Acta Math. Univ. Comenianae
Vol. LXIV, 2(1995), pp. 227–234

A CHOICE OF CRITERION PARAMETERS IN A LINEARIZATION OF REGRESSION MODELS

A. JENČOVÁ

Introduction

There are many results which are obtained in the theory of nonlinear regression models; nevertheless much more and simpler inferences may be made in linear models. Thus it is of some importance to analyze situations where a nonlinear model can be substituted by a linear one. Some rules how to proceed in a linearization of regression models are given in [**K1**], [**K2**]. Several parameters (criterion parameters) have to be chosen in the mentioned procedures. In the following it is shown that some relations among these parameters and some natural restrictions on them exist.

1. Notations and Preliminaries

Let $Y \sim N_n[f(\beta), \Sigma]$, where Y is an n-dimensional normally distributed random vector with mean value $E_\beta[Y] = f(\beta)$ and a known positively definite covariance matrix $\mathrm{Var}\,[Y] = \Sigma$. Here $\beta \in R^k$ is an unknown k-dimensional parameter and $f(\cdot)\colon R^k \to R^n$ is a known function with continuous second derivatives. According to [**K1**], the model is investigated only on a neighbourhood $\mathcal{O}(\beta_0)$ (which will be specified in the following) of a chosen point $\beta_0 \in R^k$. It is assumed that:

(i) it is known that the true value $\bar\beta$ lies in $\mathcal{O}(\beta_0)$
(ii) the terms $\delta\beta_i\delta\beta_j\delta\beta_m \frac{\partial^3 f_l(\beta)}{\partial\beta_i\partial\beta_j\partial\beta_m}|_{\beta=\beta_0}$, $l = 1,\dots,n$ and $i,j,m = 1,\dots,k$ can be neglected if $\beta_0 + \delta\beta \in \mathcal{O}(\beta_0)$, so that f has the form

$$f(\beta) = f_0 + F\delta\beta + \frac{1}{2}\kappa_{\delta\beta}$$

for $\beta_0 + \delta\beta \in \mathcal{O}(\beta_0)$. Here $f_0 = f(\beta_0)$, $F = \frac{\partial f(\beta)}{\partial\beta'}|_{\beta=\beta_0}$ is a full rank matrix, $(\kappa_{\delta\beta})_i = \delta\beta' H_i \delta\beta$, $H_i = \frac{\partial^2 f_i(\beta)}{\partial\beta\partial\beta'}$, $i = 1,\dots,n$.

Received January 20, 1995.
1980 *Mathematics Subject Classification* (1991 *Revision*). Primary 62F10.

We will use the following notations:

$$C = F'\Sigma^{-1}F,$$

$$\Delta = \begin{pmatrix} \delta\beta' H_1 \\ \vdots \\ \delta\beta' H_n \end{pmatrix},$$

$$H_i^* = \begin{pmatrix} e_i' H_1 \\ \vdots \\ e_i' H_n \end{pmatrix}, \qquad i = 1, \dots, k,$$

$$K_1^{(h)} = \begin{pmatrix} h'C^{-1}\frac{1}{2}(H_1^*)'\Sigma^{-1} \\ \vdots \\ h'C^{-1}\frac{1}{2}(H_k^*)'\Sigma^{-1} \end{pmatrix}, \qquad h \in R^k,$$

$$K_2^{(h)} = \begin{pmatrix} \frac{1}{2}L_h' H_1^* \\ \vdots \\ \frac{1}{2}L_h' H_k^* \end{pmatrix} = \sum_{i=1}^{n} \{L_h\}_i H_i, \quad L_h' = h'C^{-1}F'\Sigma^{-1},$$

$$W^{(h)} = K_1^{(h)}(\Sigma - FC^{-1}F')(K_1^{(h)})' + K_2^{(h)}C^{-1}(K_2^{(h)})'.$$

Here $e_i \in R^k$, $e_i = (0_1, \dots, 0_{i-1}, 1, 0_{i+}, \dots, 0_k)'$. Further, $K^{(int)}$ and $K^{(par)}$ are the intrinsic and parameter effect curvatures of Bates and Watts at the point β_0, respectively (see [**BW**]).

The problem is how to decide if, under the given assumptions, it is possible to use the linear estimator $\widehat{h'\beta} = L_h'(Y - f_0)$ for estimation of the linear function $h(\beta) = h'\beta$ of the parameter. All the following criteria and regions of linearization can be found in [**K1**], resp. [**K2**].

Let $\hat{\beta}(Y, \delta\beta) = \beta_0 + [(F+\Delta)'\Sigma^{-1}(F+\Delta)]^{-1}(F+\Delta)'\Sigma^{-1}(Y - f_0)$ be the BLUE of the parameter β in the model $Y - f_0 \sim N_n[(F+\Delta)\delta\beta, \Sigma]$. The linearization criteria are based on the adequacy of the model to the measured data and the difference between the estimators $\hat{h'\beta}$ and $h'\hat{\beta}(Y, \delta\beta)$ and the difference between their variances.

Definition 1.1. The model is (with respect to the function $h(\beta) = h'\beta$)

(i) c_b-linearizable in the domain $\mathcal{O}_b^{(h)}(\beta_0)$ if

$$|E_\beta[h'\hat{\beta}(Y, \delta\beta) - h'\hat{\beta}]| = |h'b(\delta\beta)| = \left|\frac{1}{2}L_h'\kappa_{\delta\beta}\right| \le c_b\sqrt{h'C^{-1}h}$$

for $\beta = \beta_0 + \delta\beta \in \mathcal{O}_b^{(h)}(\beta_0)$,

(ii) c_d-linearizable in the domain $\mathcal{O}_d(\beta_0)$ if

$$\left|\delta\beta' \frac{\partial(\mathrm{Var}\,[h'\hat{\beta}(Y, \delta\beta)] - \mathrm{Var}\,[h'\hat{\beta}])}{\partial\delta\beta}\Big|_{\delta\beta=0}\right| \le c_d^2 h'C^{-1}h$$

for $\beta_0 + \delta\beta \in \mathcal{O}_d(\beta_0)$,

(iii) c_U-linearizable in the domain $\mathcal{O}_U(\beta_0)$ if

$$\mathrm{Var}\left[\delta\beta' \frac{\partial(h'\hat{\beta}(Y,\delta\beta)-h'\hat{\beta})}{\partial\delta\beta}\Big|_{\delta\beta=0}\right] \leq c_U^2 h' C^{-1} h$$

for $\beta_0+\delta\beta \in \mathcal{O}_U(\beta_0)$,

(iv) the model is (γ,α)-linearizable with respect to its adequacy to the measured data in the domain $\mathcal{O}_{(\gamma\alpha)}(\beta_0)$ if

$$\frac{1}{4}(\kappa_{\delta\beta})'(M_F\Sigma M_F)^+\kappa_{\delta\beta} \leq \delta_t$$

for $\beta_0+\delta\beta \in \mathcal{O}_{(\gamma\alpha)}(\beta_0)$, where $(M_F\Sigma M_F)^+ = \Sigma^{-1}-\Sigma^{-1}FC^{-1}F'\Sigma^{-1}$ is the Moore-Penrose inverse of the matrix $M_F\Sigma M_F$ and δ_t is the threshold value of the noncentrality parameter for which

$$P\{\chi^2_{n-k}(\delta_t) \geq \chi^2_{n-k}(0,1-\alpha)\} = \gamma(>\alpha)$$

where $\chi^2_{n-k}(\delta_t)$ is a random variable with noncentral chi-square distribution with $n-k$ degrees of freedom and with the parameter of noncentrality equal to δ_t; $\chi^2_{n-k}(0,1-\alpha)$ is the $(1-\alpha)$-quantile of the central chi-square distribution with $n-k$ degrees of freedom.

The linearization domains are determined as follows:

Proposition 1.1.

(i) *(a)* $\mathcal{O}_b^{(h)}(\beta_0) = \left\{\beta_0+\delta\beta \ : \ |\delta\beta' K_2^{(h)}\delta\beta| \leq c_b\sqrt{h'C^{-1}h}\right\}$ *for one function* $h(\beta)=h'\beta$

(b) $\mathcal{O}_b(\beta_0) = \left\{\beta_0+\delta\beta \ : \ \delta\beta' C\delta\beta \leq \frac{2c_b}{K^{(par)}}\right\}$ *for every function* $h(\beta)=h'\beta$

(ii) $\mathcal{O}_{(\gamma,\alpha)}(\beta_0) = \left\{\beta_0+\delta\beta \ : \delta\beta' C\delta\beta \leq \frac{2\sqrt{\delta_t}}{K^{(int)}}\right\}$

(iii) $\mathcal{O}_d(\beta_0) = \left\{\beta_0+\delta\beta \ : \ \|\delta\beta\| \leq \frac{c_d^2 h'C^{-1}h}{2\|K_2^{(h)}C^{-1}h\|}\right\}$, *(here* $\|\cdot\|$ *is the Euclidean norm)*

(iv) $\mathcal{O}_U(\beta_0) = \left\{\beta_0+\delta\beta \ : \ \delta\beta' W^{(h)}\delta\beta \leq c_U^2 h'C^{-1}h\right\}$

According to [**K2**], the domain $\mathcal{O}_b$ in Proposition 1.1(i)(a) is replaced by the ellipsoid $\{\beta_0+\delta\beta \ : \ \delta\beta' K^{(h)}\delta\beta \leq c_b\sqrt{h'C^{-1}h}\}$, where $K^{(h)} = \sum_{i=1}^k |\eta_i| p_i p_i'$, if $K_2^{(h)} = \sum_{i=1}^k \eta_i p_i p_i'$ is the spectral decomposition of the matrix $K_2^{(h)}$. All the ellipsoids resulting from Proposition 1.1 are written in the form {center, a_1, f_1, $\dots$, a_k, f_k}, where a_i and f_i are the length and the direction vector of the i-th semi-axis, respectively, or in the form {center, radius} in the case of a ball.

In the case when $\mathcal{O}(\beta_0)$ is given, the use of the linear estimator is appropriate if $\mathcal{O}(\beta_0)$ is contained in all the linearization domains, with the constants c_b, c_d, c_U and (γ, α) chosen according to the requirements of the user. However, in the examples mentioned in [**K1**], [**K2**] or [**P**], another problem is described: the aim is to find a region $\mathcal{O} = \mathcal{O}(\beta_0)$ for given model and point $\beta_0 \in R^k$, such that if we know that the true value $\bar{\beta} \in \mathcal{O}$, then the model can be linearized in β_0. This problem is solved by putting $\mathcal{O} = \mathcal{O}_b \cap \mathcal{O}_d \cap \mathcal{O}_U \cap \mathcal{O}_{(\gamma,\alpha)}$ for some values of the constants. As will be shown below, the values of the constants cannot be chosen arbitrarily.

2. The Choice of the Constants

We can see that Proposition 1.1 allows us to find the corresponding linearization domains for any values of the criterion parameters. But if, for example, the chosen value of the parameter c_b is large, then the possible bias can make the probability that the estimate $\hat{\beta}$ lies inside the domain $\mathcal{O}_b(\beta_0)$ quite small. On the other hand, if the value of c_b is very small, then the situation may occur that the confidence region of $\hat{\beta}$ is greater than $\mathcal{O}(\beta_0)$. Regarding the assumption (i) from Section 1, it is convenient to exclude both of these cases. Now we will put it more precisely.

The assumption (i) from Section 1 means that we have some a priori information about the position of the parameter in the parameter space, given by the domain $[\mathcal{O}](\beta_0)$, and that we can regard β_0 as an estimate of the true value $\bar{\beta}$. The information we get using the estimator $\hat{\beta}$ is given by its confidence region. The question is if the estimation shows the position of the parameter more precisely than β_0.

First, let $\mathcal{O} = \mathcal{O}_b$ for every function $h(\beta) = h'\beta$. Without any loss of generality, we put $\beta_0 = 0$. Let $\mathcal{E} = \{\beta : (\beta - \hat{\beta})' C (\beta - \hat{\beta}) \le \chi^2_k(1-\alpha_1)\}$. Then $\mathcal{E}$ is a $(1-\alpha_1)$ confidence region for $E_{\bar{\beta}}[\hat{\beta}] = \bar{\beta} + b(\bar{\beta})$. According to the proof of Theorem 2.9. in [**K1**], we have:

Proposition 2.1. *For each $\beta \in \mathcal{O}_b$, $b'(\beta) C b(\beta) \le c_b^2$, i.e.*

$$\bar{\beta} \in \left\{ \beta : (\beta - E_{\bar{\beta}}[\hat{\beta}])' C (\beta - E_{\bar{\beta}}[\hat{\beta}]) \le c_b^2 \right\}.$$

Thus the ellipsoid $\mathcal{C} = \left\{ \hat{\beta} \ ; \ \frac{c_b}{\sqrt{\lambda_i}} + \sqrt{\frac{\chi^2_k(1-\alpha_1)}{\lambda_i}} \ , \ f_i ; \ i = 1, \dots, k \right\}$, where $C = \sum_{i=1}^k \lambda_i f_i f_i'$ is the spectral decomposition of C, contains the true value $\bar{\beta}$ with probability at least $1 - \alpha_1$. On the other hand, $\bar{\beta} \in \mathcal{O}_b = \left\{ \beta_0 \ ; \ \sqrt{\frac{2c_b}{K^{(par)}\lambda_i}} \ , \ f_i ; \ i = 1, \dots, k \right\}$. Hence if $\mathcal{C}$ is larger than $\mathcal{O}_b$, i.e. $\frac{c_b}{\sqrt{\lambda_i}} + \sqrt{\frac{\chi^2_k(1-\alpha_1)}{\lambda_i}} \ge \sqrt{\frac{2c_b}{K^{(par)}\lambda_i}}$, for $i = 1, \dots, k$, then the a priori estimate β_0 is more precise than $\hat{\beta}$ and the estimation has no sense.

The aim is to find such value of the criterion parameter c_b that

$$c_b + \sqrt{\chi_k^2(1-\alpha_1)} < \sqrt{\frac{2c_b}{K^{(par)}}}\,. \tag{1}$$

Proposition 2.2. *Condition* (1) *can be satisfied iff*

$$\delta^2 = 2K^{(par)}\sqrt{\chi_k^2(1-\alpha_1)} < 1$$

i.e.

$$K^{(par)} < \frac{1}{2\sqrt{\chi_k^2(1-\alpha_1)}}\,.$$

The needed $c_b \in (c_{b1}, c_{b2})$, *where* $c_{b1,2} = \frac{1}{2K^{(par)}}[1 \pm \sqrt{1-\delta^2}]^2$.

Proof.

$$(1) \Leftrightarrow c_b^2 + 2c_b\left(\sqrt{\chi_k^2(1-\alpha_1)} - \frac{1}{K^{(par)}}\right) + \chi_k^2(1-\alpha_1) < 0\,.$$

The rest of the proof is obvious. □

Remark 2.1. Here the question may arise of how more accurate than β_0 the estimate $\hat{\beta}$ may possibly be, i.e. what is the smallest ratio of the lengths of the semi-axes of $\mathcal{C}$ to that of $\mathcal{O}_b$. As can be easily verified, the smallest ratio is δ and it is attained for $c_b = \sqrt{\chi_k^2(1-\alpha_1)}$.

Now, let $\mathcal{O} = \mathcal{O}_b \cap \mathcal{O}_{(\gamma\alpha)}$. Similarly as for $\mathcal{O}_b$ we get a condition

$$c_b + \sqrt{\chi_k^2(1-\alpha_1)} < \sqrt{\frac{2\sqrt{\delta_t}}{K^{(int)}}} \tag{3}$$

where $c_b \in (c_{b1}, c_{b2})$.

Proposition 2.3. *Condition* (3) *can be satisfied iff*

$$K^{(int)} < \frac{K^{(par)}\sqrt{\delta_t}}{c_{b1}} = \frac{\delta^2\sqrt{\delta_t}}{c_{b1}2\sqrt{\chi_k^2(1-\alpha_1)}}\,. \tag{4}$$

Proof. The condition in question can be satisfied iff it is satisfied for the smallest possible $c_b = c_{b1}$, i.e. iff

$$\frac{1}{2K^{(par)}}\left[1 - \sqrt{1-\delta^2}\right]^2 + \sqrt{\chi_k^2(1-\alpha_1)} < \sqrt{\frac{2\sqrt{\delta_t}}{K^{(int)}}}\,.$$

Using the equality $K^{(par)} = \frac{\delta^2}{2\sqrt{\chi_k^2(1-\alpha_1)}}$ we get, after some rearrangements

$$\sqrt{K^{(int)}} < \frac{\sqrt[4]{\delta_t}\delta^2}{\sqrt{2}\sqrt{\chi_k^2(1-\alpha_1)}[1-\sqrt{1-\delta^2}]}$$

which yields the inequality (4). $\square$

In the case when only one function $h(\beta) = h'\beta$ is estimated, we consider the interval $I_h = [\min_{\beta\in\mathcal{O}} h'\beta, \max_{\beta\in\mathcal{O}} h'\beta]$. Obviously $h'\bar{\beta} \in I_h$. This interval determines the a priori information about the true value of $h'\bar{\beta}$ contained in $\mathcal{O}$. For reasons described in the beginning of this section, the length of I_h will be compared to the length of the $(1-\alpha_1)$ confidence interval for $h'\bar{\beta}$, given by the estimator $h'\hat{\beta}$.

Let $\mathcal{O} = \mathcal{O}_b$ for $h(\beta)$. We suppose that $h \in \mathcal{M}(K^{(h)})$, because if $h \notin \mathcal{M}(K^{(h)})$ then $I_h = (-\infty, \infty)$ and $\mathcal{O}$ contains no information about the value of $h'\bar{\beta}$.

Proposition 2.4. *Under the above assumptions,*

$$I_h = \left\{ x : |x| \leq \sqrt{c_b\sqrt{h'C^{-1}h}h'[K^{(h)}]^+h} \right\}$$

where $[K^{(h)}]^+ = \sum_{i=1}^k \frac{1}{|\eta_i|}p_ip_i'$ *is the Moore- Penrose inverse of the matrix* $K^{(h)} = \sum |\eta_i|p_ip_i'$ *(the spectral decomposition of* $K^{(h)}$*).*

Proof. It is obvious, that $\min_{\beta\in\mathcal{O}_b} h'\beta = \min_{\beta:\beta'K^{(h)}\beta=c_b\sqrt{h'C^{-1}h}} h'\beta$. (The same holds for maximum). By the use of the Lagrange multipliers:

$$\Phi(\beta) = h'\beta - \lambda(\beta'K^{(h)}\beta - c_b\sqrt{h'C^{-1}h})$$

$$\frac{\partial\Phi(\beta)}{\partial\beta} = h - 2\lambda K^{(h)}\beta = 0$$

$$\beta'K^{(h)}\beta = c_b\sqrt{h'C^{-1}h}\,.$$

Using the supposition $h \in \mathcal{M}(K^{(h)})$ we easily get the statement of the proposition. $\square$

The $(1-\alpha_1)$ confidence interval for $E_{\bar{\beta}}[h'\hat{\beta}]$ is $\mathcal{E}_h = \{x : |x - h'\hat{\beta}| < u(1-\alpha_1/2)\sqrt{h'C^{-1}h}\}$; here $u(1-\alpha_1/2)$ is the $(1-\alpha_1/2)$ quantile of the normal distribution $N[0,1]$. Moreover, for $\bar{\beta} \in \mathcal{O}$, $h'\bar{\beta}$ lies in the interval $\{x : |x - E_{\bar{\beta}}[h'\hat{\beta}]| \leq c_b\sqrt{h'C^{-1}h}\}$. It follows that $|h'\bar{\beta} - h'\hat{\beta}| \leq \sqrt{h'C^{-1}h}(c_b + u(1-\alpha_1/2))$ with probability at least $1-\alpha_1$. This leads to the condition

$$c_b\sqrt{h'C^{-1}h} + u(1-\alpha_1/2)\sqrt{h'C^{-1}h} < \sqrt{c_b}\sqrt[4]{h'C^{-1}h}\sqrt{h'[K^{(h)}]^+h}\,. \tag{5}$$

Let us denote

$$C_h = \begin{cases} \frac{2\sqrt{h'C^{-1}h}}{h'[K^{(h)}]^+h} & \text{for } h \in \mathcal{M}(K^{(h)}) \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 2.5. *Condition* (5) *can be fulfilled iff* $\delta^2 = 2C_h u(1-\alpha_1/2) < 1$, *i.e.*

$$C_h < \frac{1}{2u(1-\alpha_1/2)}\,. \tag{6}$$

The constant c_b *must then lie in the interval* (c_{b1}, c_{b2}), *where* $c_{b1,2} = \frac{1}{2C_h}[1 \pm \sqrt{1-\delta^2}]^2$.

Proof. The same as Proposition 2.2. □

Remark 2.2. As can be seen, the value C_h plays a similar role here as the parametric curvature $K^{(par)}$ in Proposition 2.2. For a one-parameter model, i.e. for $k = 1$, these values are equal (for each $h \in R$). The value δ has the same interpretation for $c_b = u(1-\alpha_1/2)$ $(= \sqrt{\chi_1^2(1-\alpha_1)})$ as in Remark 2.1.

If $\mathcal{O} = \mathcal{O}_b \cap \mathcal{O}_d$ for the function $h'\beta$, then according to Proposition 1.1, $|h'\bar{\beta}| \leq \|h\| \frac{c_d^2 h' C^{-1} h}{2\|K_2^{(h)} C^{-1} h\|}$ and the condition for c_b and c_d is

$$c_d^2 > \frac{2\|K_2^{(h)} C^{-1} h\|}{\sqrt{h' C^{-1} h}\|h\|}(c_b + u(1-\alpha_1/2)). \tag{7}$$

Proposition 2.6. *Condition* (7) *can be satisfied iff*

$$c_d^2 > \frac{1}{C_h}\left[1 - \sqrt{1-\delta^2}\right] \frac{2\|K_2^{(h)} C^{-1} h\|}{\sqrt{h' C^{-1} h}\|h\|}\,. \tag{8}$$

Proof. It can be easily verified, taking into account that $c_b \in (c_{b1}, c_{b2})$. □

If $\mathcal{O} = \mathcal{O}_b \cap \mathcal{O}_d \cap \mathcal{O}_U$ for the function $h'\beta$, similar arguments as in the proof of Proposition 2.5 give

$$I_h = \left\{ x : |x| \leq c_U \sqrt{h' C^{-1} h h' [W^{(h)}]^+ h} \right\}$$

and the condition is

$$c_U > \frac{c_b + u(1-\alpha_1/2)}{\sqrt{h' \left[W^{(h)}\right]^+ h}} \tag{9}$$

for $h \in \mathcal{M}(W^{(h)})$ and $c_b \in (c_{b1}, c_{b2})$.

Proposition 2.7. *Condition* (9) *can be satisfied iff*

$$c_U > \frac{1}{C_h} \frac{[1-\sqrt{1-\delta^2}]}{\sqrt{h' \left[W^{(h)}\right]^+ h}}. \tag{10}$$

Proof. Exactly the same as before. □

We can conclude that (2) resp. (6) gives a necessary condition for a model to be linearized with respect to the bias, while (4) gives a necessary condition for linearization with respect to the adequacy of the model. Further, the conditions (8) and (10) give the lower bounds for c_d^2 and c_U, respectively. If the necessary conditions for bias and adequacy are fulfilled and the values of c_d^2 and c_U satisfying (8) and (10) are acceptable for the users, then the values of the criterion parameters that will be used in the procedure of finding the linearization domains may be found using (1) resp. (5), (3), (7) and (9).

The values of the parameters obtained in the described way determine neighbourhoods of the point β_0 with the following properties:

(i) the linear estimator and its characteristics linked with any point of this neighbourhood differ from the best estimator and its characteristics in bounds which are admissible for the user;

(ii) the used linear estimators give better value of β that β_0 (the a priori information on β, given by β_0, is smaller than the information given by Y.

References

[**BW**] Bates D. M. and Watts D. G., *Relative Curvature Measures of Nonlinearity*, J. R. Statist. Soc. **B 42**, no. 1 (1980), 1–25.

[**K1**] Kubáček L., *On a Linearization of Regression Models*, Applications of Mathematics **40** (1995), 61–78.

[**K2**] ______, *Models with a Low Nonlinearity*, Proceedings of Probastat 94, Tatra Mountains, Math. publ., (to appear).

[**P**] Pulmannová A., *Lineárne odhady v regresných modeloch*, Diploma thesis, Faculty of Mathematics and Physics, Comenius University, Bratislava, 1994.

A. Jenčová, Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, 814 73 Bratislava, Slovakia