[4] E. Sverdrup, *Multiple decision theory*, Aarhus Universitet, Matematisk Institutt, Lecture Notes Series No. 11 (1969).

Added in proofreading

Another construction of multiple test procedures often leading to tests equivalent to those suggested here is given in:

[5] R. Marcus, E. Peritz and K. R. Gabriel, *On closed testing procedures with special reference to ordered analysis of variance*, Biometrika 63 (1976), pp. 655–660.

Further development of sequentially rejective multiple test procedures can be found in:

[6] S. Holm, *A simple sequentially rejective multiple test procedure*, Scand. Journ. of Stat. 6 (1979), pp. 65–70.

[7] J. P. Shaffer, *Control of directional errors with stagewise multiple test procedures*, Ann. Stat., to appear.

[8] S. Holm, *A stagewise directional test for the normal regression situation*, Conference report from The sixth conference on probability theory, Brasov, Romania, 11–15 Sept. 1979, to appear.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

ROBUST ESTIMATION IN A LINEAR REGRESSION MODEL

JANA JUREČKOVÁ

Charles University, Prague, Czechoslovakia

1. Introduction

A detailed text concerning robust estimation in a linear model will appear in the monograph: K. M. S. Humak: *Statistische Methoden der Modellbildung*, Band II (Academic Verlag, Berlin). Here we shall only give a brief survey of the most usual types of robust estimates of the regression parameter vector and mention some of their asymptotic properties.

2. Robust alternatives to the method of least squares

We shall consider the problem of estimating the regression parameters of a linear model. We want to estimate β after observing $X_n' = (X_{1n}, \ldots, X_{nn})$ where

$$X_n = C_n\beta + E, \tag{2.1}$$

$\beta = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown regression parameters, $E = (E_1, \ldots, E_n)'$ is a vector of errors and $C_n = ((c_{ij}))_{i=1,\ldots,n}^{j=1,\ldots,p}$ is a matrix of known regression constants (design matrix) of the rank p. Most of our considerations will be asymptotic as the number of observations n grows and the number of regression parameters p remains fixed. Thus, the coordinates of X_n and of C_n depend on n; we shall not indicate explicitly this dependence provided no confusion arises.

We shall suppose throughout that E_i, $i = 1, \ldots, n$, are independent and identically distributed with a common distribution function F and density f with respect to the Lebesgue measure; F and f are generally unspecified.

If F is normal with the mean 0, the appropriate procedure is to minimize the sum of squares

$$\sum_{i=1}^{n}\Big(X_i - \sum_{j=1}^{p} c_{ij}\beta_j\Big)^2 = \min \tag{2.2}$$

or, equivalently, to solve the system of equations

$$\sum_{i=1}^{n}\Big(X_i - \sum_{k=1}^{p} c_{ik}\beta_k\Big)c_{ij} = 0, \quad j = 1, \ldots, p. \tag{2.3}$$

The least-squares estimate

$$\hat{\beta} = \Sigma_n^{-1} C_n' X_n \quad \text{where} \quad \Sigma_n = C_n' C_n \tag{2.4}$$

is admissible with respect to the quadratic loss if and only if F is normal (see Kagan–Linnik–Rao [17]).

For the location submodel ($p = 1$, $c_{ij} \equiv 1$) three different classes of estimation procedures alternative to (2.4) have been considered: *M-estimates* (estimates of the maximum likelihood type), *R-estimates* (based on ranks of observations) and *L-estimates* (linear combinations of order statistics). These procedures lead — in a more or less straightforward way — to extensions to a linear regression model.

We shall work with the residuals

$$\delta_i(\beta) = X_i - \sum_{j=1}^{p} c_{ij}\beta_j, \quad i = 1, \ldots, n. \tag{2.5}$$

The common idea of all these procedures is to replace function (2.2), to be minimized, by some other function less sensitive to the extreme values of the residuals (2.5).

2.1. *L-estimates*

In the location submodel, L-estimates are the linear combinations of order statistics. If $X^{(1)} < \ldots < X^{(n)}$ are the ordered observations, the estimates are of the form

$$\beta^{**} = \sum_{i=1}^{n} \lambda_i X^{(i)}. \tag{2.6}$$

If the coefficients λ_i are generated by a suitably chosen weight function J such that $\int_0^1 J(u)F^{-1}(u)\,du = 0$ so that $\lambda_i = n^{-1}J(i/(n+1))$, $i = 1, \ldots, n$, and various other regularity conditions are satisfied (see Bickel [2], Chernoff, Gastwirth, Johns [4], Shorack [21], Stigler [22], then $n^{1/2}(\beta^{**} - \beta)$ is asymptotically normal with the mean 0 and the variance

$$K_1(J, F) = \iint J(F(x))J(F(y))\,[F(\min(x, y)) - F(x)F(y)]\,dx\,dy. \tag{2.7}$$

If F is known, then

$$J(t) = \frac{d\varphi(t,f)}{dt} \cdot f(F^{-1}(t)) \left[\int_0^1 \frac{d\varphi(t,f)}{dt} f(F^{-1}(t))\,dt \right]^{-1}$$

where

$$\varphi(t,f) = -\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}, \quad 0 < t < 1,$$

yields an asymptotically efficient estimate, i.e. one which achieves the information inequality lower bound as $n \to \infty$ (Jung [14]).

Of particular interest from the point of view of robustness are the α-trimmed means corresponding to

$$J(t) = \begin{cases} \dfrac{1}{1-2\alpha} & \text{if} \quad \alpha \leqslant t \leqslant 1-\alpha, \quad 0 < \alpha < \dfrac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

The first extension of the L-procedures to the linear model is due to Bickel [3]. In the location model it coincides with the procedure defined in (2.6).

In the general case, this is not so straightforward: the procedure starts with a preliminary reasonably good estimate β^*. The resulting estimate is then equal to β^* plus an additive term depending on the ordered residuals $\delta_i(\beta^*)$ (see (2.5)). For instance, to get an analogue of the trimmed mean, all observations corresponding to residuals with the "position index" less than α or greater than $(1-\alpha)$ are trimmed off; the usual least-squares estimate is then determined from the remaining observations.

The estimates are, under the regularity conditions, asymptotically normal with the covariance matrix $K_1(J, F)\Sigma_n^{-1}$ with K_1 given in (2.7) and $\Sigma_n = C_n'C_n$. We see that the relative efficiencies of the estimates are independent of the design matrix C_n and thus the robustness results carry over from the location problem (we shall find the same for M and R-estimates).

2.2. *M-estimates*

We obtain M-estimates of regression parameters if we minimize, instead of (2.2),

$$\sum_{i=1}^{n} \varrho\Big(X_i - \sum_{j=1}^{p} c_{ij}\beta_j\Big) = \min, \tag{2.8}$$

where ϱ is some (usually convex) function. If we differentiate (2.8) we obtain (with $\psi = \varrho'$) the following analogue of (2.3):

$$\sum_{i=1}^{n} \psi\Big(X_i - \sum_{k=1}^{p} c_{ik}\beta_k\Big) c_{ij} = 0, \quad j = 1, \ldots, p, \tag{2.9}$$

which is equivalent to (2.8) if ϱ is convex.

The class (M) has been originated by Huber ([10], [11]) for the location model and extended by Relles [20] and Huber [12] to the regression model.

If f is smooth and $\psi = -f'/f$, then the M-estimate coincides with the maximum likelihood estimate. Moreover, if f is normal with the mean 0, we obtain the least squares estimate (2.4).

Under various regularity conditions, the above authors proved that the M-estimate is asymptotically normal (as $n \to \infty$ and p is fixed) with the mean β and the covariance matrix $K_2(\psi, F) \cdot \Sigma_n^{-1}$ where

$$K_2(\psi, F) = \int \psi^2(x)\,dF(x) \cdot \Big[\int \psi'(x)\,dF(x)\Big]^{-2} \tag{2.10}$$

2.3. *R-estimates*

Hodges and Lehmann [8] suggested estimates of location based on Wilcoxon and other rank tests; they showed that their asymptotic variances could be computed from the power functions of the tests, and that the estimates never have much lower but sometimes infinitely higher efficiencies than the sample mean.

Adidie (1967), following the ideas of Hodges and Lehmann, defined the estimates of β_1 and β_2 in the regression model $X_i = \beta_1 + \beta_2 c_i + E_i$, $i = 1, \ldots, n$, based on the Wilcoxon test and found their asymptotic distribution. Jurečková [15], Koul [18] and Jaeckel [13] then extended the procedure to the p-parameters regression and to the general rank tests. The three respective estimates are asymptotically equivalent and thus have the same asymptotic distribution and efficiency.

Roughly speaking, we obtain R-estimates if we minimize, instead of (2.2),

$$\sum_{i=1}^{n} a_n(R_i)\,\delta_i(\beta) = \min, \tag{2.11}$$

with respect to $\beta = (\beta_1, \ldots, \beta_p)'$. Here R_i is the rank of $\delta_i(\beta)$ in $(\delta_1(\beta), \ldots, \delta_n(\beta))$ and $a_n(\cdot)$ is some monotone score function (for simplicity normed so that $\sum_{i=1}^{n} a_n(i) = 0$). If we differentiate (2.11), which is a piecewise linear convex function of β, we obtain the approximate equalities at the minimum:

$$\sum_{i=1}^{n} a_n(R_i)\,c_{ij} \approx 0, \quad j = 1, \ldots, p. \tag{2.12}$$

These approximate equations can in turn be reconverted into a minimum problem e.g.

$$\sum_{j=1}^{p} \Big|\sum_{i=1}^{n} a_n(R_i)\,c_{ij}\Big| = \min. \tag{2.13}$$

The variant (2.13) was investigated by Jurečková [15] and (2.11) by Jaeckel, who also proved the asymptotic equivalence of both.

The score function $a_n(i)$ is supposed to be generated by a non-constant, non-decreasing square-integrable function $\varphi(t)$, $0 < t < 1$, in the following way:

$$a_n(i) = \varphi\left(\frac{i}{n+1}\right), \quad i = 1, \ldots, n. \tag{2.14}$$

If f is known and smooth, then

$$\varphi(t) = \varphi(t, f) = -\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}, \quad 0 < t < 1, \tag{2.15}$$

yields an asymptotically efficient estimate.

Under some regularity conditions, the estimates are asymptotically normal with the mean β and the covariance matrix $K_3(\varphi, F)\cdot \Sigma_n^{-1}$, where

$$K_3(\varphi, F) = \Big[\int_0^1 \varphi^2(t)\,dt - \Big(\int_0^1 \varphi(t)\,dt\Big)^2\Big]\Big[\int_0^1 \varphi(t)\varphi(t,f)\,dt\Big]^{-2}. \tag{2.16}$$

Besides the solution of (2.11) of (2.13), the estimates allow one step versions: start with some reasonably good preliminary estimate, and then apply one step of Newton's method to the corresponding system of equations. Such an estimate was investigated by Kraft and van Eeden [23].

Finally, some adaptive estimation procedures are worth of mentioning. If the underlying distribution F is unknown, the procedure is accomplished by estimating the optimal score function $\varphi(t, f)$ (or ψ, or J) either from a part of or from all observations and by calculating the estimate generated by this function. For the location submodel, such estimates have been considered e.g. by Hájek–Šidák [6], Hájek [7], van Eeden [23], Beran [1]. Analogous estimates for the general linear model have been investigated by Dionne [5]. Despite the fact that such schemes have excellent asymptotic properties, the convergence is too slow and the sample size must be estremely large if the results are to be satisfactory. Hogg [9] mentioned that instead of the direct application of Hájek and van Eeden, a rougher approximation of φ may be useful.

We have seen from the above remarks that the three estimation procedures follow the same idea: to decrease the possible influence of outlying observations. Any one of them could lead to an asymptotically efficient estimate in the case where the basic distribution is known. In fact, as $n \to \infty$, the estimates are closely related to one another. For instance, suppose that the respective J, ψ and φ-functions are smooth and connected in the following way:

$$\begin{aligned} J(t) &= \varphi'(t) f(F^{-1}(t))\Big[\int_0^1 \varphi'(t) f(F^{-1}(t))\,dt\Big]^{-1}, \\ \psi(x) &= c\varphi(F(x)), \quad c > 0; \end{aligned} \tag{2.17}$$

then the corresponding L, M and R-estimates are asymptotically equivalent in probability.

References

[1] R. Beran, *Asymptotically efficient adaptive rank estimates in location models*, Ann. Statist. 2 (1974), pp. 63–74.

[2] P. V. Bickel, *Some contributions to the theory of order statistics*, Proc. 5th Berkeley Symp. 1 (1967), pp. 575–591.

[3] —, *On some analogues to linear combinations of order statistics in the linear model*, Ann. Statist. 1 (1973), pp. 597–616.

[4] H. Chernoff, J. Gastwirth, M. V. Johns, *Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation*, Ann. Math. Statist. 38 (1967), pp. 52–72.

[5] L. Dionne, *Efficient nonparametric estimators of parameters in the general linear hypothesis*, to appear in Ann. Statist. (1976).

[6] J. Hájek, Z. Šidák, *Theory of Rank Tests*, Academia, Prague 1967.

[7] J. Hájek, *Miscellaneous problems of rank test theory*, in (M. L. Puri, ed.) *Nonparametric techniques in statistical inference*, Cambridge Univ. Press., Cambridge 1970.

[8] V. L. Hodges, E. L. Lehmann, *Estimates of location based on rank tests*, Ann. Math. Statist. 34 (1963), pp. 598–611.

[9] R. V. Hogg, *Adaptive robust procedures. A partial review and some suggestions for future. Applications and theory*, U. Amer. Statist. Assoc. 69 (1974), pp. 909–923.
[10] P. J. Huber, *Robust estimation of a location parameter*, Ann. Math. Statist. 25 (1964), pp. 73–101.
[11] P. J. Huber, *The behavior of maximum likelihood estimates under nonstandard conditions*, Proc. 5th Berkeley Symp. 1 (1967), pp. 121–133.
[12] P. J. Huber, *Robust regression: Asymptotics conjectures and Monte Carlo. The 1972 Wald Memorial Lecture*, Ann. Statist. 1 (1973), pp. 799–821.
[13] L. A. Jaeckel, *Estimating regression coefficients by minimizing the dispersion of the residuals*, ibid. 43 (1972), pp. 1449–1458.
[14] J. Jung, *On linear estimates defined by a continuous weight function*, Ark. Math. 3 (1955), pp. 199–209.
[15] J. Jurečková, *Nonparametric estimates of regresion coefficients*, Ann. Math. Statist. 42 (1971), pp. 1328–1338.
[16] —, *Robust statistical inference in linear models*, to appear in *Statische Methoden der Modellbildung*, Band II (Academie Verlag, Berlin) (1976).
[17] A. M. Kagan, Ju. V. Linnik, C. R. Rao, *On a characterization of the normal law, based on a property of the sample average*, Sankhya A 27 (1965), pp. 405–406.
[18] H. L. Koul, *Asymptotic behavior of a class of confidence regions based on ranks in regression*, Ann. Statist. 42 (1971), pp. 466–476.
[19] C. Kraft, C. van Eeden, *Linearized rank estimates and signed-rank estimates for the general linear hypothesis*, ibid. 43 (1972), pp. 42–57.
[20] D. Relles, *Robust regression by modified least squares*, PhD Thesis, New York 1968.
[21] Shorack, *Asymptotic normality of linear combinations of order statistics*, Ann. Math. Statist. 40 (1969), pp. 2041–2050.
[22] S. M. Stigler, *Linear functions of order statistics with smooth weight functions*, Ann. Statist. 2 (1974), pp. 676–693.
[23] C. van Eeden, *Efficiency — robust estimation of location*, Ann. Math. Statist. 41 (1970), pp. 172–181.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

A REMARK ON THE CONDITIONING IN LIMIT THEOREMS FOR DEPENDENT RANDOM VECTORS IN R^d

ANDRZEJ KŁOPOTOWSKI

Institute of Mathematics, Nicholas Copernicus University, Toruń, Poland

A main problem of the classical theory of probability concerns limit distributions for sums of infinitesimal systems of independent random variables. There exists a complete solution of this problem given by necessary and sufficient conditions for the convergence in law of such systems to arbitrarily fixed, infinitely divisible probability measure.

Now the following problem is still open. Let there be given an infinitely divisible probability measure Q on $\boldsymbol{R}^d$ with the characteristic function

$$\varphi_Q(\vec{t}) = \exp\left\{i(\vec{t},\vec{a})-\tfrac{1}{2}(\vec{t},A\vec{t})+\int_{R^d}\left(e^{i(\vec{t},\vec{x})}-1-\frac{i(\vec{t},\vec{x})}{1+\|\vec{x}\|^2}\right)\frac{1+\|\vec{x}\|^2}{\|\vec{x}\|^2}d\mu(\vec{x})\right\},\qquad \vec{t}\in\boldsymbol{R}^d,$$

where $\vec{a}\in\boldsymbol{R}^d$, A is nonnegative definite $d\times d$-matrix, μ is the finite measure on $\boldsymbol{R}^d$, $\mu(\{\vec{0}\}) = 0$. Describe double sequences of random vectors $\{\{\vec{X}_{nk}\}_{1\leqslant k\leqslant k_n}\}_{n\in N}$ which converge in law to Q, i.e. the distributions of sums $S_n = \sum_{k=1}^{k_n}\vec{X}_{nk}$, $n\in N$, are weakly convergent to Q. First, for row-wise independent systems their infinitesimality can be replaced by more general condition

$$\lim_{n\to\infty}\sum_{k=1}^{k_n}|E\,e^{i(\vec{t},\vec{X}_{nk}-\vec{a}_{nk})}-1|^2 = 0,\qquad \vec{t}\in\boldsymbol{R}^d,$$

where $\vec{a}_{nk} := E(\vec{X}_{nk}I(\|\vec{X}_{nk}\| < \varepsilon))$, $\varepsilon > 0$, and the above conditions remain necessary and sufficient.

Next, let us admit a dependence between vectors in the same row. One way to find sufficient conditions for the convergence in law to Q which generalise the classical case is the following: we replace all mean values in known necessary and sufficient conditions by conditional mean values with respect to suitably chosen σ-fields and the ordinary convergence we replace by the convergence in probability of such obtained random vectors. Now we give an example of such result.