NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

SOME ADAPTIVE PROCEDURES FOR REGRESSION MODELS

M. Huskova

June 1985
CP-85-30

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

Within the framework of the Economic Structural Change Program, a cooperative research activity of IIASA and the University of Bonn, FRG, a project is carried out on "Statistical and Econometric Identification of Structural Change"; the project involves studies on the formal aspects of the analysis of structural changes. On the one hand, they include statistical methods to detect non-constancies, such as stability tests, detection criteria, etc., and on the other hand, methods which are suitable for models which incorporate non-constancy of the parameters, such as estimation techniques for time-varying parameters, adaptive methods, etc.

The present paper surveys adaptive methods for regression analysis, i.e., methods which are dependent on the data to be analyzed. A final chapter sketches some problems related to the use of adaptive regression methods in the context of structural changes, such as the investigation of properties of an adaptive version of Quandt's switching regression procedure.

Anatoli Smyshlyaev
Acting Leader
Economic Structural Change Program

SOME ADAPTIVE PROCEDURES FOR REGRESSION MODELS

Marie Huskova

Charles University, Sokolovska 83, 186 00 Prague 8, Czechoslovakia

INTRODUCTION

Regression models belong to those statistical models, which are applied to extremely diverse types of data in many fields of quantitative relationships. Normally distributed errors are usually assumed and least squares estimates are applied. It is known that for normally distributed errors the least squares estimates are optimal in several respects, while for nonnormally distributed errors these estimates are ineffective and, moreover, they are sensitive to outlying observations.

Classes of estimators were developed which show a reasonable behavior for comparatively large families of error distributions and which are not too sensitive to the outliers. Such estimators are usually called *robust*. Some of these estimators can be adapted with respect to the data in such a way that the resulting estimates are in some sense optimal; these estimators are called *adaptive*.

The aim of this paper is to present some adaptive estimates for regression models.

Consider the linear model:

$$Y_{ni} = \alpha + \sum_{j=1}^{p} c_{n,ij}\theta_j + X_{ni} \quad , \quad 1 \le i \le n \quad , \tag{1}$$

or equivalently,

$$\underset{\sim}{Y}_n = \alpha(1,\ldots,1)' + \underset{\sim}{C}_n\underset{\sim}{\theta} + \underset{\sim}{X}_n \quad , \tag{2}$$

where $\underset{\sim}{Y}_n = (Y_{n1},\ldots,Y_{nn})'$ is a vector of observations, $\underset{\sim}{X}_n = (X_{n1},\ldots,X_{nn})'$ is a vector of independent identically distributed random errors, α and $\underset{\sim}{\theta} = (\theta,\ldots,\theta)'$ are unknown parameters, and $\underset{\sim}{C}_n = (c_{n,ij})_{\substack{i = 1,\ldots,n \\ j = 1,\ldots,p}}$ is a design matrix nxp of full rank (=p). Moreover, it is assumed that X_{ni} has a distribution function F and a density f (with respect to the Lebesgue measure) belonging to some class $\mathcal{F}$ of densities.

The problem is that of estimating $\underset{\sim}{\theta}$.

If F is normal with mean zero, then the least squares estimate $\underset{\sim}{\theta}_n = (\theta_{n1},\ldots,\theta_{np})'$ is optimal. More precisely, it is unbiased

$$E\, \underset{\sim}{\theta}_n = \underset{\sim}{\theta} \tag{3}$$

and has minimal variance

$$\mathrm{var}\{\sum_{i=1}^{p} u_i\theta_{ni}\} \le \mathrm{var}\{\sum_{i=1}^{p} u_i\theta^*_{ni}\} \tag{4}$$

$$\text{for all } u_1,\ldots,u_p \quad ,$$

where $\underset{\sim}{\theta}^*_n = (\theta_{n1},\ldots,\theta_{np})'$ is an arbitrary unbiased estimate of $\underset{\sim}{\theta}$. Recall the definition of the least squares estimate:

$$\underset{\sim}{\theta}_n = \arg\min_{\underset{\sim}{\theta}} \sum_{i=1}^{n} \delta_i^2(\underset{\sim}{\theta}) \quad , \tag{5}$$

where $\delta_i(\underset{\sim}{\theta}) = Y_{ni} - \sum_{j=1}^{p} (c_{n,ij}-\bar{c}_{n,j})\theta_j$, sometimes called residuals, and $\bar{c}_{n,j} = n^{-1}\sum_{i=1}^{n} c_{n,ij}$; then the variance matrix can be rewritten as follows:

$$\operatorname{var} \underset{\sim}{\theta}_n = \operatorname{var} X_{n1} (\underset{\sim}{C}_n^* \underset{\sim}{C}_n^*)^{-1} \quad , \tag{6}$$

$$\underset{\sim}{C}_n^* = (c_{n,ij}-\bar{c}_{n,j})_{\substack{i = 1,\dots,n \\ j = 1,\dots,p}} \quad .$$

If the error distribution function F is nonnormal, the least squares estimate $\underset{\sim}{\theta}_n$ is in most cases not even reasonable (see, e.g., Huber 1972). Since the true underlying distribution is seldom exactly known, it is sensible to use procedures which work well for a variety of possible situations. Such procedures are called robust. More information on this issue can be found, e.g., in Huber (1981), Jureckova (1985).

The typical robust estimates are M- and R-estimates. The *M-estimate* (estimate of the maximum likelihood type) $\underset{\sim}{\theta}_M(\Psi)$ is defined as follows:

$$\underset{\sim}{\theta}_M(\Psi) = \arg\min_{\underset{\sim}{\theta}} \sum_{i=1}^{n} \rho(\delta_i(\underset{\sim}{\theta})) \tag{7}$$

or, equivalently, it is the solution of the system of equations

$$\sum_{i=1}^{n} \Psi(\delta_i(\underset{\sim}{\theta}))(c_{n,ij}-\bar{c}_{n,j}) = 0 \quad , \qquad j = 1,\dots,p \tag{8}$$

with respect to $\underset{\sim}{\theta}$, where ρ is a convex function and $\rho' = \Psi$. The choice of $\Psi(x) = x$ and $\Psi(x) = \Psi_f(x) = -f'(x)/f(x)$ leads to the least squares estimate and to the maximum likelihood estimate, respectively. The *R-estimate* (estimate based on ranks) can be defined in either of the following ways:

$$\tilde{\theta}_R(\psi) = \arg\min_{\tilde{\theta}} \sum_{i=1}^{n} \psi(R_{ni}(\tilde{\theta})(n+1)^{-1})\delta_i(\tilde{\theta}) \quad , \tag{9}$$

$$\tilde{\theta}_R^*(\psi) = \arg\min_{\tilde{\theta}} \sum_{j=1}^{p} \left| \sum_{i=1}^{n} \psi(R_{ni}(\tilde{\theta})(n+1)^{-1}) \cdot \right.$$

$$\left. \cdot (c_{n,ij} - \bar{c}_{n,j}) \right| \quad , \tag{10}$$

where ψ is a monotone function on (0,1) and $R_{ni}(\tilde{\theta})$ is the rank of $\delta_i(\tilde{\theta})$ among $\delta_1(\tilde{\theta}),\ldots,\delta_n(\tilde{\theta})$. Both estimates are asymptotically equivalent.

Both the M- and R-estimates allow a one-step version, i.e. to start with some reasonably good preliminary estimate and then to apply one step of the Newton method to the corresponding system of equations.

Generally, the M- and R-estimates are under very mild conditions asymptotically unbiased and consistent. If the error distribution F is known and some regularity conditions are fulfilled, the estimates $\tilde{\theta}_M(\Psi_f)$, $\tilde{\theta}_R(\psi_f)$, and $\tilde{\theta}_R^*(\psi_f)$ with $\Psi_f(x) = -f'(x)/f(x)$, $x \in R_1$, and $\psi_f(u) = -f'(F^{-1}(u))/f(F^{-1}(u))$, $u \in (0,1)$ (where F^{-1} is the quantile function corresponding to F) are asymptotically optimal, i.e. they are asymptotically unbiased and have asymptotically the smallest variance matrix. The latter property means that the asymptotic variance matrix should be closed to $(\tilde{C}_n^*\tilde{C}_n^*)^{-1}I(f)$, where $I(f) = \int (f'(x))^2/f(x)dx$ is the Fisher information.

If F is unknown and we are still interested in having an asymptotically optimal estimate, at least in some class of error distributions, we can either construct quite new estimates (which is a difficult problem and solved only in very special cases) or adapt the already known estimates with respect to the data. Attention was mainly paid to the latter case. To adapt M- and R-estimates means either to replace Ψ_f and ψ_f, respectively, by suitable estimates, or--assuming that the true density belongs to family F of error distributions--to choose a density $f_o \in F$ according to a decision rule that fits the data.

A simple form of such adaptive estimates was already intuitively used by many scholars in the field of applied statistics; e.g., with regard to the problem of estimating θ in the model $Y_{ni} = \theta + X_{ni}$, $i = 1,\ldots,n$, where X_{ni} has a symmetric distribution, they used either the arithmetic mean or the median, depending on the data to be analyzed.

In the next section some typical adaptive M- and R-estimates are introduced.

For more detailed information on adaptive procedures for various models and other statistical problems see review papers by Hogg (1974), Hogg and Lenth (1984), and Huskova (1985). General considerations on adaptive procedures can be found in the paper by Bickel (1982).

ADAPTIVE M- AND R-ESTIMATES

The basic steps in the procedure are the following:

a. Find a reasonable robust preliminary estimate $\bar{\theta}_n$ of θ.

b. Choose a reasonable family F of error distributions and a decision rule for selecting a density $f_o \in F$ as a possible true density or the type of estimate for $\Psi_f(\psi_f)$.

c. Using the residuals $\delta_1(\theta_n),\ldots,\delta_n(\theta_n)$, select $f_o \in F$ according to the decision rule or find an estimate $\hat{\Psi}_f(\hat{\psi}_f)$ of $\Psi_f(\psi_f)$.

d. Compute the one-step version of the M- (R-)estimate using the preliminary estimator $\bar{\theta}$ and replacing $\Psi_f(\psi_f)$ by either $\Psi_{f_o}(\psi_{f_o})$ or by its estimate $\hat{\Psi}_f(\hat{\psi}_f)$ from step c.

As preliminary estimates either M-estimates with $\Psi(x) = x\alpha$, $x \in R_1$, $1 \leq \alpha < 2$, or R-estimates with $\psi(u) = u$, $u \in (0,1)$ are recommended.

Moberg et al. (1980) proposed a decision rule based on the measure of skewness Q_3 and the measure of tailweight Q_4, where

$$Q_3 = \frac{\bar{U}(0.05)-\bar{M}(0.50)}{\bar{M}(0.50)-\bar{L}(0.05)} \quad , \quad Q_4 = \frac{\bar{U}(0.05)-\bar{L}(0.05)}{\bar{U}(0.50)-\bar{L}(0.50)} \qquad (11)$$

with $\bar{L}(\alpha)$, $\bar{M}(\alpha)$, $\bar{U}(\alpha)$ being the arithmetic means of the smallest, the medium, and the largest $[n\alpha]$ of the order statistics $Z_{(1)} \leq \dots \leq Z_{(n)}$ corresponding to the residuals $\delta_1(\bar{\theta}_n), \dots, \delta_n(\bar{\theta}_n)$.

Starting from the generalized λ-family of distribution (the quantile function can be expressed as $F^{-1}(p) = \lambda_1 + (p^{\lambda_2} - (1-p)^{\lambda_3})/\lambda_4$, $p \in (0,1)$, $\lambda_i \in R_1$, $i = 1,\dots 3$, $\lambda_4 > 0$), and using the Monte Carlo method, they proposed partitioning of distributions into five classes (light-tailed and symmetric (I), medium-tailed and symmetric (II), heavy-tailed and symmetric (III), light-tailed and skewed to the right (IV), moderate-tailed and skewed to the right (V)) according to Q_3 and Q_4. For each class they recommend a proper choice of the function Ψ.

Jones (1979) developed an adaptive procedure based on ranks and order statistics, originally for testing of symmetry. This can easily be modified to the estimation problem. The author assumes that the family $\mathcal{F}$ consists of densities f with ψ_f expressed as follows:

$$\psi_f(u) = (\lambda-1)(u^{\lambda-2} - (1-u)^{\lambda-2})(u^{\lambda-1} + (1-u)^{\lambda-1})^{-2} \quad ,$$

$$u \in (0,1) \quad , \quad \lambda \in R_1 \quad , \qquad (12)$$

which contains densities ranging from light-tailed ($\lambda > 0$) to heavy-tailed ($\lambda < 0$) densities. The estimate of λ was defined through the ordered sample $Z_{(1)},\dots,Z_{(n)}$ corresponding to $\delta_1(\bar{\theta}_n),\dots,\delta_n(\bar{\theta}_n)$, namely,

$$\hat{\lambda} = (\log 2)^{-1} \log (Z_{(n-2M+1)} - Z_{(n-4M+1)}) \cdot$$

$$\cdot (Z_{(n-M+1)} - Z_{(n-2M-1)})^{-1} , \tag{13}$$

where M is chosen in a proper way to reflect the behavior of the tail.

Koul and Susarla (1983) constructed the estimate

$$\hat{\Psi}_f(x) = - \frac{\hat{f}'(x;r_n)}{\max(\hat{f}(x;r_n), a_n)} , \quad x \in R_1 , \tag{14}$$

where $f(x;r_n)$ is the kernel estimate of the density f (with kernel $N(0,r_n^2)$) based on $\delta_1(\bar{\theta}_n),\dots,\delta_n(\bar{\theta}_n)$, $a_n \searrow 0$, $r_n \searrow 0$, and as a resulting estimate they propose a slightly modified one-step version of $\theta_M(\hat{\Psi}_f)$.

Huskova (1984) made use of the fact that for $\psi_f \in L_2(0,1)$ one can write

$$\psi_f(u) = \sum_{k=0}^{n} d_k P_k(u) , \quad u \in (0,1) , \tag{15}$$

where $\{P_k(u)\}_{k=o}^{\infty}$ is the system of Legendre's polynomials on (0,1) and

$$d_k = \|P_k\|^{-1} \int_o^1 P_k(u)\psi_f(u)du , \quad k = 0,1,2,\dots \tag{16}$$

$$\|P_k\|^2 = \int_o^1 P_k^2(u)du ,$$

and suggested the following estimator of ψ_f:

$$\hat{\psi}_f(u) = \sum_{k=o}^{M_n} \hat{d}_k P_k(u) , \quad u \in (0,1) \tag{17}$$

with $\hat{d}_k$ being an estimate of d_k obtained by means of the asymptotic linearity of rank statistics, $M_n \to \infty$ as $n \to \infty$.

The procedure proposed by Moberg et al. (1980) can be easily applied in practice; the Monte Carlo study supports this procedure, but from the asymptotical point of view it is not optimal. Several modifications of this procedure were developed.

The procedure of Jones (1979) is asymptotically optimal, if the true density belongs to the λ-family of distributions.

The last two remaining procedures lead to asymptotically optimal estimates, but due to computational problems their practical application is--in their present form--not very appealing.

ADAPTIVE PROCEDURES FOR DETECTING CHANGE

Consider the regression model:

$$Y_i = Y(t_i) = \alpha + I\{t_i \le \tau\} \cdot \sum_{j=1}^{p} \theta_j c_j(t_i) + $$

$$+ I\{t_i > \tau\} \cdot \sum_{j=1}^{p} \beta_j c_j(t_i) + X_i \quad , \qquad (18)$$

$$1 \le i \le n \quad ,$$

where $Y(t_i)$ is the observation taken at time $t_i, t_1 \le t_2 \le \ldots \le t_n$ (not all equal), $\alpha, \theta_1, \ldots, \theta_p, \beta_1, \ldots, \beta_p$ are unknown parameters, $\tau \in (t_1, t_n]$ is an unknown time point, $X_1, \ldots, X_n$ are independent random variables with a distribution function F, and $(c_j(t_i))_{\substack{i=1,\ldots,n \\ j=1,\ldots,p}}$ is a design matrix.

The problem is concerned with testing the constancy of the regression relationship over time, i.e., $H_o : \theta_j = \beta_j$, $1 \le j \le p$ against $H_1 : \theta_j \ne \beta_j$ for at least one j.

Sen (1980, 1982) proposed some test procedures based on rank statistics, or, more exactly, on the statistics

$$\sum_{i=1}^{k} (c_j(t_i)-\bar{c}_j)(R_{ki}(\bar{\theta}_n)(n+1)^{-1}) \quad , \quad k = 1,\ldots,n. \quad (19)$$

Sen (1983) developed a procedure for a more general testing problem: $Y(t_1),\ldots,Y(t_n)$ are independent random variables, $Y(t_i)$ has a distribution function F_i, $i = 1,\ldots,n$, and

$$H_o : F_1 = \ldots = F_n \quad \text{against}$$

$$H_1 : F_1 = \ldots = F_q \neq F_{q+1} = \ldots = F_n \quad ,$$

where q is unknown, $1 \leq q < n$. The test procedure is based on U-statistics, i.e.

$$\sum_{1 \leq i_1 < i_2 < \ldots < i_m \leq k} \sum \ldots \sum h(Y(t_{i_1}),\ldots,Y(t_{i_n})), \quad 1 \leq k \leq n \quad ,$$

where h is a symmetric function on R_m, m is fixed, $1 \leq m \leq n$. Both types of procedures mentioned belong to the robust procedures. Adaptive procedures were not yet developed.

The problems to be solved (first for a simple linear model and then for the general regression model) are:

1. The development of adaptive procedures combining already existing robust procedures (i.e. based on ranks) with the methods of adaptation and the investigation of their asymptotic properties.

2. The development of robust procedures based on M-estimates (modification of Quandt's log-likelihood ratio procedure) and the investigation of their asymptotic properties.

3. The development of adaptive procedures corresponding to the robust procedures of point 2, and again the investigation of their asymptotic properties.

4. The development of robust and adaptive procedures for a more general problem, namely, to admit in regression model (18) X_i with different distributions for $t_i \leq \tau$ and $t_i > \tau$.

5. The development of suitable algorithms for the procedures of points 1-4.

REFERENCES

Bickel, P. (1982). On adaptive estimation. Annals of Statistics 10:647-671.

Hogg, R.V. (1974). Adaptive robust procedures: partial review and some suggestions for future applications and theory. J. Amer. Statist. Assoc. 69:909-923.

Hogg, R.V., and R.V. Lenth (1984). A review of some adaptive statistical techniques. Commun. in Statist. A 13:1551-1579.

Huber, P.J. (1972). Robust statistics: a review. Ann. Math. Statist. 43:1041-1067.

Huber, P.J. (1981). Robust Statistics. New York: Wiley.

Huskova, M. (1984). Adaptive procedures for the two-sample location model.. Commun. in Statist. Sequential Analysis 2:387-401.

Huskova, M. (1985). Adaptive methods. Handbook of Statistics, P.R. Krishnaiah and P.K. Sen, eds., 4:347-358.

Jones, D.H. (1979). An efficient adaptive distribution-free test for location. J. Amer. Statist. Assoc. 74:822-828.

Jureckova, J. (1985). M-, L- and R-estimators. Handbook of Statistics, P.R. Krishnaiah and P.K. Sen, eds., 4:463-485.

Koul, H.L., and V. Susarla (1983). Adaptive estimation in linear regression. Statistics and Decision 1:379-400.

Moberg, T.F., J.S. Ramberg, and R.H. Randles (1980). An adaptive regression procedure based on M-estimators. Technometrics 22:213-224.

Sen, P.K. (1980). Asymptotic theory of some tests for a possible change in the regression slope occuring at an unknown time-point. Z. f. Wahrscheinlichkeitstheorie verw. Gebiete, 52:203-218.

Sen, P.K. (1982). Asymptotic theory of some tests for constancy of regression relationships over time. Math. Operationsforsch. Statist., Statistics, 13:21-31.

Sen, P.K. (1983). Tests for change-points based on recursive U-statistics. Commun. in Statist. Sequential Analysis, 1:263-284.