UDC 519.21

I. Fazekas[*] (Inst. Informatics, Univ. Debrecen, Hungary),
A. Kukush[**] (Kyiv Nat. Taras Shevchenko Univ., Ukraine),
S. Zwanzig (Uppsala Univ., Sweden)

CORRECTION OF NONLINEAR ORTHOGONAL REGRESSION ESTIMATOR

КОРЕКЦІЯ ОЦІНКИ НЕЛІНІЙНОЇ ОРТОГОНАЛЬНОЇ РЕГРЕСІЇ

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 is presented, which is based on the implicit-function theorem. For small measurement errors, a leading term of asymptotic expansion of the estimator is derived. A corrected estimator is constructed, which has smaller asymptotic deviation for small measurement errors.

Для довільної нелінійної функції регресії показано, що оцінка ортогональної регресії є неконзистентною. Застосовано нову техніку доведення неконзистентності, яка ґрунтується на теоремі про неявну функцію. Для випадку малих похибок вимірювання виписано головний член асимптотичного розкладу оцінки. Побудовано виправлену оцінку, що має менше асимптотичне відхилення у випадку малих похибок вимірювання.

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 known in numerical literature also under the name of total least squares, compare the works by Boggs, Byrd and Schnabel [1], Schwetlick and Tiller [2], and the references there. The numerical algorithms are globally and locally convergent and already implemented in software packages ODRPACK, FUNKE, GaussFit as discussed by Boggs and Rogers [3] (ODRPACK), by Strebel, Sourlier and Gander [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], Jefferys [6] (GaussFit), in biology by Van Huffel [7], in robotics by Mallick [8].

For linear errors-in-variables model, 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 summary of the linear errors-in-variables models was given by Fuller [9].

In the nonlinear case, the consistency of the least-squares estimator is only given under additional conditions, which ensure that the unknown design points are consistently estimable. This is fulfilled, for instance, under an 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 the statistical literature, the inconsistency of the unrestricted nonlinear orthogonal distance estimator has been known for a long time and several adjusting proposals are given by Wolter and Fuller [11], Stefanski [12], Stefanski and Carroll

[*] Partially supported by the Hungarian Foundation of Scientific Researches under Grant № OTKA T032361 / 2000 and Grant № OTKA T032658 / 2000.

[*] The research was partially realized while this author was visiting University of Debrecen.

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

Nevertheless, inconsistency results are proved for special cases only. 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 implicit defined functions and to derive an expansion of the respected score functions

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

This expansion includes terms which 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 a 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. Especially, they require that the variances decrease quicker than the sample sizes increase and obtain another nonvanishing leading term under their asymptotic approach.

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

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

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

This assumption is done for convenience. In [24], the proof of inconsistency is given for arbitrary error distributions with a 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, $i = 1, \ldots, n$, and σ^2. The parameter of interest is $\beta^0 \in \Theta \subset \mathbb{R}^p$. The variables $\xi_1, \ldots, \xi_n$ are the nuisance parameters, whose number grows up with the sample size n.

We assume that the variables $\xi_1, \ldots, \xi_n$ 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 suppose the smoothness condition

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

Derivatives will be 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 will show that, under this setup, $\hat{\beta}$ is inconsistent.

The sum of the projected squares is denoted by

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

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:

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

with the initial condition

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

Because

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

with

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

the implicit-functions theorem implies the following. Under the smoothness condition (9), there exist a constant ν_0 and an ε-neighborhood $U_\varepsilon(\beta^0)$ of β^0 such that

$$h(\cdot,\cdot,\cdot): [\xi - \nu_0, \xi + \nu_0] \times [g(\xi, \beta^0) - \nu_0, g(\xi, \beta^0) + \nu_0] \times U_\varepsilon(\beta^0) \to \mathbb{R} \quad (13)$$

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

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

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 following theorem, 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 the model* (1) *and* (2), *assumptions* (6), (7), (8), (9) *are satisfied. Then, for each positive constant* $\nu \le \nu_0$, ν_0 *from* (13),

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

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

with

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

and, for all constants $c > 0$,

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

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

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

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 the expansion (16) is nonvanishing. Actually, the following Theorem 2 states much more than inconsistency.

Theorem 2. *Suppose that, for the model* (1) *and* (2), *the conditions of Theorem* 1 *are satisfied. Assume additionally that*

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

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

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

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

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

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

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

***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 necessary contrast condition for the consistency of the orthogonal regression estimator is not satisfied.

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

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

If $\beta^0 \neq 0$ and the design points ξ_i are positive, bounded, and separated from zero, then (20) holds and, under the 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 each* $\varepsilon > 0$ *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 each* $\varepsilon > 0$, *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 each $\delta > 0$,

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

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 following result is very close to the Lemma 1 in [23]. We give it without proof. Remind 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 the model* (1) *and* (2), *assumptions* (6), (7), (8), *and* (21) *are satisfied and* $g \in C(\mathbb{R} \times \Theta)$. *Then*

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

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} respects 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 the model* (1) *and* (2), *the conditions of Theorem 2 are satisfied. Assume additionally the validity of* (21) *and*

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

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

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

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

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

5. The corrected estimator. Relation (23) enables us to define a corrected estimator $\tilde{\beta}$ by

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

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

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

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

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

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

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

Lemma 2. *Suppose that the conditions of Lemma* 1 *hold. 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* $A > 0$,

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

Then, for the model (1) *and* (2),

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

Applying 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

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

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

Lemma 3. *Suppose that the conditions of Theorem* 3 *hold. Assume additionally 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., $D^2\varepsilon_{1i} \ne D^2\varepsilon_{2i}$, 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 leading term of order σ^2 involving $g^{\beta\beta}$. The correction demands stronger restrictions than (30) on the derivatives of g. For instance, a bound for the third derivative with respect to β is needed. Recall that, in a linear model, $\hat{\beta}_{\text{naive}}$ is inconsistent, while $\hat{\beta}$ is consistent.

6. Conclusions. We considered an orthogonal regression estimator $\hat{\beta}$ in a nonlinear functional errors-in-variables model. In the situation when the model is strictly separated from a linear model, we gave a mathematical proof of 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 the 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 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 separate 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

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

Now we 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 term $\text{rest}_{(1)}$ satisfies (18). Because of (11), we have

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

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

$$\le 2\varepsilon_{1i}^2 + \varepsilon_{2i}^2 + \text{const}$$

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

Hence,

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

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

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

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

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

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

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

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

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

and

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

with

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

Introduce Δ with the equality

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

where $h(x, y, \beta^0)$ is defined in (12). Under (7), (8), (9), the expansions of the regression function and of its derivatives at the point $h = h(x, y, \beta^0)$ are:

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

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

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

Because of (37), all the variables in (38) – (40) belong to some compact set. So, (9) implies that, for $k = 2, 3$, the remainder terms satisfy the inequality

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

We put (35), (36), (37), (38) and (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 of $h(x, y, \beta)$, we obtain

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

Now (41) gives

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

Thus,

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

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

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

and, 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

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

From (11) we have

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

where the regression function g and its derivatives are taken at the point $(h(x, y, \beta), \beta)$. As $h(x, y, \beta)$ 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 put (38), (40) into (48) and apply (44). Thus,

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

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

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

Here, L is the linear term and has the form

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

V is the quadratic term and has the form

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

with

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

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, m_i depend only on bounded partial derivatives of the regression function.

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

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

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

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

Therefore,

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

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

By (47) and (50), we get

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

with

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

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). By (6)

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

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). From (51) – (53), we get the expansion (16) for $Q^{\beta}_{\text{Lead}}(\beta^0)$.

Theorem 1 is proved.

7.2. Proof of Theorem 2. $Q^{\beta\beta}_{\text{Lead}}(\beta)$ is bounded. Let

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

with $x, y, u \in \mathbb{R}$, $\beta \in \Theta$. Then , by (11),

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

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

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

The second derivative is

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

From formulas (54) and (14), we obtain that, under condition (9), for γ, ν being small enough but positive,

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

with a deterministic constant Λ depending only on γ and ν.

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

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

Reminding (8) and (9), for $\beta \in U_\gamma(\beta^0)$ and $\Delta\beta = \beta - \beta^0$, we have

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

where $\overline{\beta}$ 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 $Q_{\text{Lead}}^{\beta\beta}(\beta)$ imply

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

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

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

Put both $\Delta\varphi = -t\kappa_n$ and $\Delta\hat{\varphi}$ into (57) and remember $Q_{\text{Proj}}\left(\hat{\beta}\right) \le Q_{\text{Proj}}(\beta_t)$. We obtain

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

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 the following, κ can be different in different statements but it can be chosen to be arbitrary close to 1.) By (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 $n > n_0$,

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

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

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

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 for all $n \ge 1$,

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

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

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

with probability greater than κ. As the coefficients in 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 used (20) in proving (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 statement in Theorem 2 remain valid. This proves Corollary 1.

8. Appendix 2: Proofs for the correction. *8.1. **Proof of Theorem*** **3.** According to Lemma 1 we can consider $\sigma \in (0, \sigma_{\varepsilon\gamma}]$ and $n \ge n_{\varepsilon\gamma}$, such that $\hat\beta \in U_\gamma(\beta^0)$. (It has probability greater than $1 - \varepsilon$.)

First we shall prove the following. For some $\nu_0 > 0$, if $0 < \nu \le \nu_0$, then

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

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

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

From (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

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

$$\le \text{const}\,\nu. \tag{63}$$

We have

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

Because $\| g^\beta(\xi, \beta^0)\| \le \text{const}$, as $|\beta| \le a$, 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

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

Now, relations (62) – (66) imply (61).

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

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

for small positive γ, 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. From (15) and (16) of Theorem 1 and from (60) and (67) we get

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

with

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

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 $\hat\beta \in U_\gamma(\beta^0)$. From (68) and $Q_{\text{Proj}}\left(\hat\beta\right) \le Q_{\text{Proj}}(\beta^0)$ we obtain

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

with

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

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

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

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

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

But the value ν in (32) can be chosen small enough and from (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

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

with

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

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

By the mean value theorem

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

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

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

Here ε_{2i}/σ, $i = 1, \ldots, n$, 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 have

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

with probability 1.

By 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 $\hat{\beta}$. We recall that, by (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 $n_{\varepsilon\gamma} < \infty$ such that if $\sigma \in (0, \sigma_{\varepsilon\gamma}]$ and $n > n_{\varepsilon\gamma}$ then, with probability greater than $1 - \varepsilon$, we have: $\hat{\beta} \in \overline{U}_\gamma(\beta^0) = \left\{\beta : \left\|\beta - \beta^0\right\| \le \gamma\right\}$ and therefore

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

So $r_2 = o_{P_\sigma}(1)$. This and (73) give the result.

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

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

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

with

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

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

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

Estimate now R_1. Applying the mean value theorem twice, we obtain

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

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

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

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

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

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

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

We obtain similarly that $R_{13} = \sigma^2 o_{P_\sigma}(1)$.

Now, using the mean value theorem, we obtain

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

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

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

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

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

Therefore, we obtain

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

Applying again the mean value theorem, we get that

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

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

Summarizing, we obtain from (74) that

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

According to Lemma 2,

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

(75) and (76) imply the statement of Lemma 3.

1. *Boggs P. T., Byrd H. R., Schnabel R. B.* A stable and efficient algorithm for nonlinear orthogonal distance regression // SIAM J. Sci. Comput. – 1987. – **8**, № 6. – P. 1052 – 1078.
2. *Schwetlick H., Tiller V.* Numerical methods for estimating parameters in nonlinear models with errors in variables // Technometrics. – 1985. – **27**, № 1. – P. 17 – 24.
3. *Boggs P. T., Rogers J. E.* Orthogonal distance estimators // Contemp. Math. – 1990. – **112**. – P. 183 – 194.
4. *Strebel R., Sourlier D., Gander W.* A comparison of orthogonal least squares fitting in coordinate meteorology // Recent Adv. Total Least Squares Techn. and Errors-in-Variables Modeling (Leuven, 1996). – Philadelphia: SIAM, 1997. – P. 249 – 258.
5. *Branham R. L.* Total least squares in astronomy // Ibid. – P. 371 – 380.
6. *Jefferys W. H.* Robust estimation when more than one variable per equation of condition has error // Biometrika. – 1990. – **77**, № 3. – P. 597 – 607.
7. *van Huffel S.* TLS applications in biometrical signal processing // Recent Adv. Total Least Squares Techn. and Errors-in-Variables Modeling (Leuven, 1996). – Philadelphia: SIAM, 1997. – P. 307 – 318.
8. *Mallick M. K.* Applications of nonlinear orthogonal distance regression in 3D Motion estimation // Ibid. – P. 273 – 284.
9. *Fuller W. A.* Measurement errors models. – New York: Wiley, 1987. – 440 p.
10. *Zwanzig S.* Estimation in nonlinear functional errors-in-variables models. – Habilitation, Univ. Hamburg, 1998.
11. *Wolter K. M., Fuller W. A.* Estimation of nonlinear errors-in-variables models // Ann. Statist. – 1982. – **10**, № 2. – P. 539 – 548.
12. *Stefanski L. A.* The effect of measurement error on parameter estimation // Biometrika. – 1985. – **72**, № 3. – P. 583 – 592.
13. *Stefanski L. A., Carroll R. J.* Covariate measurement error in logistic regression // Ann. Statist. – 1985. – **13**, № 4. – P. 1335 – 1351.
14. *Nagelkerke N. J. D.* Maximum likelihood estimation of functional relationship // Lect. Notes Statist. – 1992. – **69**. – 110 p.
15. *Armstrong B.* Measurement error in the generalized linear model // Commun. Statist.-Simula. Computa. – 1985. – **14**, № 3. – P. 529 – 544.
16. *Schafer D. W.* Covariate measurement error in generalized linear models // Boimetrika. – 1987. – **72**, № 3. – P. 583 – 592.
17. *Hillegers L. T. M. E.* The estimation in functional relationship models // Proefschrift. – Eindhoven: Techn. Univ., 1986.
18. *Amemiya Y.* Generalization of the TLS approach in the errors-in-variables problem // Recent Adv. Total Least Squares Techn. and Errors-in-Variables. Modeling (Leuven, 1996). – Philadelphia: SIAM, 1997. – P. 77 – 86.
19. *Gleser L. J.* Some new approaches to estimation in linear and nonlinear errors-in-variables regression models // Ibid. – P. 69 – 76.
20. *Kukush A. G., Zwanzig S.* On the adaptive minimum contrast estimator in a model with nonlinear functional relations // Ukr. Math. J. – 2001. – **53**, № 9. – P. 1204 – 1209.
21. *Carroll R. J., Spiegelman C. H., Lan G. K. K., Bailey K. T., Abbott R. D.* On errors-in-variables for binary regression models // Biometrika. – 1984. – **71**, № 1. – P. 1925 – 1984.
22. *Stefanski L. A.* Correcting data for measurement errors in generalized linear models // Commun. Statist. Theory Meth. – 1989. – **18**, № 5. – P. 1715 – 1733.
23. *Amemiya Y., Fuller W. F.* Estimation for the nonlinear functional relationship // Ann. Statist. – 1988. – **16**, № 1. – P. 147 – 160.
24. *Fazekas I., Kukush A. G., Zwanzig S.* On inconsistency of the least squares estimator in nonlinear functional relations. – Odense Univ., Denmark, 1998. – 30 p. – Preprint № 1.

Received 16.12.2003,
after revision – 20.04.2004