A new estimator for nonlinear regression models*

S. Baran
Institute of Informatics, University of Debrecen
P. O. Box 12, H-4010 Debrecen, Hungary
barans@inf.unideb.hu

Abstract

In this paper an estimator for the general (nonlinear) regression model is studied which is based on the Fourier transform of a certain weight function. Consistency and asymptotic normality of the estimators are established and simulation results are presented to illustrate the theoretical ones.
Mathematics subject classification. Primary: 62F12; secondary: 62J05.
Key words and phrases. Nonlinear regression, consistency, asymptotic normality, strong mixing, Fourier transform.

1. Introduction

In statistical work one often encounters such situations when the linear model is not appropriate to describe the relationship between the observed processes. Throughout this paper we will deal with the model

$$y_i = g(x_i, \beta_0) + \varepsilon_i, \qquad i \in \mathbb{N}, \tag{1.1}$$

where g is a known function and β_0 denotes the true value of the unknown parameter β to be estimated on the basis of observations on y_i and x_i.

The usual method of estimation in this case is the least squares one, however, the least squares estimator (LSE) is very sensitive to the outlying values of the error

Acknowledgement. I am grateful to my wife Ágnes Baran for her suggestions and remarks concerning the numerical part of the computer simulations.

I would like to thank also István Fazekas for helpful discussions about this topic and for his useful remarks.

*Supported by the Hungarian National Science Foundation OTKA under Grant No. F 032060/2000 and by the Bolyai Grant of the Hungarian Academy of Sciences.

terms. Another popular estimator is the least absolute deviations estimator (LADE), that has been the subject of much attention because of its robust properties.

For the linear model, i.e. when $g(x, \beta) = x^\top \beta$, An *et al.* [1] introduced a new estimator of the unknown parameter, based on the Fourier transform of a symmetric weight function and they proved its strong consistency in case of i.i.d. observations. In [3] the author generalized the estimator proposed in [1] for the linear measurement error model by using the so-called deconvolution method (see e.g. [11] or [2]), and assuming i.i.d. observations he proved its strong consistency. Later, in [4] the author generalized the results of [1] and [3] for the case of dependent (strong mixing) processes.

In this paper we extend the method of An *et al.* [1] to the model (1.1). In Section 2 we describe the new estimator and we give some models where this estimator can be applied. Theorems about the consistency of this new estimator are formulated in Section 3, while Section 4 deals with the asymptotic normality. Proofs of the theorems of these two sections are given in Section 6. At the end, in Section 5 we present some simulation results where we compare the performance of our estimator with the performances of the LSE and LADE.

2. The estimator and the idea behind it

Consider the model (1.1) and assume that the parameter set is p-dimensional: $\beta \in \Theta \subseteq \mathbb{R}^p$; x_i is q-dimensional, y_i and ε_i are scalar random variables ($i \in \mathbb{N}$) that have the same distribution as random variables x, y and ε, respectively. Moreover, assume that sequences $\{x_i\}$ and $\{\varepsilon_i\}$ are independent of each other.

Our results will be derived and analyzed under the following assumption:

Assumption A. For all $\beta \in \Theta$, $\beta \neq \beta_0$, random variable $g(x, \beta) - g(x, \beta_0)$ is non-degenerate, i.e.

$$\mathrm{Var}\left(g(x, \beta) - g(x, \beta_0)\right) \neq 0.$$

Assumption A is needed in order to ensure that our model can recognize the true value of the parameter β.

Now, let $w(t)$ be a continuous density kernel function that satisfies

$$w(t) = w(-t) \geq 0, \quad t > 0, \qquad \text{and} \qquad \int_{-\infty}^{\infty} |t| w(t) dt < \infty \tag{2.1}$$

and let $\varphi_w(v)$ denote the Fourier transform of w, that is

$$\varphi_w(v) := \int_{-\infty}^{\infty} e^{itv} w(t) dt.$$

Let us define the estimator $\widehat{\beta}_n$ of β_0 as the maximum point of

$$A_n(\beta) := \frac{1}{n^2}\sum_{\ell=1}^{n}\sum_{s=1}^{n}\varphi_w\big(y_\ell - y_s - (g(x_\ell,\beta) - g(x_s,\beta))\big). \tag{2.2}$$

Here the kernel function $w(t)$ should be chosen in such a way that its Fourier transform has a closed form. The usual kernels are the densities of the normal distribution $\mathcal{N}(0,a^2)$, the uniform distribution on $[-a,a]$, the symmetric exponential (Laplace) distribution, that is

$$w(t) := \frac{a}{2}e^{-a|t|},$$

or a mixture of two Laplace distributions, e.g.

$$w(t) := \frac{a}{3}(2e^{-a|t|} - e^{-2a|t|}). \tag{2.3}$$

The idea behind this method is the same as in [1]. Let

$$\varphi(t,\beta) := \mathsf{E}\exp(it(y - g(x,\beta))$$

and

$$A(\beta) := \int_{-\infty}^{\infty} |\varphi(\beta,t)|^2 w(t)dt.$$

As x and ε are independent,

$$\big|\varphi(t,\beta)\big| = \big|\mathsf{E}e^{it\varepsilon}\mathsf{E}e^{it(g(x,\beta_0)-g(x,\beta))}\big| \le \big|\mathsf{E}e^{it(g(x,\beta_0)-g(x,\beta))}\big| \le 1,$$

where by Assumption A, if $\beta \ne \beta_0$, the last term is strictly less then 1 except for countable many values of t. Hence,

$$A(\beta_0) = \sup_{\beta\in\mathbb{R}^p} A(\beta) > A(\alpha), \qquad \text{for every } \alpha \ne \beta_0.$$

In fact, $A_n(\beta)$ is the empirical analogue of $A(\beta)$ and, as we will see in Section 3, under certain conditions it converges to $A(\beta)$, uniformly in β.

At the end, we give some examples of regression models, where Assumption A is satisfied.

Example 2.1 *Polynomial model.* Let

$$g(u,\beta) = \beta_0 + \beta_1 u + \beta_2 u^2 + \cdots + \beta_p u^p,$$

where $u, \beta_i \in \mathbb{R}$, $i = 0, 1, \ldots, p$, and the parameter is $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^\top$, and assume that random vector $(1, x, \ldots, x^p)^\top$ is non-degenerate (e.g. if x is absolutely continuous).

Example 2.2 *Trigonometric polynomial model.* Let

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

where $u,\alpha_i,\beta_j\in\mathbb{R}$, $i=0,1,\ldots,k$, $j=1,2,\ldots,k$, and the parameter is $\beta=(\alpha_0,\alpha_1,\beta_1,\ldots,\alpha_k,\beta_k)^\top$, and similarly to Example 2.1 assume that random vector $(1,\cos(x),\sin(x),\ldots,\cos(kx),\sin(kx))^\top$ is non-degenerate.

Example 2.3 *Exponential model.* Let

$$g(u,\beta)=\exp(u^\top\gamma)\tau,$$

where $u,\gamma\in\mathbb{R}^q$, $\tau\in\mathbb{R}$ and the parameter is $\beta=(\gamma^\top,\tau)^\top$, and assume that x is a non-degenerate random vector.

Example 2.4 *Periodic model (sinusoid).* Let

$$g(u,\beta)=R\sin(\omega u),$$

where $u,R,\omega\in\mathbb{R}$ and the parameter is $\beta=(R,\omega)^\top$, and assume that x is an absolutely continuous random variable.

Example 2.5 *Gaussian regression curve.* Let

$$g(u,\beta)=\exp(-\|u-m\|^2/\gamma)\tau,$$

where $u,m\in\mathbb{R}^q$, γ and τ are positive real numbers and the parameter is $\beta=(m^\top,\gamma,\tau)^\top$, and assume that x is a non-degenerate, absolutely continuous random vector with non-symmetric components.

3. Consistency

In the present paper we assume that the sequences $\{x_i\}$ and $\{\varepsilon_i\}$ are strongly mixing. If $\mathcal{A}$ and $\mathcal{B}$ are σ-algebras, then

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

Let $\mathcal{M}_k^\ell$ denote the σ-algebra generated by $\{x_i^\top,\varepsilon_i\ :\ k\le i\le\ell\}$ and let

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

The mixing condition we will use for the dependence structure of $\{x_i\}$ and $\{\varepsilon_i\}$ is

$$\sum_{n=1}^{\infty}\alpha(n)<\infty. \tag{3.1}$$

Theorem 3.1 *Suppose that* $(y_i, x_i^\top)$, $i \in \mathbb{N}$, *satisfy model equation* (1.1) *and Assumption A, the parameter set* Θ *is compact,* g *is continuous in* β,

$$\mathsf{E} \sup_{\beta \in \Theta} |g(x, \beta)| < \infty, \tag{3.2}$$

and mixing condition (3.1) *is satisfied for* $(x_i^\top, \varepsilon_i)$, $i \in \mathbb{N}$. *Then* $\widehat{\beta}_n$ *is a consistent estimator of the true parameter* β_0.

Remark. If Θ is convex and g is differentiable with respect to β then Theorem 3.1 remains valid if (3.2) is changed for

$$\mathsf{E} \sup_{\beta \in \Theta} \left\| \frac{\partial g(x, \beta)}{\partial \beta} \right\| < \infty. \tag{3.3}$$

Example 3.2 Consider the special polynomial model $g(u, \beta) = (u + \beta)^2$, where $u \in \mathbb{R}$ and $\beta \in [a, b] \subset \mathbb{R}$, $-\infty < a < \beta_0 < b < \infty$. Assume that $\mathsf{E}|x| < \infty$, but the second moment of x is not finite. It is easy to see, that in this case (3.3) is satisfied, but (3.2) is not valid.

Naturally, in the case of i.i.d observations $(y_i, x_i^\top)$, $i \in \mathbb{N}$, Theorem 3.1 remains valid without changes, so consistency of the estimator $\widehat{\beta}_n$ is proved. However, in this case we can prove strong consistency, too.

Theorem 3.3 *Suppose that* $(y_i, x_i^\top)$, $i \in \mathbb{N}$, *are i.i.d observations on* $(y, x^\top)$, *where the random vector* $(y, x^\top)$ *satisfies model equation* (1.1) *and Assumption A, and its distribution is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^{q+1}$. *Assume that the parameter set* Θ *is compact,* $g(u, \beta)$ *is differentiable with respect to* u *and* $\frac{\partial g(u,\beta)}{\partial u}$ *is continuous in both of its variables. Moreover, assume that* $\varphi_w(v)$ *is differentiable and*

$$\int_{-\infty}^{\infty} \left| \frac{d\varphi_w(v)}{dv} \right| dv < \infty. \tag{3.4}$$

Then $\widehat{\beta}_n$ *is a strongly consistent estimator of the true parameter* β_0.

Remark. Compactness of Θ, differentiability of g and condition on the distribution of $(y, x^\top)$ are needed to verify that the collection $\mathcal{G}$ of sets

$$D(v, \beta) = \left\{ (z, u^\top) \in \mathbb{R}^{q+1} \;\middle|\; z - g(u, \beta) < v \right\}, \qquad v \in \mathbb{R},\ \beta \in \Theta, \tag{3.5}$$

forms a Glivenko-Cantelli class with respect to the distribution of $(y, x^\top)$. For example, in the linear model, i.e. if $g(u, \beta) = u^\top \beta$, for all $v \in \mathbb{R}$ and $\beta \in \Theta$ the set $D(v, \beta)$ is a half plane of $\mathbb{R}^{q+1}$, so $\mathcal{G}$ is automatically Glivenko-Cantelli (see e.g. [9]).

4. Asymptotic normality

To prove asymptotic normality we need stronger conditions. The following mixing condition will be appropriate for our purposes:

$$b(\alpha, \nu) := \sum_{k=1}^{\infty} \alpha^{\nu/(2+\nu)}(k) < \infty, \tag{4.1}$$

where $\nu > 0$.

We remark that condition (4.1) implies (3.1).

Theorem 4.1 *Suppose that $(y_i, x_i^\top)$, $i \in \mathbb{N}$, satisfy model equation (1.1) and Assumption A, the parameter set Θ is compact and convex, β_0 is an interior point of Θ, g is twice differentiable with respect to β,*

$$\mathrm{Var}\left(\frac{\partial g(x, \beta_0)}{\partial \beta}\right)$$

is a regular matrix and

$$\mathsf{E} \sup_{\beta \in \Theta} \left\| \frac{\partial g(x, \beta)}{\partial \beta} \right\|^2 < \infty, \tag{4.2}$$

$$\mathsf{E} \sup_{\beta \in \Theta} \left\| \frac{\partial^2 g(x, \beta)}{\partial \beta \partial \beta^\top} \right\| < \infty. \tag{4.3}$$

Assume that

$$\lim_{\gamma \to 0} \mathsf{E} \sup_{\|\beta_1 - \beta_2\| \le \gamma} \left\| \frac{\partial g(x, \beta_1)}{\partial \beta} \right\|^2 |g(x, \beta_1) - g(x, \beta_2)| = 0, \tag{4.4}$$

$$\lim_{\gamma \to 0} \mathsf{E} \sup_{\|\beta_1 - \beta_2\| \le \gamma} \left\| \frac{\partial^2 g(x, \beta_1)}{\partial \beta \partial \beta^\top} \right\| |g(x, \beta_1) - g(x, \beta_2)| = 0, \tag{4.5}$$

$$\lim_{\gamma \to 0} \mathsf{E} \sup_{\|\beta_1 - \beta_2\| \le \gamma} \left\| \frac{\partial^2 g(x, \beta_1)}{\partial \beta \partial \beta^\top} - \frac{\partial^2 g(x, \beta_2)}{\partial \beta \partial \beta^\top} \right\| = 0. \tag{4.6}$$

Suppose also that for some positive constant ν

$$\mathsf{E} \left\| \frac{\partial^2 g(x, \beta)}{\partial \beta \partial \beta^\top} \right\|^{1+2\nu} < \infty, \tag{4.7}$$

$$\mathsf{E} \left\| \frac{\partial g(x, \beta)}{\partial \beta} \right\|^{2+4\nu} < \infty, \tag{4.8}$$

for all $\beta \in \Theta$, and mixing condition (4.1) holds for $(x_i^\top, \varepsilon_i)$, $i \in \mathbb{N}$, with the same constant.

Moreover, assume that density kernel $w(t)$ has finite third absolute moment, and

$$\lim_{n\to\infty} n^{-1}\operatorname{Var}\Big(\sum_{k=1}^{n} h(x_k, \varepsilon_k)\Big) = \Sigma, \tag{4.9}$$

where

$$h(u,v) := \int_{-\infty}^{\infty} t\Big(\frac{\partial g(u,\beta_0)}{\partial \beta} - \mathsf{E}\frac{\partial g(x,\beta_0)}{\partial \beta}\Big)\big(\sin(tv)\mathsf{E}\cos(t\varepsilon) - \cos(tv)\mathsf{E}\sin(t\varepsilon)\big)w(t)dt,$$

$u \in \mathbb{R}^q$, $v \in \mathbb{R}$, and Σ is a positive definite matrix. In this case, if $\widehat{\beta}_n$ is consistent then it is asymptotically normal.

Remark. Observe, that if g is twice continuously differentiable, then condition (4.3) implies (4.6). Moreover, conditions

$$\mathsf{E}\sup_{\beta\in\Theta}\Big\|\frac{\partial g(x,\beta)}{\partial \beta}\Big\|^3 < \infty \qquad \text{and} \qquad \mathsf{E}\sup_{\beta\in\Theta}\Big\|\frac{\partial^2 g(x,\beta)}{\partial \beta \partial \beta^\top}\Big\|^2 < \infty$$

imply (4.2)–(4.5).

5. Simulation results

To study the behavior of the alternative estimator (AE) $\widehat{\beta}_n$ defined by (2.2) we performed Monte Carlo simulations using the MATLAB system. Similarly to [1], as a comparison we also calculated the LSE and the LADE of β_0 that are the minimum points of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i,\beta))^2 \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n}|y_i - g(x_i,\beta)|,$$

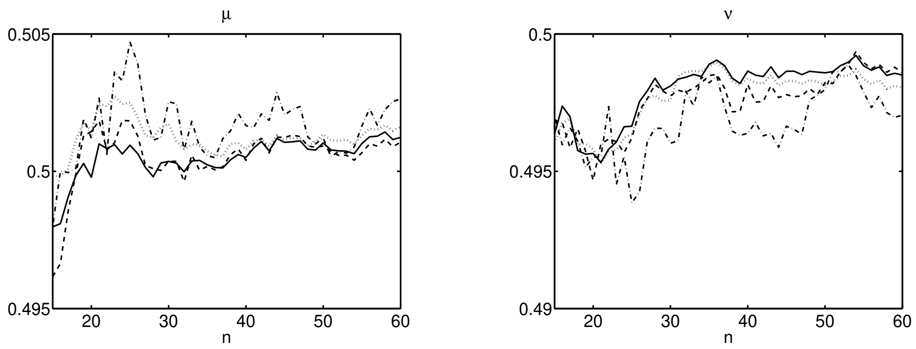
respectively.

In each of our examples 500 independent samples were simulated and using subsamples of increasing size the estimates of the parameters were calculated. For each sample size n and for each particular parameter the mean and the mean squared error (MSE) of these 500 estimates were calculated.

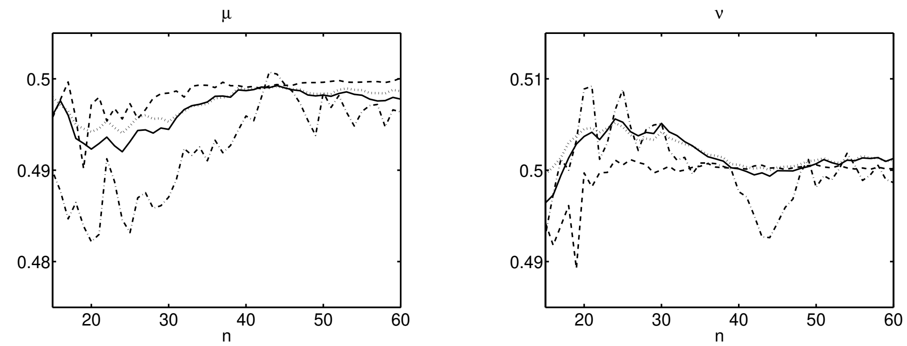
Example 5.1 Consider a special case of the model of Example 2.3

$$y = \exp(\mu x^{(1)} + \nu x^{(2)}) + \varepsilon,$$

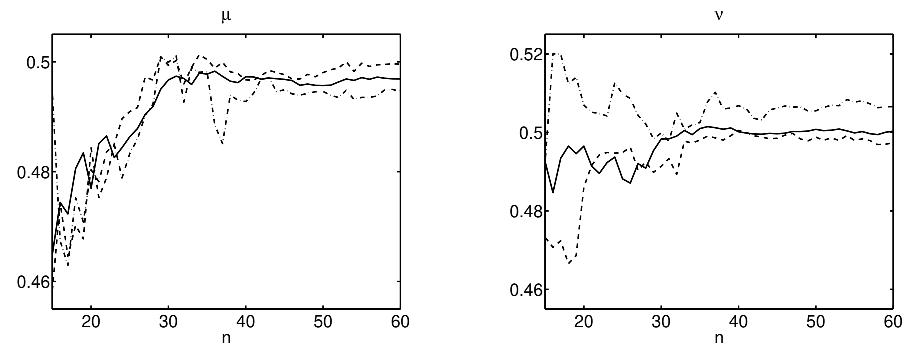
where the parameter vector $\beta = (\mu,\nu)^\top$ is in $\mathbb{R}^2$, the original parameters are $\mu_0 = \nu_0 = 0.5$ and the samples on $x = (x^{(1)}, x^{(2)})^\top$ and ε are i.i.d. The components of the



a) normal error

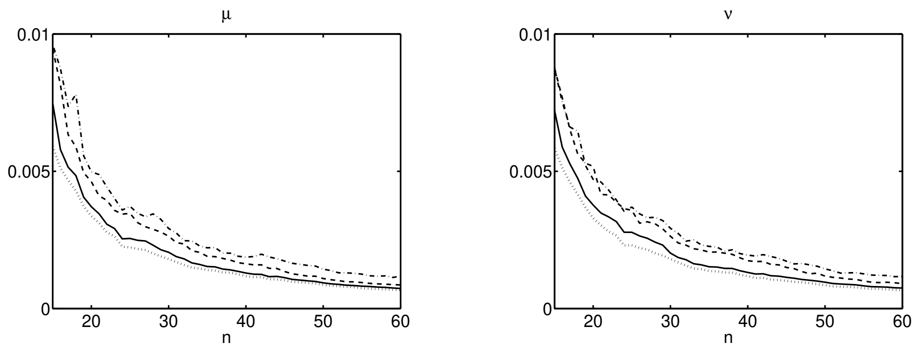


b) Bernoulli error

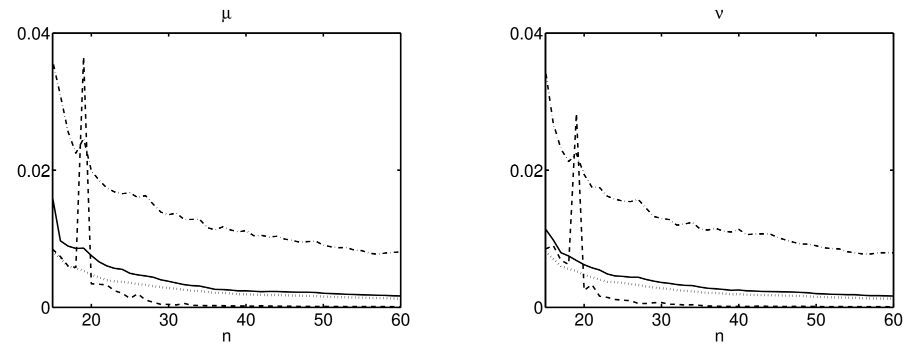


c) Cauchy error

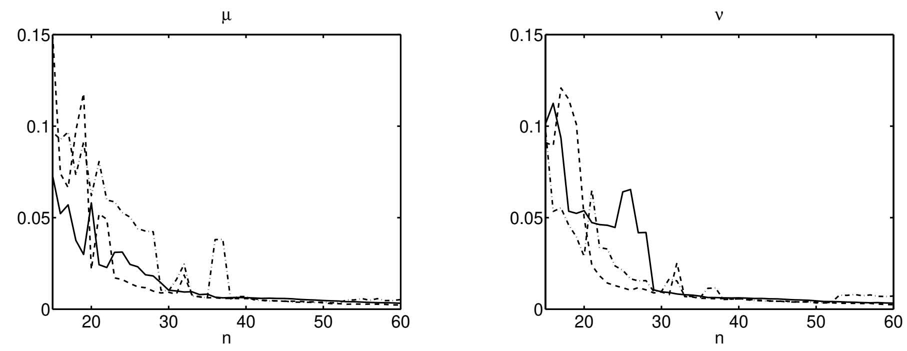
Figure 1: The means of the estimates of μ and ν in Example 5.1. Solid line: AE with normal kernel; dashed line: AE with Laplace kernel; dash-dot line: LADE; dotted line: LSE.



a) normal error



b) Bernoulli error



c) Cauchy error

Figure 2: The MSEs of the estimates of μ and ν in Example 5.1. Solid line: AE with normal kernel; dashed line: AE with Laplace kernel; dash-dot line: LADE; dotted line: LSE.

	mean		mean squared error	
n	μ	ν	μ	ν
20	0.2934	0.3049	0.9234	0.9891
40	0.2977	0.3559	0.8128	0.8259
60	0.3343	0.3482	0.7378	0.7867

Table 1: The mean and MSE of the LSE of μ and ν in Case c) of Example 5.1.

explanatory variable x are both normal with mean 2 and variance 1 and they are independent of each other. Data are taken at $1 \leq i \leq n$, where n ranges from 15 to 60.

We consider two different kernel functions $w(t)$: the density of the $\mathcal{N}(0, a^2)$ distribution with tuning parameter $a = 0.1$ and a mixture of two Laplace distributions defined by (2.3). In the latter case $a = 5$. The error term ε:

a) has normal distribution with mean 0 and variance 9;

b) has Bernoulli distribution with $\mathsf{P}(\varepsilon = -4) = \mathsf{P}(\varepsilon = 4) = \frac{1}{2}$;

c) $\varepsilon = 3\varepsilon^*$, where ε^* has Cauchy distribution with parameters $(0, 1)$.

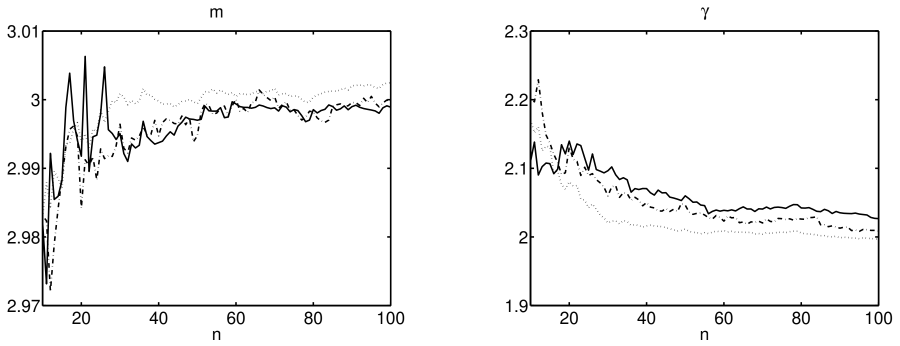
On Figure 1 the means of the estimates of the two parameters are plotted versus the sample size, while Figure 2 shows the MSEs. These figures clearly show the consistency of the AE for both kernel functions in all of the three cases. For normal and Bernoulli error terms its performance is similar to the LSE and LADE but computationally it is not so efficient. In the Cauchy error case the LSE gives completely false estimates and in this way huge MSEs, so they are not indicated on Figures 1c and 2c. On Table 1 the means and the MSEs of this estimator are given for three different sample sizes. On figures 1c and 2c one can see that in the Cauchy error case, especially for large samples, the AE is a bit better than the LADE, so it can be well applied in models where the regression error has large variance.

Example 5.2 Consider a special case of the Gaussian regression model of Example 2.5

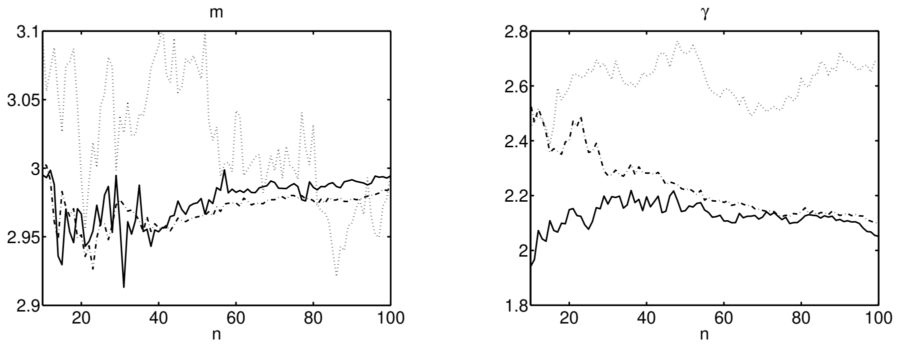
$$y = 5 \exp\big(- (x - m)^2/\gamma \big) + \varepsilon,$$

where the parameter vector $\beta = (m, \gamma)^\top$ is in $\mathbb{R}^2$ and the original parameters are $m_0 = 3,\ \gamma_0 = 2$. Data are taken at $1 \leq i \leq n$, where n ranges from 10 to 100. We consider a normal kernel function with tuning parameter $a = 2$. The explanatory variables are i.i.d normal with mean 4 and variance 1 and we have i.i.d. observations of the error term ε, too. The distribution of ε is:

a) normal with mean 0 and variance 0.5;



a) normal error



b) Cauchy error

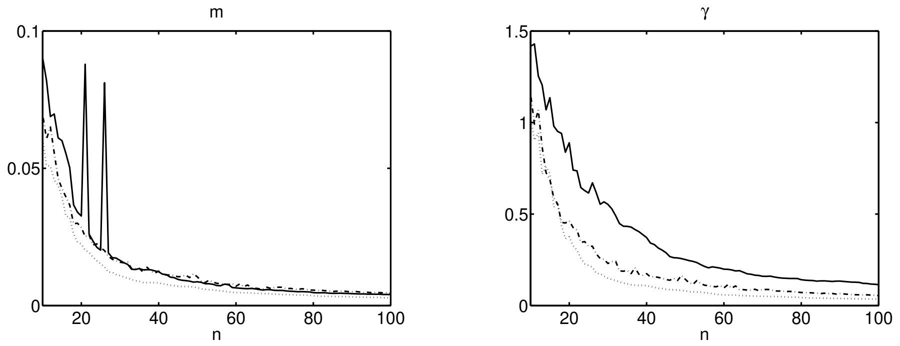
Figure 3: The means of the estimates of m and γ in Example 5.2. Solid line: AE; dash-dot line: LADE; dotted line: LSE.

b) Cauchy with parameters $(0, 1)$.

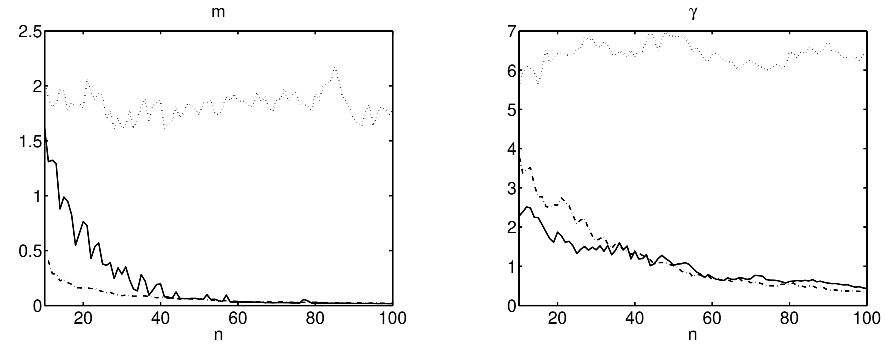
Figure 3 shows the means of the estimates of the two parameters, while on Figure 4 the MSEs are plotted versus the sample size. Similarly to Example 5.1, one can see that when the variance of the regression error is small, then the performance of the three estimators are rather similar. However, Figures 3b and 4b clearly show the advantage of the AE in the Cauchy error case.

6. Proofs

To prove Theorems 3.1 and 4.1 we need the following lemma, the proof of which can be found e.g. in [7].



a) normal error



b) Cauchy error

Figure 4: The MSEs of the estimates of m and γ in Example 5.2 .Solid line: AE; dash-dot line: LADE; dotted line: LSE.

Lemma 6.1 *Let $U_n(\beta)$ be random vectors, $n \in \mathbb{N}$, $\beta \in \Theta$, where Θ is a compact set and assume that $\lim_{n\to\infty} U_n(\beta) = 0$ in probability for each $\beta \in \Theta$. Suppose that for each $\delta > 0$*

$$\lim_{\gamma\to 0} \limsup_{n\to\infty} \mathsf{P}^*\left\{ \sup_{\|\beta_1-\beta_2\|\le\gamma} \|U_n(\beta_1) - U_n(\beta_2)\| > \delta \right\} = 0, \tag{6.1}$$

where $\mathsf{P}^*(A)$ 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 $\delta > 0$

$$\lim_{n\to\infty} \mathsf{P}^*\left\{ \sup_{\beta\in\Theta} \|U_n(\beta)\| > \delta \right\} = 0.$$

Proof of theorem 3.1 Let $U_n(\beta) := A_n(\beta) - A(\beta)$. We are going to show that $U_n(\beta) \to 0$ in probability as $n \to \infty$, uniformly in β. Let $\delta > 0$ and $\beta \in \Theta$ fixed.

$$\begin{aligned}
\mathsf{P}(|U_n(\beta)| > \delta) &\le \mathsf{P}\left(\left| \int_{-\infty}^{\infty} \left(\left(\frac{1}{n}\sum_{k=1}^{n} e^{it(y_k - g(x_k,\beta))}\right) V_n(\beta,-t)\right.\right.\right. \\
&\qquad \left.\left.\left. + \left(\mathsf{E}e^{-it(y-g(x,\beta))}\right) V_n(\beta,t)\right) w(t)dt \right| > \delta\right) \\
&\le \frac{1}{\delta}\int_{-\infty}^{\infty} \left(\mathsf{E}|V_n(\beta,-t)| + \mathsf{E}|V_n(\beta,t)|\right) w(t)dt,
\end{aligned}$$

where

$$V_n(\beta,t) := \frac{1}{n}\sum_{k=1}^{n}\left(e^{it(y_k-g(x_k,\beta))} - \mathsf{E}e^{it(y-g(x,\beta))}\right). \tag{6.2}$$

As

$$\mathsf{E}|V_n(\beta,t)| \le \left(\frac{32}{n}\sum_{k=0}^{n-1}\alpha(k)\right)^{1/2},$$

mixing condition (3.1) implies that

$$\mathsf{E}|V_n(\beta,-t)| \to 0 \qquad \text{and} \qquad \mathsf{E}|V_n(\beta,t)| \to 0$$

as $n \to \infty$. Since $|V_n(\beta,t)| \le 2$ and $w(t)$ is a density kernel function, we can apply the dominated convergence theorem. Thus, $A_n(\beta) \to A(\beta)$ in probability as $n \to \infty$.

According to Lemma 6.1, it remains to prove (6.1) for $U_n(\beta)$. For all $\delta > 0$ and $\gamma > 0$,

$$\begin{aligned}
\mathsf{P}\Big(\sup_{\|\beta_1-\beta_2\|\le\gamma} |U_n(\beta_1) - U_n(\beta_2)| > \delta\Big) &\le \frac{1}{\delta}\sup_{\|\beta_1-\beta_2\|\le\gamma} |A(\beta_1) - A(\beta_2)| \\
&\quad + \frac{1}{\delta}\mathsf{E}\sup_{\|\beta_1-\beta_2\|\le\gamma} |A_n(\beta_1) - A_n(\beta_2)|.
\end{aligned}$$

Let us deal with the first term. It is easy to see that

$$\sup_{\|\beta_1-\beta_2\|\le\gamma} |A(\beta_1) - A(\beta_2)| \le 2\sup_{\|\beta_1-\beta_2\|\le\gamma} \mathsf{E}|(g(x,\beta_1) - g(x,\beta_2)|\int_{-\infty}^{\infty}|t|w(t)dt. \tag{6.3}$$

For the second term we have

$$\begin{aligned}
&\mathsf{E}\sup_{\|\beta_1-\beta_2\|\le\gamma} |A_n(\beta_1) - A_n(\beta_2)| \qquad (6.4)\\
&\le \mathsf{E}\sup_{\|\beta_1-\beta_2\|\le\gamma} \frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\int_{-\infty}^{\infty}\left|e^{it(g(x_\ell,\beta_1)-g(x_k,\beta_1))} - e^{it(g(x_\ell,\beta_2)-g(x_k,\beta_2))}\right| w(t)dt \\
&\le 2\mathsf{E}\sup_{\|\beta_1-\beta_2\|\le\gamma} |(g(x,\beta_1) - g(x,\beta_2)|\int_{-\infty}^{\infty}|t|w(t)dt.
\end{aligned}$$

As $g(u,\beta)$ is continuous in β and Θ is compact,

$$\lim_{\gamma\to 0}\sup_{\|\beta_1-\beta_2\|\le\gamma} |(g(u,\beta_1)-g(u,\beta_2)|=0, \qquad u\in\mathbb{R}^p.$$

Due to conditions (2.1) and (3.2) we can apply the dominated convergence theorem that together with inequalities (6.3) and (6.4) implies (6.1), that completes the proof. □

Proof of the remark to Theorem 3.1 If g is differentiable and Θ is convex then using (6.3) and (6.4) we have

$$\sup_{\|\beta_1-\beta_2\|\le\gamma} |A(\beta_1)-A(\beta_2)| \le 2\gamma \sup_{\beta\in\Theta} \mathsf{E}\Big\|\frac{\partial g(x,\beta)}{\partial\beta}\Big\| \int_{-\infty}^{\infty} |t|w(t)dt,$$

$$\mathsf{E}\sup_{\|\beta_1-\beta_2\|\le\gamma} |A_n(\beta_1)-A_n(\beta_2)| \le 2\gamma \mathsf{E}\sup_{\beta\in\Theta} \Big\|\frac{\partial g(x,\beta)}{\partial\beta}\Big\| \int_{-\infty}^{\infty} |t|w(t)dt.$$

These inequalities together with condition (3.3) imply (6.1). □

The proof of Theorem 3.3 is based on the following lemma, that can be found e.g. in [5, Appendix I, Theorem 2]. The statement can also be derived from the results discussed in [12, Chapter 2].

Lemma 6.2 *Let $\mathcal{G}$ be a collection of subsets of $\mathbb{R}^m$ and for all $M>0$ let*

$$K_M := \big\{u=(u_1,\dots,u_m)^\top\in\mathbb{R}^m \mid \|u\|_\infty = \max_{1\le i\le m}|u_i|\le M\big\}.$$

Assume that for all $\delta>0$ and $D\in\mathcal{G}$ the δ-neighborhood Γ_D^δ of the boundary of $D\cap K_M$ has Lebesgue measure $\lambda(\Gamma_D^\delta)\le\psi(\delta,M)$, where $\psi(\delta,M)$ does not depend on D and for each fixed M, $\psi(\delta,M)\to 0$, as $\delta\to 0$. If ξ is an m-dimensional random vector that has an absolutely continuous distribution with respect to the m-dimensional Lebesgue measure, then $\mathcal{G}$ is a Glivenko-Cantelli class with respect to the distribution P_ξ of ξ, i.e.

$$\lim_{n\to\infty}\sup_{D\in\mathcal{G}} \big|\mathsf{P}_{\xi,n}(D)-\mathsf{P}_\xi(D)\big| = 0, \qquad a.s.$$

where $\mathsf{P}_{\xi,n}$ is the empirical measure of the i.i.d. sample $\xi_1,\xi_2,\dots,\xi_n$ drawn from the distribution P_ξ.

Proof of Theorem 3.3 The proof of this theorem is based on the ideas given in [1] and [3]. Let $F_\beta(u)$ denote the distribution function of $y-g(x,\beta)$, where

$y = g(x, \beta_0) + \varepsilon$ and denote by $\widehat{F}_{\beta,n}(u)$ the empirical distribution function of the sample $y_i - g(x_i, \beta),\ \ i = 1, 2, \ldots, n$. It is easy to see, that

$$A(\beta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi_w(u - v)dF_\beta(u)dF_\beta(v),$$
$$\widetilde{A}_n(\beta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi_w(u - v)d\widehat{F}_{\beta,n}(u)d\widehat{F}_{\beta,n}(v),$$

where $\varphi_w(u)$ denotes the Fourier transform of the density kernel $w(t)$. Let

$$F^\circ_\beta(v) := \int_{-\infty}^{\infty} F_\beta(v + u)dF_\beta(u), \qquad \text{and} \qquad \widehat{F}^\circ_{\beta,n}(v) := \int_{-\infty}^{\infty} \widehat{F}_{\beta,n}(v + u)d\widehat{F}_{\beta,n}(u).$$

Then

$$A(\beta) = \int_{-\infty}^{\infty} \varphi_w(v)dF^\circ_\beta(v) \qquad \text{and} \qquad \widetilde{A}_n(\beta) = \int_{-\infty}^{\infty} \varphi_w(v)d\widehat{F}^\circ_{\beta,n}(v).$$

Due to condition (3.4) we can use integration by parts, so using the Riemann-Lebesgue lemma (see e.g. [10, Theorem 7.5]) we have

$$A(\beta) = -\int_{-\infty}^{\infty} F^\circ_\beta(v)\frac{d\varphi_w(v)}{dv}dv \quad \text{and} \quad \widetilde{A}_n(\beta) = -\int_{-\infty}^{\infty} \widehat{F}^\circ_{\beta,n}(v)\frac{d\varphi_w(v)}{dv}dv.$$

Thus,

$$\sup_\beta |\widetilde{A}_n(\beta) - A(\beta)| \le \sup_{v,\beta} |\widehat{F}^\circ_{\beta,n}(v) - F^\circ_\beta(v)| \int_{-\infty}^{\infty} \left|\frac{d\varphi_w(v)}{dv}\right| dv,$$

and it is easy to see that

$$\sup_{v,\beta} |\widehat{F}^\circ_{\beta,n}(v) - F^\circ_\beta(v)| \le 2 \sup_{v,\beta} |\widehat{F}_{\beta,n}(v) - F_\beta(v)|.$$

Hence, condition (3.4) implies that to show the almost sure convergence of $\widehat{\beta}_n$ to β_0 it is enough to prove

$$\lim_{n\to\infty} \sup_{v,\beta} |\widehat{F}_{\beta,n}(v) - F_\beta(v)| = 0, \qquad \text{a.s.} \tag{6.5}$$

Consider the vector $\xi = (y, x^\top)$. For a given $v \in \mathbb{R}$ and $\beta \in \Theta$ the event $\{y - g(x, \beta) < v\}$ is equivalent to the event $\{\xi \in D(v, \beta)\}$, where $D(v, \beta)$ is the set defined by (3.5). Thus, to verify (6.5) we have to show that the collection of sets

$$\mathcal{G} = \big\{D(v, \beta) \,\big|\, v \in \mathbb{R},\ \beta \in \Theta\big\}$$

forms a Glivenko-Cantelli class with respect to the distribution of ξ. To do it we make use of Lemma 6.2. Let K_M be defined as in Lemma 6.2 with $m = q + 1$. For a given $v \in \mathbb{R}$ and $\beta \in \Theta$ the surface area of $D(v, \beta) \cap K_M$ is less than or equal to the sum of the surfaces of the sides of the hypercube K_M, which equals $2(q+1)(2M)^q$, plus the area of the surface

$$\big\{(z, u^\top) \in \mathbb{R}^{q+1} \,\big|\, z - g(u, \beta) = v,\ |z| \le M,\ |u_i| \le M,\ i = 1, \dots, q\big\}.$$

The latter quantity does not depend on v and it is equal to

$$\int\limits_{\|u\|_\infty \le M} \sqrt{1 + \|\nabla g(u, \beta)\|^2} du \le \int\limits_{\|u\|_\infty \le M} \sqrt{1 + \sup_{\beta \in \Theta} \|\nabla g(u, \beta)\|^2} du =: \rho(M).$$

The existence of such a function ρ is ensured by the compactness of Θ and continuity of $\frac{\partial g(u,\beta)}{\partial u}$ in β and u. Thus, the Lebesgue measure of the δ-neighborhood of the boundary of $D(v, \beta) \cap K_M$ is less than or equal to $\psi(\delta, M) := 2\delta\big(2(q+1)(2M)^q + \rho(M)\big)$. This means that conditions of Lemma 6.2 are satisfied, that completes the proof. $\square$

In the proof of Theorem 4.1 we use a CLT of Herrndorf [8, Corollary 1] given below.

Lemma 6.3 *Let η_i, $i \in \mathbb{N}$, be a centered vector valued stochastic process (i.e. $\mathsf{E}\eta_i = 0$, $i \in \mathbb{N}$), $S_n = \sum_{i=1}^n \eta_i$ and let Σ_n denote the variance matrix of S_n. Assume that mixing condition (4.1) is satisfied for some $\nu > 0$ and with the same ν*

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

Then

$$\limsup_n \frac{1}{n} \sum_{i,j=1}^n \|\operatorname{Cov}(\eta_i, \eta_j)\| < \infty.$$

If, moreover,

$$\lim_{n \to \infty} \frac{1}{n} \Sigma_n = \Sigma,$$

where Σ is a positive definite matrix, then

$$(\Sigma_n)^{-1/2} S_n \Rightarrow \mathcal{N}(0, I)$$

as $n \to \infty$.

Besides this we need the following form of the Rosenthal inequality for mixing processes (see e.g. [6]).

Lemma 6.4 *Let η_i, $i \in \mathbb{N}$, be a centered stochastic process and let $\nu > 0$. Then for all $1 \le t \le 2$,*

$$\mathsf{E}\left\|\sum_{i=1}^{n} \eta_i\right\|^t \le cb(\alpha, \nu)\sum_{i=1}^{n}(\|\eta_i\|_{t+\nu})^t. \tag{6.6}$$

PROOF OF THEOREM 4.1 To prove the theorem we apply Taylor series expansion and Lemmata 6.1 – 6.4. Since β_0 is an interior point of Θ and $\widehat{\beta}_n$ is a consistent estimator of β_0, for n large enough $\widehat{\beta}_n$ is in the interior of Θ with great probability. As $\widehat{\beta}_n$ is the maximum point of $A_n(\beta)$,

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

By Taylor series expansion around β_0 we get

$$0 = \frac{\partial A_n(\beta_0)}{\partial \beta} + \frac{\partial^2 A_n(\widetilde{\beta}_n)}{\partial \beta \partial \beta^\top}(\widehat{\beta}_n - \beta_0), \tag{6.7}$$

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 notation we do not indicate it. Obviously, if $\widehat{\beta}_n$ converges to β_0 in probability, so does $\widetilde{\beta}_n$. Equation (6.7) can be rewritten in the following way

$$\sqrt{n}\frac{\partial A_n(\beta_0)}{\partial \beta} = -\frac{\partial^2 A_n(\widetilde{\beta}_n)}{\partial \beta \partial \beta^\top}\sqrt{n}(\widehat{\beta}_n - \beta_0). \tag{6.8}$$

First we prove that the left hand side of (6.8) has a normal limit distribution. It is easy to see that

$$\begin{aligned}
\frac{\partial A_n(\beta_0)}{\partial \beta} &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\int_{-\infty}^{\infty} it\Big(\frac{\partial g(x_\ell, \beta_0)}{\partial \beta} - \frac{\partial g(x_k, \beta_0)}{\partial \beta}\Big)e^{it(\varepsilon_k - \varepsilon_\ell)}w(t)dt \\
&= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n} G(x_k, \varepsilon_k, x_\ell, \varepsilon_\ell),
\end{aligned}$$

where

$$G(u_1, v_1, u_2, v_2) := \int_{-\infty}^{\infty} t\Big(\frac{\partial g(u_1, \beta_0)}{\partial \beta} - \frac{\partial g(u_2, \beta_0)}{\partial \beta}\Big)\sin\big(t(v_1 - v_2)\big)w(t)dt,$$

$u_1, u_2 \in \mathbb{R}^q$, $v_1, v_2 \in \mathbb{R}$. Short calculation shows that

$$\frac{\partial A_n(\beta_0)}{\partial \beta} = U_n + R_n.$$

where

$$U_n := \frac{2}{n}\sum_{k=1}^{n} h(x_k,\varepsilon_k),$$

$$R_n := \frac{2}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\int_{-\infty}^{\infty} t\big(g_1(x_k,\varepsilon_k,\varepsilon_\ell) - g_2(x_k,\varepsilon_k,\varepsilon_\ell)\big)w(t)dt,$$

and

$$g_1(u,v_1,v_2) = \Big(\frac{\partial g(u,\beta_0)}{\partial\beta} - \mathsf{E}\frac{\partial g(x,\beta_0)}{\partial\beta}\Big)\sin(tv_1)(\cos(tv_2) - \mathsf{E}\cos(t\varepsilon)),$$

$$g_2(u,v_1,v_2) = \Big(\frac{\partial g(u,\beta_0)}{\partial\beta} - \mathsf{E}\frac{\partial g(x,\beta_0)}{\partial\beta}\Big)\cos(tv_1)(\sin(tv_2) - \mathsf{E}\sin(t\varepsilon)),$$

$u \in \mathbb{R}^q,\ v_1, v_2 \in \mathbb{R}$. Obviously,

$$\begin{aligned}
&\mathsf{E}\Big\|\frac{1}{n^{3/2}}\sum_{k=1}^{n}\sum_{\ell=1}^{n} g_1(x_k,\varepsilon_k,\varepsilon_\ell)\Big\| \\
&\quad\le \mathsf{E}\bigg(\Big\|\frac{1}{n^{1/2}}\sum_{k=1}^{n}\Big(\frac{\partial g(u,\beta_0)}{\partial\beta} - \mathsf{E}\frac{\partial g(x,\beta_0)}{\partial\beta}\Big)\sin(t\varepsilon_k)\Big\|\Big|\frac{1}{n}\sum_{\ell=1}^{n}(\cos(t\varepsilon_\ell) - \mathsf{E}\cos(t\varepsilon))\Big|\bigg) \\
&\quad\le \Big\|\frac{1}{n^{1/2}}\sum_{k=1}^{n}\Big(\frac{\partial g(u,\beta_0)}{\partial\beta} - \mathsf{E}\frac{\partial g(x,\beta_0)}{\partial\beta}\Big)\sin(t\varepsilon_k)\Big\|_2\Big\|\frac{1}{n}\sum_{\ell=1}^{n}(\cos(t\varepsilon_\ell) - \mathsf{E}\cos(t\varepsilon))\Big\|_2.
\end{aligned}$$

Using inequality (6.6) and condition (4.8) we obtain

$$\begin{aligned}
&\mathsf{E}\Big\|\frac{1}{n^{1/2}}\sum_{k=1}^{n}\Big(\frac{\partial g(u,\beta_0)}{\partial\beta} - \mathsf{E}\frac{\partial g(x,\beta_0)}{\partial\beta}\Big)\sin(t\varepsilon_k)\Big\|^2 \\
&\qquad\qquad \le \frac{c}{n}\sum_{k=1}^{n}\bigg(\mathbb{E}\Big\|\frac{\partial g(u,\beta_0)}{\partial\beta} - \mathsf{E}\frac{\partial g(x,\beta_0)}{\partial\beta}\Big\|^{2+\nu}|\sin(t\varepsilon_k)|^{2+\nu}\bigg)^{\frac{2}{2+\nu}} \le c, \\
&\mathsf{E}\Big|\frac{1}{n}\sum_{\ell=1}^{n}(\cos(t\varepsilon_\ell) - \mathsf{E}\cos(t\varepsilon))\Big|^2 \le \frac{c}{n^2}\sum_{\ell=1}^{n}\Big(\mathbb{E}|(\cos(t\varepsilon_\ell) - \mathsf{E}\cos(t\varepsilon))|^{2+\nu}\Big)^{\frac{2}{2+\nu}} \le \frac{c}{n},
\end{aligned}$$

and the second term is less than c and converges to 0 as $n \to \infty$. Similar result can be proved for the other term of R_n. Due to condition (2.1) on the density kernel $w(t)$ we can apply the dominated convergence theorem, thus $\lim_{n\to\infty}\sqrt{n}R_n = 0$, in probability.

Conditions (2.1), (4.8) and (4.9) ensure that we can apply Lemma 6.3, so $\sqrt{n}U_n$ and in this way the left hand side of (6.8) has a normal limit distribution.

Now, let us deal with the right hand side of (6.8). First we show that

$$\frac{\partial^2 A_n(\beta)}{\partial\beta\partial\beta^\top} \to \frac{\partial^2 A(\beta)}{\partial\beta\partial\beta^\top} \tag{6.9}$$

in probability as $n\to\infty$, uniformly in β. In order to do this we make use of Lemma 6.1. Let $r,s\in\{1,2,\dots,p\}$. An easy calculation shows that

$$\frac{\partial^2 A_n(\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}} - \frac{\partial^2 A(\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}} = 2U_1^{(r,s)}(\beta) + 2U_2^{(r,s)}(\beta) - 2U_3^{(r,s)}(\beta),$$

where

$$\begin{aligned}
U_1^{(r,s)}(\beta) &:= \int_{-\infty}^{\infty} it\Bigg(\Big(\frac{1}{n}\sum_{k=1}^n \frac{\partial^2 g(x_k,\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}} e^{it(y_k-g(x_k,\beta))}\Big)\Big(\frac{1}{n}\sum_{\ell=1}^n e^{-it(y_\ell-g(x_\ell,\beta))}\Big)\\
&\qquad - \mathsf{E}\frac{\partial^2 g(x,\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}} e^{it(y-g(x,\beta))}\mathsf{E}e^{-it(y-g(x,\beta))}\Bigg)w(t)dt,\\
U_2^{(r,s)}(\beta) &:= \int_{-\infty}^{\infty} t^2\Bigg(\Big(\frac{1}{n}\sum_{k=1}^n \frac{\partial g(x_k,\beta)}{\partial\beta^{(r)}} e^{it(y_k-g(x_k,\beta))}\Big)\Big(\frac{1}{n}\sum_{\ell=1}^n \frac{\partial g(x_\ell,\beta)}{\partial\beta^{(s)}} e^{-it(y_\ell-g(x_\ell,\beta))}\Big)\\
&\qquad - \mathsf{E}\frac{\partial g(x,\beta)}{\partial\beta^{(r)}} e^{it(y-g(x,\beta))}\mathsf{E}\frac{\partial g(x,\beta)}{\partial\beta^{(s)}}e^{-it(y-g(x,\beta))}\Bigg)w(t)dt,\\
U_3^{(r,s)}(\beta) &:= \int_{-\infty}^{\infty} t^2\Bigg(\Big(\frac{1}{n}\sum_{k=1}^n \frac{\partial g(x_k,\beta)}{\partial\beta^{(r)}}\frac{\partial g(x_k,\beta)}{\partial\beta^{(s)}} e^{it(y_k-g(x_k,\beta))}\Big)\Big(\frac{1}{n}\sum_{\ell=1}^n e^{-it(y_\ell-g(x_\ell,\beta))}\Big)\\
&\qquad - \mathsf{E}\frac{\partial g(x,\beta)}{\partial\beta^{(r)}}\frac{\partial g(x,\beta)}{\partial\beta^{(s)}} e^{it(y-g(x,\beta))}\mathsf{E}e^{-it(y-g(x,\beta))}\Bigg)w(t)dt.
\end{aligned}$$

For all $r,s\in\{1,2,\dots,p\}$ let

$$\begin{aligned}
W_n^{(r,s)}(\beta,t) &:= \frac{1}{n}\sum_{k=1}^n\Big(\frac{\partial^2 g(x_k,\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}}e^{it(y_k-g(x_k,\beta))} - \mathsf{E}\frac{\partial^2 g(x,\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}}e^{it(y-g(x,\beta))}\Big),\\
Q_n^{(r,s)}(\beta,t) &:= \frac{1}{n}\sum_{k=1}^n\Big(\frac{\partial g(x_k,\beta)}{\partial\beta^{(r)}}\frac{\partial g(x_k,\beta)}{\partial\beta^{(s)}}e^{it(y_k-g(x_k,\beta))}\\
&\qquad - \mathsf{E}\frac{\partial g(x,\beta)}{\partial\beta^{(r)}}\frac{\partial g(x_k,\beta)}{\partial\beta^{(s)}}e^{it(y-g(x,\beta))}\Big),\\
Z_n^{(r)}(\beta,t) &:= \frac{1}{n}\sum_{k=1}^n\Big(\frac{\partial g(x_k,\beta)}{\partial\beta^{(r)}}e^{it(y_k-g(x_k,\beta))} - \mathsf{E}\frac{\partial g(x,\beta)}{\partial\beta^{(r)}}e^{it(y-g(x,\beta))}\Big).
\end{aligned}$$

It is not difficult to see, that for a positive constant δ

$$
\begin{aligned}
\mathsf{P}\big(\big|U_1^{(r,s)}(\beta)\big| \geq \delta\big) &\leq \frac{1}{\delta}\int_{-\infty}^{\infty} |t| \bigg(\mathsf{E}\big|W_n^{(r,s)}(\beta,t)\big| + \mathsf{E}\Big|\frac{\partial^2 g(x,\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}}\Big|\mathsf{E}\big|V_n(\beta,-t)\big|\bigg) w(t)dt,\\
\mathsf{P}\big(\big|U_2^{(r,s)}(\beta)\big| \geq \delta\big) &\leq \frac{1}{\delta}\int_{-\infty}^{\infty} t^2 \bigg(\mathsf{E}\Big(\big|Z_n^{(r)}(\beta,t)\big|\Big|\frac{1}{n}\sum_{\ell=1}^{n}\frac{\partial g(x_\ell,\beta)}{\partial\beta^{(s)}}e^{-it(y_\ell - g(x_\ell,\beta))}\Big|\Big)\\
&\qquad + \mathsf{E}\Big|\frac{\partial g(x,\beta)}{\partial\beta^{(r)}}\Big|\mathsf{E}\big|Z_n^{(s)}(\beta,-t)\big|\bigg) w(t)dt\\
&\leq \frac{1}{\delta}\int_{-\infty}^{\infty} t^2 \bigg(\Big\|\frac{\partial g(x,\beta)}{\partial\beta^{(s)}}\Big\|_2 \big\|Z_n^{(r)}(\beta,t)\big\|_2\\
&\qquad + \big\|Z_n^{(s)}(\beta,-t)\big\|_2 \Big\|\frac{\partial g(x,\beta)}{\partial\beta^{(r)}}\Big\|_2\bigg) w(t)dt,\\
\mathsf{P}\big(\big|U_3^{(r,s)}(\beta)\big| \geq \delta\big) &\leq \frac{1}{\delta}\int_{-\infty}^{\infty} t^2 \bigg(\mathsf{E}\big|Q_n^{(r,s)}(\beta,t)\big|\\
&\qquad + \mathsf{E}\Big|\frac{\partial g(x,\beta)}{\partial\beta^{(r)}}\Big|\Big|\frac{\partial g(x,\beta)}{\partial\beta^{(s)}}\Big|\mathsf{E}\big|V_n(\beta,-t)\big|\bigg) w(t)dt,
\end{aligned}
$$

where $V_n(\beta,t)$ is defined by (6.2).

Let $\beta \in \Theta$ fixed. As $w(t)$ has a finite third absolute moment and conditions (4.2) and (4.3) are satisfied, to prove that

$$
\lim_{n\to\infty} \frac{\partial^2 A_n(\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}} = \frac{\partial^2 A(\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}}, \qquad r,s \in \{1,2,\ldots,p\}, \tag{6.10}
$$

in probability, it is enough to show

$$
\begin{aligned}
&\lim_{n\to\infty} \mathsf{E}\big|W_n^{(r,s)}(\beta,t)\big| = 0, && (6.11)\\
&\lim_{n\to\infty} \mathsf{E}\big|Q_n^{(r,s)}(\beta,t)\big| = 0, && (6.12)\\
&\lim_{n\to\infty} \mathsf{E}\big|Z_n^{(r)}(\beta,t)\big|^2 = 0, && (6.13)\\
&\lim_{n\to\infty} \mathsf{E}\big|V_n(\beta,t)\big| = 0, \qquad t \in \mathbb{R},\ r,s \in \{1,2,\ldots,p\}. && (6.14)
\end{aligned}
$$

Obviously,

$$
\mathsf{E}\big|W_n^{(r,s)}(\beta,t)\big| \leq \big\|W_n^{(r,s)}(\beta,t)\big\|_{1+\nu}.
$$

Using inequality (6.6) we obtain

$$
\mathsf{E}\big|W_n^{(r,s)}(\beta,t)\big|^{1+\nu} \leq \frac{c}{n^{1+\nu}}\sum_{k=1}^{n}\left(\mathsf{E}\Big|\frac{\partial^2 g(x_k,\beta)}{\partial\beta^{(r)}\partial\beta^{(s)}}\Big|^{1+2\nu}\right)^{\frac{1+\nu}{1+2\nu}}
$$

that together with (4.7) implies (6.11). In a similar way (using (4.8) instead of (4.7) when it is necessary) we can verify (6.12)–(6.14).

It remains to check condition (6.1) of Lemma 6.1 for functions $U_1^{(r,s)}(\beta)$, $U_2^{(r,s)}(\beta)$ and $U_3^{(r,s)}(\beta)$, $r,s\in\{1,2,\ldots,p\}$. Consider e.g. $U_2^{(r,s)}(\beta)$. For all $\delta>0$,

$$\mathsf{P}\Big(\sup_{\|\beta_1-\beta_2\|\le\gamma}\big|U_2^{(r,s)}(\beta_1)-U_2^{(r,s)}(\beta_2)\big|\ge\delta\Big)\le\frac{1}{\delta}\big(C_n^{(r,s)}(\gamma)+C_n^{(s,r)}(\gamma)+D^{(r,s)}(\gamma)+D^{(s,r)}(\gamma)\big),$$

where

$$\begin{aligned}C_n^{(r,s)}(\gamma):=\mathsf{E}\sup_{\|\beta_1-\beta_2\|\le\gamma}\int_{-\infty}^{\infty}t^2\bigg(\Big|\frac{\partial g(x,\beta_1)}{\partial\beta^{(r)}}e^{-itg(x_k,\beta_1)}-\frac{\partial g(x,\beta_2)}{\partial\beta^{(r)}}e^{-itg(x,\beta_2)}\Big|\\ \times\Big|\frac{\partial g(x,\beta_1)}{\partial\beta^{(s)}}\Big|\bigg)w(t)dt,\\ D^{(r,s)}(\gamma):=\sup_{\|\beta_1-\beta_2\|\le\gamma}\int_{-\infty}^{\infty}t^2\,\mathsf{E}\Big|\frac{\partial g(x,\beta_1)}{\partial\beta^{(r)}}e^{-itg(x,\beta_1)}-\frac{\partial g(x,\beta_2)}{\partial\beta^{(r)}}e^{-itg(x,\beta_2)}\Big|\\ \times\mathsf{E}\Big|\frac{\partial g(x,\beta_1)}{\partial\beta^{(s)}}\Big|w(t)dt.\end{aligned}$$

It is easy to see, that

$$\begin{aligned}C_n^{(r,s)}(\gamma)\le\mathsf{E}\sup_{\|\beta_1-\beta_2\|\le\gamma}\Big|\frac{\partial g(x,\beta_1)}{\partial\beta^{(s)}}\Big|\Big|\frac{\partial g(x,\beta_1)}{\partial\beta^{(r)}}-\frac{\partial g(x,\beta_2)}{\partial\beta^{(r)}}\Big|\int_{-\infty}^{\infty}t^2w(t)dt\\ +\mathsf{E}\sup_{\|\beta_1-\beta_2\|\le\gamma}\Big|\frac{\partial g(x,\beta_1)}{\partial\beta^{(r)}}\Big|\Big|\frac{\partial g(x,\beta_1)}{\partial\beta^{(s)}}\Big|\big|g(x,\beta_1)-g(x,\beta_2)\big|\int_{-\infty}^{\infty}|t|^3w(t)dt.\end{aligned}\tag{6.15}$$

Moment condition on $w(t)$ and (4.4) immediately imply that the second term of the right hand side of (6.15) converges to zero as $\gamma\to0$. As g is twice differentiable with respect to β and Θ is compact, $\frac{\partial g(u,\beta)}{\partial\beta^{(r)}}$, $u\in\mathbb{R}^p$ is uniformly continuous on Θ. Thus,

$$\lim_{\gamma\to0}\sup_{\|\beta_1-\beta_2\|\le\gamma}\Big|\frac{\partial g(u,\beta_1)}{\partial\beta^{(s)}}\Big|\Big|\frac{\partial g(u,\beta_1)}{\partial\beta^{(r)}}-\frac{\partial g(u,\beta_2)}{\partial\beta^{(r)}}\Big|=0.$$

Due to condition (4.2) we can apply the dominated convergence theorem, so the first term of the right hand side of (6.15) also converges to zero as $\gamma\to0$. Thus,

$$\lim_{\gamma\to0}\limsup_{n\to\infty}C_n^{(r,s)}(\gamma)=0,\qquad r,s\in\{1,2,\ldots,p\}.$$

For $D^{(r,s)}(\gamma)$ we have

$$\begin{aligned}D^{(r,s)}(\gamma)\le\gamma\sup_{\beta\in\Theta}\mathsf{E}\Big\|\frac{\partial g(x,\beta)}{\partial\beta}\Big\|\sup_{\beta\in\Theta}\mathsf{E}\Big\|\frac{\partial^2 g(x,\beta)}{\partial\beta\partial\beta^\top}\Big\|\int_{-\infty}^{\infty}t^2w(t)dt\\ +\gamma\sup_{\beta\in\Theta}\mathsf{E}\Big\|\frac{\partial g(x,\beta)}{\partial\beta}\Big\|\sup_{\beta\in\Theta}\mathsf{E}\Big\|\frac{\partial g(x,\beta)}{\partial\beta}\Big\|^2\int_{-\infty}^{\infty}|t|^3w(t)dt,\end{aligned}$$

that together with moment condition on $w(t)$ and conditions (4.2) and (4.3) implies

$$\lim_{\gamma\to 0} D^{(r,s)}(\gamma) = 0, \qquad r,s \in \{1,2,\dots,p\}.$$

This means that (6.1) is satisfied for $U_2^{(r,s)}(\beta)$. In a similar way (using also conditions (4.5) and (4.6)) we can show that (6.1) is satisfied for $U_1^{(r,s)}(\beta)$ and $U_3^{(r,s)}(\beta)$, $r,s \in \{1,2,\dots,p\}$, too. Hence, as $\lim_{n\to\infty} \widetilde{\beta}_n = \beta_0$ in probability, from Lemma 6.1 we have

$$\lim_{n\to\infty} \frac{\partial^2 A_n(\widetilde{\beta}_n)}{\partial\beta\partial\beta^\top} = \frac{\partial^2 A(\beta_0)}{\partial\beta\partial\beta^\top} = -2\,\mathrm{Var}\left(\frac{\partial g(x,\beta_0)}{\partial\beta}\right)\int_{-\infty}^{\infty} t^2\varphi_\varepsilon(-t)\varphi_\varepsilon(t)w(t)dt \tag{6.16}$$

in probability, where $\varphi_\varepsilon(t)$ denotes the characteristic function of ε. According to the conditions of the theorem the right hand side of (6.16) is a regular matrix, that completes the proof. $\square$

References

[1] An, H-Z., Hickernell, F. J. and Zhu, L-X., A new class of consistent estimators for stochastic linear regressive models. *J. Multivariate Anal.* **63** (1997), no. 2, 242–258.

[2] Baran, S., A consistent estimator in general functional errors-in-variables models. *Metrika* **51** (2000), no. 2, 117–132.

[3] Baran, S., A new consistent estimator for linear errors-in-variables models. *Comput. Math. Appl.* **41** (2001), no. 7-8, 821–833.

[4] Baran, S., A consistent estimator for linear models with dependent observations. Institute of Mathematics and Informatics, University of Debrecen, *Preprint* No. 286 (*Technical Report* No. 2002/10) (submitted to *Comm. Statist. Theory Methods*).

[5] Borovkov, A. A., *Mathematical Statistics.* Nauka, Moscow, 1984 (in Russian).

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

[7] Fazekas, I., Kukush, A. G., Asymptotic properties of an estimator in nonlinear functional errors-in-variables models with dependent error terms. *Comput. Math. Appl.* **34** (1997), no. 10, 23–39.

[8] Herrndorf, N., A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probability* **12** (1984), no. 1, 141–153.

[9] Pollard, D., *Convergence of Stochastic Processes*, Springer, New York, 1984.

[10] Rudin, W., *Functional Analysis.* McGraw-Hill, New York, 1973.

[11] Stefanski, L. A., Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Comm. Statist. Theory Method* **18** (1989), no. 12, 4335–4358.

[12] van der Vaart, A. W., Wellner, J. A., *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer, New York, 1996.