Anna Jenčová

Linearization conditions for regression models with unknown variance parameter

Applications of Mathematics, Vol. 45 (2000), No. 2, 145–160

Persistent URL: `http://dml.cz/dmlcz/134433`

LINEARIZATION CONDITIONS FOR REGRESSION MODELS WITH UNKNOWN VARIANCE PARAMETER

Anna Jenčová, Bratislava

(Received November 23, 1998)

Abstract. In the case of the nonlinear regression model, methods and procedures have been developed to obtain estimates of the parameters. These methods are much more complicated than the procedures used if the model considered is linear. Moreover, unlike the linear case, the properties of the resulting estimators are unknown and usually depend on the true values of the estimated parameters. It is sometimes possible to approximate the nonlinear model by a linear one and use the much more developed linear methods, but some procedure is needed to recognize such situations. One attempt to find such a procedure, taking into account the requirements of the user, is given in [4], [5], [3], where the existence of an a priori information on the parameters is assumed. Here some linearization criteria are proposed and the linearization domains, i.e. domains in the parameter space where these criteria are fulfilled, are defined. The aim of the present paper is to use a similar approach to find simple conditions for linearization of the model in the case of a locally quadratic model with unknown variance parameter σ^2. Also a test of intrinsic nonlinearity of the model and an unbiased estimator of this parameter are derived.

Keywords: nonlinear regression models, linearization domains, linearization conditions

MSC 2000: 62F10

1. Linearization criteria

Let us consider the nonlinear regression model

$$Y = f(\beta) + \varepsilon, \quad \varepsilon \sim N_n[0, \sigma^2 W] \tag{1}$$

where $f\colon \mathbb{R}^k \to \mathbb{R}^n$ is a known function with continuous second derivatives, $\beta \in \mathbb{R}^k$ and $\sigma \in \mathbb{R}^+$ are unknown parameters and W is a known positively definite matrix. In accordance with [4], it is further assumed that

(1) the true value $\overline{\beta}$ of the parameter β is known to lie in a neighbourhood $\mathcal{O}(\beta_0)$ of a given point $\beta_0 \in \mathbb{R}^k$,

(2) the third derivatives of the function f can be neglected for $\beta \in \mathcal{O}(\beta_0)$,

(3) the model is regular at the point β_0, i.e. the matrix $F = \frac{\partial f(\beta)}{\partial \beta'}|_{\beta=\beta_0}$ has the full rank.

It follows from the assumptions (1) and (2) that the parameter space of the model can be restricted to the set $\mathcal{O}(\beta_0)$ and model (1) has the form

$$Y = f_0 + F(\beta - \beta_0) + \frac{1}{2} h^{ij}(\beta_i - \beta_{0i})(\beta_j - \beta_{0j}) + \varepsilon \tag{2}$$

where $f_0 = f(\beta_0)$, $h^{ij} = \frac{\partial}{\partial \beta_i \partial \beta_j} f(\beta)|_{\beta=\beta_0}$ and the expression $h^{ij} v_i v_j$ denotes the sum over the indices i and j; this convention will be used throughout the text. Without loss of generality, we will assume that $\beta_0 = 0$ and $f_0 = 0$.

The linearization of model (2) at the point $\beta = 0$ usually means the approximation of the function f by the linear part of its Taylor formula at this point, i.e. the model is replaced by

$$Y = F\beta + \varepsilon. \tag{3}$$

This means that the solution locus $\mathcal{E} = \{f(\beta)\colon\ \beta \in \mathcal{O}\}$ is replaced by $\mathcal{E}^* = \{F\beta,\ \beta \in \mathcal{O}\}$, i.e. a part of its tangent space at the point $\beta = 0$. But this may be inadequate to the measured data, if the distance of the true mean value $f(\overline{\beta})$ of Y from the set $\mathcal{E}^*$ is large. It is clear that this distance is caused only by the intrinsic curvature of the model (see [1] for the definition of the Bates-Watts intrinsic and parameter effect curvatures). Let us suppose that T is a test of the intrinsic linearity of the model with significance level α, one such test will be given below. This test can then be used to derive a linearization criterion.

Definition 1.1. Model (2) is $(\alpha, d\alpha)$-linearizable with respect to the adequacy of the model to the measured data if

$$P\{\text{the test } T \text{ rejects the hypothesis of intrinsic linearity of the model}\} \leqslant \alpha + d\alpha$$

where $d\alpha \ll \alpha$.

Further, let the estimation of a linear functional $h(\beta) = h'\beta$ of the parameter be considered. It is known from the linear theory that the BLUE of this functional in model (3) is given by $h'\hat{\beta}(Y, 0) = h'(F'W^{-1}F)^{-1}F'W^{-1}Y$. In the original model, however, this estimator is biased, but this bias can be neglected, if it is small compared to the square root of the variance of the estimator. Moreover, if the model is linearized at a different point β, then the resulting estimator, denoted by $h'\hat{\beta}(Y, \beta)$, and its properties may depend heavily on β. These considerations lead to the following linearization criteria.

Definition 1.2. Model (2) is on the set $\mathcal{O}$

(i) c_b-linearizable with respect to the bias for the functional h if

$$\forall \overline{\beta} \in \mathcal{O}\ |E_{\overline{\beta}}[h'\hat{\beta}(Y,0) - h'\overline{\beta}]| \leqslant c_b \sqrt{\operatorname{Var}[h'\hat{\beta}(Y,0)]},$$

(ii) c_m-linearizable with respect to the mean for the functional h if

$$\forall \overline{\beta}, \beta \in \mathcal{O}\ |E_{\overline{\beta}}[h'\hat{\beta}(Y,\beta) - h'\hat{\beta}(Y,0)]| \leqslant c_m \sqrt{\operatorname{Var}[h'\hat{\beta}(Y,0)]},$$

(iii) c_d-linearizable with respect to the variance for the functional h if

$$\forall \beta \in \mathcal{O}\ |\operatorname{Var}[h'\hat{\beta}(Y,\beta)] - \operatorname{Var}[h'\hat{\beta}(Y,0)]| \leqslant c_d^2 \operatorname{Var}[h'\hat{\beta}(Y,0)],$$

(iv) c_U-linearizable with respect to the estimator for the functional h if

$$\forall \beta \in \mathcal{O}\ \operatorname{Var}[h'\hat{\beta}(Y,\beta) - h'\hat{\beta}(Y,0)] \leqslant c_U^2 \operatorname{Var}[h'\hat{\beta}(Y,0)].$$

The criterion parameters α, $d\alpha$, c_b, c_m, c_d and c_U should be chosen by a statistician according to the requirements of the user.

Remark 1.1. Similar linearization criteria, based partially on slightly different considerations, were given in [4], [5].

2. Linearization domains

Linearization criteria, defined in the previous section, can be now used to find linearization domains, which are defined as sets on which linearization criteria are satisfied for some choice of the criterion parameters. To do this easily, we will first find a suitable parametrization of the model.

Let us consider the n-dimensional vector space $\mathbb{R}^n$ with the inner product $\langle x, y\rangle_{W^{-1}} = x'W^{-1}y$. We denote by $\mathcal{M} = \operatorname{span}\{F_{\cdot 1}, \ldots, F_{\cdot k}\}$ the tangent space to the solution locus $\mathcal{E}$ at the point $\beta = 0$ and by $\mathcal{M}^{\perp}$ its orthogonal complement—the ancillary space. Let P and M be the corresponding orthogonal projectors. We define subspaces $\mathcal{M}_1 = \operatorname{span}\{Ph^{ij},\ i,j = 1,\ldots,k\} \subseteq \mathcal{M}$ and $\mathcal{M}_2 = \operatorname{span}\{Mh^{ij},\ i,j = 1,\ldots,k\} \subseteq \mathcal{M}^{\perp}$. Let the columns of the $n \times k$ matrix $J = (J_1, J_2)$ form an orthonormal basis of $\mathcal{M}$, such that $J_1 = (p_1, \ldots, p_d)$ is an orthonormal basis of $\mathcal{M}_1$. Similarly, let $\Omega = (\Omega_1, \Omega_2)$ be an orthonormal basis of $\mathcal{M}^{\perp}$, such that $\Omega_1 = (m_1, \ldots, m_p)$ is an orthonormal basis of $\mathcal{M}_2$. Let L be a $k \times k$ nonsingular matrix such that $F = JL$. We introduce a new parameter $\theta = L\beta$ to obtain the model

$$Y = J\theta + \frac{1}{2} g^{ij}\theta_i\theta_j + \varepsilon \tag{4}$$

where $g^{ij} = h^{ml} r_{mi} r_{lj}$, $i, j = 1, \ldots, k$ and $\{r_{ij}\} = R = L^{-1}$. Since the vectors g^{ij}, $i, j = 1, \ldots, k$ are linear combinations of the vectors h^{ij}, we may write

$$g^{ij} = J_1 J_1' W^{-1} g^{ij} + \Omega_1 \Omega_1' W^{-1} g^{ij} = \gamma_\alpha^{ij} p^\alpha + \delta_\beta^{ij} m^\beta.$$

Hence the 3-dimensional arrays $\{\gamma_l^{ij}\}$ and $\{\delta_l^{ij}\}$ fully describe the nonlinearity of the model on the set $\mathcal{O}$. Let $\gamma^{ij} = (\gamma_1^{ij}, \ldots, \gamma_d^{ij})'$ and $\delta^{ij} = (\delta_1^{ij}, \ldots, \delta_p^{ij})'$. Then we may put

$$K^{\text{int}} = \sup_v \frac{\|\delta^{ij} v_i v_j\|}{v'v} \sigma,$$
$$K^{\text{par}} = \sup_v \frac{\|\gamma^{ij} v_i v_j\|}{v'v} \sigma$$

for the intrinsic and parameter effect curvatures at the point $\beta = 0$. It is obvious that this is equivalent to the definition of the curvatures given in [1] (for the case of a known parameter σ). In what follows, the unknown parameter σ^2 will be replaced by its estimate s^2, which will be specified later.

Now we may proceed to the determination of the linearization domains. First, let us consider the criterion of adequacy of the model to the data. In [4], the statistic $R_0^2 = \sigma^{-2}\|MY\|_{W^{-1}}^2$ is used to obtain the required test. But not all of the components of the residual vector MY are influenced by the intrinsic nonlinearity of the model. Moreover, this test cannot be used unless the parameter σ is known. Therefore we propose a test statistic

$$F = \frac{\|\Omega_1' W^{-1} Y\|^2}{ps^2}$$

where $s^2 = \frac{1}{n-k-p} Y' W^{-1} \Omega_2 \Omega_2' W^{-1} Y$. It is easy to see that F has the noncentral $F_{k,\nu}(\delta)$ distribution with the noncentrality parameter $\delta = \delta(\theta) = \frac{1}{4}\sigma^{-2}\|\delta^{ij}\theta_i\theta_j\|^2$, where $\nu = n - k - p$. It is also clear that under the assumptions of Section 1, s^2 is an unbiased estimator of the parameter σ^2 with ν degrees of freedom, but it can be used only if ν is sufficiently large. But $p \leqslant \dim\{g^{ij},\ i, j = 1, \ldots, k\} \leqslant \frac{k}{2}(k+1)$, and therefore this estimator can always be used if $n \gg k + \frac{k}{2}(k+1)$.

The criterion from Definition 1.1 can now be restated as follows.

Definition 2.1. Model (4) is $(\alpha, d\alpha)$-linearizable with respect to the adequacy of the model to the measured data and for the test statistic F if for each $\theta \in \mathcal{O}$

$$P_\theta\{F \geqslant F_{p,\nu}(\alpha)\} \leqslant \alpha + d\alpha$$

where $F_{p,\nu}(\alpha)$ is the critical value of the central $F_{p,\nu}$ distribution.

Let δ_t be the threshold value of the noncentrality parameter for which the criterion from Definition 2.1 is satisfied. Then the linearization domain corresponding to this criterion can be defined as the set satisfying the condition $s^{-2}\|\delta^{ij}\theta_i\theta_j\|^2 \leqslant \delta_t$.

Proposition 2.1. *Let the domain $\mathcal{O}$ be given by the condition*

$$s^{-2}\theta'\theta \leqslant \frac{2\sqrt{\delta_t}}{K^{\text{int}}}.$$

Then $\mathcal{O}$ is the linearization domain corresponding to the criterion from Definition 2.1.

P r o o f. The proof is analogous to that of Proposition 4.1 in [5]. □

Next, estimators of linear functionals $h(\theta) = h'\theta$ will be considered. Using the above mentioned orthogonal matrices, we obtain

$$\begin{pmatrix} J_1' \\ J_2' \\ \Omega_1' \\ \Omega_2' \end{pmatrix} W^{-1}Y = \begin{pmatrix} I \\ 0 \end{pmatrix} \theta + \frac{1}{2} \begin{pmatrix} \gamma^{ij}\theta_i\theta_j \\ 0 \\ \delta^{ij}\theta_i\theta_j \\ 0 \end{pmatrix} + \xi$$

where $\xi \sim N_n[0, \sigma^2 I]$. If we denote $\eta(\theta) = f(R\theta)$ then the estimator $h'\hat{\theta}(Y, \theta)$ has the form

$$\widehat{h'\theta}(Y,\theta) = h'\theta + h'\left[I + \begin{pmatrix} \Delta_p \\ 0 \end{pmatrix} + (\Delta_p', 0) + \Delta_p'\Delta_p + \Delta_m'\Delta_m\right]^{-1}$$
$$\left[(I,0) + (\Delta_p', 0, \Delta_m', 0)\right] \begin{pmatrix} J' \\ \Omega' \end{pmatrix} W^{-1}(Y - \eta(\theta))$$

where $\Delta_p(\theta) = (\gamma^{i1}\theta_i, \ldots, \gamma^{ik}\theta_i)$ and $\Delta_m(\theta) = (\delta^{i1}\theta_i, \ldots, \delta^{ik}\theta_i)$.

Lemma 2.1. *If the estimator $h'\hat{\theta}(Y, \theta)$ is approximated by the linear part of its Taylor expansion at the point $\theta = 0$, then*

$$h'\hat{\theta}(Y,\theta) \doteq h'J'W^{-1}Y + \theta'(G^h)'\Omega_1'W^{-1}Y - \theta'K^hJ'W^{-1}Y$$

where K^h is a $k \times k$ symmetric matrix with elements $K_{ij}^h = h^l\gamma_l^{ij}$ and G^h is a $p \times k$ matrix such that $G_{ij}^h = \delta_i^{jl}h_l$, $i = 1 \ldots, p$, $j = 1, \ldots, k$.

P r o o f. It follows easily from the equality $\partial A^{-1} = -A^{-1}\partial A A^{-1}$ and from linearity of the functions $\theta \mapsto \Delta_p(\theta)$ and $\theta \mapsto \Delta_m(\theta)$. $\square$

Lemma 2.2. *If the approximation from the previous lemma is used, then*

$$\mathrm{Var}[\widehat{h'\theta}(Y,\theta)] \doteq \sigma^2 h'h + \sigma^2\theta'(G^h)'G^h\theta + \sigma^2\theta'K^hK^h\theta - 2\sigma^2\theta'K^hh$$

and

$$\mathrm{Var}[\widehat{h'\theta}(Y,\theta) - \widehat{h'\theta}(Y,0)] = \sigma^2\theta'(G^h)'G^h\theta + \sigma^2\theta'K^hK^h\theta.$$

P r o o f. The statement follows from the independence of the vectors $\Omega_1'W^{-1}Y$ and $J'W^{-1}Y$. $\square$

For the case of model (4), the linearization criteria given in Definition 1.2 can now be restated using these approximations. The resulting linearization domains are given in the next proposition. Let A be a symmetric matrix with the spectral decomposition $A = \sum_i \mu_i p_i p_i'$, then the matrix $|A|$ is defined by $|A| = \sum_i |\mu_i| p_i p_i'$.

Proposition 2.2. *Let $\mathcal{O}$ be such that*

(i) $\theta'|K^h|\theta \leqslant 2c_b s\sqrt{h'h}$

(ii) $\theta'(G^h)'G^h\theta \leqslant c^2 h'h$

for all $\theta \in \mathcal{O}$. Let $k_h = \frac{|\lambda_h|}{\sqrt{h'h}}s$, where λ_h is an eigenvalue of the matrix K^h with the greatest absolute value. Then, for the functional h, the model is on $\mathcal{O}$

(a) *c_b-linearizable with respect to the bias,*

(b) *c_d-linearizable with respect to the variance, where $c_d^2 = c^2 + 2c_bk_h + 2\sqrt{2c_bk_h}$,*

(c) *c_U-linearizable with respect to the estimator, where $c_U^2 = c^2 + 2c_bk_h$.*

Moreover, if $\sup_{\mathcal{O}} s^{-1}\|\delta^{ij}\bar{\theta}_i\bar{\theta}_j\| = M_\delta < \infty$ and $\sup_{\mathcal{O}} s^{-1}\|\gamma^{ij}\bar{\theta}_i\bar{\theta}_j\| = M_\gamma < \infty$, then the model is on $\mathcal{O}$

(d) *c_m-linearizable with respect to the mean, where $c_m = \frac{1}{2}cM_\delta + \frac{1}{2}\sqrt{2c_bk_h}M_\gamma + 2c_b$.*

P r o o f. The statement (a) follows from the fact that $|E_\theta[h'\hat{\theta}(Y,0) - h'\theta]| = |\theta'K^h\theta| \leqslant \theta'|K^h|\theta$. (c) is proved similarly as (b). (b) From Schwarz inequality and Lemma 2.2, it is clear that it suffices to prove that

$$\theta'(G^h)'G^h\theta + \theta'K^hK^h\theta + 2\sqrt{\theta'K^hK^h\theta}\sqrt{h'h} \leqslant c^2h'h.$$

Let $K^h = \sum_i \lambda_i f_i f_i'$ be the spectral decomposition of the matrix K^h and let $\mathcal{O}_1$ be the ellipsoid given by condition (i). Let $\theta \in \mathcal{O}_1$ and $\theta = \sum_i x_i f_i$. Then

$$\theta'K^hK^h\theta = \sum_i x_i^2\lambda_i^2 \leqslant |\lambda_h| \sum_i x_i^2|\lambda_i| = k_h\theta'|K^h|\theta\frac{\sqrt{h'h}}{s} \leqslant 2k_hc_bh'h.$$

The rest of the proof is obvious.

(d) From Lemma 2.1, we have

$$\begin{aligned}|E_{\bar\theta}[\widehat{h'\theta}(Y,\theta)-\widehat{h'\theta}(Y,0)]| &= \Big|\frac{1}{2}\theta'(G^h)'\delta^{ij}\bar\theta_i\bar\theta_j-\theta'K^h\Big(\bar\theta-\frac{1}{2}\gamma^{ij}\bar\theta_i\bar\theta_j\Big)\Big|\\ &\leqslant \frac{1}{2}s\sqrt{\theta'(G^h)'G^h\theta}s^{-1}\|\delta^{ij}\bar\theta_i\bar\theta_j\|+\frac{1}{2}s\sqrt{\theta'K^hK^h\theta}s^{-1}\|\gamma^{ij}\bar\theta_i\bar\theta_j)\|\\ &\quad + ss^{-1}\theta'|K^h|\bar\theta.\end{aligned}$$

The proof is completed by using the proof of (b) and Lemma 2.3 below. □

Next we will find domains on which the linearization criteria are satisfied for all linear functionals $h(\theta)=h'\theta$. We prove some auxiliary statements first. The domains are given in Proposition 2.3. The proof of the following lemma can be found e.g. in [2], p. 180.

Lemma 2.3. *Let A be an arbitrary symmetric matrix. Then*

$$\|A\|=\sup_{\|x\|=\|y\|=1}|y'Ax|=\sup_{\|x\|=1}|x'Ax|.$$

Lemma 2.4. *Let $\{t_l^{ij}\}$ be a 3-dimensional array such that $t_l^{ij}=t_l^{ji}$ $\forall i,j,l$ and suppose that C is a p.d. matrix. Then*

$$\sup_{\|v\|=\|u\|=1}\|t^{ij}v_iu_j\|_C=\sup_{\|v\|=1}\|t^{ij}v_iv_j\|_C.$$

P r o o f. Let R^x be a matrix with elements $R^x_{ij}=x'Ct^{ij}$, $\|x\|_C=1$. Then R^x is a symmetric matrix and $\|t^{ij}u_iv_j\|_C=\sup_x|u'R^xv|$. According to Lemma 2.3 we have

$$\begin{aligned}\sup_x\Big(\sup_{\|u\|=\|v\|=1}|u'R^xv|\Big) &= \sup_x\Big(\sup_{\|v\|=1}|v'R^xv|\Big)\\ &= \sup_{\|v\|=1}\Big(\sup_x|v'R^xv|\Big)=\sup_{\|v\|=1}\|t^{ij}v_iv_j\|_C.\end{aligned}$$

□

Lemma 2.5. *If $\mathcal{O}$ is given by $s^{-2}\theta'\theta\leqslant M^2$, then*

$$M_\gamma\leqslant K^{\mathrm{par}}M^2,\quad M_\delta\leqslant K^{\mathrm{int}}M^2.$$

P r o o f. It is easy to prove that ([6]) $s^{-1}\|\gamma^{ij}\theta_i\theta_j\|\leqslant K^{\mathrm{par}}s^{-2}\theta'\theta$. The statement for M_δ is proved analogously. □

Proposition 2.3. *Let*

$$\mathcal{O} = \{\theta\colon\ s^{-2}\|\theta\|^2 \leqslant M^2\}.$$

Then the model (4) *is for all functionals* $h'\theta$

(a) c_b*-linearizable with respect to the bias, where* $c_b = \frac{1}{2}M^2K^{\mathrm{par}}$,

(b) c_d*-linearizable with respect to the variance, where* $c_d^2 = M^2((K^{\mathrm{int}})^2 + (K^{\mathrm{par}})^2) + 2MK^{\mathrm{par}}$,

(c) c_U*-linearizable with respect to the estimator, where* $c_U^2 = M^2((K^{\mathrm{int}})^2 + (K^{\mathrm{par}})^2)$,

(d) c_m*-linearizable with respect to the mean, where* $c_m = M^3\frac{1}{2}((K^{\mathrm{int}})^2 + (K^{\mathrm{par}})^2) + M^2K^{\mathrm{par}}$.

Proof. (a) Can be found in [4].

(b) According to (a) $|\theta' K^h\theta| \leqslant M^2K^{\mathrm{par}}s\|h\|$ holds for all h. From the proof of Proposition 2.2 (b) and using the fact that

$$k_h = \frac{|\lambda_h|s}{\sqrt{h'h}} = \sup_\theta \frac{|h^l\gamma_l^{ij}\theta_i\theta_j|}{\theta'\theta}\frac{s}{\sqrt{h'h}} \leqslant \sup_\theta \frac{\|\gamma^{ij}\theta_i\theta_j\|}{\theta'\theta}s = K^{\mathrm{par}}$$

we get

$$\theta' K^h K^h \theta \leqslant M^2(K^{\mathrm{par}})^2 h'h.$$

Further, it is clear from the definition of the matrix G^h that $\theta'(G^h)'G^h\theta = \|\delta^{ij}h_i\theta_j\|^2$. According to Lemma 2.4, $\|\delta^{ij}h_i\theta_j\| \leqslant K^{\mathrm{int}}\|h\|s^{-1}\|\theta\|$. It follows that for $\theta \in \mathcal{O}$ and $\forall h$,

$$\theta'(G^h)'G^h\theta \leqslant (K^{\mathrm{int}})^2M^2\|h\|^2.$$

(c) The same as (b).

(d) From part (b) and Lemma 2.5 we get

$$s\sqrt{\theta'(G^h)'G^h\theta}M_\delta \leqslant sK^{\mathrm{int}}M\sqrt{h'h}K^{\mathrm{int}}M^2$$

and

$$s\sqrt{\theta' K^h K^h\theta}M_\gamma \leqslant (K^{\mathrm{par}})^2M^3s\sqrt{h'h}.$$

It is easy to prove, using Lemmas 2.3 and 2.5, that $|\theta' K^h\bar{\theta}| \leqslant M^2K^{\mathrm{par}}s\|h\|$ for all $h \in \mathbb{R}^k$ and $\theta \in \mathcal{O}$. $\square$

Propositions 2.2 and 2.3 allow us to find a linearization domain for any set of criterion parameters. These domains have the following interpretation. If the linearization approach is to be used in a given model and the error caused by the nonlinearity of the model should be less than the tolerance of the user, specified by given criterion parameters, then the a priori information domain should be contained in the corresponding linearization domain. As was pointed out in [3], this approach leads to certain natural restrictions on the model as well as on the values of the criterion parameters. The idea here is that the information on the parameters yielded from the estimation procedure, which is given by the confidence region based on the estimator, is required to be more precise that the a priori information given by $\mathcal{O}$. These restrictions will be stated in the next section.

3. Criterion parameters and linearization conditions

Let us now suppose that the true value of the parameter $\overline{\beta}$ lies in the domain $\mathcal{O}$ from Proposition 2.3. Then it is clear that

$$\mathcal{C} = \Big\{\theta\colon\ s^{-2}\big(\hat{\theta}(Y,0)-\theta\big)'\big(\hat{\theta}(Y,0)-\theta\big) \leqslant \Big(\frac{1}{2}M^2K^{\mathrm{par}} + \sqrt{kF_{k,\nu}(1-\alpha)}\Big)^2\Big\}$$

is a $1-\alpha$ confidence region for θ. As was said at the end of the previous section, we now require that

$$\frac{1}{2}M^2K^{\mathrm{par}} + \sqrt{kF_{k,\nu}(1-\alpha)} \leqslant M. \tag{5}$$

Solving this equation, we easily get the restriction on M.

Proposition 3.1. *Let*

$$K^{\mathrm{par}} = \frac{\omega^2}{2\sqrt{kF_{k,\nu}(1-\alpha)}}, \quad \omega^2 \leqslant 1$$

and

$$\begin{aligned} 1-\sqrt{1-\omega^2} \leqslant MK^{\mathrm{par}} \leqslant 1+\sqrt{1-\omega^2} \quad &\textit{if } \omega^2>0,\\ \sqrt{kF_{k,\nu}(1-\alpha)} \leqslant M \quad &\textit{otherwise.}\end{aligned}$$

Then the condition (5) *is satisfied.*

Let us now assume that $K^{\mathrm{par}} > 0$. The bounds on $M = \sup_{\mathcal{O}} s^{-2}\theta'\theta$, given by Proposition 3.1, may be compared to the results of Propositions 2.1 and 2.3. We get

$$\frac{\sqrt{\delta_t}}{q_2} K^{\mathrm{par}} \leqslant K^{\mathrm{int}} \sqrt{kF_{k,\nu}(1-\alpha)} \leqslant \frac{\sqrt{\delta_t}}{q_1} K^{\mathrm{par}}$$

and

$$q_1 = \frac{1}{\omega^2}[1 - \sqrt{1-\omega^2}]^2 \leqslant q_b \leqslant \frac{1}{\omega^2}[1 + \sqrt{1-\omega^2}]^2 = q_2$$

where $q_b = \frac{c_b}{\sqrt{kF_{k,\nu}(1-\alpha)}}$. Similar inequalities are obtained for the parameters c_d^2, c_U^2 and c_m.

If $K^{\mathrm{par}} = 0$, the condition (5) becomes $\sqrt{kF_{k,\nu}(1-\alpha)} \leqslant M$, thus

$$K^{\mathrm{int}} \leqslant \frac{2\sqrt{\delta_t}}{kF_{k,\nu}(1-\alpha)}.$$

In this case $c_b = 0$, which means that the criterion with respect to the bias is satisfied for any value of the parameter c_b. Clearly, $q_1 \to 0$ and $q_2 \to \infty$ as $\omega^2 \to 0$.

If the estimation of only one linear functional $h'\theta$ is of interest, we use, for $\bar{\theta}$ in the domain $\mathcal{O}$ defined in Proposition 2.2, the $1-\alpha$ confidence interval

$$\mathcal{I} = \Big\{x\colon\ |\widehat{h'\theta}(Y,0) - x| \leqslant (1+q_b)t_\nu\Big(\frac{\alpha}{2}\Big)s\sqrt{h'h}\Big\}$$

where $t_\nu(\frac{\alpha}{2})$ is a critical value of Student's t-distribution with ν degrees of freedom and $q_b t_\nu(\frac{\alpha}{2}) = c_b$. The a priori interval, given by $\mathcal{O}$, will be, as in [3], defined by $I_h = \{x\colon\ |x| \leqslant \max_{\mathcal{O}} h'\theta\}$.

Lemma 3.1. $I_h = \{x\colon\ |x| \leqslant \min\{M_K, M_G\}\}$, *where*

$$M_K = \begin{cases} \sqrt{2c_b s\sqrt{h'h}h'[K^h]^+h} & \text{if } h \in \mathcal{M}(K^h) \\ \infty & \text{otherwise,} \end{cases}$$

$$M_G = \begin{cases} c\sqrt{h'hh'[(G^h)'G^h]^+h} & \text{if } h \in \mathcal{M}((G^h)'G^h) \\ \infty & \text{otherwise} \end{cases}$$

where $A^+ := \sum_i \mu_i^{-1} f_i f_i'$ *denotes the Moore-Penrose generalized inverse of the symmetric matrix* A *with the spectral decomposition* $A = \sum_i \mu_i f_i f_i'$.

Proof. It is clear that

$$\max_{\mathcal{O}} h'\theta = \min\Big\{\max_{\theta' K^h\theta \leqslant 2c_b s\sqrt{h'h}} h'\theta,\ \max_{\theta'(G^h)'G^h\theta \leqslant c^2 h'h} h'\theta\Big\}.$$

The proof of the statement for K^h can be found in [3], the proof for G^h is exactly the same. $\square$

Comparing intervals $\mathcal{I}$ and $\mathcal{I}_h$ we get the conditions

$$(1+q_b)t_\nu\Big(\frac{\alpha}{2}\Big) \leqslant \sqrt{\frac{2q_b t_\nu(\frac{\alpha}{2})h'[K^h]^+h}{s\sqrt{h'h}}} \tag{6}$$

and

$$(1+q_b)t_\nu\Big(\frac{\alpha}{2}\Big) \leqslant \frac{c\sqrt{h'[(G^h)'G^h]^+h}}{s}. \tag{7}$$

Proposition 3.2. *Let*

$$C_h^{\mathrm{par}} = \begin{cases} \frac{\sqrt{h'h}}{h'[K^h]^+h}s & \text{if } h \in \mathcal{M}(K^h) \\ 0 & \text{otherwise,} \end{cases}$$
$$C_h^{\mathrm{int}} = \begin{cases} \frac{s}{\sqrt{h'[(G^h)'G^h]^+h}} & \text{if } h \in \mathcal{M}((G^h)'G^h) \\ 0 & \text{otherwise.} \end{cases}$$

Let

$$C_h^{\mathrm{par}} = \frac{\omega_h^2}{2t_\nu(\frac{\alpha}{2})}, \qquad \omega_h^2 \leqslant 1$$

and

$$q_1 = \frac{1}{\omega_h^2}\Big[1-\sqrt{1-\omega_h^2}\Big]^2 \leqslant q_b \leqslant \frac{1}{\omega_h^2}\Big[1+\sqrt{1-\omega_h^2}\Big]^2 = q_2,$$
$$C_h^{\mathrm{int}} t_\nu\Big(\frac{\alpha}{2}\Big) \leqslant \frac{c}{1+q_b}.$$

Then the conditions (6) *and* (7) *are satisfied.*

From this, restrictions on all criterion parameters follow.

Remark 3.1. The values of C_h^{par} and C_h^{int} can be considered as some measures of nonlinearity of the model for the functional h. It is easy to see that in the case of $k=1$, we have $C_h^{\mathrm{par}} = K^{\mathrm{par}}$ and $C_h^{\mathrm{int}} = K^{\mathrm{int}}$ for all $h \in \mathbb{R}$. In general, these are different from K_h^{int} and K_h^{par}.

Remark 3.2. According to [3], the smallest ratio of the confidence region to the linearization domain is attained for c_b equal to $\sqrt{kF_{k,\nu}(\alpha)}$ or $c_b = t_\nu(\frac{\alpha}{2})$, and is equal to ω or ω_h, respectively. From this we see that we should require more than in Proposition 3.1 or 3.2, for example, $\omega \leqslant 0.5$.

The conditions stated in this section allow us to decide whether a given model can be linearized with a given set of criterion parameters, before the a priori information is specified. These conditions can be used as a characterization of models that allow the use of linear methods. If some values of the criterion parameters are given by the user to specify the tolerance of the error caused by the nonlinearity of the model, and if the conditions are fulfilled, it means that on some subset of the parameter space the model can be treated as linear with the errors within the tolerance, and the estimation still makes sense. Further, the intervals given by the restrictions on the criterion parameters indicate which criteria are influenced by the nonlinearity. If the linearization domain is sufficiently large, so that the required a priori information can be achieved, then the linear methods can be used.

4. An application

To illustrate our results, we will use the Michaelis-Menten model

$$Y_i = \frac{\beta_1 x_i}{\beta_2 + x_i} + \varepsilon_i, \;\; i = 1, \dots, n$$

where $\varepsilon \sim N_n[0, \sigma^2 I]$, with the design points x_i and realizations y_i given in [1]. This model will be linearized at the point $\beta_0 = (0.10579, 1.7007)$ (the least squares estimate of β). It should be said here that a quadratic approximation of the model at this point will be used; the assumption (2) of Section 1 might be violated if the domain $\mathcal{O}$ is too large. We will find the linearization domains for the estimators of the parameters β_1 and β_2 as well as for the whole parameter β. The domains will be given in the original coordinates. As in [3], the ellipses will be written in the form $\{p;\; a, b\}$, where p is the direction of the semiaxis with length a.

First, let $\theta = L(\beta - \beta_0)$ be the reparametrization from Section 2. We have

$$L = \begin{pmatrix} -0.9168 & 0.0353 \\ 0.2551 & 0 \end{pmatrix},$$

$d = 2$, $p = 1$ and $s = 0.004725$. Let us consider the functionals $h_1'\theta = \beta_1 - \beta_{01}$, $h_2'\theta = \beta_2 - \beta_{02}$, then $u_1 = h_1/\|h_1\| = (0, 1)'$ and $u_2 = h_2/\|h_2\| = (0.268, 0.963)'$ and

$$\begin{aligned}
\gamma_{1}^{\cdot\cdot} &= -\begin{pmatrix} 22.2 & 42.7244 \\ 42.7244 & 20.399 \end{pmatrix}, \\
\gamma_{2}^{\cdot\cdot} &= -\begin{pmatrix} 5.674 & 20.391 \\ 20.391 & 73.273 \end{pmatrix} = -\frac{k_{h_1}}{s} u_2 u_2', \\
\delta_{1}^{\cdot\cdot} &= -\begin{pmatrix} 0.735 & 2.641 \\ 2.641 & 9.49 \end{pmatrix} = -\frac{K^{\mathrm{int}}}{s} u_2 u_2'.
\end{aligned}$$

We see that no linear functional can be unbiasedly estimated. Moreover, the intrinsic curvature of the model is caused only by the parameter β_2. The curvature measures are given in the next table.

	K^{par} (C_h^{par})	K^{int} (C_h^{int})	ω^2	k_h
whole par	0.444	0.0483	2.5919	–
$h = h_1$	0	0	0	0.3730
$h = h_2$	0.2119	0.0483	0.9591	0.4186

Let us now find the linearization domains. Using the above expression for $\delta^{\cdot\cdot}$, we obtain a linearization domain for the adequacy criterion, given by

$$|\beta_2 - \beta_{02}| = |h_2'\theta| \leqslant \sqrt{\frac{\sqrt{\delta_t}}{K^{\text{int}}} s\|h_2\|} = 2.2725\delta_t^{\frac{1}{4}}$$

where the values of δ_t for $\alpha = 0.05$ and some values of $d\alpha$ are given in Table 1.

$d\alpha$	0.001	0.005	0.01	0.015	0.02	0.025	0.03	0.04
$\nu = 9$	0.0108	0.0539	0.1076	0.1610	0.2141	0.267	0.3197	0.4245

Table 1. Values of δ_t for $F_{1,\nu}$-distribution.

Let now $h = h_1$. The linearization domain from Proposition 2.2 has the form

$$\mathcal{E}_1^1 = \{\beta \mid |\beta_2 - \beta_{02}| \leqslant 1.1567\sqrt{c_b}\} \text{ and } \mathcal{E}_2^1 = \{\beta \mid |\beta_2 - \beta_{02}| \leqslant 10.7286c\}$$

where $\mathcal{E}_1^1$ and $\mathcal{E}_2^1$ are the sets given by the conditions (i) and (ii), respectively. If we put $\mathcal{O}_1 = \mathcal{E}_1^1$ then $c^2 = 0.0116c_b$ and we see that

$$c_d^2 = 0.7576c_b + 1.727\sqrt{c_b}, \quad c_U^2 = 0.7576c_b, \tag{8}$$

but, as the norm of $\gamma^{ij}(L\beta)_i(L\beta)_j$ is not bounded on this strip, the model is on $\mathcal{O}_1$ not linearizable with respect to the mean for any value of c_m.

If the estimator of $\beta_2 - \beta_{02} = h_2'\theta$ is considered, then the domains are

$$\mathcal{E}_1^2 = \{(0.0794, 0.9968)';\ 1.5393\sqrt{c_b}, 0.0686\sqrt{c_b}\},$$
$$\mathcal{E}_2^2 = \{\beta \mid |\beta_2 - \beta_{02}| \leqslant 10.3375c\}.$$

Compute $\sup_{\mathcal{E}_1^2} |\beta_2 - \beta_{02}| = M_K = 1.5287\sqrt{c_b}$, so that if we put $\mathcal{O}_2 = \mathcal{E}_1^2$, we have $c = 0.1484\sqrt{c_b}$. Further, $s^{-1}\|\delta^{ij}\theta_i\theta_j\| = s^{-2}\|h_2\|^{-2}K^{\text{int}}(h_2'\theta)^2$, so that

$$M_\delta = \|h_2\|^{-2}\frac{K^{\text{int}}}{s^2}M_K^2 = 0.4557c_b.$$

To get M_γ using an iteration algorithm, we compute $\sup\limits_{\theta} \frac{\|\gamma^{ij}\theta_i\theta_j\|}{\theta'|K^{h_2}|\theta}\|h_2\| = 8.1837$, thus $M_\gamma = 16.3674 c_b$.

We get

$$c_d^2 = 0.8592 c_b + 1.8300\sqrt{c_b}, \quad c_m = 7.5219 c_b^{\frac{3}{2}} + 2c_b \tag{9}$$

where the value of c_U^2 was omitted because it is always equal to the first summand of the expression for c_d^2.

Return now to the linearization domain for $h = h_1$. If the criterion with respect to the mean is to be satisfied, consider the expression

$$E_{\bar\theta}[u_1'(\hat\theta(Y,\theta) - \hat\theta(Y,0))] = \frac{1}{2}\Big(\frac{K^{\text{int}}}{s}\Big)^2 u_1' u_2 (u_2'\theta)(u_2'\bar\theta)^2 + \frac{k_{u_1}}{s} u_2'\theta\Big(u_2'\bar\theta + \frac{1}{2}\bar\theta' K^{u_2}\bar\theta\Big)$$

obtained from Lemma 2.1 and the expressions for $\gamma_2^{\cdot\cdot}$ and $\delta^{\cdot\cdot}$. We see that it is sufficient to use a domain where $\bar\theta' K^{u_2}\bar\theta$ is bounded, i.e. a linearization domain for the functional $h_2'\theta$ for some value c_{b2} of the criterion parameter c_b. Take the domain $\mathcal{E}_1^2$ with c_{b2} such that $\delta\mathcal{E}_1^2 \cap \delta\mathcal{E}_1^1 \neq \emptyset$, where $\delta\mathcal{E}$ denotes the boundary of the domain $\mathcal{E}$, i.e. $c_{b2} \geqslant 0.5725 c_b$. Then it is easy to see that if $\mathcal{O}_1 = \mathcal{E}_1^1 \cap \mathcal{E}_1^2$ then $\mathcal{O}_1$ is also a linearization domain for $h = h_1$ with the values of criterion parameters as in (8) and, moreover

$$c_m = 0.9129\sqrt{c_b}\, c_{b2} + 2c_b. \tag{10}$$

Finally, let us find the linearization domain for the whole vector β for some value of c_b. According to Proposition 2.3,

$$\begin{aligned}\mathcal{O}_3 &= \{\beta\colon\ \beta' L' L\beta \leqslant s^2 M^2 = 0.0001 c_b\}\\ &= \{(0.0357, 0.9994)';\ 1.0608\sqrt{c_b}, 0.0105\sqrt{c_b}\}\end{aligned}$$

and

$$c_d^2 = 0.8992 c_b + 1.8854\sqrt{c_b},\ c_m = 0.9546 c_b^{\frac{3}{2}} + 2c_b. \tag{11}$$

Let us now consider the linearization conditions. We see that if the domain $\mathcal{E}_1^1$ is used as the linearization domain, linearization is possible and no restrictions on the parameter occur. As for the domain $\mathcal{E}_1^1 \cap \mathcal{E}_1^2$, it will again be compared to the confidence interval for β_1 and a linearization condition follows, as in Section 3, with

$$\omega_{h_1}^2 = 2\frac{\sqrt{h_1' h_1}}{\frac{c_{b2}}{c_b} h_1' |K^{u_2}| h_1} s t_\nu\Big(\frac{\alpha}{2}\Big) = 0.3264$$

if $c_{b2} = 0.5725 c_b$. For the criterion parameters, we get the restrictions

$$0.2228 \leqslant c_b \leqslant 22.9759, \qquad 0.9840 \leqslant c_d^2 \leqslant 25.6852,$$
$$0.1688 \leqslant c_U^2 \leqslant 17.4070, \qquad 0.5006 \leqslant c_m \leqslant 103.5048.$$

We see that in this case, linearization is possible and the values of the criterion parameters are reasonable. Moreover, with respect to the relatively small standard deviation of the estimator, greater values of the criterion parameters might be tolerated, to obtain greater domains.

In the case of the functional $h = h_2$, we get

$$1.4890 \leqslant c_b \leqslant 3.4160, \qquad 3.5191 \leqslant c_d^2 \leqslant 6.3173,$$
$$1.2793 \leqslant c_U^2 \leqslant 2.9350, \qquad 16.6444 \leqslant c_m \leqslant 54.3206.$$

In this case, the conditions are much more restrictive and the domains obtained are very small. Moreover, the value of ω_{h_2} is still very large, indicating that the linearization domain is small compared to the confidence interval (see Remark 3.2).

As for the estimator of the whole vector β, we see that linearization is impossible.

Appendix. The computation of orthonormal bases

In this section we will indicate the procedure to find the orthonormal matrices $J = (J_1, J_2)$ and $\Omega = (\Omega_1, \Omega_2)$ and the arrays $\{\gamma_l^{ij}\}$ and $\{\delta_l^{ij}\}$.

Let X be a matrix with columns $x^{ij} = (F'W^{-1}F)^{-\frac{1}{2}} F'W^{-1}h^{ij}$, $i \leqslant j$, $j = 1, \dots, k$ and let $X = U \begin{pmatrix} T \\ 0 \end{pmatrix}$ be the QR decomposition of the matrix X, i.e. U is orthonormal and T upper triangular. Let $J = F(F'W^{-1}F)^{-\frac{1}{2}} U$ and $J \begin{pmatrix} T \\ 0 \end{pmatrix} = J_1 T$. Then it is easy to see that $J = (J_1, J_2)$ is the orthonormal matrix defined in Section 2. Further, if $R = L^{-1} = (J'W^{-1}F)^{-1}$, then $\gamma^{ij} = T^{ml} r_{mi} r_{lj}$, where $T^{ml} = T^{lm}$ is the column of T corresponding to the vector x^{ij}.

Similarly as above, let now Z be a matrix with columns $W^{-\frac{1}{2}} M h^{ij}$ and let $Z = Q \begin{pmatrix} S \\ 0 \end{pmatrix}$ be its QR decomposition. Let Q_1 be such that $Q \begin{pmatrix} S \\ 0 \end{pmatrix} = Q_1 S$. Then $\Omega_1 = W^{\frac{1}{2}} Q_1$ and $\delta^{ij} = S^{ml} r_{mi} r_{lj}$. The matrix Ω_2 can be found by completing the vectors (J_1, J_2, Ω_1) to an orthonormal basis of $\mathbb{R}^n$.

References

[1] *D. M. Bates, D. G. Watts*: Relative curvature measures of nonlinearity. J. Roy. Statist. Soc. B *42* (1980), 1–25.

[2] *P. R. Halmos*: Finite-dimensional Vector Spaces. Springer-Verlag, New York-Heidelberg-Berlin, 1974.

[3] *A. Jenčová*: A choice of criterion parameters in linearization of regression models. Acta Math. Univ. Comenianae, Vol LXIV, 2 (1995), 227–234.

[4] *L. Kubáček*: On a linearization of regression models. Appl. Math. *40* (1995), 61–78.

[5] *L. Kubáček*: Models with a low nonlinearity. Tatra Mountains Math. Publ. *7* (1996), 149–155.

[6] *A. Pázman*: Nonlinear Statistical Models. Kluwer Acad. Publishers, Dordrecht-Boston-London, and Ister Science Press, Bratislava, 1993.

Author's address: *Anna Jenčová*, Mathematical Institute of the Slovak Academy of Sciences, Štefánikova 49, 814 73 Bratislava, Slovakia, e-mail: `jenca@mat.savba.sk`.