On consistent estimators in nonlinear functional errors-in-variables models

Alexander Kukush* K.U. Leuven

Silvelyn Zwanzig† Uppsala University

Abstract

For the implicit nonlinear functional relation model a new contrast estimation procedure is proposed, where the deconvolution idea is used for eliminating the nuisance parameters in the usual minimum contrast function. Several examples are considered including L1- and L2- methods. Sufficient conditions for consistency are given.

Keywords: nonlinear errors-in-variables models, functional relationship, minimum contrast estimation, estimation function, total least squares.

1 Introduction

The functional relation model is characterized by an increasing number of nuisance parameters and a fixed number of parameters of interest. It is an old fact that in models with an increasing number of nuisance parameters usual estimation procedures may fail, compare for instant the classical paper of Neyman, Scott (1948), [**?**]. The main problem is the elimination of the nuisance parameters in the estimating function, in such a way that the estimating procedure delivers consistent estimators. In linear models the least squares approach fulfills that and gives consistent and efficient estimators for the parameter of interest. The situation is different in nonlinear models.

In this paper we will present a method for construction of consistent estimators in nonlinear functional relation models. The main idea is a correction of a general minimum contrast estimation criterion. Under the assumption that the minimal point is given, we are interested in the convergence of probability for increasing sample size. Our method is strongly related to Stefanski's corrected-score method, see Carroll et al. (1995), [**?**] Chapter 6.

*Dep. ESAT/SISTA-COSIC, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, akukush@esat.kuleuven.ac.be

†Dep. of Mathematics, Box 480, S-75106 Uppsala, Sweden, 20@math.uu.se

2 Corrected minimum contrast estimators

In order to emphasize the symmetric character of the observations we consider the implicit version of errors-in-variables models, which makes no difference between independent and dependent variables.

$$G(\eta_i, \beta) = 0, \quad Z_i = \eta_i + \varepsilon_i, \quad i = 1, .., n \,, \tag{1}$$

where ε_i are *i.i.d.* with expected value zero and covariance matrix Γ. The functional relation $G : \mathrm{R}^q \times \mathrm{R}^p \longrightarrow \mathrm{R}$ is a known smooth function. The expected values of the observations $\eta_i \in \mathcal{D} \subset \mathrm{R}^q$, form the nuisance parameter $\eta = (\eta_1, ..., \eta_n) \in \mathcal{D}^n$. The parameter of interest $\beta \in \mathcal{B} \subset \mathrm{R}^p$ has fixed dimension. The parameter consisting of the nuisance parameter and the parameter of interest is denoted by θ.

We apply the principle of *minimum contrast estimators* $\widetilde{\theta}$, which are defined as a measurable solution of

$$\widetilde{\theta} \in \arg\min_{\theta \in \Theta} \widetilde{C}_n(Z, \theta) \,. \tag{2}$$

Because of the model structure in (**??**) we will consider estimation criteria of the following structure.

$$\widetilde{C}_n(Z, \theta) = \frac{1}{n} \sum_{i=1}^{n} \widetilde{c}(Z_i, \eta_i, \beta) \,. \tag{3}$$

Examples for $\widetilde{C}_n(Z, \theta)$ are the total least squares estimator and L1- type estimator with

$$\frac{1}{n} \sum_{i=1}^{n} \|Z_i - \eta_i\|^2 \,, \quad \frac{1}{n} \sum_{i=1}^{n} |Z_i - \eta_i| \,. \tag{4}$$

A contrast function $\widetilde{C}_n(Z, \theta)$ is reasonable if for $C_n(\theta_0, \theta) = E_{\theta_0} \widetilde{C}_n(Z_i, \theta)$, with $\theta_0 = \left(\eta_1^0, ..., \eta_n^0, \beta_0\right)^T$, the following contrast condition is fulfilled:

(i) For all θ_0, θ

$$\Delta C_n(\theta) = C_n(\theta_0, \theta) - C_n(\theta_0, \theta_0) \geq \rho(d(\theta_0, \theta)) \,, \tag{5}$$

where $d(\theta_0, \theta)$ denotes a seminorm in Θ and $\rho(.)$ a strong monotone function with $\rho(0) = 0$. In the least squares case we have $\Delta C_n(\theta) = \left\|\eta - \eta^0\right\|^2$. The consistency of the minimum contrast estimators can be shown with help of the following inequality

$$\forall \epsilon > 0 \quad P\left(d\left(\widetilde{\theta}, \theta_0\right) > \epsilon\right) \leq P\left(\sup_{\theta \in \Theta(\epsilon)} \frac{\left|\Delta C_n(\theta) - \Delta \widetilde{C}_n(\theta)\right|}{\rho(d(\theta, \theta_0))} \geq 1\right), \tag{6}$$

where $\Theta(\epsilon) = \Theta \cap \{\theta : d(\theta_0, \theta) > \epsilon\}$ and $\Delta \widetilde{C}_n(\theta) = \widetilde{C}_n(\theta) - \widetilde{C}_n(\theta_0)$. The main problem of this approach is that the dimension of the parameter θ grows in the same order as the sample size. Under additional assumptions on the nuisance parameter set, which imply the consistent estimation of the nuisance parameters in Zwanzig (1997) [?] and Zwanzig (1998) [?] the consistency of the L1- and L2-norm estimators is shown.

But we are interested in an estimation of the parameter of interest only. We set $d(\theta_0, \theta) = \|\beta - \beta_0\|$. Thus we search for estimation criteria, where the nuisance parameter is eliminated and which fulfill (**??**). The naive approach is to use Z_i instead of η_i in (**??**) and to consider the estimator

$$\widetilde{\beta}_{naive} \in \arg\min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^{n} \widetilde{c}(Z_i, Z_i, \beta).$$

But this estimator is inconsistent. The total least squares approach is

$$\widetilde{\beta}_{TLS} \in \arg\min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^{n} \min_{\eta_i : G(\eta_i, \beta) = 0} \|Z_i - \eta_i\|^2. \tag{7}$$

That works fine for linear models. But for nonlinear models (**??**) is not fulfilled. Our proposal is now to take instead of $\widetilde{C}(Z, \theta)$ another function $Q(Z, \beta)$, which is independent of the nuisance parameter η - but which has the same expectation as $\widetilde{C}(Z, \theta)$:

$$E_{\theta_0} Q(Z, \beta) = C_n(\theta_0, \theta). \tag{8}$$

Under

$$Q(Z, \beta) = \frac{1}{n} \sum_{i=1}^{n} q(Z_i, \beta), \quad C_n(\theta_0, \theta) = \frac{1}{n} \sum_{i=1}^{n} c(\eta_i, \beta)$$

it is sufficient to solve (**??**) for each summand separately. Rewrite this as deconvolution problem: we want $q(Z_i, \beta)$ with

$$\forall \beta \in \mathcal{B} \quad \forall \eta_i \in \mathcal{D} \qquad \int q(\eta_i + \varepsilon, \beta)\, p(\varepsilon)\, d\varepsilon = c(\eta_i, \beta), \tag{9}$$

where $p(\varepsilon)$ is the density of the i.i.d. errors in (**??**).

The new *corrected minimum contrast estimator* for the parameter of interest $\widetilde{\beta}$ is then defined as a measurable solution of

$$\widetilde{\beta} \in \arg\min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^{n} q(Z_i, \beta). \tag{10}$$

Because the explicit solution of (**??**) is not given in every case, we require the existence of an approximate solution only. We call $q_\mu(Z_i, \beta)$ an approximate solution of (**??**) if

$$\textit{for all } \mu > 0 \quad \sup_{\beta \in \mathcal{B}} \sup_{\eta_i \in \mathcal{D}} \left| \int q_\mu(\eta_i + \varepsilon, \beta)\, p(\varepsilon)\, d\varepsilon - c(\eta_i, \beta) \right| \leq \mu. \tag{11}$$

Then the related *approximate corrected minimum contrast estimator* for the parameter of interest $\widetilde{\beta}_\mu$ is defined as a measurable solution of

$$\widetilde{\beta}_\mu \in \arg\min_{\beta\in\mathcal{B}} \frac{1}{n}\sum_{i=1}^{n} q_\mu\left(Z_i,\beta\right). \tag{12}$$

3 Special cases

We present examples, where (**??**) is solvable or where we can find a solution of (**??**). In both cases we have to consider integral equations. The solutions have to be given explicitly for each parameter in order to determine the minimal point. The key point is to find combinations of contrast functions $c\left(\eta_i,\beta\right)$ fulfilling (i) and model functions G and error densities $p\left(\varepsilon\right)$, which have explicit solutions at least of (**??**). For details see Kukush and Zwanzig (1997), [**?**], and Kukush and Zwanzig (2001), [**?**].

3.1 Corrected least squares estimator

Let us consider the explicit model in (**??**) with $\Gamma=\sigma^2 I$ and with $q=2$ and

$$G\left(\eta_i,\beta\right)=\eta_{1i}-g(\eta_{2i},\beta). \tag{13}$$

We set for the contrast $C_n(\theta_0,\theta)$ the sum of squares

$$\frac{1}{n}\sum_{i=1}^{n}\left(g(\eta_{2i}^0,\beta)-g(\eta_{2i}^0,\beta_0)\right)^2. \tag{14}$$

Then the contrast condition (i) is of Jennrich type: for some $c_0>0$

$$\frac{1}{n}\sum_{i=1}^{n}\left(g(\eta_{2i}^0,\beta)-g(\eta_{2i}^0,\beta_0)\right)^2 \geq c_0\left\|\beta-\beta_0\right\|^2. \tag{15}$$

The corrected least squares estimator $\widetilde{\beta}_{LSE}$ is given as measurable solution of

$$\widetilde{\beta}_{LSE} \in \arg\min_{\beta\in\mathcal{B}} \frac{1}{n}\sum_{i=1}^{n}\left(\left(Z_{1i}-f(Z_{2i},\beta)\right)^2+h(Z_{2i},\beta)-f(Z_{2i},\beta)^2\right),$$

where f and h satisfy for all $\beta\in\mathcal{B}$ and all η_{2i}

$$E_{\eta_{2i}}f(Z_{2i},\beta)=g(\eta_{2i},\beta), \qquad E_{\eta_{2i}}h(Z_{2i},\beta)=g^2(\eta_{2i},\beta). \tag{16}$$

The estimator $\widetilde{\beta}_{LSE}$ is investigated in Kukush and Zwanzig (1997), [**?**] and in Fazekas and Kukush (1997), [**?**], see also Zwanzig(1998), [**?**]. In a polynomial explicit model $\widetilde{\beta}_{LSE}$ coincides with the adjusted least squares estimator proposed by Cheng and Schneeweiss (1998), [**?**], this was mentioned in Baran (2000), [**?**]. Astrometric applications of that estimator are considered in Zwanzig (1997), [**?**] and in de Vegt and Zwanzig (2001), [**?**].

3.2 An approximative corrected L2-estimator

Consider the model (**??**) with $g(\eta_{2i},\beta)=\beta\left|\eta_{2i}\right|,\beta\in[-d,d]$. Suppose that ε_{2i} is Laplace distributed, i.e. its density equals $\frac{1}{2}\exp(-\left|t\right|)$. We will apply the contrast (**??**). The first equation in (**??**), $Ef(Z_{2i})=\beta\left|\eta_{2i}\right|$, has the formal solution involving the generalized $\delta-$function: $f(z)=\beta(|z|-2\delta(z))$. We approximate the $\delta-$function by $\delta_{\mu}(z)=\mu^{-1}\varpi(\mu^{-1}z)$, where $\varpi(.)$ is a smooth probability density over $[-1,1]$. Then the approximate corrected least squares estimator is defined by

$$\widetilde{\beta}_{\mu}\in\arg\min_{\beta\in\mathcal{B}}\frac{1}{n}\sum_{i=1}^{n}\left(Z_{1i}^{2}-2Z_{1i}\beta(|Z_{2i}|-2\delta_{\mu}(Z_{2i}))+\beta^{2}(Z_{2i}^{2}-2)\right).$$

3.3 Regression functions of quotient form

Consider the explicit model (**??**) with a regression function of quotient form $g(\eta_{2i},\beta)=g_{1}(\eta_{2i},\beta)/g_{2}(\eta_{2i},\beta)$ and transform it to an implicit model with $G(\eta_{i},\beta)=g_{2}(\eta_{2i},\beta)\eta_{1i}-g_{1}(\eta_{2i},\beta)$. We set for the contrast $C_{n}(\theta_{0},\theta)$:

$$\frac{1}{n}\sum_{i=1}^{n}G(\eta_{i},\beta)^{2}.$$

The contrast condition (i) is fulfilled if $g_{2}(\eta_{2i},\beta)>0$ and if $g_{1}(\eta_{2i},\beta)$ satisfies the Jennrich type condition (**??**). Models of quotient form occur in biometry for growth curves and also in chemistry. The following model is used for the determination of $\beta=(r_{1},r_{2})$, where r_{1} and r_{2} are the reaction rates in the copolymerization process

$$g(\eta_{2i},\beta)=\frac{\eta_{2i}(r_{1}\eta_{2i}+1)}{r_{2}+\eta_{2i}}.$$

Assume that the moments $E\varepsilon_{ji}^{k}=m_{k}$ of the error distribution in (**??**) are known. Then a solution $q(Z_{i},\beta)$ of (**??**) is given by $Z_{i}=(y_{i},x_{i})$ and

$$r_{1}^{2}f_{4}(x_{i})-2r_{1}r_{2}f_{2}(x_{i})y_{i}+r_{2}^{2}f_{2}(y_{i})+2r_{1}f_{3}(x_{i})(1-y_{i})+2r_{2}x_{i}(f_{2}(y_{i})-1),$$

where $Ef_{k}(m+\varepsilon)=m^{k}$, especially

$$f_{2}(z)=z^{2}-\sigma^{2},\ \ f_{3}(z)=z^{3}-3z\sigma^{2}-m_{3},\ \ f_{4}(z)=z^{4}-6z^{2}\sigma^{2}-4zm_{3}+6\sigma^{4}-m_{4}$$

For details see Keeler and Reilly (1992), [**?**], Kukush and Zwanzig (1997), [**?**], and Zwanzig (2000), [**?**].

3.4 Corrected naive L1-estimators

Consider the model (**??**), where ε_{2i} has canonical Laplace distribution and $g(.,\beta)$ has continuous first and second derivatives $g^{\xi},g^{\xi\xi}$ and $\eta_{2i}\in[-a,a]$, for all i . The estimator is based on the L1- estimation in ordinary nonlinear regression. We will apply the contrast $C_n(\theta_0, \theta)$:

$$\frac{1}{n}\sum_{i=1}^{n} E_{\theta_0}\left|Z_{1i} - g(\eta_{2i}^0, \beta)\right|, \tag{17}$$

which satisfies the contrast condition (i), for details see Zwanzig (1997), [**?**]. Sufficient for the approximative corrected criteria function q_μ is the condition:

$$\sup_{Z_{1i}\in \mathrm{R}}\sup_{\beta\in\mathcal{B}}\sup_{\eta_{2i}\in[-a,a]}\left|\int q_\mu\left(Z_{1i}, \eta_{2i}+t, \beta\right)\frac{1}{2}\exp(-\left|t\right|)dt - \left|Z_{1i} - g(\eta_{2i}, \beta)\right|\right| \leq \mu. \tag{18}$$

The formal "exact" solution of (**??**) is

$$q_0\left(Z_i, \beta\right) = \left|R_i\right| - g^{\xi\xi}\ \ sign\left(R_i\right) - 2\left(g^{\xi}\right)^2\delta\left(R_i\right),$$

where the derivatives are taken at (Z_{2i}, β) and where $R_i = Z_{1i} - g(Z_{2i}, \beta)$ is the naive residuum. We introduce approximations for the $\delta-$function, for the $sign(.)$ and for the $|.|$ function as well:

$$\delta_\mu\left(z\right) = \mu^{-1}\varpi(\mu^{-1}z),\ \ \Delta_\mu(s) = 2\int_0^s\delta_\mu\left(z\right)dz,\ \ A_\mu(t) = \int_0^t\Delta_\mu\left(s\right)ds,$$

where $\varpi(z)$ is a smooth even probability density over $[-1, 1]$, fulfilling some regularity conditions. Then the approximate function satisfying (**??**) is

$$q_\mu\left(Z_i, \beta\right) = A_\mu(R_i) - g^{\xi\xi}\Delta_\mu\left(R_i\right) - 2\left(g^{\xi}\right)^2\delta_\mu\left(R_i\right).$$

3.5 Corrected symmetric L1 approach

Consider the model as above but we require additionally, that both errors are Laplace distributed. As contrast $C_n(\theta_0, \theta)$ function we choose now the empirical L1 distance between regression functions

$$\frac{1}{n}\sum_{i=1}^{n}\left|g(\eta_{2i}^0, \beta) - g(\eta_{2i}^0, \beta_0)\right|.$$

The respected deconvolution problem is to find q_μ with

$$\sup_{\beta\in\mathcal{B}}\sup_{\eta_{2i}^0\in[-a,a]}\left|E_{\theta_0}q_\mu\left(Z_i, \beta\right) - \left|g(\eta_{2i}^0, \beta) - g(\eta_{2i}^0, \beta_0)\right|\right| \leq\ \mu. \tag{19}$$

The formal "exact" solution of (**??**) can also be given by generalized functions. Using the approximations above we get

$$\begin{aligned} q_\mu\left(Z_i, \beta\right) &= A_\mu(R_i) + g^{\xi\xi}\Delta_\mu\left(R_i\right) - 2\left(\left(g^{\xi}\right)^2 + 1\right)\delta_\mu\left(R_i\right) \\ &\quad -2g^{\xi\xi}\delta_\mu'\left(R_i\right) + 2\left(g^{\xi}\right)^2\delta_\mu''\left(R_i\right). \end{aligned}$$

4 Consistency results

Our main theoretical results are the strong consistency of the estimators defined in (**??**) and (**??**).

Introduce two moment assumptions on $\overline{q}_i(\beta) = q(Z_i,\beta) - E_{\theta_0} q(Z_i,\beta)$. They are related to the Whittle inequality, see Whittle (1960), [**?**].

(ii) There exist constants $c > 0$, $k \geq 1$, such that for all $\beta \in \mathcal{B}$

$$\frac{1}{n}\sum_{i=1}^{n}\left(E_{\theta_0}\overline{q}_i(\beta)^{2k}\right)^{\frac{1}{k}} \leq c^{\frac{1}{k}}.$$

(iii) There exist a random variable $M_{(n)}$, a constant C and a real number $k \geq 1$, such that for all n and for all $\beta, \beta' \in \mathcal{B}: E_{\theta_0} M_{(n)}{}^k \leq C$ and

$$\frac{1}{n}\sum_{i=1}^{n}\left(\overline{q}_i(\beta) - \overline{q}_i(\beta')\right)^2 \leq M_{(n)} \left\|\beta - \beta^{'}\right\|^2.$$

Theorem 1 *Suppose Condition (i), (ii), (iii) with fixed $k \geq 1$.*

1. *Then for all $\tau > 0$ and for all $n \geq 1$*

$$P_{\theta_0}\left(\left\|\widetilde{\beta} - \beta_0\right\| > \tau\right) \leq const\ \rho(\tau)^{-2k}\, n^{-\frac{2k^2}{2k+p}}$$

2. *If $k > \frac{1+\sqrt{1+2p}}{2}$ then*

$$\widetilde{\beta} \to \beta_0 \quad a.s.$$

Introduce the related versions of (ii) and (iii) for the normalized approximative corrected estimation function $\overline{q}_{i,\mu}(\beta) = q_\mu(Z_i,\beta) - E_{\theta_0} q_\mu(Z_i,\beta)$.

(iv) There exist constants $c > 0$, $1 < k < \infty$, $\gamma_1 = \gamma_1(k) > 0$, $\mu_0 > 0$ such that for all $\mu < \mu_0$ and for all β

$$\frac{1}{n}\sum_{i=1}^{n}\left(E_{\theta_0}\left|\overline{q}_{i,\mu}(\beta)\right|^{2k}\right)^{\frac{1}{k}} \leq c^{\frac{1}{k}}\mu^{-\frac{\gamma_1(k)}{k}}.$$

(v) There exist a random variable $M_{(n)}$, a constant C and a real number $k \geq 1$ and $\gamma_2 = \gamma_2(k) > 0$, $\mu_0 > 0$, such that for all n, $\mu < \mu_0$ and for all $\beta, \beta' \in \mathcal{B}: E_{\theta_0} M_{(n)}{}^k \leq C\ \mu^{-\gamma_2(k)}$ and

$$\frac{1}{n}\sum_{i=1}^{n}\left(\overline{q}_{i,\mu}(\beta) - \overline{q}_{i,\mu}(\beta')\right)^2 \leq M_{(n)} \left\|\beta - \beta^{'}\right\|^2.$$

Theorem 2 *Suppose (i), (iv), (v) with fixed $k \geq 1$ and $\gamma_2(k) \geq \gamma_1(k)$. Set in (**??**) for $a > 0$*

$$\mu = \mu(n) = a\, n^{-r}, \quad r = \left(\frac{2k^2}{2k+p} - \varkappa\right)\frac{2k}{2k\ \gamma_1(k)\ +\ p\ \gamma_2(k)}.$$

1. *If $\varkappa > 0$ then for all $\tau > 0$ and for all $n \geq \left(\frac{2a}{\rho(\tau)}\right)^{\frac{1}{r}}$*

$$P_{\theta_0}\left(\left\|\widetilde{\beta}_{\mu(n)} - \beta_0\right\| > \tau\right) \leq const\ \left(\rho\left(\tau\right) - an^{-r}\right)^{-2k} n^{-\varkappa}.$$

2. *If $\varkappa > 1$ and $k > \frac{1+\sqrt{1+2p}}{2}$, then*

$$\widetilde{\beta}_{\mu(n)} \to \beta_0 \quad a.s.$$

5 Proofs

The following lemma gives an inequality analoguos to (**??**) for the approximative criterion function.

Denote $S_n\left(\beta\right) = \frac{1}{n}\sum_{i=1}^{n} \overline{q}_{i,\mu}(\beta)$ and $\mathcal{B}\left(\tau\right) = \left\{\beta \in \mathcal{B} : \|\beta - \beta_0\| > \tau\right\}.$

Lemma 3 *Let suppose (i). Then for all $\tau > 0$*

$$P_{\theta_0}\left(\left\|\widetilde{\beta}_{\mu} - \beta_0\right\| > \tau\right) \leq P_{\theta_0}\left(\sup_{\beta\in\mathcal{B}(\tau)} \left|S_n\left(\beta\right)\right| > \rho(\tau) - 2\mu\right).$$

Proof. Let $Q_\mu\left(\beta\right) = \frac{1}{n}\sum_{i=1}^{n} q_\mu\left(Z_i, \beta\right)$ and $\Delta Q_\mu\left(\beta\right) = Q_\mu\left(\beta\right) - Q_\mu\left(\beta_0\right)$ and $\Delta C_n\left(\beta\right) = C_n\left(\eta_0, \beta\right) - C_n\left(\eta_0, \beta_0\right).$

Condition (i) implies $\inf_{\beta\in\mathcal{B}(\tau)} \Delta C_n\left(\beta\right) \geq \rho\left(\tau\right).$

From (**??**) we have $\left|\Delta C_n\left(\beta\right) - E_{\theta_0}\Delta Q_\mu\left(\beta\right)\right| \leq 2\mu$. Thus

$$\inf_{\beta\in\mathcal{B}(\tau)} E_{\theta_0}\Delta Q_\mu\left(\beta\right) \geq \rho\left(\tau\right) - 2\mu. \tag{20}$$

From the definition of $\widetilde{\beta}_\mu$ we have $\Delta Q_\mu\left(\widetilde{\beta}_\mu\right) \leq 0$. Under $\widetilde{\beta}_\mu \in \mathcal{B}\left(\tau\right),$ (**??**) implies that

$$\begin{aligned}\rho(\tau) - 2\mu &\leq \inf_{\beta\in\mathcal{B}(\tau)} E_{\theta_0}\Delta Q_\mu\left(\beta\right) - \inf_{\beta\in\mathcal{B}(\tau)} \Delta Q_\mu\left(\beta\right)\\ &\leq \sup_{\beta\in\mathcal{B}(\tau)} \left|\Delta Q_\mu\left(\beta\right) - E_{\theta_0}\Delta Q_\mu\left(\beta\right)\right|,\end{aligned}$$

which proves the lemma, for $S_n\left(\beta\right) = \Delta Q_\mu\left(\beta\right) - E_{\theta_0}\Delta Q_\mu\left(\beta\right).\square$

The next lemma delivers the uniform convergence of the approximative criterion function.

Lemma 4 *Assume (iv) and (v) with $\gamma_2(k) \geq \gamma_1(k)$. Then there exist a constant const, such that for all $\epsilon > 0, n \geq 1$*

$$P_{\theta_0}\left(\sup_{\beta\in\mathcal{B}} \left|S_n\left(\beta\right)\right| > \epsilon\right) \leq const\ \epsilon^{-2k}\ \mu^{-\frac{\gamma_1(k)\ 2k\ +\ \gamma_2(k)\ p}{2k+p}}\ n^{-\frac{2k^2}{2k\ +p}}.$$

Proof. Imbed $\mathcal{B}$ into a ball $K(\beta_0, R) = \{\beta : \|\beta - \beta_0\| \le R\}$. For $0 < d < C_1$ the compact set $\mathcal{B}$ can be covered by a finite number of balls $K\left(\beta^l, d\right)$, $l = 1, ..., N(d)$, with $N(d) \le C_2\, d^{-p}$, where C_1 and C_2 are constants. Then

$$\sup_{\beta\in\mathcal{B}} |S_n(\beta)| \le \max_{l=1,..,N(d)} \left|S_n\left(\beta^l\right)\right| + \sup_{\|\beta''-\beta'\|\le d} |S_n(\beta'') - S_n(\beta')| = S_1 + S_2.$$

$$P = P_{\theta_0}\left(\sup_{\beta\in\mathcal{B}} |S_n(\beta)| \ge \epsilon\right) \le \sum_{l=1}^{N(d)} P_{\theta_0}\left(\left|S_n\left(\beta^l\right)\right| \ge \frac{\epsilon}{2}\right) + P_{\theta_0}\left(S_2 \ge \frac{\epsilon}{2}\right) = P_1 + P_2. \tag{21}$$

We apply the Whittle inequality for a real number $s, s \ge 2$, then

$$P_1 \le const\, d^{-p} n^{-k} \mu^{-\gamma_1(k)} \epsilon^{-2k}.$$

From condition (v) we have $S_2^2 \le M_{(n)} d^2$ and applying the Chebychev inequality we obtain

$$P_2 \le \epsilon^{-2k} 4^k d^{2k} E M_{(n)}^k \le const\ \epsilon^{-2k} \mu^{-\gamma_2(k)} d^{2k}.$$

Summarizing we obtain $P \le const\, \epsilon^{-2k}\left[\mu^{-\gamma_1(k)} d^{-p}\ n^{-k} + \mu^{-\gamma_2(k)} d^{2k}\right]$. Choosing d in an optimal way by $d_{opt}^{2k+p} = const_1\ n^{-k} \mu^{(\gamma_2(k)-\gamma_1(k))} \le const_2$, we obtain the statement. □

Proof of Theorem 2:

For $\mu = an^{-r},\ n > \left(\frac{2a}{\rho(\tau)}\right)^{\frac{1}{r}}$ we have $\rho(\tau) - 2\mu > 0$. Lemma 3 implies

$$P_{\theta_0}\left(\left\|\widetilde{\beta}_\mu - \beta_0\right\| > \tau\right) \le P_{\theta_0}\left(\sup_{\beta\in\mathcal{B}} |S_n(\beta)| \ge \frac{1}{2}(\rho(\tau) - 2\mu)\right).$$

We apply Lemma 4 with $2\epsilon = \rho(\tau) - 2\mu$ and get the first statement of Theorem 2. The second statement follows by the Borel Cantelli Lemma.□

Proof of Theorem 1: The proof goes along the line above. We can interpret condition (ii) as condition (iv) with $\gamma_1(k) = 0$ and condition (iii) as condition (v) with $\gamma_1(k) = 0$. Therefore the following analogue to Lemma 4 holds

$$P_{\theta_0}\left(\sup_{\beta\in\mathcal{B}} |S_n(\beta)| > \epsilon\right) \le const\ \epsilon^{-2k}\ n^{-\frac{2k^2}{2k+p}}.$$

By Lemma 3 with $\mu = 0$ we get Theorem 1.□

References

[1] Baran, S. (2000) A consistent estimator in general functional errors-in-variables model. Metrika 51 (2), 117-132.

[2] Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). Measurement Errors in Nonlinear Models. Chapman and Hall, London.

[3] Cheng, C.-L. and Schneeweiss, H. (1998). The polynomial regression with errors in the variables. J. R. Statist. Soc. B, 60 (1), 189-199.

[4] de Vegt, Ch. and Zwanzig, S. (2001). Comparison of reference frame linking methods, Preprint-No. 2001-03, April 2001, University of Hamburg, Schwerpunkt Mathematische Statistik und Stochastische Prozesse, 23p..

[5] Fazekas I. and Kukush, A. (1997) Asymptotic properties of an estimator in nonlinear functional errors-in-variables models. Computers Math. Appl. 34 (10), 23-39.

[6] Keeler, S.E. and Reilly, P.M. (1992) The design of experiments when there are errors in all the variables. Can. J. Chem. Eng. 70, 774-778.

[7] Kukush, A. and Zwanzig, S. (1997) On a corrected contrast estimator in the implicit nonlinear functional relation model. Preprint No. 97-11, Institute of Mathematical Stochastics. University of Hamburg.

[8] Kukush, A. and Zwanzig, S. (2001) On adaptive minimum contrast estimators in the implicit nonlinear functional regression models. Ukr. Math. Journal. 53, (9).

[9] Neyman,J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. Econometrica 16 (1), 1-32.

[10] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. Theory Prob. and Appl. 5, 302-305.

[11] Zwanzig, S. (1997). On L1-norm estimators in nonlinear regression and in nonlinear error-in-variables models. In: IMS Lecture Notes-Monograph Series, 31, 101-118.

[12] Zwanzig, S. (1997): Application of Hipparcos dates: A new statistical method for star reduction. Journees 1997, Academy of Sciences of the Czech Republic, Astronomical Institute, p.142-144.

[13] Zwanzig, S.(1998) Estimation in nonlinear functional error-in-variables models. Habilitationsschrift. University of Hamburg. Department of Mathematics.

[14] Zwanzig, S. (2000) Experimental design for nonlinear error-in-variables models, in "Advances in Stochastic Simulation Methods" Ed. Balakrishnan, Melas, Ermakov. p. 153-163.