L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

On L_1-norm estimators in nonlinear regression and in nonlinear error-in-variables models

Silvelyn Zwanzig

University of Hamburg, Germany

Abstract: Strong consistency results for the L_1-norm estimator of the parameter of interest are shown in nonlinear models, namely in the nonlinear regression model, in the nonlinear error-in-variables model and in a nonlinear semiparametric model.

Key words: Nonlinear semiparametric regression, nonlinear functional relation, minimum contrast estimates, L_1-norm, consistency.

AMS subject classification: 62F12.

1 Introduction

In regression models the L_1-norm estimators receive their justification from robustness theory. For the error distribution only assumptions on the behavior around the median, which should be zero, are required for consistency results.

For nonlinear regression the consistence of the L_1-norm estimators is shown by Oberhofer (1982). Richardson and Bhattacharyya (1987) extended this result to general noncompact parameter sets by using a sieves method. The general approach of M-estimators in Liese and Vajda (1994) for nonlinear regression models includes also the L_1-norm estimator. They obtained a similar result as Richardson and Bhattacharyya (1987) with a different method and conditions that are statistically more transparent.

The concept of minimum contrast estimators and the method of sieves are studied in the nonparametric theory by van de Geer (1990), van de Geer (1995), Birge and Massart (1991) and Birge and Massart (1994). They

apply their general results for consistency rates of minimum contrast estimators to the L_1-norm estimator of the nonparametric regression function. (Compare also the Section 3.4.4. in the recent book of van der Vaart and Wellner, 1996).

The aim of this paper is to give consistency results for L_1-norm estimator in semiparametric regression models, where the parameter of interest is of fixed dimension and the nuisance parameter is either given as an unknown function from a nonparametric function space or has a dimension which increases with the sample size. In this paper we will study simultaneously three models, namely nonlinear regression, nonlinear functional relation, nonlinear semiparametric regression, all to be introduced in Section 2. This allows us to demonstrate the line of proving the L_1-norm consistency results and to emphasize the underlying problems. For illustration reasons the L_1-approach is also embedded in the minimum contrast context, see Section 3. The proof consists mainly of two steps given in Section 4: first the approximation of the empirical L_1-norm by its expected value, which will be done in Subsection 4.1, and second the identification of the parameter by using the expected values of the difference of empirical L_1-norms in Subsection 4.2. The identification problem is specific for the individual models. This is not characteristic for the L_1-norm approach, since the same problem also occurs in the L_2-theory. The technique of approximation is more or less standard and is based on results of the increments of sub-Gaussian processes. In Section 5 the consistency results for each model are separately given. For the nonlinear regression model this is a known strong consistency result in form of an exponential probability inequality. The consistency result for the nonlinear functional relation models is new. Under this kind of entropy condition on the nuisance parameter space only results are known for the least squares estimator, see Zwanzig (1990). The nonlinear semiparametric model is of a special structure, because of the constrains imposed by the identification problem. Linton (1995) studied this model with a linear parametric part. The result given here for this model seems to be new as well.

2 The general setting

Here we introduce a general semiparametric regression model to be specified in the following. Suppose we have independent and in general not identically distributed two dimensional real valued observations $(y_1, x_1), ..., (y_n, x_n)$, generated by

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

$$x_i = h(\xi_i, \beta) + \varepsilon_{2i}, \tag{2}$$

with $i = 1, ..., n$. The probability of each observation (y_i, x_i) is $P_{\xi_i\beta}$. The common distribution of the whole sample is denoted by $P_{\xi\beta} = \prod_{i=1}^{n} P_{\xi_i\beta}$ and dominated by a σ-finite measure μ_n.

The errors ε_{ji} are i.i.d. with distribution P_ε, expected value zero and positive variances σ^2. The error distribution does not depend of the parameter β. To get the consistency of L_1-norm estimators we will need the assumption **E** on the error distribution that the median is zero and that the distribution has enough mass in the local neighborhood of it:

E $\exists D_0 \, \exists \kappa_\varepsilon \, \forall d \leq D_0$ such that

$$P_\varepsilon(-d \leq \varepsilon \leq 0) \geq \kappa_\varepsilon d \quad , \quad P_\varepsilon(0 \leq \varepsilon \leq d) \geq \kappa_\varepsilon d. \tag{3}$$

$$P_\varepsilon(\varepsilon \geq 0) = \frac{1}{2}$$

The functions $h(.,.)$ and $g(.,.)$ are continuous and known. The regression parameter $\beta \in \mathcal{B} \subset \mathrm{R}^p$ is the parameter of interest. The dimension p of β does not depend on the sample size n. The design points or variables $\{\xi_1, ..., \xi_n\} \subset \mathrm{R}$ are unknown and fixed. They are the nuisance parameters, whose number grows with the sample size n. We write the nuisance parameters as components of a column vector of dimension n: $\xi = \xi^{(n)} = (\xi_1, ..., \xi_n)^T \in \mathcal{X}^{(n)} \subseteq \mathrm{R}^n$. The common unknown parameter is

$$\theta = (\xi, \beta) \in \Theta = \mathcal{X}^{(n)} \times \mathcal{B} \subseteq \mathrm{R}^{n+p}. \tag{4}$$

The model assumptions (1), (2) above include different models for different specification of $\mathcal{X}^{(n)} \subseteq \mathrm{R}^n$ and of the functions h and g.

2.1 The nonlinear regression model

Suppose the design is known. That is, we have $\mathcal{X}^{(n)} = \left\{\xi^0 = (\xi_1^0, ..., \xi_n^0)^T\right\}$. We consider only the first equation (1) and obtain the nonlinear regression model with n observations

$$y_i = g\left(\xi_i^0, \beta\right) + \varepsilon_{1i}, \quad \text{with} \quad i = 1, ..., n. \tag{5}$$

Note that in this model no nuisance parameters occur.

2.2 The nonlinear error-in-variables model

For $h(\xi_i, \beta) = \xi_i$ in (2) we obtain the nonlinear error-in-variables model,

$$y_i = g(\xi_i, \beta) + \varepsilon_{1i}, \tag{6}$$

$$x_i = \xi_i + \varepsilon_{2i}, \tag{7}$$

with $i = 1, ..., n$. This is exactly the functional one, because the variables are considered as fixed and unknown and play the role of a nuisance parameter with increasing dimension. For consistency we need additional assumptions on the set of nuisance parameters $\mathcal{X}^{(n)}$, because we know that for $\mathcal{X}^{(n)} = [0,1]^n$ the least squares estimator is inconsistent, see Kukush and Zwanzig (1996). One interesting additional information may be

$$\mathcal{X}^{(n)} = \left\{\xi = (\xi_1, ..., \xi_n)^T : 0 \le \xi_1 \le \xi_2 \le ... \le \xi_{n-1} \le \xi_n \le 1\right\}. \tag{8}$$

On the first view this assumption seems to be artificial; it is, however, useful in applications, for instance in biology or chemistry. There the unknown design points ξ_i often stand for different levels of concentration. The experimenter measures these concentration levels with error ε_{2i}. But he does have some influence on the level that the concentration lies at and he can guarantee with high security that the concentration level of the next experiment will be higher.

The assumption (8) can be rewritten

$$\mathcal{X}^{(n)} = \left\{\xi = (f(z_1), ..., f(z_i), ..., f(z_n))^T : f : [0,1] \to [0,1],\ f \text{ increasing}\right\}. \tag{9}$$

where the $z_1 \le, ..., \le z_n$ are fixed design points satisfying the following design condition:

D

$$\lim_{n\to\infty} \max_i |z_i - z_{i-1}| = 0.$$

Another possibility is to consider

$$\mathcal{X}^{(n)} = \left\{\xi = (f(z_1), ..., f(z_i), ..., f(z_n))^T : f \in \mathcal{X}_{m,\alpha}(C, L)\right\} \tag{10}$$

with

$$\mathcal{X}_{m,\alpha}(C, L) = \left\{ f \in C_{m,\alpha}[0,1] : \begin{array}{c} \left|f^{(k)}(x)\right| \le C, k = 0, 1, ..., m \\ \left|f^{(m)}(x_1) - f^{(m)}(x_2)\right| \le L|x_1 - x_2|^\alpha \end{array} \right\}. \tag{11}$$

This kind of additional information is also used in the following semiparametric model.

2.3 The nonlinear semiparametric model

Assume that $\{z_1, ..., z_n\} \subset [0,1]$ are known design points with **D.** Under the specification $\xi_i = f(z_i)$ and $g(\xi_i, \beta) = m(z_i, \beta) + \xi_i$, where the functions m and f satisfy the following identification condition,

ID $\exists n_0 \forall n > n_0 \; \exists \{w_i\} > 0, \sum_{i=1}^n w_i = 1 \; \forall f \in \mathcal{X}_{m,\alpha}(C, L) \; \forall \beta \in \mathcal{B}^c$ such that

$$\sum_{i=1}^{n} w_i m(z_i, \beta) f(z_i) = 0, \tag{12}$$

we consider only the first equation (1) and obtain the model

$$y_i = m(z_i, \beta) + f(z_i) + \varepsilon_{1i}. \tag{13}$$

The condition **ID** contains the orthogonality in the sense of the empirical measure generated by the weighted design points $z_1, ..., z_n$. This model (13) describes alternatives in the context of model choice.

3 Minimum contrast estimates

The L_1-norm estimator will be considered as a special minimum contrast estimator. We call a nonrandom positive real function $C_n : \theta \in \Theta^c \to \mathrm{R}_+$ a *contrast for* θ *at* $\theta_{(n)}$ iff it is lower semicontinuous and

$$\theta_{(n)} = \arg\min_{\theta \in \Theta^c} C_n(\theta), \tag{14}$$

where Θ^c denotes the compactification of the parameter set $\mathcal{X}^{(n)} \times \mathcal{B}$ in $\overline{\mathrm{R}}^{n+p}$. The contrast may depend on the sample size n. Examples are the empirical L_q-contrast

$$C_n(\theta) = \sum_{i=1}^{n} w_i \left| g\left(\xi_i^0, \beta\right) - g\left(\xi_i^0, \beta^0\right) \right|^q \tag{15}$$

or the asymptotic L_q-contrast

$$C_n(\theta) = C(\beta) = \int \left| g(x, \beta) - g\left(x, \beta^0\right) \right|^q dG, \tag{16}$$

with the $\theta_{(n)} = (\beta^0, \xi^0)$ satisfying (1) and (2). The first depends on the unknown design points $\xi \in \mathcal{F}^{(n)}$, the other on an asymptotic design G. Under a unique parameterization of the regression function each distance

measure d for functions $g(.,\beta) \in \mathcal{M}$ seems to be a useful contrast at $\beta_{(n)} = \beta^0$, namely $C_n(\beta) = d(g(.,\beta), g(.,\beta^0))$.

We call a measurable function

$$\widetilde{C}_n(.,.,.) : \mathrm{R}^n \times \mathrm{R}^n \times \Theta^c :\to \mathrm{R}_+ \tag{17}$$

a *contrast function* and require that it is continuous with respect to θ. In the general statistical experiment, which includes random processes and random fields as well, Liese and Vajda (1995) introduced a more general concept and called the function corresponding to (17) a contrast principle. Note in order to simplify the denotation we will suppress the dependence on the sample and let the tilde hints to this: $\widetilde{C}_n(X,Y,\theta) =: \widetilde{C}_n(\theta)$.We then define the corresponding estimator as follows.

A measurable solution $\widetilde{\theta} : \mathrm{R}^n \times \mathrm{R}^n \to \Theta^c$ is called *a minimum contrast estimator* iff

$$\widetilde{\theta} \in \arg\min_{\theta \in \Theta^c} \widetilde{C}_n(\theta). \tag{18}$$

Under the model assumptions above the existence of minimum contrast estimators are given by the Lemma 2 of Liese and Vajda (1995).

The following lemma gives the connection between the consistency of the minimum contrast estimator and the uniform consistent approximation of the contrast $C_n(\theta)$ by the contrast function $\widetilde{C}_n(\theta)$. It is a version of an "argmin" result, like the argmax theorem for i.i.d. experiments in van der Vaart and Wellner (1996). Consider the differences of the contrast and of the contrast function,

$$\Delta C_n(\theta) = C_n(\theta) - C_n\left(\theta_{(n)}\right) \quad \text{and} \quad \Delta\widetilde{C}_n(\theta) = \widetilde{C}_n(\theta) - \widetilde{C}_n\left(\theta_{(n)}\right). \tag{19}$$

Lemma 1 *Let $\rho : \mathrm{R}_+ \to \mathrm{R}_+$ be strictly increasing , with $\rho(0) = 0$. Let $d(.,.)$ be a semimetric on Θ^c. For any $\epsilon > 0$ define the set*

$$\overline{\Theta}_n(\epsilon) = \Theta^c \cap \left\{\theta : d\left(\theta, \theta_{(n)}\right) > \epsilon\right\}. \tag{20}$$

Let C_n be a contrast with

$$\Delta C_n(\theta) = C_n(\theta) - C_n\left(\theta_{(n)}\right) \geq \rho\left(d\left(\theta, \theta_{(n)}\right)\right). \tag{21}$$

Then

$$\forall \epsilon > 0 \quad P\left(d\left(\widetilde{\theta}, \theta_{(n)}\right) > \epsilon\right) \leq P\left(\sup_{\theta \in \overline{\Theta}_n(\epsilon)} \frac{\left|\Delta C_n(\theta) - \Delta\widetilde{C}_n(\theta)\right|}{\rho\left(d\left(\theta, \theta_{(n)}\right)\right)} \geq 1\right). \tag{22}$$

Proof: From (21) follows that

$$\min_{\theta \in \overline{\Theta}_n(\epsilon)} \frac{\Delta C_n(\theta)}{\rho\left(d\left(\theta, \theta_{(n)}\right)\right)} \geq 1. \tag{23}$$

Under $\widetilde{\theta} \in \overline{\Theta}_n(\epsilon)$ we have $\widetilde{C}_n\left(\widetilde{\theta}\right) \leq \widetilde{C}_n\left(\theta_{(n)}\right)$ and $\rho\left(d\left(\widetilde{\theta}, \theta_{(n)}\right)\right) > \rho(\epsilon) > 0$, thus

$$\frac{\Delta \widetilde{C}_n\left(\widetilde{\theta}\right)}{\rho\left(d\left(\widetilde{\theta}, \theta_{(n)}\right)\right)} \leq 0. \tag{24}$$

Hence under $\widetilde{\theta} \in \overline{\Theta}_n(\epsilon)$ we obtain from (23) and (24) the following chain of inequalities

$$1 \leq \min_{\theta \in \overline{\Theta}_n(\epsilon)} \frac{\Delta C_n(\theta)}{\rho\left(d\left(\theta, \theta_{(n)}\right)\right)} - \frac{\Delta \widetilde{C}_n\left(\widetilde{\theta}\right)}{\rho\left(d\left(\widetilde{\theta}, \theta_{(n)}\right)\right)} \leq \sup_{\theta \in \overline{\Theta}_n(\epsilon)} \frac{\left|\Delta C_n(\theta) - \Delta \widetilde{C}_n(\theta)\right|}{\rho\left(d\left(\theta, \theta_{(n)}\right)\right)}, \tag{25}$$

which yields the statement of the lemma. □

Note that the rate of convergence of the minimum contrast estimator given by this lemma depends mainly on the separation property of the contrast (21) and the semimetric $d(.,.)$ chosen in (21).

4 The L_1-estimator

In the following we focus our attention on the L_1-contrast function only. This means, let from now on

$$\Delta \widetilde{C}_n(\theta) = \sum_{i=1}^{n} w_{1i}\left(|\varepsilon_{1i} + \Delta g(\xi_i, \beta)| - |\varepsilon_{1i}|\right) + \sum_{i=1}^{n} w_{2i}\left(|\varepsilon_{2i} + \Delta h(\xi_i, \beta)| - |\varepsilon_{2i}|\right) \tag{26}$$

with

$$\Delta g(\xi_i, \beta) = g\left(\xi_i^0, \beta^0\right) - g(\xi_i, \beta) \text{ and } \Delta h(\xi_i, \beta) = h\left(\xi_i^0, \beta^0\right) - h(\xi_i, \beta).$$

We will see that the L_1-contrast is

$$\Delta C_n(\theta) = E_{\xi^0 \beta^0} \Delta \widetilde{C}_n(\theta). \tag{27}$$

It is easily checked that the L_1-contrast does not coincide with the contrast defined by the empirical L_1-metric (15) or by the asymptotic L_1-metric (16), see Lemma 4.

Applying Lemma 1 for the consistency proof of the L_1-estimator, we have to do two steps: first the uniform approximation of the contrast by the contrast function with an appropriate rate, second the study of the separation condition (21).

4.1 The approximation

One of the main advantages of the L_1-approach is that the difference

$$Z(\theta) = \Delta C_n(\theta) - \Delta \widetilde{C}_n(\theta) \tag{28}$$

as a stochastic process $\{Z(\theta), \theta \in \Theta^c\}$ with index set $\Theta^c \subseteq \overline{\mathrm{R}}^{n+p}$ forms an sub-Gaussian process without additional assumptions on the tails of the error distribution P_ε. The only things we need are Lipschitz conditions and a convenient semimetric d in Θ^c. Henceforth, we will use the following denotation for the sum of weighted squares with known normalized weights

$$w_{ji} > 0, \quad \sum_{i=1}^{n} w_{ji} = 1, \quad w_{\max} = \max_{ij} w_{ji} \tag{29}$$

$$\sum_{i=1}^{n} w_{1i} |g(\xi_i, \beta)|^2 = \int |g(x, \beta)|^2 \, dG_{w_1}(\xi) = |g(\xi, \beta)|^2_{w_1},$$

where $G_{w_1}(\xi)$ is the weighted empirical measure generated by the design points $\xi \in \mathcal{X}^{(n)}$. We use also the corresponding notation for the scalar product $(.,.)_{w_1}$. Note that in the unweighted case $w_{\max} = n^{-1}$. The Lipschitz conditions are used with respect to both types of parameters.

L1 $\exists n_0 \, \exists L_1, \; L_1 < \infty \; \forall n \geq n_0 \, \forall \beta \in \mathcal{B}^c \, \forall \xi, \xi' \in \mathcal{X}^{(n)c}$, such that

$$|g(\xi, \beta) - g(\xi', \beta)|^2_{w_1} + |h(\xi, \beta) - h(\xi', \beta)|^2_{w_2} \leq L_1 \, |\xi - \xi'|^2_{w_2}. \tag{30}$$

L2 $\exists n_0 \, \exists L_2, \; L_2 < \infty, \; \forall n \geq n_0 \;\; \forall \xi \in \mathcal{X}^{(n)c} \;\; \forall \beta, \beta' \in \mathcal{B}^c$, such that

$$|g(\xi, \beta) - g(\xi, \beta')|^2_{w_1} + |h(\xi, \beta) - h(\xi, \beta')|^2_{w_2} \leq L_2 \, \|\beta - \beta'\|^2. \tag{31}$$

In (31) $\|.\|$ denotes the Euclidean norm in R^p.

Lemma 2 *Suppose* **L1, L2.** *Set*

$$d(\theta, \theta')^2 = w_{\max} \left[|\xi - \xi'|^2_{w_2} + \|\beta - \beta'\|^2 \right]. \tag{32}$$

Then there is a constant $\alpha = \alpha(L_1, L_2)$ *independent of* n *and* θ *such that for all* $t \geq 0$ *and all* $n \geq n_0$

$$P_\varepsilon \left(|Z(\theta) - Z(\theta')| \geq t \right) \leq 2 \exp\left(- \frac{\alpha t^2}{d(\theta, \theta')^2} \right). \tag{33}$$

Proof: We have for $Z(\theta)$ as defined in (28) with (26), (27) that $Z_n(\theta) = \sum_{i=1}^n X_i(\theta)$ and $EX_i(\theta) = 0$. Using the inequality $||a+b| - |a+c|| \leq |b+c|$ we obtain

$$|X_i(\theta) - X_i(\theta')| \leq 2w_{1i} |g(\xi_i, \beta) - g(\xi_i', \beta')| + 2w_{2i} |h(\xi_i, \beta) - h(\xi_i', \beta')|,$$

and using the triangle inequality we get further

$$|X_i(\theta) - X_i(\theta')| \leq d_i(\theta, \theta') \tag{34}$$

where $d_i(\theta, \theta')$ denotes the following seminorm $d_i(\theta, \theta') = d_{1i}(\theta, \theta') + d_{2i}(\theta, \theta')$, with

$$d_{1i}(\theta, \theta') = 2w_{1i} |g(\xi_i, \beta) - g(\xi_i', \beta)| + 2w_{1i} |g(\xi_i', \beta) - g(\xi_i', \beta')|,$$

$$d_{2i}(\theta, \theta') = 2w_{2i} |h(\xi_i, \beta) - h(\xi_i', \beta)| + 2w_{2i} |h(\xi_i', \beta) - h(\xi_i', \beta')|.$$

Because of (34), we can apply the Corollary 3.2 of van de Geer (1990). Thus there is a constant α' independent of n and θ such that

$$P_{\xi^0 \beta^0}(|Z(\theta) - Z(\theta')| \geq t) \leq 2 \exp\left(-\frac{t^2 \alpha'}{2\overline{d}(\theta, \theta')^2}\right).$$

with $\overline{d}(\theta, \theta')^2 = \sum_{i=1}^n d_i(\theta, \theta')^2$. Under the conditions **L1, L2** we estimate

$$\sum_{i=1}^n d_i(\theta, \theta')^2 \leq 16 w_{\max}(L_1 + L_2) \left(|\xi - \xi'|_{w_2}^2 + \|\beta - \beta'\|^2\right)$$

and obtain the statement with $\alpha = \alpha' (32(L_1 + L_2))^{-1}$. □

In order to formulate the entropy condition we need a few more definitions. Let us introduce them for a general set A with a metric d, because inside the proof we will use the notion of entropy in the context of several different sets. A family of subsets $U_1, ...U_N$ is called an *ϵ-covering* of A with respect to a metric d, if the diameter of each U_k does not exceed 2ϵ and if the sets cover A, $A \subseteq \cup_i^N U_i$. The *ϵ-covering* number $N(\epsilon)$ is the minimal number of U_k's in any *ϵ-covering* of A. The *ϵ-entropy* $H(\epsilon)$ of A is given by the logarithm $H(\epsilon) = \ln N(\epsilon)$. The entropy depends on the metric d and on A. We therefore denote the local *ϵ-entropy* of $A \cap (a : d(a, a_0) \leq D)$ by $H_{A,d}(\epsilon, D)$. We will require a condition on the local entropy of the nuisance parameter set only, that is $A = \mathcal{X}^{(n)}$ and $d(\xi, \xi') = |\xi - \xi'|_{w_2}$. For abbreviation we write $H_{\mathcal{X}^{(n)}, |\cdot|_{w_2}}(\epsilon, D) = H_\xi(\epsilon, D)$. The entropy condition we need is:

Ent For all $\delta, \eta_n > 0$, with $\delta\,\eta_n\sqrt{w_{\max}} \geq 1$

$$\lim_{L\to\infty} \frac{\int_0^1 \sqrt{H_\xi\left(Lu\delta, L\delta D\right)}du}{L\delta\,\eta_n\sqrt{w_{\max}}} = 0. \tag{35}$$

The following lemma is an application of a modified result of van de Geer (1990), which is an adaptation of the chaining method of Pollard (1984), on page 144.

Lemma 3 *Suppose* **L1** *with* L_1, **L2** *with* L_2, **Ent** *with* η_n *and* δ, *then there exist constants* L_0 *and* $C_0 = C\left(L_1, L_2\right)$, *such that for all* $L \geq L_0$ *and all* $n \geq n_0$

$$P_{\xi^0,\beta^0}\left(\sup_{\theta\in\overline{\Theta}(L\delta)} \frac{\left|\Delta\widetilde{C}_n\left(\theta\right) - \Delta C_n\left(\theta\right)\right|}{|\xi-\xi'|_{w_2}^2 + \|\beta-\beta'\|^2} \geq w_{\max}\eta_n\right) \leq \exp\left(-C_0L^2\eta_n^2\delta^2 w_{\max}\right) \tag{36}$$

where $\overline{\Theta}_n\left(\epsilon\right) = \left((\xi,\beta) : |\xi-\xi'|_{w_2}^2 + \|\beta-\beta'\|^2 \geq \varepsilon^2\right)$ *and* $\Delta\widetilde{C}_n(\theta)$, $\Delta C_n\left(\theta\right)$ *given in (26) and (27).*

Proof: We will apply a small modified version of Lemma 3.4. of van de Geer (1990), with $\Lambda = \Theta^c$ and the semimetric d given in (32) and $Z_n\left(\lambda\right)$ from (28) with $Z_n\left(\lambda^0\right) = 0$. The used modified version is: Under the entropy condition on Θ^c with respect to the metric d, that is for all $\delta', \eta_n > 0$, with $\delta'\eta_n > 1$

$$\lim_{L\to\infty} \frac{\int_0^1 \sqrt{H_{\Theta^c,d}\left(uL\delta', L\delta' D\right)}du}{L\delta'\eta_n} = 0, \tag{37}$$

it holds

$$P_{\xi^0,\beta^0}\left(\sup_{d(\theta,\theta^0)>L\delta'} \frac{|Z_n\left(\theta\right)|}{d^2\left(\theta,\theta^0\right)} \geq \eta_n\right) \leq \exp\left(-\eta_n^2 C_0 L^2\delta'^2\right). \tag{38}$$

For $\delta' = \delta\sqrt{w_{\max}}$ from (38) follows the result. The difference to the lemma of van de Geer is that we have $\sqrt{n} = \eta_n$. The proof of this modification has the same steps, but we have to change $\sqrt{n}$ to η_n in the entropy condition and in the exponential rate. Her assumptions on $Z_n\left(\lambda\right)$ are not needed in the L_1-context, since the main property she used in the proof is (33), that the process is sub-Gaussian. It remains to check the entropy condition on Θ^c (37). For Cartesian products $A = A_1 \times A_2$ with $a = (a_1, a_2)$ and $d_A^2\left(a, a'\right) \leq d_{A_1}^2\left(a_1, a_1'\right) + d_{A_2}^2\left(a_2, a_2'\right)$, we know the following inequality:

$$H_{A,d_A}\left(\epsilon, D\right) \leq H_{A_1,d_{A_1}}\left(\frac{\epsilon}{2}, D\right) + H_{A_2,d_{A_2}}\left(\frac{\epsilon}{2}, D\right). \tag{39}$$

It can be derived in a similar way as the formula (7) in Lorentz (1966), on page 152. Here we have $A = \mathcal{X}^{(n)} \times \mathcal{B}$ and $d^2_{A_1}(a_1, a_1') = w_{\max} |\xi - \xi'|^2_{w_2}$, and $d^2_{A_2}(a_2, a_2') = w_{\max} \|\beta - \beta'\|^2$,

$$H_{A_1, d_{A_1}}(\epsilon\sqrt{w_{\max}}, D\sqrt{w_{\max}}) = H_{\mathcal{X}^{(n)}, |.|_{w_2}}(\epsilon, D).$$

$\mathcal{B}$ is a set of fixed dimension p, therefore the local entropy is bounded:

$$H_{A_2, d_{A_2}}(\epsilon\sqrt{w_{\max}}, D\sqrt{w_{\max}}) = H_{\mathcal{B}, \|.\|}(\epsilon, D) \leq p \ln\left(\frac{2D}{\epsilon}\sqrt{p}\right).$$

Thus it suffices to require the entropy condition (37) for $H_{A_1, d_{A_1}}$ with $\delta' = \delta\sqrt{w_{\max}}$ only, that is, assume **Ent.** □

4.2 The separation condition

The aim of this subsection is to verify the separation condition (21) for the L_1-contrast in (27). First we quote a result by Oberhofer (1982) in the form given by van de Geer (1990).

Lemma 4 *Suppose ε r.v. whose distribution satisfies* **E** *with constants D_0 and κ_ε then for all $|\Delta| \leq \Delta_{\max}$*

$$\kappa_\varepsilon \frac{D_0}{\Delta_{\max}} |\Delta|^2 \leq E(|\varepsilon + \Delta| - |\varepsilon|) \leq |\Delta|. \tag{40}$$

Proof: This is the i.i.d. version of Lemma 4.2. of van de Geer (1990). □

Applying Lemma 4 to the L_1-contrast with

$$\Delta_{\max} = \max_i \max_{\xi_i} \max_{\beta \in \mathcal{B}^c} \left\{ \left| g\left(\xi_i^0, \beta^0\right) - g(\xi_i, \beta) \right| + \left| h\left(\xi_i^0, \beta^0\right) - h(\xi_i, \beta) \right| \right\},$$

we obtain

$$C_n(\theta) - C_n\left(\theta^0\right) \geq \kappa_\varepsilon \frac{D_0}{\Delta_{\max}} D_n(\xi, \beta), \tag{41}$$

with

$$D_n(\xi, \beta) = \left| g(\xi, \beta) - g\left(\xi^0, \beta^0\right) \right|^2_{w_1} + \left| h(\xi, \beta) - h\left(\xi^0, \beta^0\right) \right|^2_{w_2}. \tag{42}$$

Now we need separation conditions on g and h also, such that it is possible to estimate

$$D_n(\xi, \beta) \geq \rho\left(d\left(\theta, \theta_{(n)}\right)\right) \tag{43}$$

for an appropriately chosen metric d on the parameter space. Deriving (43) means solving the identification problem in semiparametric models. In the same way this problem occurs also in the L_2-norm theory. This is

the main reason, why we are not able to obtain a nice consistency result for the parameter of interest β in the general setting (1), (2). From now on we consider the models separately. We say a regression function g fulfills the contrast condition **Con** iff

Con $\exists n_0 \ \forall n \geq n_0 \ \exists a_n, 0 < a_n < \infty, \ \forall \xi \in \mathcal{X}^{(n)} \ \ \forall \beta, \beta' \in \Theta$

$$\left| g\left(\xi, \beta\right) - g\left(\xi', \beta\right) \right|_{w_1} \geq a_n \left\| \beta - \beta' \right\|.$$

Under **Con** for g we have in the nonlinear regression model (5), that

$$D_n\left(\xi, \beta\right) = \left| g\left(\xi^0, \beta\right) - g\left(\xi^0, \beta^0\right) \right|_{w_1}^2 \geq a_n^2 \left\| \beta - \beta^0 \right\|^2. \tag{44}$$

In the nonlinear semiparametric model (13) the identification condition **ID** implies

$$2\left(m\left(z, \beta\right) - m\left(z, \beta^0\right), f\left(z\right) - f^0\left(z\right)\right)_{w_1} = 0,$$

and under **Con** for m we have here also

$$D_n\left(\xi, \beta\right) = \left| m\left(z, \beta\right) - m\left(z, \beta^0\right) + f\left(z\right) - f^0\left(z\right) \right|_{w_1}^2$$

$$= \left| m\left(z, \beta\right) - m\left(z, \beta^0\right) \right|_{w_1}^2 + \left| f\left(z\right) - f^0\left(z\right) \right|_{w_1}^2 \geq a_n^2 \left\| \beta - \beta^0 \right\|^2. \tag{45}$$

For the nonlinear functional relation model (6), (7) the following lemma helps to solve the identification problem. This result is strongly related to the Lemma 1 in Zwanzig (1990). Define

$$L_n\left(\xi, \beta\right) = \left| g\left(\xi, \beta\right) - g\left(\xi^0, \beta\right) \right|_{w_1}^2 + \left| g\left(\xi^0, \beta\right) - g\left(\xi^0, \beta^0\right) \right|_{w_1}^2 + \left| \xi - \xi^0 \right|_{w_2}^2. \tag{46}$$

Lemma 5 *Under* **L1**, $\exists n_0 \ \exists \tau > 0 \ \forall n \geq n_0 \ \ \forall \xi, \xi^0 \in \left(\mathcal{X}^{(n)}\right)^c \ \ \forall \beta, \beta^0 \in \mathcal{B}^c$ *such that*

$$\left| g\left(\xi, \beta\right) - g\left(\xi^0, \beta^0\right) \right|_{w_1}^2 + \left| \xi - \xi^0 \right|_{w_2}^2 > \tau \, L_n\left(\xi, \beta\right). \tag{47}$$

Proof: Inside of this proof let us use the abbreviations $g\left(\xi^0, \beta^0\right) = g^{00}$ and $g\left(\xi^0, \beta\right) = g^0$. By adding $\pm g^0$ in $\left| g - g^{00} \right|_{w_1}^2$, we obtain

$$D_n\left(\xi, \beta\right) = L_n\left(\xi, \beta\right)\left(1 - 2\Delta_n\left(\xi, \beta\right)\right) \tag{48}$$

with

$$\Delta_n\left(\xi, \beta\right) = \frac{\left(g^0 - g^{00}, g^0 - g\right)_{w_1}}{\left| g - g^0 \right|_{w_1}^2 + \left| g^{00} - g^0 \right|_{w_1}^2 + \left| \xi - \xi^0 \right|_{w_2}^2} \tag{49}$$

for $L_n(\xi,\beta) > 0$ and $\Delta_n(\xi,\beta) = 0$ otherwise. It remains to show, that there exists a constant $\tau > 0$ such that

$$\sup_{\xi\in\left(\mathcal{X}^{(n)}\right)^c}\sup_{\beta\in\Theta^c}\Delta_n(\xi,\beta) \le \frac{1}{2} - \tau. \tag{50}$$

Let $c = \frac{\left(\frac{1}{2}-\tau_1\right)^2}{L_1}$ with L_1 from (30) and τ_1 such that $0 < \tau_1 < \frac{1}{2}$. We will distinguish two cases:

$$\text{i) } \left|\xi-\xi^0\right|_{w_2}^2 \le c\left|g^0-g^{00}\right|_{w_1}^2 \quad , \quad \text{ii) } \left|\xi-\xi^0\right|_{w_2}^2 \ge c\left|g^0-g^{00}\right|_{w_1}^2 .$$

i) We apply the Cauchy-Schwarz inequality and the assumption (30)

$$\Delta_n(\xi,\beta)^2 \le \frac{\left|g^0-g\right|_{w_1}^2\left|g^0-g^{00}\right|_{w_1}^2}{L_n^2} \le \frac{L_1\left|\xi-\xi^0\right|_{w_2}^2\left|g^0-g^{00}\right|_{w_1}^2}{L_n^2}$$

with $L_n = L_n(\xi,\beta)$ given in (46). Note $\left|g^0-g^{00}\right|_{w_1}^2 \le L_n$. Under i) we have

$$\Delta_n(\xi,\beta)^2 \le L_1 c\left(\frac{\left|g^0-g^{00}\right|_{w_1}^2}{L_n}\right)^2 \le L_1 c \le \left(\frac{1}{2}-\tau_1\right)^2,$$

and thus in case i) (50) follows. Consider case ii). Because of

$$\begin{aligned} 0 &\le \left|\left(g^0-g^{00}\right)-\left(g^0-g\right)\right|_{w_1}^2 \\ &= \left|g^0-g^{00}\right|_{w_1}^2 + \left|g^0-g\right|_{w_1}^2 - 2\left(g^0-g^{00},g^0-g\right)_{w_1}, \end{aligned}$$

we have

$$2\left(g^0-g^{00},g^0-g\right)_{w_1} \le \left|g^0-g^{00}\right|_{w_1}^2 + \left|g^0-g\right|_{w_1}^2 .$$

Using this and (46) we obtain for $\Delta_n = \Delta_n(\xi,\beta)$

$$2\Delta_n \le \frac{\left|g^0-g^{00}\right|_{w_1}^2 + \left|g^0-g\right|_{w_1}^2}{L_n} \le 1 - \frac{\left|\xi-\xi^0\right|_{w_2}^2}{L_n}.$$

From the assumption (30) we get

$$2\Delta_n \le 1 - \frac{\left|\xi-\xi^0\right|_{w_2}^2}{\left|g^0-g^{00}\right|_{w_1}^2 + (1+L_1)\left|\xi-\xi^0\right|_{w_2}^2}.$$

For positive a, the function $f(x) = \frac{x}{a+(1+L_1)x}$ is increasing in x. Using ii) we have $ca \le x$ and $f(x) \ge \frac{c}{1+c+cL_1}$. We obtain $2\Delta_n \le 1 - \frac{c}{1+(1+L_1)c}$. For τ_2 such that $2\tau_2 = \frac{c}{1+(1+L_1)c}$, $0 > \tau_2 > \frac{1}{2}$ one has under ii) $\Delta_n \le \frac{1}{2} - \tau_2$. We choose $\tau = \min(\tau_1,\tau_2)$ and get (50). □

5 The consistency of the L_1-estimators

In this section we summarize the results above and obtain the consistency of the L_1-estimators in the different submodels.

5.1 The nonlinear regression model

Consider the model (5). Then we have for

$$\widetilde{\beta} = \arg\min_{\beta \in \mathcal{B}^c} \sum_{i=1}^{n} w_{1i} \left| y_i - g\left(\xi_i^0, \beta\right) \right| \tag{51}$$

the following strong consistency result. Set

$$G_{\max} = \max_i \sup_{\beta \in \mathcal{B}^c} \left| g\left(\xi_i^0, \beta\right) - g\left(\xi_i^0, \beta^0\right) \right|. \tag{52}$$

Theorem 1 *Suppose for the error distribution* $\mathbf{E}$ *with the constants* κ_ε, D_0 *and suppose for the regression function* g *that*

$$\exists n_0 \, \exists L_2 \forall n \geq n_0 \; \exists a_n, a_n > 0 \; \forall \beta, \beta' \in \mathcal{B}^c$$

$$a_n^2 \left\|\beta - \beta'\right\|^2 \leq \left| g\left(\xi^0, \beta\right) - g\left(\xi^0, \beta'\right) \right|_{w_1}^2 \leq L_2 \left\|\beta - \beta'\right\|^2. \tag{53}$$

Then there exists a positive constant C_0 *such that for all* $L > 0$ *and all* $n \geq n_0$

$$P_{\xi^0, \beta^0}\left(\left\| \widetilde{\beta} - \beta^0 \right\| > L \right) \leq \exp\left(- \frac{C_0}{w_{\max}} \left(\frac{a_n^2 \kappa_\epsilon D_o L}{G_{\max}} \right)^2 \right). \tag{54}$$

Proof: Under (53) from (44) and from (41) with (52) follows that the separation condition (21) is fulfilled with $\rho\left(\left\|\beta - \beta^0\right\|\right) = \frac{\kappa_\epsilon D_o a_n^2}{G_{\max}} \left\|\beta - \beta^0\right\|^2$. Lemma 1 gives

$$P_{\xi^0 \beta^0}\left(\left\| \widetilde{\beta} - \beta^0 \right\| > L \right) \leq P_{\xi^0, \beta^0}\left(\sup_{\theta \in \overline{\Theta}(L)} \frac{\left| \Delta C_n(\theta) - \Delta \widetilde{C}_n(\theta) \right|}{\left\|\beta - \beta^0\right\|^2} \geq \frac{\kappa_\epsilon D_o a_n^2}{G_{\max}} \right).$$

The entropy condition **Ent** is satisfied for the one point set. Then the result (54) is a consequence of Lemma 3 with $\eta_n = \frac{\kappa_\epsilon D_o a_n^2}{G_{\max} w_{\max}}$. □

Note the result is interesting only for $\frac{a_n^2}{G_{\max}} \geq \sqrt{w_{\max}}$.

5.2 The nonlinear error-in-variables model

Consider the model (6), (7). The L_1-norm estimator is defined as

$$\widetilde{\beta} = \arg \min_{\beta \in \mathcal{B}^c} \min_{\xi \in \mathcal{X}^{(n)c}} \sum_{i=1}^{n} w_{1i} \left| y_i - g\left(\xi_i, \beta\right) \right| + w_{2i} \left| x_i - \xi_i \right| . \tag{55}$$

Set

$$G_{\max} = \max_i \sup_{\xi \in \mathcal{X}^{(n)c}} \sup_{\beta \in \mathcal{B}^c} \left| g\left(\xi_i, \beta\right) - g\left(\xi_i^0, \beta^0\right) \right| + \left| \xi_i - \xi_i^0 \right| . \tag{56}$$

Then we have the following exponential probability inequality.

Theorem 2 *Suppose for the error distribution* **E** *with the constants* κ_ε, D_0 *and suppose for the nuisance parameter set the entropy condition* **Ent** *is satisfied with* $\eta_n = \frac{a_n^2 \kappa_\epsilon D_o \tau}{w_{\max} G_{\max}}$ *and* δ_n.

Suppose for the regression function g that

$$\exists n_0 \; \exists L_1, L_2 < \infty, \; \forall n \geq n_0 \; \exists a_n, a_n > 0 \quad \forall \beta, \beta' \in \mathcal{B}^c \;\; \forall \xi, \xi' \in \mathcal{X}^{(n)c} \tag{57}$$

$$\begin{gathered} a_n^2 \left\| \beta - \beta' \right\|^2 \leq \left| g\left(\xi, \beta\right) - g\left(\xi, \beta'\right) \right|_{w_1}^2 \leq L_2 \left\| \beta - \beta' \right\|^2 \\ \text{and} \quad \left| g\left(\xi, \beta\right) - g\left(\xi', \beta\right) \right|_{w_1}^2 \leq L_1 \left| \xi - \xi' \right|_{w_2}^2 . \end{gathered} \tag{58}$$

Then there exist positive constants C_0, L_0 such that for all $L > L_0$ and all $n \geq n_0$

$$P_{\xi^0, \beta^0} \left(\left\| \widetilde{\beta} - \beta^0 \right\| > L\delta_n \right) \leq \exp \left(- \frac{\delta_n^2}{w_{\max}} \left(\frac{\kappa_\epsilon D_o a_n^2}{G_{\max}} \right)^2 C_0 L^2 \right) . \tag{59}$$

Proof: Without loss of generality we set $a_n < 1$. Under (56) from Lemma 4 and under (58) from Lemma 5 follows that the separation condition (21) is fulfilled with

$$\rho \left(\sqrt{\left\| \beta - \beta^0 \right\|^2 + \left| \xi - \xi^0 \right|_{w_2}^2} \right) = \kappa_\epsilon D_o \frac{a_n^2 \tau}{G_{\max}} \left(\left\| \beta - \beta^0 \right\|^2 + \left| \xi - \xi^0 \right|_{w_2}^2 \right) .$$

Lemma 1 gives

$$\begin{aligned} P_{\xi^0 \beta^0} \left(\left\| \widetilde{\beta} - \beta^0 \right\| > L\delta_n \right) &\leq P_{\xi^0 \beta^0} \left(\left\| \widetilde{\beta} - \beta^0 \right\|^2 + \left| \widetilde{\xi} - \xi^0 \right|_{w_2}^2 > L^2 \delta_n^2 \right) \\ &\leq P_{\xi^0, \beta^0} \left(\sup_{\theta \in \overline{\Theta}(L\delta_n)} \frac{\left| \Delta C_n\left(\theta\right) - \Delta \widetilde{C}_n\left(\theta\right) \right|}{\frac{\kappa_\epsilon D_o}{G_{\max}} a_n^2 \tau \left(\left\| \beta - \beta^0 \right\|^2 + \left| \xi - \xi^0 \right|_{w_2}^2 \right)} \geq 1 \right) . \end{aligned}$$

The entropy condition **Ent** is assumed above explicitly. Hence the result (59) is a consequence of Lemma 3 with $\eta_n = \frac{\kappa_\epsilon D_o \tau a_n^2}{w_{\max} G_{\max}}$. □

Consider now special cases and the unweighted case $w_{\max} = n^{-1}$.

Corollary 1 *Suppose for the error distribution* **E** *and suppose for the regression function g that (57),(58) is valid with*

$$\frac{G_{\max}}{a_n^2} \leq const, \text{ for all } n. \tag{60}$$

Suppose i) $\mathcal{X}^{(n)}$ *defined in (9) or ii)* $\mathcal{X}^{(n)}$ *defined in (10). Then*

$$\widetilde{\beta} \to \beta^0 \quad P_{\xi^0,\beta^0} - a.s.. \tag{61}$$

Proof: Under (60) in the unweighted case we have $\eta_n \sqrt{w_{\max}} \leq const\sqrt{n}$.

i) For $\mathcal{X}^{(n)}$ defined in (8), the entropy is $H_\xi(\delta, D) \leq const \frac{1}{\delta} \ln^+\left(\frac{1}{\delta}\right)$ and the entropy condition **Ent** is satisfied with $\delta = n^{-\frac{1}{3}} (\ln n)^{\frac{1}{3}}$, (see Example 2.1 of van de Geer (1990)). From Theorem 2 follows that there exists a d, $0 < d < 1$, for all $\epsilon > 0$

$$\sum_{n=1}^{\infty} P_{\xi^0,\beta^0}\left(\left\|\widetilde{\beta} - \beta^0\right\| > \epsilon\right) \leq \sum_{n=1}^{\infty} \exp\left(-n^d \, \epsilon \, const\right) \tag{62}$$

$$< const\,(n_0) \sum_{n=n_0}^{\infty} n^{-2} < \infty. \tag{63}$$

We obtain the statement by the Lemma of Borel Cantelli.

ii) For $\mathcal{X}^{(n)}$ defined in (10) the entropy can by derived from the classical result of Kolmogorov and Tichomirov (1960), for the sup-norm $|f - f^0|_{\sup} = \max_{x\in[0,1]} |f(x) - f^0(x)|$

$$H_{\mathcal{X}^{(n)},|\cdot|_{\sup}}(\delta, D) \leq const \left(\frac{C}{\delta}\right)^{\frac{1}{m+\alpha}}.$$

Since under the design assumption **D** $\max_i |z_i - z_{i-1}| \leq c_d$ for $n \geq n_0$ and since

$$\begin{aligned} \left|\xi - \xi^0\right|_{w_2} &\leq \max_i \left|f(z_i) - f^0(z_i)\right| \leq \left|f - f^0\right|_{\sup} \\ &\leq \max_i \left|f(z_i) - f^0(z_i)\right| + 2Lc_d \end{aligned}$$

we have $\mathcal{X}^{(n)} \subseteq \left\{f : |f - f^0|_{\sup} \leq 3(C + L)\right\}$. Thus

$$H_\xi(\delta, D) \leq const \left(\frac{1}{\delta}\right)^{\frac{1}{m+\alpha}}. \tag{64}$$

Then the entropy condition **Ent** is fulfilled for $\delta = n^{-b}$ with $b = \frac{m+\alpha}{2(m+\alpha)+1} > 0$ and the result (61) follows by the same arguments as in (62). □

5.3 The nonlinear semiparametric model

Consider the model (13) with design condition **D** and identification condition **ID.** Then we can derive on the same way as above the strong consistency of the L_1-estimator

$$\widetilde{\beta} = \arg\min_{\beta \in \mathcal{B}^c} \min_{f \in \mathcal{X}_{m,\alpha}(C,L)} \sum_{i=1}^{n} w_i \left| y_i - m\left(z_i, \beta\right) - f\left(z_i\right) \right| .$$

Theorem 3 *Suppose for the error distribution* **E** *and suppose for the function* m *that*

$$\exists n_0 \; \exists L_2 < \infty, \; \exists a > 0 \; \forall n \geq n_0 \quad \forall \beta, \beta' \in \mathcal{B}^c$$

$$a^2 \left\| \beta - \beta' \right\|^2 \leq \left| m\left(z, \beta\right) - m\left(z, \beta'\right) \right|_{w_1}^2 \leq L_2 \left\| \beta - \beta' \right\|^2 \tag{65}$$

and

$$G_{\max} = \max_i \sup_{\beta \in \mathcal{B}^c} \left| m\left(z_i, \beta\right) - m\left(z_i, \beta^0\right) \right| \leq const. \tag{66}$$

Then

$$\widetilde{\beta} \to \beta^0 \quad P_{\xi^0, \beta^0} - a.s. \tag{67}$$

Proof: Under **ID** and (65) from (41) and (45) the separation condition (21) of Lemma 1 is satisfied with $\Delta_{\max} = G_{\max} + 2C$, where $G_{\max}$ from (66) and where C from (11),

$$\rho\left(\left\|\beta - \beta^0\right\|\right) = \kappa_\epsilon D_o \frac{a^2}{G_{\max} + 2C} \left\|\beta - \beta^0\right\|^2 .$$

Because of (64) for $\eta_n \sqrt{w_{\max}} \leq const\sqrt{n}$ and for $\delta = n^{-b}$ with $b = \frac{m+\alpha}{2(m+\alpha)+1} > 0$ the entropy condition **Ent** of Lemma 3 is valid. Then from both lemmata follows an inequality of type (62) and we obtain (67) by the same arguments as in (62). □

References

[1] L. Birge and P. Massart (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Relat. Fields* **77,** 115–150.

[2] L. Birge and P. Massart (1994). Minimum contrast estimators on sieves. Technical Report 94, Universite de Paris-Sud.

[3] S. van de Geer (1995). The method of sieves and minimum contrast estimators. *Math. Meth, Statist.* **4,** 20–38.

[4] S.van de Geer (1990). Estimating a regression function. *Ann. Statist.* **18,** 907–924.

[5] A. N. Kolmogoroff and W. M. Tichomirow (1960). *Arbeiten zur Informationstheorie III.* Mathematische Forschungsberichte, Verlag der Wissenschaften, Berlin.

[6] A. Kukush and S. Zwanzig (1996). On an alternative estimator in nonlinear functional relations. Unpublished manuscript.

[7] A. Kukush and S. Zwanzig (1996). On inconsistency of the least squares estimator in nonlinear functional error-in-variables models. Technical Report 12, Institute of Mathematical Stochastics, University of Hamburg.

[8] F. Liese and Vajda I (1994). Consistency of M-estimates in general regression models. *J. Multiv. Anal.* **50,** 93–114.

[9] F. Liese and Vajda I. (1995). Necessary and sufficient conditions for consistency of generalized M-estimates. *Metrika*, **42**, 291–324.

[10] O. Linton (1995). Second order approximation in the partially linear regression model. *Econometrica* **63,** 1079–1112.

[11] G. G. Lorentz (1966). *Approximation of Functions.* Holt, Rinehart and Winston.

[12] W. Oberhofer (1982). The consistency of nonlinear regression minimizing the L1-norm. *Ann. Statist.* **10,** 316–319.

[13] D. Pollard (1984). *Convergence of Stochastic Processes.* Springer Series in Statistics.

[14] G. D. Richardson and B. B. Bhattacharyya (1987). Consistent L1-estimators in nonlinear regression for a noncompact parameter space. *Sankhyā* A **49,** 377–387.

[15] A. W. van der Vaart and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

[16] S. Zwanzig (1980). The choice of approximative models in nonlinear regression. *Statistics* **11,** 23–47.

[17] S. Zwanzig (1990). On consistency in nonlinear functional relations. P-MATH P-Math 10-90, Institute of Mathematics, Academy of Science GDR.