PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 15 (1995), pp. 397–407

AFFINE- AND SCALE-EQUIVARIANT M-ESTIMATORS IN LINEAR MODEL*

BY

JANA JUREČKOVÁ (PRAGUE)

Abstract. M-estimators of regression parameter vector in linear model, studentized by a suitable affine-invariant and scale-equivariant scale statistic, become affine- and scale-equivariant. We study some asymptotic properties of studentized M-estimators and give a brief review of suitable studentizing statistics, proposed in the literature.

1. Introduction. Consider the linear regression model

$$Y = X\beta + E, \tag{1.1}$$

where $Y = (Y_1, \ldots, Y_n)'$ is a vector of observations, $X = X_n$ is a (known or observable) design matrix of order $n \times p$, $\beta = (\beta_1, \ldots, \beta_p)'$ is an unknown unobservable parameter, and $E = (E_1, \ldots, E_n)'$ is a vector of i.i.d. errors with an (unknown) distribution function (d.f.) F. Among various robust estimators of β, the M-*estimators* and GM-*estimators* apparently attained the most of the popularity.

Given an absolutely continuous function $\varrho\colon \mathscr{R}^1 \to \mathscr{R}^1$ with derivative ψ, we define an M-*estimator* M_n of β as a solution of the minimization

$$\sum_{i=1}^{n} \varrho(Y_i - x_i' t) := \min \tag{1.2}$$

with respect to $t \in \mathscr{R}^p$, where x_i' is the i-th row of X_n, $i = 1, \ldots, n$. Such an estimator is robust with respect to deviations in the distribution of E_1 (see [8]) and its influence function is bounded provided X_n is a fixed suitably bounded matrix. However, M_n is not robust with respect to X; it is sensitive to eventual leverage points in X and the influence function of M_n is unbounded in the case of random X_n (see [6]). The latter shortage is usually treated by using the *generalized* M-*estimators* (GM-*estimators*; see, e.g., [16]).

* Research supported by the Grant of the Charles University GAUK-365.

The M-estimator defined in (1.2) is *affine-equivariant*, i.e.,

$$M_n(Y+Xb) = M_n(Y)+b \quad \text{for all } Y \in \mathscr{R}^n,\ b \in \mathscr{R}^p. \tag{1.3}$$

However, such an M-estimator is generally not *scale-equivariant*, i.e., the following condition does not hold in general:

$$M_n(aY) = aM_n(Y) \quad \text{for all } a > 0,\ Y \in \mathscr{R}^n. \tag{1.4}$$

This could be regarded as a lack of robustness, because such an estimator of β is uncontrollably sensitive to eventual changes in the measurement precision of the laboratory, etc. For this reason, M-estimators or GM-estimators of β are calculated simultaneously with some measure of scale; for instance, given the observations (x_i', Y_i), $i = 1, \ldots, n$, a GM-*estimator* M_n of β is calculated as a solution of the system of equations

$$\sum_{i=1}^{n} \psi\big(x_i', (Y_i - x_i'M_n)/S_n\big)x_i = 0, \tag{1.5}$$

where $\psi\colon \mathscr{R}^{p+1} \to \mathscr{R}^1$ and S_n is an estimator of scale which is computed simultaneously with M_n by means of the equation

$$\sum_{i=1}^{n} \chi\big((Y_i - x_i'M_n)/S_n\big) = 0 \tag{1.6}$$

with some $\chi\colon \mathscr{R}^1 \to \mathscr{R}^1$. Hence, we supplement the GM-estimator by an estimator S_n of scale and in this way we obtain an affine- and scale-equivariant estimator of β. As an alternative to the proposal (1.5)–(1.6), we may consider *studentized* M-*estimators* (or GM-*estimators*) defined as a solution of the minimization

$$\sum_{i=1}^{n} \varrho\big((Y_i - x_i't)/S_n\big) := \min \tag{1.7}$$

with an appropriate scale statistic $S_n = S_n(Y)$, not necessarily one defined in (1.5) and (1.6), which is *affine-invariant* and *scale-equivariant* in the sense that

$$S_n(Y+Xb) = S_n(Y) \quad \text{for all } Y \in \mathscr{R}^n,\ b \in \mathscr{R}^p, \tag{1.8}$$

and

$$S_n(aY) = aS_n(Y) \quad \text{for all } a > 0,\ Y \in \mathscr{R}^n. \tag{1.9}$$

Generally, S_n consistently estimates its population counterpart $S(F)$ and in this way brings an additional information about the model. However, the exact value of $S(F)$ is not so important in the equivariant situation; our ultimate goal is to obtain an affine- and scale-equivariant studentized M-estimator. On the other hand, it may be of interest to estimate the covariance matrix of M_n itself.

In the present paper, we shall characterize some asymptotic properties of studentized M-estimators. In Section 2 we shall formulate the asymptotic representations of M_n, which were derived in [13] with the aid of the second-order uniform asymptotic linearity of M-statistics of Jurečková and Sen [10]. These asymptotic representations yield many interesting properties of studentized M-estimators, some of them even not yet explicitly mentioned in the literature.

We shall notice that the asymptotic representation of a studentized M-estimator generally involves S_n unless F is symmetric or in some other special cases. The covariance matrix of the asymptotic distribution of $n^{1/2}(M_n-\beta)$ includes $S(F)$, and hence (and this applies already to the location model) the studentized M-estimator is not any more asymptotically minimax over a contaminated distribution family [8]. However, the asymptotic minimax property of nonstudentized M-estimator concerns the contamination "neighborhood" of one fixed distribution and not that of a distribution law; the sensitivity of an estimator to a "contaminated distribution shape" would be an interesting and challenging problem.

We shall also examine which scale statistics are suitable for the studentization in a regression setup. Surprisingly, not many affine-invariant and scale-equivariant scale statistics for the linear model were described in the literature. The classical scale statistic, the root of the residual sum of squares,

$$S_n(Y) = (Y'(I_n-\hat{H}_n)Y)^{1/2} \tag{1.10}$$

with the projection matrix $\hat{H}_n = X(X'X)^{-1}X'$, satisfies (1.8) and (1.9) but is sensitive to deviations from the normal shape, and hence nonrobust. In Section 3 we shall give a brief review of suitable alternative studentizing statistics for the linear model, which were proposed in the literature.

2. Asymptotic representation of studentized M-estimators. We shall work with the model (1.1) and with the studentized M-estimator of β defined as a solution of the minimization (1.7). Assume that ϱ: $\mathscr{R}^1 \to \mathscr{R}^1$ is absolutely continuous with the nonconstant derivative $\psi = \varrho'$ which could be decomposed into the sum

$$\psi = \psi_a+\psi_c+\psi_s, \tag{2.1}$$

where ψ_a is an absolutely continuous function with absolutely continuous derivative $\psi_a'(z) = d\psi_a(z)/dz$, ψ_c is a continuous, piecewise linear function which is constant in a neighborhood of $\pm\infty$, and ψ_s is a nondecreasing step-function. Either of functions ψ_a, ψ_c, ψ_s may vanish.

Moreover, we shall impose the following conditions on the quantities in (1.7):

(M.1) $S_n = S_n(Y) > 0$ a.s., satisfies (1.8) and (1.9) (affine-invariance and scale-equivariance) and there exists a functional $S = S(F) > 0$ such that

$$n^{1/2}(S_n - S) = O_p(1).$$

(M.2) The function $h(t) = \int \varrho((z-t)/S)dF(z)$ has the unique minimum at $t = 0$.

(M.3) For some $\delta > 0$ and $\eta > 1$,

$$\int_{-\infty}^{\infty} \{|z| \sup_{|u| \leqslant \delta} \sup_{|v| \leqslant \delta} |\psi_a''(e^{-v}(z+u)/S)|\}^\eta dF(z) < \infty$$

and

$$\int_{-\infty}^{\infty} \{z^2 \sup_{|u| \leqslant \delta} \psi_a''((z+u)/S)\}^\eta dF(z) < \infty.$$

(M.4) ψ_c is a continuous, piecewise linear function with knots at $r_1, \ldots, r_k$, which is constant in a neighborhood of $\pm\infty$; hence the derivative ψ_c' of ψ_c is a step-function,

$$\psi_c'(z) = \alpha_\nu \quad \text{for } r_\nu < z < r_{\nu+1},\ \nu = 0, 1, \ldots, k,$$

where $\alpha_0, \alpha_1, \ldots, \alpha_k \in \mathscr{R}^1$, $\alpha_0 = \alpha_k = 0$ and $-\infty < r_0 < r_1 < \ldots < r_k < r_{k+1} = \infty$. We assume that $f(z) = dF(z)/dz$ exists and is bounded in neighborhoods of $Sr_1, \ldots, Sr_k$.

(M.5) $\psi_s(z) = \lambda_\nu$ for $q_\nu < z \leqslant q_{\nu+1}$, $\nu = 1, \ldots, m$, where

$$-\infty = q_0 < q_1 < \ldots < q_m < q_{m+1} = \infty, \qquad -\infty < \lambda_0 < \lambda_1 < \ldots < \lambda_m < \infty.$$

We assume that $f(z) = dF(z)/dz$ and $f'(z) = d^2F(z)/dz^2$ exist and are bounded in neighborhoods of $Sq_1, \ldots, Sq_m$.

Remark. Condition (M.3) is essentially a moment condition which holds, e.g., if ψ_a'' is bounded and either

(2.2) $$\psi_a''(z) = 0 \quad \text{for } z < a \text{ or } z > b,\ -\infty < a < b < \infty,$$

or

(2.3) $$\int_{-\infty}^{\infty} |z|^{2+\varepsilon} dF(z) < \infty \quad \text{for some } \varepsilon > 0.$$

Conditions (M.4) and (M.5) show explicitly the trade-off between the smoothness of ψ and smoothness of F. The class of ψ_c-functions covers the usual Huber's and Hampel's proposals (see [5]–[7]).

Moreover, we shall impose the following conditions on the matrix X_n:

(X.1) $x_{i1} = 1$, $i = 1, \ldots, n$.

(X.2) $n^{-1}\sum_{i=1}^n \|x_i\|^4 = O(1)$ as $n \to \infty$.

(X.3) $\lim_{n\to\infty} \boldsymbol{Q}_n = \boldsymbol{Q}$, where $\boldsymbol{Q}_n = n^{-1}\boldsymbol{X}_n'\boldsymbol{X}_n$ and $\boldsymbol{Q}$ is a positive definite $(p \times p)$-matrix.

Let $\boldsymbol{M}_n$ be a solution of the minimization (1.7). The definition of $\boldsymbol{M}_n$ should be supplemented by a rule as defined in the case $S_n = 0$; however, this event happens with probability 0 and the choice of the rule does not affect the asymptotic behavior of $\boldsymbol{M}_n$. If $\psi = \varrho'$ is continuous (i.e., $\psi_s = 0$), then $\boldsymbol{M}_n$ is a solution of the system of equations

$$\sum_{i=1}^{n} \boldsymbol{x}_i \psi((Y_i - \boldsymbol{x}_i'\boldsymbol{t})/S_n) = 0. \tag{2.4}$$

The system (2.4) may have more roots corresponding to the local minima, etc.; however, we may show that, under our conditions, there exists a root of (2.4) which is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$. In situations of nonconvex ϱ we are rather able to work with this root than with the global minimum of (1.7).

On the other hand, if ψ is a step-function and it is nondecreasing, then $\boldsymbol{M}_n$ minimizes a convex function and its consistency and asymptotic normality may be proved by using a different argument.

The asymptotic representations of $\boldsymbol{M}_n$ will involve the functionals

$$\gamma_1 = S^{-1} \int_{-\infty}^{\infty} [\psi_a'(z/S) + \psi_c'(z/S)]\, dF(z), \tag{2.5}$$

$$\gamma_2 = S^{-1} \int_{-\infty}^{\infty} z[\psi_a'(z/S) + \psi_c'(z/S)]\, dF(z), \tag{2.6}$$

$$\gamma_1^* = \sum_{\nu=1}^{m} (\lambda_\nu - \lambda_{\nu-1}) f(Sq_\nu), \tag{2.7}$$

$$\gamma_2^* = S \sum_{\nu=1}^{m} (\lambda_\nu - \lambda_{\nu-1})\, q_\nu f(Sq_\nu). \tag{2.8}$$

The asymptotic representations are formulated in the following theorems.

THEOREM 2.1. *Consider the model* (1.1) *under the conditions* (M.1)–(M.5), (X.1)–(X.3) *and assume that* $\gamma_1 > 0$ *and* $\psi_s \equiv 0$.

(i) *Then there exists a root* $\boldsymbol{M}_n$ *of the system* (2.4) *such that*

$$n^{1/2} \|\boldsymbol{M}_n - \boldsymbol{\beta}\| = O_p(1) \quad \text{as } n \to \infty. \tag{2.9}$$

(ii) *Any root of* (2.4) *satisfying* (2.9) *admits the asymptotic representation*

$$\boldsymbol{M}_n - \boldsymbol{\beta} = (n\gamma_1)^{-1} \boldsymbol{Q}_n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \psi(E_i/S) - (\gamma_2/\gamma_1)(S_n/S - 1)\boldsymbol{e}_1 + \boldsymbol{R}_n, \tag{2.10}$$

where $\|\boldsymbol{R}_n\| = O_p(n^{-1})$ *and* $\boldsymbol{e}_1 = (1, 0, \ldots, 0)' \in \mathcal{R}^p$.

THEOREM 2.2. *Consider the linear model* (1.1) *under the conditions* (M.1), (M.2), (X.1)–(X.3) *and assume* $\gamma_1^* > 0$ *and* $\psi_a = \psi_c = 0$. *Let* M_n *be a point of the global minimum of* (1.7). *Then*

$$n^{1/2}\|M_n - \beta\| = O_p(1) \quad \text{as } n \to \infty \tag{2.11}$$

and M_n *admits the representation*

$$M_n - \beta = (n\gamma_1^*)^{-1} Q_n^{-1} \sum_{i=1}^{n} x_i \psi(E_i/S) - (\gamma_2^*/\gamma_1^*)(S_n/S - 1)e_1 + R_n, \tag{2.12}$$

where $\|R_n\| = O_p(n^{-3/4})$ *as* $n \to \infty$.

Remark. Notice that S_n affects only the first component (intercept) of M_n in the representations (2.10) and (2.12).

Combining the above results, we immediately obtain the following result for the general class of M-estimators.

THEOREM 2.3. *Consider the model* (1.1) *under the conditions* (M.1)–(M.5) *and* (X.1)–(X.3). *Let* ψ *be either continuous or monotone and let* $\gamma_1 + \gamma_1^* > 0$. *Then, for any* M-*estimator* M_n *satisfying* $n^{1/2}\|M_n - \beta\| = O_p(1)$,

$$M_n - \beta = \big(n(\gamma_1 + \gamma_1^*)\big)^{-1} Q_n^{-1} \sum_{i=1}^{n} x_i \psi(E_i/S) - \frac{\gamma_2 + \gamma_2^*}{\gamma_1 + \gamma_1^*}\left(\frac{S_n}{S} - 1\right)e_1 + R_n, \tag{2.13}$$

where

$$\|R_n\| = \begin{cases} O_p(n^{-1}) & \text{if } \psi_s \equiv 0, \\ O_p(n^{-3/4}) & \text{otherwise.} \end{cases} \tag{2.14}$$

Theorems 2.1–2.3 follow, after a slight modification, from Jurečková and Welsh [13]; hence, we omit the details. We shall rather concentrate on various interesting asymptotic properties of studentized M-estimators, which are implied by the above representations.

PROPOSITION 2.1. *Under the conditions of Theorem* 2.1, *let* $M_n^{(1)}$ *and* $M_n^{(2)}$ *be any pair of roots of the system of equations* (2.4) *both satisfying* (2.9). *Then*

$$\|M_n^{(1)} - M_n^{(2)}\| = O_p(n^{-1}). \tag{2.15}$$

Proof. By Theorem 2.1, every root of (2.4) satisfying (2.9) admits the asymptotic representation (2.10); this immediately implies (2.15). ∎

PROPOSITION 2.2. *Assume that*

$$\sigma^2 = \int_{-\infty}^{\infty} \psi^2(z/S)\,dF(z) < \infty. \tag{2.16}$$

Then, under the conditions of Theorems 2.1–2.3,

$$n^{1/2}\{\tilde{\gamma}_1(M_n - \beta) + \tilde{\gamma}_2(S_n/S - 1)e_1\} \tag{2.17}$$

has the asymptotic p-dimensional normal distribution $\mathcal{N}_p(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1})$; *here* $\tilde{\gamma}_i$ *stands for* γ_i, γ_i^* *or* $\gamma_i + \gamma_i^*$, *respectively*, $i = 1, 2$.

Proof. The asymptotic distribution follows from the respective representations in Theorems 2.1–2.3 via the central limit theorem. ∎

PROPOSITION 2.3. *Assume* (2.16) *and the symmetry of* F *and* ϱ, *i.e.*,

$$F(z)+F(-z) = 1 \quad \text{and} \quad \varrho(-z) = \varrho(z), \quad z \in \mathscr{R}^1. \tag{2.18}$$

Then, under the conditions of either of Theorems 2.1–2.3,

$$\mathbf{M}_n - \boldsymbol{\beta} = (n\tilde{\gamma}_1)^{-1} \mathbf{Q}_n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \psi(E_i/S) + \mathbf{R}_n, \tag{2.19}$$

where $\|\mathbf{R}_n\| = O_p(n^{-1})$ *provided* $\psi_s = 0$ *and* $\|\mathbf{R}_n\| = O_p(n^{-3/4})$ *otherwise. Moreover,* $n^{1/2}(\mathbf{M}_n - \boldsymbol{\beta})$ *has the asymptotic normal distribution*

$$\mathcal{N}_p(O, (\sigma^2/\tilde{\gamma}_1^2) \mathbf{Q}^{-1}) \tag{2.20}$$

with the same $\tilde{\gamma}_1$ *as in Proposition* 2.2.

Proof. Notice that $\gamma_2^* = 0$ in the symmetric case; the proposition then follows from the asymptotic representations in Theorems 2.1–2.3 and from Proposition 2.2. ∎

Without symmetry conditions, but assuming that S_n itself admits an asymptotic representation, we could still obtain a representation and asymptotic distribution for $\mathbf{M}_n$:

PROPOSITION 2.4. *Assume that* S_n *admits an asymptotic representation*

$$S_n/S - 1 = n^{-1} \sum_{i=1}^{n} \varphi(E_i) + o_p(n^{-1/2}), \tag{2.21}$$

where $\int_{-\infty}^{\infty} \varphi(z)dF(z) = 0$, $\eta^2 = \int_{-\infty}^{\infty} \varphi^2(z)dF(z) < \infty$. *Then, under the conditions of Theorems* 2.1–2.3,

$$\mathbf{M}_n - \boldsymbol{\beta} = (n\tilde{\gamma}_1)^{-1} \mathbf{Q}_n^{-1} \sum_{i=1}^{n} [\mathbf{x}_i \psi(E_i/S) - \mathbf{e}_1 \tilde{\gamma}_2 \varphi(E_i)] + o_p(n^{-1/2}), \tag{2.22}$$

and $n^{1/2}(\mathbf{M}_n - \boldsymbol{\beta})$ *has the asymptotic normal distribution*

$$\mathcal{N}_p(\mathbf{0}, \tilde{\gamma}_1^{-2}\{\sigma^2 \mathbf{Q}^{-1} + [\sigma_2^2 - 2\sigma_{12}] \mathbf{e}_1 \mathbf{e}_1'\}) \tag{2.23}$$

with

$$\sigma_{12} = \int_{-\infty}^{\infty} \tilde{\gamma}_2 \varphi(z) \psi(z/S) dF(z), \quad \sigma_2^2 = \int_{-\infty}^{\infty} (\tilde{\gamma}_2 \varphi(z))^2 dF(z). \tag{2.24}$$

Proof. The asymptotic representation (2.22) follows from the respective representations in Theorems 2.1–2.3 combined with (2.21). The asymptotic distribution then follows from the central limit theorem. ∎

Remark 1. In the location submodel corresponding to $X_n = 1_n$, the studentized M-estimator of the location parameter is asymptotically normally distributed with zero expectation and with the variance

$$\eta^2 = \tilde{\gamma}_1^{-2} \int_{-\infty}^{\infty} \{\psi(z/S) - \tilde{\gamma}_2 \varphi(z)\}^2 dF(z).$$

Remark 2. Notice that, in the linear regression model (1.1) with intercept and with asymmetric distribution of errors, the studentization has an impact only on the intercept component of the estimator; the variance of the intercept component in the asymptotic distribution coincides with η^2. On the other hand, the marginal covariance matrix corresponding to $(\beta_2, \ldots, \beta_p)'$ in the asymptotic distribution is $Q^{(1)}(\sigma^2/\tilde{\gamma}_1^2)$, where $Q^{(1)}$ is a submatrix of Q^{-1}.

3. Some studentizing scale statistics. In this section we shall briefly describe some scale statistics S_n which are affine-invariant and scale-equivariant (see (1.7) and (1.8)). The root of residual sum of squares (1.10) satisfies both (1.7) and (1.8) but is closely connected with the classical normal model and sensitive to the deviations from the same.

(i) *An extension of* MAD (*median absolute deviation from the median*) to the model (1.1) was proposed by Welsh [18]. Starting with an initial $\sqrt{n}$-consistent and affine- and scale-quivariant estimator $\hat{\beta}^0$ of β, let us put

$$Y_i(\hat{\beta}^0) = Y_i - x_i'\hat{\beta}^0, \quad i = 1, \ldots, n, \tag{3.1}$$

$$\zeta_{1/2}(\hat{\beta}^0) = \operatorname*{med}_{1 \leqslant i \leqslant n} Y_i(\hat{\beta}^0), \tag{3.2}$$

$$S_n = \operatorname*{med}_{1 \leqslant i \leqslant n} |Y_i(\hat{\beta}^0) - \zeta_{1/2}(\hat{\beta}^0)|. \tag{3.3}$$

Then S_n satisfies (1.8) and (1.9) provided $\hat{\beta}_0$ is affine- and scale-equivariant. Moreover, under some regularity conditions on F, S_n is a $\sqrt{n}$-consistent estimator of the population median deviation; Welsh [18] also derived its asymptotic representation of type (2.21), hence Proposition 2.4 applies.

(ii) L-*statistics based on regression quantiles.* The α-regression quantile $\hat{\beta}(\alpha)$ for the model (1.1) was introduced by Koenker and Bassett [14] as a solution of the minimization problem

$$\sum_{i=1}^{n} \varrho_\alpha(Y_i - x_i't) := \min \quad \text{with respect to } t \in \mathscr{R}^p, \tag{3.4}$$

where

$$\varrho_\alpha(z) = |z|\{\alpha I[z > 0] + (1-\alpha) I[z < 0]\}, \quad z \in \mathscr{R}^1. \tag{3.5}$$

The Euclidean distance of two regression quantiles

(3.6) $$S_n = \|\hat{\boldsymbol{\beta}}_n(\alpha_2)-\hat{\boldsymbol{\beta}}_n(\alpha_1)\|, \quad 0 < \alpha_1 < \alpha_2 < 1,$$

satisfies (1.8) and (1.9) and $S_n \xrightarrow{P} S(F) = F^{-1}(\alpha_2)-F^{-1}(\alpha_1)$; its asymptotic representation follows, e.g., from that for $\hat{\boldsymbol{\beta}}_n(\alpha)$ derived by Ruppert and Carroll [17]. The Euclidean norm may be replaced by L_p-norm or by another appropriate norm. An alternative statistic is the deviation of the first components of regression quantiles, $S_n = \hat{\beta}_{n1}(\alpha_2)-\hat{\beta}_{n1}(\alpha_1)$, $0 < \alpha_1 < \alpha_2 < 1$, with the same population counterpart.

More generally, Bickel and Lehmann [1] proposed various measures of spread of the distribution F, which could also serve as the scale functional $S(F)$; the corresponding scale statistic is then an estimator of $S(F)$ based on regression quantiles. As an example, we could consider

(3.7) $$S(F) = \{\int_{1/2}^{1} [F^{-1}(u)-F^{-1}(1-u)]^2 d\Lambda(u)\}^{1/2},$$

where Λ is the uniform distribution on $(\frac{1}{2}, 1-\delta)$, $0 < \delta < \frac{1}{2}$; then

(3.8) $$S_n(F) = \{\int_{1/2}^{1} \|\hat{\boldsymbol{\beta}}(u)-\hat{\boldsymbol{\beta}}(1-u)\|^2 d\Lambda(u)\}^{1/2}.$$

(iii) Falk [3] proposed a histogram- and kernel-type estimators of the value $1/f(F^{-1}(\alpha))$, $0 < \alpha < 1$, in the location model, $f(x) = dF(x)/dx$. Dodge and Jurečková [2] extended Falk's estimators to the linear model in the following way.

First,

(3.9) $$H_n^{(\alpha)} = \{\hat{\beta}_{n1}(\alpha+\nu_n)-\hat{\beta}_{n1}(\alpha-\nu_n)\}/(2\nu_n),$$

where

(3.10) $$\nu_n = o(n^{-1/3}) \quad \text{and} \quad n\nu_n \to \infty \quad \text{as } n \to \infty,$$

is the histogram-type $(n\nu_n)^{1/2}$-consistent estimator of $1/f(F^{-1}(\alpha))$, satisfying (1.8) and (1.9).

Second, considering the kernel function $k\colon \mathscr{R}^1 \to \mathscr{R}^1$ with a compact support, which is continuous on its support and satisfies

(3.11) $$\int k(x)dx = 0 \quad \text{and} \quad \int xk(x)dx = -1,$$

we could construct the following kernel-type estimator of $1/f(F^{-1}(\alpha))$:

(3.12) $$\kappa_n^{(\alpha)} = \nu_n^{-2} \int_0^1 \hat{\beta}_{n1}(u)k\left(\frac{\alpha-u}{\nu}\right)du,$$

where

$$\text{(3.13)} \qquad \nu_n \to 0, \quad n\nu_n^2 \to \infty \quad \text{and} \quad n\nu_n^3 \to 0 \quad \text{as } n \to \infty;$$

then, again, $\kappa_n^{(\alpha)}$ is an $(n\nu_n)^{1/2}$-consistent estimator of $1/f(F^{-1}(\alpha))$, whose asymptotic variance may be less than that of (3.9) for some kernels. Regarding their lower rates of consistency, we shall not use the above estimators just for a simple studentization, but rather in an inference on the population quantiles, based on the regression data.

(iv) Jurečková and Sen [11] constructed scale statistics based on *regression rank scores*, which are dual to the regression quantiles in the linear programming sense and represent an extension of the rank scores to the linear model (see [4] for a detailed account of this concept). More precisely, regression rank scores $\hat{\boldsymbol{a}}_n(\alpha) = (\hat{a}_{n1}(\alpha), \ldots, \hat{a}_{nn}(\alpha))'$, $0 < \alpha < 1$, for the model (1.1) are defined as a solution of the maximization

$$\text{(3.14)} \qquad \boldsymbol{Y}'\hat{\boldsymbol{a}}_n(\alpha) := \max$$

under the restriction

$$\text{(3.15)} \qquad \boldsymbol{X}'(\hat{\boldsymbol{a}}_n(\alpha) - \boldsymbol{1}_n(1-\alpha)) = \boldsymbol{0}, \quad \hat{\boldsymbol{a}}_n(\alpha) \in [0, 1]^n, \quad 0 < \alpha < 1.$$

In the location model, they reduce to the rank scores considered in [5]. The proposed scale statistic is of the form

$$\text{(3.16)} \qquad S_n = n^{-1} \sum_{i=1}^{n} Y_i \hat{b}_{ni},$$

where

$$\text{(3.17)} \qquad \hat{b}_{ni} = -\int_{\alpha_0}^{1-\alpha_0} \varphi(\alpha)\, d\hat{a}_{ni}(\alpha),$$

$0 < \alpha_0 < \frac{1}{2}$, and φ: $[0, 1] \to \mathscr{R}^1$ is a nondecreasing, square-integrable and antisymmetric score generating function standardized so that $\int_{\alpha_0}^{1-\alpha_0} \varphi^2(\alpha)\, d\alpha = 1$. Then S_n satisfies (1.8) and (1.9) and is a $\sqrt{n}$-consistent estimator of

$$S(F) = \int_{\alpha_0}^{1-\alpha_0} \varphi(\alpha) F^{-1}(\alpha)\, d\alpha.$$

In the location model, S_n reduces to the Jaeckel [9] measure of dispersion of ranks of residuals.

REFERENCES

[1] P. J. Bickel and E. L. Lehmann, *Descriptive statistics for nonparametric models*, IV. *Spread*, in: *Contributions to Statistics, Jaroslav Hájek Memorial Volume*, J. Jurečková (Ed.), Reidel, Dordrecht 1979, pp. 33–40.

[2] Y. Dodge and J. Jurečková, *Estimation of quantile density function based on regression quantiles.* Statist. Probab. Lett. 23 (1993), pp 73–78.
[3] M. Falk, *On the estimation of the quantile density function,* ibidem 4 (1986), pp. 69–73.
[4] C. Gutenbrunner and J. Jurečková, *Regression rank scores and regression quantiles,* Ann. Statist. 20 (1992), pp. 305–330.
[5] J. Hájek and Z. Šidák, *Theory of Rank Tests,* Academic Press, New York 1967.
[6] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics,* Wiley, New York 1986.
[7] P. J. Huber, *Robust regression: Asymptotics, conjectures and Monte Carlo,* Ann. Statist. 1 (1973), pp. 799–821.
[8] — *Robust Statistics,* Wiley, New York 1981.
[9] L. A. Jaeckel, *Estimating the regression coefficients by minimizing the dispersion of residuals,* Ann. Math. Statist. 43 (1972), pp. 1449–1458.
[10] J. Jurečková and P. K. Sen, *Uniform second order asymptotic linearity of M-statistics in linear models,* Statist. Decisions 7 (1989), pp. 263–276.
[11] — *Regression rank scores scale statistics and studentization in linear models,* Proc. 5th Prague Conf. on Asympt. Statist., M. Hušková and P. Mandl (Eds.), Physica Verlag, Wien 1994, pp. 111–121.
[12] — *Robust Statistical Inference: Asymptotics and Interrelations,* Wiley, 1995 (to appear).
[13] J. Jurečková and A. H. Welsh, *Asymptotic relations between L- and M-estimators in the linear model,* Ann. Inst. Statist. Math. 42 (1990), pp. 671–698.
[14] R. Koenker and G. Bassett, *Regression quantiles,* Econometrica 46 (1978), pp. 33–50.
[15] R. A. Maronna, O. H. Bustos and V. J. Yohai, *Bias- and efficiency robustness of general M-estimators for regression with random carriers,* in: *Smoothing Techniques for Curve Estimation,* Lecture Notes in Math. 757, Springer, Berlin 1981, pp. 91–116.
[16] R. D. Maronna and V. J. Yohai, *Asymptotic behavior of general M-estimates for regression and scale with random carriers,* Z. Wahrsch. verw. Gebiete 58 (1981), pp. 7–20.
[17] D. Ruppert and R. J. Carroll, *Trimmed least squares estimation in the linear model,* J. Amer. Statist. Assoc. 75 (1980), pp. 828–838.
[18] A. H. Welsh, *Bahadur representation for robust scale estimators based on regression residuals,* Ann. Statist. 14 (1986), pp. 1246–1251.

Charles University
Department of Probability and Statistics
Sokolovská 83, 18600 Prague 8
Czech Republic

Received on 14.12.1993