A consistent estimator in general functional errors-in-variables models*

S. Baran
Institute of Mathematics and Informatics, Lajos Kossuth University
P. O. Box 12, H-4010 Debrecen, Hungary
barans@math.klte.hu

Abstract

The problem of estimation in nonlinear functional errors-in-variables model is considered. A modified least squares estimator is studied, its consistency and asymptotic normality is established both for temporal and for spatial observations. Simulation results are also presented showing the good performance of the estimator in small sample situations.
Mathematics subject classification. Primary: 62F12; secondary: 62J02.
Key words and phrases. Errors-in-variables, functional model, estimator, consistency, asymptotic normality, α-mixing error term, temporal model, spatial model, increasing domain.

1. Introduction

In statistical work one often encounters such non-standard situations when the explanatory variables of the regression models (e.g. location, temperature, pressure) can only be observed with errors. Throughout this paper we will deal with a model of such kind. It has the following form:

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

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

Models of the above form are called errors-in-variables models and have been studied for about half a century. A summary of results is given in Fuller's textbook ([9]).

There are two main classes of the errors-in-variables models: structural one, when $\xi_1, \xi_2, \ldots$ are random and functional one, when they are non-random. The latter is the case in this paper.

*Supported by the Hungarian Ministry of Education under Grant No. FKFP 0121/1999.

The problem of estimation in linear functional relationship has quite a wide literature but relatively less is said about the case when g is nonlinear. It has been known for a long time that the least squares estimator of β_0 that minimizes the function

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

is not consistent (see e.g. [8, 12]). In [3] general ideas are given to construct a consistent estimator of the parameter in the polynomial model in the case when the error terms have jointly normal distribution. Cheng and Schneeweiss [2] generalizes this idea for arbitrary i.i.d. pairs (ε_i,δ_i), $i=1,2,\ldots$ and gets an estimator that is easy to calculate.

In this paper an alternative estimator is considered that minimizes the function $Q_n(\beta)$ determined by (2.3). A slightly simpler version of this estimator has been studied by Kukush and Zwanzig [13] in the case when ε_i,δ_i, $i=1,2,\ldots,$ are i.i.d. one-dimensional random variables. Their result was extended in [6] for vector valued x_i and dependent error terms $\{\delta_i\}$ and in [5] for vector valued x_i, y_i and mixing error terms $\{\delta_i\}$, $\{\varepsilon_i\}$. Both in [6] and [13] the independence of δ_i and ε_i for each i was a key assumption. In the present paper the situation of vector valued x_i and y_i is examined when $\{\delta_i\}$ and $\{\varepsilon_i\}$ are mixing and not necessarily independent. Example 2.1 shows that the estimator proposed in [2] is a special case of the estimator considered here.

The following notation is used. $\mathbb{N}$ is the set of positive integers, $\mathbb{Z}$ is the set of all integers, $\mathbb{R}$ is the real line, $\mathbb{R}^p$ is the p-dimensional Euclidean space with norm $\|\ \|$. Vectors are always considered to be column vectors. Sign $^\top$ denotes the transpose of a matrix. I stands for the identity matrix and given a matrix D, $\lambda_{\min}(D)$ denotes its smallest eigenvalue. $C(A\to\mathbb{R}^r)$ is the set of $\mathbb{R}^r$-valued continuous functions defined on A. If $y=(y_1,\ldots,y_r)^\top\in\mathbb{R}^r$, $\ell=(\ell_1,\ldots,\ell_r)^\top:\mathbb{R}^p\to\mathbb{R}^r$ and $\beta=(\beta_1,\ldots,\beta_p)^\top$, $\widetilde{\beta}=(\widetilde{\beta}_1,\ldots,\widetilde{\beta}_p)^\top\in\mathbb{R}^p$, then we shall use the following notation:

$$y^\top\frac{\partial\ell(\beta)}{\partial\beta}=\frac{\partial y^\top\ell(\beta)}{\partial\beta}\quad\text{and}\quad y^\top\frac{\partial^2\ell(\beta)}{\partial\beta^2}=\frac{\partial^2 y^\top\ell(\beta)}{\partial\beta^2}.$$

More precisely,

$$z=\left(y^\top\frac{\partial^2\ell(\beta)}{\partial\beta^2}\right)\widetilde{\beta}$$

means that the ith coordinate of z is:

$$z_i=\sum_{k=1}^{r}y_k\left(\sum_{j=1}^{p}\frac{\partial^2\ell_k(\beta)}{\partial\beta_i\partial\beta_j}\cdot\widetilde{\beta}_j\right),\qquad i=1,2,\ldots,p.$$

We shall denote different constants with the same letter c. The cardinality of the set A is denoted by $|A|$. We shall suppose the existence of an underlying probability space $(\Omega,\mathcal{F},\mathbb{P})$; $\omega\in\Omega$ denotes an elementary event. $\mathbb{E}$ stands for the expectation.

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

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

2. The model and the estimator

Let us consider the following model

$$y_i = g(\xi_i, \beta_0) + \delta_i, \tag{2.1}$$
$$x_i = \xi_i + \varepsilon_i, \tag{2.2}$$

$i = 1, 2, ...$, where design points $\xi_1, \xi_2, ...$, are non-stochastic but not observed, y_i, x_i are observed, ε_i, δ_i are (non-observed, not necessarily independent) random error terms with zero expectation ($i = 1, 2, ...$). 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 r-dimensional vectors ($i = 1, 2, ...$), $g : V \times \Theta \to \mathbb{R}^r$ is a known function where $V \subseteq \mathbb{R}^q$, $\xi_i \in V$ for each i.

We shall suppose the existence of three sequences of auxiliary functions:
(Af) There exist an open set $U \supset \Theta$ and functions $f_i \in \mathrm{C}(\mathbb{R}^q \times U \to \mathbb{R}^r)$ such that

$$\mathbb{E} f_i(\xi + \varepsilon_i, \beta) = g(\xi, \beta)$$

for each $\xi \in V$, $\beta \in \Theta$, $i \in \mathbb{N}$.
(Ah) There exist functions $h_i \in \mathrm{C}(\mathbb{R}^q \times U \to \mathbb{R})$ such that

$$\mathbb{E} h_i(\xi + \varepsilon_i, \beta) = \|g(\xi, \beta)\|^2$$

for each $\xi \in V$, $\beta \in \Theta$, $i \in \mathbb{N}$.
(Ad) There exist functions $d_i \in \mathrm{C}(\mathbb{R}^q \times U \to \mathbb{R})$ such that

$$\mathbb{E} d_i(\xi + \varepsilon_i, \beta) = \mathbb{E} \delta_i^\top f(\xi + \varepsilon_i, \beta)$$

for each $\xi \in V$, $\beta \in \Theta$, $i \in \mathbb{N}$.

We consider the following modification of the least squares method which is a natural generalization of the estimation method described in [5]. Let the estimator $\widehat{\beta} = \widehat{\beta}_n$ of β_0 be the minimum point of

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

By [14], there exists a measurable solution of the above minimum problem if the parameter set Θ is compact. Compared to [5] and [6] $2d_i(x_i, \beta)$ is an extra term that enables us to handle the case where δ_i and ε_i are not independent.

Our first example shows that the polynomial regression method developed in [2] is a special case of the above estimator.

Example 2.1 Consider the polynomial model $g(x, \beta) = \beta_0 + \beta_1 x + \cdots + \beta_k x^k$, where $x \in \mathbb{R}$. The parameter vector here is $\beta = (\beta_0, \beta_1, \ldots, \beta_k) \in \mathbb{R}^{k+1}$. Suppose that we know every moment $\mathbb{E}\varepsilon_i^\ell$, $\ell = 1, 2, \ldots, 2k$ and $\mathbb{E}(\varepsilon_i^\ell \delta_i)$, $\ell = 1, 2, \ldots, k$. Using the method of [2] we can find uniquely defined polynomials $t_r^{(i)}(x)$ of order r such that

$\mathbb{E}t_r^{(i)}(x+\varepsilon_i) = x^r, \;\; r = 0, 1, \ldots, 2k$. This can be obtained by solving the following recursive system of equations:

$$x^r = \sum_{j=0}^{r} \binom{r}{j} t_j^{(i)}(x) \mathbb{E}\varepsilon_i^{r-j},$$

$r = 0, 1, \ldots, 2k$. Thus,

$$t_r^{(i)}(x) = \sum_{j=0}^{r} a_{rj}^{(i)} x^j,$$

where the $a_{rj}^{(i)}$ are functions of $\mathbb{E}\varepsilon_i^s, \;\; s = 0, 1, \ldots, r$. Now let

$$\begin{aligned} f_i(x,\beta) &= \sum_{j=0}^{k} \beta_j t_j^{(i)}(x), \\ h_i(x,\beta) &= \sum_{j=0}^{k} \left(\sum_{\ell=0}^{j} \beta_\ell \beta_{j-\ell} \right) t_j^{(i)}(x) + \sum_{j=k+1}^{2k} \left(\sum_{\ell=j-k}^{k} \beta_\ell \beta_{j-\ell} \right) t_j^{(i)}(x). \end{aligned}$$

In order to obtain the third sequence of auxiliary functions define

$$d_r^{(i)}(x) = \sum_{j=0}^{r} \left(\sum_{s=j}^{r} a_{rs}^{(i)} \binom{s}{j} \mathbb{E}(\varepsilon_i^{s-j}\delta_i) \right) t_j^{(i)}(x),$$

$r = 0, 1, \ldots, k$, and let

$$d_i(x,\beta) = \sum_{j=0}^{k} \beta_j d_j^{(i)}(x).$$

It is easy to verify that the three sequences of functions defined above satisfy conditions (Af), (Ah) and (Ad), respectively. Actually, these facts are direct consequences of the results in [2] despite that the general model (1.1)–(1.2) and estimator defined by (2.3) are not considered there. The system of equations

$$\frac{\partial Q_n(\beta)}{\partial \beta_j} = 0, \quad\;\; j = 0, 1, \ldots, k,$$

that provides a necessary condition for a β to be the minimum of Q_n is exactly the same system of linear equations that gives the estimator of β_0 in [2]. It is the following:

$$\sum_{s=0}^{k} \beta_s \Big(\frac{1}{n} \sum_{i=1}^{n} t_{s+j}^{(i)}(x) \Big) = \frac{1}{n} \sum_{i=1}^{n} y_i t_j^{(i)}(x) - \frac{1}{n} \sum_{i=1}^{n} d_j^{(i)}(x),$$

$j = 0, 1, \ldots, k$.

Example 2.2 Consider the trigonometric polynomial model

$$g(x,\beta) = \frac{\alpha_0}{2} + \sum_{\ell=1}^{k} \bigl(\alpha_\ell \cos(\ell x) + \beta_\ell \sin(\ell x)\bigr),$$

where $x \in \mathbb{R}$, with parameter vector $\beta = (\alpha_0, \alpha_1, \beta_1, \ldots, \alpha_k, \beta_k)^\top \in \mathbb{R}^{2k+1}$. Suppose we know all values $\mathbb{E}\cos(\ell\varepsilon_i)$, $\mathbb{E}\sin(\ell\varepsilon_i)$, $\ell = 1, 2, \ldots, 2k$ and $\mathbb{E}\delta_i \cos(\ell\varepsilon_i)$, $\mathbb{E}\delta_i \sin(\ell\varepsilon_i)$, $\ell = 1, 2, \ldots, k$.

We are seeking for functions f_i in the following form:

$$f_i(x,\beta) = \frac{a_0^{(i)}}{2} + \sum_{\ell=1}^{k} \bigl(a_\ell^{(i)} \cos(\ell x) + \beta_\ell^{(i)} \sin(\ell x)\bigr).$$

Comparing the coefficients of $\mathbb{E} f_i(\xi_i + \varepsilon_i, \beta)$ with the coefficients of $g(\xi_i, \beta)$ we get the formulae: $a_0^{(i)} = \alpha_0$ and

$$\begin{pmatrix} a_\ell^{(i)} \\ b_\ell^{(i)} \end{pmatrix} = \frac{1}{(\mathbb{E}\cos(\ell\varepsilon_i))^2 + (\mathbb{E}\sin(\ell\varepsilon_i))^2} \begin{pmatrix} \mathbb{E}\cos(\ell\varepsilon_i) & -\mathbb{E}\sin(\ell\varepsilon_i) \\ \mathbb{E}\sin(\ell\varepsilon_i) & \mathbb{E}\cos(\ell\varepsilon_i) \end{pmatrix} \begin{pmatrix} \alpha_\ell \\ \beta_\ell \end{pmatrix},$$

$\ell = 1, 2, \ldots, k$.

As $g^2(x,\beta)$ is a trigonometric polynomial of order $2k$ with coefficients that are functions of β, $h_i(x,\beta)$ can be determined similarly to $f_i(x,\beta)$.

Short straightforward calculation shows that

$$d_i(x,\beta) = \sum_{\ell=1}^{k} \bigl(\mu_\ell^{(i)} \cos(\ell x) + \nu_\ell^{(i)} \sin(\ell x)\bigr),$$

where

$$\begin{aligned} \begin{pmatrix} \mu_\ell^{(i)} \\ \nu_\ell^{(i)} \end{pmatrix} &= \frac{1}{(\mathbb{E}\cos(\ell\varepsilon_i))^2 + (\mathbb{E}\sin(\ell\varepsilon_i))^2} \begin{pmatrix} \mathbb{E}\cos(\ell\varepsilon_i) & -\mathbb{E}\sin(\ell\varepsilon_i) \\ \mathbb{E}\sin(\ell\varepsilon_i) & \mathbb{E}\cos(\ell\varepsilon_i) \end{pmatrix} \\ &\quad \times \begin{pmatrix} \mathbb{E}\delta_i \cos(\ell\varepsilon_i) & \mathbb{E}\delta_i \sin(\ell\varepsilon_i) \\ -\mathbb{E}\delta_i \sin(\ell\varepsilon_i) & \mathbb{E}\delta_i \cos(\ell\varepsilon_i) \end{pmatrix} \begin{pmatrix} a_\ell^{(i)} \\ b_\ell^{(i)} \end{pmatrix}, \end{aligned}$$

$\ell = 1, 2, \ldots, k$.

3. Consistency and asymptotic normality

To prove the consistency of the estimator $\widehat{\beta}_n$ we need some conditions that are listed below:

(A1) The parameter set Θ is compact.

(A2) g is bounded on $V \times \Theta$, $\xi_i \in V$ for each i.

(A3) For each sequence $\{\xi_n\}$ and for each $\beta^* \in \Theta$ and $b > 0$

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

(A4) For each $b > 0$ there exists $\ell > 0$ such that for each $s \in V$

$$\|g(s,\beta_1) - g(s,\beta_2)\| < b \quad \text{if} \quad \|\beta_1 - \beta_2\| < \ell.$$

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

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

(A7)
$$\lim_{\ell\to 0} \limsup_{n\to\infty} \mathbb{E} \sup_{\|\beta_1-\beta_2\|\le \ell} |d_n(x_n,\beta_1) - d_n(x_n,\beta_2)| = 0.$$

(A8)
$$\lim_{\ell\to 0} \limsup_{n\to\infty} \mathbb{E} \sup_{\|\beta_1-\beta_2\|\le \ell} |\delta_n^\top f_n(x_n,\beta_1) - \delta_n^\top f_n(x_n,\beta_2)| = 0.$$

(A9)
$$\sup_n \mathbb{E}\, \|f_n(x_n,\beta)\|^{t+a} < \infty$$

for each $\beta \in \Theta$.

(A10)
$$\sup_n \mathbb{E}\, |h_n(x_n,\beta)|^{t+a} < \infty$$

for each $\beta \in \Theta$.

(A11)
$$\sup_n \mathbb{E}\, |d_n(x_n,\beta)|^{t+a} < \infty$$

for each $\beta \in \Theta$.

(A12)
$$\sup_n \mathbb{E}\, |\delta_n^\top f_n(x_n,\beta)|^{t+a} < \infty$$

for each $\beta \in \Theta$.

We remark that instead of condition (A12) we can demand $\sup_n \mathbb{E}\, \|\delta_n\|^{2t+a} < \infty$ and $\sup_n \mathbb{E}\, \|f_n(x_n,\beta)\|^{2t+a} < \infty$.

We shall use mixing conditions for the dependence structure of $\{\delta_i, \varepsilon_i, i = 1, 2, \dots\}$. If $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras, then

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

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

$$\alpha(n) = \sup_{1\le k<\infty} \alpha\left(\mathcal{M}_1^k, \mathcal{M}_{k+n}^\infty\right),$$

$$j(t) = 2\min\{k \in \mathbb{N} \ : \ 2k \geq t\}.$$

The following condition will be appropriate for our purpose:

$$\text{(A}\alpha\text{)} \qquad b(\alpha, t, a) = \sum_{k=1}^{\infty} \alpha^{a/(j(t)+a)}(k)(k+1)^{j(t)-2} < \infty,$$

where $a > 0$.

In our proofs we will often use the following two lemmas:

Lemma 3.1 *Let η_i be measurable with respect to the σ-algebra generated by the pair δ_i, ε_i and assume that $\mathbb{E}\eta_i = 0$, $i = 1, 2, \ldots$.*

Let $a > 0$, then

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

if $1 \leq t \leq 2$;

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

if $2 \leq t$. In particular

$$\mathbb{E} \left\| \sum_{i=1}^{n} a_i \eta_i \right\|^t \leq c(t) b(\alpha, t, a) \left(\sum_{i=1}^{n} a_i^2 \right)^{t/2} \max_{1 \leq i \leq n} (\|\eta_i\|_{t+a})^t,$$

if $2 \leq t$. (Here $c(t)$ depends on t (and on the dimension) but it does not depend on n.)

The proof can be found in [4, Theorem 2, p. 26].

Lemma 3.2 *Let $U_n(\beta)$ be random vectors, $n = 1, 2, \ldots$, $\beta \in \Theta$, where Θ is a compact set. Let $\lim_{n\to\infty} U_n(\beta) = 0$ in probability for each $\beta \in \Theta$. Suppose that for each $\varepsilon > 0$*

$$\lim_{\ell \to 0} \limsup_{n\to\infty} \mathbb{P}^* \left\{ \sup_{\|\beta_1 - \beta_2\| \leq \ell} \|U_n(\beta_1) - U_n(\beta_2)\| > \varepsilon \right\} = 0, \tag{3.1}$$

where $\mathbb{P}^*(A)$ means the outer probability of the set $A \subseteq \Omega$. Then

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

in probability, that is for any $\varepsilon > 0$

$$\mathbb{P}^* \left\{ \sup_{\beta \in \Theta} \|U_n(\beta)\| > \varepsilon \right\} \longrightarrow 0,$$

as $n \to \infty$.

The proof can be found e.g. in [6].

Now, let us formulate the theorem that concerns the consistency of $\widehat{\beta}_n$ in the α-mixing case.

Theorem 3.3 *Suppose that for the model (2.1)–(2.2) conditions (Af), (Ah), (Ad) and (A1)–(A8) are satisfied. Moreover, assume that for some $t > 1$ and $a > 0$ conditions (Aα), and (A9)–(A12) are satisfied. Then $\lim_{n\to\infty} \widehat{\beta}_n = \beta_0$ in probability, where $\widehat{\beta}_n$ is the estimator defined by (2.3).*

PROOF. Considering that $y_i = g(\xi_i, \beta_0) + \delta_i$ let us take the following form of $Q_n(\beta)$:

$$\begin{aligned} Q_n(\beta) &= \frac{1}{n}\sum_{i=1}^{n} \{\|g(\xi_i,\beta) - g(\xi_i,\beta_0)\|^2 + 2\delta_i^\top g(\xi_i,\beta_0) + \|\delta_i\|^2 \\ &\quad -2\big(f_i(x_i,\beta) - g(\xi_i,\beta)\big)^\top g(\xi_i,\beta_0) - (\|g(\xi_i,\beta)\|^2 \\ &\quad + h_i(x_i,\beta)) - 2\delta_i^\top f_i(x_i,\beta) + 2d_i(x_i,\beta)\} . \end{aligned}$$

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

$$\begin{aligned} \widetilde{Q}_n(\beta) &= \frac{1}{n}\sum_{i=1}^{n} \{\|g(\xi_i,\beta) - g(\xi_i,\beta_0)\|^2 \\ &\quad -2\big(f_i(x_i,\beta) - g(\xi_i,\beta)\big)^\top g(\xi_i,\beta_0) \\ &\quad - \|g(\xi_i,\beta)\|^2 + h_i(x_i,\beta) - 2\delta_i^\top f_i(x_i,\beta) + 2d_i(x_i,\beta)\} . \end{aligned}$$

Later we shall prove that

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \Big\{2\big(f_i(x_i,\beta) - g(\xi_i,\beta)\big)^\top g(\xi_i,\beta_0) + 2\big(\delta_i^\top f_i(x_i,\beta) - d_i(x_i,\beta)\big) \\ &\qquad + \big(\|g(\xi_i,\beta)\|^2 - h_i(x_i,\beta)\big)\Big\} = \mathrm{o}_P(1), \end{aligned} \tag{3.2}$$

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

Now, we prove that (3.2) and condition (A3) imply that the minimum point $\widehat{\beta}_n$ of $\widetilde{Q}_n(\beta)$ is consistent. From (3.2) and $\widetilde{Q}_n(\widehat{\beta}_n) \le \widetilde{Q}_n(\beta_0)$ follows that

$$\frac{1}{n}\sum_{i=1}^{n} \Big\|g(\xi_i,\widehat{\beta}_n) - g(\xi_i,\beta_0)\Big\|^2 = \mathrm{o}_P(1), \tag{3.3}$$

as $n \to \infty$.

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

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

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

$$A_{n_0}(b) = \Big\{\omega :\ \|\widehat{\beta}_{n_0}(\omega) - \beta_0\| \ge b\Big\} .$$

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

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

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

$$\mathbb{P}\left(\|\widehat{\beta}_{n_0} - \beta_0\| \geq b\right) \leq \mathbb{P}\left(\frac{1}{n_0}\sum_{i=1}^{n_0}\left\|g(\xi_i,\widehat{\beta}_{n_0}) - g(\xi_i,\beta_0)\right\|^2 > \bar{\varepsilon}\right) \tag{3.4}$$

if $n_0 > n_{\bar{\varepsilon}}$. Now, (3.3) and (3.4) imply that $\widehat{\beta}_n \to \beta_0$ in probability, as $n \to \infty$.

To prove (3.2) we shall use Lemma 3.2 for the components of the sum in (3.2). We restrict our attention to the case $1 < t \leq 2$.

First let

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(f_i(x_i,\beta) - g(\xi_i,\beta)\right)^\top g(\xi_i,\beta_0). \tag{3.5}$$

The summands are α-mixing with zero expectation.

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

as $n \to \infty$, because of assumptions (A2) and (A9). Therefore, $U_n(\beta) \to 0$ in probability.

Now, we prove condition (3.1) for $U_n(\beta)$ defined by (3.5).

$$\begin{aligned}
&\sup_{\|\beta_1-\beta_2\|\leq\ell}|U_n(\beta_1) - U_n(\beta_2)| \qquad (3.6)\\
&\leq \sup_{\|\beta_1-\beta_2\|\leq\ell}\frac{1}{n}\sum_{i=1}^{n}\|g(\xi_i,\beta_0)\|\cdot\|f_i(x_i,\beta_1) - f_i(x_i,\beta_2)\| \\
&+ \sup_{\|\beta_1-\beta_2\|\leq\ell}\frac{1}{n}\sum_{i=1}^{n}\|g(\xi_i,\beta_0)\|\cdot\|g(\xi_i,\beta_1) - g(\xi_i,\beta_2)\|.
\end{aligned}$$

The second summand in the right hand side of (3.6) tends to zero as $\ell \to 0$, because of assumption (A4). The expectation of the first summand in the right hand side of (3.6) is majorized by

$$c\sup_i \mathbb{E}\sup_{\|\beta_1-\beta_2\|\leq\ell}\|f_i(x_i,\beta_1) - f_i(x_i,\beta_2)\|.$$

Therefore, assumption (A5) and Chebyshev's inequality imply that assumption (3.1) is satisfied for $U_n(\beta)$ defined by (3.5).

Now, we have to prove that

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{h_i(x_i,\beta) - \|g(\xi_i,\beta)\|^2\right\} \tag{3.7}$$

satisfies the conditions of Lemma 3.2. As the summands are mixing with zero expectation, we have

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

and, by (A10), the last expression tends to 0, as $n\to\infty$. So $U_n(\beta)\to 0$ in probability.

To check condition (3.1) consider

$$\sup_{\|\beta_1-\beta_2\|\le\ell}|U_n(\beta_1) - U_n(\beta_2)| \le \sup_{\|\beta_1-\beta_2\|\le\ell}\frac{1}{n}\sum_{i=1}^{n}|h_i(x_i,\beta_1) - h_i(x_i,\beta_2)|$$
$$+\sup_{\|\beta_1-\beta_2\|\le\ell}\frac{1}{n}\sum_{i=1}^{n}\left|\|g(\xi_i,\beta_1)\|^2 - \|g(\xi_i,\beta_2)\|^2\right| = B_{n,\ell} + C_{n,\ell},$$

Here $C_{n,\ell}$ tends to 0, as $\ell\to 0$, because of conditions (A2) and (A4). Furthermore,

$$\lim_{\ell\to 0}\limsup_{n\to\infty}\mathbb{E}|B_{n,\ell}| = 0$$

by (A6). Therefore, Chebyshev's inequality implies that condition (3.1) is satisfied for $U_n(\beta)$ defined by (3.7).

At the end, we show that

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i^\top f_i(x_i,\beta) - d_i(x_i,\beta)\right)\to 0 \tag{3.8}$$

in probability, as $n\to\infty$, uniformly in β. By the choice of d_i the summands have zero mean and satisfy α-mixing conditions, so

$$\begin{aligned}\mathbb{E}|U_n(\beta)|^t &\le c\frac{1}{n^t}\sum_{i=1}^{n}\left(\mathbb{E}|\delta_i^\top f_i(x_i,\beta) - d_i(x_i,\beta)|^{t+a}\right)^{\frac{t}{t+a}}\\ &\le c\frac{1}{n^t}\sum_{i=1}^{n}\left(c\mathbb{E}|\delta_i^\top f_i(x_i,\beta)|^{t+a} + c\mathbb{E}|d_i(x_i,\beta)|^{t+a}\right)^{\frac{t}{t+a}}.\end{aligned}$$

Due to conditions (A11) and (A12) the above expression converges to zero when $n\to\infty$ for each β, hence $U_n(\beta)\to 0$ in probability.

$$\sup_{\|\beta_1-\beta_2\|\le\ell}|U_n(\beta_1) - U_n(\beta_2)| \le \sup_{\|\beta_1-\beta_2\|\le\ell}\frac{1}{n}\sum_{i=1}^{n}|\delta_i^\top f_i(x_i,\beta_1) - \delta_i^\top f_i(x_i,\beta_2)|$$
$$+\sup_{\|\beta_1-\beta_2\|\le\ell}\frac{1}{n}\sum_{i=1}^{n}|d_i(x_i,\beta_1) - d_i(x_i,\beta_2)| = V_{n,\ell} + W_{n,\ell}$$

Conditions (A7) and (A8) imply that

$$\lim_{\ell\to 0}\limsup_{n\to\infty}\mathbb{E}|V_{n,\ell}| = 0 \quad \text{and} \quad \lim_{\ell\to 0}\limsup_{n\to\infty}\mathbb{E}|W_{n,\ell}| = 0,$$

so condition (3.1) is satisfied for $U_n(\beta)$ defined by (3.8).

We remark that the case $t > 2$ follows from the case $1 < t \le 2$. □

The following CLT can be found in [11, Corollary 1.]. Let η_i, $i \in \mathbb{N}$, be a centered (vector valued) stochastic process. Let $S_n = \sum_{i=1}^n \eta_i$ and denote the variance (matrix) of S_n by Σ_n. Consider the following assumptions.

There exists $a > 0$ such that

$$\sup_{i\in\mathbb{N}} \|\eta_i\|_{2+a} < \infty. \tag{3.9}$$

$$\lim_{n\to\infty} n^{-1}\Sigma_n = \Sigma, \tag{3.10}$$

where Σ is a positive definite matrix.

Lemma 3.4 *Assume that conditions (Aα) with $t = 2$, and (3.9) are satisfied (with the same a).*

(a) Then $\limsup_n n^{-1}\sum_{i,j=1}^n \|\operatorname{cov}(\eta_i,\eta_j)\| < \infty$.

(b) If, moreover, (3.10) is satisfied, then $(\Sigma_n)^{-1/2}S_n \Rightarrow \mathcal{N}(0,I)$, as $n\to\infty$.

The following lemma is quite a technical one, its proof can be found in [5].

Lemma 3.5 *Let Θ be a compact set, $\eta_n(\beta)$ a sequence of random variables, $\beta\in\Theta$, and let $a(\beta) : \Theta \to \mathbb{R}^d$ be a function, continuous in a point β_0 of Θ. If $\widehat{\beta}_n \to \beta_0$ in probability and $\eta_n(\beta) \to a(\beta)$ in probability uniformly in β ($n\to\infty$) then $\eta_n(\widehat{\beta}_n) \to a(\beta_0)$ in probability, as $n\to\infty$.*

To prove asymptotic normality we need some additional conditions.

(B1) Θ is compact and convex. β_0 is in the interior of Θ.

(B2) $f_i(\xi,\beta)$, $h_i(\xi,\beta)$ and $d_i(\xi,\beta)$ are twice differentiable with respect to β for each i and each ξ.

(B3) The $(2+a)$'th absolute moments of $\frac{\partial f_i(\xi_i+\varepsilon_i,\beta_0)}{\partial\beta}$, $\frac{\partial h_i(\xi_i+\varepsilon_i,\beta_0)}{\partial\beta}$, $\frac{\partial d_i(\xi_i+\varepsilon_i,\beta_0)}{\partial\beta}$ and $\delta_i^\top\frac{\partial f_i(\xi_i+\varepsilon_i,\beta_0)}{\partial\beta}$ are bounded with respect to i.

(B4) $\mathbb{E}\sup_\beta\left\|\frac{\partial f_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta}\right\| < \infty$, $\mathbb{E}\sup_\beta\left\|\frac{\partial h_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta}\right\| < \infty$, $\mathbb{E}\sup_\beta\left\|\frac{\partial d_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta}\right\| < \infty$ and $\mathbb{E}\sup_\beta\left\|\delta_i^\top\frac{\partial f_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta}\right\| < \infty$ for each i.

(B5) $\mathbb{E}\sup_\beta\left\|\frac{\partial^2 f_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta^2}\right\| < \infty$, $\mathbb{E}\sup_\beta\left\|\frac{\partial^2 h_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta^2}\right\| < \infty$, $\mathbb{E}\sup_\beta\left\|\frac{\partial^2 d_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta^2}\right\| < \infty$ and $\mathbb{E}\sup_\beta\left\|\delta_i^\top\frac{\partial^2 f_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta^2}\right\| < \infty$ for each i.

(B6) The $(1+a)$'th absolute moments of $\frac{\partial^2 f_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta^2}$, $\frac{\partial^2 h_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta^2}$, $\frac{\partial^2 d_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta^2}$ and $\delta_i^\top\frac{\partial^2 f_i(\xi_i+\varepsilon_i,\beta)}{\partial\beta^2}$ are bounded with respect to both i and β.

(B7) $\lim_{\ell\to 0}\limsup_n \mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|\le\ell}\left\|\frac{\partial^2 f_n(\xi_n+\varepsilon_n,\beta_1)}{\partial\beta^2} - \frac{\partial^2 f_n(\xi_n+\varepsilon_n,\beta_2)}{\partial\beta^2}\right\|\right) = 0.$

(B8) $\lim_{\ell\to 0}\limsup_n \mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|\le\ell}\left\|\frac{\partial^2 h_n(\xi_n+\varepsilon_n,\beta_1)}{\partial\beta^2}-\frac{\partial^2 h_n(\xi_n+\varepsilon_n,\beta_2)}{\partial\beta^2}\right\|\right)=0.$

(B9) $\lim_{\ell\to 0}\limsup_n \mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|\le\ell}\left\|\frac{\partial^2 d_n(\xi_n+\varepsilon_n,\beta_1)}{\partial\beta^2}-\frac{\partial^2 d_n(\xi_n+\varepsilon_n,\beta_2)}{\partial\beta^2}\right\|\right)=0.$

(B10) $\lim_{\ell\to 0}\limsup_n \mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|\le\ell}\left\|\delta_i^\top\frac{\partial^2 f_n(\xi_n+\varepsilon_n,\beta_1)}{\partial\beta^2}-\delta_i^\top\frac{\partial^2 f_n(\xi_n+\varepsilon_n,\beta_2)}{\partial\beta^2}\right\|\right)=0.$

(B11) The function $m_n(\beta)=\frac{1}{n}\sum_{i=1}^n\frac{\partial g(\xi_i,\beta)}{\partial\beta}^\top\frac{\partial g(\xi_i,\beta)}{\partial\beta}$ converges to $m_\infty(\beta)$ uniformly in β and $m_\infty(\beta)$ is invertible and continuous at β_0.

Theorem 3.6 *Suppose that for the model (2.1)–(2.2) conditions (Af), (Ah), (Ad) and (A2) are satisfied. Moreover, assume that for some $a>0$ conditions (B1)–(B11) and (Aα) with $t=2$ are satisfied. Assume that (3.10) is satisfied with*

$$S_n=\sum_{i=1}^n\left(\frac{\partial h_i(\xi_i+\varepsilon_i,\beta_0)}{\partial\beta}-2\frac{\partial f_i(\xi_i+\varepsilon_i,\beta_0)}{\partial\beta}^\top[g(\xi_i,\beta_0)+\delta_i]+2\frac{\partial d_i(\xi_i+\varepsilon_i,\beta_0)}{\partial\beta}\right).$$

Suppose that $\widehat{\beta}_n$ is consistent. Then $\widehat{\beta}_n$ is asymptotically normal.

Proof. To prove the theorem we apply the Taylor series expansion and Lemmas 3.1, 3.2, 3.4 and 3.5.

$$\begin{aligned}Q_n(\beta) &= \frac{1}{n}\sum_{i=1}^n\{\| y_i-f_i(x_i,\beta)\|^2+h_i(x_i,\beta)-\| f_i(x_i,\beta)\|^2+2d_i(x_i,\beta)\}\\ &= \frac{1}{n}\sum_{i=1}^n\{y_i^\top y_i-2y_i^\top f_i(x_i,\beta)+h_i(x_i,\beta)+2d_i(x_i,\beta)\}.\end{aligned}$$

Since β_0 is an interior point of Θ and $\widehat{\beta}_n$ is consistent, $\widehat{\beta}_n$ is in the interior of Θ with great probability if n is large enough. As $\widehat{\beta}_n$ is the minimum point of $Q_n(\beta)$,

$$\frac{\partial Q_n(\widehat{\beta}_n)}{\partial\beta}=0.$$

By the Taylor series expansion around β_0 we get:

$$0=\frac{\partial Q_n(\beta_0)}{\partial\beta}+\frac{\partial^2 Q_n(\widetilde{\beta}_n)}{\partial\beta^2}(\widehat{\beta}_n-\beta_0),$$

that is

$$\begin{aligned}0 &= \sum_{i=1}^n\Big\{\frac{\partial h_i(x_i,\beta_0)}{\partial\beta}-2y_i^\top\frac{\partial f_i(x_i,\beta_0)}{\partial\beta}+2\frac{\partial d_i(x_i,\beta_0}{\partial\beta}\Big\}\\ &\quad+\sum_{i=1}^n\Big\{\frac{\partial^2 h_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}-2y_i^\top\frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}+2\frac{\partial^2 d_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}\Big\}(\widehat{\beta}_n-\beta_0),\end{aligned}$$

where $\widetilde{\beta}_n$ is a point between $\widehat{\beta}_n$ and β_0. Actually $\widetilde{\beta}_n$ can depend on the coordinate, but to simplify the notation we do not indicate it. Obviously, if $\widehat{\beta}_n$ converges to β_0

in probability so does $\widetilde{\beta}_n$. As $y_i = g(\xi_i, \beta_0) + \delta_i$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big\{ 2(g(\xi_i,\beta_0)+\delta_i)^\top \frac{\partial f_i(x_i,\beta_0)}{\partial \beta} - \frac{\partial h_i(x_i,\beta_0)}{\partial \beta} - 2\frac{\partial d_i(x_i,\beta_0)}{\partial \beta} \Big\} \tag{3.11}$$
$$= \frac{1}{n} \sum_{i=1}^{n} \Big\{ \frac{\partial^2 h_i(x_i,\widetilde{\beta}_n)}{\partial \beta^2} - 2(g(\xi_i,\beta_0)+\delta_i)^\top \frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial \beta^2} + 2\frac{\partial^2 d_i(x_i,\widetilde{\beta}_n)}{\partial^2 \beta} \Big\}$$
$$\times \sqrt{n}(\widehat{\beta}_n - \beta_0).$$

First we have to check the conditions of Lemma 3.4 for the left hand side of (3.11). By virtue of conditions (B4) and (B5) we can use the dominated convergence theorem and interchange the differentiation and integration, so

$$\begin{aligned}
\mathbb{E}\frac{\partial f_i(x_i,\beta)}{\partial \beta} &= \frac{\partial g(\xi_i,\beta)}{\partial \beta}, \\
\mathbb{E}\frac{\partial h_i(x_i,\beta)}{\partial \beta} &= 2g^\top(\xi_i,\beta)\frac{\partial g(\xi_i,\beta)}{\partial \beta}, \\
\mathbb{E}\frac{\partial d_i(x_i,\beta)}{\partial \beta} &= \mathbb{E}\delta_i^\top \frac{\partial f(\xi_i,\beta)}{\partial \beta},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\frac{\partial^2 f_i(x_i,\beta)}{\partial \beta^2} &= \frac{\partial^2 g(\xi_i,\beta)}{\partial \beta^2}, \\
\mathbb{E}\frac{\partial^2 h_i(x_i,\beta)}{\partial \beta^2} &= 2\frac{\partial g^\top(\xi_i,\beta)}{\partial \beta}\frac{\partial g(\xi_i,\beta)}{\partial \beta} + 2g^\top(\xi_i,\beta)\frac{\partial^2 g(\xi_i,\beta)}{\partial \beta^2}, \\
\mathbb{E}\frac{\partial^2 d_i(x_i,\beta)}{\partial \beta^2} &= \mathbb{E}\delta_i^\top \frac{\partial^2 f(\xi_i,\beta)}{\partial \beta^2}.
\end{aligned}$$

Hence

$$\mathbb{E}\Big\{ 2(g(\xi_i,\beta_0)+\delta_i)^\top \frac{\partial f_i(x_i,\beta_0)}{\partial \beta} - \frac{\partial h_i(x_i,\beta_0)}{\partial \beta} - 2\frac{\partial d_i(x_i,\beta_0)}{\partial \beta} \Big\} = 0.$$

Let $\delta > 0$.

$$\begin{aligned}
&\mathbb{E}\left\| 2(g(\xi_i,\beta_0)+\delta_i)^\top \frac{\partial f_i(x_i,\beta_0)}{\partial \beta} - \frac{\partial h_i(x_i,\beta_0)}{\partial \beta} - 2\frac{\partial d_i(x_i,\beta_0)}{\partial \beta} \right\|^{2+\delta} \\
&\quad \le c\mathbb{E}\left\| \frac{\partial h_i(x_i,\beta_0)}{\partial \beta} \right\|^{2+\delta} + c \parallel g(\xi_i,\beta_0) \parallel^{2+\delta} \mathbb{E}\left\| \frac{\partial f_i(x_i,\beta_0)}{\partial \beta} \right\|^{2+\delta} \\
&\quad + c\mathbb{E}\left\| \delta_i^\top \frac{\partial f_i(x_i,\beta_0)}{\partial \beta} \right\|^{2+\delta} + c\mathbb{E}\left\| \frac{\partial d_i(x_i,\beta_0)}{\partial \beta} \right\|^{2+\delta},
\end{aligned}$$

which is bounded due to conditions (A2) and (B3). Therefore the left hand side of (3.11) has normal limit distribution.

Let us deal with the right hand side of (3.11).

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\partial^2 h_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}-2(g(\xi_i,\beta_0)+\delta_i)^\top\frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}+2\frac{\partial^2 d_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}\Big\}$$
$$=\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\partial^2 h_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}-\mathbb{E}\frac{\partial^2 h_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}+2\frac{\partial g^\top(\xi_i,\widetilde{\beta}_n)}{\partial\beta}\frac{\partial g(\xi_i,\widetilde{\beta}_n)}{\partial\beta}$$
$$+2g^\top(\xi_i,\widetilde{\beta}_n)\frac{\partial^2 g(\xi_i,\widetilde{\beta}_n)}{\partial\beta^2}-2g^\top(\xi_i,\beta_0)\frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}+2g^\top(\xi_i,\beta_0)\mathbb{E}\frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}$$
$$-2g^\top(\xi_i,\beta_0)\frac{\partial^2 g(\xi_i,\widetilde{\beta}_n)}{\partial\beta^2}-2\delta_i^\top\frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}+\frac{\partial^2 d_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}\Big\}.$$

As $\widehat{\beta}_n$ is consistent and, by (B3) and (B6), the first and second derivatives of g are bounded, we have

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{g^\top(\xi_i,\widetilde{\beta}_n)\frac{\partial^2 g(\xi_i,\widetilde{\beta}_n)}{\partial\beta^2}-g^\top(\xi_i,\beta_0)\frac{\partial^2 g(\xi_i,\widetilde{\beta}_n)}{\partial\beta^2}\Big\}\to 0$$

in probability, as $n\to\infty$.

Condition (B11), the consistency of $\widehat{\beta}_n$ and Lemma 3.5 imply that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g^\top(\xi_i,\widetilde{\beta}_n)}{\partial\beta}\frac{\partial g(\xi_i,\widetilde{\beta}_n)}{\partial\beta}\to m_\infty(\beta_0)$$

in probability, where $m_\infty(\beta_0)$ is a regular matrix.

Using Lemma 3.1 it is easy to prove that the following stochastic convergences hold for each $\beta\in\Theta$, as $n\to\infty$:

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\partial^2 h_i(x_i,\beta)}{\partial\beta^2}-\mathbb{E}\frac{\partial^2 h_i(x_i,\beta)}{\partial\beta^2}\Big\}\to 0, \tag{3.12}$$
$$\frac{1}{n}\sum_{i=1}^{n}\Big\{g^\top(\xi_i,\beta_0)\frac{\partial^2 f_i(x_i,\beta)}{\partial\beta^2}-g^\top(\xi_i,\beta_0)\mathbb{E}\frac{\partial^2 f_i(x_i,\beta)}{\partial\beta^2}\Big\}\to 0, \tag{3.13}$$
$$\frac{1}{n}\sum_{i=1}^{n}\Big\{\delta_i^\top\frac{\partial^2 f_i(x_i,\beta)}{\partial\beta^2}-\frac{\partial^2 d_i(x_i,\beta)}{\partial\beta^2}\Big\}\to 0. \tag{3.14}$$

For example, let $s, k \in \{1, 2, \ldots, r\},\ 0 < \delta < 2,\ \varepsilon > 0$. Then

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i^\top \frac{\partial^2 f_i(x_i,\beta)}{\partial\beta_s\partial\beta_k} - \frac{\partial^2 d_i(x_i,\beta)}{\partial\beta_s\partial\beta_k}\right\}\right| > \varepsilon\right) \\
&\quad \le \frac{1}{(n\varepsilon)^{1+\delta/2}}\mathbb{E}\left|\sum_{i=1}^{n}\left\{\delta_i^\top \frac{\partial^2 f_i(x_i,\beta)}{\partial\beta_s\partial\beta_k} - \frac{\partial^2 d_i(x_i,\beta)}{\partial\beta_s\partial\beta_k}\right\}\right|^{1+\delta/2} \\
&\quad \le \frac{c}{(n\varepsilon)^{1+\delta/2}}\sum_{i=1}^{n}\left(\mathbb{E}\left|\delta_i^\top \frac{\partial^2 f_i(x_i,\beta)}{\partial\beta_s\partial\beta_k} - \frac{\partial^2 d_i(x_i,\beta)}{\partial\beta_s\partial\beta_k}\right|^{1+\delta}\right)^{\frac{1+\delta/2}{1+\delta}} \\
&\quad \le \frac{c}{(n\varepsilon)^{1+\delta/2}}\sum_{i=1}^{n}\left(c\mathbb{E}\left\|\delta_i^\top \frac{\partial^2 f_i(x_i,\beta)}{\partial\beta_s\partial\beta_k}\right\|^{1+\delta} + c\mathbb{E}\left\|\frac{\partial^2 d_i(x_i,\beta)}{\partial\beta_s\partial\beta_k}\right\|^{1+\delta}\right)^{\frac{1+\delta/2}{1+\delta}},
\end{aligned}$$

where c is a positive constant which does not depend on n. By condition (B6) the quantity above converges to zero as $n \to \infty$, which proves (3.14).

We have to prove uniform convergence in (3.12)-(3.14). Let $s, k \in \{1, 2, \ldots, r\}$ and $\varepsilon > 0$.

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\|\beta-\overline{\beta}\|\le h}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\delta_i^\top \frac{\partial^2 f_i(x_i,\beta)}{\partial\beta_s\partial\beta_k} - \frac{\partial^2 d_i(x_i,\beta)}{\partial\beta_s\partial\beta_k} - \delta_i^\top \frac{\partial^2 f_i(x_i,\overline{\beta})}{\partial\beta_s\partial\beta_k} + \frac{\partial^2 d_i(x_i,\overline{\beta})}{\partial\beta_s\partial\beta_k}\right)\right| > \varepsilon\right) \\
&\qquad \le \frac{1}{n\varepsilon}\mathbb{E}\sup_{\|\beta-\overline{\beta}\|\le h}\sum_{i=1}^{n}\left\|\delta_i^\top \frac{\partial^2 f_i(x_i,\beta)}{\partial\beta_s\partial\beta_k} - \delta_i^\top \frac{\partial^2 f_i(x_i,\overline{\beta})}{\partial\beta_s\partial\beta_k}\right\| \\
&\qquad\quad + \frac{1}{n\varepsilon}\mathbb{E}\sup_{\|\beta-\overline{\beta}\|\le h}\sum_{i=1}^{n}\left\|\frac{\partial^2 d_i(x_i,\beta)}{\partial\beta_s\partial\beta_k} - \frac{\partial^2 d_i(x_i,\overline{\beta})}{\partial\beta_s\partial\beta_k}\right\|.
\end{aligned}$$

Taking first the lim sup in n of both sides and after that the limit as $h \to 0$, conditions (B9) and (B10) ensure that (3.1) holds for

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i^\top \frac{\partial^2 f_i(x_i,\beta)}{\partial\beta_s\partial\beta_k} - \frac{\partial^2 d_i(x_i,\beta)}{\partial\beta_s\partial\beta_k}\right\}.$$

Hence, by Lemma 3.2, the convergence in (3.14) holds uniformly in β. Using (B7) and (B8), the same result can be proved for (3.12) and (3.13). By Lemma 3.5, this implies that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial^2 h_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2} - \mathbb{E}\frac{\partial^2 h_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}\right\} \to 0, \\
&\frac{1}{n}\sum_{i=1}^{n}\left\{g^\top(\xi_i,\beta_0)\frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2} - g^\top(\xi_i,\beta_0)\mathbb{E}\frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}\right\} \to 0, \\
&\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i^\top \frac{\partial^2 f_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2} - \frac{\partial^2 d_i(x_i,\widetilde{\beta}_n)}{\partial\beta^2}\right\} \to 0
\end{aligned}$$

in probability ($n \to \infty$). Therefore, the coefficient on the right hand side of (3.11) converges to $m_\infty(\beta_0)$. This completes the proof. $\square$

4. Consistency and asymptotic normality for increasing domain in spatial models

First we describe the pattern of observations in the case of spatial processes when increasing domain asymptotic properties are to be studied. We suppose that we observe the process on a subset of the d-dimensional integer lattice points $\mathbb{Z}^d$. Let $\mathcal{P}_n$ be the set of points in $\mathbb{Z}^d$ where the process is observed. To find asymptotic behaviour of the estimator we shall suppose that $\mathcal{P}_n$, $n = 1, 2, \dots$, is an increasing sequence of finite subsets of $\mathbb{Z}^d$ with $|\mathcal{P}_n| \to \infty$, as $n \to \infty$. Let $\mathcal{P}_\infty = \cup_{n=1}^{\infty} \mathcal{P}_n$.

Our n'th set of observations is (y_i, x_i), $i \in \mathcal{P}_n$, where (y_i, x_i) satisfies model equations (2.1)–(2.2). $\widehat{\beta}_n$ is the minimum point of

$$Q_n(\beta) = \frac{1}{|\mathcal{P}_n|} \sum_{i \in \mathcal{P}_n} \left\{ \|y_i - f_i(x_i, \beta)\|^2 + h_i(x_i, \beta) - \|f_i(x_i, \beta)\|^2 + 2d_i(x_i, \beta) \right\}, \tag{4.1}$$

where f_i, h_i and d_i satisfy assumptions (Af), (Ah) and (Ad), respectively.

As the formulae in the spatial case are obtained from the formulae used in the temporal case by systematic modifications, we do not repeat all of them. To get formulae for the spatial case one has to substitute in temporal formulae expression $\frac{1}{|\mathcal{P}_n|} \sum_{i \in \mathcal{P}_n}$ for $\frac{1}{n} \sum_{i=1}^{n}$, $\sup_{n \in \mathcal{P}_\infty}$ for $\sup_n$, and $\lim_{n\to\infty} \sup_{i \in (\mathcal{P}_\infty - \mathcal{P}_n)}$ for $\limsup_{n\to\infty}$.

We shall suppose that random error terms $(\delta_i, \varepsilon_i)$, $i \in \mathcal{P}_\infty$, satisfy α-mixing conditions. We shall use the notion of α-mixing for random fields in the sense of [4]. Let η_s, $s \in \mathbb{Z}^d$, be a random field. For $\Lambda \subseteq \mathbb{Z}^d$ let $\mathcal{A}_\Lambda$ denote the σ-algebra generated by $\{\eta_s \ : \ s \in \Lambda\}$. Let ϱ denote the max distance in $\mathbb{Z}^d$: $\varrho(\mathbf{u}, \mathbf{v}) = \max_{1 \le i \le d} |u_i - v_i|$, for $\mathbf{u} = (u_1, \dots, u_d)^\top, \mathbf{v} = (v_1, \dots, v_d)^\top \in \mathbb{Z}^d$. Let $|\Lambda|$ denote the cardinality of Λ and let $\varrho(\Lambda_1, \Lambda_2)$ denote the distance of Λ_1 and Λ_2 for Λ_1, $\Lambda_2 \subseteq \mathbb{Z}^d$. Let

$$\alpha_\eta(n; u, v) =$$

$$= \sup\{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| \ : \ A \in \mathcal{A}_{\Lambda_1}, B \in \mathcal{A}_{\Lambda_2}, |\Lambda_1| \le u, |\Lambda_2| \le v, \varrho(\Lambda_1, \Lambda_2) \ge n\}.$$

We need the following condition for mixing coefficients α_η of the random field $\{\eta_s = (\delta_s, \epsilon_s) : \ s \in \mathcal{P}_\infty\}$. For $a > 0$,

$$\sum_{i=1}^{\infty} i^{d-1} [\alpha_\eta(i; 1, 1)]^{a/(2+a)} < \infty. \tag{4.2}$$

To prove the consistency of $\widetilde{\beta}_n$ we need the spatial version of Lemma 3.1.

Lemma 4.1 *Let η_s, $s \in \mathcal{P}_\infty$, be a centered random field, $1 < t \le 2$ and $a > 0$. Suppose that condition (4.2) is satisfied. Assume that $\mathbb{E}|\eta_s|^{t+a}$ is finite for every $s \in \mathcal{P}_\infty$. Then there exists a constant K depending only on t and mixing coefficients $\alpha_\eta(i; 1, 1)$ such that*

$$\mathbb{E} \left| \sum_{s \in \mathcal{P}} \eta_s \right|^t \le K \sum_{s \in \mathcal{P}} (\mathbb{E}|\eta_s|^{t+a})^{t/(t+a)} \tag{4.3}$$

for any finite set $\mathcal{P} \subset \mathcal{P}_\infty$.

The proof can be found e.g. in [7].

In the following theorem all formulae are to be understood as spatial versions of the temporal ones.

Theorem 4.2 *Suppose that observations (y_i, x_i), $i \in \mathcal{P}_n$, satisfy model equations (2.1)–(2.2) and suppose that conditions (Af), (Ah), (Ad) and (A1)–(A8) are satisfied. Moreover, assume that for some $1 < t \leq 2$ and $a > 0$ conditions (4.2), and (A9)–(A12) are satisfied. Then $\lim_{n\to\infty} \widehat{\beta}_n = \beta_0$ in probability, where $\widehat{\beta}_n$ is the estimator defined by (4.1).*

Proof. The proof is the same as that of Theorem 3.3 if we use Lemma 4.1 instead of Lemma 3.1. □

The following lemma is a CLT for random fields (see [1, 10]). Let η_s, $s \in \mathbb{Z}^d$, be a centered (vector valued) random field. and let $\mathcal{P}_n, n = 1, 2, \ldots$, be a strictly increasing sequence of finite subsets of $\mathbb{Z}^d$. Let $S_n = \sum_{s\in\mathcal{P}_n} \eta_s$ and denote the variance (matrix) of S_n by Σ_n. Consider the following assumptions:

$$\sum_{i=1}^{\infty} i^{d-1}\alpha_\eta(i; u, v) < \infty \quad \text{if} \quad u + v \leq 4 \quad \text{and} \quad \alpha_\eta(i; 1, \infty) = \mathrm{o}(i^{-d}). \tag{4.4}$$

$$\exists\, a > 0 \quad \text{such that } \{\|\eta_s\|^{2+a} \,:\, s \in \mathbb{Z}^d\} \text{ are uniformly integrable.} \tag{4.5}$$

$$\liminf_{n\to\infty} \lambda_{\min}\left(|\mathcal{P}_n|^{-1}\Sigma_n\right) > 0. \tag{4.6}$$

Lemma 4.3 (Guyon [10, Theorem (3.3.1)].) *(a) Assume that condition (4.2) and the uniform integrability condition (4.5) are satisfied (with the same a). Then*

$$\limsup_n |\mathcal{P}_n|^{-1} \sum_{u,v\in\mathcal{P}_n} \|\operatorname{cov}(\eta_u, \eta_v)\| < \infty.$$

(b) If, moreover, (4.4) and (4.6) are satisfied, then $(\Sigma_n)^{-1/2} S_n \Rightarrow \mathcal{N}(0, I)$, as $n \to \infty$.

Remark that in [10] only the uniform boundedness of $(2+a)$'th moments of $\{\eta_s\}$ was demanded. But actually the stronger condition (4.5) is needed in order to be able to reduce by truncation the CLT to the case of uniformly bounded random vectors.

Theorem 4.4 *Suppose that observations (y_i, x_i), $i \in \mathcal{P}_n$, satisfy model equations (2.1)–(2.2) and suppose that conditions (Af), (Ah), (Ad), (A2), (4.4) and for some $a > 0$ condition (4.2) are fulfilled. Moreover, assume that conditions (B1) – (B11) are satisfied, in (B3) with $a + \lambda$ ($\lambda > 0$) instead of a, and (4.6) satisfied with*

$$S_n = \sum_{i\in\mathcal{P}_n} \left(\frac{\partial h_i(\xi_i + \varepsilon_i, \beta_0)}{\partial\beta} - 2\frac{\partial f_i(\xi_i + \varepsilon_i, \beta_0)}{\partial\beta}^{\top} [g_i(\xi_i, \beta_0) + \delta_i] + 2\frac{\partial d_i(\xi_i + \varepsilon_i, \beta_0)}{\partial\beta} \right).$$

If $\widehat{\beta}_n$ is consistent then it is asymptotically normal.

Proof. The proof is the same as that of Theorem 3.6 if we use Lemma 4.3 instead of Lemma 3.4. □

5. Results for the polynomial model

In this section the appropriate versions of Theorems 3.3 and 3.6 are formulated for the polynomial model of Example 2.1. Let $(\beta_{00}, \beta_{01}, \ldots, \beta_{0k})^\top$ denote the true parameter.

Theorem 5.1 *Consider the polynomial model*

$$y_i = \sum_{j=0}^{k} \beta_{0j} \xi_i^j + \delta_i, \tag{5.1}$$

$$x_i = \xi_i + \varepsilon_i, \tag{5.2}$$

and suppose that condition (B1) is satisfied, V is bounded, $\xi_i \in V$ for each i and

$$\liminf_{n\to\infty} \lambda_{\min}\Big(\frac{1}{n}\sum_{i=1}^{n}(1, \xi_i, \ldots, \xi_i^k)^\top (1, \xi_i, \ldots, \xi_i^k)\Big) > 0. \tag{5.3}$$

Moreover, assume that for some $t > 1$ and $a > 0$ condition (Aα) is satisfied and

$$\mathbb{E}|\delta_i|^{2t+a} \le K \quad \text{and} \quad \mathbb{E}|\varepsilon_i|^{2kt+a} \le L, \qquad i \in \mathbb{N}, \tag{5.4}$$

for some $K \ge 0$ and $L \ge 0$. Then the estimator defined by (2.3) is a consistent estimator of $(\beta_{00}, \beta_{01}, \ldots, \beta_{0k})^\top$

Proof. In the proof we have to check conditions (A2)–(A12) for the polynomial model (5.1)–(5.2). Conditions (A2) and (A4) follow from the assumptions concerning V.

Next, we prove that (A3) holds, that is for each sequence $\{\xi_n\}$ and for each $\beta^* \in \Theta$ and $b > 0$

$$\liminf_{n\to\infty} \inf_{\|\beta-\beta^*\|\ge b} \frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j=0}^{k}(\beta_j - \beta_j^*)\xi_i^j\Big)^2 > 0. \tag{5.5}$$

As the expression in (5.5) is a quadratic form of $\beta - \beta^*$, we get

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j=0}^{k}(\beta_j - \beta_j^*)\xi_i^j\Big)^2 &\ge \|\beta-\beta^*\|^2 \lambda_{\min}\Big(\frac{1}{n}\sum_{i=1}^{n}(1, \xi_i, \ldots, \xi_i^k)^\top (1, \xi_i, \ldots, \xi_i^k)\Big) \\ &\ge b^2 \lambda_{\min}\Big(\frac{1}{n}\sum_{i=1}^{n}(1, \xi_i, \ldots, \xi_i^k)^\top (1, \xi_i, \ldots, \xi_i^k)\Big).\end{aligned}$$

Hence, (5.3) ensures that (A3) is satisfied.

As V is bounded and

$$t_r^{(i)}(x) \quad \text{and} \quad d_s^{(i)}(x)$$

are polynomials of x of order r and s, respectively, (5.4) implies that

$$\mathbb{E}t_r^{(i)}(\xi_i + \varepsilon_i) \le \widetilde{K} \quad \text{and} \quad \mathbb{E}d_s^{(i)}(\xi_i + \varepsilon_i) \le \widetilde{L} \tag{5.6}$$

for some $\widetilde{K} \geq 0$ and $\widetilde{L} \geq 0$, $i \in \mathbb{N}$, $r = 0, 1, \ldots, 2k$, $s = 0, 1, \ldots, k$.

$$\begin{aligned}
\mathbb{E} \sup_{\|\beta_1 - \beta_2\| \leq \ell} |f(x_i, \beta_1) - f(x_i, \beta_2)| &= \mathbb{E} \sup_{\|\beta_1 - \beta_2\| \leq \ell} \Big| \sum_{j=0}^{k} (\beta_{1j} - \beta_{2j}) t_j^{(i)}(\xi_i + \varepsilon_i) \Big| \\
&\leq \ell \sum_{j=0}^{k} \mathbb{E} \big| t_j^{(i)}(\xi_i + \varepsilon_i) \big|,
\end{aligned}$$

so (5.6) implies that (A5) is satisfied. We can check (A6) and (A8) in a similar way.

$$\begin{aligned}
\mathbb{E} \sup_{\|\beta_1 - \beta_2\| \leq \ell} |h(x_i, \beta_1) - h(x_i, \beta_2)| &= \mathbb{E} \sup_{\|\beta_1 - \beta_2\| \leq \ell} \Big| \sum_{j=0}^{k} \frac{\partial h(x_n, \widetilde{\beta})}{\partial \beta_j} (\beta_{1j} - \beta_{2j}) \Big| \\
= \mathbb{E} \sup_{\|\beta_1 - \beta_2\| \leq \ell} \Big| \sum_{j=0}^{k} \Big(\sum_{u=0}^{k} \widetilde{\beta}_u t_{u+j}^{(i)}(\xi_i + \varepsilon_i) \Big) (\beta_{1j} - \beta_{2j}) \Big| \\
\leq \ell \sum_{j=0}^{k} \mathbb{E} \Big| \sum_{u=0}^{k} \widetilde{\beta}_u t_{u+j}^{(i)}(\xi_i + \varepsilon_i) \Big|,
\end{aligned}$$

where $\widetilde{\beta}$ is an element of Θ between β_1 and β_2. Now, (A7) follows from (B1) and (5.6).

At the end, (5.4) and the fact that V is assumed to be bounded directly imply that (A9)–(A12) are satisfied. □

Now, let us formulate the asymptotic normality result for the polynomial model (5.1)–(5.2).

Theorem 5.2 *Suppose that for the model (5.1)–(5.2) condition (B1) is satisfied, V is bounded, $\xi_i \in V$ for each i and $\frac{1}{n}\sum_{i=1}^{n}(1, \xi_i, \ldots, \xi_i^k)^\top (1, \xi_i, \ldots, \xi_i^k)$ has a regular limit as $n \to \infty$. Moreover, assume that for a positive real number a*

$$\mathbb{E}|\delta_i|^{4+a} \leq K \quad \text{and} \quad \mathbb{E}|\varepsilon_i|^{4k+a} \leq L, \qquad i \in \mathbb{N}, \tag{5.7}$$

for some $K \geq 0$ and $L \geq 0$, and condition (Aα) with $t = 2$ is satisfied. Assume that (3.10) is satisfied with $\eta_i = (\eta_0^{(i)}, \eta_1^{(i)}, \ldots, \eta_k^{(i)})^\top$, where

$$\eta_j^{(i)} = \sum_{s=0}^{k} \beta_{0s} t_{j+s}^{(i)}(\xi_i + \varepsilon_i) - 2t_j^{(i)}(\xi_i + \varepsilon_i) \Big(\sum_{r=0}^{k} \beta_{0r} \xi_i^r + \delta_i \Big) + 2d_j^{(i)}(\xi_i + \varepsilon_i),$$

$j = 0, 1, \ldots, k$, $i \in \mathbb{N}$. Suppose that $\widehat{\beta}_n$ is consistent. Then $\widehat{\beta}_n$ is asymptotically normal.

Proof. To prove the theorem we have to check conditions (B2)–(B11). We can verify that (B3)–(B6) are satisfied with a reasoning, similar to the one in the proof of Theorem 5.1. The result follows from (B1), (5.7) and the fact that V is assumed to be bounded.

As the second derivatives $\frac{\partial^2 f_i(\xi, \beta)}{\partial \beta^2}$, $\frac{\partial^2 h_i(\xi, \beta)}{\partial \beta^2}$ and $\frac{\partial^2 d_i(\xi, \beta)}{\partial \beta^2}$ do not depend on β, conditions (B7)–(B10) are trivially satisfied. □

6. Simulation results

All simulations were made with the MATLAB system.

In each of the following cases 2000 repetitions were made and the main characteristics of the alternative estimator (2.3) and the naive estimator

$$\widetilde{\beta} = \arg\min_{\beta} \frac{1}{|\mathcal{P}|} \sum_{i\in\mathcal{P}} \|y_i - g(x_i,\beta)\|^2, \tag{6.1}$$

were calculated, where $\mathcal{P}$ is the set of points where the considered process is observed.

We remark that in the following two examples the estimates were found by solving the appropriate systems of equations of the form

$$\frac{\partial Q_n(\beta)}{\partial \beta} = 0.$$

These systems occurred to be linear both for the polynomial and for the trigonometric polynomial model.

Example 6.1 Consider the polynomial model of Example 2.1 with $k = 2$

$$\begin{aligned} y_i &= \mu + \nu\xi_i + \varrho\xi_i^2 + \delta_i, \\ x_i &= \xi_i + \varepsilon_i, \qquad\qquad i \in \mathbb{N}, \end{aligned}$$

where the parameter vector $\beta = (\mu, \nu, \varrho)^\top$ is in $\mathbb{R}^3$, all the other variables are one-dimensional.

The error terms ε_i and δ_i are components of a two-dimensional stationary AR(1) process, that is

$$\begin{pmatrix} \varepsilon_i \\ \delta_i \end{pmatrix} = A \begin{pmatrix} \varepsilon_{i-1} \\ \delta_{i-1} \end{pmatrix} + B \begin{pmatrix} \eta_i \\ \zeta_i \end{pmatrix}, \qquad i \in \mathbb{N},$$

where the eigenvalues of A are inside the unit circle, η_i and ζ_i are independent standard normal and $(\varepsilon_1, \delta_1)^\top$ is chosen according to stationarity. This ensures that condition (5.4) is satisfied.

Case 1. The data are taken at locations $\mathcal{P} = \{1 \le i \le 31\}$, the original parameters are: $\mu = 3,\ \nu = -1$ and $\varrho = 2$. The autoregressive parameter matrices A and B are:

$$A = \begin{pmatrix} 1/5 & 1/3 \\ 1/2 & -1/4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/20 & 2/20 \\ 3/20 & 1/20 \end{pmatrix}.$$

The design points $\xi_i,\ i \in \mathbb{N}$, are the following functions of i

$$\xi_i = 10\sin(i-1).$$

The results of the simulation are the following:

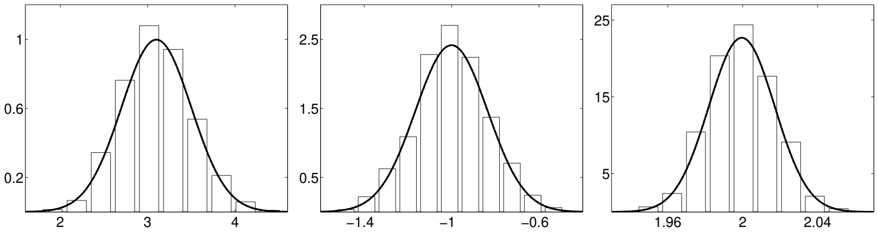

Figure 1: The histograms of the estimators of $\mu,\ \nu$ and ϱ in Case 1 of Example 6.1 for $\mathcal{P}=\{1\le i\le 31\}$.

Alternative estimator		Naive estimator	
$\overline{\mu}=\frac{1}{2000}\sum\widehat{\mu}_i$	3.099371	$\overline{\mu}=\frac{1}{2000}\sum\widetilde{\mu}_i$	3.198841
$\overline{\nu}=\frac{1}{2000}\sum\widehat{\nu}_i$	−1.000639	$\overline{\nu}=\frac{1}{2000}\sum\widetilde{\nu}_i$	−0.997519
$\overline{\varrho}=\frac{1}{2000}\sum\widehat{\varrho}_i$	1.999499	$\overline{\varrho}=\frac{1}{2000}\sum\widetilde{\varrho}_i$	1.995682
$\frac{1}{2000}\sum(\widehat{\mu}_i-\overline{\mu})^2$	0.159847	$\frac{1}{2000}\sum(\widetilde{\mu}_i-\overline{\mu})^2$	0.157116
$\frac{1}{2000}\sum(\widehat{\nu}_i-\overline{\nu})^2$	0.027301	$\frac{1}{2000}\sum(\widetilde{\nu}_i-\overline{\nu})^2$	0.027112
$\frac{1}{2000}\sum(\widehat{\varrho}_i-\overline{\varrho})^2$	0.000309	$\frac{1}{2000}\sum(\widetilde{\varrho}_i-\overline{\varrho})^2$	0.000307
$\frac{1}{2000}\sum(\widehat{\mu}_i-\mu_0)^2$	0.169722	$\frac{1}{2000}\sum(\widetilde{\mu}_i-\mu_0)^2$	0.196653
$\frac{1}{2000}\sum(\widehat{\nu}_i-\nu_0)^2$	0.027302	$\frac{1}{2000}\sum(\widetilde{\nu}_i-\nu_0)^2$	0.027118
$\frac{1}{2000}\sum(\widehat{\varrho}_i-\varrho_0)^2$	0.000309	$\frac{1}{2000}\sum(\widetilde{\varrho}_i-\varrho_0)^2$	0.000325

Now, data are taken at locations $\mathcal{P}=\{1\le i\le 801\}$. The results of the simulation are the following:

Alternative estimator		Naive estimator	
$\overline{\mu}=\frac{1}{2000}\sum\widehat{\mu}_i$	2.995037	$\overline{\mu}=\frac{1}{2000}\sum\widetilde{\mu}_i$	3.250658
$\overline{\nu}=\frac{1}{2000}\sum\widehat{\nu}_i$	−0.999173	$\overline{\nu}=\frac{1}{2000}\sum\widetilde{\nu}_i$	−0.998395
$\overline{\varrho}=\frac{1}{2000}\sum\widehat{\varrho}_i$	2.000112	$\overline{\varrho}=\frac{1}{2000}\sum\widetilde{\varrho}_i$	1.994229
$\frac{1}{2000}\sum(\widehat{\mu}_i-\overline{\mu})^2$	0.007014	$\frac{1}{2000}\sum(\widetilde{\mu}_i-\overline{\mu})^2$	0.006839
$\frac{1}{2000}\sum(\widehat{\nu}_i-\overline{\nu})^2$	0.001030	$\frac{1}{2000}\sum(\widetilde{\nu}_i-\overline{\nu})^2$	0.001022
$\frac{1}{2000}\sum(\widehat{\varrho}_i-\overline{\varrho})^2$	0.000011	$\frac{1}{2000}\sum(\widetilde{\varrho}_i-\overline{\varrho})^2$	0.000011
$\frac{1}{2000}\sum(\widehat{\mu}_i-\mu_0)^2$	0.007039	$\frac{1}{2000}\sum(\widetilde{\mu}_i-\mu_0)^2$	0.069668
$\frac{1}{2000}\sum(\widehat{\nu}_i-\nu_0)^2$	0.001030	$\frac{1}{2000}\sum(\widetilde{\nu}_i-\nu_0)^2$	0.001024
$\frac{1}{2000}\sum(\widehat{\varrho}_i-\varrho_0)^2$	0.000011	$\frac{1}{2000}\sum(\widetilde{\varrho}_i-\varrho_0)^2$	0.000044

On Figures 1 and 2 the histograms of the estimators are showed for Case 1 for two different sample sizes together with the normal probability densities with parameters estimated from the samples.

CASE 2. The data are taken at locations $\mathcal{P}=\{1\le i\le 51\}$, the original parameters are: $\mu=2,\ \nu=1$ and $\varrho=1$. The autoregressive parameter matrices A and B are:

$$A=\begin{pmatrix}1/2 & 1/2\\ 1/3 & 1/2\end{pmatrix},\qquad B=\begin{pmatrix}1/20 & 2/25\\ 3/20 & 1/25\end{pmatrix}.$$

The design points $\xi_i,\ i\in\mathbb{N}$, are the following functions of i

$$\xi_i=10\sin(i-1).$$

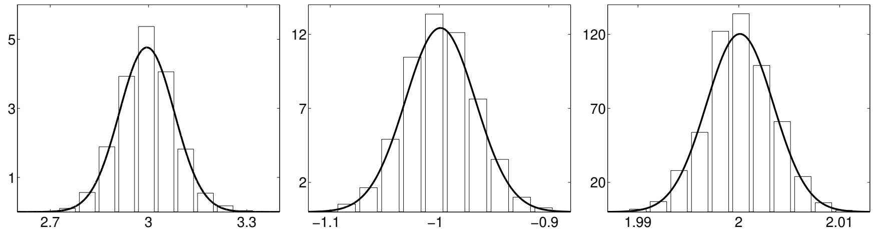

Figure 2: The histograms of the estimators of μ, ν and ϱ in Case 1 of Example 6.1 for $\mathcal{P} = \{1 \leq i \leq 801\}$.

The results of the simulation are the following:

Alternative estimator		Naive estimator	
$\overline{\mu} = \frac{1}{2000}\sum \widehat{\mu}_i$	2.031506	$\overline{\mu} = \frac{1}{2000}\sum \widetilde{\mu}_i$	2.416457
$\overline{\nu} = \frac{1}{2000}\sum \widehat{\nu}_i$	1.003096	$\overline{\nu} = \frac{1}{2000}\sum \widetilde{\nu}_i$	1.002312
$\overline{\varrho} = \frac{1}{2000}\sum \widehat{\varrho}_i$	1.000048	$\overline{\varrho} = \frac{1}{2000}\sum \widetilde{\varrho}_i$	0.991097
$\frac{1}{2000}\sum(\widehat{\mu}_i - \overline{\mu})^2$	0.071584	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \overline{\mu})^2$	0.063543
$\frac{1}{2000}\sum(\widehat{\nu}_i - \overline{\nu})^2$	0.123333	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \overline{\nu})^2$	0.121048
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \overline{\varrho})^2$	0.000070	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \overline{\varrho})^2$	0.000067
$\frac{1}{2000}\sum(\widehat{\mu}_i - \mu_0)^2$	0.072576	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \mu_0)^2$	0.236979
$\frac{1}{2000}\sum(\widehat{\nu}_i - \nu_0)^2$	0.123343	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \nu_0)^2$	0.121053
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \varrho_0)^2$	0.000070	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \varrho_0)^2$	0.000146

Case 3. The data are taken at locations $\mathcal{P} = \{10 \leq i \leq 90\}$, the original parameters are: $\mu = 2$, $\nu = 1$ and $\varrho = 1$. The autoregressive parameter matrices A and B are:

$$A = \begin{pmatrix} 1/7 & 2/5 \\ 1 & 1/4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/20 & 2/20 \\ 3/20 & 1/20 \end{pmatrix}.$$

The design points ξ_i, $i \in \mathbb{N}$, are the following functions of i:

$$\xi_i = \frac{10\cos(i)}{\log(100i)}.$$

The results of the simulation are the following:

Alternative estimator		Naive estimator	
$\overline{\mu} = \frac{1}{2000}\sum \widehat{\mu}_i$	1.959971	$\overline{\mu} = \frac{1}{2000}\sum \widetilde{\mu}_i$	2.145917
$\overline{\nu} = \frac{1}{2000}\sum \widehat{\nu}_i$	0.998530	$\overline{\nu} = \frac{1}{2000}\sum \widetilde{\nu}_i$	1.011423
$\overline{\varrho} = \frac{1}{2000}\sum \widehat{\varrho}_i$	1.076130	$\overline{\varrho} = \frac{1}{2000}\sum \widetilde{\varrho}_i$	0.774954
$\frac{1}{2000}\sum(\widehat{\mu}_i - \overline{\mu})^2$	0.036386	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \overline{\mu})^2$	0.003811
$\frac{1}{2000}\sum(\widehat{\nu}_i - \overline{\nu})^2$	0.020903	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \overline{\nu})^2$	0.009291
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \overline{\varrho})^2$	0.049241	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \overline{\varrho})^2$	0.004741
$\frac{1}{2000}\sum(\widehat{\mu}_i - \mu_0)^2$	0.037989	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \mu_0)^2$	0.025103
$\frac{1}{2000}\sum(\widehat{\nu}_i - \nu_0)^2$	0.020906	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \nu_0)^2$	0.009421
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \varrho_0)^2$	0.055037	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \varrho_0)^2$	0.055387

It is easy to see that in each of the above cases conditions of Theorem 5.1 are fulfilled, so the estimators of β are consistent. Unfortunately, conditions of the asymptotic normality can not be checked even in this simple situation.

Example 6.2 Consider the trigonometric polynomial model of Example 2.2 with $k = 1$

$$\begin{aligned} y_i &= \mu/2 + \nu\cos(\xi_i) + \varrho\sin(\xi_i) + \delta_i, \\ x_i &= \xi_i + \varepsilon_i, \qquad\qquad\qquad i \in \mathbb{Z}, \end{aligned}$$

where the parameter vector $\beta = (\mu, \nu, \varrho)^\top$ is in $\mathbb{R}^3$, all the other variables are one-dimensional.

The error terms are components of a two dimensional stationary AR(1) process that is now defined on $\mathbb{Z}$.

CASE 1. The data are taken at locations $\mathcal{P} = \{0 \le i \le 20\}$, the original parameters are: $\mu = 3$, $\nu = -1$ and $\varrho = 2$. The autoregressive parameter matrices A and B are:

$$A = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 1/2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/10 & 2/10 \\ 3/10 & 1/10 \end{pmatrix}.$$

The design points are the following functions:

$$\xi_i = 10i.$$

The results of the simulation are the following:

Alternative estimator		Naive estimator	
$\overline{\mu} = \frac{1}{2000}\sum\widehat{\mu}_i$	3.018565	$\overline{\mu} = \frac{1}{2000}\sum\widetilde{\mu}_i$	2.963042
$\overline{\nu} = \frac{1}{2000}\sum\widehat{\nu}_i$	−1.011594	$\overline{\nu} = \frac{1}{2000}\sum\widetilde{\nu}_i$	−0.803079
$\overline{\varrho} = \frac{1}{2000}\sum\widehat{\varrho}_i$	1.973366	$\overline{\varrho} = \frac{1}{2000}\sum\widetilde{\varrho}_i$	1.681032
$\frac{1}{2000}\sum(\widehat{\mu}_i - \overline{\mu})^2$	0.690820	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \overline{\mu})^2$	0.673594
$\frac{1}{2000}\sum(\widehat{\nu}_i - \overline{\nu})^2$	0.912250	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \overline{\nu})^2$	0.628640
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \overline{\varrho})^2$	0.343584	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \overline{\varrho})^2$	0.196739
$\frac{1}{2000}\sum(\widehat{\mu}_i - \mu_0)^2$	0.691165	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \mu_0)^2$	0.674959
$\frac{1}{2000}\sum(\widehat{\nu}_i - \nu_0)^2$	0.912384	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \nu_0)^2$	0.667418
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \varrho_0)^2$	0.344294	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \varrho_0)^2$	0.298480

CASE 2. The data are taken at locations $\mathcal{P} = \{0 \le i \le 30\}$, the original parameters are: $\mu = -5$, $\nu = 2$ and $\varrho = -1$. The autoregressive parameter matrices A and B are:

$$A = \begin{pmatrix} 1/5 & 1/3 \\ -1/2 & 1/4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/5 & 2/5 \\ 3/5 & 1/5 \end{pmatrix}.$$

The design points are the following functions:

$$\xi_i = 10i.$$

The results of the simulation are the following:

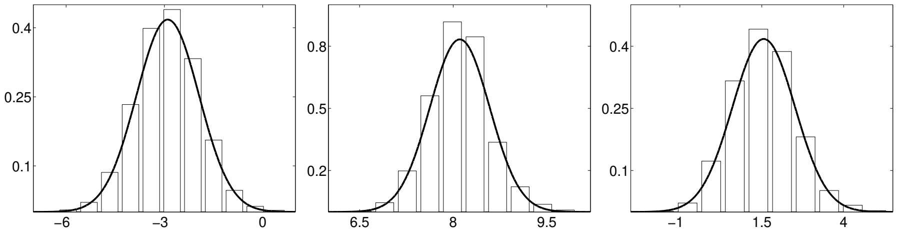

Figure 3: The histograms of the estimators of μ, ν and ϱ in Case 4 of Example 6.2 for $\mathcal{P} = \{0 \le i \le 20\}$.

Alternative estimator		Naive estimator	
$\overline{\mu} = \frac{1}{2000}\sum \widehat{\mu}_i$	-5.013017	$\overline{\mu} = \frac{1}{2000}\sum \widetilde{\mu}_i$	-4.971107
$\overline{\nu} = \frac{1}{2000}\sum \widehat{\nu}_i$	2.149887	$\overline{\nu} = \frac{1}{2000}\sum \widetilde{\nu}_i$	1.745426
$\overline{\varrho} = \frac{1}{2000}\sum \widehat{\varrho}_i$	-1.076837	$\overline{\varrho} = \frac{1}{2000}\sum \widetilde{\varrho}_i$	-0.832176
$\frac{1}{2000}\sum(\widehat{\mu}_i - \overline{\mu})^2$	0.087195	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \overline{\mu})^2$	0.056343
$\frac{1}{2000}\sum(\widehat{\nu}_i - \overline{\nu})^2$	0.245944	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \overline{\nu})^2$	0.056280
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \overline{\varrho})^2$	0.261768	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \overline{\varrho})^2$	0.079877
$\frac{1}{2000}\sum(\widehat{\mu}_i - \mu_0)^2$	0.087364	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \mu_0)^2$	0.057178
$\frac{1}{2000}\sum(\widehat{\nu}_i - \nu_0)^2$	0.268410	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \nu_0)^2$	0.121088
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \varrho_0)^2$	0.267672	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \varrho_0)^2$	0.108042

Case 3. The data are taken at locations $\mathcal{P} = \{0 \le i \le 50\}$, the original parameters are: $\mu = 3$, $\nu = -4$ and $\varrho = -2$. The autoregressive parameter matrices A and B are:

$$A = \begin{pmatrix} 1/5 & 1/3 \\ -1/2 & 1/4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/15 & 2/15 \\ 3/15 & 1/15 \end{pmatrix}.$$

The design points ξ_i, $i \in \mathbb{Z}$, are the following functions of i:

$$\xi_i = 10\exp(-i).$$

The results of the simulation are the following:

Alternative estimator		Naive estimator	
$\overline{\mu} = \frac{1}{2000}\sum \widehat{\mu}_i$	3.085045	$\overline{\mu} = \frac{1}{2000}\sum \widetilde{\mu}_i$	3.292866
$\overline{\nu} = \frac{1}{2000}\sum \widehat{\nu}_i$	-4.050121	$\overline{\nu} = \frac{1}{2000}\sum \widetilde{\nu}_i$	-4.243325
$\overline{\varrho} = \frac{1}{2000}\sum \widehat{\varrho}_i$	-1.760243	$\overline{\varrho} = \frac{1}{2000}\sum \widetilde{\varrho}_i$	-0.785537
$\frac{1}{2000}\sum(\widehat{\mu}_i - \overline{\mu})^2$	0.126423	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \overline{\mu})^2$	0.097756
$\frac{1}{2000}\sum(\widehat{\nu}_i - \overline{\nu})^2$	0.044969	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \overline{\nu})^2$	0.026129
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \overline{\varrho})^2$	0.912295	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \overline{\varrho})^2$	0.146251
$\frac{1}{2000}\sum(\widehat{\mu}_i - \mu_0)^2$	0.133656	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \mu_0)^2$	0.183527
$\frac{1}{2000}\sum(\widehat{\nu}_i - \nu_0)^2$	0.047481	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \nu_0)^2$	0.085336
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \varrho_0)^2$	0.969778	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \varrho_0)^2$	1.621170

Case 4. The data are taken at locations $\mathcal{P} = \{0 \le i \le 20\}$, the original parameters are: $\mu = -3$, $\nu = 8$ and $\varrho = 1.5$. The autoregressive parameter matrices A and B

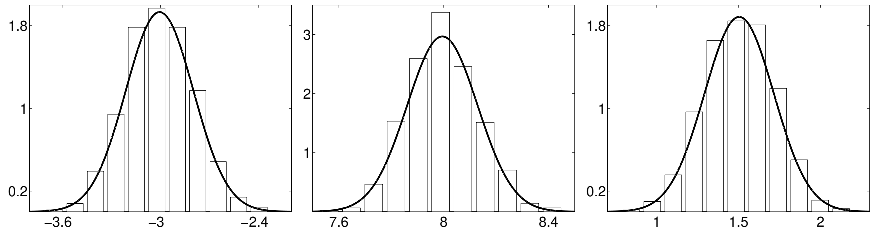

Figure 4: The histograms of the estimators of μ, ν and ϱ in Case 4 of Example 6.2 for $\mathcal{P} = \{-200 \leq i \leq 200\}$.

are:

$$A = \begin{pmatrix} 1/5 & 1/3 \\ -1/2 & 1/4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/8 & 2/8 \\ 3/8 & 1/8 \end{pmatrix}.$$

The design points here are bounded. They have the following form:

$$\xi_i = 100 \sin(i), \qquad i \in \mathbb{Z}.$$

The results of the simulation are the following:

Alternative estimator		Naive estimator	
$\overline{\mu} = \frac{1}{2000}\sum \widehat{\mu}_i$	-2.896695	$\overline{\mu} = \frac{1}{2000}\sum \widetilde{\mu}_i$	-3.403995
$\overline{\nu} = \frac{1}{2000}\sum \widehat{\nu}_i$	8.104989	$\overline{\nu} = \frac{1}{2000}\sum \widetilde{\nu}_i$	7.660003
$\overline{\varrho} = \frac{1}{2000}\sum \widehat{\varrho}_i$	1.558654	$\overline{\varrho} = \frac{1}{2000}\sum \widetilde{\varrho}_i$	1.215840
$\frac{1}{2000}\sum(\widehat{\mu}_i - \overline{\mu})^2$	0.912097	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \overline{\mu})^2$	0.730054
$\frac{1}{2000}\sum(\widehat{\nu}_i - \overline{\nu})^2$	0.229426	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \overline{\nu})^2$	0.123231
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \overline{\varrho})^2$	0.913180	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \overline{\varrho})^2$	0.630973
$\frac{1}{2000}\sum(\widehat{\mu}_i - \mu_0)^2$	0.922769	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \mu_0)^2$	0.893266
$\frac{1}{2000}\sum(\widehat{\nu}_i - \nu_0)^2$	0.240449	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \nu_0)^2$	0.238829
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \varrho_0)^2$	0.916620	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \varrho_0)^2$	0.711720

Now, the data are taken at locations $\mathcal{P} = \{-200 \leq i \leq 200\}$. The results of the simulation are the following:

Alternative estimator		Naive estimator	
$\overline{\mu} = \frac{1}{2000}\sum \widehat{\mu}_i$	-3.005029	$\overline{\mu} = \frac{1}{2000}\sum \widetilde{\mu}_i$	-2.990759
$\overline{\nu} = \frac{1}{2000}\sum \widehat{\nu}_i$	7.994931	$\overline{\nu} = \frac{1}{2000}\sum \widetilde{\nu}_i$	7.576244
$\overline{\varrho} = \frac{1}{2000}\sum \widehat{\varrho}_i$	1.503078	$\overline{\varrho} = \frac{1}{2000}\sum \widetilde{\varrho}_i$	1.422239
$\frac{1}{2000}\sum(\widehat{\mu}_i - \overline{\mu})^2$	0.042654	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \overline{\mu})^2$	0.034586
$\frac{1}{2000}\sum(\widehat{\nu}_i - \overline{\nu})^2$	0.018069	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \overline{\nu})^2$	0.008739
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \overline{\varrho})^2$	0.044800	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \overline{\varrho})^2$	0.032841
$\frac{1}{2000}\sum(\widehat{\mu}_i - \mu_0)^2$	0.042680	$\frac{1}{2000}\sum(\widetilde{\mu}_i - \mu_0)^2$	0.034671
$\frac{1}{2000}\sum(\widehat{\nu}_i - \nu_0)^2$	0.018095	$\frac{1}{2000}\sum(\widetilde{\nu}_i - \nu_0)^2$	0.188308
$\frac{1}{2000}\sum(\widehat{\varrho}_i - \varrho_0)^2$	0.038888	$\frac{1}{2000}\sum(\widetilde{\varrho}_i - \varrho_0)^2$	0.044800

Again, using the same ideas as for the polynomial model, it is not difficult to prove that in Cases 1, 2 and 4 the estimator of the parameter vector β is consistent. In Case 3 condition (5.3) is not satisfied.

From Examples 6.1 and 6.2 we can conclude that the mean of the alternative estimator is always closer to the original parameter to be estimated than that of the naive one. It can be used even in small sample cases but its dispersion is sometimes greater than the dispersion of the naive estimator.

References

[1] Bolthausen, E., On the central limit theorem for stationary mixing random fields. *Ann. Probability*, **10**, 1047-1050 (1982).

[2] Cheng, C.-L. and Schneeweiss, H., Polynomial regression with errors in the variables. *J. R. Statist. Soc.* B, **60**, 189-199 (1998).

[3] Chan, L. K. and Mak, T. K., On the polynomial functional relationship. *J. R. Statist. Soc.* B, **47**, 510-518 (1985).

[4] Doukhan, P., *Mixing. Properties and Examples.* Lecture Notes in Statistics **85**, Springer, New York (1994).

[5] Fazekas, I., Baran, S., Kukush, A., Lauridsen, J., Asymptotic properties in space and time of an estimator in non-linear functional errors-in-variables models. *Technical Report* No. 6/1998, Kossuth University, Debrecen, to appear in *Random Operators and Stochastic Equations* (1999).

[6] Fazekas, I. and Kukush, A. G., Asymptotic properties of an estimator in nonlinear functional errors-in-variables models with dependent error terms. *Computers Math. Applic.* Vol. 34, No. 10, 23-39 (1997).

[7] Fazekas, I., Kukush, A. G. and Tómács, T., On the Rosenthal inequality for mixing fields. *Technical Report* No. 5/1998, Kossuth University, Debrecen, submitted to the *Ukrainian Math. J.* (1998).

[8] Fazekas, I., Kukush, A. G. and Zwanzig, S., On inconsistency of least squares estimator in nonlinear functional relations. *Research Report* No. 1/1998, Department of Statistics and Demography, Odense University, submitted to *Mathematical Methods of Statistics* (1998).

[9] Fuller, W. A., *Measurement Error Models.* Wiley, New York (1987).

[10] Guyon, X., *Random Fields on a Network. Modeling, Statistics, and Applications.* Springer, New York (1995).

[11] Herrndorf, N., A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probability*, **12**, 141-153 (1984).

[12] Kukush, A. G. and Zwanzig, S., On inconsistency of the least squares estimator in nonlinear functional error-in-variables models. *Preprint No. 96-12*, Universität Hamburg (1996).

[13] Kukush, A. G. and Zwanzig, S., On a corrected contrast estimator in the implicit nonlinear functional relation model. *Preprint No. 97-11*, Universität Hamburg (1997).

[14] Pfanzagl, J., On the measurability and consistency of minimum contrast estimates. *Metrika*, **14**, 249-272 (1969).