

ON ADAPTIVE SCALE-EQUIVARIANT M-ESTIMATORS IN LINEAR MODELS

by

Jana Jurečková<sup>v</sup>

Department of Statistics and Probability  
Charles University, Prague, Czechoslovakia

and

Pranab Kumar Sen

Department of Biostatistics  
University of North Carolina, Chapel Hill, NC 27514

Institute of Statistics Mimeo Series No. 1422

December 1982

# On Adaptive Scale-Equivariant M-Estimators in Linear Models

JANA JUREČKOVÁ and PRANAB KUMAR SEN\*

Based on the concept of regression quantiles (due to Koenker and Bassett ), an adaptive, scale-equivariant version of M-estimators in linear models is considered. Various properties of the proposed estimator are studied.

KEY WORDS: Adaptive estimator; linear models; M-estimators; regression quantiles; scale-equivariance.

## 1. INTRODUCTION

Let  $X_1, \dots, X_n$  be  $n$  independent random variables (r.v.) with distribution functions (d.f.)  $F_1, \dots, F_n$ , respectively, all defined on the real line  $R$ . Consider the simple linear model where

$$F_i(x) = F(x - \beta' c_i) \quad , \quad x \in R \quad , \quad i=1, \dots, n, \quad (1.1)$$

$\beta' = (\beta_1, \dots, \beta_p)$  is an unknown vector of  $p (\geq 1)$  parameters, the  $c_i$  are known vectors of regression constants, and the d.f.  $F$  is assumed to be absolutely continuous and symmetric about 0. Let  $M_n = (M_{n1}, \dots, M_{np})'$  be the Huber M-estimator of  $\beta$ , i.e., it is a solution of

$$\sum_{i=1}^n c_i \psi(X_i - \beta' c_i) = 0 \quad , \quad (1.2)$$

where the *score-function*  $\psi$  is defined by

$$\psi(x) = \begin{cases} x, & |x| \leq k, \\ k \operatorname{sgn} x, & |x| > k, \end{cases} \quad (1.3)$$

for some specified value of  $k (> 0)$ . We write

$$X_n = (X_1, \dots, X_n)' \quad \text{and} \quad M_n = M_n(X_n) \quad . \quad (1.4)$$

$M_n$  is known to be a robust estimator of  $\beta$  having other desirable properties

---

\* Jana Jurečková is a Professor, Department of Statistics and Probability, Charles University, Prague, Czechoslovakia, and Pranab Kumar Sen is Cary C. Boshamer Professor, Department of Biostatistics, University of North Carolina, Chapel Hill. The second author's research was supported by the National Institutes of Health, Contract NIH-NHLBI-71-2243-L.

too. However, it has one discouraging property, namely, unlike the R- or L-estimators,  $M_n$  is not scale-equivariant i.e.,  $M_n(dX_n)$  is generally not equal to  $d.M_n(X_n)$ , for all  $d > 0$ . This is particularly due to the fact that  $\psi(dx) \neq d\psi(x)$  for all  $x$ , whenever  $k < \infty$ . On the other hand, the choice of  $k$  in (1.3), linked to the minimax property of  $M_n$  for the error-contamination model, may generally depend on the scale parameter of  $F$  or on the amount of contamination through the percentile points of  $F$ , which may not be precisely known. For this reason, Huber (1973,1981) suggested the use of a studentized version, viz.,

$$\sum_{i=1}^n c_i \psi((X_i - t'c_i)/s_n) , \quad (1.5)$$

where  $s_n$  is some suitable estimator of the scale parameter. We consider here an adaptive (scale-equivariant) estimator which does not require the location-scale model for the d.f.  $F$  and is somewhat related to the studentized version. The proposed estimator is based on the concept of regression quantiles due to Koenker and Bassett (1978) and the same is introduced in Section 2. The main results on the properties of the proposed estimator are presented in Section 3. The concluding section deals with a general class of M-estimators and their corresponding adaptive versions.

## 2. AN ADAPTIVE M-ESTIMATOR

Following Koenker and Bassett (1978), we define for every  $\alpha: 0 < \alpha < 1$ ,

$$\phi_\alpha(x) = \alpha - I(x \leq 0) , \quad x \in R , \quad (2.1)$$

$$\rho_\alpha(x) = x\phi_\alpha(x) , \quad x \in R. \quad (2.2)$$

The  $\alpha$ -th regression quantile  $\hat{\beta}_{(\alpha)}$  is then defined as any solution ( $t$ ) for which

$$\sum_{i=1}^n \rho_\alpha(X_i - t'c_i) \text{ is a minimum.} \quad (2.3)$$

Consider the residuals

$$\hat{X}_{i(\alpha)} = X_i - \hat{\beta}_{(\alpha)}'c_i , \quad i=1, \dots, n \quad (2.4)$$

and, for every  $\alpha: 0 < \alpha < 1/2$ , define  $A_{n(\alpha)} = \text{Diag}(a_{n1(\alpha)}, \dots, a_{nn(\alpha)})$ , by letting

$$a_{ni(\alpha)} = \begin{cases} 1, & \text{if } \hat{X}_{i(1-\alpha)} < 0 < \hat{X}_{i(\alpha)} \\ 0, & \text{otherwise, for } i=1, \dots, n. \end{cases} \quad (2.5)$$

Also, let  $\underline{c}_n = (c_1, \dots, c_n)'$ , and let

$$\underline{L}_n(\alpha) = (\underline{C}'_n \underline{A}_n(\alpha) \underline{C}_n)^{-1} (\underline{C}'_n \underline{A}_n(\alpha) \underline{X}_n) . \quad (2.6)$$

Thus,  $\underline{L}_n(\alpha)$  is the ordinary least squares estimator of  $\underline{\beta}$  based on the  $n-2$  [an] observations for which  $a_{ni}(\alpha) = 1$ , and is therefore termed the trimmed least squares estimator of  $\underline{\beta}$ . Having obtained this scale-equivariant estimator  $\underline{L}_n(\alpha)$  of  $\underline{\beta}$ , we consider the residuals

$$\hat{e}_i = X_i - \underline{L}'_n(\alpha) \underline{c}_i, \quad i=1, \dots, n, \quad (2.7)$$

and the corresponding empirical d.f.

$$\hat{F}_n(x) = n^{-1} \sum_{i=1}^n I(\hat{e}_i \leq x), \quad x \in R. \quad (2.8)$$

Further, let

$$\hat{F}_n^{-1}(p) = \begin{cases} \sup\{x : \hat{F}_n(x) \leq p\}, & 0 < p < \frac{1}{2}, \\ \inf\{x : \hat{F}_n(x) \geq p\}, & \frac{1}{2} < p < 1, \end{cases} \quad (2.9)$$

with an arbitrary convention for  $p = \frac{1}{2}$ . Let then,

$$k_n = (\hat{F}_n^{-1}(1 - \alpha) - \hat{F}_n^{-1}(\alpha)) / 2, \quad \alpha \in (0, \frac{1}{2}). \quad (2.10)$$

Finally, keeping (1.3) in mind, we define

$$\hat{\psi}_n(x) = \begin{cases} x, & |x| \leq k_n \\ k_n \operatorname{sgn} x, & |x| > k_n, \end{cases} \quad (2.11)$$

where  $k_n$  is defined by (2.10). Then, the adaptive estimator  $M_n^*$  of  $\underline{\beta}$  is defined as the solution of

$$\sum_{i=1}^n \underline{c}_i \hat{\psi}_n(X_i - \underline{t}' \underline{c}_i) = \underline{0}. \quad (2.12)$$

In the back of our mind, in (1.3), we let  $k = F^{-1}(1-\alpha)$ , so that by the assumed symmetry of  $F$ ,  $F^{-1}(\alpha) = -k$ . The relationship between  $k_n$  and  $k$  provides the basic key to our subsequent analysis.

### 3. ASYMPTOTIC PROPERTIES OF $M_n^*$

The estimator  $\underline{L}_n(\alpha)$  has been studied in further detail by Ruppert and Carroll (1980), and its relation with  $M_n$  by Jurečková (1983a,b). We intend to employ these results along with the Bahadur-representation type results for weighted empirical d.f. (due to Ghosh and Sen (1972)) to study the asymptotic properties of  $M_n^*$  and its relationship with  $M_n$ . Note that by the results of

Ruppert and Carroll (1980), under appropriate regularity conditions, including

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n c_i c_i' = Q \text{ (positive definite) } \quad (3.1)$$

and

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \{ n^{-1/2} \|c_i\| \} = 0, \quad (3.2)$$

for every  $\alpha : 0 < \alpha < 1/2$ ,

$$n^{1/2} ( \underline{L}_n(\alpha) - \underline{\beta} ) \rightsquigarrow \mathcal{N}_p( 0, Q^{-1} \sigma^2(\alpha, F) ), \quad (3.3)$$

where  $\sigma^2(\alpha, F)$  is the asymptotic variance of the  $\alpha$ -trimmed mean, and is given by

$$(1-2\alpha)^{-1} \{ \int_{-\xi_\alpha}^{\xi_\alpha} x^2 dF(x) + 2\alpha \xi_\alpha^2 \}, \text{ where } F(\xi_\alpha) = 1 - \alpha. \quad (3.4)$$

Further,

$$n^{1/2} \| \underline{L}_n(\alpha) - \underline{M}_n(\alpha) \| \xrightarrow{p} 0, \text{ as } n \rightarrow \infty, \quad (3.5)$$

where  $\underline{M}_n(\alpha)$  is defined as in (1.1)-(1.4) with  $k = k_\alpha = F^{-1}(1-\alpha)$ . Actually, for

$p=2$ ,  $\underline{c}_i = (1, c_i)'$ ,  $i=1, \dots, n$ , (3.5) has been improved by Jurečková (1983b) to

$$n^{1/2} \| \underline{M}_n(\alpha) - \underline{L}_n(\alpha) \| = O_p(n^{-1/4}), \text{ as } n \rightarrow \infty. \quad (3.6)$$

It may be noted that (3.6) holds in the general case of  $p \geq 2$ , and, further, under appropriate regularity conditions, we shall show that

$$n^{1/2} \| \underline{M}_n^*(\alpha) - \underline{M}_n(\alpha) \| = O_p(n^{-1/4}), \text{ as } n \rightarrow \infty, \quad (3.7)$$

where  $\underline{M}_n^*(\alpha)$  is defined by (2.10)-(2.12). This would establish the equivalence

of all the three estimates  $\underline{M}_n(\alpha)$ ,  $\underline{L}_n(\alpha)$  and  $\underline{M}_n^*(\alpha)$  (of  $\underline{\beta}$ ) upto the order  $O_p(n^{-3/4})$ , so that they all share the common properties of the Huber estimator

$\underline{M}_n(\alpha)$ . In addition,  $\underline{M}_n(\alpha)$  is not scale-equivariant but  $\underline{M}_n^*(\alpha)$  is so. Hence,

the proposed adaptive estimator combines the desirable properties of the Huber estimator and, at the same time, is scale-equivariant.

We define  $e_i = X_i - \underline{\beta}' \underline{c}_i$ ,  $i=1, \dots, n$ , so that these are i.i.d.r.v. with the d.f.  $F$ . Let then  $K^* = [-k, k]^p$ , and, for every  $\underline{t} \in K^*$ , let

$$\underline{Z}_n(\alpha)(\underline{t}) = n^{-1/2} \sum_{i=1}^n \underline{c}_i \psi_\alpha(e_i - n^{-1/2} \underline{t}' \underline{c}_i), \quad (3.8)$$

where  $\psi_\alpha$  is defined by (1.3) with  $k = k_\alpha$ . Also, let

$$\underline{S}_n^*(x; \underline{t}) = n^{-1/2} \sum_{i=1}^n \underline{c}_i I(e_i \leq x + n^{-1/2} \underline{t}' \underline{c}_i), \quad x \in R, \underline{t} \in K^*. \quad (3.9)$$

Let then

$$\omega_n(x, \underline{t}) = \left\| S_n^*(x; \underline{t}) - S_n^*(x; 0) - f(x)Q_n \underline{t} \right\|, \quad (3.10)$$

where  $Q_n = n^{-1} \sum_{i=1}^n C_i$  ( $\rightarrow Q$ , by (3.1)) and  $f=F'$ . Also, we assume that

$$\bar{C}_n = n^{-1} \sum_{i=1}^n C_i \rightarrow \bar{C} \text{ as } n \rightarrow \infty, \text{ where } \|\bar{C}\| < \infty. \quad (3.11)$$

Then, we have the following

*Lemma 1. Under (3.1), (3.2), (3.11) and the assumed absolute continuity of the density function  $f$ , over  $[-k, k]$ , as  $n \rightarrow \infty$ ,*

$$\sup\{ \omega_n(x, \underline{t}) : \underline{t} \in K^*, -k \leq x \leq k \} = O_p(n^{-1/4}). \quad (3.12)$$

The proof runs parallel to Theorem 3.1 of Ghosh and Sen(1972), and hence, is omitted. Actually, Ghosh and Sen considered an almost sure order, and hence, needed more stringent regularity conditions. In view of Theorem 3.1 of Springer and Sen (1983), the Bahadur-type representation holds even without the concordance-discordance condition of Jurečková (1977), so that (3.12) also holds without such extra conditions.

Note that by (3.8) and (3.9),

$$Z_{n(\alpha)}(\underline{t}) = \int_{-\infty}^{\infty} \psi_{\alpha}(x) dS_n^*(x; \underline{t}), \quad \underline{t} \in K^*, \quad (3.13)$$

while, by the classical central limit theorem (for independent summands),

$$Z_{n(\alpha)}(0) \sim \mathcal{N}(0, Q \int_{-\infty}^{\infty} \psi_{\alpha}^2(x) dF(x)), \quad (3.14)$$

and hence,

$$\|Z_{n(\alpha)}(0)\| = O_p(1). \quad (3.15)$$

Further, by (3.10), (3.12) and (3.13),

$$\begin{aligned} & \sup\{ \|Z_{n(\alpha)}(\underline{t}) - Z_{n(\alpha)}(0) + Q_n \underline{t}(1-2\alpha) \| : \underline{t} \in K^* \} \\ &= \sup\{ \|Z_{n(\alpha)}(\underline{t}) - Z_{n(\alpha)}(0) + Q_n \underline{t} \cdot \{ \int_{-\infty}^{\infty} f(x) d\psi_{\alpha}(x) \} \| : \underline{t} \in K^* \} \\ &= \sup\{ \| \int_{-k_{\alpha}}^{k_{\alpha}} \{ S_n^*(x; 0) - S_n^*(x; \underline{t}) + f(x)Q_n \underline{t} \} d\psi_{\alpha}(x) \| : \underline{t} \in K^* \} \\ &\leq \sup\{ \omega_n(x; \underline{t}) : \underline{t} \in K^*, -k_{\alpha} \leq x \leq k_{\alpha} \} = O_p(n^{-1/4}). \end{aligned} \quad (3.16)$$

Therefore, on letting  $\underline{t} = n^{1/2}(M_{n(\alpha)} - \beta)$ , we obtain from (1.2), (3.15) and (3.16),

$$n^{1/2}(M_{n(\alpha)} - \beta) = (1-2\alpha)^{-1} Q_n^{-1} Z_{n(\alpha)}(0) + O_p(n^{-1/4}). \quad (3.17)$$

If we define  $F_n(x) = n^{-1} \sum_{i=1}^n I(e_i \leq x)$  and let

$$k_n^0 = ( F_n^{-1}(1-\alpha) - F_n^{-1}(\alpha) )/2 , \quad (3.18)$$

where  $F_n^{-1}$  is defined as in (2.9), then, noting that  $k_\alpha = F^{-1}(1-\alpha)$ , we have by the classical results on sample quantiles that whenever  $f(k_\alpha) > 0$ ,

$$n^{1/2} | k_n^0 - k_\alpha | = O_p(1) . \quad (3.19)$$

Also, if we let

$$S_n^0(x; \tilde{t}) = n^{-1} \sum_{i=1}^n I(e_i \leq x + n^{-1/2} \tilde{t}' c_i) , \quad x \in R, \tilde{t} \in K^* , \quad (3.20)$$

then,  $S_n^0(x; 0) = F_n(x)$ , and parallel to (3.12), we have

$$\sup\{ n^{1/2} | S_n^0(x; \tilde{t}) - S_n^0(x; 0) - f(x) \tilde{c}_n^{-1} \tilde{t} | : x \in [-k, k], \tilde{t} \in K^* \} = O_p(n^{-1/4}) . \quad (3.21)$$

Using (2.7) - (2.10), (3.3) and (3.21), we conclude that

$$n^{1/2} | k_n - k_n^0 | = O_p(1) . \quad (3.22)$$

Thus, by (3.19) and (3.22),

$$n^{1/2} | k_n - k_\alpha | = O_p(1) . \quad (3.23)$$

We may virtually repeat the steps in (3.16) and verify that for every  $\epsilon \in (0, 1/2)$ ,

$$\sup_{\epsilon < \alpha < 1/2} \sup_{\tilde{t} \in K^*} || Z_{\tilde{n}(\alpha)}(\tilde{t}) - Z_{\tilde{n}(\alpha)}(0) + Q_{\tilde{n}}^{-1} \tilde{t} (1-2\alpha) || = O_p(n^{-1/4}) . \quad (3.24)$$

Further, if we let for  $u \in [-k, k]$ ,

$$\begin{aligned} W_n(u) &= n^{1/2} \{ Z_{\tilde{n}(\alpha)}(0) - Z_{\tilde{n}(\alpha+n^{-1/2}u)}(0) \} \\ &= \sum_{i=1}^n c_i \{ \psi_\alpha(e_i) - \psi_{\alpha+n^{-1/2}u}(e_i) \} , \end{aligned} \quad (3.25)$$

then, by routine computations, we have, for every  $u$ ,

$$E W_n(u) = 0 \quad \text{and} \quad E (W_n(u))' Q_n^{-1} (W_n(u)) = u^2 \{ f(k_\alpha) \}^{-2} + O(n^{-1/2} |u|^3) , \quad (3.26)$$

and a similar result holds for  $W_n(u) - W_n(u')$ . Thus, visualizing  $W_n$  as a (multivariate) process on  $C[-k, k]$  and using the Kolmogorov existence theorem, we conclude that for every  $k$  ( $0 < k < \infty$ ),

$$\sup\{ || W_n(u) || : |u| \leq k \} = O_p(1) . \quad (3.27)$$

By (3.24), (3.27) and (1.2), we have the following : for every  $k : 0 < k < \infty$ ,

$$\begin{aligned} \eta_n(\alpha, k) &= \sup\{ || n^{1/2} (M_{\tilde{n}(\alpha+n^{-1/2}u)} - \beta) - \\ &\quad (1-2(\alpha+n^{-1/2}u))^{-1} Q_n^{-1} Z_{\tilde{n}(\alpha+n^{-1/2}u)}(0) || : |u| \leq k \} = O_p(n^{-1/4}) . \end{aligned} \quad (3.28)$$

As such, we have

$$\sup\{ || n^{1/2} (M_{\tilde{n}(\alpha+n^{-1/2}u)} - \beta) - (1-2\alpha)^{-1} Q_n^{-1} Z_{\tilde{n}(\alpha)}(0) || : |u| \leq k \}$$

$$\begin{aligned} &\leq \eta_n(\alpha, k) + 2n^{-\frac{1}{2}}(1-2\alpha)^{-1}(1-2\alpha-2n^{-\frac{1}{2}}k)^{-1} \cdot \sup\{ \|Q_n^{-1}Z_{n(\alpha+n^{-\frac{1}{2}}u)}(0)\| : |u| \leq k \} \\ &\quad + n^{-\frac{1}{2}}(1-2\alpha)^{-1} \cdot \sup\{ \|Q_n^{-1}W_n(u)\| : |u| \leq k \} \\ &= O_p(n^{-\frac{1}{4}}) + O_p(n^{-\frac{1}{2}}) + O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{4}}). \end{aligned} \quad (3.29)$$

Therefore, by (2.11), (3.23) and (3.29), we conclude that

$$n^{\frac{1}{2}}(M_n^* - \underline{\beta}) = (1-2\alpha)^{-1}Q_n^{-1}Z_{n(\alpha)}(0) + O_p(n^{-\frac{1}{4}}), \quad (3.30)$$

so that (3.7) follows from (3.17) and (3.30).

#### 4. REMARKS ON GENERAL SCORE FUNCTIONS

In addition to the Huber M-estimator, we may also consider other M-estimators and obtain the corresponding adaptive versions based on the trimmed least squares estimators. For example, if we consider the score function

$$\psi_\alpha(x) = \begin{cases} F^{-1}(\alpha) - \alpha/f(F^{-1}(\alpha)), & x < F^{-1}(\alpha), \\ x, & F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha), \\ F^{-1}(1-\alpha) + \alpha/f(F^{-1}(1-\alpha)), & x > F^{-1}(1-\alpha), \end{cases} \quad (4.1)$$

then, for the location model, the corresponding M-estimator reduces to the Winsorized mean  $n^{-1}\{ [n\alpha]X_{n:[n\alpha]} + [n\alpha]X_{n:n-[n\alpha]} + \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{n:i} \}$ , where the  $X_{n:i}$  are the ordered r.v. corresponding to  $X_1, \dots, X_n$ . Estimate of the regression vector  $\underline{\beta}$  based on the score function in (4.1) can also be considered; its refined relation with the trimmed least squares estimate has also been studied by Jurečková (1983a). We may easily obtain the adaptive version of this M-estimator by replacing in (4.1),  $F^{-1}(\alpha) = -F^{-1}(1-\alpha)$  by  $-k_n$ , defined by (2.10). In such a case (3.7) holds under no extra regularity conditions. The adaptive estimator will be scale-equivariant, while, the original M-estimator (based on some assumed value of  $F^{-1}(\alpha)$ ) will not be so. In general, for an arbitrary score function having a constant value outside a compact interval, the proposed adaptive version will work out well, and a relation like (3.7) will follow under quite general conditions.



## REFERENCES

- Ghosh, M. and Sen, P.K. (1972). On bounded length confidence interval for the regression coefficient based on a class of rank statistics. Sankhya, Ser.A 34, 33-52.
- Huber, P.J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. Ann. Statist. 1, 799-821.
- Huber, P.J. (1981). Robust Statistics. New York : John Wiley.
- Jurečková, J. (1971). Asymptotic relations of M-estimates and R-estimates in linear regression model. Ann.Statist. 5, 464-472.
- Jurečková, J. (1983a). Winsorized least squares estimator and its M-estimator counterpart. In Contributions to Statistics : Essays in honour of Norman L. Johnson (ed. P.K. Sen), North -Holland, Amsterdam,pp.
- Jurečková, J. (1983b). Robust estimators of location and regression parameters and their second order asymptotic relations. To appear in the Proc. 9th Prague Conference on Statist. Dec. Theory and Inform.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. Econometrica 46, 33-50.
- Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. Jour. Amer. Statist. Asso. 75, 828-838.
- Springer, J.M. and Sen, P.K. (1983). On M-methods in multivariate linear models. (to appear).