Applications of Mathematics

Lubomír Kubáček
Linearized models with constraints of type I

Applications of Mathematics, Vol. 48 (2003), No. 2, 81–95

Persistent URL: `http://dml.cz/dmlcz/134520`

LINEARIZED MODELS WITH CONSTRAINTS OF TYPE I*

Lubomír Kubáček, Olomouc

(Received August 16, 2000)

Abstract. In nonlinear regression models with constraints a linearization of the model leads to a bias in estimators of parameters of the mean value of the observation vector. Some criteria how to recognize whether a linearization is possible is developed. In the case that they are not satisfied, it is necessary to decide whether some quadratic corrections can make the estimator better. The aim of the paper is to contribute to the solution of the problem.

MSC 2000: 62J05, 62F10

Keywords: nonlinear regression model with constraints, linearization, quadratization

1. Introduction

How to proceed in estimation of parameters in nonlinear models with constraints is a frequently occurring problem. Some problems without constraints are investigated in [2], [3], [5], [6], [7], [9], [12]; models with constraints are investigated in [8].

The aim of the paper is to find out some simple rules how to proceed in the situation when a nonlinear model is constrained by a nonlinear condition. The quadratic approximation only is considered.

* This work was supported by the grant No. 201/99/0327 of the Grant Agency of the Czech Republic and by the Council of the Czech Government J 14/98: 153 100011.

2. Notation and preliminaries

The notation

$$\mathbf{Y} \sim N_n(\mathbf{f}(\boldsymbol{\beta}), \boldsymbol{\Sigma}), \quad \boldsymbol{\beta} \in \mathcal{V} = \{\boldsymbol{\beta} \colon\ \mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}\} \tag{1}$$

means the following. The n-dimensional vector $\mathbf{Y}$ (observation vector) is random with normal distribution, its mean value $E(\mathbf{Y})$ is $\mathbf{f}(\boldsymbol{\beta})$ where $\mathbf{f}(\cdot)$ is an n-dimensional vector function of a known analytical form with continuous second derivatives, the k-dimensional parameter $\boldsymbol{\beta}$ is unknown, its value is constrained by the condition $\boldsymbol{\beta} \in \{\boldsymbol{\beta} \colon\ \mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}\}$, where $\mathbf{g}(\cdot)$ is a q-dimensional function with continuous second derivatives. The covariance matrix $\operatorname{var}(\mathbf{Y})$ of the vector $\mathbf{Y}$ is given and equal to $\boldsymbol{\Sigma}$.

This form of constraints, i.e. $\{\boldsymbol{\beta} \colon\ \mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}\}$ is called the constraints of type I. They frequently occur in practice. Another type of constraints (of type II) is

$$\left\{\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\kappa} \end{pmatrix} \colon\ \mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\kappa}) = \mathbf{0}\right\}.$$

The unknown parameter $\boldsymbol{\kappa}$ occurs in the constraints only. A model with such type of constraints is investigated in another paper.

A simple example of a model with constraints of type I can be

$$\begin{aligned} E(Y_i) &= \beta_1 \exp(-\beta_2 x_i), \ \ x_i < T \ \ (T \text{ is a given number}), \ \ i = 1, \dots, n_1, \\ E(Y_j) &= \beta_3 + \beta_4 x_j + \beta_5 x_j^2, \ \ x_j > T, \ \ j = 1, \dots, n_2, \\ \beta_1 \exp(-\beta_2 T) &= \beta_3 + \beta_4 T + \beta_5 T^2 \ \ (\text{continuity at the point } T), \\ -\beta_1 \beta_2 \exp(-\beta_2 T) &= \beta_4 + 2\beta_5 T \ \ (\text{continuity of the derivative at the point } T), \\ \operatorname{var}(Y_i) &= \operatorname{var}(Y_j) = \sigma^2, \ \ i = 1, \dots, n_1, \ \ j = 1, \dots, n_2, \\ \operatorname{cov}(Y_k, Y_l) &= 0, \ \ k \neq l. \end{aligned}$$

If in the estimation of the parameters $\beta_1, \dots, \beta_5$ linearization is used, it is not clear in advance whether this procedure does not lead to estimators with a non negligible bias, non tolerable size of the dispersion, etc. However, the use of the nonlinear least squares method is relatively complicated and thus it is useful to find some simple rules which enable us to decide whether the linearization is possible or it is better to use another procedure.

The linearized and quadratized approximation of the model with constraints of type I, i.e.

$$\mathbf{Y} = \mathbf{f}_0 + \mathbf{F}\delta\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \delta\boldsymbol{\beta} \in \{\delta\boldsymbol{\beta}\colon\ \mathbf{G}\delta\boldsymbol{\beta} = \mathbf{0}\} \tag{2}$$

and

$$\mathbf{Y} = \mathbf{f}_0 + \mathbf{F}\delta\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\kappa}(\delta\boldsymbol{\beta}), \quad \delta\boldsymbol{\beta} \in \left\{\delta\boldsymbol{\beta}\colon\ \mathbf{G}\delta\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\gamma}(\delta\boldsymbol{\beta}) = \mathbf{0}\right\}, \tag{3}$$

respectively, will be under consideration. Here $\boldsymbol{\varepsilon}$ is an error vector, $\boldsymbol{\beta}_0$ is an approximate value of the actual value $\boldsymbol{\beta}^*$ of the vector $\boldsymbol{\beta}$ and

$$\begin{aligned}
\mathbf{f}_0 &= \mathbf{f}(\boldsymbol{\beta}_0), \quad \mathbf{F} = \partial\mathbf{f}(\mathbf{u})/\partial\mathbf{u}'|_{u=\beta_0},\\
\boldsymbol{\kappa}(\delta\boldsymbol{\beta}) &= (\kappa_1(\delta\boldsymbol{\beta}), \ldots, \kappa_n(\delta\boldsymbol{\beta}))', \quad \kappa_i(\delta\boldsymbol{\beta}) = \delta\boldsymbol{\beta}'\mathbf{F}_i\delta\boldsymbol{\beta}, \quad i = 1, \ldots, n,\\
\mathbf{F}_i &= \partial^2 f_i(\mathbf{u})/\partial\mathbf{u}\partial\mathbf{u}'|_{u=\beta_0}, \quad i = 1, \ldots, n, \quad \mathbf{G} = \partial\mathbf{g}(\mathbf{u})/\partial\mathbf{u}'|_{u=\boldsymbol{\beta}_0},\\
\boldsymbol{\gamma}(\delta\boldsymbol{\beta}) &= (\gamma_1(\delta\boldsymbol{\beta}), \ldots, \gamma_q(\delta\boldsymbol{\beta}))', \quad \gamma_i(\delta\boldsymbol{\beta}) = \delta\boldsymbol{\beta}'\mathbf{G}_i\delta\boldsymbol{\beta}, \quad i = 1, \ldots, q,\\
\mathbf{G}_i &= \partial^2 g_i(\mathbf{u})/\partial\mathbf{u}\partial\mathbf{u}'|_{u=\beta_0}, \quad i = 1, \ldots, q.
\end{aligned}$$

In the following we assume

$$r(\mathbf{F}) = k, \quad r(\mathbf{G}) = q < k, \quad \boldsymbol{\Sigma} \text{ is positive definite.}$$

A solution of the equation $\mathbf{G}\delta\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\gamma}(\delta\boldsymbol{\beta}) = \mathbf{0}$ is not known in general. One quadratic approximation of $\delta\boldsymbol{\beta}$ can be expressed as

$$\delta\boldsymbol{\beta} = \mathbf{K}_G\delta\mathbf{s} - \frac{1}{2}\mathbf{G}^-\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s}),$$

where $\mathbf{K}_G$ is such a $k \times (k-q)$ matrix that

$$\mathcal{M}(\mathbf{K}_G) = \{\mathbf{K}_G\mathbf{u}\colon\ \mathbf{u} \in \mathbb{R}^{k-q}\} = \mathrm{Ker}(\mathbf{G}) = \{\mathbf{s}\colon\ \mathbf{G}\mathbf{s} = \mathbf{0}\}$$

and $\mathbf{G}^-$ is any generalized inverse of the matrix $\mathbf{G}$ (in detail cf. [11]). Then the model (2) can be rewritten with the new parameter $\delta\mathbf{s}$ as the model without constraints

$$\mathbf{Y} - \mathbf{f}_0 \sim N_n(\mathbf{F}\mathbf{K}_G\delta\mathbf{s}, \boldsymbol{\Sigma}) \tag{4}$$

and the BLUE (best linear unbiased estimator) of the parameter $\delta\mathbf{s}$ in this model is

$$\delta\hat{\mathbf{s}} = (\mathbf{K}_G'\mathbf{C}\mathbf{K}_G)^{-1}\mathbf{K}_G'\mathbf{F}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{f}_0). \tag{5}$$

The quadratic version of the model (3) can be rewritten with the new parameter $\delta\mathbf{s}$ as the model without constraints

$$\mathbf{Y} \sim N_n\Big(\mathbf{f}_0 + \mathbf{F}\mathbf{K}_G\delta\mathbf{s} + \frac{1}{2}[\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s}) - \mathbf{F}\mathbf{G}^-\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s})], \boldsymbol{\Sigma}\Big). \tag{6}$$

Let $\mathbf{M}^C_{\mathrm{Ker}(G)} = \mathbf{I} - \mathbf{P}^C_{\mathrm{Ker}(G)}$ where $\mathbf{P}^C_{\mathrm{Ker}(G)}$ is the projection matrix on $\mathrm{Ker}(\mathbf{G})$ in the norm given by the matrix $\mathbf{C} = \mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F}$ ($\|\mathbf{u}\| = \sqrt{\mathbf{u}'\mathbf{C}\mathbf{u}}$) and $\mathbf{P}_{G'} = \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}\mathbf{G}$, $\mathbf{M}_{G'} = \mathbf{I} - \mathbf{P}'_G$. The relation (5) can be rewritten with help of

$$\begin{aligned}
\mathbf{K}_G(\mathbf{K}'_G\mathbf{C}\mathbf{K}_G)^{-1}\mathbf{K}'_G\mathbf{F}'\boldsymbol{\Sigma}^{-1} &= \mathbf{M}_{G'}(\mathbf{M}_{G'}\mathbf{C}\mathbf{M}_{G'})^{+}\mathbf{M}_{G'}\mathbf{F}'\boldsymbol{\Sigma}^{-1} \\
&= [\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}\mathbf{C}^{-1}]\mathbf{F}'\boldsymbol{\Sigma}^{-1} \\
&= \mathbf{P}^C_{\mathrm{Ker}(G)}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}.
\end{aligned}$$

Thus

$$\delta\hat{\hat{\boldsymbol{\beta}}} = \widehat{\mathbf{K}_G\delta\mathbf{s}} = \mathbf{P}^C_{\mathrm{Ker}(G)}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{f}_0).$$

Here the notation $\widehat{\ }$ means the estimator in the model without constraints and $\hat{\hat{\ }}$ means the estimator in the model with constraints.

The linear estimator $\delta\hat{\hat{\boldsymbol{\beta}}}$, which is the BLUE in (2) but not in (1), is the estimator most frequently used in practice. Its quality depends on the nonlinearity of the model and on the choice of $\boldsymbol{\beta}_0$. If $\boldsymbol{\beta}_0 = \boldsymbol{\beta}^*$ (the actual value of the parameter $\boldsymbol{\beta}$), then $\delta\hat{\hat{\boldsymbol{\beta}}}$ is unbiased independently of the nonlinearity of the model. If $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}^*$, then the nonlinearity can intervene essentially. Thus the proper choice of $\boldsymbol{\beta}_0$ is important and therefore let us make a comment on its choice. In the subsequent steps the influence of the nonlinearity on the quality of the estimator $\delta\hat{\hat{\boldsymbol{\beta}}}$ will be investigated.

R e m a r k 2.1. The vector $\boldsymbol{\beta}_0$ is to be chosen in such a way that $\mathbf{g}(\boldsymbol{\beta}_0) = \mathbf{0}$ (in detail cf. [8]) and

$$\delta\hat{\hat{\boldsymbol{\beta}}}'\mathbf{C}\delta\hat{\hat{\boldsymbol{\beta}}} \leqslant \chi^2_{k-q}(0; 1-\alpha),$$

where

$$\begin{aligned}
&\delta\hat{\hat{\boldsymbol{\beta}}} = [\mathbf{I} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\delta\hat{\boldsymbol{\beta}} \quad \text{(the BLUE in the model (2))}, \\
&\delta\hat{\boldsymbol{\beta}} = \mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{f}_0) \quad \text{(the BLUE in the model (2) without} \\
&\qquad \text{linearized constraints)}.
\end{aligned}$$

This requirement is implied by the following fact.

As a consequence of Proposition 2.9.1 in [10] the $(1-\alpha)$-confidence region in the model (1) is

$$\{\boldsymbol{\beta}\colon\ [\mathbf{Y}-\mathbf{f}(\boldsymbol{\beta})]'\boldsymbol{\Sigma}^{-1}\mathbf{P}^{\Sigma^{-1}}_{F(\beta)K_{G(\beta)}}[\mathbf{Y}-\mathbf{f}(\boldsymbol{\beta})]\leqslant\chi^2_{k-q}(0;1-\alpha),\mathbf{g}(\boldsymbol{\beta})=\mathbf{0}\},$$

where

$$\mathbf{F}(\boldsymbol{\beta})=\frac{\partial\mathbf{f}(\mathbf{u})}{\partial\mathbf{u}'}\Big|_{u=\boldsymbol{\beta}},\quad \mathcal{M}(\mathbf{K}_{G(\beta)})=\mathrm{Ker}[\mathbf{G}(\boldsymbol{\beta})],\quad \mathbf{G}(\boldsymbol{\beta})=\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}'}\Big|_{u=\boldsymbol{\beta}}.$$

Since $\mathbf{F}\delta\hat{\hat{\boldsymbol{\beta}}}=\mathbf{P}^{\Sigma^{-1}}_{FK_G}(\mathbf{Y}-\mathbf{f}_0)$, we have

$$\delta\hat{\hat{\boldsymbol{\beta}}}'\mathbf{C}\delta\hat{\hat{\boldsymbol{\beta}}}=(\mathbf{Y}-\mathbf{f}_0)'\boldsymbol{\Sigma}^{-1}\mathbf{P}^{\Sigma^{-1}}_{FK_G}(\mathbf{Y}-\mathbf{f}_0)$$

and thus $\boldsymbol{\beta}_0$ is an element of the $(1-\alpha)$-confidence region of the model (1).

Now let the influence of the nonlinearity be commented.

Lemma 2.2. *The bias of the estimator $\delta\hat{\hat{\boldsymbol{\beta}}}$ is*

$$\begin{aligned}E(\delta\hat{\hat{\boldsymbol{\beta}}})-\delta\boldsymbol{\beta}&=\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\frac{1}{2}\boldsymbol{\gamma}(\delta\boldsymbol{\beta})\\&\quad+[\mathbf{I}-\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\frac{1}{2}\boldsymbol{\kappa}(\delta\boldsymbol{\beta}).\end{aligned}$$

Proof.

$$\begin{aligned}E(\delta\hat{\hat{\boldsymbol{\beta}}})&=[\mathbf{I}-\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}E(\mathbf{Y}-\mathbf{f}_0)\\&=[\mathbf{I}-\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\Big[\mathbf{F}\delta\boldsymbol{\beta}+\frac{1}{2}\boldsymbol{\kappa}(\delta\boldsymbol{\beta})\Big]\\&=\delta\boldsymbol{\beta}-\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}\delta\boldsymbol{\beta}\\&\quad+[\mathbf{I}-\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\frac{1}{2}\boldsymbol{\kappa}(\delta\boldsymbol{\beta})\\&=\delta\boldsymbol{\beta}+\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\frac{1}{2}\boldsymbol{\gamma}(\delta\boldsymbol{\beta})\\&\quad+[\mathbf{I}-\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\frac{1}{2}\boldsymbol{\kappa}(\delta\boldsymbol{\beta}).\end{aligned}$$

□

3. Simple quadratic estimator

The simplest way how to correct the bias of the estimator $\delta\hat{\hat{\boldsymbol{\beta}}}$ is to use the estimator

$$
\begin{aligned}
(7)\qquad \delta\tilde{\tilde{\boldsymbol{\beta}}} = \delta\hat{\hat{\boldsymbol{\beta}}} &- \frac{1}{2}[\mathbf{I} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\kappa}(\delta\hat{\hat{\boldsymbol{\beta}}}) \\
&- \frac{1}{2}\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\boldsymbol{\gamma}(\delta\hat{\hat{\boldsymbol{\beta}}}) \\
&+ \frac{1}{2}[\mathbf{I} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\mathrm{Tr}[\mathbf{F}_1\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})] \\ \vdots \\ \mathrm{Tr}[\mathbf{F}_n\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\end{pmatrix} \\
&+ \frac{1}{2}\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\begin{pmatrix}\mathrm{Tr}[\mathbf{G}_1\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})] \\ \vdots \\ \mathrm{Tr}[\mathbf{G}_q\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\end{pmatrix}.
\end{aligned}
$$

Since

$$
\begin{aligned}
(8)\quad \mathbf{G}\delta\tilde{\tilde{\boldsymbol{\beta}}} + \frac{1}{2}\boldsymbol{\gamma}(\delta\tilde{\tilde{\boldsymbol{\beta}}}) &= -\frac{1}{2}\boldsymbol{\gamma}(\delta\hat{\hat{\boldsymbol{\beta}}}) + \frac{1}{2}\begin{pmatrix}\mathrm{Tr}[\mathbf{G}_1\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})] \\ \vdots \\ \mathrm{Tr}[\mathbf{G}_q\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\end{pmatrix} + \frac{1}{2}\boldsymbol{\gamma}(\delta\tilde{\tilde{\boldsymbol{\beta}}}) \\
&= \frac{1}{2}\begin{pmatrix}\mathrm{Tr}[\mathbf{G}_1\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})] \\ \vdots \\ \mathrm{Tr}[\mathbf{G}_q\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\end{pmatrix} + \text{terms of the 3rd and higher orders,}
\end{aligned}
$$

the estimator $\delta\tilde{\tilde{\boldsymbol{\beta}}}$ does not satisfy the constraints as far as the second order terms are concerned. (It is to be said that the mean value of the estimator $\delta\tilde{\tilde{\boldsymbol{\beta}}}$ from (7) satisfies the constraints $\mathbf{G}\delta\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\gamma}(\delta\boldsymbol{\beta}) = \mathbf{0}$.)

There are two possibilities. Either the term

$$\frac{1}{2}(\mathrm{Tr}[\mathbf{G}_1\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})], \ldots, \mathrm{Tr}[\mathbf{G}_q\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})])'$$

in the estimator $\delta\tilde{\tilde{\boldsymbol{\beta}}}$ will cancel, or the constraints will not be satisfied, however the bias will be of the third and higher orders only.

If a function $h(\cdot)$, i.e. $h(\boldsymbol{\beta}) = \mathbf{h}'\boldsymbol{\beta}$, is under consideration, the following notation will be used:

$$
\begin{aligned}
\mathbf{F}_{h(\cdot)} &= \frac{1}{2}\sum_{i=1}^{n}\{\mathbf{h}'\mathbf{P}^{C}_{\mathrm{Ker}}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\}_i\mathbf{F}_i, \quad \mathbf{h} \in \mathbb{R}^k,\\
\mathbf{G}_{h(\cdot)} &= \frac{1}{2}\sum_{i=1}^{q}\{\mathbf{h}'\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\}_i\mathbf{G}_i, \quad \mathbf{h} \in \mathbb{R}^q,\\
\mathbf{b} &= E(\delta\hat{\hat{\boldsymbol{\beta}}}) - \delta\boldsymbol{\beta}\\
&= \frac{1}{2}\mathbf{P}_{\mathrm{Ker}(G)}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s})\\
&\quad + \frac{1}{2}\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s})\\
&= (\delta\boldsymbol{\beta}'(\mathbf{F}_{e_1(\cdot)} + \mathbf{G}_{e_1(\cdot)})\delta\boldsymbol{\beta}, \ldots, \delta\boldsymbol{\beta}'(\mathbf{F}_{e_k(\cdot)} + \mathbf{G}_{e_k(\cdot)})\delta\boldsymbol{\beta})'.
\end{aligned}
$$

Here $\mathbf{e}_i \in \mathbb{R}^k$ and $\{\mathbf{e}_i\}_j = \delta_{i,j}$ (the Kronecker delta). Thus $e_i(\boldsymbol{\beta}) = \mathbf{e}_i'\boldsymbol{\beta} = \beta_i$, $i = 1, \ldots, k$.

From the viewpoint of practice it seems that it is more natural to prefer the constraints to the bias. Thus in the sequel we will analyze the estimator

$$
\begin{aligned}
(9)\quad \mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}} = \mathbf{h}'\Big\{&\delta\hat{\hat{\boldsymbol{\beta}}} - \frac{1}{2}[\mathbf{I} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\kappa}(\delta\hat{\hat{\boldsymbol{\beta}}})\\
&- \frac{1}{2}\mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\boldsymbol{\gamma}(\delta\hat{\hat{\boldsymbol{\beta}}})\\
&+ \frac{1}{2}[\mathbf{I} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\mathrm{Tr}[\mathbf{F}_1\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\\ \vdots\\ \mathrm{Tr}[\mathbf{F}_n\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\end{pmatrix}\Big\}\\
= \mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}} &- \delta\hat{\hat{\boldsymbol{\beta}}}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\hat{\hat{\boldsymbol{\beta}}} + \mathrm{Tr}[\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{F}_{h(\cdot)}].
\end{aligned}
$$

If the ratio (cf. the relationship (8))

$$
(10)\qquad \mathrm{Tr}[\mathbf{G}_{h(\cdot)}\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]/\sqrt{\mathbf{h}'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}\mathbf{C}^{-1}]\mathbf{h}}
$$

is sufficiently small by the opinion of the user, then it is possible to tolerate the bias of the second order, i.e. to use the estimator $\delta\overline{\overline{\boldsymbol{\beta}}}$.

Under the given notation we can write (cf. (9))

$$
\begin{aligned}
(11)\qquad E(\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}) = \mathbf{h}'\delta\boldsymbol{\beta} &- \mathrm{Tr}[\mathbf{G}_{h(\cdot)}\,\mathrm{var}(\delta\hat{\hat{\boldsymbol{\beta}}})] - 2\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta}\\
&- \mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b}.
\end{aligned}
$$

Thus the bias of the linearized estimator of a linear function $h(\cdot)$ and of the quadratic estimator, respectively, is

$$
\begin{aligned}
|E(\mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}}) - \mathbf{h}'\delta\boldsymbol{\beta}| &= |\delta\boldsymbol{\beta}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta}|,\\
|E(\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}) - \mathbf{h}'\delta\boldsymbol{\beta}| &= |\operatorname{Tr}[\mathbf{G}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})] + 2\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta}\\
&\qquad + \mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b}|.
\end{aligned}
$$

In [8] it is described how to recognize whether the bias of the linearized estimator $\mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}}$ is smaller than $c\sqrt{\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h}}$, where c is a constant chosen by users.

The following statement is proved there (Theorem 2.2.8 in [8]). If

$$
\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} \leqslant \frac{2c}{C^{(\mathrm{par})}_{I,\delta\beta}}(\boldsymbol{\beta}_0), \quad \mathbf{G}\delta\boldsymbol{\beta} = \mathbf{0},
$$

then

$$
\forall\{\mathbf{h} \in \mathbb{R}^k\}\big|\mathbf{h}'[E(\delta\hat{\hat{\boldsymbol{\beta}}}) - \delta\beta]\big| \leqslant c\sqrt{\mathbf{h}'\mathbf{C}^{-1}\mathbf{h}}.
$$

The quantity $C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)$ will be defined in below in (14).

It is necessary to know with practical certainty that a shift $\delta\beta$ is in a linearization region for the bias, i.e. the confidence region of $\delta\beta$ for a sufficiently high confidence level $1-\alpha$ is included into the linearization region. If this condition is not satisfied, then it is necessary to compare the bias of $\delta\hat{\hat{\boldsymbol{\beta}}}$ and $\delta\overline{\overline{\boldsymbol{\beta}}}$. However, the bias and also the variance of the estimator depend on the possible shift $\delta\boldsymbol{\beta} = \boldsymbol{\beta}^* - \boldsymbol{\beta}_0$ ($\boldsymbol{\beta}^*$ means the actual value of the parameter $\boldsymbol{\beta}$). If the dimension k is large, then an investigation of the bias and the variance in different directions of the shift $\delta\boldsymbol{\beta}$ is tedious. Thus some knowledge of upper bounds independent of directions can be useful in practice.

4. Upper bounds for the bias and variance of the linear and quadratic estimators

The MSE (mean square error) of the estimator $\mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}}$ is

$$
\operatorname{MSE}(\mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}}) = \mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h} + [\delta\boldsymbol{\beta}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta}]^2. \tag{12}
$$

The MSE (mean square error) of the estimator $\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}$ is

$$
\begin{aligned}
\operatorname{MSE}(\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}) = \mathbf{h}'\operatorname{var}(\delta\overline{\overline{\boldsymbol{\beta}}})\mathbf{h} + \{&\operatorname{Tr}[\mathbf{G}_{h(\cdot)}\operatorname{var}(\hat{\hat{\boldsymbol{\beta}}})] + 2\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta}\\
&+ \mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b}\}^2.
\end{aligned} \tag{13}
$$

We have

$$
\begin{aligned}
\mathbf{h}' \operatorname{var}(\delta\overline{\overline{\boldsymbol{\beta}}})\mathbf{h} &= \mathbf{h}' \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h} + 2\operatorname{Tr}\{[(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)}) \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]^2\} \\
&\quad - 4\mathbf{h}' \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta} - 4\mathbf{h}' \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b} \\
&\quad + 4\delta\boldsymbol{\beta}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)}) \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\beta \\
&\quad + 8\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)}) \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\beta \\
&\quad + 4\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)}) \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b}, \\
\operatorname{bias}^2(\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}) &= \{\operatorname{Tr}[\mathbf{G}_{h(\cdot)} \operatorname{var}(\hat{\hat{\boldsymbol{\beta}}})] + 2\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\beta + \mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b}\}^2 \\
&= 4[\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\beta]^2 \\
&\quad + 4\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta}\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b} \\
&\quad + [\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b}]^2 + 4\operatorname{Tr}[\mathbf{G}_{h(\cdot)} \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta} \\
&\quad + 2\operatorname{Tr}[\mathbf{G}_{h(\cdot)} \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\mathbf{b}'(\mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)})\mathbf{b} + \{\operatorname{Tr}[\mathbf{G}_{h(\cdot)} \operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\}^2.
\end{aligned}
$$

In order to find an upper bound for the individual terms of the variance and the bias, it is necessary to go back to the definition of the parametric measure of nonlinearity [1] and its generalization.

The Bates and Watts parametric measure of nonlinearity [1] in the model (6) with respect to the new parameter $\delta\mathbf{s}$ is

$$
K^{(\mathrm{par})}_{I,\delta s}(\boldsymbol{\beta}_0) = \sup\left\{ \frac{\sqrt{A}}{\delta\mathbf{s}'\mathbf{K}_G'\mathbf{C}\mathbf{K}_G\delta\mathbf{s}} :\ \delta\mathbf{s} \in \mathbb{R}^{k-q} \right\},
$$

where

$$
\begin{aligned}
A &= [\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s}) - \mathbf{F}\mathbf{G}^-\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s})]'\boldsymbol{\Sigma}^{-1}\mathbf{P}^{\Sigma^{-1}}_{FK_G}[\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s}) - \mathbf{F}\mathbf{G}^-\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s})] \\
&= [\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s})]'\boldsymbol{\Sigma}^{-1}\mathbf{P}^{\Sigma^{-1}}_{FK_G}[\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s})],
\end{aligned}
$$

since

$$
\begin{aligned}
&\mathbf{P}^{\Sigma^{-1}}_{FK_G}\mathbf{F}\mathbf{G}^-\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s}) \\
&\quad = \mathbf{F}[\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}\mathbf{C}^{-1}]\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F}\mathbf{G}^-\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s}) \\
&\quad = \mathbf{F}[\mathbf{G}^- - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')]\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s}) = \mathbf{0}
\end{aligned}
$$

when a choice $\mathbf{G}^- = \mathbf{G}^-_{m(\mathbf{C})} = \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}$ is realized. The quantity gives information on the bias of the linear estimator of $\delta\mathbf{s}$, but not on the bias of the linear estimator of the parameter $\delta\boldsymbol{\beta}$. Thus (in detail cf. [8]) a generalization of the quantity $K_{I,\delta s}^{(\mathrm{par})}(\boldsymbol{\beta}_0)$ given by the definition

$$C_{I,\delta\beta}^{(\mathrm{par})}(\boldsymbol{\beta}_0) = \sup\left\{\frac{\sqrt{B}}{\delta\mathbf{s}'\mathbf{K}_G'\mathbf{C}\mathbf{K}_G\delta\mathbf{s}} : \delta\mathbf{s} \in \mathbb{R}^{k-q}\right\}, \tag{14}$$

where

$$B = A + \boldsymbol{\gamma}'(\mathbf{K}_G\delta\mathbf{s})(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s}),$$

seems to be more suitable.

The second term in the last equality occurs in an intrinsic measure of curvature of the constraints (in detail cf. Definition 2.2.7 and Theorem 2.2.8 in [8]).

Lemma 4.1. *The inequality*

$$\mathbf{b}'\mathbf{C}\mathbf{b} \leqslant \frac{1}{4}[C_{I,\delta\beta}^{(\mathrm{par})}]^2(\delta\mathbf{s}'\mathbf{K}_G'\mathbf{C}\mathbf{K}_G\delta\mathbf{s})^2 + \text{ terms of the 5th order}$$

is valid.

Proof. When the terms of the fifth order are neglected, then

$$\begin{aligned}
\mathbf{b}'\mathbf{C}\mathbf{b} &= \frac{1}{4}[\mathbf{P}_{\mathrm{Ker}(G)}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s}) \\
&\quad + \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s})]'\mathbf{C} \\
&\quad \times [\mathbf{P}_{\mathrm{Ker}(G)}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s}) \\
&\quad + \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s})] \\
&= \frac{1}{4}(\boldsymbol{\kappa}'(\mathbf{K}_G\delta\mathbf{s})\boldsymbol{\Sigma}^{-1}\mathbf{P}_{FK+G}^{\Sigma^{-1}}\boldsymbol{\kappa}(\mathbf{K}_G\delta\mathbf{s}) \\
&\quad + \boldsymbol{\gamma}'(\mathbf{K}_G\delta\mathbf{s})(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\boldsymbol{\gamma}(\mathbf{K}_G\delta\mathbf{s}))
\end{aligned}$$

and thus

$$\frac{\sqrt{4\mathbf{b}'\mathbf{C}\mathbf{b}}}{\delta\mathbf{s}'\mathbf{K}_G'\mathbf{C}\mathbf{K}_G\delta\mathbf{s}} \leqslant C_{I,\delta\beta}^{(\mathrm{par})},$$

which implies the proof. $\square$

Theorem 4.2. *For the sake of simplicity, let* $\mathbf{A}_{h(\cdot)} = \mathbf{F}_{h(\cdot)} + \mathbf{G}_{h(\cdot)}$. *We have (in the term* $\mathbf{b}'\mathbf{C}\mathbf{b}$ *the terms of the fifth order are neglected)*

$$\begin{aligned}
\delta\boldsymbol{\beta}'\mathbf{A}_{h(\cdot)}\delta\boldsymbol{\beta} &\leqslant \sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}\,\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|-4\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{A}_{h(\cdot)}\delta\boldsymbol{\beta}| &\leqslant 4\sigma_h\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}},\\
|4\delta\boldsymbol{\beta}'\mathbf{A}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{A}_{h(\cdot)}\delta\boldsymbol{\beta}| &\leqslant 4(k-q)^{1/4}\{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^8]\}^{1/4}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|-4\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{A}_{h(\cdot)}\mathbf{b}| &\leqslant 2\sigma_h\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|8\mathbf{b}'\mathbf{A}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{A}_{h(\cdot)}\delta\boldsymbol{\beta}| &\leqslant 4(k-q)^{1/4}\{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^8]\}^{1/4}\\
&\quad\times C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2},\\
|4\operatorname{Tr}[\mathbf{G}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\mathbf{b}'\mathbf{A}_{h(\cdot)}\delta\boldsymbol{\beta}| &\leqslant 2\operatorname{Tr}[\mathbf{G}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}\\
&\quad\times C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2},\\
|4\mathbf{b}'\mathbf{A}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{A}_{h(\cdot)}\mathbf{b}| &\leqslant (k-q)^{1/4}\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^4]}\\
&\quad\times [C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2,\\
|2\operatorname{Tr}[\mathbf{G}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\mathbf{b}'\mathbf{A}_{h(\cdot)}\mathbf{b}| &\leqslant \frac{1}{2}\operatorname{Tr}[\mathbf{G}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}\\
&\quad\times [C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2,\\
|4(\mathbf{b}'\mathbf{A}_{h(\cdot)}\delta\boldsymbol{\beta})^2| &\leqslant \operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2][C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}\mathbf{C}\delta\boldsymbol{\beta})^3,\\
|4\mathbf{b}'\mathbf{A}_{h(\cdot)}\delta\boldsymbol{\beta}\mathbf{b}'\mathbf{A}_{h(\cdot)}\mathbf{b}| &\leqslant \frac{1}{2}\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2][C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)]^3(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{7/2},\\
(\mathbf{b}'\mathbf{A}_{h(\cdot)}\mathbf{b})^2 &\leqslant \frac{1}{16}\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2][C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol{\beta}_0)]^4(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^4.
\end{aligned}$$

Here $\sigma_h = \sqrt{\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h}}$.

P r o o f. For the sake of simplicity only some of the given inequalities are proved. The other can be proved analogously.

$$\begin{aligned}
\delta\boldsymbol{\beta}'\mathbf{A}_{h(\cdot)}\delta\boldsymbol{\beta} &= \delta\boldsymbol{\beta}'\mathbf{C}^{1/2}\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\\
&= \operatorname{Tr}[\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\delta\boldsymbol{\beta}'\mathbf{C}^{1/2}]\\
&\leqslant \sqrt{\operatorname{Tr}[\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}]}\\
&\quad\times\sqrt{\operatorname{Tr}[\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\delta\boldsymbol{\beta}'\mathbf{C}^{1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\delta\boldsymbol{\beta}'\mathbf{C}^{1/2}]}\\
&= \sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}\,\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},
\end{aligned}$$

$$
\begin{aligned}
|-4\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{A}_{h(\cdot)}\delta\boldsymbol\beta| &= 4|\operatorname{Tr}[\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{C}^{1/2}\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol\beta| \\
&= 4|\operatorname{Tr}[\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol\beta\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{C}^{1/2}] \\
&\leqslant 4\sigma_h\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}\sqrt{\delta\boldsymbol\beta'\mathbf{C}\delta\boldsymbol\beta}, \\
&\quad \text{since}\ \ \operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{C}\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}}) = \operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}}), \\
|-4\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{A}_{h(\cdot)}\mathbf{b}| &= 4|\operatorname{Tr}[\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\mathbf{b}\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{C}^{1/2}]| \\
&\leqslant 4\sigma_h\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}\sqrt{\mathbf{b}'\mathbf{C}\mathbf{b}} \\
&\leqslant 2\sigma_h\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}C^{(\mathrm{par})}_{I,\delta\beta}(\boldsymbol\beta_0)\delta\boldsymbol\beta'\mathbf{C}\delta\boldsymbol\beta \\
&\quad \text{by virtue of Lemma 4.1}, \\
4\delta\boldsymbol\beta'\mathbf{A}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{A}_{h(\cdot)}\delta\boldsymbol\beta &= 4\operatorname{Tr}(\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol\beta\delta\boldsymbol\beta'\mathbf{C}^{1/2}) \\
&\leqslant 4\sqrt{\operatorname{Tr}[(\mathbf{A}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1})^2]}\delta\boldsymbol\beta'\mathbf{C}\delta\boldsymbol\beta \\
&= 4\{\operatorname{Tr}[(\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{C}^{1/2} \\
&\quad \times\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2})^2]\}^{1/2}\delta\boldsymbol\beta'\mathbf{C}\delta\boldsymbol\beta \\
&\leqslant 4\{\operatorname{Tr}[(\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2})^8]\}^{1/4} \\
&\quad \times\{\operatorname{Tr}[(\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{C}]\}^{1/4}\delta\boldsymbol\beta'\mathbf{C}\delta\boldsymbol\beta \\
&= 4(k-q)^{1/4}\{\operatorname{Tr}[(\mathbf{C}^{-1/2}\mathbf{A}_{h(\cdot)}\mathbf{C}^{-1/2})^8]\}^{1/4} \\
&\quad \times\delta\boldsymbol\beta'\mathbf{C}\delta\boldsymbol\beta\quad\text{since}\ \ \operatorname{Tr}[\operatorname{var}(\delta\hat{\hat{\boldsymbol\beta}})\mathbf{C}] = k-q.
\end{aligned}
$$

We can complete the proof in an analogous way. □

In practice the terms of orders higher than fourth are of small importance. If they are inadmissibly large, then the quadratic theory cannot be used and we have to study the problem from the viewpoint of the general theory given in [10].

Now some simple rules for a decision on the linearization will be given in the next section.

5. Conclusion

In the preceding sections linear and quadratic estimators have been studied. In the expressions for the variance and the bias the terms with higher powers (in $\delta\boldsymbol\beta$) are given and this seems to be superfluous for the linear and quadratic estimators. Several presented terms are of higher powers than the neglected terms in the Taylor series of the model. A reason for it is an endeavour to investigate in the first step the behaviour of the quadratic estimators only, since the estimators of orders higher than two have a small chance to be used in practice because of the complexity of their statistical properties. Even in the case of quadratic estimators the power of a single term in the MSE can attain the value eight and of course in practice it is of no sense. Nevertheless, they are presented to ensure the completeness of the theory of quadratic estimators. Results given in the paper form only a small part of knowledge necessary for an efficient application of quadratic estimators. Further research is necessary.

A procedure, based on partial results obtained in the paper, how to deal with models with constraints of type I can be described as follows.

(i) The value $\boldsymbol{\beta}_0$ is chosen in the sense of Remark 2.1.

(ii) The value of the quantity $C^{(\mathrm{par})}_{I,\delta\beta}$ is determined at the point $\boldsymbol{\beta}_0$ and the region $\mathcal{L}_b = \left\{\delta\boldsymbol{\beta}\colon\ \delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} \leqslant \frac{2c}{C^{(\mathrm{par})}_{I,\delta\beta}}\right\}$ is determined. If $\mathcal{L}_b \supset \mathcal{E}$, where

$$\mathcal{E} = \{\delta\boldsymbol{\beta}\colon\ \delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} \leqslant \chi^2_{k-q}(0; 1-\alpha)\}$$

for a sufficiently small value α, then the estimator $\delta\hat{\hat{\boldsymbol{\beta}}}$ can be used.

(iii) If $\mathcal{L}_b \not\supset \mathcal{E}$, then it is useful to investige a possibility to make variances of the observation vector $\mathbf{Y}$ smaller and thus attain the inclusion $\mathcal{E} \subset \mathcal{L}_b$. If this way cannot be realized, then it is reasonable to construct the estimator $\delta\overline{\overline{\boldsymbol{\beta}}}$ and to compare the values UBMSE($\mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}}$) and UBMSE($\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}$) for different important functions $h(\cdot)$. If, e.g., UBMSE($\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}$) $\leqslant$UBMSE($\mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}}$) and the value UBMSE($\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}$) is acceptable, the estimator $\delta\overline{\overline{\boldsymbol{\beta}}}$ can be used.

(iv) For a more detailed investigation of the situation it is useful to use the relationships (12) and (13). This procedure is time consuming and it will be used probably in special cases only.

Some more detailed comments follow.

If we want to compare the estimators $\delta\hat{\hat{\boldsymbol{\beta}}}$ and $\delta\overline{\overline{\boldsymbol{\beta}}}$, we have several possibilities.

If we are interested in the bias only, then we can compare the bias in different directions of the shift $\delta\boldsymbol{\beta}$ or we can compare the upper bounds for $|\mathbf{h}'[E(\delta\hat{\hat{\boldsymbol{\beta}}}) - \delta\boldsymbol{\beta}|$ and $|\mathbf{h}'[E(\delta\overline{\overline{\boldsymbol{\beta}}}) - \delta\boldsymbol{\beta}|$ with help of Theorem 4.2. However, in models with a weak nonlinearity the bias of the quadratic estimator is essentially smaller than the bias of the linear estimator.

If we are intersted in the MSE, then we find out such a shift $\delta\boldsymbol{\beta}$ in a given direction which gives the equality MSE($\mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}}$) =MSE($\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}$). For a larger shift the linear estimator cannot be used. If this distance is out of the confidence ellipsoid we prefer the linear estimator to the quadratic one. However, for a large dimension k of the parameter $\boldsymbol{\beta}$ this procedure is not suitable.

We find out the Mahalanobis distance $d = \sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}}$ which gives the equality UBMSE($\mathbf{h}'\delta\hat{\hat{\boldsymbol{\beta}}}$) =UBMSE($\mathbf{h}'\delta\overline{\overline{\boldsymbol{\beta}}}$) (UBMSE= upper bound of the mean square error).

For the first orientation the following rule can be also used. In a small neighbourhood of the point $\boldsymbol{\beta}_0$ the most dangerous term is $-4\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)}+\mathbf{G}_{h(\cdot)})\delta\boldsymbol{\beta}$, i.e. $\delta\boldsymbol{\beta}$ is proportional to the vector $(\mathbf{F}_{h(\cdot)}+\mathbf{G}_{h(\cdot)})\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h}$. Let us find such a number t that

$$\begin{aligned}\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h} &+ t^4[\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)}+\mathbf{G}_{h(\cdot)})^3\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h}]^2\\ &= \mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h} + 2\operatorname{Tr}\{[(\mathbf{F}_{h(\cdot)}+\mathbf{G}_{h(\cdot)})\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]^2\}\\ &\quad + \{\operatorname{Tr}[\mathbf{G}_{h(\cdot)}\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]\}^2\\ &\quad + 4t\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)}+\mathbf{G}_{h(\cdot)})^2\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h}.\end{aligned}$$

The smallest positive solution t_{crit} of this equation gives the vector $t_{\text{crit}}(\mathbf{F}_{h(\cdot)}+\mathbf{G}_{h(\cdot)})\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h}$ with the Mahalanobis distance equal to

$$d^2_{\text{crit}} = t^2_{\text{crit}}\mathbf{h}'\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})(\mathbf{F}_{h(\cdot)}+\mathbf{G}_{h(\cdot)})\mathbf{C}(\mathbf{F}_{h(\cdot)}+\mathbf{G}_{h(\cdot)})\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})\mathbf{h}.$$

If $d^2_{\text{crit}} > \chi^2_{k-q}(0;1-\alpha)$ for a sufficiently small α, the linear estimator is to be preferred.

In what follows several remarks to the utilization of the term $\mathbf{b}'\mathbf{C}\mathbf{b}$ are necessary.

If no constraints are under consideration, then the term $\mathbf{b}'\mathbf{C}\mathbf{b}$ is obviously suitable for a decision whether the bias is or is not acceptable. If constraints are under consideration, then the covariance matrix $\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}}) = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}\mathbf{C}^{-1}$ is singular and thus its inverse does not exist. The term $\mathbf{b}'[\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]^-\mathbf{b}$ is not determined unambiguously, since the relation $\mathbf{b}\in\mathcal{M}(\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}}))$ need not be valid. However, one version of $[\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]^-$ is $\mathbf{C}$ which is positive definite and thus the term $\mathbf{b}'\mathbf{C}\mathbf{b}$ for $\mathbf{b}\in\mathcal{M}(\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}}))$ is equal to $\mathbf{b}'[\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})]^-\mathbf{b}$. If

$$\mathbf{b}\notin\mathcal{M}(\operatorname{var}(\delta\hat{\hat{\boldsymbol{\beta}}})),$$

then another reason exists for a utilization of the term $\mathbf{b}'\mathbf{C}\mathbf{b}$. Let us investigate the projection $\mathcal{PE}$ of the ellipsoid $\mathcal{E} = \{\mathbf{x}\colon\ \mathbf{x}'\mathbf{C}\mathbf{x}\leqslant t^2\}$ in the Mahalanobis norm on $\operatorname{Ker}(\mathbf{G})$. We have

$$\begin{aligned}\mathcal{PE} &= \{\mathbf{P}^C_{\operatorname{Ker}(G)}\mathbf{x}\colon\ \mathbf{x}'(\mathbf{P}^C_{\operatorname{Ker}(G)})'\mathbf{C}\mathbf{P}^C_{\operatorname{Ker}(G)}\mathbf{x} + \mathbf{x}'(\mathbf{M}^C_{\operatorname{Ker}(G)})'\mathbf{C}\mathbf{M}^C_{\operatorname{Ker}(G)}\mathbf{x}\leqslant t^2\}\\ &= \{\mathbf{P}^C_{\operatorname{Ker}(G)}\mathbf{x}\colon\ \mathbf{x}'[\mathbf{C}-\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{x}\leqslant t^2\},\end{aligned}$$

since $(\mathbf{M}^C_{\operatorname{Ker}(G)})'\mathbf{C}\mathbf{P}^C_{\operatorname{Ker}(G)} = \mathbf{0}$.

If $\mathbf{x} \in \operatorname{Ker}(\mathbf{G})$, then

$$\mathbf{x}'[\mathbf{C} - \mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}]\mathbf{x} = \mathbf{x}'\mathbf{C}\mathbf{x},$$

and thus the projection is the section of the ellipsoid $\mathcal{E}$ by the subspace $\operatorname{Ker}(\mathbf{G})$. Further,

$$\begin{aligned}
&[\delta\hat{\hat{\boldsymbol{\beta}}} - E(\delta\hat{\hat{\boldsymbol{\beta}}})]'[\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{G}'(\mathbf{G}\mathbf{C}^{-1}\mathbf{G}')^{-1}\mathbf{G}\mathbf{C}^{-1}]^{-}[\delta\hat{\hat{\boldsymbol{\beta}}} - E(\delta\hat{\hat{\boldsymbol{\beta}}})] \\
&\quad = [\delta\hat{\hat{\boldsymbol{\beta}}} - E(\delta\hat{\hat{\boldsymbol{\beta}}})]'\mathbf{C}[\delta\hat{\hat{\boldsymbol{\beta}}} - E(\delta\hat{\hat{\boldsymbol{\beta}}})] \sim \chi^2_{k-q}
\end{aligned}$$

and thus the term $\mathbf{b}'\mathbf{C}\mathbf{b}$ is suitable for a decision whether the bias $\mathbf{b}$ is or is not acceptable.

References

[1] *D. M. Bates, D. G. Watts*: Relative curvature measures of nonlinearity. J. Roy. Statist. Soc. Ser. B *42* (1980), 1–25.
[2] *A. Jenčová*: A comparison of linearization and quadratization domains. Appl. Math. *42* (1997), 279–291.
[3] *L. Kubáček*: On a linearization of regression models. Appl. Math. *40* (1995), 61–78.
[4] *L. Kubáček, L. Kubáčková, J. Volaufová*: Statistical Models with Linear Structures. Veda, Bratislava, 1995.
[5] *L. Kubáček*: Models with a low nonlinearity. Tatra Mountains Math. Publ. *7* (1996), 149–155.
[6] *L. Kubáček*: Quadratic regression models. Math. Slovaca *46* (1996), 111–126.
[7] *L. Kubáček*: Corrections of estimators in linearized models. Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica *37* (1998), 69–80.
[8] *L. Kubáček, L. Kubáčková*: Regression Models with a Weak Nonlinearity. Technical Reports. University of Stuttgart, 1998, pp. 1–64.
[9] *L. Kubáček*: Linear versus quadratic estimators in linearized models. Submitted to Appl. Math.
[10] *A. Pázman*: Nonlinear Statistical Models. Kluwer Academic Publishers, Dordrecht-Boston-London and Ister Science Press, Bratislava, 1993.
[11] *C. R. Rao, S. K. Mitra*: Generalized Inverse of the Matrix and Its Applications. J. Wiley, New York, 1971.
[12] *E. Tesaříková, L. Kubáček*: How to deal with regression models with a weak nonlinearity. Discuss. Math. Probab. Stat. *21* (2001), 21–48.

Author's address: *L. Kubáček*, Department of Mathematical Analysis and Applied Mathematics, Faculty of Science, Palacký University, Tomkova 40, 779 00 Olomouc, Czech Republic, e-mail: `kubacekl@risc.upol.cz`.