, relation (59) and the statement of Theorem 2 remain valid. This proves Corollary 1.

8. Appendix 2: Proofs for Correction

8.1. Proof of Theorem 3. According to Lemma 1, we can consider $\sigma \in (0, \sigma_{\varepsilon\gamma}]$ and $n \geq n_{\varepsilon\gamma}$ such that $\hat{\beta} \in U_\gamma(\beta^0)$. (It has probability greater than $1 - \varepsilon$.)

First, we prove the following: For some $\nu_0 > 0$, if $0 < \nu \leq \nu_0$, then

$$Q_{\text{Lead}}^{\beta\beta}(\beta^0) = 2V_n + \sigma^4 o_{P_\sigma}(1) + \text{rest}_4, \tag{60}$$

where $|\text{rest}_4| \leq \text{const}\,\nu$. Here, ν comes from (32). To obtain this, recall that

$$Q_{\text{Lead}}^{\beta\beta}(\beta^0) = \frac{1}{n}\sum_{i\in B_n(\nu)} q^{\beta\beta}(x_i, y_i, \beta^0). \tag{61}$$

Using (48), we obtain

$$\frac{1}{2}q^{\beta\beta} = (g^\beta g^{\beta T} + g^\xi g^\beta h^{\beta T}) + (y - g)(g^{\beta\xi} h^{\beta T} + g^{\beta\beta}), \tag{62}$$

where h^β is given in (14). For the first summand in (62), we have

$$\frac{1}{n}\sum_{i\in B_n(\nu)} \left\| \left[g^\beta g^{\beta T} + g^\xi g^\beta h^{\beta T} \right]\Big|_{(x_i,y_i,\beta^0)} - \left[g^\beta g^{\beta T} + g^\xi g^\beta \left\{ -\frac{g^\xi g^{\beta T}}{1 + (g^\xi)^2} \right\} \right]\Big|_{(\xi_i,\beta^0)} \right\| \leq \text{const}\,\nu. \tag{63}$$

We get

$$g^\beta g^{\beta T} - g^\xi g^\beta \frac{g^\xi g^{\beta T}}{1 + (g^\xi)^2} = \frac{g^\beta g^{\beta T}}{1 + (g^\xi)^2}. \tag{64}$$

Since $\| g^\beta(\xi, \beta^0) \| \le \text{const}$ and $|\beta| \le a$, we obtain

$$\frac{1}{n} \sum_{i \in B_n(\nu)} \left[\frac{g^\beta g^{\beta T}}{1 + (g^\xi)^2} \right] \bigg|_{(\xi_i, \beta^0)} = V_n + \sigma^4 o_{P_\sigma}(1). \tag{65}$$

For $i \in B_n(\nu)$, we have $|y_i - g(x_i, \beta^0)| \le \text{const } \nu$. Therefore,

$$\left\| \frac{1}{n} \sum_{i \in B_n(\nu)} (y - g)(g^{\beta\xi} h^{\beta T} + g^{\beta\beta}) \bigg|_{(x_i, y_i, \beta^0)} \right\| \le \text{const } \nu. \tag{66}$$

Relations (62)–(66) now yield (61).

By virtue of the smoothness condition (9), for $\beta \in U_\gamma(\beta^0)$ the third derivative $Q_{\text{Lead}}^{\beta\beta\beta}(\beta)$ satisfies the boundedness relation

$$\left\| Q_{\text{Lead}}^{\beta\beta\beta}(\beta) \right\| < \text{const}$$

for small positive γ and ν.

For $\beta \in U_\gamma(\beta^0)$, we use the Taylor expansion

$$Q_{\text{Lead}}(\beta) = Q_{\text{Lead}}(\beta^0) + Q_{\text{Lead}}^{\beta T}(\beta^0)\Delta\beta + \frac{1}{2}\Delta\beta^T Q_{\text{Lead}}^{\beta\beta}(\beta^0)\Delta\beta + \frac{1}{6}\sum_{i,j,k=1}^{p} Q_{\text{Lead}}^{\beta_i\beta_j\beta_k}(\overline{\beta})\Delta\beta_i \Delta\beta_j \Delta\beta_k, \tag{67}$$

where $\overline{\beta}$ is an intermediate point between β^0 and β. Using relations (15) and (16) of Theorem 1 and relations (60) and (67), we get

$$Q_{\text{Proj}}(\beta) = Q_{\text{Proj}}(\beta^0) + \sigma^2 \kappa_n^T \Delta\beta + \Delta\beta^T V_n \Delta\beta + \text{rest},$$

where

$$\text{rest} = \nu\sigma^2 O_P(1)\|\Delta\beta\| + \sigma^4 o_{P_\sigma}(1) + \nu O(1)\|\Delta\beta\|^2 + O(1)\|\Delta\beta\|^3.$$

We set $\beta = \beta_\varphi = \beta^0 + \sigma^2\Delta\varphi$. Then

$$Q_{\text{Proj}}(\beta_\varphi) = Q_{\text{Proj}}(\beta^0) + \sigma^4\left(\kappa_n^T\Delta\varphi + \Delta\varphi^T V_n \Delta\varphi\right) + \text{rest}(\varphi), \tag{68}$$

where

$$\text{rest}(\varphi) = \nu\sigma^4 O_P(1)\|\Delta\varphi\| + \sigma^4 o_{P_\sigma}(1) + \nu\sigma^4 O(1)\|\Delta\varphi\|^2 + \sigma^6 O(1)\|\Delta\varphi\|^3.$$

Let $\hat{\beta} = \beta^0 + \sigma^2 \Delta \hat{\varphi}$. By Lemma 1, we have $\sigma^2 \|\Delta \hat{\varphi}\| = o_{P_\sigma}(1)$. Recall that we consider $\hat{\beta} \in U_\gamma(\beta^0)$. Using (68) and the inequality $Q_{\text{Proj}}(\hat{\beta}) \leq Q_{\text{Proj}}(\beta^0)$, we obtain

$$\kappa_n^T \Delta \hat{\varphi} + \Delta \hat{\varphi}^T V_n \Delta \hat{\varphi} + \sigma^{-4} \text{rest}(\hat{\varphi}) \leq 0, \tag{69}$$

where

$$\sigma^{-4} \text{rest}(\hat{\varphi}) = o_{P_\sigma}(1) \|\Delta \hat{\varphi}\|^2 + \nu O_P(1) \|\Delta \hat{\varphi}\| + o_{P_\sigma}(1) + \nu O(1) \|\Delta \hat{\varphi}\|^2.$$

Consider $\nu > 0$ such that $\nu |O(1)| \leq (\liminf_{n \to \infty} \lambda_{\min}(V_n))/2$ in the last summand. Then, using the boundedness condition for κ_n and relations (22) and (69), we get $\|\Delta \hat{\varphi}\| = O_{P_\sigma}(1)$. This yields $\sigma^2 \|\Delta \hat{\varphi}\|^3 = o_{P_\sigma}(1)$ and $\nu O(1) \|\Delta \hat{\varphi}\|^2 = \nu O_{P_\sigma}(1) \|\Delta \hat{\varphi}\|$. Therefore, it follows from (68) that

$$Q_{\text{Proj}}(\hat{\beta}) = Q_{\text{Proj}}(\beta^0) + \sigma^4 \left(\kappa_n^T \Delta \hat{\varphi} + \Delta \hat{\varphi}^T V_n \Delta \hat{\varphi} + \nu O_{P_\sigma}(1) \|\Delta \hat{\varphi}\|\right) + \sigma^4 o_{P_\sigma}(1). \tag{70}$$

Let $z_n = -\left(V_n^{-1} \kappa_n / 2\right)$. According to (20) and (22), $\|z_n\|$ is bounded and separated from zero. By the definition of $\hat{\beta}$, we have $Q_{\text{Proj}}(\hat{\beta}) \leq Q_{\text{Proj}}(\beta^0 + \sigma^2 z_n)$. Therefore, relations (70) and (68) yield

$$\left\|V_n^{1/2}(\Delta \hat{\varphi} - z_n)\right\|^2 + \nu O_{P_\sigma}(1) \|\Delta \hat{\varphi}\| \leq \nu O_P(1) + o_{P_\sigma}(1). \tag{71}$$

Taking into account that the value ν in (32) can be chosen small enough and using relations (71) and (22), we obtain $\|(\Delta \hat{\varphi} - z_n)\| = o_{P_\sigma}(1)$, which proves (23).

8.2. Proof of Lemma 2. We have

$$\frac{1}{n} \sum_{i=1}^n \left(F(x_i, \hat{\beta}) - F(\xi_i, \beta^0)\right) = r_1 + r_2, \tag{72}$$

where

$$r_1 = \frac{1}{n} \sum_{i=1}^n \left(F(x_i, \hat{\beta}) - F(\xi_i, \hat{\beta})\right),$$

$$r_2 = \frac{1}{n} \sum_{i=1}^n \left(F(\xi_i, \hat{\beta}) - F(\xi_i, \beta^0)\right).$$

By virtue of the mean-value theorem, we have

$$r_1 = \frac{1}{n} \sum_{i=1}^n F^\xi(\overline{\xi}_i, \hat{\beta}) \varepsilon_{2i},$$

where $\bar{\xi}_i$ is an intermediate point between ξ_i and $x_i = \xi_i + \varepsilon_{2i}$. Therefore, by virtue of (27), we get

$$|r_1| \leq \frac{1}{n}\sum_{i=1}^{n} c\exp\left(A\left|\bar{\xi}_i\right|\right)|\varepsilon_{2i}| \leq c\sigma\exp(Aa)\frac{1}{n}\sum_{i=1}^{n} \exp\left(\sigma A\left|\frac{\varepsilon_{2i}}{\sigma}\right|\right)\left|\frac{\varepsilon_{2i}}{\sigma}\right|,$$

where ε_{2i}/σ, $i = 1,\dots,n$, are i.i.d. standard normal. Therefore, the expectation of $\exp\left(\sigma A\left|\frac{\varepsilon_{2i}}{\sigma}\right|\right)\left|\frac{\varepsilon_{2i}}{\sigma}\right|$ is bounded by $(c_1 + c_2 A\sigma)\exp\left(\frac{\sigma^2 A^2}{2}\right)$. Using the law of large numbers, we establish that

$$\limsup_{n\to\infty}|r_1| \leq \sigma\exp(Aa)(c_1 + c_2 A\sigma)\exp\left(\frac{\sigma^2 A^2}{2}\right) \tag{73}$$

with probability 1.

By virtue of the mean-value theorem,

$$r_2 = \frac{1}{n}\sum_{i=1}^{n} F^\beta(\xi_i, \beta_{(i)})^T\left(\hat{\beta} - \beta^0\right),$$

where $\beta_{(i)}$ is an intermediate point between β^0 and $\hat{\beta}$. Recall that, according to (8), β^0 lies in the interior of the compact set Θ. By Lemma 1, $\left\|\hat{\beta} - \beta^0\right\| = o_{P_\sigma}(1)$. Therefore, for fixed $\varepsilon > 0$ and some $\gamma > 0$, we can choose $\sigma_{\varepsilon\gamma} > 0$ and $n_{\varepsilon\gamma} < \infty$ so that if $\sigma \in (0, \sigma_{\varepsilon\gamma}]$ and $n > n_{\varepsilon\gamma}$, then, with probability greater than $1 - \varepsilon$, we have $\hat{\beta} \in \overline{U}_\gamma(\beta^0) = \left\{\beta: \left\|\beta - \beta^0\right\| \leq \gamma\right\}$ and, therefore,

$$|r_2| \leq \sup_{|\xi|\leq a, \beta\in\Theta}\left\|F^\beta(\xi, \beta)\right\|\left\|\Delta\hat{\beta}\right\| \leq K\gamma.$$

Thus, $r_2 = o_{P_\sigma}(1)$. This and (73) yield the required result.

8.3. Proof of Lemma 3. The numerator of $\hat{\sigma}^2$ is

$$\frac{1}{n}\sum_{i=1}^{n}\left(g(x_i, \hat{\beta}) - y_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(g(\xi_i + \varepsilon_{2i}, \hat{\beta}) - \varepsilon_{1i} - g(\xi_i, \beta^0)\right)^2 = R_1 + R_2 - 2R_3, \tag{74}$$

where

$$R_1 = \frac{1}{n}\sum_{i=1}^{n}\left[g(\xi_i + \varepsilon_{2i}, \hat{\beta}) - g(\xi_i, \beta^0)\right]^2, \qquad R_2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{1i}^2,$$

$$R_3 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{1i}\left[g(\xi_i + \varepsilon_{2i}, \hat{\beta}) - g(\xi_i, \beta^0)\right].$$

Here $R_2 = \sigma^2 (1 + o_P(1))$.

Let us estimate R_1. Applying the mean-value theorem twice, we obtain

$$g\big(\xi_i + \varepsilon_{2i}, \hat{\beta}\big) \;=\; g(\xi_i, \beta^0) \;+\; g^\xi\big(\overline{\xi}_i, \beta^0\big)\varepsilon_{2i} \;+\; g^\beta\big(\xi_i + \varepsilon_{2i}, \overline{\beta}_{(i)}\big)^T \Delta\hat{\beta},$$

where $\overline{\xi}_i$ is an intermediate point between ξ_i and $\xi_i + \varepsilon_{2i}$, while $\overline{\beta}_{(i)}$ is an intermediate point between $\hat{\beta}$ and β^0. Therefore,

$$R_1 \;=\; \frac{1}{n}\sum_{i=1}^{n}\Big[\,g^\xi\big(\overline{\xi}_i, \beta^0\big)\Big]^2 \varepsilon_{2i}^2 \;+\; 2\frac{1}{n}\sum_{i=1}^{n} g^\xi\big(\overline{\xi}_i, \beta^0\big)\varepsilon_{2i}\,g^\beta\big(\xi_i + \varepsilon_{2i}, \overline{\beta}_{(i)}\big)^T \Delta\hat{\beta}$$

$$+\; \frac{1}{n}\sum_{i=1}^{n}\Big[\,g^\beta\big(\xi_i + \varepsilon_{2i}, \overline{\beta}_{(i)}\big)^T \Delta\hat{\beta}\Big]^2 \;=\; R_{11} + R_{12} + R_{13}.$$

By Theorem 3, $\Delta\hat{\beta} = \sigma^2 O_{P_\sigma}(1)$. Therefore, using (30), we obtain

$$\big|R_{12}\big| \;\le\; \sigma^2 O_{P_\sigma}(1)\, c\sigma\exp(2Aa)\frac{1}{n}\sum_{i=1}^{n}\exp\!\Big(2\sigma A\Big|\frac{\varepsilon_{2i}}{\sigma}\Big|\Big)\Big|\frac{\varepsilon_{2i}}{\sigma}\Big|.$$

Here, $\dfrac{\varepsilon_{2i}}{\sigma}$, $i = 1, \dots, n$, are i.i.d. standard normal. Therefore, by the law of large numbers, we get $R_{12} = \sigma^2 o_{P_\sigma}(1)$.

By analogy, we establish that $R_{13} = \sigma^2 o_{P_\sigma}(1)$.

Using the mean-value theorem, we now obtain

$$R_{11} \;=\; \frac{1}{n}\sum_{i=1}^{n}\Big[\,g^\xi\big(\xi_i, \beta^0\big) + g^{\xi\xi}\big(\tilde{\xi}_i, \beta^0\big)\tilde{\varepsilon}_{2i}\,\Big]^2 \varepsilon_{2i}^2$$

$$=\; \frac{1}{n}\sum_{i=1}^{n}\Big[\,g^\xi\big(\xi_i, \beta^0\big)\Big]^2 \varepsilon_{2i}^2 \;+\; 2\frac{1}{n}\sum_{i=1}^{n} g^\xi\big(\xi_i, \beta^0\big)g^{\xi\xi}\big(\tilde{\xi}_i, \beta^0\big)\tilde{\varepsilon}_{2i}\varepsilon_{2i}^2 \;+\; \frac{1}{n}\sum_{i=1}^{n}\Big[\,g^{\xi\xi}\big(\tilde{\xi}_i, \beta^0\big)\tilde{\varepsilon}_{2i}\varepsilon_{2i}\Big]^2,$$

where $\tilde{\xi}_i$ is an intermediate point between ξ_i and $\xi_i + \varepsilon_{2i}$, and $|\tilde{\varepsilon}_{2i}| \le |\varepsilon_{2i}|$. Using (30), and the law of large numbers, we conclude that the second and the third terms in the last expression are $\sigma^2 o_{P_\sigma}(1)$. However, the Cantelli strong law of large numbers implies that, for the first term, we have

$$\frac{1}{n}\sum_{i=1}^{n}\Big[\,g^\xi(\xi_i, \beta^0)\Big]^2\big(\varepsilon_{2i}^2 - \sigma^2\big) \;=\; \sigma^2 o_{P_\sigma}(1).$$

Therefore,

$$R_1 \;=\; \frac{1}{n}\sum_{i=1}^{n}\sigma^2\Big[\,g^\xi(\xi_i, \beta^0)\Big]^2 \;+\; \sigma^2 o_{P_\sigma}(1).$$

Applying the mean-value theorem again, we get

$$R_3 = \frac{1}{n}\sum_{i=1}^{n} g^{\xi}(\overline{\xi}_i, \beta^0)\varepsilon_{2i}\varepsilon_{1i} + \frac{1}{n}\sum_{i=1}^{n} g^{\beta}(\xi_i + \varepsilon_{2i}, \overline{\beta}_{(i)})^T \Delta\hat{\beta}\varepsilon_{1i}.$$

Here, for the first term, we have used (30) and the Cantelli strong law of large numbers, while, for the second term, we have used (30) and Theorem 3 to obtain $R_3 = \sigma^2 o_{P_\sigma}(1)$.

Using (74), we get

$$\frac{1}{n}\sum_{i=1}^{n}\left[g(x_i, \hat{\beta}) - y_i\right]^2 = \sigma^2\left[1 + \frac{1}{n}\sum_{i=1}^{n}\left(g^{\xi}(\xi_i, \beta^0)\right)^2\right] + \sigma^2 o_{P_\sigma}(1). \tag{75}$$

According to Lemma 2,

$$\frac{1}{n}\sum_{i=1}^{n}\left[g^{\xi}(\xi_i, \hat{\beta})\right]^2 = \frac{1}{n}\sum_{i=1}^{n}\left[g^{\xi}(\xi_i, \beta^0)\right]^2 + o_{P_\sigma}(1). \tag{76}$$

Relations (75) and (76) yield the statement of Lemma 3.

REFERENCES

1. P. T. Boggs, H. R. Byrd, and R. B. Schnabel, "A stable and efficient algorithm for nonlinear orthogonal distance regression," *SIAM J. Sci. Comput.,* **8**, No. 6, 1052–1078 (1987).
2. H. Schwetlick and V. Tiller, "Numerical methods for estimating parameters in nonlinear models with errors in variables," *Technometrics,* **27**, No. 1, 17–24 (1985).
3. P. T. Boggs and J. E. Rogers, "Orthogonal distance estimators," *Contemp. Math.,* **112**, 183–194 (1990).
4. R. Strebel, D. Sourlier, and W. Gander, "A comparison of orthogonal least squares fitting in coordinate meteorology," in: *Recent Adv. Total Least Squares Techn. and Errors-in-Variables Modeling (Leuven, 1996),* SIAM, Philadelphia (1997), pp. 249–258.
5. R. L. Branham, "Total least squares in astronomy," in: *Recent Adv. Total Least Squares Techn. and Errors-in-Variables Modeling (Leuven, 1996),* SIAM, Philadelphia (1997), pp. 371–380.
6. W. H. Jefferys, "Robust estimation when more than one variable per equation of condition has error," *Biometrika,* **77**, No. 3, 597–607 (1990).
7. S. van Huffel, "TLS applications in biometrical signal processing," in: *Recent Adv. Total Least Squares Techn. and Errors-in-Variables Modeling (Leuven, 1996),* SIAM, Philadelphia (1997), pp. 307–318.
8. M. K. Mallick, "Applications of nonlinear orthogonal distance regression in 3D Motion estimation," in: *Recent Adv. Total Least Squares Techn. and Errors-in-Variables Modeling (Leuven, 1996),* SIAM, Philadelphia (1997), pp. 273–284.
9. W. A. Fuller, *Measurement Errors Models,* Wiley, New York (1987).
10. S. Zwanzig, *Estimation in Nonlinear Functional Errors-in-Variables Models,* Habilitation, Univ. Hamburg (1998).
11. K. M. Wolter and W. A. Fuller, "Estimation of nonlinear errors-in-variables models," *Ann. Statist.,* **10**, No. 2, 539–548 (1982).
12. L. A. Stefanski, "The effect of measurement error on parameter estimation," *Biometrika,* **72**, No. 3, 583–592 (1985).
13. L. A. Stefanski and R. J. Carroll, "Covariate measurement error in logistic regression," *Ann. Statist.,* **13**, No. 4, 1335–1351 (1985).

14. N. J. D. Nagelkerke, "Maximum likelihood estimation of functional relationship," in: *Lect. Notes Statist.,* **69** (1992).

15. B. Armstrong, "Measurement error in the generalized linear model," *Commun. Statist.-Simula. Computa,* **14**, No. 3, 529–544 (1985).

16. D. W. Schafer, "Covariate measurement error in generalized linear models," *Boimetrika,* **72**, No. 3, 583–592 (1987).

17. L. T. M. E. Hillegers, *The Estimation in Functional Relationship Models,* Proefschrift, Techn. Univ., Eindhoven (1986).

18. Y. Amemiya, "Generalization of the TLS approach in the errors-in-variables problem," in: *Recent Adv. Total Least Squares Techn. and Errors-in-Variables Modeling (Leuven, 1996),* SIAM, Philadelphia (1997), pp. 77–86.

19. L. J. Gleser, "Some new approaches to estimation in linear and nonlinear errors-in-variables regression models," in: *Recent Adv. Total Least Squares Techn. and Errors-in-Variables Modeling (Leuven, 1996),* SIAM, Philadelphia (1997), pp. 69–76.

20. A. G. Kukush and S. Zwanzig, "On the adaptive minimum contrast estimator in a model with nonlinear functional relations," *Ukr. Mat. Zh.,* **53**, No. 9, 1204–1209 (2001).

21. R. J. Carroll, C. H. Spiegelman, G. K. K. Lan, K. T. Bailey, and R. D. Abbott, "On errors-in-variables for binary regression models," *Biometrika,* **71**, No. 1, 1925–1984 (1984).

22. L. A. Stefanski, "Correcting data for measurement errors in generalized linear models," *Commun. Statist. Theory Meth.,* **18**, No. 5, 1715–1733 (1989).

23. Y. Amemiya and W. F. Fuller, "Estimation for the nonlinear functional relationship," *Ann. Statist.,* **16**, No. 1, 147–160 (1988).

24. I. Fazekas, A. G. Kukush, and S. Zwanzig, *On Inconsistency of the Least Squares Estimator in Nonlinear Functional Relations,* Preprint No. 1, Odense Univ., Denmark (1998).