IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 194–203

DOI: 10.1214/10-IMSCOLL719

Second-order asymptotic representation of M-estimators in a linear model

Marek Omelka[1,*]

Charles University in Prague

Abstract: The asymptotic properties of fixed-scale as well as studentized M-estimators in linear models with fixed carriers are studied. A two term von Mises expansion (second order asymptotic representation) is derived and verified. Possible applications of this result are shortly discussed.

1. Introduction

Suppose that observations $\mathbf{Y} = (Y_1, \ldots, Y_n)^\mathsf{T}$ follow a linear model

$$Y_i = \beta_1\, x_{i1} + \ldots + \beta_p\, x_{ip} + e_i = \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_i + e_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\mathsf{T}$ is a vector of unknown parameters, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\mathsf{T}$ $(i = 1, \ldots, n)$ are rows of a known matrix $\mathbf{X}_n$, and $e_1, \ldots, e_n$ are independent, identically distributed random variables with an unknown cumulative distribution function (cdf) F.

Given an absolutely continuous loss function ρ, a fixed scale (studentized) M-estimator $\hat{\boldsymbol{\beta}}_n$ of the parameter $\boldsymbol{\beta}$ is defined as a solution of the minimisation

$$\sum_{i=1}^n \rho\left(Y_i - \mathbf{t}^\mathsf{T}\mathbf{x}_i\right) := \min, \qquad \left(\text{or } \sum_{i=1}^n \rho\left(\tfrac{Y_i - \mathbf{t}^\mathsf{T}\mathbf{x}_i}{S_n}\right) := \min\right),$$

where S_n is an estimator of scale.

If the function ρ is differentiable with $\psi = \rho'$ being continuous, then the estimator $\hat{\boldsymbol{\beta}}_n$ may be found as a solution of the system of equations

$$\sum_{i=1}^n \mathbf{x}_i\, \psi(Y_i - b^\mathsf{T}\mathbf{x}_i) = \mathbf{0} \qquad \left(\text{or } \sum_{i=1}^n \mathbf{x}_i\, \psi(\tfrac{Y_i - b^\mathsf{T}\mathbf{x}_i}{S_n}) = \mathbf{0}\right). \tag{1.2}$$

As the defining equation (1.2) gives more flexibility to tune properties of M-estimators by a choice of a function ψ, $\hat{\boldsymbol{\beta}}_n$ is usually defined as a carefully chosen root of (1.2).

*The work was supported by the grant MSM 0021620839.
[1]Marek Omelka, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Prague, Czech Republic. e-mail: omelka@karlin.mff.cuni.cz
AMS 2000 subject classifications: Primary 62G05; secondary 62F05.
Keywords and phrases: M-estimator, empirical processes.

It is well known (see e. g. Jurečková and Sen [14]) that provided some standard regularity assumptions are met, then the M-estimator $\hat{\boldsymbol{\beta}}_n$ admits the following representation

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \frac{\mathbf{V}_n^{-1}}{\gamma_1 \sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \, \psi(e_i) + \mathbf{R}_n, \qquad (\text{or } (3.4)), \tag{1.3}$$

with $\gamma_1 = \mathsf{E}\,\psi'(e_1)$ and $\mathbf{V}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}}$, where the remainder term $\mathbf{R}_n$ is of order $o_p(1)$. The equation (1.3) is sometimes called the first order asymptotic representation of the estimator $\hat{\boldsymbol{\beta}}_n$ or asymptotic linearity of $\hat{\boldsymbol{\beta}}_n$ or a Bahadur-Kiefer representation. Let us recall that the interest in the behaviour of the remainder term $\mathbf{R}_n$ goes back to the work of Bahadur [6] and Kiefer [15], where a similar expansion for a sample quantile was considered. Provided that the function ψ and the distribution of the errors F are sufficiently smooth, in Jurečková and Sen [12] it was proved that $\mathbf{R}_n = O_p(\frac{1}{\sqrt{n}})$. The asymptotic distribution of the random variable $\sqrt{n}\,\mathbf{R}_n$ was studied by Boos [7] for the special case of a location model and by Jurečková and Sen [13] for an M-estimator of a general scalar parameter. The case of a discontinuous (score) function $\psi = \rho'$ was treated in Jurečková and Sen [11].

Many interesting results about the distributional as well as almost sure behavior of the reminder term $\mathbf{R}_n$ can be found in the work of Arcones. Among others let us mention results for U-quantiles in Arcones [1], multivariate location M-estimators in Arcones and Mason [5], and the two dimensional spatial medians in Arcones [3].

Important contributions to the study of the behavior of the reminder term $\mathbf{R}_n$ in the context of a linear model (1.1) are the results of Jurečková and Sen [12] from which the O_P-rate for a general M-estimator of $\boldsymbol{\beta}$ can be deduced. Arcones [2] considered L_p-regression estimators (i. e. $\rho(x) = |x|^p$, $p \geq 1$) and found the almost sure behavior of $\mathbf{R}_n$. Further, Arcones [4] and Knight [16] focused on the least absolute deviation regression estimator (i. e. $\rho(x) = |x|$) and derived the limiting distribution of $n^{1/4}\,\mathbf{R}_n$.

Our paper extends the results of Boos [7], and Jurečková and Sen [13] in the following way. We derive a two term von Mises expansion (a second order asymptotic representation) of the M-estimator in the linear model (1.1) and we rigorously verify that the second term of the von Mises expansion $\mathbf{T}_n^{(2)}$ satisfies

$$|\mathbf{T}_n^{(2)} - \mathbf{R}_n|_2 = o_p(\tfrac{1}{\sqrt{n}}),$$

where $|\cdot|_2$ stands for the Euclidean norm. That yields not only the asymptotic distribution of $\sqrt{n}\,\mathbf{R}_n$, but it also enables a finer comparison of an M-estimator with another estimator (e. g. an R-estimator) that is asymptotically equivalent.

Moreover, our approach can be easily modified to verify higher order von Mises expansions of one-step M-estimators that were derived in Welsh and Ronchetti [19] in a heuristic way.

In Section 2, we state some auxiliary results on asymptotic behaviour of M-processes, which may be of independent interest. In Section 3, we derive a two term von Mises expansions of an M-estimator. We finish with a short discussion of possible applications of our results. The proofs are to be found in Omelka [18].

2. Auxiliary results

In this section some auxiliary results concerning the asymptotic behaviour of certain processes associated with M-estimation in the model (1.1) are stated. It is useful to distinguish whether an M-estimator is studentized or not.

2.1. Fixed scale

Let $\{c_{in}, i = 1,\dots,n\}$ and $\{\mathbf{x}_{in}, i = 1,\dots,n\}$ be triangular arrays of scalars and vectors in $\mathbb{R}^p$ respectively, and $\mathbf{t} = (t_1,\dots,t_p)^\mathsf{T}$. Our interest is in the (fixed scale) M-process

$$M_n(\mathbf{t}) = \sum_{i=1}^{n} c_{in}\left[\psi(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_{in}}{\sqrt{n}}) - \psi(e_i) + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_{in}}{\sqrt{n}}\,\psi'(e_i)\right], \tag{2.1}$$

where $\mathbf{t} \in T = \{\mathbf{s} \in \mathbb{R}^p : |\mathbf{s}|_2 \le M\}$ and M is an arbitrarily large but fixed constant.

We will make the following assumptions:

X.1

$$\frac{1}{n}\sum_{i=1}^{n} c_{in}^2 = O(1), \qquad \lim_{n\to\infty}\frac{\max_{1\le i\le n}|c_{in}|}{\sqrt{n}} = 0,$$

X.2

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_{in}|_2^2 = O(1), \qquad \lim_{n\to\infty}\frac{\max_{1\le i\le n}|\mathbf{x}_{in}|_2}{\sqrt{n}} = 0,$$

X.3

$$\lim_{n\to\infty}\max_{1\le i\le n}\frac{|c_{in}|\,|\mathbf{x}_{in}|_2}{\sqrt{n}} = 0,$$

X.4

$$B_n^2 = \frac{1}{n}\sum_{i=1}^{n} c_{in}^2\,|\mathbf{x}_{in}|_2^2 = O(1), \quad \text{as } n\to\infty.$$

While assumptions **X.1** $-$ **3** are analogous to the assumptions used in Jurečková [9] to deal with Wilcoxon rank process, the last assumption **X.4** is purely for convenience. If $B_n^2 = O(1)$ were not satisfied, we would work with the process $M_n'(\mathbf{t}) = \frac{M_n(\mathbf{t})}{B_n}$ and derive analogous results.

In Section 3 we will substitute x_{ij} $(j = 1,\dots,p)$ for c_{in} to find the second order asymptotic distributions of the regression M-estimator $\hat{\boldsymbol{\beta}}_n$. For $c_{in} = |\mathbf{x}_{in}|_2$, assumptions **X.1** $-$ **4** may be summarised as

XX.1

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_{in}|_2^4 = O(1), \qquad \lim_{n\to\infty}\frac{\max_{1\le i\le n}|\mathbf{x}_{in}|_2^2}{\sqrt{n}} = 0. \tag{2.2}$$

For notational simplicity, in the following we will write simply c_i and $\mathbf{x}_i$ instead of c_{in} and $\mathbf{x}_{in}$.

The distribution function F of the errors in the model (1.1) and the function ψ used to construct an M-estimator through (1.2) are assumed to satisfy the following regularity conditions.

Fix. 1 ψ is absolutely continuous with a derivative ψ' such that

$$\mathsf{E}[\psi'(e_1)]^2 < \infty.$$

Fix. 2 The (random) function $p(t) = \psi'(e_1 + t)$ is continuous in the quadratic mean at the point zero, that is

$$\lim_{t\to 0}\mathsf{E}\,[p(t) - p(0)]^2 = \lim_{t\to 0}\mathsf{E}\,[\psi'(e_1 + t) - \psi'(e_1)]^2 = 0.$$

Fix. 3 The second derivative of the function $\lambda(t) = \mathsf{E}\,\psi(e_1 + t)$ is finite and continuous at the point 0.

Inspecting **Fix. 1 − 3** one sees that the more is assumed about the function ψ, the less is needed to be assumed about F and the other way around. In robust statistics it is quite common to put restrictive conditions on the function ψ, as the distribution F of the errors is generally unknown. For instance if the function ψ is twice differentiable, then it is not difficult to verify that assumptions **Fix. 1 − 3** are met if both ψ' and ψ'' are bounded and ψ'' is continuous F-almost everywhere. This includes e. g. Tukey's biweight function

$$\psi(x) = x(1 - \tfrac{x^2}{k^2})^2\, \mathbb{I}\{|x| \le k\}.$$

An important class of ψ functions which do not posses a second derivative everywhere are piecewise linear functions. This class includes e. g. Huber's function

$$\psi(x) = \max\{\min\{x, k\}, -k\}.$$

Assumptions **Fix. 1 − 3** are satisfied provided that:

A.1 ψ is a continuous piecewise linear function with the derivative

$$\psi'(x) = \alpha_j, \quad \text{for } r_j < x \le r_{j+1},\ j = 0, \ldots, k,$$

where $\alpha_0, \alpha_1, \ldots, \alpha_k$ are real numbers, $\alpha_0 = \alpha_k = 0$ and $-\infty = r_0 < r_1 < \ldots < r_k < r_{k+1} = \infty$.

A.2 The cdf F is absolutely continuous with a derivative which is continuous at the points $r_1, \ldots, r_k$.

Note that assumption **A.1** trivially implies **Fix. 1** and **A.2** ensures both **Fix. 2** and **Fix.3**.

Many of the following results (in particular for studentized M-estimators) simplify significantly if the distribution of the errors is symmetric. For the sake of later reference let us state this assumption explicitly.

Sym The distribution of the errors is symmetric and the ψ-function is antisymmetric, that is $F(x) = 1 - F(-x)$ and $\psi(x) = -\psi(-x)$ for all $x \in \mathbb{R}$.

Put γ_2 for the second derivative of the function $\lambda(t) = \mathsf{E}\,\psi(e_1 + t)$ at the point 0. That is $\gamma_2 = \sum_{j=1}^{k} \alpha_j\, [f(r_{j+1}) - f(r_j)]$ in the case of a piecewise linear ψ and $\gamma_2 = \mathsf{E}\,\psi''(e_1)$ for a sufficiently smooth and integrable ψ. Note that if **Sym** holds then $\gamma_2 = 0$.

Theorem 1. *Put* $\mathbf{W}_{c,n} = \frac{1}{n}\sum_{i=1}^{n} c_i\, \mathbf{x}_i \mathbf{x}_i^{\mathsf{T}}$. *If* **X.1 − 4** *and* **Fix. 1 − 3** *hold, then*

$$\mathsf{E} \sup_{\mathbf{t}\in T} \Big| M_n(\mathbf{t}) - \tfrac{\gamma_2}{2}\, \mathbf{t}^{\mathsf{T}} \mathbf{W}_{c,n} \mathbf{t} \Big| = o(1). \tag{2.3}$$

Later it will be useful to rewrite the statement of Theorem 1 (with the help of Chebychev's inequality) as

$$\begin{aligned} \sum_{i=1}^{n} c_i\, \psi(e_i - \tfrac{\mathbf{t}^{\mathsf{T}}\mathbf{x}_i}{\sqrt{n}}) - \sum_{i=1}^{n} c_i\, \psi(e_i) + \tfrac{\gamma_1 \mathbf{t}^{\mathsf{T}}}{\sqrt{n}} \sum_{i=1}^{n} c_i\, \mathbf{x}_i \\ = -\tfrac{\mathbf{t}^{\mathsf{T}}}{\sqrt{n}} \sum_{i=1}^{n} c_i\, \mathbf{x}_i\, [\psi'(e_i) - \gamma_1] + \tfrac{\gamma_2}{2}\, \mathbf{t}^{\mathsf{T}} \mathbf{W}_{c,n} \mathbf{t} + o_p(1) \end{aligned} \tag{2.4}$$

uniformly in $\mathbf{t} \in T$.

2.2. Studentized M-processes

As the M-estimator is not in general scale invariant, in practice it is usually studentized. To investigate properties of the studentized M-estimators, it is useful to study the asymptotic properties of the 'studentized' M-process

$$M_n(\mathbf{t},u) = \sum_{i=1}^{n} c_i \Big[\psi\Big(e^{-u\,n^{-1/2}}(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\Big) - \psi(e_i/S) + \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{S\sqrt{n}}\,\psi'(e_i/S) + \tfrac{u}{\sqrt{n}}\tfrac{e_i}{S}\,\psi'(e_i/S)\Big],$$

where $(\mathbf{t},u) \in T = \{(\mathbf{s},v) : |\mathbf{s}|_2 \le M, |v| \le M\}$ $(\subset \mathbb{R}^{p+1})$ with M being an arbitrarily large but fixed constant.

As the studentization brings in perturbations in scale, more restrictive assumptions on the function ψ and the distribution of the errors than in the fixed scale case are needed.

St.1 ψ is absolutely continuous with a derivative ψ' such that

$$\mathsf{E}\left[\psi'\left(\tfrac{e_1}{S}\right)\right]^2 < \infty.$$

St.2 The (random) function $p(t,v) = \psi'(\frac{e_1+t}{Se^v})$ is continuous in the quadratic mean at the point $(0,0)$, that is

$$\lim_{(t,v)\to(0,0)} \mathsf{E}\left[p(t,v) - p(0,0)\right]^2 = \lim_{(t,v)\to(0,0)} \mathsf{E}\left[\psi'\left(\tfrac{e_1+t}{Se^v}\right) - \psi'\left(\tfrac{e_1}{S}\right)\right]^2 = 0.$$

St.3 The function $\lambda(t,v) = \mathsf{E}\,\psi(\frac{e_1+t}{Se^v})$ is twice differentiable and the second partial derivatives are continuous and bounded in a neighbourhood of the point $(0,0)$.

If the function ψ is twice differentiable almost everywhere then it is not difficult to show that assumptions **St. 1 − 3** are met if the following functions $\psi'(x)$, $x\,\psi'(x)$, $\psi''(x)$, $x\,\psi''(x)$ and $x^2\,\psi''(x)$ are bounded and continuous F-almost everywhere.

If ψ is a piecewise linear function, then the assumptions **St.1-3** are met provided **A.1-2** hold with the only modification that the points $r_1,\dots,r_k$ in **A.2** are replaced by the points $S\,r_1,\dots,S\,r_k$.

Before we proceed, it will be useful to introduce the following notation. Let the partial derivatives of the functions $\lambda(t,v) = \mathsf{E}\,\psi(\frac{e_1+t}{Se^v})$ and $\delta(t,v) = \mathsf{E}\,\frac{e_1}{S}\,\psi'(\frac{e_1+t}{Se^v})$ be indicated by subscripts. Put

$$\gamma_1 = \lambda_t(0,0)\ \left(= \tfrac{1}{S}\,\mathsf{E}\,\psi'\left(\tfrac{e_1}{S}\right)\right), \qquad \gamma_{1e} = -\lambda_v(0,0)\ \left(= \mathsf{E}\,\tfrac{e_1}{S}\,\psi'\left(\tfrac{e_1}{S}\right)\right), \tag{2.5}$$

$$\gamma_2 = \lambda_{tt}(0,0)\ \left(= \tfrac{1}{S^2}\,\mathsf{E}\,\psi''\left(\tfrac{e_1}{S}\right)\right), \qquad \gamma_{2e} = \delta_t(0,0)\ \left(= \mathsf{E}\,\tfrac{e_1}{S^2}\,\psi''\left(\tfrac{e_1}{S}\right)\right),$$

$$\gamma_{2ee} = -\delta_v(0,0)\ \left(= \mathsf{E}\left(\tfrac{e_1}{S}\right)^2\psi''\left(\tfrac{e_1}{S}\right)\right).$$

The formulas in the brackets are for the case of ψ sufficiently smooth and appropriately integrable. We do not give formulas for the case of a piecewise linear ψ as they are rather complicated in general case. According to the assumptions **St. 1 − 3** all these quantities are finite. Note that $\lambda_{tv}(0,0) = \gamma_1 + \gamma_{2e}$ and $\lambda_{vv}(0,0) = \gamma_{1e} + \gamma_{2ee}$.

Theorem 2. *If* **X.1-4** *and* **St. 1 − 3** *hold, then*

$$\mathsf{E} \sup_{(\mathbf{t},u)\in T} \Big| M_n(\mathbf{t}, u) - \tfrac{\gamma_2}{2}\, \mathbf{t}^\mathsf{T} \mathbf{W}_{c,n} \mathbf{t} - \tfrac{(\gamma_{2e}+\gamma_1) u\, \mathbf{t}^\mathsf{T}}{n} \sum_{i=1}^n c_i\, \mathbf{x}_i - \tfrac{(\gamma_{2ee}+\gamma_{1e})\, u^2}{2n} \sum_{i=1}^n c_i \Big| = o(1), \tag{2.6}$$

where $\mathbf{W}_{c,n}$ *was defined in Theorem 1.*

Remark 1. Note that if $\sum_{i=1}^n c_i = 0$, the last term (corresponding to small perturbations in scale) on the left-hand side of (2.6) vanishes. If assumption **Sym** (of symmetry) is satisfied, then $\gamma_2 = \gamma_{1e} = \gamma_{2ee} = 0$ and even the second term on the left-hand side of (2.6) disappears. Thus under assumption **Sym** Theorem 2 implies that

$$\begin{aligned} \sum_{i=1}^n c_i\, \psi \Big(e^{-u\, n^{-1/2}} (e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}})/S\Big) - \sum_{i=1}^n c_i\, \psi(e_i/S) + \tfrac{\gamma_1\, \mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\, \mathbf{x}_i \\ = -\tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n c_i\, \mathbf{x}_i \left[\tfrac{1}{S}\, \psi'(e_i/S) - \gamma_1\right] - \tfrac{u}{\sqrt{n}} \sum_{i=1}^n c_i \left[\tfrac{e_i}{S}\, \psi'(e_i/S)\right] \\ + \tfrac{(\gamma_{2e}+\gamma_1) u\, \mathbf{t}^\mathsf{T}}{n} \sum_{i=1}^n c_i\, \mathbf{x}_i + o_p(1), \end{aligned} \tag{2.7}$$

uniformly in $(\mathbf{t}, u) \in T$.

3. Second order asymptotic representation of M-estimators

In Section 2 technical results on approximation of linear processes associated with M-estimation in linear models were presented. One of the possible applications of these results is deriving a two term von Mises expansion of M-estimators defined in (1.2).

3.1. First order asymptotic representation (FOAR)

Deriving the second order asymptotic representation of a fixed scale M-estimator is very straightforward provided one is allowed to substitute the parameter $\mathbf{t}$ in the asymptotic expansion (2.4) with $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. To justify this substitution the estimator $\hat{\boldsymbol{\beta}}_n$ has to be $\sqrt{n}$-root consistent, that is $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = O_p(1)$. That is guaranteed by the following two assumptions:

Fix. 4 (St.4) The function $h(t) = \mathsf{E}\, \rho(e_1 - t)$ (or $h(t) = \mathsf{E}\, \rho(\frac{e_1 - t}{S})$) has a unique minimum at $t = 0$, that is for every $\delta > 0$: $\inf_{|t|>\delta} h(t) > h(0)$.

XX.2 $\mathbf{V} = \lim_{n\to\infty} \mathbf{V}_n$, where $\mathbf{V}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i\, \mathbf{x}_i^\mathsf{T}$ and $\mathbf{V}$ is a positive definite $p \times p$ matrix.

With the help of **Fix. 4**, **XX.2** and Theorem 1 which implies

$$\sup_{\mathbf{t}\in T} \left| \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \left[\psi(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \psi(e_i)\right] + \gamma_1 \mathbf{V}_n \mathbf{t} \right| = o_p(1), \tag{3.1}$$

one can use the technique of the proof of Theorem 5.5.1 of Jurečková and Sen [14] to show that there exists a root $\hat{\boldsymbol\beta}_n$ of system of equations (1.2) such that

$$\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta) = O_p(1). \tag{3.2}$$

Now inserting $\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta)$ for the parameter $\mathbf{t}$ in (3.1) gives the first order asymptotic representation (1.3).

3.1.1. FOAR for a studentized M-estimator

To be able to be as explicit as possible we will concentrate on models (1.1) that include an intercept, that is $x_{i1} = 1$ for $i = 1, \dots, n$. Let us also assume the scale estimator S_n to be $\sqrt{n}$-consistent, that is there exists a finite positive constant S such that

$$\sqrt{n}(\tfrac{S_n}{S} - 1) = O_p(1). \tag{3.3}$$

Similarly as for a fixed scale M-estimator one can derive the first order asymptotic representation

$$\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta) = \frac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\,\psi\left(\tfrac{e_i}{S}\right) - \tfrac{\gamma_{1e}}{\gamma_1}\sqrt{n}(\tfrac{S_n}{S} - 1)\,\mathbf{u}_1 + o_p(1), \tag{3.4}$$

where $\mathbf{u}_1 = (1, 0, \dots, 0)^\mathsf{T} \in \mathbb{R}_p$ and γ_1, γ_{1e} are defined in (2.5) of Section 2.2. Note that the FOAR of the slope part of $\hat{\boldsymbol\beta}_n$ does not depend on the asymptotic distribution of the scale estimator S_n. This holds true also for the intercept provided the assumption of symmetry **Sym** is satisfied, which implies $\gamma_{1e} = 0$.

3.2. Second order asymptotic representation (SOAR)

3.2.1. SOAR for a fixed-scale M-estimator

For our convenience let us restate expansion (2.4) for the vector case. For $l = 1, \dots, p$ put $\mathbf{W}_{nl} = \frac{1}{n}\sum_{i=1}^n x_{li}\,\mathbf{x}_i\mathbf{x}_i^\mathsf{T}$ and let $\mathbf{W}_n$ be a bilinear form from $\mathbb{R}^p \times \mathbb{R}^p$ to $\mathbb{R}^p$ given by

$$\mathbf{W}_n(\mathbf{t}, \mathbf{s}) = (\mathbf{t}^\mathsf{T}\mathbf{W}_{n1}\,\mathbf{s}, \dots, \mathbf{t}^\mathsf{T}\mathbf{W}_{np}\,\mathbf{s})^\mathsf{T}.$$

Corollary 1. *Assume* **XX.1** *and* **Fix. 1 − 3**, *then it holds uniformly in* $\mathbf{t} \in T$

$$\begin{aligned}\sum_{i=1}^n \mathbf{x}_i\,\psi(e_i - \tfrac{\mathbf{t}^\mathsf{T}\mathbf{x}_i}{\sqrt{n}}) - \sum_{i=1}^n \mathbf{x}_i\,\psi(e_i) + \gamma_1\sqrt{n}\,\mathbf{V}_n\,\mathbf{t} \\ = -\tfrac{\mathbf{t}^\mathsf{T}}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\,\mathbf{x}_i^\mathsf{T}[\psi'(e_i) - \gamma_1] + \tfrac{\gamma_2}{2}\,\mathbf{W}_n(\mathbf{t}, \mathbf{t}) + o_p(1).\end{aligned} \tag{3.5}$$

The proof follows by applying Theorem 1 to each of the coordinate separately.

As the estimator $\hat{\boldsymbol\beta}_n$ satisfies (3.2), $\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta)$ can be substituted for $\mathbf{t}$ in (3.5). The first order asymptotic representation (1.3) and some algebraic manipulations yield

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \tag{3.6}$$
$$= -\tfrac{1}{\sqrt{n}} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \mathbf{x}_i^{\mathsf{T}} [\psi'(e_i) - \gamma_1] \right\} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \right\}$$
$$+ \tfrac{\gamma_2\, \mathbf{V}_n^{-1}}{2\,\gamma_1\,\sqrt{n}}\, \mathbf{W}_n \left(\tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i),\ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i) \right) + o_p(\tfrac{1}{\sqrt{n}}).$$

If the symmetry assumption **Sym** is satisfied, then the second term on the right-hand side vanishes and both factors in the first term are asymptotically normal as well as asymptotically independent. This is in agreement with the results of Jurečková and Sen [13] where the asymptotic distribution of the second term in the von Mises expansion is shown to be a product of two normal distributions.

3.2.2. SOAR for a studentized M-estimator

If $\sqrt{n}$-consistency of S_n as expressed by (3.3) holds and assumptions **St.1-4** and **XX.1-2** are satisfied, one can proceed very similarly as for the fixed scale M-estimators. Informally speaking, the second order asymptotic representation for studentized M-estimators may be found by substituting $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$, $\sqrt{n}\log(\frac{S_n}{S})$ for u and $\mathbf{x}_i$ for c_i in (2.7). But as the resulting expression is rather long, we will write it down only when the assumption of symmetry **Sym** holds. After some algebra we get

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i/S) \tag{3.7}$$
$$= -\tfrac{1}{\sqrt{n}} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \mathbf{x}_i^{\mathsf{T}} [\psi'(e_i/S) - \gamma_1] \right\} \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i/S) \right\}$$
$$- \tfrac{1}{\sqrt{n}} \left\{ \sqrt{n}(\tfrac{S_n}{S} - 1)\, \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \left[\tfrac{e_i}{S}\, \psi'(e_i/S) \right] \right\}$$
$$+ \tfrac{\gamma_{2e}+\gamma_1}{\gamma_1\,\sqrt{n}}\, \sqrt{n}(\tfrac{S_n}{S} - 1) \left\{ \tfrac{\mathbf{V}_n^{-1}}{\gamma_1\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i\, \psi(e_i/S) \right\} + o_p(\tfrac{1}{\sqrt{n}}).$$

Inspecting (3.7) it may be of interest to note that although the first order asymptotic distribution of a studentized M-estimator of the slope parameters does not depend on the asymptotic distribution of S_n, the second order distribution does, even if the assumption **Sym** is satisfied. Thus when excluding artificial or pathological examples, the studentized M-estimator cannot be asymptotically equivalent of second order with an R-estimator or a fixed scale M-estimator.

4. Conclusions

We have presented a way how to derive a second order asymptotic representation of an M-estimator in a linear model with fixed carriers. This representation may be used e. g. to compare the M-estimator $\hat{\boldsymbol{\beta}}_n$ with another estimator that is asymptotically equivalent to $\hat{\boldsymbol{\beta}}_n$. This may be for example a one-step M-estimator (see e. g. Welsh and Ronchetti [19]) or an appropriate R-estimator (see Hušková and Jurečková [8] and Jurečková [10]). For instance, it is well known that if $\psi(x)$ is proportional to $(F(x) - \frac{1}{2})$, then the fixed-scale M-estimator is asymptotically equivalent to an R-estimator based on the Wilcoxon scores. Our results can be used for a finer comparison of those estimators. The second order asymptotic results also proved to be useful when investigating 'Rao Score type' confidence interval, see Omelka [17].

Acknowledgements

The author wish to express his thanks to Prof. Jana Jurečková for her encouragement, guidance and support when supervising his PhD thesis.

The author is also thankful to two anonymous referees for their remarks and comments.

References

[1] Arcones, M. A. (1996). The Bahadur–Kiefer representation for U-quantiles. *Ann. Statist.* **24** 1400–1422.

[2] Arcones, M. A. (1996). The Bahadur–Kiefer Representation of L_p Regression Estimators. *Econometric Theory* **12** 257–283.

[3] Arcones, M. A. (1998). The Bahadur–Kiefer representation of two dimensional spatial medians. *Ann. Inst. Stat. Math* **50** 71–86.

[4] Arcones, M. A. (1998). Second order representations of the least absolute deviation regression estimator. *Ann. Inst. Stat. Math* **50** 87–117.

[5] Arcones, M. A. and Mason, D. M. (1997). A general approach to Bahadur–Kiefer representations for M-estimators. *Mathematical Methods of Statistics* **6** 267–292.

[6] Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

[7] Boos, D. D. (1977). Comparison of L- and M-estimators using the second term of the von Mises expansion. *Tech. rep.*, North Carolina State University, Raleigh, North Carolina.

[8] Hušková, M. and Jurečková, J. (1981). Second order asymptotic relations of M-estimators and R-estimators in two-sample location model. *J. Statist. Plann. Inference* **5** 309–328.

[9] Jurečková, J. (1973). Central limit theorem for Wilcoxon rank statistics process. *Ann. Statist.* **1** 1046–1060.

[10] Jurečková, J. (1977). Asymptotic Relations of M-estimates and R-estimates in linear regression model. *Ann. Statist.* **5** 464–472.

[11] Jurečková, J. and Sen, P. K. (1989). A second-order asymptotic distributional representation of M-estimators with discontinuous score functions. *Annals of Probability* **15** 814–823.

[12] Jurečková, J. and Sen, P. K. (1989). Uniform second order asymptotic linearity of M-statistics in linear models. *Statist. Dec.* **7** 263–276.

[13] Jurečková, J. and Sen, P. K. (1990). Effect of the initial estimator on the asymptotic behavior of one-step M-estimator. *Ann. Inst. Statist. Math.* **42** 345–357.

[14] Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* Wiley, New York.

[15] Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Statist.* **38** 1323–1342.

[16] Knight, K. (1997). Asymptotics for L_1 regression estimators under general conditions. Technical Report 9716, Dept. Statistics, Univ. Toronto.

[17] Omelka, M. (2006). An alternative method for constructing confidence intervals from M-estimators in linear models. In: *Proceedings of Prague Stochastics 2006*, 568–578.

[18] Omelka, M. (2006). Second order properties of some M-estimators and R-estimators. *Ph.D. thesis*, Charles University in Prague, available at http://www.karlin.mff.cuni.cz /~omelka/Soubory/omelka_thesis.pdf.

[19] Welsh, A. H. and Ronchetti, E. (2002). A journey in single steps: robust one-step M-estimation in linear regression. *J. Statist. Plann. Inference* **103** 287–310.