Friedrich Liese; Igor Vajda

M-estimators of structural parameters in pseudolinear models

Applications of Mathematics, Vol. 44 (1999), No. 4, 245–270

Persistent URL: `http://dml.cz/dmlcz/134412`

M-ESTIMATORS OF STRUCTURAL PARAMETERS IN PSEUDOLINEAR MODELS[1]

Friedrich Liese, Rostock, and Igor Vajda, Praha

(Received September 2, 1997)

Abstract. Real valued M-estimators $\hat\theta_n := \min \sum_1^n \varrho(Y_i - \tau(\theta))$ in a statistical model with observations $Y_i \sim F_{\theta_0}$ are replaced by $\mathbb{R}^p$-valued M-estimators $\hat\beta_n := \min \sum_1^n \varrho(Y_i - \tau(u(z_i^T \beta)))$ in a new model with observations $Y_i \sim F_{u(z_i^t \beta_0)}$, where $z_i \in \mathbb{R}^p$ are regressors, $\beta_0 \in \mathbb{R}^p$ is a structural parameter and $u : \mathbb{R} \to \mathbb{R}$ a structural function of the new model. Sufficient conditions for the consistency of $\hat\beta_n$ are derived, motivated by the sufficiency conditions for the simpler "parent estimator" $\hat\theta_n$. The result is a general method of consistent estimation in a class of nonlinear (pseudolinear) statistical problems. If F_θ has a natural exponential density $\mathrm{e}^{\theta x - b(x)}$ then our pseudolinear model with $u = (g \circ \mu)^{-1}$ reduces to the well known generalized linear model, provided $\mu(\theta) = \mathrm{d}b(\theta)/\mathrm{d}\theta$ and g is the so-called link function of the generalized linear model. General results are illustrated for special pairs ϱ and τ leading to some classical M-estimators of mathematical statistics, as well as to a new class of generalized α-quantile estimators.

Keywords: M-estimator, generalized linear models, pseudolinear models

MSC 2000: 62F10, 62F12, 62F35

1. Introduction

We consider a sequence of random vectors $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$ with independent components $Y_i \sim F_{\theta_0}$ where the distribution function F_{θ_0} is from a family $\mathscr{F} = \{F_\theta\colon \theta \in \Theta\}$ parametrized by means of an interval $\Theta \subset \mathbb{R}$. A sequence of Θ-valued measurable functions $\hat\theta_n = \hat\theta_n(\mathbf{Y}_n)$ is an *M-estimator* if

$$\hat\theta_n \in \operatorname{argmin}_{\theta\in\Theta} M_n(\theta) \text{ a.s.} \quad \text{for} \quad M_n(\theta) = \frac{1}{n}\sum_{i=1}^n \varrho(Y_i - \tau(\theta)),$$

[1] Supported by the grants AS CR 2075703 and GA CR 102/99/1137.

where $\varrho\colon \mathbb{R} \to [0,\infty)$ and $\tau\colon \mathbb{R} \to \mathbb{R}$. The last condition can be rewritten into the form

$$\hat{\theta}_n \in \operatorname{argmin}_{\theta\in\Theta} \int_{\mathbb{R}} \varrho(y-\tau(\theta))\,\mathrm{d}\hat{F}_n(y), \tag{1}$$

where

$$\hat{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{(Y_i,\infty)}(y)$$

is the empirical distribution function.

Let us consider the function

$$M^*_{\theta_0}(\theta) = \to E\, M_n(\theta) = \int_{\mathbb{R}} \varrho(y-\tau(\theta))\,\mathrm{d}F_{\theta_0}(y) \tag{2}$$

of variables θ_0, $\theta \in \Theta$. In Section 2 regularity conditions are formulated under which $M^*_{\theta_0}(\theta)$ is for every $\theta_0 \in \Theta$ continuous in θ. By Theorem 1 in Section 3 below, under these regularity conditions, the estimator (1) is consistent if for every $\theta_0 \in \Theta$

(A) the function (2) satisfies the condition

$$\inf_{|\theta-\theta_0|>\varepsilon} M^*_{\theta_0}(\theta) > M^*_{\theta_0}(\theta_0) \quad \text{for all } \varepsilon > 0$$

and

(B) there exists $c = c(\theta_0) > 0$ such that

$$\lim_n \to P\Big(\inf_{|\theta-\theta_0|>c} M_n(\theta) > M_n(\theta_0)\Big) = 1.$$

Note that the continuity of ϱ and τ assumed in Section 2 implies that the infima considered in (B) are measurable functions of $\mathbf{Y}_n$ (cf. e.g. Liese and Vajda [8]).

For estimators defined by minimization or maximization, the proof of consistency is a first step necessary for further asymptotic analysis. While this step seems to tell us relatively little about the asymptotics of the estimators, it is often formally much more difficult than the steps leading to more detailed information about the asymptotics such as e.g. the asymptotic linearity or asymptotic normality. The case where the ϱ-function is convex is an exception. Then the pointwise convergence of $M_n(\theta)$ to $M^*_{\theta_0}(\theta)$ implies the locally uniform convergence and the convexity of $M_{\theta_0}(\theta)$. The locally uniform convergence together with the identifiability (A) already implies consistency. Convex ϱ-functions were considered e.g. by Koenker amd Basset [6], Pollard [14], Morgenthaler [10], Jurečková and Procházka [4], and Zwanzig [16]. Authors dealing with nonconvex $M_n(\theta)$ impose some verifiable regularity conditions guaranteeing the locally uniform convergence of $M_n(\theta)$, see e.g. Pollard [12], Pfanzagl [11], Liese and Vajda [7, 8], van der Vaart and Wellner [15]. The condition (B) is then needed too, because it guarantees that the estimates are asymptotically localized in a compact neighborhood of the true parameter θ_0. In our model this condition is not needed in cases where the special structure enables us to avoid the need of such a localization. In the remaining cases we formulate verifiable regularity conditions which imply (B).

In this paper we verify (A), (B) for several M-estimators (functions ϱ and τ) and families $\mathscr{F}$. In particular, we do it for the L_2-estimators defined by $\varrho(y) = y^2$ and L_α-estimators, $0 < \alpha < 1$, defined by

$$\varrho_\alpha(y) = |y| \big(\alpha\, 1_{[0,\infty)}(y) + (1-\alpha)\, 1_{(-\infty,0)}(y)\big) \tag{3}$$

(α-quantile estimators, least absolute deviation estimators if $\alpha = \frac{1}{2}$), and for the natural exponential families defined by densities

$$f_\theta(y) = \frac{\mathrm{d}F_\theta(y)}{\mathrm{d}\nu(y)} = \mathrm{e}^{\theta y - b(\theta)},$$

where ν is a σ-finite measure and

$$b(\theta) = \ln \int_{\mathbb{R}} \mathrm{e}^{\theta y}\, \mathrm{d}\nu(y)$$

is finite on $\Theta \subset \mathbb{R}$. The set $\{\theta \in \mathbb{R}\colon\ 0 < \int \mathrm{e}^{\theta y}\, \mathrm{d}\nu(y) < \infty\}$ as well as the function $b(\theta)$ are known to be convex, and in the interior of this set there exist all derivatives of $b(\theta)$ and all moments of the natural exponential distributions. If $b(\theta)$ is strictly convex in the interior of Θ, then the mean

$$\mu(\theta) = \int_{\mathbb{R}} y\, \mathrm{d}F_\theta(y) = \frac{\mathrm{d}b(\theta)}{\mathrm{d}\theta} \tag{4}$$

is increasing and continuous on Θ. Since the family $\mathscr{F}$ is stochastically increasing, the quantile functions

$$\mu_\alpha(\theta) = F_\theta^{-1}(\alpha), \quad 0 < \alpha < 1, \tag{5}$$

are nondecreasing. Under certain assumptions concerning $b(\theta)$ and the dominating measure ν, also the quantile functions are continuous and increasing on Θ. For the L_2-estimators it is assumed that $\tau(\theta) = \mu(\theta)$ and for the L_α-estimators, $\tau(\theta) = \mu_\alpha(\theta)$.

The consistency of M-estimators in families $\mathscr{F} = (F_\theta\colon\ \theta \in \Theta)$ based on conditions (A), (B) can be extended to statistical models with observables $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ where the components are independent, given by the formula

$$Y_i \sim F_{\theta_i}, \qquad \theta_i = u(z_i^t\,\beta_0), \tag{6}$$

for a vector parameter $\beta_0 = (\beta_{01}, \dots, \beta_{0p})^t \in B$, $B \subset \mathbb{R}^p$, (nonrandom) regressors $z_i = (z_{i1}, \dots, z_{ip})^t \in \mathscr{X}$, $\mathscr{X} \subset \mathbb{R}^p$, and a function $u\colon \mathbb{R} \to \Theta$. Statistical models of this pseudolinear structure will be called *pseudolinear models* with *structural functions* u and *structural parameters* β_0.

Our concept of pseudolinear model is wider than the classical *generalized linear model* of mathematical statistics. The latter is using formula (6) too but it assumes that $\mathscr{F}$ is a natural exponential family with a one-to-one mean function $\mu(\theta)$. The structural function is then defined by the formula $u = (g \circ \mu)^{-1}$, where g is an injective *link function* ($g = \mu^{-1}$ is the so-called *natural link function*, see Fahrmeir and Kaufmann [2]). If e.g. F_θ is normal with mean θ and variance 1, i.e. if we have a natural exponential family with $b(\theta) = \theta^2/2$ and the dominating measure $\nu \sim N(0,1)$, then the classical generalized linear model (6) reduces to the model of nonlinear regression

$$Y_i = u(z_i^t\,\beta_0) + \varepsilon_i, \quad \to E\,\varepsilon_i = 0, \quad \to E\,\varepsilon_i^2 = 1, \tag{7}$$

where the errors ε_i are independent normal (the natural link function leads to linear regression). If the errors are independent logistic then the regression model (7) still satisfies (6) for the logistic family

$$F_\theta(y) = \left(1 + \mathrm{e}^{-\pi(y-\theta)/\sqrt{3}}\right)^{-1}, \quad \theta \in \mathbb{R}, \tag{8}$$

i.e. it remains to be pseudolinear but, since the logistic family is not natural exponential, it is not generalized linear. Similarly, a linear model for scale in the logistic family is not generalized linear. It is even not a regression model.

These examples at the same time demonstrate that the classes of pseudolinear models and of the models of nonlinear regression have a nonempty intersection, and that neither of these classes is included in the other.

By an *M-estimator of structural parameters* β_0 in a generalized linear model we mean a sequence of $\mathbb{R}^p$-valued measurable functions $\hat\beta_n = \hat\beta_n(\mathbf{Y}_n, \mathbf{z}_n)$ of observables $\mathbf{Y}_n$ and $p \times n$ regressor matrices $\mathbf{z}_n = (z_1, \dots, z_n)$ such that

$$\hat\beta_n \in \operatorname{argmin} \frac{1}{n} \sum_{i=1}^n \varrho(Y_i - \tau \circ u(z_i^t\,\beta)), \tag{9}$$

where the minimization extends over all $\beta = (\beta_1, \dots, \beta_p)^t \in B$.

The motivation of M-estimators, in particular of the L_α-estimators, $0 < \alpha \leqslant 1$, is in our case the same as in the case of linear and nonlinear regression (cf. e.g. Pollard [14], Morgenthaler [10], Jurečková and Procházka [4]). Our paper extends the method of M-estimation to some of the models where the measurements are distorted by random errors in a manner which is not necessarily additive.

In this paper we present alternatives to conditions (A), (B) which are sufficient for the consistency of M-estimators of structural parameters of generalized linear models. Our results extend the consistency theorems in Fahrmeir and Kaufman [2], Liese and Vajda [7, 8], Jurečková and Procházka [4] and some other papers. Note that for consistent M-estimators one can establish the asymptotic normality relatively easily by using the methods developed recently by Pollard [12, 14] and others (see Chap. 7 in Pfanzagl [11] or Sec. 3.2 in van der Vaart and Wellner [15]). Thus in some sense the present paper opens the possibility to study also the higher order asymptotic properties of estimators under consideration, like the efficiency and robustness.

2. Regularity conditions

In this section we formulate regularity conditions imposed in the sequel on the estimators and models under consideration.

(E1) The function $\varrho\colon \mathbb{R} \to [0,\infty)$ is continuous and satisfies for every $y \in \mathbb{R}$ and $s > 0$ the relation

$$\sup_{|t|\leqslant s} \varrho(y+t) \leqslant \varrho(y+s) + \varrho(y-s). \tag{10}$$

Obviously, (E1) follows from the following stronger condition.

(E1+) $\varrho\colon \mathbb{R} \to [0,\infty)$ is continuous, nondecreasing on $(0,\infty)$ and nonincreasing on $(-\infty,0)$, with $\varrho(0)=0$.

(E2) The function $\tau\colon \mathbb{R} \to \mathbb{R}$ is strictly monotone and continuous.

In the sequel we consider for $N > 0$ the functions

$$\chi_N(y) = 1_{[-N,N]}(y), \quad y \in \mathbb{R}, \tag{11}$$

and

$$\varphi_N(y,t) = \varrho(y-t)\,\chi_N(y), \quad y \in \mathbb{R},\ t \in \mathbb{R}. \tag{12}$$

(EM1) For every $t \in \mathbb{R}$ and $a < b$ belonging to the interval Θ we have

$$\lim_{N\to\infty} \sup_{a<\theta<b} \int \varrho(y-t)\,(1-\chi_N(y))\,\mathrm{d}F_\theta(y) = 0.$$

In the next condition we need the function

$$\overline{\varrho}(y) = \min\{\varrho(y), \varrho(-y)\}, \quad y \geqslant 0, \tag{13}$$

and the extended real valued constants

$$\varrho(\infty) = \lim_{y\to\infty} \varrho(y) \quad \text{and} \quad \varrho(-\infty) = \lim_{y\to-\infty} \varrho(y). \tag{14}$$

If (E1+) holds then $\varrho(\infty)$ and $\varrho(-\infty)$ exist and we can extend the definition (13) to $y = \infty$. Under (E1+) we consider the following condition.

(EM2) We have $\varrho(\infty) = \varphi(-\infty) = \overline{\varrho}(\infty)$ and

$$\int \varrho(y - \tau(\theta_0))\mathrm{d}F_{\theta_0}(y) < \overline{\varrho}(\infty).$$

This condition follows automatically from (E1+) when $\varrho(\infty) = \varrho(-\infty) < \infty$ and $\varrho(y) < \overline{\varrho}(\infty)$ for all $y \in \mathbb{R}$. By Lemma 1 below, it follows also from (E1+), (E2) and (EM1) when $\varrho(\infty) = \varrho(-\infty) = \infty$.

(E1+), (E2), (EM1) and (EM2) hold for practically all M-estimators and families $\mathscr{F}$ considered in statistics (cf. Lehman [9] and Hampel et al [3]). Next we give several conditions concerning the model which are also satisfied in typical situations.

(M1) The family $\mathscr{F}$ is weakly continuous, i.e. the mapping $\theta \to F_\theta$ is continuous in the sense of weak convergence of distributions.

(M2) The structural function $u\colon \mathbb{R} \to \Theta$ is strictly monotone and continuous.

(M3) All regressors $z_1, z_2, \ldots$ belong to a closed set $\mathscr{Z} \subset \mathbb{R}^p$ bounded in the norm by $\gamma > 0$.

(M4) The probability measure

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{z_i}$$

defined on $\mathscr{Z}$ by the Dirac measures concentrated on regressors z_i tends for $n \to \infty$ weakly to a probability measure μ on $\mathscr{Z}$, in symbols

$$\mu_n \models\Longrightarrow \mu.$$

Note that (M4) in some sense follows from (M3). Indeed, under (M3) there exists a probability measure μ and a subsequence μ_{n_k} such that $\mu_{n_k} \models\Longrightarrow \mu$ for $k \to \infty$.

Next, an important auxiliary result follows.

Lemma 1. *If* (E1), (E2) *and* (EM1) *hold then the functions* $M^*_{\theta_0}(\theta)$, $\theta_0 \in \Theta$, *are finitely valued and continuous on* Θ.

P r o o f. Let θ_0, $\theta_* \in \Theta$ be arbitrary fixed. The finiteness of integrals

$$\int \varrho(y-t)\,\mathrm{d}F_{\theta_0}(y), \quad t \in \mathbb{R},$$

follows from (EM1). If $\theta_n \in \Theta$ is a sequence tending to θ_* as $n \to \infty$ then the continuity in (E1) and (E2) implies that the functions $\varrho(y-\tau(\theta_n))$ tend pointwise to $\varrho(y-\tau(\theta_*))$ and (10) implies that

$$\sup_{|t|<s} \varrho(y-\tau(\theta_*)+t)$$

is an F_{θ_0}-integrable majorant for all functions $\varrho(y-\tau(\theta_n))$ with sufficiently large index n. □

3. Parent M-estimators

Let the domain B of possible values of structural parameters considered in (6) and (9) be open in $\mathbb{R}^p$, and let

$$\Theta = \{u(z^t\beta) \colon\ z \in \mathscr{Z},\ \beta \in B\}$$

be an open interval in $\mathbb{R}$.

The family $\mathscr{F}$ is a *parent model* for the generalized linear model given by (6) and by the respective regressors, structural parameters, and structural function. Obviously, in the particular one-dimensional case $B \subset \mathbb{R}$ with constant regressors $z_1 = z_2 = \ldots = z \in \mathbb{R}$ and structural function $u(y) = y$ the pseudolinear model reduces to the parent model with the rescaled parameter $\theta = z\beta$, $\beta \in B$.

Similarly, the estimator $\hat{\theta}_n$ of the parameter of the parent model given by (1) is a *parent M-estimator* of the estimator given by (9). Let us recall that in (1) it is assumed that

$$Y_i \sim F_{\theta_0}$$

for the independent observations $Y_1, \ldots, Y_n$ figuring in the definition of $M_n(\theta)$. Thus in this section the convergence in probability is considered with respect to the probability measure defined in the usual way by the distribution function $F_{\theta_0}(y)$.

It is obvious that the consistency of parent M-estimators is necessary for the consistency of all the corresponding M-estimators of structural parameters. Therefore we start with the simpler problem of consistency of parent M-estimators (1). The main result of this section is based on the following lemma.

Lemma 2. *Let* (E1), (E2) *and* (EM1) *hold. Then for every* $c > 0$

$$\sup_{|\theta-\theta_0|\leqslant c} |M_n(\theta) - M^*_{\theta_0}(\theta)| \xrightarrow{P} 0 \quad \textit{as } n \to \infty.$$

If τ *is bounded on* Θ *then*

$$\sup_{\theta\in\Theta} |M_n(\theta) - M^*_{\theta_0}(\theta)| \xrightarrow{P} 0 \quad \textit{as } n \to \infty.$$

Proof. (I) Let τ be bounded on Θ. By using the assumed nonnegativity of ϱ and (E1), one obtains for $s = \sup\{|\tau(\theta)|\colon \theta \in \Theta\}$ and for the function (11) with arbitrary $N > 0$ the inequality

$$0 \leqslant W_n(N) \triangleq \sup_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{n} \varrho(Y_i - \tau(\theta))\,[1 - \chi_N(Y_i)]$$
$$\leqslant \frac{1}{n} \sum_{i=1}^{n} [\varrho(Y_i - s) + \varrho(Y_i + s)]\,[1 - \chi_N(Y_i)].$$

Therefore (EM1) implies

$$\lim_{N\to\infty} \Bigl(\sup_n \to E\,W_n(N)\Bigr) = 0.$$

Further, (E2) implies that τ is strictly monotone on the interval Θ. Thus the boundedness of τ implies for every $\delta > 0$ the existence of a finite net $\theta_1, \ldots, \theta_k$ in Θ decomposing Θ by the nearest neighbor rule into the so-called Voronoi cells $V(\theta_1), \ldots, V(\theta_k)$ such that

$$\sup_{\theta\in V(\theta_j)} |\tau(\theta) - \tau(\theta_j)| < \delta, \quad 1 \leqslant j \leqslant k.$$

Let now N be fixed. Denote by T the interval of values of the mapping τ. Since the function $\varphi_N(y, t)$ of (12) is continuous on $[-N, N] \times T$, the last result means that for any $\varepsilon > 0$ one can find a finite net $\theta_1, \ldots, \theta_k$ in Θ such that

$$\sup_{y\in\mathbb{R}\,:\,\theta\in V(\theta_j)} |\varphi_N(y, \tau(\theta)) - \varphi_N(y, \tau(\theta_j))| < \varepsilon, \quad 1 \leqslant j \leqslant k.$$

Finally, for each $1 \leqslant j \leqslant k$ the bounded random variables $X_i = \varphi_N(Y_i, \tau(\theta_j))$, $1 \leqslant i \leqslant n$, satisfy the law of large numbers

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \to E\,X_i) \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

By combining the last three facts, one obtains the desired uniform law of large numbers:

$$\sup_{\theta\in\Theta}|M_n(\theta)-M^*_{\theta_0}(\theta)| = \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}[\varrho(Y_i-\tau(\theta))-\!\!\longrightarrow E\,\varrho(Y_i-\tau(\theta))]\right| \xrightarrow{P} 0 \quad \text{as } n\to\infty.$$

(II) Let now τ be unbounded on Θ. By (E2), it is bounded on the bounded set $\Theta_0 = \{\theta \in \Theta\colon\ |\theta-\theta_0| \leqslant c\}$ for any $c > 0$. Thus all steps of part (I) are applicable with Θ replaced by Θ_0. □

Theorem 1. *Let* (E1), (E2) *and* (EM1) *hold. Then* (A) *and* (B) *imply the consistency of estimator* (1). *If τ is bounded on Θ then* (A) *alone implies this consistency.*

Proof. (I) Let τ be bounded and let $\varepsilon > 0$ be arbitrary. Then

$$\{|\hat\theta_n-\theta_0|>\varepsilon\} \subset \Big\{\inf_{|\theta-\theta_0|>\varepsilon} M_n(\theta) \leqslant M_n(\theta_0)\Big\} \subset A_n,$$

where

$$A_n = \Big\{\inf_{|\theta-\theta_0|>\varepsilon} M^*_{\theta_0}(\theta) \leqslant M^*_{\theta_0}(\theta_0) + 2\sup_{\theta\in\Theta}|M_n(\theta)-M^*_{\theta_0}(\theta)|\Big\}.$$

Since (A) implies

$$\inf_{|\theta-\theta_0|>\varepsilon} M^*_{\theta}(\theta) - M^*_{\theta_0}(\theta_0) > 0,$$

the desired relation $\lim_n \rightarrow P(A_n) = 0$ follows from Lemma 2.

(II) Let now τ be unbounded and $\delta > 0$ arbitrary. Choose any ε between 0 and c figuring in (B). Similarly as in part (I),

$$\{|\theta_n-\theta_0|>\varepsilon\} \subset \{|\hat\theta_n-\theta_0|>c\} \cup \{\varepsilon<|\hat\theta_n-\theta_0|\leqslant c\}.$$

Since

$$\{|\hat\theta_n-\theta_0|>c\} \subset \Big\{\inf_{|\theta-\theta_0|>c} M_n(\theta) \leqslant M_n(\theta_0)\Big\},$$

(B) implies for the first event on the right-hand side

$$\lim_n \rightarrow P(|\hat\theta_n-\theta_0|>c) = 0.$$

The second right-hand side event can be bounded as follows:

$$\{\varepsilon<|\hat\theta_n-\theta_0|\leqslant c\} \subset \Big\{\inf_{\varepsilon<|\theta-\theta_0|\leqslant c} M_n(\theta) < M_n(\theta_0)\Big\} \subset \tilde A_n$$

where

$$\tilde{A}_n = \Big\{ \inf_{\varepsilon < |\theta - \theta_0| \leqslant c} M^*_{\theta_0}(\theta) < M^*_{\theta_0}(\theta_0) + 2 \sup_{|\theta - \theta_0| \leqslant c} |M_n(\theta) - M^*_{\theta_0}(\theta)| \Big\}.$$

Since by Lemma 2

$$\sup_{|\theta-\theta_0|\leqslant c} |M_n(\theta) - M^*_{\theta_0}(\theta)| \xrightarrow{P} 0 \quad \text{as } n \to \infty,$$

the same argument as in part (I) leads to the desired relation $\lim_n \to P(\tilde{A}_n) = 0$. □

The identifiability condition (A) must be verified separately for each M-estimator (i.e. for each pair ϱ and τ) and each parent model (i.e. family $\mathscr{F}$). If τ is unbounded on Θ, then one can prove a relatively weak sufficient condition for (B).

Lemma 3. *If τ is unbounded on Θ then the conditions* (E1+), (E2), (EM1) *and* (EM2) *imply* (B).

Proof. For τ unbounded on Θ the set $S_N = \{\theta \in \Theta\colon |\tau(\theta) - \tau(\theta_0)| \geqslant N\}$ is nonempty for every $N > 0$. By (E2), the open interval Θ must be unbounded and S_N is either a semiclosed unbounded interval, or the union of two disjoint semiclosed intervals of which at least one is unbounded. Define

$$\gamma_{\theta_0}(N) = \begin{cases} \theta_0 - a_N & \text{if } S_N = (-\infty, a_N] \\ b_N - \theta_0 & \text{if } S_N = [b_N, \infty) \\ \max\{\theta_0 - a_N, b_N - \theta_0\} & \text{if } S_N = (a, a_N] \cup [b_N, b), \end{cases}$$

where $-\infty \leqslant a < a_N < b_N < b \leqslant \infty$. Then $|\theta - \theta_0| \geqslant \gamma_{\theta_0}(2N)$ implies $|\tau(\theta) - \tau(\theta_0)| \geqslant 2N$. Further, if the function

$$d(y,\theta) = y - \tau(\theta), \quad y \in \mathbb{R},\ \theta \in \Theta,$$

satisfies for some $y \in \mathbb{R}$ and $\theta_0 \in \Theta$ the condition

$$|d(y,\theta_0)| < N \quad (\text{i.e. } \chi_N(d(y,\theta_0)) = 1, \text{ cf. (11)})$$

then $|\theta - \theta_0| \geqslant \gamma_{\theta_0}(2N)$ implies

$$|d(y,\theta)| \geqslant |\tau(\theta) - \tau(\theta_0)| - |d(y,\theta_0)| \geqslant N.$$

Finally, (E1+) implies that the function (13) satisfies for all $d \in \mathbb{R}$ with $|d| \geqslant N$ the inequality

$$\varrho(d) \geqslant \overline{\varrho}(N). \tag{15}$$

Therefore, if for an arbitrary fixed $N > 0$ we define random variables

$$W_i = \chi_N(d(Y_i, \theta_0)), \quad 1 \leqslant i \leqslant n,$$

then for $|\theta - \theta_0| \geqslant \gamma_{\theta_0}(2N)$, (15) implies

$$\begin{aligned} M_n(\theta) &= \frac{1}{n}\sum_{i=1}^n \varrho(d(Y_i, \theta)) \geqslant \frac{1}{n}\sum_{i=1}^n W_i \, \varrho(d(Y_i, \theta)) \\ &\geqslant \frac{1}{n}\sum_{i=1}^n W_i \, \overline{\varrho}(N) = \overline{\varrho}(N)\frac{1}{n}\sum_{i=1}^n W_i. \end{aligned}$$

Consequently,

$$\inf_{|\theta-\theta_0| \geqslant \gamma_{\theta_0}(2N)} M_n(\theta) \geqslant \overline{\varrho}(N)\frac{1}{n}\sum_{i=1}^n W_i. \tag{16}$$

Here, by the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^n W_i \xrightarrow{P} w(N) \quad \text{as } n \to \infty,$$

where

$$w(N) = \rightarrow E\, W = \rightarrow E\, \chi_N(Y - \tau(\theta_0)) = F_{\theta_0}(\tau(\theta_0) + N) - F_{\theta_0}(\tau(\theta_0) - N). \tag{17}$$

It follows from (14) and (17) that

$$\lim_{N \to \infty} \overline{\varrho}(N)\, w(N) = \overline{\varrho}(\infty).$$

On the other hand, (EM2) implies

$$M^*_{\theta_0}(\theta_0) = \rightarrow E\, \varrho(Y - \tau(\theta_0)) < \overline{\varrho}(\infty).$$

It is clear from here and from (16) that there exists $N > 0$ such that (15) holds for

$$c(\theta_0) = \gamma_{\theta_0}(2N).$$

□

Corollary 1. *If the regularity conditions* (E1+), (E2), (EM1) *and* (EM2) *hold then the identifiability assumption* (A) *implies the consistency of the parent M-estimator* (1).

The proof of consistency for as wide class of M-estimators of real parameter as considered in Corollary 1 seems to be a new result of the parametric statistics (cf. Lehman [9], Hampel et al [3], Pfanzagl [11], van der Vaart and Wellner [15]). In the next section this result and the related proofs are adapted to the more complicated reality of pseudolinear problems. Now, some examples illustrating the applicability of our results in the context of present as well as of the following section follow.

E x a m p l e 1 (least square error estimator). Let us consider an arbitrary family $\mathscr{F} = (F_\theta\colon\ \theta \in \mathbb{R})$ with mean $\mu(\theta)$ strictly monotone and continuous, and variance

$$\sigma^2(\theta) = \int_{\mathbb{R}} (y - \mu(\theta))^2 \,\mathrm{d}F_\theta(y) < \infty,$$

the L_2-estimator for $\tau(\theta) = \mu(\theta)$. Then

$$M^*_{\theta_0}(\theta) = \sigma^2(\theta_0) + (\mu(\theta) - \mu(\theta_0))^2, \qquad M_n(\theta) = \frac{1}{n}\sum_{i=1}^n Y_i^2 - \frac{2\mu(\theta)}{n}\sum_{i=1}^n Y_i + \mu^2(\theta),$$

and the assumptions of Corollary 1 are satisfied. Hence the estimator $\hat\theta_n$ is in this case consistent. The value $\hat\theta_n$ can be explicitly evaluated, namely

$$\hat\theta_n = \mu^{-1}\left(\int_{\mathbb{R}} y \,\mathrm{d}F_n\right) = \mu^{-1}\left(\frac{1}{n}\sum_{i=1}^n Y_i\right) \quad \text{a.s.},$$

so that the consistency can thus be established also directly, via the law of large numbers. All these facts are well known. If $\mathscr{F}$ is natural exponential then the estimator (10) satisfies the maximum likelihood condition,

$$\hat\theta_n \in \operatorname{argmax}\left(\frac{\theta}{n}\sum_{i=1}^n Y_i - b(\theta)\right) \quad \text{a.s.}$$

To this end it suffices to take into account the formula (4) for $\mu(\theta)$ and the fact that $b(\theta)$ is continuously differentiable and convex. The consistency of maximum likelihood estimators in natural exponential families is also well known (cf. e.g. Brown [1]).

E x a m p l e 2 (α-quantile estimator). Let $0 < \alpha < 1$ and let $\mathscr{F} = \{F_\theta \colon \theta \in \Theta\}$, $\Theta = \mathbb{R}$ be a stochastically increasing family. Then the α-quantile function $\mu_\alpha(\theta)$ given by (5) is nondecreasing on Θ. Assume in addition that $\mu_\alpha(\theta)$ is continuous and increasing. Consider the L_α-estimator for $\tau(\theta) = \mu_\alpha(\theta)$. Then (E1+), (E2), (EM1) and (EM2) hold. If a distribution function $F_{\theta_0}(y)$ is continuous and increasing in the neighborhood of $y = \mu_\alpha(\theta_0)$ then, by Proposition 1 in Appendix, also assumption (A) is satisfied. Therefore, by Corollary 1, the L_α-estimator of $\theta_0 \in \Theta$ is consistent. Some of these facts are well known (cf. e.g. Koenker and Basset [6] in the case of the location model $\mathscr{F} = \{F(x-\theta) \colon \theta \in \mathbb{R}\}$ with absolutely continuous parent distribution function $F(x)$). We have coined the term "α-quantile estimator" for the L_α-estimator $\hat{\theta}_n$ because it employs the α-quantile function $\mu_\alpha(\theta)$ at the place of the general $\tau(\theta)$ figuring in the definition of $M_n(\theta)$. Let us point out that the estimates $\hat{\theta}_n$ identify distributions from $\mathscr{F}$ with the α-quantiles least distant from the sample α-quantiles.

4. Main results

Unless otherwise explicitly stated, throughout this section we assume the validity of regularity conditions (E1), (E2), (EM1) and (M1)–(M4).

In a pseudolinear model the expectation $\rightarrow E\, m_n(\beta)$ of the function $m_n(\beta)$ minimized in (9) depends not only on the true parameter β_0 figuring in the definition (6), but also on the matrix of regressors $\mathbf{z}_n = (z_1, \ldots, z_n) \in \mathscr{Z}^n$. We formulate conditions (a), (b) analogous to the above considered (A), (B). The new conditions are essentially obtained from the previous ones by replacing the functions $M_n(\theta)$ by

$$(18) \qquad m_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \varrho(Y_i - \varphi(z_i^t\,\beta)) \quad \text{where} \quad \varphi = \tau \circ u \quad \text{(cf. (9))}$$

and the expectations (2) by the corresponding alternatives

$$(19) \qquad m^*_{n,\beta_0}(\beta) = \rightarrow E\, m_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \int \varrho(y - \varphi(z_i^t\beta))\,\mathrm{d}F_{u(z_i^t\beta_0)}(y).$$

Note that by (E2) and (M2), $\varphi \colon \mathbb{R} \to \mathbb{R}$ is continuous and strictly monotone. The expectation (19) can be expressed in terms of the function $M^*_{\theta_0}(\theta)$,

$$m^*_{n,\beta_0}(\beta) = \frac{1}{n}\sum_{i=1}^{n} M^*_{u(z_i^t\,\beta_0)}(u(z_i^t\,\beta)).$$

One can argue similarly as in the proof of Lemma 1 that the functions (19) are finitely valued and continuous in β.

It is convenient to introduce on Borel sets $C \subset \mathbb{R}^2$ probability measures Γ_n and Γ by the formulas

$$\Gamma_n(C) = \int_C \mathrm{d}F_{u(z^t\beta_0)}(y)\,\mathrm{d}\mu_n(z)$$

and

$$\Gamma(C) = \int_C \mathrm{d}F_{u(z^t\beta_0)}(y)\,\mathrm{d}\mu(z),$$

where μ_n and μ are the probability measures figuring in (M4). By (M1), for every continuous and bounded $f\colon\ \mathbb{R} \times \mathscr{Z} \to \mathbb{R}$, the function

$$\psi(z) = \int_{\mathbb{R}} f(y,z)\,\mathrm{d}F_{u(z^t\beta_0)}(y)$$

is bounded and continuous on $\mathscr{Z}$. Therefore (M4) implies

$$\int_{\mathbb{R}\times\mathscr{Z}} f(y,z)\,\mathrm{d}\Gamma_n(y,z) = \int_{\mathscr{Z}} \psi(z)\,\mathrm{d}\mu_n(z) \longrightarrow \int_{\mathbb{R}\times\mathscr{Z}} f(y,z)\,\mathrm{d}\Gamma(y,z) \quad \text{as } n \to \infty.$$

Let us define

$$m^*_{\beta_0}(\beta) = \int_{\mathbb{R}\times\mathscr{Z}} \varrho(y - \varphi(z^t\beta))\,\mathrm{d}F_{u(z^t\beta_0)}(y)\,\mathrm{d}\mu(z), \quad \beta,\ \beta_0 \in B. \tag{20}$$

Lemma 4. *For any $\beta_0 \in B$ and $\beta \in B$ the functions* (19) *and* (20) *satisfy the relation*

$$\lim\nolimits_n m^*_{n,\beta_0}(\beta) = m^*_{\beta_0}(\beta),$$

where the convergence is locally uniform in β and the limit function is continuous in β.

P r o o f. Let β_0 and β be fixed. Then

$$\varphi(y,z,\beta) = \varrho(y - \varphi(z^t\beta)) \tag{21}$$

is continuous on $\mathbb{R} \times \mathscr{Z}$ and the functions (19) and (20) satisfy the relations

$$m^*_{n,\beta_0}(\beta) = \int_{\mathbb{R}\times\mathscr{Z}} \varphi(y,z,\beta)\,\mathrm{d}\Gamma_n(y,z) \quad \text{and} \quad m^*_{\beta_0}(\beta) = \int_{\mathbb{R}\times\mathscr{Z}} \varphi(y,z,\beta)\,\mathrm{d}\Gamma(y,z).$$

If ϱ is bounded then $f(y,z)$ is bounded, too and the desired convergence follows directly from the limit relation preceding (20). For an unbounded ϱ this argument still applies since, under (E1), (E2), (EM1) and (M3), the function

$$\psi(z) = \int_{\mathbb{R}} \varphi(y,z,\beta)\,\mathrm{d}F_{u(z^t\beta_0)}(y)$$

remains to be continuous and bounded on $\mathscr{Z}$. As $\varphi(y, z, \beta)$ is continuous on $\mathbb{R} \times \mathscr{Z} \times B$, for each bounded subset $B_0 \subset B$ the class of functions

$$\{f_\beta(y, z) = \varphi(y, z, \beta)\colon\ \beta \in B_0\}$$

is equicontinuous on $\mathbb{R} \times \mathscr{Z}$. Hence the convergence is locally uniform. This convergence thus extends the above mentioned continuity of functions $m^*_{n,\beta_0}(\beta)$ to the limit $m^*_{\beta_0}(\beta)$. $\square$

Next we establish a property of the function φ defined in (18) which is useful in the sequel.

Lemma 5. *If $\varphi(z^t\beta)$ is bounded on $\mathscr{Z} \times B$ then for every $\delta > 0$ there exists a finite decomposition $\{B_1, \dots, B_\ell\}$ of B and points $\beta_k \in B_k$ such that*

$$\max_{1 \leqslant k \leqslant \ell} \sup_{\beta \in B_k, z \in \mathscr{Z}} |\varphi(z^t\beta) - \varphi(z^t\beta_k)| < \delta.$$

Proof. (I) As said above, φ is continuous and strictly monotone on $\mathbb{R}$. We shall prove that for every $\varepsilon > 0$ and $\gamma > 0$ there exists intervals $I_0 = (-\infty, a_0]$, $I_1 = (a_0, a_1], \dots, I_m = (a_m, \infty)$ such that for every $b, b^* \in I_j$

$$\sup_{b, b^* \in I_j} \sup_{a \in \mathbb{R}\,:\,|c| \leqslant \gamma} |\varphi(a + bc) - \varphi(a + b^*c)| < \varepsilon, \quad 0 \leqslant j \leqslant m.$$

It follows from the properties of φ that there exist intervals $J_0 = (-\infty, t_0]$, $J_1 = (t_0, t_1], \dots, J_n = (t_n, \infty)$ such that $|\varphi(x) - \varphi(y)| < \varepsilon/2$ if both x, y belong to the same interval. Let $\Delta > 0$ and a natural r satisfy the conditions

$$\gamma\Delta = \min_{1 \leqslant i \leqslant n} (t_i - t_{i-1}), \quad r\Delta \geqslant \max\{|t_0|, |t_n|\},$$

and let

$$(-\infty, -r\Delta],\ (-r\Delta, -(r-1)\,\Delta], \dots, (r\Delta, \infty)$$

be the intervals I_j. If $b, b^* \in I_j$ then either $|b - b^*| < \Delta$, i.e. $|a + bc - (a + b^*c)| < \gamma\Delta$, or both b and b^* belong to the same boundary interval J_0 or J_n. In either case $x = a + bc$ and $y = a + b^*c$ belong to the same J_k, or to the neighboring intervals J_k, J_{k+1}, so that $|\varphi(x) - \varphi(y)| < \varepsilon$.

(II) Consider now arbitrary $\beta, \beta^* \in B$ and $z \in \mathscr{Z}$ and denote by $(\beta)_s$, $(\beta^*)_s$ and $(z)_s$ the respective s-th coordinates of these vectors for $1 \leqslant s \leqslant p$. Further, define for $1 \leqslant s \leqslant p + 1$ new vectors

$$\beta(s) = \big((\beta^*)_1, \dots, (\beta^*)_{s-1}(\beta)_s, \dots, (\beta)_p\big) \in \mathbb{R}^p.$$

Then

$$\begin{aligned}\varphi(z^t\beta) - \varphi(z^t\beta^*) &= \varphi(z^t\beta(1)) - \varphi(z^t\beta(p+1)) \\ &= \sum_{i=1}^{p}\Big[\varphi(z^t\beta(s)) - \varphi(z^t\beta(s+1))\Big].\end{aligned}$$

Let a_s be defined for $1 \leqslant s \leqslant p$ by the formula

$$a_s = z^t\beta(s) - (z)_s\,(\beta)_s \equiv z^t\beta(s+1) - (z)_s(\beta^*)_s.$$

Then

$$\begin{aligned}|\varphi(z^t\beta) - \varphi(z^t\beta^*)| &= \left|\sum_{s=1}^{p}[\varphi(a_s + (z)_s(\beta)_s) - \varphi(a_s - (z)_s\,(\beta^*)_s)]\right| \\ &\leqslant \sum_{s=1}^{p} |\varphi(a_s + b_s c_s) - \varphi(a_s + b_s^* c_s)|,\end{aligned}$$

where

$$b_s = (\beta)_s, \quad b_s^* = (\beta^*)_s \quad \text{and} \quad c_s = (z)_s.$$

By applying to each summand the result of part (I), one obtains intervals I_{sj} such that

$$\sup_{b_s, b_s^* \in I_{sj}} \; \sup_{a_s \in \mathbb{R}\,:\,|c_s| \leqslant \gamma} |\varphi(a_s + b_s c_s) - \varphi(a_s + b_s^* c_s)| < \varepsilon, \quad 1 \leqslant j \leqslant m.$$

If γ is the same as in the regularity assumption (M3) then this result implies for any integers $1 \leqslant j_1, \ldots, j_p \leqslant m$ the inequality

$$\sup_{\beta, \beta^* \in I_{1j_1} \times \ldots \times I_{pj_p}} \; \sup_{z \in \mathscr{Z}} |\varphi(z^t\beta) - \varphi(z^t\beta^*)| < p\varepsilon.$$

Thus if $\mathscr{A}$ is the algebra of subsets of B induced by all rectangles $I_{1j_1} \times \ldots \times I_{pj_p}$ then the class $\mathscr{B} = \{B_1, \ldots, B_\ell\}$ of atoms of $\mathscr{A}$ is the desired decomposition. The points $\beta_1, \ldots, \beta_\ell$ from the respective atoms may be arbitrary. □

Next we introduce conditions (a), (b) analogous to the consistency conditions (A), (B) for the parent estimator. As proved in Theorem 2 below, if for every $\beta_0 \in B$

(a) the function (20) satisfies the condition

$$\inf_{\|\beta - \beta_0\| > \varepsilon} m^*_{\beta_0}(\beta) > m^*_{\beta_0}(\beta_0) \quad \text{for every } \varepsilon > 0$$

and

(b) for every $\delta > 0$ there exists $c = c(\beta_0, \delta)$ such that

$$\liminf_n \to P\Big(\inf_{\|\beta-\beta_0\|>c} m_n(\beta) > m_n(\beta_0)\Big) > 1-\delta$$

then the estimator (9) is consistent. Theorem 2 is based on the following analogue of Lemma 2.

Lemma 6. *For all* $c > 0$,

$$\sup_{\|\beta-\beta_0\|\leqslant c} |m_n(\beta) - m^*_{\beta_0}(\beta)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

If $\varphi(z^t\beta)$ *is bounded on* $\mathscr{Z} \times B$ *then*

$$\sup_{\beta\in B} |m_n(\beta) - m^*_{\beta_0}(\beta)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Proof. (I) Let $\varphi(z^t\beta)$ be bounded. Similarly as in the proof of Lemma 2, one can use

$$s = \sup_{z\in\mathscr{Z},\beta\in B} |\varphi(z^t\beta)| < \infty$$

to prove that the random variables

$$W_n(N) = \sup_{\beta\in B} \frac{1}{n}\sum_{i=1}^n \varrho(Y_i - \varphi(z^t\beta))\,[1-\chi_N(Y_i)], \quad N > 0,$$

satisfy the relation

$$\lim_{N\to\infty}\Big(\sup_n \to E\,W_n(N)\Big) = 0.$$

Since

$$\varphi_N(y,z,\beta) \stackrel{\triangle}{=} \varphi(y,z,\beta)\,\chi_N(y) \quad \text{(cf. (11) and (21))}$$

is continuous in $(y,z,\beta) \in [-N,N] \times \mathscr{Z} \times B$, Lemma 5 implies that for every $\varepsilon > 0$ one can find points $\beta_1,\ldots,\beta_\ell$ in the corresponding sets of a decomposition $B_1,\ldots,B_\ell$ of B such that

$$\max_{1\leqslant k\leqslant \ell} \sup_{[-N,N]\times\mathscr{Z}\times B_k} |\varphi_N(y,z,\beta)\,\chi_N(y) - \varphi_N(y,z,\beta_k)| < \varepsilon.$$

This together with the fact that for any fixed $N > 0$ and $1 \leqslant k \leqslant \ell$ the random variables

$$X_i = \varphi_N(Y_i, z_i, \beta_k), \quad 1 \leqslant i \leqslant n,$$

satisfy the law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \longrightarrow E\, X_i) \xrightarrow{P} 0 \quad \text{as } n \to \infty,$$

implies the first stochastic convergence of Lemma 6 with $m^*_{\beta_0}(\beta)$ replaced by $m^*_{n,\beta_0}(\beta)$. However, by Lemma 4 this convergence holds without this replacement as well.

(II) Let $\varphi(z^t\beta)$ be unbounded on $\mathscr{X} \times B$. Due to the continuity, it is bounded on the bounded set

$$\mathscr{X} \times B_0 \quad \text{for} \quad B_0 = \{\beta \in B\colon\ \|\beta - \beta_0\| \leqslant c\}$$

for any $c > 0$. Thus all steps of part (I) are applicable with B replaced by B_0. □

In the proof of Theorem 2 below the following sharpening of Lemma 4 is needed.

Lemma 7. *If $\varphi(z^t\beta)$ is bounded on $\mathscr{X} \times B$ then the convergence in Lemma 4 is uniform in $\beta \in B$.*

P r o o f. It follows from Lemma 5 that in this case the whole class

$$\{f_\beta(y,z) = \varphi(y,z,\beta)\colon\ \beta \in B\}$$

considered in the proof of Lemma 4 is equicontinuous. The rest is the same as in the proof of Lemma 4. □

Theorem 2. *Let the regularity conditions* (E1), (E2), (EM1) *and* (M1)–(M4) *hold. Then assumptions* (a) *and* (b) *imply the consistency of the estimator* (9). *If $\varphi(z^t\beta)$ is bounded on $\mathscr{X} \times B$ then assumption* (a) *alone is sufficient for this consistency.*

P r o o f. (I) Let $\varphi(z^t\beta)$ be bounded. Then similarly as in the proof of Theorem 1

$$\{\|\hat\beta_n - \beta_0\| > \varepsilon\} \subset A_n,$$

where

$$A_n = \Big\{\inf_{\|\beta-\beta_0\|>\varepsilon} m^*_{n,\beta_0}(\beta) < m^*_{n,\beta_0}(\beta_0) + 2\sup_{\beta\in B}|m_n(\beta) - m^*_{n,\beta_0}(\beta)|\Big\}.$$

Obviously, A_n is dominated by

$$\tilde{A}_n = \Big\{ \inf_{\|\beta-\beta_0\|>\varepsilon} m^*_{\beta_0}(\beta) < m^*_{\beta_0}(\beta_0) + 2 \sup_{\beta\in B} |m^*_{n,\beta_0}(\beta) - m^*_{\beta_0}(\beta)| + 2 \sup_{\beta\in B} |m_n(\beta) - m^*_{n,\beta_0}(\beta)| \Big\}.$$

Assumption (a) implies

$$\inf_{\|\beta-\beta_0\|>\varepsilon} m^*_{\beta_0}(\beta) - m^*_{\beta_0}(\beta_0) > 0.$$

By Lemmas 6 and 7, both suprema in $\tilde{A}_n$ tend in probability to zero. Therefore $\rightarrow P(\tilde{A}_n) \rightarrow 0$ as $n \rightarrow \infty$.

(II) Let now $\varphi(z^t\beta)$ be unbounded and choose ε between 0 and c from the assumption (b). Since (b) implies

$$\liminf{}_n \rightarrow P(\|\hat{\beta}_n - \beta_0\| > c) > 1 - \delta,$$

it suffices to prove analogously as in part (II) of the proof of Theorem 1 that the probability of

$$\tilde{A}_n = \Big\{ \inf_{\varepsilon<\|\beta-\beta_0\|\leqslant c} m^*_{\beta_0}(\beta) < m^*_{\beta_0}(\beta_0) + 2 \sup_{\|\beta-\beta_0\|\leqslant c} \|m^*_{n,\theta_0}(\theta) - m^*_{\theta_0}(\theta)\| + 2 \sup_{\|\beta-\beta_0\|\leqslant c} |m_n(\beta) - m^*_{n,\theta_0}(\beta)| \Big\}$$

tends to zero. Similarly as at the end of part (I), by using (a) this can be reduced to the proof that both the suprema tend to zero. But this was already proved in Lemmas 4 and 6. □

Assumption (b) is needed only when $\varphi(z^t\beta)$ is unbounded on $\mathscr{Z} \times B$. We shall show in Lemma 9 that then (b) is almost always fulfilled, similarly as it has been observed in the parent model in Lemma 3. In the present case we need the additional assumptions

$$\lim_{s\rightarrow\infty} \lim_{c\rightarrow\infty} \operatorname{liminf}_n \mathscr{M}_n(c,s) > 0 \tag{22}$$

or

$$\lim_{s\rightarrow\infty} \lim_{c\rightarrow\infty} \operatorname{liminf}_n \mathscr{M}_n(c,s) = 1, \tag{23}$$

where

$$\mathscr{M}_n(c,s) = \inf_{\|\beta\|>c} \mu_n(z \in \mathscr{Z}\colon\ |z^t\beta| > s) \tag{24}$$

and the limits exist due to the monotonicity of the corresponding functions.

The next result provides simple sufficient conditions for (22), (23) in terms of the measure μ figuring in the regularity assumption (M4).

Lemma 8. *Relation (22) follows from the condition*

$$\mu(z \in \mathscr{Z}\colon\ z^t\beta = 0) < 1 \quad \text{for every nonzero } \beta \in B, \tag{25}$$

and relation (23) *from the condition*

$$\mu(z \in \mathscr{Z}\colon\ z^t\beta = 0) = 0 \quad \text{for every nonzero } \beta \in B. \tag{26}$$

Proof. Denote

$$\alpha = \lim_{\varepsilon\downarrow 0} \operatorname{liminf}_n \inf_{\|\beta\|=1} \mu_n(z \in \mathscr{Z}\colon\ |z^t\beta| > \varepsilon).$$

By Lemma 5 in Liese and Vajda [7], (25) is equivalent to $\alpha > 0$ and (26) is equivalent to $\alpha = 1$. Taking into account that for $\beta \in \mathbb{R}^p$ with $\|\beta\| > c > 0$

$$\mu_n(z \in \mathscr{Z}\colon\ |z^t\beta| > s) \geqslant \mu_n\Big(z \in \mathscr{Z}\colon\ \frac{|z^t\beta|}{\|\beta\|} \geqslant \frac{s}{c}\Big) \quad \text{for all } s > 0,$$

one obtains from (24) for all positive s and c

$$\mathscr{M}_n(s,c) \geqslant \inf_{\|\beta\|=1} \mu_n(z \in \mathscr{Z}\colon\ |z^t\beta| > s/c).$$

Consequently,

$$\lim_{c\to\infty} \operatorname{liminf}_n \mathscr{M}_n(s,c) \geqslant \alpha \quad \text{for all } s > 0.$$

It is clear from here that (25) implies (22) and (26) implies (23). □

In the next lemma we consider the function $\overline{\varrho}$ defined by (13) and (14), and also $\overline{\varphi}$ defined by analogous formulas

$$\overline{\varphi}(y) = \min\{|\varphi(y)|,\, |\varphi(-y)|\}, \quad y > 0,$$
$$\overline{\varphi}(\infty) = \lim_{y\to\infty} \overline{\varphi}(y),$$

for φ introduced in (18). As said at the beginning of this section, φ is continuous and strictly monotone. Therefore $\overline{\varphi}$ is continuous and increasing in the domain $(0,\infty)$. If $\overline{\varphi}(\infty)=\infty$, then the inverse $\overline{\varphi}^{-1}$ is continuous and increasing in the same domain and

$$\psi(s) \stackrel{\triangle}{=} \overline{\varphi}^{-1}(s) \to \infty \quad \text{as } s \to \infty. \tag{27}$$

Lemma 9. *Let* $\overline{\varphi}(\infty)=\infty$ *and let also the regularity assumptions* (E1+) *and (EM2) hold. Then* (23) *implies* (b). *The weaker assumption* (22) *implies* (b) *if* $\overline{\varrho}(\infty)=\infty$.

Proof. Let $\beta\in B$ be arbitrary and define for $N>0$

$$\overline{m}_n^N(\beta)=\frac{1}{n}\sum_{i=1}^n \overline{\varrho}(|y_i-\varphi(z_i^t\beta)|)\,\chi_N(Y_i) \quad \text{(cf. (11))}.$$

Then $m_n(\beta)\geqslant \overline{m}_n^N(\beta)$. If $M>N>0$ then similar arguments as in the proof of Lemma 3 lead to the inequalities

$$\begin{aligned} m_n(\beta) &\geqslant \overline{\varrho}(M-N)\,\frac{1}{n}\sum_{i:\ |\varphi(z_i^t\beta)|>M}\chi_N(Y_i)\\ &\geqslant \overline{\varrho}(M-N)\,\frac{1}{n}\sum_{i:\ \overline{\varphi}(z_i^t\beta)>M}\chi_N(Y_i)\\ &\geqslant \overline{\varrho}(M-N)\Big[-1+\frac{1}{n}\sum_{i=1}^n\chi_N(Y_i)+\mu_n(z\in\mathscr{X}\colon\ \overline{\varphi}(z^t\beta)>M)\Big]. \end{aligned}$$

Hence for any $c>0$ and $\psi(M)=\overline{\varphi}^{-1}(M)$ we have

$$\inf_{\|\beta\|>c} m_n(\beta)\geqslant \overline{\varrho}(M-N)\Big[\mathscr{M}_n(c,\psi(M))-\frac{1}{n}\sum_{i=1}^n(1-\chi_N(Y_i))\Big] \quad \text{(cf. (24))}. \tag{28}$$

For $\overline{\varrho}(\infty)<\infty$ it follows from (EM2) and (20) that one can find $\alpha>0$ with the property

$$\overline{\varrho}(\infty)\,(\mathscr{M}-\alpha)>m^*_{\beta_0}(\beta_0)+\alpha, \tag{29}$$

where $\mathscr{M}=1$ is the limit figuring in (23). Similarly for $\overline{\varrho}=\infty$ it follows from the finiteness of $m^*_{\beta_0}(\beta_0)$ (see Lemma 4) that there exists $\alpha>0$ such that (29) remains satisfied for the limit $0<\mathscr{M}\leqslant 1$ figuring in (22). If we define for α figuring in (29)

$$A_n^N=\Big\{\frac{1}{n}\sum_{i=1}^n(1-\chi_N(Y_i))\leqslant\alpha\Big\}$$

then for every $\delta > 0$ there exist n_0 and N_0 such that $N > N_0$, $n > n_0$ implies $\to P(A_n^N) \geqslant 1 - \delta/2$. Similarly if

$$A_n^* = \{m_n(\beta_0) \leqslant m_{\beta_0}^*(\beta_0) + \alpha\}$$

then Lemma 4 implies the existence of n_1 such that $\to P(A_n^*) > 1 - \delta/2$ for $n > n_1$. Further, by (28) and the definition of A_n^N and A_n^*, for all c, n, M and N under consideration we have

$$\begin{aligned}
&\Big\{\inf_{\|\beta\|>c} m_n(\beta) > m_n(\beta_0)\Big\} \\
&\supset \Big\{\overline{\varrho}(M-N)\Big[\mathscr{M}_n(c,\psi(M)) - \frac{1}{n}\sum_{i=1}^{N}(1-\chi_N(Y_i))\Big] > m_n(\beta_0)\Big\} \\
&\supset \Big\{\overline{\varrho}(M-N)\left[\mathscr{M}_n(c,\psi(M)) - \alpha\right] > m_{\beta_0}^*(\beta_0) + \alpha\Big\} \cap A_n^N \cap A_n^*.
\end{aligned}$$

Suppose now that (23) holds, or that (22) holds and $\overline{\varrho}(\infty) = \infty$. Then (27) and (29) imply

$$\lim_{N\to\infty}\lim_{c\to\infty}\liminf{}_n\, \overline{\varrho}(N^2 - N)\left[\mathscr{M}_n(c, \psi(N^2) - \alpha)\right] > m_{\beta_0}^*(\beta_0) + \alpha.$$

Hence there exist $N_* > N_0$ and $c_* > 0$ such that for all sufficiently large n

$$\overline{\varrho}(N_*^2 - N_*)\left[\mathscr{M}_n(c_*, \psi(N_*^2) - \alpha)\right] > m_{\beta_0}^*(\beta_0) + \alpha.$$

Thus it follows from the above inclusions that

$$\liminf{}_n \to P\Big(\inf_{\|\beta\|>c} m_n(\beta) > m_n(\beta_0)\Big) \geqslant \liminf{}_n\, P(A_n^{N_*} \cap A_n^*) \geqslant 1 - \delta,$$

i.e. that (b) holds. □

Next we introduce two corollaries of the main result of this paper.

Corollary 2. *Let* $\overline{\varphi}(\infty) = \infty$ *and let the regularity conditions* (E1+), (E2), (EM1), (EM2) *and* (M1)–(M4) *hold. Then the identifiability assumption* (a) *together with* (23) *implies the consistency of the* M*-estimator* (9). *If* $\overline{\varrho}(\infty) = \infty$ *then the last statement remains true with* (23) *replaced by the weaker condition* (22).

P r o o f. Clear from Theorem 2 and Lemma 9. □

Corollary 3. *Corollary* 2 *holds with* (23) *replaced by* (26) *and* (22) *replaced by* (25).

P r o o f. Clear from Lemma 8. □

Corollaries 2 and 3 present verifiable regularity conditions which are sufficient for consistency, and which are applicable to a great variety of M-estimators and pseudolinear models. For the pseudolinear and generalized linear models these conditions are new. Their nontriviality can be checked when they are applied to a special class of linear models where they can be compared with similar conditions derived in literature directly for this special case. One can easily verify that they either directly coincide with, or are very close to the conditions which can be found for the linear case in Hampel et al [3], Pollard [13, 14], Jurečková and Procházka [4], Jurečková and Sen [5], Liese and Vajda [7, 8] and other references. This is illustrated in Example 5 of the next section.

5. Applications

Corollaries 2 and 3 are of considerable practical importance. As we have mentioned (and practically illustrated by several examples) in Section 2, regularity conditions (E1+), (E2), (EM1) and (EM2) hold for practically all M-estimators and families $\mathscr{F}$ considered in statistical applications. The families $\mathscr{F}$ considered in applications usually fulfil also conditions (M1) and (M2). The other conditions (M3), (M4) and (22), (23) (or their simpler alternatives (25), (26)) characterize designs of experiments reducing consistency to the identifiability condition (a). This condition is then in fact necessary and sufficient for consistency (cf. the linear regression example in Liese and Vajda [9]). The design of an identifiable experiment and the choice of a class of estimation procedures able to identify the true parameter in the sense of (a), are the primary tasks of the statistician. To this end he has at his disposal two universal tools: the functions ϱ and τ. The only restrictions on these tools are the quite tolerant regularity conditions (E1+), (E2), (EM1) and (EM2). As soon as this task is fulfilled, further criteria can be involved in the process of final specification of the estimation procedure, e.g. the efficiency or robustness.

Now we illustrate the applicability of Corollaries 2 and 3, and of the results summarized in Examples 1–4, to pseudolinear statistical models.

Example 5 (α-quantile estimator). Let $0 < \alpha < 1$ and let $\mathscr{F} = \{F(y - \theta), \theta \in \mathbb{R}\}$ be a location family with $F(y)$ strictly increasing in the neighborhood of $y = F^{-1}(\alpha)$. By the result of Example 2, in this case the conditions (E1+), (E2), (EM1) and (EM2) hold. The generalized linear model (6) with $\beta_0 \in B = \mathbb{R}^p$ reduces to the nonlinear regression model (7) where the independent errors ε_i are distributed by $F(y)$. Suppose that the function $u\colon \mathbb{R} \to \mathbb{R}$ figuring in our model (7) is strictly monotone and continuous, so that (M2) holds. If u is the identity mapping then we are within the framework of the classical linear regression with i.i.d. errors.

Obviously, also (M1) holds in our case. If (M3), (M4) and (25) hold then for all $\beta_0 \in \mathbb{R}^p$

$$\mu\Big(z \in \mathscr{X} \colon \inf_{\|\beta-\beta_0\|>\varepsilon} |z^t(\beta-\beta_0)| > 0\Big) > 0, \quad \varepsilon > 0.$$

Since u is strinctly monotone on $\mathbb{R}$, this implies

$$\mu\Big(z \in \mathscr{X} \colon \inf_{\|\beta-\beta_0\|>\varepsilon} |u(z^t\,\beta) - u(z^t\,\beta_0)| > 0\Big) > 0, \quad \varepsilon > 0.$$

One can deduce from here that also

$$\inf_{\|\beta-\beta_0\|>\varepsilon} m^*_{\beta_0}(\beta) - m^*_{\beta_0}(\beta_0) > 0,$$

i.e. that the assumption (a) is satisfied. Therefore, by Corollary 3, the consistency of the α-quantile estimator of the regression parameter β_0 takes place if the conditions (M3), (M4) and (25) hold. These conditions are simpler and in most cases weaker than the consistency conditions which can be obtained from Theorem 1 in Jurečková and Procházka [4] establishing, however, not only the consistency but also the asymptotic normality of the estimator under consideration.

Appendix: Quantiles

We say that a real valued function $M^*(\theta)$ defined on an interval $\Theta \subset \mathbb{R}$ is *unimodal* with the *mode* at θ_0 in $M^*(\theta)$ if it is nondecreasing in the subdomain $\theta \geqslant \theta_0$ and nonincreasing in the subdomain $M^*(\theta)$.

Consider an arbitrary $0 < \alpha < 1$ and a probability distribution $F(x)$ continuous from the left on $\mathbb{R}$. The α-quantile

$$x_\alpha = x_\alpha(F)$$

is the real number satisfying the condition

$$F(x_\alpha) \leqslant \alpha \leqslant F(x_\alpha + 0). \tag{30}$$

The α-quantiles are not unique in general. For every F and α under consideration, the quantile function (generalized inverse)

$$F^{-1}(\alpha) = \inf\{x \in \mathbb{R} \colon\; F(x) \geqslant \alpha\}$$

represents the smallest α-quantile.

Each α-quantile is a solution of a minimization problem. To this end consider

$$\varrho_\alpha(x) = \begin{cases} \alpha x & \text{if } x > 0 \\ -(1-\alpha)\,x & \text{if } x \leqslant 0 \end{cases} \qquad \text{(cf. (3))}$$

and the function

$$\Phi_\alpha(x) = \int \varrho_\alpha(y-x)\,\mathrm{d}F(y),$$

and suppose that $\int |x|\,\mathrm{d}F(x) < \infty$. Since $\varrho_\alpha(x)$ is piecewise linear, there exists $c > 0$ such that

$$\Big|\frac{1}{h}(\varrho_\alpha(x+h) - \varrho_\alpha(x))\Big| \leqslant c|x| \quad \text{for all } h \neq 0.$$

If D_+ and D_- denote the right and left hand derivatives then, by the Lebesgue theorem,

$$\begin{aligned} D_+\,\Phi_\alpha(x) &= \int_{(x,\infty)} (-\alpha)\,\mathrm{d}F(y) + \int_{(-\infty,x]} (1-\alpha)\,\mathrm{d}F(y) \\ &= (1-\alpha)\,F(x+0) - \alpha(1 - F(x+0)) \\ &= F(x+0) - \alpha \end{aligned}$$

and, similarly,

$$D_-\,\Phi_\alpha(x) = F(x) - \alpha.$$

It follows from here that x_α minimizes $\Phi_\alpha(x)$ on $\mathbb{R}$. This result has been established previously by Koenker and Basset [6] for absolutely continuous distribution functions $F(x)$.

Proposition 1. *Let $F(x)$ be an arbitrary distribution function on $\mathbb{R}$. If $\int |x|\,\mathrm{d}F(x) < \infty$ then every α-quantile x_α minimizes $\Phi_\alpha(x)$ on $\mathbb{R}$.*

Now we can formulate the main result of Appendix which guarantees the consistency condition (A) of Section 1 for functions (2) with $\varrho = \varrho_\alpha$ and $0 < \alpha < 1$. Let $\mathscr{F} = (F_\theta\colon\, \theta \in \Theta)$, $\Theta \subset \mathbb{R}$, be an arbitrary family of probability distribution functions on $\mathbb{R}$ satisfying for some $\theta_0 \in \Theta$ the condition $\int |y|\,\mathrm{d}F_{\theta_0}(y) < \infty$. Define for fixed $0 < \alpha < 1$ functions

$$\mu_\alpha(\theta) = F_\theta^{-1}(\alpha)$$

and

$$M^*_{\theta_0}(\theta) = \int \varrho_\alpha(y - \mu_\alpha(\theta))\,\mathrm{d}F_{\theta_0}(y)$$

of the variable $\theta \in \Theta$. The next proposition follows from Proposition 1.

Proposition 2. *Let $0 < \alpha < 1$ be arbitrary. If $\mu_\alpha(\theta)$ is monotone on Θ then the function $M^*_{\theta_0}(\theta)$ is unimodal with the mode at $\theta = \theta_0$. If $\mu_\alpha(\theta)$ is strictly monotone on Θ and $F^{-1}_{\theta_0}(\alpha)$ or $F^{-1}_{\theta_0}(\alpha^*)$ belongs to the support of the distribution F_{θ_0} then the respective mode is unique.*

References

[1] *L. D. Brown*: Fundamentals of Statistical Exponential Families. Lecture Notes No. 9. Institute of Mathematical Statistics, Hayward, California, 1986.

[2] *L. Fahrmeir, H. Kaufman*: Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. Annals of Statistics *13* (1985), 342–368.

[3] *F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti, W. A. Stahel*: Robust Statistics: The Approach Based on Influence Functions. Wiley, New York, 1986.

[4] *J. Jurečková, B. Procházka*: Regression quantiles and trimmed least squares estimator in nonlinear regression model. Nonparametric Statistics *3* (1994), 201–222.

[5] *J. Jurečková, P. K. Sen*: Robust Statistical Procedures. Wiley, New York, 1996.

[6] *R. Koenker, G. Basset*: Regression quantiles. Econometrica *46* (1978), 33–50.

[7] *F. Liese, I. Vajda*: Consistency of M-estimates in general regression models. J. Multivar. Analysis *50* (1994), 93–114.

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

[9] *E. L. Lehman*: Theory of Point Estimation. Wiley, New York, 1983.

[10] *S. Morgenthaler*: Least-absolute-deviations fits for generalized linear models. Biometrika *79* (1992), 747–754.

[11] *J. Pfanzagl*: Parametric Statistical Theory. De Gruyter, Berlin, 1994.

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

[13] *D. Pollard*: Empirical Processes: Theory and Applications. IMS, Hayward, 1990.

[14] *D. Pollard*: Asymptotics for least absolute deviation regression estimators. Econometric Theory *7* (1991), 186–199.

[15] *A. van der Vaart, J. A. Wellner*: Weak Convergence and Empirical Processes. Springer, New York, 1996.

[16] *S. Zwanzig*: On L_1-norm estimators in nonlinear regression and in nonlinear error-in-variables models. IMS Lecture Notes 31, 101–118, Hayward, 1997.

Authors' addresses: *Friedrich Liese*, Department of Mathematics, University of Rostock, Universitätsplatz 1, D-180 55 Rostock, Germany; *Igor Vajda*, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, CZ-182 08 Praha, Czech Republic.