Pergamon

Computers Math. Applic. Vol. 34, No. 10, pp. 23–39, 1997

0898-1221/97 $17.00 + 0.00

PII: S0898-1221(97)00204-6

Asymptotic Properties of an Estimator in Nonlinear Functional Errors-in-Variables Models with Dependent Error Terms

I. FAZEKAS
Institute of Mathematics and Informatics, Kossuth University
P.O. Box 12, 4010 Debrecen, Hungary
fazekasi@math.klte.hu

A. G. KUKUSH*
Department of Mechanics and Mathematics, Kiev University
252601 Vladimirskaya st. 60, Kiev, Ukraine

(Received April 1997; accepted June 1997)

Abstract—Nonlinear functional errors-in-variables models are studied. An estimator for regression parameters is proposed. Consistency of the estimator is established.

Keywords—Functional errors-in-variables, Estimator, Dependent errors.

1. INTRODUCTION

Throughout the paper, the model

$$y_i = g(\xi_i, \beta_0) + \delta_i, \tag{1.1}$$
$$x_i = \xi_i + \varepsilon_i, \tag{1.2}$$

is considered, where y_i and x_i are observed, ε_i and δ_i are random error terms, ξ_i is (nonstochastic) nuisance parameter, $i = 1, 2, \dots$, g is a known function and β_0 is the true value of the unknown parameter β to be estimated.

Models having the form (1.1)–(1.2) are called errors-in-variables regression models. Errors-in-variables models have been considered for about half a century. A summary of results is given in Fuller's textbook [1].

In the literature two main classes of errors-in-variables models are distinguished, the functional case, where explanatory variables ξ_i are nonrandom, and the structural case, where ξ_i are random. In this paper discussion will be confined to the functional case.

The functional case was studied in 2.3.1 of [1] when g is linear. Asymptotic properties of the least squares estimator in the linear functional model were discussed in [2].

Concerning the structural case we refer to the following recent papers. In [3] an improvement of the naive approach was studied. In [4] a consistent procedure assuming validation data and based on least squares methods was proposed. In [5] the asymptotic properties of so-called SIMEX estimator were derived.

*This research was supported by Hungarian Foundation for Scientific Researches under Grant No. OTKA-T016933-1966 and by Hungarian Ministry of Culture and Education under Grant No. 179–1995.

Typeset by AMS-TEX

The least squares estimator of β minimizes the function

$$\frac{1}{n}\sum_{i=1}^{n}\min_{\xi\in\mathbb{R}^q}\left[(y_i-g(\xi,\beta))^2+(x_i-\xi)^2\right]. \tag{1.3}$$

In [6] an explicit proof was given for the fact that in the case of i.i.d. one-dimensional error terms under certain conditions the least squares estimator is not consistent. This result was extended for vector valued and dependent error terms $\{\varepsilon_i\}$ in [7].

Instead of the least squares estimator an alternative estimator is studied. The alternative estimator is $\hat{\beta}_n$ that minimizes expression $Q_n(\beta)$ in (2.3). This estimator has been studied in [8] in the case when ε_i, δ_i, $i = 1, 2, \dots$, are i.i.d. one-dimensional random variables. In the present paper a more general setting is studied, i.e., the case of vector valued $\{\varepsilon_i\}$ and mixing $\{\delta_i\}$. We remark that a similar approach was studied in [9].

In Section 2 the model and the estimator are described. In Section 3 theorems about the asymptotic behaviour of $\hat{\beta}_n$ are presented. Proofs are given in Sections 4 and 5.

Theorem 3.1 is a consistency result: $\hat{\beta}_n \to \beta_0$, in probability. The main tool of the proof is Utev's inequality for mixing sequences of random variables.

In Theorem 3.6 it is proved that under certain assumptions

$$P\left(\sqrt{n}\left\|\hat{\beta}_n-\beta_0\right\|>\varrho\right)\le\frac{c}{\varrho^r}.$$

In the proof Utev's inequalities for mixing sequences and inequalities for large deviations of fluctuations of random fields (see [10,11]) are used. The method of the proof contains a division of the parameter set into small parts. That approach was used by Dorogovtsev [12] and Ibragimov and Hasminskii [10]. Ivanov [13] and Prakasa Rao [14] investigated the least squares estimator in nonlinear regression models by that method.

The following notation is used. $\mathbb{N}$ is the set of positive integers, $\mathbb{R}$ is the real line, $\mathbb{R}^p$ is the p-dimensional Euclidean space with norm $\|\ \|$, $C(A)$ is the set of real valued continuous functions defined on A. We shall denote different constants with the same letter c. We shall suppose the existence of an underlying probability space $(\Omega, \mathcal{F}, P)$. $\omega \in \Omega$ denotes an elementary event. $\mathbb{E}$ stands for the expectation.

$$\|\eta\|_p = \{\mathbb{E}|\eta|^p\}^{1/p}, \qquad 1 \le p < \infty,$$

is the norm in L_p. $o_P(1)$ denotes a quantity converging to zero in probability.

2. THE MODEL AND THE ESTIMATOR

Let us consider the following model:

$$y_i = g(\xi_i, \beta_0) + \delta_i, \tag{2.1}$$

$$x_i = \xi_i + \varepsilon_i, \tag{2.2}$$

$i = 1, 2, \dots$, where design points $\xi_1, \xi_2, \dots$, are nonstochastic but not observed, y_i, x_i are observed, ε_i, δ_i are (nonobserved) random error terms ($i = 1, 2, \dots$). β_0 is the true value of the unknown parameter β to be estimated.

Suppose that the parameter set is p-dimensional: $\beta_0 \in \Theta \subset \mathbb{R}^p$, x_i, ξ_i, ε_i are q-dimensional vectors, y_i, δ_i are scalars ($i = 1, 2, \dots$), $g : \mathbb{R}^q \times \Theta \to \mathbb{R}$ is a known function.

Consider the following assumptions.

(Ai) The set $\{\delta_i : i = 1, 2, \dots\}$ is independent from $\{\varepsilon_i : i = 1, 2, \dots\}$, $\mathbb{E}\varepsilon_i = 0$, $\mathbb{E}\delta_i = 0$, $i = 1, 2, \dots$.

(Aid) $\varepsilon_1, \varepsilon_2, \dots$, are independent identically distributed (i.i.d.) random variables.

We shall suppose the existence of two auxiliary functions.

(Af) There exist an open set $U \supset \Theta$ and a function $f \in C(\mathbb{R}^q \times U)$ such that

$$\mathbb{E} f(\xi + \varepsilon_1, \beta) = g(\xi, \beta),$$

for each $\xi \in \mathbb{R}^q$, $\beta \in \Theta$.

(Ah) There exists a function $h \in C(\mathbb{R}^q \times U)$ such that

$$\mathbb{E} h(\xi + \varepsilon_1, \beta) = g^2(\xi, \beta),$$

for each $\xi \in \mathbb{R}^q$, $\beta \in \Theta$.

We consider the following modification of the least squares method. Let $\hat{\beta} = \hat{\beta}_n$ be the minimum point of

$$Q_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \left\{ (y_i - f(x_i, \beta))^2 + h(x_i, \beta) - f^2(x_i, \beta) \right\}. \tag{2.3}$$

By Pfanzagl [15] there exists a measurable solution of the above minimum problem.

Contrary to $\{\varepsilon_i\}$ the sequence $\{\delta_i\}$ is not supposed to be i.i.d. We shall use mixing conditions for the dependence of $\{\delta_i\}$. If $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras, then

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(AB) - P(A)P(B)|,$$
$$\varphi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(B \mid A) - P(B)|.$$

Let $\mathcal{M}_k^l$ denote the σ-algebra generated by $\{\delta_i : k \le i \le l\}$. Let

$$\alpha(n) = \sup_{1 \le k < \infty} \alpha\left(\mathcal{M}_1^k, \mathcal{M}_{k+n}^\infty\right),$$
$$\varphi(n) = \sup_{1 \le k < \infty} \varphi\left(\mathcal{M}_1^k, \mathcal{M}_{k+n}^\infty\right),$$
$$j(t) = 2 \min\{k \in \mathbb{N} : 2k \ge t\}.$$

The following conditions will be appropriate for our purpose:

$$a(\varphi, t) = \sum_{k=1}^{\infty} \varphi^{1/j(t)}(k)(k+1)^{j(t)-2} < \infty, \tag{Aφ}$$

$$b(\alpha, t, d) = \sum_{k=1}^{\infty} \alpha^{d/(j(t)+d)}(k)(k+1)^{j(t)-2} < \infty, \tag{Aα}$$

where $d > 0$.

We close this section with examples of functions f and h.

EXAMPLE 2.1. In the case of linear model $g(x; \tau, \gamma) = \tau + \gamma x$, where the parameter is $\beta = (\tau, \gamma)$, it is easy to see that functions $f(x; \tau, \gamma) = \tau + \gamma x$ and $h(x; \tau, \gamma) = (\tau + \gamma x)^2 - \gamma^2 \mathbb{E}\varepsilon_1^2$ satisfy assumptions (Af) and (Ah), respectively.

EXAMPLE 2.2. In the case of exponential model $g(x; \tau, \gamma) = \tau e^{\gamma x}$, where the parameter is $\beta = (\tau, \gamma)$, it is easy to see that functions $f(x; \tau, \gamma) = (\mathbb{E}e^{\gamma\varepsilon})^{-1} \tau e^{\gamma x}$ and $h(x; \tau, \gamma) = (\mathbb{E}e^{2\gamma\varepsilon})^{-1} \tau^2 e^{2\gamma x}$ satisfy assumptions (Af) and (Ah), respectively.

EXAMPLE 2.3. Consider the case of Gaussian regression curve

$$g(x; \tau, m, \gamma) = \tau \exp\left(-\frac{(x - m)^2}{2\gamma^2}\right),$$

where the parameter is $\beta = (\tau, m, \gamma)$. Assume that ε_1 is normally distributed with variance σ^2 and $\gamma > \sigma\sqrt{2}$. An easy calculation shows that functions

$$f(x; \tau, m, \gamma) = \frac{\tau\gamma}{\sqrt{\gamma^2 - \sigma^2}} \exp\left(-\frac{(x-m)^2}{2(\gamma^2 - \sigma^2)}\right), \qquad \text{and}$$

$$h(x; \tau, m, \gamma) = \frac{\tau^2\gamma}{\sqrt{\gamma^2 - 2\sigma^2}} \exp\left(-\frac{(x-m)^2}{\gamma^2 - 2\sigma^2}\right),$$

satisfy Assumptions (Af) and (Ah), respectively.

Remark that [8] gives a method to derive f and h. The method is based on Fourier transform.

3. ASYMPTOTIC BEHAVIOUR OF ESTIMATOR $\hat{\beta}_n$

3A. Consistency

First we list conditions for consistency of $\hat{\beta}_n$ in the φ-mixing case.

(A1) The parameter set Θ is compact.

(A2) g is bounded.

(A3) For each sequence $\{\xi_n\}$ and for each $\beta_* \in \Theta$ and $a > 0$

$$\liminf_{n\to\infty} \inf_{\|\beta-\beta_*\|\geq a} \frac{1}{n}\sum_{i=1}^{n} (g(\xi_i, \beta) - g(\xi_i, \beta_*))^2 > 0.$$

(A4) For each $d > 0$ there exists $l > 0$ such that for each $s \in \mathbb{R}^q$

$$\|g(s, \beta_1) - g(s, \beta_2)\| < d, \qquad \text{if } \|\beta_1 - \beta_2\| < l.$$

(A5)
$$\lim_{l\to 0} \limsup_{n\to\infty} \mathbb{E} \sup_{\|\beta_1-\beta_2\|\leq l} |f(x_n, \beta_1) - f(x_n, \beta_2)| = 0.$$

(A6)
$$\lim_{l\to 0} \limsup_{n\to\infty} \mathbb{E} \sup_{\|\beta_1-\beta_2\|\leq l} |h(x_n, \beta_1) - h(x_n, \beta_2)| = 0.$$

(A7)
$$\sup_n \mathbb{E} \, |f(x_n, \beta)|^t < \infty,$$

for each $\beta \in \Theta$.

(A8)
$$\sup_n \mathbb{E} \, |h(x_n, \beta)|^t < \infty,$$

for each $\beta \in \Theta$.

(A9)
$$\sup_n \mathbb{E} \, |\delta_n|^t < \infty.$$

The following theorem concerns the consistency of $\hat{\beta}_n$ in the φ-mixing case.

THEOREM 3.1. *Suppose that for the model (2.1),(2.2) conditions (Ai), (Aid), (Af), (Ah), (A1)–(A6) are satisfied. Moreover, assume that for a $t > 1$ conditions (Aφ), and (A7)–(A9) are satisfied. Then $\lim_{n\to\infty} \hat{\beta}_n = \beta_0$ in probability, where $\hat{\beta}_n$ is the estimator defined by* (2.3).

To prove Theorem 3.1 moment inequalities for φ-mixing sequences of random variables are used. Applying appropriate inequalities for α-mixing sequences the same result can be proved.

REMARK 3.2. Theorem 3.1 remains valid if conditions (A9) and (Aφ) are replaced with the following conditions: there exist $t > 1$ and $d > 0$ such that (A9) with exponent $t + d$ and (Aα) are satisfied.

REMARK 3.3. The independence of $\varepsilon_1, \varepsilon_2, \ldots,$ in Theorem 3.1 and Remark 3.2 can be replaced with appropriate mixing conditions.

We remark that the identical distribution of $\varepsilon_1, \varepsilon_2, \ldots,$ is not necessary, however, if $\varepsilon_1, \varepsilon_2, \ldots,$ have different distributions then instead of f and h one has to use sequences $\{f_n\}$ and $\{h_n\}$.

REMARK 3.4. Consistency of $\hat{\beta}_n$ can be proved for vector valued g, too. In that case f will be also vector valued.

REMARK 3.5. Instead of (A2), one can suppose that $\xi_1, \xi_2, \ldots,$ belong to a compact set and g is continuous.

Instead of (A5), one can suppose that

(A5′)
$$\lim_{l \to 0} \sup_{\xi} \mathbb{E} \sup_{\|\beta_1 - \beta_2\| \le l} |f(\xi + \varepsilon_1, \beta_1) - f(\xi + \varepsilon_1, \beta_2)| = 0.$$

Instead of (A7), one can suppose that

(A7′)
$$\sup_{\xi} \mathbb{E}\, |f(\xi + \varepsilon_1, \beta)|^t < \infty,$$

for each $\beta \in \Theta$.

The same applies for conditions (A6) and (A8), too.

3B. Large Deviation

Now, we list assumptions for large deviation result.

(A1*) The parameter set Θ is compact and convex.

Put
$$\Psi_n(\beta_1, \beta_2) = \frac{1}{n} \sum_{i=1}^{n} (g(\xi_i, \beta_1) - g(\xi_i, \beta_2))^2,$$
where $\beta_1,\ \beta_2 \in \Theta$.

(A3*) There exist $0 < K_1 \le K_2 < \infty$ such that for each $\beta_1,\ \beta_2 \in \Theta,\ n \in \mathbb{N}$
$$K_1 \|\beta_1 - \beta_2\|^2 \le \Psi_n(\beta_1, \beta_2) \le K_2 \|\beta_1 - \beta_2\|^2.$$

(A5*)
$$\mathbb{E} \sup_{\beta \in \Theta} \left\| \frac{\partial f(\xi + \varepsilon_1, \beta)}{\partial \beta} \right\|^r$$
is a bounded function of ξ.

(A6*)
$$\mathbb{E} \sup_{\beta \in \Theta} \left\| \frac{\partial h(\xi + \varepsilon_1, \beta)}{\partial \beta} \right\|^r$$
is a bounded function of ξ.

THEOREM 3.6. *Suppose that for the model (2.1),(2.2) assumptions (Ai), (Aid), (Af), (Ah), (A1*), (A2), (A3*) are satisfied. Moreover, assume that there exists r with $2 \le r$, $p < r$ such that (Aφ) with $t = r$, (A5*), (A6*), and (A9) with $t = r$ are satisfied. Then, there exists $c > 0$ such that for each $\varrho > 0$ and $n \in \mathbb{N}$*
$$P\left(\sqrt{n}\left\|\hat{\beta}_n - \beta_0\right\| > \varrho\right) \le \frac{c}{\varrho^r}.$$

In the α-mixing case the following modification of Theorem 3.6 is valid.

REMARK 3.7. Suppose that for the model (2.1)–(2.2) assumptions (Ai), (Aid), (Af), (Ah), (A1*), (A2), (A3*) are satisfied. Moreover, assume that there exist $d > 0$ and r with $2 \le r$, $p < r$ such that (Aα) with $t = r$, (A9) with $t = r + d$, (A5*), and (A6*) are satisfied. Then, there exists $c > 0$ such that for each $\varrho > 0$ and $n \in \mathbb{N}$
$$P\left(\sqrt{n}\left\|\hat{\beta}_n - \beta_0\right\| > \varrho\right) \le \frac{c}{\varrho^r}.$$

REMARK 3.8. The independence of $\varepsilon_1, \varepsilon_2, \ldots,$ in Theorem 3.6 and Remark 3.7 can be replaced with appropriate mixing conditions.

4. PROOF OF CONSISTENCY

We need the following inequalities of Utev [16].

LEMMA 4.1. *Let η_i be measurable with respect to the σ-algebra generated by δ_i, $\mathbb{E}\eta_i = 0$, $i = 1, 2, \ldots$.*

(i) *In the φ-mixing case:*

$$\mathbb{E}\left|\sum_{i=1}^{n}\eta_i\right|^t \le c_1(t)a(\varphi,t)\sum_{i=1}^{n}\mathbb{E}\,|\eta_i|^t,$$

if $1 \le t \le 2$;

$$\mathbb{E}\left|\sum_{i=1}^{n}\eta_i\right|^t \le c_1(t)a(\varphi,t)\max\left\{\sum_{i=1}^{n}\mathbb{E}\,|\eta_i|^t, \left(\sum_{i=1}^{n}\mathbb{E}\,(\eta_i)^2\right)^{t/2}\right\},$$

if $2 \le t$. In particular,

$$\mathbb{E}\left|\sum_{i=1}^{n}a_i\eta_i\right|^t \le c_1(t)a(\varphi,t)\left(\sum_{i=1}^{n}a_i^2\right)^{t/2}\max_{1\le i\le n}\mathbb{E}\,|\eta_i|^t, \tag{4.1}$$

if $2 \le t$.

(ii) *In the α-mixing case let $d > 0$, then*

$$\mathbb{E}\left|\sum_{i=1}^{n}\eta_i\right|^t \le c_2(t)b(\alpha,t,d)\sum_{i=1}^{n}\|\eta_i\|_{t+d}^t,$$

if $1 \le t \le 2$;

$$\mathbb{E}\left|\sum_{i=1}^{n}\eta_i\right|^t \le c_2(t)b(\alpha,t,d)\max\left\{\sum_{i=1}^{n}\|\eta_i\|_{t+d}^t, \left(\sum_{i=1}^{n}\|\eta_i\|_{2+d}^2\right)^{t/2}\right\},$$

if $2 \le t$. In particular

$$\mathbb{E}\left|\sum_{i=1}^{n}a_i\eta_i\right|^t \le c_2(t)b(\alpha,t,d)\left(\sum_{i=1}^{n}a_i^2\right)^{t/2}\max_{1\le i\le n}\|\eta_i\|_{t+d}^t, \tag{4.2}$$

if $2 \le t$. (Here $c_1(t)$ and $c_2(t)$ depend on t (and $c_2(t)$ on the dimension in the multidimensional case) but they do not depend on n.)

The proof can be found in [16], Theorems 2.1 and 2.2, and Corollaries 2.1 and 2.2. ∎

We shall intensively use c_p-, Chebyshev's, and Jensen's inequalities without explicitly mentioning them.

LEMMA 4.2. *Let $\lim_{n\to\infty} U_n(\beta) = 0$ in probability for each $\beta \in \Theta$, where Θ is a compact set. Suppose that for each $\varepsilon > 0$*

$$\lim_{l\to 0}\limsup_{n\to\infty} P\left\{\sup_{\|\beta_1-\beta_2\|\le l}|U_n(\beta_1) - U_n(\beta_2)| > \varepsilon\right\} = 0. \tag{4.3}$$

Then

$$\lim_{n\to\infty}\sup_{\beta\in\Theta}|U_n(\beta)| = 0, \tag{4.4}$$

in probability.

PROOF OF LEMMA 4.2. Let $l > 0$. By compactness of Θ there exist $\beta_1, \ldots, \beta_m \in \Theta$ such that

$$\Theta \subseteq \bigcup_{k=1}^{m} \{\beta : \|\beta - \beta_k\| \le l\}.$$

Then

$$\begin{aligned} P\left\{\sup_{\beta\in\Theta} |U_n(\beta)| > \varepsilon\right\} &\le P\left\{\sup_{k=1,\ldots,m} |U_n(\beta_k)| > \frac{\varepsilon}{2}\right\} \\ &\quad + P\left\{\sup_{k=1,\ldots,m}\ \sup_{\|\beta-\beta_k\|\le l} |U_n(\beta) - U_n(\beta_k)| > \frac{\varepsilon}{2}\right\} \\ &\le \sum_{k=1}^{m} P\left\{|U_n(\beta_k)| > \frac{\varepsilon}{2}\right\} + P\left\{\sup_{\|\beta_1-\beta_2\|\le l} |U_n(\beta_1) - U_n(\beta_2)| > \frac{\varepsilon}{2}\right\}. \end{aligned}$$

Therefore

$$\limsup_{n\to\infty} P\left\{\sup_{\beta\in\Theta} |U_n(\beta)| > \varepsilon\right\} \le \limsup_{n\to\infty} P\left\{\sup_{\|\beta_1-\beta_2\|\le l} |U_n(\beta_1) - U_n(\beta_2)| > \frac{\varepsilon}{2}\right\}.$$

By $l \to 0$ we obtain the result. The lemma is proved.

PROOF THEOREM 3.1. Using abbreviations

$$g_i(\beta) = g(\xi_i, \beta), \qquad f_i(\beta) = f(x_i, \beta), \qquad h_i(\beta) = h(x_i, \beta), \tag{4.5}$$

we have $y_i = g_i(\beta_0) + \delta_i$ and

$$\begin{aligned} Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} &\Big\{(g_i(\beta) - g_i(\beta_0))^2 - 2\,(f_i(\beta) - g_i(\beta))\,g_i(\beta_0) \\ &- \left(g_i^2(\beta) - h_i(\beta)\right) + 2\delta_i g_i(\beta_0) - 2\delta_i f_i(\beta) + \delta_i^2\Big\}. \end{aligned} \tag{4.6}$$

As δ_i and $g_i(\beta_0)$ do not depend on β, we shall minimize the following expression:

$$\tilde{Q}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \{(g_i(\beta) - g_i(\beta_0))^2 - 2(f_i(\beta) - g_i(\beta))g_i(\beta_0) - (g_i^2(\beta) - h_i(\beta)) - 2\delta_i f_i(\beta)\}. \tag{4.7}$$

Later we shall prove that

$$\frac{1}{n}\sum_{i=1}^{n} \{2(f_i(\beta) - g_i(\beta))g_i(\beta_0) + 2\delta_i f_i(\beta) + (g_i^2(\beta) - h_i(\beta))\} = o_P(1), \tag{4.8}$$

as $n \to \infty$, uniformly in $\beta \in \Theta$.

Now we prove that (4.8) and condition (A3) imply that the minimum point $\hat{\beta}_n$ of $\tilde{Q}_n(\beta)$ is consistent. (4.8) and $\tilde{Q}_n(\hat{\beta}_n) \le \tilde{Q}_n(\beta_0)$ imply that

$$\frac{1}{n}\sum_{i=1}^{n} \left(g\left(\xi_i, \hat{\beta}_n\right) - g(\xi_i, \beta_0)\right)^2 \le o_P(1), \tag{4.9}$$

as $n \to \infty$.

Condition (A3) implies that for arbitrary but fixed $\{\xi_i,\ i = 1, 2, \ldots\}$ and $a > 0$ there exists $\bar{\varepsilon} > 0$ such that

$$\inf_{\|\beta-\beta_0\|\ge a} \frac{1}{n}\sum_{i=1}^{n} [g(\xi_i, \beta) - g(\xi_i, \beta_0)]^2 > \bar{\varepsilon}, \tag{4.10}$$

if $n > n_{\bar{\varepsilon}}$. For fixed $a > 0$ and $n_0 > n_{\bar{\varepsilon}}$, let $A_{n_0}(a)$ denote the following event:

$$A_{n_0}(a) = \left\{\omega : \left\|\hat{\beta}_{n_0}(\omega) - \beta_0\right\| \geq a\right\}. \tag{4.11}$$

Then for $\omega \in A_{n_0}(a)$ we have

$$\frac{1}{n_0}\sum_{i=1}^{n_0}\left[g\left(\xi_i, \hat{\beta}_{n_0}\right) - g(\xi_i, \beta_0)\right]^2 > \bar{\varepsilon}.$$

Therefore, for each $a > 0$ there exists $\bar{\varepsilon} > 0$ such that

$$P\left(\left\|\hat{\beta}_{n_0} - \beta_0\right\| \geq a\right) \leq P\left(\frac{1}{n_0}\sum_{i=1}^{n_0}\left[g\left(\xi_i, \hat{\beta}_{n_0}\right) - g(\xi_i, \beta_0)\right]^2 > \bar{\varepsilon}\right), \tag{4.12}$$

if $n_0 > n_{\bar{\varepsilon}}$. Now, (4.9) and (4.12) imply that $\hat{\beta}_n \to \beta_0$ in probability, as $n \to \infty$.

It remained to prove (4.8). To prove (4.8), we shall use Lemma 4.2 for components of the sum in (4.8).

First, let

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(f(x_i, \beta) - g(\xi_i, \beta)\right) g(\xi_i, \beta_0). \tag{4.13}$$

The summands are independent with zero expectation. For $1 < t \leq 2$

$$\begin{aligned}
\mathbb{E}|U_n(\beta)|^t &\leq c\frac{1}{n^t}\sum_{i=1}^{n}\mathbb{E}\left|\left(f(x_i, \beta) - g(\xi_i, \beta)\right) g(\xi_i, \beta_0)\right|^t \\
&\leq c\frac{1}{n^{t-1}}\max_{1\leq i\leq n}|g(\xi_i, \beta_0)|^t \max_{1\leq i\leq n}\mathbb{E}|f(x_i, \beta) - g(\xi_i, \beta)|^t \\
&\leq c\frac{1}{n^{t-1}}\max_{1\leq i\leq n}|g(\xi_i, \beta_0)|^t\left\{\max_{1\leq i\leq n}\mathbb{E}|f(x_i, \beta)|^t + \max_{1\leq i\leq n}|g(\xi_i, \beta)|^t\right\} \to 0,
\end{aligned}$$

as $n \to \infty$, because of assumptions (A2) and (A7). Therefore, $U_n(\beta) \to 0$ in probability.

Now, we prove condition (4.3) for $U_n(\beta)$ defined in (4.13).

$$\begin{aligned}
\sup_{\|\beta_1-\beta_2\|\leq l}|U_n(\beta_1) - U_n(\beta_2)| \leq{}& \sup_{\|\beta_1-\beta_2\|\leq l}\frac{1}{n}\sum_{i=1}^{n}|g(\xi_i, \beta_0)|\,|f(x_i, \beta_1) - f(x_i, \beta_2)| \\
&+ \sup_{\|\beta_1-\beta_2\|\leq l}\frac{1}{n}\sum_{i=1}^{n}|g(\xi_i, \beta_0)|\,|g(\xi_i, \beta_1) - g(\xi_i, \beta_2)|.
\end{aligned} \tag{4.14}$$

The second summand in the right-hand side of (4.14) tends to zero, as $l \to 0$, because of assumption (A4). The expectation of the first summand in the right-hand side of (4.14) is majorized by

$$c\sup_i\mathbb{E}\sup_{\|\beta_1-\beta_2\|\leq l}|f(x_i, \beta_1) - f(x_i, \beta_2)|. \tag{4.15}$$

Therefore, assumption (A5) and Chebyshev's inequality imply that assumption (4.3) is satisfied for $U_n(\beta)$ defined by (4.13).

Now, we have to prove that

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{h(x_i, \beta) - g^2(\xi_i, \beta)\right\}, \tag{4.16}$$

satisfies conditions of Lemma 4.2. As the summands are independent with zero expectation, we have for $1 < t \le 2$

$$\begin{aligned}\mathbb{E}|U_n(\beta)|^t &\le c\frac{1}{n^t}\sum_{i=1}^{n}\mathbb{E}|h(x_i,\beta)-g^2(\xi_i,\beta)|^t \\ &\le c\frac{1}{n^{t-1}}\max_{1\le i\le n}\mathbb{E}|h(x_i,\beta)|^t,\end{aligned} \tag{4.17}$$

and by (A8) the last expression tends to 0, as $n \to \infty$. So, $U_n(\beta) \to 0$ in probability.

To prove condition (4.3), consider

$$\begin{aligned}\sup_{\|\beta_1-\beta_2\|\le l}|U_n(\beta_1)-U_n(\beta_2)| &\le \sup_{\|\beta_1-\beta_2\|\le l}\frac{1}{n}\sum_{i=1}^{n}|h(x_i,\beta_1)-h(x_i,\beta_2)| \\ &+ \sup_{\|\beta_1-\beta_2\|\le l}\frac{1}{n}\sum_{i=1}^{n}|g^2(\xi_i,\beta_1)-g^2(\xi_i,\beta_2)| = B_1+B_2,\end{aligned} \tag{4.18}$$

say. Here, B_2 tends to 0, as $l \to 0$, because of conditions (A2) and (A4). Furthermore,

$$\lim_{l\to 0}\limsup_{n\to\infty}\mathbb{E}|B_1| = 0,$$

by (A6). Therefore, Chebyshev's inequality implies that condition (4.3) is satisfied for $U_n(\beta)$ defined by (4.16).

Now, we prove that

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} f(x_i,\beta)\delta_i, \tag{4.19}$$

satisfies conditions of Lemma 4.2. As $\{\varepsilon_i\}$ and $\{\delta_i\}$ are independent and $\mathbb{E}\delta_i = 0$, the summands have zero expectation. By independence and (4.1)

$$\begin{aligned}\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i,\beta)\delta_i\right|^t &= \mathbb{E}_\varepsilon\left(\mathbb{E}_\delta\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i,\beta)\delta_i\right|^t\right) \\ &\le \mathbb{E}_\varepsilon\left(\frac{1}{n^t}\sum_{i=1}^{n}|f(x_i,\beta)|^t\,\mathbb{E}_\delta|\delta_i|^t\right) = \frac{1}{n^t}\sum_{i=1}^{n}\mathbb{E}|f(x_i,\beta)|^t\mathbb{E}|\delta_i|^t,\end{aligned} \tag{4.20}$$

where $\mathbb{E}_\varepsilon$ (respectively, $\mathbb{E}_\delta$), denotes the expectation with respect to $\{\varepsilon_i : i = 1,2,\dots\}$ (respectively, $\{\delta_i : i = 1,2,\dots\}$). By (A7) and (A9) the last expression in (4.20) converges to zero, if $n \to \infty$. Therefore, $U_n(\beta) \to 0$ in probability.

By independence

$$\mathbb{E}\sup_{\|\beta_1-\beta_2\|\le l}|U_n(\beta_1)-U_n(\beta_2)| \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|\le l}|f(x_i,\beta_1)-f(x_i,\beta_2)|\right)\mathbb{E}|\delta_i|.$$

Therefore, by (A9) and (A5)

$$\lim_{l\to 0}\limsup_{n\to\infty}\mathbb{E}\sup_{\|\beta_1-\beta_2\|\le l}|U_n(\beta_1)-U_n(\beta_2)| = 0,$$

so conditions of Lemma 4.2 are satisfied. ∎

5. PROOF OF UPPER BOUNDS FOR LARGE DEVIATIONS

PROOF OF THEOREM 3.6. Remind notation used in the proof of Theorem 3.1.

$$g_i(\beta) = g(\xi_i, \beta), \qquad f_i(\beta) = f(x_i, \beta), \qquad h_i(\beta) = h(x_i, \beta), \tag{5.1}$$

for $\beta \in \Theta$. Then, our model is

$$\begin{aligned} y_i &= g_i(\beta_0) + \delta_i, \\ x_i &= \xi_i + \varepsilon_i. \end{aligned} \tag{5.2}$$

$Q_n(\beta)$ can be written in the form:

$$\begin{aligned} Q_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \Big\{ &(g_i(\beta) - g_i(\beta_0))^2 + 2\delta_i g_i(\beta_0) - 2\delta_i f_i(\beta) \\ &- 2g_i(\beta_0)\left(f_i(\beta) - g_i(\beta)\right) + \left(h_i(\beta) - g_i^2(\beta)\right) + \delta_i^2 \Big\}. \end{aligned} \tag{5.3}$$

We shall use the following notation:

$$\Psi_n(\beta_1, \beta_2) = \frac{1}{n} \sum_{i=1}^{n} (g_i(\beta_1) - g_i(\beta_2))^2, \tag{5.4}$$

$$Y_n(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \left(f_i(\beta) - f_i(\beta_0)\right), \tag{5.5}$$

$$Z_n^{(1)}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i(\beta_0) \left(f_i(\beta) - g_i(\beta)\right), \tag{5.6}$$

$$Z_n^{(2)}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h_i(\beta) - g_i^2(\beta)\right). \tag{5.7}$$

As $\hat\beta = \hat\beta_n$ is minimum point of $Q_n(\beta)$, we have

$$Q_n\left(\hat\beta\right) \le Q_n(\beta_0). \tag{5.8}$$

Using (5.3)–(5.7), we get from inequality (5.8) that

$$\sqrt{n}\Psi_n\left(\hat\beta, \beta_0\right) - 2Y_n\left(\hat\beta\right) - 2\left[Z_n^{(1)}\left(\hat\beta\right) - Z_n^{(1)}(\beta_0)\right] + \left[Z_n^{(2)}\left(\hat\beta\right) - Z_n^{(2)}(\beta_0)\right] \le 0. \tag{5.9}$$

Let $\varrho > 0$. On event $\{\sqrt{n}\|\hat\beta - \beta_0\| > \varrho\}$, by (A3*), $\Psi_n(\hat\beta, \beta_0) > 0$, therefore (5.9) gives

$$\frac{Y_n\left(\hat\beta\right)}{\sqrt{n}\Psi_n\left(\hat\beta, \beta_0\right)} + \frac{Z_n^{(1)}\left(\hat\beta\right) - Z_n^{(1)}(\beta_0)}{\sqrt{n}\Psi_n\left(\hat\beta, \beta_0\right)} - \frac{Z_n^{(2)}\left(\hat\beta\right) - Z_n^{(2)}(\beta_0)}{2\sqrt{n}\Psi_n\left(\hat\beta, \beta_0\right)} \ge \frac{1}{2}. \tag{5.10}$$

Introduce the following events:

$$B_1 = \left\{ \frac{\left|Y_n\left(\hat\beta\right)\right|}{\sqrt{n}\Psi_n\left(\hat\beta, \beta_0\right)} \ge \frac{1}{4} \right\}, \tag{5.11}$$

$$B_2 = \left\{ \frac{\left|Z_n^{(1)}\left(\hat\beta\right) - Z_n^{(1)}(\beta_0)\right|}{\sqrt{n}\Psi_n\left(\hat\beta, \beta_0\right)} \ge \frac{1}{8} \right\}, \tag{5.12}$$

$$B_3 = \left\{ \frac{\left|Z_n^{(2)}\left(\hat\beta\right) - Z_n^{(2)}(\beta_0)\right|}{2\sqrt{n}\Psi_n\left(\hat\beta, \beta_0\right)} \ge \frac{1}{8} \right\}, \tag{5.13}$$

$$A_{n\varepsilon} = \left\{ \left\|\hat\beta - \beta_0\right\| > \varepsilon \right\}, \qquad \varepsilon > 0. \tag{5.14}$$

Then

$$P(A_{n\varepsilon}) \le P(A_{n\varepsilon}\cap B_1) + P(A_{n\varepsilon}\cap B_2) + P(A_{n\varepsilon}\cap B_3)$$
$$\le P\left(\sup_{\|\beta-\beta_0\|>\varepsilon} \frac{|Y_n(\beta)|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \ge \frac{1}{4}\right)$$
$$+ P\left(\sup_{\|\beta-\beta_0\|>\varepsilon} \frac{\left|Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0)\right|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \ge \frac{1}{8}\right) \quad (5.15)$$
$$+ P\left(\sup_{\|\beta-\beta_0\|>\varepsilon} \frac{\left|Z_n^{(2)}(\beta) - Z_n^{(2)}(\beta_0)\right|}{2\sqrt{n}\Psi_n(\beta,\beta_0)} \ge \frac{1}{8}\right).$$

We shall put $\varepsilon = \varrho/\sqrt{n}$ into (5.15) and estimate each summand separately. For each summand in (5.15), we shall use the following decomposition:

$$P\left(\sup_{\|\beta-\beta_0\|>(\varrho/\sqrt{n})} \frac{|T_n(\beta)|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \ge c\right) \le P\left(\sup_{\|\beta-\beta_0\|>\varrho} \frac{|T_n(\beta)|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \ge c\right)$$
$$+ P\left(\sup_{(\varrho/\sqrt{n})\le\|\beta-\beta_0\|\le\varrho} \frac{|T_n(\beta)|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \ge c\right). \quad (5.16)$$

In (5.16) we shall choose $T_n(\beta)$ as $Y_n(\beta)$, $Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0)$, and $Z_n^{(2)}(\beta) - Z_n^{(2)}(\beta_0)$.

Let us start with the first term in (5.16).

Consider the case of $T_n(\beta) = Y_n(\beta)$. By Cauchy's inequality

$$\frac{Y_n^2(\beta)}{n\Psi_n(\beta,\beta_0)} \le \frac{1}{n}\sum_{i=1}^n \delta_i^2 \frac{1}{n}\sum_{i=1}^n \frac{(f_i(\beta)-f_i(\beta_0))^2}{\Psi_n(\beta,\beta_0)}$$
$$\le \frac{1}{n}\sum_{i=1}^n \delta_i^2 \frac{1}{n}\sum_{i=1}^n \left\|\frac{\partial f_i(\bar{\beta})}{\partial\beta}\right\|^2 \frac{\|\beta-\beta_0\|^2}{\Psi_n(\beta,\beta_0)}. \quad (5.17)$$

By (A3*), $(\|\beta-\beta_0\|^2)/(\Psi_n(\beta,\beta_0))$ is bounded, therefore (5.17) implies

$$\mathbb{E}\sup_{\|\beta-\beta_0\|>\varrho}\left(\frac{|Y_n(\beta)|}{\sqrt{n}\sqrt{\Psi_n(\beta,\beta_0)}}\right)^r \le c\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n\delta_i^2\right)^{r/2}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n\sup_\beta\left\|\frac{\partial f_i(\beta)}{\partial\beta}\right\|^2\right)^{r/2}$$
$$\le c\frac{1}{n}\sum_{i=1}^n\mathbb{E}|\delta_i|^r\frac{1}{n}\sum_{i=1}^n\mathbb{E}\sup_\beta\left\|\frac{\partial f_i(\beta)}{\partial\beta}\right\|^r < \infty, \quad (5.18)$$

where we used independence, (A5*), and Jensen's inequality. Now by (A3*), (5.18), and Chebyshev's inequality:

$$P\left(\sup_{\|\beta-\beta_0\|>\varrho}\frac{|Y_n(\beta)|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \ge c\right) \le P\left(\sup_{\|\beta-\beta_0\|>\varrho}\frac{|Y_n(\beta)|}{\sqrt{n}\sqrt{\Psi_n(\beta,\beta_0)}} \ge c\varrho\right) \le \frac{c}{\varrho^r}. \quad (5.19)$$

Now, we consider $T_n(\beta) = Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0)$. By c_p-, Jensen's, and Lyapunov's inequalities

$$\left(\frac{\left|Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0)\right|}{\sqrt{n}\sqrt{\Psi_n(\beta,\beta_0)}}\right)^r$$
$$\le \frac{c}{(\Psi_n(\beta,\beta_0))^{r/2}}\left\{\left(\frac{1}{n}\sum_{i=1}^n|f_i(\beta)-f_i(\beta_0)|\right)^r + \left(\frac{1}{n}\sum_{i=1}^n|g_i(\beta)-g_i(\beta_0)|\right)^r\right\} \quad (5.20)$$

$$\leq \frac{c}{(\Psi_n(\beta,\beta_0))^{r/2}} \left\{ \frac{1}{n} \sum_{i=1}^{n} |f_i(\beta) - f_i(\beta_0)|^r + \left(\frac{1}{n} \sum_{i=1}^{n} (g_i(\beta) - g_i(\beta_0))^2 \right)^{r/2} \right\}$$

$$\leq \frac{c}{(\Psi_n(\beta,\beta_0))^{r/2}} \frac{1}{n} \sum_{i=1}^{n} \sup_{\beta} \left\| \frac{\partial f_i(\beta)}{\partial \beta} \right\|^r \|\beta - \beta_0\|^r + c \tag{5.20 (cont.)}$$

$$\leq c \frac{1}{n} \sum_{i=1}^{n} \sup_{\beta} \left\| \frac{\partial f_i(\beta)}{\partial \beta} \right\|^r + c.$$

Therefore,

$$\mathbb{E} \left(\sup_{\|\beta-\beta_0\|>\varrho} \frac{\left| Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0) \right|}{\sqrt{n}\sqrt{\Psi_n(\beta,\beta_0)}} \right)^r \leq c \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \sup_{\beta} \left\| \frac{\partial f_i(\beta)}{\partial \beta} \right\|^r + c < \infty, \tag{5.21}$$

by (A5*).

So, by Chebyshev's inequality and (A3*)

$$\begin{aligned} &P\left(\sup_{\|\beta-\beta_0\|>\varrho} \frac{\left| Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0) \right|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \geq c \right) \\ &\quad \leq P\left(\sup_{\|\beta-\beta_0\|>\varrho} \frac{\left| Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0) \right|}{\sqrt{n}\sqrt{\Psi_n(\beta,\beta_0)}} \geq c\varrho \right) \leq \frac{c}{\varrho^r}. \end{aligned} \tag{5.22}$$

Now, consider $T_n(\beta) = Z_n^{(2)}(\beta) - Z_n^{(2)}(\beta_0)$.

$$\begin{aligned} &\left(\frac{\left| Z_n^{(2)}(\beta) - Z_n^{(2)}(\beta_0) \right|}{\sqrt{n}\sqrt{\Psi_n(\beta,\beta_0)}} \right)^r \\ &\leq \frac{c}{(\Psi_n(\beta,\beta_0))^{r/2}} \left\{ \left| \frac{1}{n} \sum_{i=1}^{n} (h_i(\beta) - h_i(\beta_0)) \right|^r + \left| \frac{1}{n} \sum_{i=1}^{n} (g_i^2(\beta) - g_i^2(\beta_0)) \right|^r \right\} \\ &\leq \frac{c}{(\Psi_n(\beta,\beta_0))^{r/2}} \frac{1}{n} \sum_{i=1}^{n} |h_i(\beta) - h_i(\beta_0)|^r + \frac{c}{(\Psi_n(\beta,\beta_0))^{r/2}} \left| \frac{1}{n} \sum_{i=1}^{n} (g_i^2(\beta) - g_i^2(\beta_0)) \right|^r. \end{aligned} \tag{5.23}$$

The second term in the right-hand side of (5.23) is bounded because of the Cauchy inequality, while the first term is bounded by

$$c \frac{1}{n} \sum_{i=1}^{n} \sup_{\beta} \left\| \frac{\partial h_i(\beta)}{\partial \beta} \right\|^r.$$

Therefore, using (A3*) and (A6*), Chebyshev's inequality implies

$$P\left(\sup_{\|\beta-\beta_0\|>\varrho} \frac{\left| Z_n^{(2)}(\beta) - Z_n^{(2)}(\beta_0) \right|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \geq c \right) \leq \frac{c}{\varrho^r}. \tag{5.24}$$

To estimate

$$P\left(\sup_{(\varrho/\sqrt{n}) \leq \|\beta-\beta_0\| \leq \varrho} \frac{|T_n(\beta)|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \geq c \right), \tag{5.25}$$

we need the following lemma.

LEMMA 5.1. *Let $\delta = \varrho/\sqrt{n}$, $\varrho_m = m\delta$, $m = 1, 2, \ldots$. Suppose that for fixed constants $0 < F_1 \leq 1 \leq F_2 < \infty$*

$$\mathbb{E}\left|\frac{T_n(\beta)}{\varrho_m}\right|^r < c, \qquad \text{if } F_1\varrho_m \leq \|\beta - \beta_0\| \leq F_2\varrho_{m+1}, \tag{5.26}$$

for $m = 1, 2, \ldots$, where c does not depend on m. Suppose furthermore

$$\mathbb{E}|T_n(\beta_1) - T_n(\beta_2)|^r \leq c\|\beta_1 - \beta_2\|^r. \tag{5.27}$$

Then,

$$P\left(\sup_{(\varrho/\sqrt{n})\leq\|\beta-\beta_0\|\leq\varrho} \frac{|T_n(\beta)|}{\sqrt{n}\Psi_n(\beta, \beta_0)} \geq c\right) \leq \frac{c}{\varrho^r}. \tag{5.28}$$

We postpone the proof of Lemma 5.1 and continue the proof of Theorem 3.6. Now, we shall check assumptions of Lemma 5.1 for $T_n(\beta) = Y_n(\beta)$, $Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0)$, and $Z_n^{(2)}(\beta) - Z_n^{(2)}(\beta_0)$.

Let us start with $T_n(\beta) = Y_n(\beta)$. By independence, (4.1), Jensen's inequality, (A5*), and (A9)

$$\begin{aligned}\mathbb{E}\left|\frac{Y_n(\beta)}{\varrho_m}\right|^r &= \mathbb{E}\left|\frac{1}{\varrho_m}\frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\left[f_i(\beta) - f_i(\beta_0)\right]\right|^r \\ &\leq c\frac{1}{\varrho_m^r}\max_{1\leq i\leq n}\mathbb{E}|\delta_i|^r\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n\left[f_i(\beta) - f_i(\beta_0)\right]^2\right)^{r/2} \\ &\leq c\frac{1}{\varrho_m^r}\mathbb{E}\frac{1}{n}\sum_{i=1}^n|f_i(\beta) - f_i(\beta_0)|^r \\ &\leq c\frac{1}{\varrho_m^r}\frac{1}{n}\sum_{i=1}^n\mathbb{E}\sup_\beta\left\|\frac{\partial f_i(\beta)}{\partial\beta}\right\|^r\|\beta - \beta_0\|^r \leq c\left(\frac{\varrho_{m+1}}{\varrho_m}\right)^r < c,\end{aligned} \tag{5.29}$$

if $\|\beta - \beta_0\| \leq F_2\varrho_{m+1}$. Therefore, (5.26) is satisfied.

Now, we turn to Assumption (5.27). By the same method as in (5.29), we get

$$\begin{aligned}\mathbb{E}\left|Y_n(\beta_1) - Y_n(\beta_2)\right|^r &= \mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\delta_i\left[f_i(\beta_1) - f_i(\beta_2)\right]\right|^r \\ &\leq c\frac{1}{n}\sum_{i=1}^n\mathbb{E}\sup_\beta\left\|\frac{\partial f_i(\beta)}{\partial\beta}\right\|^r\|\beta_1 - \beta_2\|^r \leq c\|\beta_1 - \beta_2\|^r.\end{aligned} \tag{5.30}$$

We check assumptions of Lemma 5.1 for

$$T_n(\beta) = Z_n^{(1)}(\beta) - Z_n^{(1)}(\beta_0).$$

We have

$$\begin{aligned}\mathbb{E}|T_n(\beta)|^r &= \mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i(\beta_0)\left[(f_i(\beta) - f_i(\beta_0)) - (g_i(\beta) - g_i(\beta_0))\right]\right|^r \\ &\leq c\max\left\{\frac{1}{n^{r/2}}\sum_{i=1}^n\mathbb{E}\left|f_i(\beta) - f_i(\beta_0) - g_i(\beta) + g_i(\beta_0)\right|^r,\right. \\ &\qquad\left.\left(\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left(f_i(\beta) - f_i(\beta_0) - g_i(\beta) + g_i(\beta_0)\right)^2\right)^{r/2}\right\}.\end{aligned} \tag{5.31}$$

The first expression behind the max sign is majorized by

$$
\begin{aligned}
& c\left\{\frac{1}{n^{r/2}}\sum_{i=1}^{n}\mathbb{E}\left|f_i(\beta)-f_i(\beta_0)\right|^r+\frac{1}{n^{r/2}}\sum_{i=1}^{n}\left|g_i(\beta)-g_i(\beta_0)\right|^r\right\} \\
&\quad \le c\frac{1}{n^{r/2}}\sum_{i=1}^{n}\mathbb{E}\sup_{\beta}\left\|\frac{\partial f_i(\beta)}{\partial\beta}\right\|^r\|\beta-\beta_0\|^r+c\frac{1}{n^{r/2}}\left(\sum_{i=1}^{n}\left(g_i(\beta)-g_i(\beta_0)\right)^2\right)^{r/2} \\
&\quad \le c\|\beta-\beta_0\|^r+c\left(\Psi(\beta,\beta_0)\right)^{r/2}\le c\|\beta-\beta_0\|^r,
\end{aligned} \tag{5.32}
$$

where we used (A5*), (A3*), and $r \geq 2$. (Remark that $(\sum_{i=1}^{n}|a_i|^r)^{1/r}$ is a decreasing function of r.)

The second expression behind the max sign in (5.31) is majorized by

$$
\begin{aligned}
& \frac{c}{n^{r/2}}\left(\sum_{i=1}^{n}\mathbb{E}\left(f_i(\beta)-f_i(\beta_0)\right)^2+\left(g_i(\beta)-g_i(\beta_0)\right)^2\right)^{r/2} \\
&\quad \le c\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(f_i(\beta)-f_i(\beta_0)\right)^2\right)^{r/2}+c\left(\frac{1}{n}\sum_{i=1}^{n}\left(g_i(\beta)-g_i(\beta_0)\right)^2\right)^{r/2} \\
&\quad \le c\|\beta-\beta_0\|^r.
\end{aligned} \tag{5.33}
$$

Inequalities (5.31)–(5.33) imply $\mathbb{E}|T_n(\beta)|^r \le c\|\beta-\beta_0\|^r$, and therefore

$$
\mathbb{E}\left|\frac{T_n(\beta)}{\varrho_m}\right|^r < c, \qquad \text{if } \|\beta-\beta_0\| \le F_2\varrho_{m+1},
$$

so (5.26) is satisfied.

Now, $T_n(\beta_1)-T_n(\beta_2) = Z_n^{(1)}(\beta_1)-Z_n^{(1)}(\beta_2)$ has the same structure as the expression analysed above, therefore

$$
\mathbb{E}\left|T_n(\beta_1)-T_n(\beta_2)\right|^r=\mathbb{E}\left|Z_n^{(1)}(\beta_1)-Z_n^{(1)}(\beta_2)\right|^r\le c\|\beta_1-\beta_2\|^r. \tag{5.34}
$$

It remained to check Assumptions (5.26) and (5.27) for $T_n(\beta) = Z_n^{(2)}(\beta) - Z_n^{(2)}(\beta_0)$. Using Assumption (A6*), by the same method as in (5.31)–(5.33), one can get

$$
\mathbb{E}\left|Z_n^{(2)}(\beta_1)-Z_n^{(2)}(\beta_2)\right|^r\le c\|\beta_1-\beta_2\|^r. \tag{5.35}
$$

Therefore

$$
\mathbb{E}\left|\frac{T_n(\beta)}{\varrho_m}\right|^r\le\frac{c\|\beta-\beta_0\|^r}{\varrho_m^r}<c, \tag{5.36}
$$

if $\|\beta-\beta_0\| \le F_2\varrho_{m+1}$, and

$$
\mathbb{E}\left|T_n(\beta_1)-T_n(\beta_2)\right|^r=\mathbb{E}\left|Z_n^{(2)}(\beta_1)-Z_n^{(2)}(\beta_2)\right|^r\le c\|\beta_1-\beta_2\|^r. \tag{5.37}
$$

So, the assumptions of Lemma 5.1 are satisfied. The proof of Theorem 3.6 is complete. ∎

PROOF OF REMARK 3.7. In the proof of inequality (5.29) instead of (4.1) use (4.2). ∎

PROOF OF LEMMA 5.1. Let $\|\ \|_\infty$ denote the sup norm in $\mathbb{R}^p$. Then, there exist finite positive constants L_1 and L_2 such that

$$
L_1\|x\| \le \|x\|_\infty \le L_2\|x\|, \quad \forall x \in \mathbb{R}^p.
$$

Therefore

$$\left\{\beta : \frac{\varrho}{\sqrt{n}} \leq \|\beta - \beta_0\| \leq \varrho\right\} \subseteq \left\{\beta : \frac{L_1\varrho}{\sqrt{n}} \leq \|\beta - \beta_0\|_\infty \leq L_2\varrho\right\} = G.$$

Let $\tilde{\delta} = \varrho(L_1/\sqrt{n})$, $\tilde{\varrho}_m = m\tilde{\delta}$, and $G_m = \{\beta : \tilde{\varrho}_m \leq \|\beta - \beta_0\|_\infty \leq \tilde{\varrho}_{m+1}\}$.

Then

$$G \subset \bigcup_{m=1}^{[L_2\sqrt{n}/L_1]} G_m,$$

where [ ] denotes the integer part. We cover G_m by balls (in sense of $\| \ \|_\infty$-norm) with radius $\tilde{\delta}$. It is possible that the number of balls covering G_m is not greater than cm^{p-1}. Let us denote a covering of this kind by $\mathcal{P}_m$. Then

$$P\left(\sup_{(\varrho/\sqrt{n})\leq\|\beta-\beta_0\|\leq\varrho} \frac{|T_n(\beta)|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \geq c\right) \leq \sum_{m=1}^{[L_2\sqrt{n}/L_1]} P_{mn}, \tag{5.38}$$

where

$$\begin{aligned}
P_{mn} &= P\left(\sup_{\tilde{\varrho}_m\leq\|\beta-\beta_0\|_\infty\leq\tilde{\varrho}_{m+1}} \frac{|T_n(\beta)|}{\sqrt{n}\Psi_n(\beta,\beta_0)} \geq c\right) \\
&\leq P\left(\sup_{\tilde{\varrho}_m\leq\|\beta-\beta_0\|_\infty\leq\tilde{\varrho}_{m+1}} |T_n(\beta)| \geq c\varrho_m^2\sqrt{n}\right) \\
&\leq \sum_{B\in\mathcal{P}_m} P\left(\sup_{\beta\in B\cap\Theta} |T_n(\beta)| \geq c\varrho_m^2\sqrt{n}\right) \\
&= \sum_{B\in\mathcal{P}_m} P\left(|T_n(\beta_B)| \geq \frac{1}{2}c\varrho_m^2\sqrt{n}\right) \\
&\quad + \sum_{B\in\mathcal{P}_m} P\left(\sup_{\beta_1,\ \beta_2\in B\cap\Theta} |T_n(\beta_1) - T_n(\beta_2)| \geq \frac{1}{2}c\varrho_m^2\sqrt{n}\right) \\
&\leq cm^{p-1} \sup_{B\in\mathcal{P}_m} P\left(|T_n(\beta_B)| \geq \frac{1}{2}c\varrho_m^2\sqrt{n}\right) \\
&\quad + cm^{p-1} \sup_{B\in\mathcal{P}_m} P\left(\sup_{\beta_1,\ \beta_2\in B\cap\Theta} |T_n(\beta_1) - T_n(\beta_2)| \geq \frac{1}{2}c\varrho_m^2\sqrt{n}\right) \\
&= P_{mn}^{(1)} + P_{mn}^{(2)},
\end{aligned} \tag{5.39}$$

say, where β_B denotes a point in the ball B.

By Chebyshev's inequality and assumption (5.26)

$$P\left(|T_n(\beta_B)| \geq c\varrho_m^2\sqrt{n}\right) \leq c\frac{\mathbb{E}\,|T_n(\beta_B)|^r}{\varrho_m^{2r}n^{r/2}} \leq \frac{c}{\varrho_m^r n^{r/2}} = \frac{c}{m^r\varrho^r}, \tag{5.40}$$

if $\beta_B \in G_m$. Therefore

$$P_{mn}^{(1)} \leq \frac{cm^{p-1}}{m^r\varrho^r} = \frac{c}{m^{r-p+1}\varrho^r},$$

and

$$\sum_{m=1}^{[L_1\sqrt{n}/L_2]} P_{mn}^{(1)} \leq \frac{c}{\varrho^r}\sum_{m=1}^{\infty} \frac{1}{m^{r-p+1}} \leq \frac{c}{\varrho^r}, \tag{5.41}$$

because $r > p$.

Now, we turn to $P^{(2)}_{mn}$. We need Theorem 3.2.3 of [11]. For convenience we shall quote it in Lemma 5.2. By (5.45)

$$P\left(\sup_{\beta_1,\ \beta_2\in B\cap\Theta}|T_n(\beta_1)-T_n(\beta_2)|\ge a\right)\le k_0 c\left((m+1)\frac{\varrho}{\sqrt{n}}\right)^p\left(\frac{2\varrho}{\sqrt{n}}\right)^{r-p}a^{-r} = c\left(\frac{\varrho}{\sqrt{n}}\right)^r(m+1)^p a^{-r}, \tag{5.42}$$

if $\beta_1,\ \beta_2\in G_m$. Therefore

$$P^{(2)}_{mn}\le cm^{p-1}\sup_{B\in\mathcal{P}_m}\left(\frac{\varrho}{\sqrt{n}}\right)^r(m+1)^p\left(\left(\frac{m\varrho}{\sqrt{n}}\right)^2\sqrt{n}\right)^{-r}\le c\frac{1}{\varrho^r}\frac{1}{m^{2r-2p+1}},$$

which implies

$$\sum_{m=1}^{[L_2\sqrt{n}/L_1]}P^{(2)}_{mn}\le\frac{c}{\varrho^r}\sum_m\frac{1}{m^{2r-2p+1}}\le\frac{c}{\varrho^r}, \tag{5.43}$$

because $p<r$. This completes the proof. ∎

LEMMA 5.2. *(See Theorem 3.2.3 of [11].) Let T_β be a separable, measurable stochastic field defined on the closed set $\Theta\subseteq\mathbb{R}^p$. Suppose that for any β, $\tilde\beta$ (for which $\beta+\tilde\beta\in\Theta$)*

$$\mathbb{E}\left|T\left(\beta+\tilde\beta\right)-T(\beta)\right|^r\le l(\beta)\left\|\tilde\beta\right\|^{r'}, \tag{5.44}$$

for some $r\ge r'>p$ and a locally bounded function $l:\mathbb{R}^p\to R^1_+$. Then for any q, h, and $a>0$

$$P\left(\sup_{\substack{\beta',\beta''\in\Theta\cap B(q)\\ \|\beta'-\beta''\|\le h}}|T(\beta')-T(\beta'')|>a\right)\le k_0\left(\sup_{\beta\in\Theta\cap B(q)}l(\beta)\right)q^p h^{r'-p}a^{-r}, \tag{5.45}$$

where k_0 depends on r, r', and p but does not depend on q, h, and a. (Here $B(q)=\{\beta\ :\ \|\beta\|\le q\}$ is the closed ball with radius q.)

REFERENCES

1. W.A. Fuller, *Measurement Error Models*, Wiley, New York, (1987).
2. L.J. Gleser, R.J. Carroll and P.P. Gallo, The limiting distribution of least squares in an errors-in-variables regression model, *Ann. Statist.* **15**, 220–233 (1987).
3. L.J. Gleser, Improvement of the naive approach to estimation in nonlinear errors-in-variables regression models, *Contemporary Mathematics* **120**, 99–114 (1990).
4. L.-F. Lee and J.H. Sepanski, Estimation of linear and nonlinear errors-in-variables models using validation data, *J. Amer. Statist. Assoc.* **90**, 130–140 (1995).
5. R.J. Carroll, H. Küchenhoff, F. Lombard and L.A. Stefanski., Asymptotics for the SIMEX estimator in nonlinear measurement error models, *J. Amer. Statist. Assoc.* **91**, 242–250 (1996).
6. A.G. Kukush and S. Zwanzig, On inconsistency of the least squares estimator in nonlinear functional error-in-variables models., Preprint No. 12/1996, Universität Hamburg (1996).
7. I. Fazekas and A.G. Kukush, On inconsistency of the least squares estimator in nonlinear functional errors-in-variables models with dependent error terms, Technical Report No. 28/1996, Kossuth University, Debrecen, (1996).
8. A.G. Kukush and S. Zwanzig, On an alternative estimator in nonlinear errors-in-variables models, Unpublished manuscipt, Universität Hamburg, (1996).
9. L.A. Stefanski, Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models, *Comm. Statist. Theory Meth.* **18** (12), 4335–4358 (1989).
10. I.A. Ibragimov and R.Z. Hasmiskii, *Asymptotic Theory of Estimation*, (in Russian), Nauka, Moscow, (1979).

11. N.N. Leonenko and A.V. Ivanov, *Statistical Analysis of Random Fields*, (in Russian), Visha Shkola, Kiev, (1986).
12. A.Ya. Dorogovtsev, On the existence and convergence of moments of the least squares estimator in a nonlinear model of autoregression (in Ukrainian), *Proc. Acad. Sci. Ukraine, Ser. A* **12**, 10–13 (1973).
13. A.V. Ivanov, An asymptotic expansion for the distribution of the least squares estimator of the nonlinear regression parameter, *Theory Probab. Appl.* **21**, 557–570 (1976).
14. B.L.S. Prakasa Rao, The rate of convergence of the least squares estimator in a nonlinear regression model with dependent errors, *J. Multivariate Anal.* **14**, 315–322 (1984).
15. J. Pfanzagl, On the measurability and consistency of minimum contrast estimates, *Metrika* **14**, 249–272 (1969).
16. S.A. Utev, Inequalities for sums of weakly dependent random variables and rate of convergence in invariance principle, (in Russian), In *Limit Theorems for Sums of Random Variables*, Nauka, Novosibirsk, 50–70 (1984).
17. I. Fazekas and A.G. Kukush, Asymptotic properties of an estimator in non-linear functional errors-in-variables models, Technical Report No. 25/1996, Kossuth University, Debrecen, (1996).
18. I. Fazekas and A.G. Kukush, Asymptotic properties of estimators in nonlinear functional errors-in-variables models with dependent error terms, In *Symposium on the Expansion Method*, (Edited by G. Kristensen), pp. 188–195, Odense University, (1996).