Lubomír Kubáček

Linear versus quadratic estimators in linearized models

Applications of Mathematics, Vol. 49 (2004), No. 2, 81–95

Persistent URL: http://dml.cz/dmlcz/134560

49 (2004) APPLICATIONS OF MATHEMATICS No. 2, 81–95

LINEAR VERSUS QUADRATIC ESTIMATORS IN LINEARIZED MODELS*

Lubomír Kubáček, Olomouc

(Received January 30, 2001)

Abstract. In nonlinear regression models an approximate value of an unknown parameter is frequently at our disposal. Then the linearization of the model is used and a linear estimate of the parameter can be calculated. Some criteria how to recognize whether a linearization is possible are developed. In the case that they are not satisfied, it is necessary to take into account either some quadratic corrections or to use the nonlinear least squares method. The aim of the paper is to find some criteria for an ordering linear and quadratic estimators.

Keywords: nonlinear regression model, linearization, quadratization

MSC 2000: 62J05, 62F10

1. Introduction and notation

How to proceed in estimation of parameters in nonlinear models is a frequently occurring problem. There are several possibilities; to linearize the model, to use the nonlinear least squares method, the maximum likelihood principle, a polynomial estimator, etc.

The aim of the paper is to find out some simple rules how to recognize whether the linearization is sufficient for the solution of the problem or whether it is necessary to use some quadratic corrections in order to obtain an estimator with smaller MSE (mean square error) than MSE of the linear estimator. Since criteria for recognizing the possibility of linearizing the model have been already developed (cf. [4], [6], [9]), the problem is what to do when such a criterion is not satisfied.

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

The notation $\mathbf{Y} \sim N_n(\mathbf{f}(\boldsymbol{\beta}), \boldsymbol{\Sigma})$, $\boldsymbol{\beta} \in \mathbb{R}^k$, 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 the known analytical form with continuous second derivatives, the k-dimensional parameter $\boldsymbol{\beta}$ is unknown and can be any element of the k-dimensional Euclidean space $\mathbb{R}^k$. The covariance matrix $\operatorname{var}(\mathbf{Y})$ of the vector $\mathbf{Y}$ is given and is denoted by $\boldsymbol{\Sigma}$.

The linearized and quadratized approximations of this model, i.e.

$$\mathbf{Y} = \mathbf{f}_0 + \mathbf{F}\delta\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

and

$$\mathbf{Y} = \mathbf{f}_0 + \mathbf{F}\delta\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\kappa}_f(\delta\boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \tag{2}$$

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

$$\begin{aligned}
\mathbf{f}_0 &= \mathbf{f}(\boldsymbol{\beta}_0),\\
\mathbf{F} &= \partial\mathbf{f}(\mathbf{u})/\partial\mathbf{u}'|_{u=\boldsymbol{\beta}_0},\\
\boldsymbol{\kappa}_f(\delta\boldsymbol{\beta}) &= (\kappa_1(\delta\boldsymbol{\beta}), \ldots, \kappa_n(\delta\boldsymbol{\beta}))',\\
\kappa_i(\delta\boldsymbol{\beta}) &= \delta\boldsymbol{\beta}'\mathbf{F}_i\delta\boldsymbol{\beta}, \quad i = 1, \ldots, n,\\
\mathbf{F}_i &= \partial^2 f_i(\mathbf{u})/\partial\mathbf{u}\partial\mathbf{u}'|_{u=\boldsymbol{\beta}_0}, \quad i = 1, \ldots, n.
\end{aligned}$$

Remark 1.1. The vector $\boldsymbol{\beta}_0$ should be chosen in such a way that the inequality

$$\mathbf{v}_0'\boldsymbol{\Sigma}^{-1}\mathbf{v}_0 \leqslant \chi_k^2(1-\alpha),$$

be satisfied for a sufficiently small α. Here

$$\mathbf{v}_0 = \mathbf{F}(\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}_0)]$$

($\mathbf{y}$ is a realization of $\mathbf{Y}$), $\chi_k^2(1-\alpha)$ is the $(1-\alpha)$ quantile of the central chi-square distribution with k degrees of freedom. By virtue of Proposition 2.6.1 in [11] this ensures that $\boldsymbol{\beta}_0$ is an element of the confidence region for the parameter $\boldsymbol{\beta}$ with probability $1-\alpha$.

The task is to estimate a function $h(\cdot)$ of the form

$$h(\boldsymbol{\beta}) = h_0 + \mathbf{h}'\delta\boldsymbol{\beta} + \frac{1}{2}\delta\boldsymbol{\beta}'\mathbf{H}_1\delta\boldsymbol{\beta},$$

where h_0 is a known number, $\mathbf{h}$ is a known k-dimensional vector and $\mathbf{H}_1$ is a known $k \times k$ symmetric matrix.

2. Preliminaries

In what follows the following well known statement will be useful.

Lemma 2.1. *Let $\boldsymbol{\eta} \sim N_k(\boldsymbol{\mu}, \mathbf{V})$, $\mathbf{h} \in \mathbb{R}^k$ and let $\mathbf{A}$ be an $k \times k$ symmetric matrix. Then*

$$\begin{aligned}
\operatorname{var}(\mathbf{h}'\boldsymbol{\eta}) &= \mathbf{h}'\mathbf{V}\mathbf{h},\\
\operatorname{cov}(\mathbf{h}'\boldsymbol{\eta}, \boldsymbol{\eta}'\mathbf{A}\boldsymbol{\eta}) &= 2\mathbf{h}'\mathbf{V}\mathbf{A}\boldsymbol{\mu},\\
E(\boldsymbol{\eta}'\mathbf{A}\boldsymbol{\eta}) &= \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + \operatorname{Tr}(\mathbf{A}\mathbf{V}),\\
\operatorname{var}(\boldsymbol{\eta}'\mathbf{A}\boldsymbol{\eta}) &= 2\operatorname{Tr}(\mathbf{A}\mathbf{V}\mathbf{A}\mathbf{V}) + 4\boldsymbol{\mu}'\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu}.
\end{aligned}$$

Proof cf., e.g., in [5]. □

In [7] a simple quadratic estimator in the model (2) is given in the form

$$\tilde{\tilde{\boldsymbol{\beta}}} = \hat{\boldsymbol{\beta}} - \mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\frac{1}{2}\boldsymbol{\kappa}_f(\delta\hat{\boldsymbol{\beta}}) + \frac{1}{2}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}(\operatorname{Tr}(\mathbf{C}^{-1}\mathbf{F}_1), \dots, \operatorname{Tr}(\mathbf{C}^{-1}\mathbf{F}_n))',$$

where

$$\begin{aligned}
\tilde{\tilde{\boldsymbol{\beta}}} &= \boldsymbol{\beta}_0 + \delta\tilde{\tilde{\boldsymbol{\beta}}},\\
\delta\hat{\boldsymbol{\beta}} &= \mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{f}_0),\\
\hat{\boldsymbol{\beta}} &= \boldsymbol{\beta}_0 + \delta\hat{\boldsymbol{\beta}},\\
\mathbf{C} &= \mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F}.
\end{aligned}$$

We suppose that the rank of the matrix $\mathbf{F}$ is $r(\mathbf{F}) = k < n$ and the matrix $\boldsymbol{\Sigma}$ is positive definite.

Let

$$\mathbf{b}(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}, \quad \mathbf{P}_F^{\Sigma^{-1}} = \mathbf{F}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}, \quad \mathbf{M}_F^{\Sigma^{-1}} = \mathbf{I} - \mathbf{P}_F^{\Sigma^{-1}}$$

and

$$\begin{aligned}
K^{(\text{int})}(\boldsymbol{\beta}_0) &= \sup\left\{\frac{\sqrt{\boldsymbol{\kappa}_f'(\mathbf{u})\boldsymbol{\Sigma}^{-1}\mathbf{M}_F^{\Sigma^{-1}}\boldsymbol{\kappa}_f(\mathbf{u})}}{\mathbf{u}'\mathbf{C}\mathbf{u}} : \mathbf{u} \in \mathbb{R}^k\right\},\\
K^{(\text{par})}(\boldsymbol{\beta}_0) &= \sup\left\{\frac{\sqrt{\boldsymbol{\kappa}_f'(\mathbf{u})\boldsymbol{\Sigma}^{-1}\mathbf{P}_F^{\Sigma^{-1}}\boldsymbol{\kappa}_f(\mathbf{u})}}{\mathbf{u}'\mathbf{C}\mathbf{u}} : \mathbf{u} \in \mathbb{R}^k\right\}
\end{aligned}$$

(cf. [1]) and let $\chi^2_k(1-\alpha)$ be the $(1-\alpha)$-quantile of the chi-square central distribution with k degrees of freedom.

If the linearization region (with respect to bias) (cf. e.g. [4]) for the estimator $\hat{\boldsymbol{\beta}}$, i.e.

$$\left\{\boldsymbol{\beta}_0 + \delta\boldsymbol{\beta}\colon\ \delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} \leqslant c_b\sqrt{\chi^2_k(1-\alpha)}/K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\right\}$$
$$\left(\Rightarrow \forall\{\mathbf{h}' \in \mathbb{R}^k\}|\mathbf{h}'\mathbf{u}(\boldsymbol{\beta})| \leqslant c_b\sqrt{\chi^2_k(1-\alpha)}\sqrt{\mathbf{h}'\mathbf{C}^{-1}\mathbf{h}}\right),$$

covers the actual value $\boldsymbol{\beta}^*$ of the parameter $\boldsymbol{\beta}$ with sufficiently high probability, there is no use to correct the linear estimator $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \delta\hat{\boldsymbol{\beta}}$. If this situation does not occur, then it may be useful to use a correction of the linear estimator. The decision whether to realize this correction or not depends, e.g., on the mean square error (MSE) of the linear estimator and on the MSE of the quadratic estimator.

In the following the aim is to find a simple rule for the above mentioned decision (another investigation of this situation is given in [3]).

In order to be a little more general, the problem is formulated as follows.

Let (2) be under consideration. The function

$$h(\boldsymbol{\beta}) = h_0 + \mathbf{h}'\delta\boldsymbol{\beta} + \frac{1}{2}\delta\boldsymbol{\beta}'\mathbf{H}_1\delta\boldsymbol{\beta}$$

is to be estimated.

In the linearized situation, i.e. (1) and $h(\boldsymbol{\beta}) = h_0 + \mathbf{h}'\delta\boldsymbol{\beta}$, the estimator is

$$h(\hat{\boldsymbol{\beta}}) = h_0 + \mathbf{h}'\delta\hat{\boldsymbol{\beta}}.$$

Its bias is

$$b_h(\hat{\boldsymbol{\beta}}) = h_0 + \mathbf{h}'E(\delta\hat{\boldsymbol{\beta}}) - h(\boldsymbol{\beta}) = \frac{1}{2}\mathbf{L}'_h\boldsymbol{\kappa}_f(\delta\boldsymbol{\beta}) - \frac{1}{2}\delta\boldsymbol{\beta}'\mathbf{H}_1\delta\boldsymbol{\beta} = \delta\boldsymbol{\beta}'\mathbf{A}_h\delta\boldsymbol{\beta},$$

where

$$\mathbf{L}'_h = \mathbf{h}'\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1},$$
$$\mathbf{B}_h = \sum_{i=1}^{n}\frac{1}{2}\{\mathbf{L}'_h\}_i\mathbf{F}_i,$$
$$\mathbf{A}_h = \mathbf{B}_h - \frac{1}{2}\mathbf{H}_1.$$

Thus the simplest quadratic unbiased (as far as the second order terms of $\delta\boldsymbol{\beta}$ are concerned) estimator of the function $h(\cdot)$ is

$$h(\tilde{\tilde{\boldsymbol{\beta}}}) = h_0 + \mathbf{h}'\delta\hat{\boldsymbol{\beta}} - \delta\hat{\boldsymbol{\beta}}'\mathbf{A}_h\delta\hat{\boldsymbol{\beta}} + \mathrm{Tr}(\mathbf{C}^{-1}\mathbf{A}_h).$$

Remark 2.1. If $\mathbf{H}_1 = \sum_{i=1}^{n} \{\mathbf{L}_h'\}_i \mathbf{F}_i$, then it is sufficient to use the linear estimator $h(\tilde{\tilde{\boldsymbol{\beta}}}) = h(\boldsymbol{\beta}_0) + \mathbf{h}'\delta\hat{\boldsymbol{\beta}}$.

Let $\mathbf{B}_{0,i} = \sum_{j=1}^{n} \{\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\}_j \frac{1}{2}\mathbf{F}_j$ and $\mathbf{B}_{0,h} = \sum_{i=1}^{k} h_i \mathbf{B}_{0,i}$, where $\mathbf{e}_i \in \mathbb{R}^k$, $\{\mathbf{e}_i\}_j = \delta_{i,j}$ (Kronecker delta).

In the following the notation $\overline{\mathbf{B}}_i$ means $\sum_{i=1}^{k} |\lambda_i| \mathbf{f}_i\mathbf{f}_i'$, where $\mathbf{B}_{0,i} = \sum_{i=1}^{k} \lambda_i \mathbf{f}_i \mathbf{f}_i'$ is the spectral decomposition of the matrix $\mathbf{B}_{0,i}$.

The bias of the simple quadratic estimator is

$$
\begin{aligned}
b_h(\tilde{\tilde{\boldsymbol{\beta}}}) &= E_{\boldsymbol{\beta}}[h_0 + \mathbf{h}'\widehat{\delta\boldsymbol{\beta}} - \widehat{\delta\boldsymbol{\beta}}\mathbf{A}_h\widehat{\delta\boldsymbol{\beta}} + \operatorname{Tr}(\mathbf{C}^{-1}\mathbf{A}_h)] \\
&\quad - h_0 - \mathbf{h}'\delta\boldsymbol{\beta} - \frac{1}{2}\delta\boldsymbol{\beta}'\mathbf{H}_1\delta\boldsymbol{\beta} \\
&= h_0 + \mathbf{h}'\delta\boldsymbol{\beta} + \delta\boldsymbol{\beta}'\mathbf{B}_{0,h}\delta\boldsymbol{\beta} \\
&\quad - (\delta\boldsymbol{\beta}'\mathbf{A}_h\delta\boldsymbol{\beta} + 2\mathbf{b}'(\hat{\boldsymbol{\beta}})\mathbf{A}_h\delta\boldsymbol{\beta} + \mathbf{b}'(\hat{\boldsymbol{\beta}})\mathbf{A}_h\mathbf{b}(\hat{\boldsymbol{\beta}})) \\
&\quad - h_0 - \mathbf{h}'\delta\boldsymbol{\beta} - \frac{1}{2}\delta\boldsymbol{\beta}'\mathbf{H}_1\delta\boldsymbol{\beta} \\
&= -2\mathbf{b}'(\hat{\boldsymbol{\beta}})\mathbf{A}_h\delta\boldsymbol{\beta} - \mathbf{b}'(\hat{\boldsymbol{\beta}})\mathbf{A}_h\mathbf{b}(\hat{\boldsymbol{\beta}}),
\end{aligned}
$$

where

$$\mathbf{u}(\hat{\boldsymbol{\beta}}) = (\delta\boldsymbol{\beta}'\mathbf{B}_{0,1}\delta\boldsymbol{\beta}, \ldots, \delta\boldsymbol{\beta}'\mathbf{B}_{0,k}\delta\boldsymbol{\beta})'.$$

Its variance is

$$
\begin{aligned}
\text{(3)}\qquad \operatorname{var}(h(\tilde{\tilde{\boldsymbol{\beta}}})) &= \mathbf{h}'\mathbf{C}^{-1}\mathbf{h} + 2\operatorname{Tr}(\mathbf{A}_h\mathbf{C}^{-1}\mathbf{A}_h\mathbf{C}^{-1}) \\
&\quad + 4E(\delta\hat{\boldsymbol{\beta}})'\mathbf{A}_h\mathbf{C}^{-1}\mathbf{A}_h E(\delta\hat{\boldsymbol{\beta}}) - 4\mathbf{h}'\mathbf{C}^{-1}\mathbf{A}_h E(\delta\hat{\boldsymbol{\beta}})
\end{aligned}
$$

(cf. Lemma 2.1).

The MSE of the linear estimator is

$$\text{(4)}\qquad \operatorname{MSE}(\mathbf{h}'\hat{\boldsymbol{\beta}}) = \operatorname{var}(\mathbf{h}'\delta\hat{\boldsymbol{\beta}}) + b_h^2(\hat{\boldsymbol{\beta}}) = \mathbf{h}'\mathbf{C}^{-1}\mathbf{h} + (\delta\boldsymbol{\beta}'\mathbf{A}_h\delta\boldsymbol{\beta})^2.$$

The MSE of the quadratic estimator is

$$
\begin{aligned}
\text{(5)}\qquad \operatorname{MSE}(\mathbf{h}'\tilde{\tilde{\boldsymbol{\beta}}}) &= \operatorname{var}(h(\tilde{\tilde{\boldsymbol{\beta}}})) + b_h^2(\tilde{\tilde{\boldsymbol{\beta}}}) \\
&= \mathbf{h}'\mathbf{C}^{-1}\mathbf{h} + 2\operatorname{Tr}(\mathbf{A}_h\mathbf{C}^{-1}\mathbf{A}_h\mathbf{C}^{-1}) \\
&\quad + 4E(\delta\hat{\boldsymbol{\beta}})'\mathbf{A}_h\mathbf{C}^{-1}\mathbf{A}_h E(\delta\hat{\boldsymbol{\beta}}) - 4\mathbf{h}'\mathbf{C}^{-1}\mathbf{A}_h E(\delta\hat{\boldsymbol{\beta}}) \\
&\quad + (2\mathbf{b}'\mathbf{A}_h\delta\boldsymbol{\beta} + \mathbf{b}'\mathbf{A}_h\mathbf{b})^2.
\end{aligned}
$$

To find a region of shifts $\delta\boldsymbol{\beta}$ where (5) $\leqslant$ (4) is tedious for a large k (dimension of $\boldsymbol{\beta}$). A rule for the first orientation is given in (6).

3. Main results

If the dimension k of the parameter $\boldsymbol{\beta}$ is large, then the investigation of the quantities $\mathbf{b}(\hat{\boldsymbol{\beta}})$, $\mathbf{b}(\tilde{\tilde{\boldsymbol{\beta}}})$, $\mathrm{MSE}(\hat{\beta}_i)$, $i=1,\dots,k$, and $\mathrm{MSE}(\tilde{\tilde{\beta}}_i)$, $i=1,\dots,k$, in this way is tedious, since the number of different directions of $\delta\boldsymbol{\beta}$ necessary for an investigation may be huge. Thus it seems that the following theorem can be useful in practice.

Theorem 3.1. *The following inequalities are valid:*

$$
\begin{aligned}
|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}},\\
|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}},\\
|\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant \sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant \mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|4\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|4\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\,\mathrm{Tr}[(\overline{\mathbf{B}}_i\mathbf{C}^{-1})^2]\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b}| &\leqslant 2\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b}| &\leqslant 2\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant \sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{(3/2)},\\
|-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant \mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{(3/2)},\\
|8\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{(3/2)},\\
|8\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\,\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{(3/2)},\\
|\mathbf{b}'\mathbf{B}_{0,i}\mathbf{b}| &\leqslant \frac{1}{4}\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2,\\
|\mathbf{b}'\mathbf{B}_{0,i}\mathbf{b}| &\leqslant \frac{1}{4}\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2,\\
|4\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b}| &\leqslant \sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2,\\
|4\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b}| &\leqslant \mathrm{Tr}[(\overline{\mathbf{B}}_i\mathbf{C}^{-1})^2][K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2.
\end{aligned}
$$

Obviously the right-hand sides depend on the quantity independent of the direction of the shift $\delta\boldsymbol{\beta}$. They depend on the Mahalanobis distance $\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}}$ only.

Proof. The Schwarz inequality and the relation implied by the definition of the quantity $K^{(\mathrm{par})}(\boldsymbol{\beta}_0)$, i.e.

$$\mathbf{b}'\mathbf{C}\mathbf{b} \leqslant \frac{1}{4}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},$$

will be used. We have

$$
\begin{aligned}
|\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &= |\delta\boldsymbol{\beta}'\mathbf{C}^{1/2}\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \\
&\leqslant \delta\boldsymbol{\beta}'\mathbf{C}^{1/2}\sum_{i=1}^{k}|\lambda_i|\mathbf{f}_i\mathbf{f}_i'\mathbf{C}^{1/2}\delta\boldsymbol{\beta} = \sum_{i=1}^{k}|\lambda_i|(\mathbf{f}'\mathbf{C}^{1/2}\delta\boldsymbol{\beta})^2 \\
&\leqslant \sum_{i=1}^{k}|\lambda_i|\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} = \mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},
\end{aligned}
$$

where $\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2} = \sum_{i=1}^{k}\lambda_i\mathbf{f}_i\mathbf{f}_i'$ is the spectral decomposition of the matrix $\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}$.

Another procedure is

$$
\begin{aligned}
|\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &= |\delta\boldsymbol{\beta}\mathbf{C}^{1/2}\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \\
&= |\,\mathrm{Tr}(\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\delta\boldsymbol{\beta}'\mathbf{C}^{1/2})| \\
&\leqslant \sqrt{\mathrm{Tr}(\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2})}\sqrt{\mathrm{Tr}[(\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\delta\boldsymbol{\beta}'\mathbf{C}^{1/2})^2]} \\
&= \sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},
\end{aligned}
$$

$$
\begin{aligned}
|-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &= 2|\mathbf{b}'\mathbf{C}^{1/2}\sum_{i=1}^{k}\lambda_i\mathbf{f}_i\mathbf{f}_i'\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \leqslant 2\sum_{i=1}^{k}|\lambda_i|\,|\mathbf{b}'\mathbf{C}^{1/2}\mathbf{f}_i|\,|\mathbf{f}_i'\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \\
&\leqslant 2\sum_{i=1}^{k}|\lambda_i|\sqrt{\mathbf{b}'\mathbf{C}\mathbf{b}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} = 2\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\mathbf{b}'\mathbf{C}\mathbf{b}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} \\
&\leqslant \mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2}.
\end{aligned}
$$

Another procedure is

$$
\begin{aligned}
|-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &= 2|\mathbf{b}'\mathbf{C}^{1/2}\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \\
&= 2|\,\mathrm{Tr}(\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\mathbf{b}'\delta\boldsymbol{\beta}\mathbf{C}^{1/2})| \\
&\leqslant 2\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\mathrm{Tr}(\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\mathbf{b}'\mathbf{C}^{1/2}\mathbf{C}^{1/2}\mathbf{b}'\delta\boldsymbol{\beta}\mathbf{C}^{1/2})} \\
&= 2\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\mathbf{b}'\mathbf{C}\mathbf{b}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} \\
&\leqslant 2\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\frac{1}{2}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2}.
\end{aligned}
$$

The other inequalities can be proved analogously. □

The term $-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}$ is dominant when $\text{MSE}(\hat{\beta}_i)$ and $\text{MSE}(\tilde{\tilde{\beta}}_i)$ is compared in the small neighbourhood of the point $\boldsymbol{\beta}_0$.

Lemma 3.1.
(i)

$$\max\{|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}|\colon\ \delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} = c^2\} = \left|-4c\frac{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}{\sqrt{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{i,i}\mathbf{C}^{-1}\}_{i,i}}}\right|.$$

(ii) *The equality*

$$|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| = (\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta})^2$$

in the direction of the vector $\{\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{\cdot,i}$ *is attained for*

$$\delta\boldsymbol{\beta} = \frac{c}{\sqrt{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}}\{\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{\cdot,i},$$

where

$$c^3 = 4\frac{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}^2\mathbf{C}^{-1}\}_{i,i}}{[\{\mathbf{C}^{-1}\mathbf{B}_{0,i}^3\mathbf{C}^{-1}\}_{i,i}]^2}\sqrt{[\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}]^3}.$$

P r o o f. The gradient of the function

$$f(\delta\boldsymbol{\beta}) = -4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}, \quad \delta\boldsymbol{\beta} \in \mathbb{R}^k,$$

is $-4\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{e}_i$ at the point $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. If the vector $\delta\boldsymbol{\beta}$, satisfying the equality $c^2 = \delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}$, is directed as the gradient, then obviously the function $f(\cdot)$ attains its maximum. Further procedure is evident. □

The distance c from the last lemma can be compared with the value

$$\sqrt{\chi_k^2(1-\alpha)}$$

(the boundary of the confidence ellipsoid). The shift $\delta\boldsymbol{\beta}$ in the direction of the gradient which attains the boundary of the confidence ellipsoid is

$$\delta\boldsymbol{\beta} = \sqrt{\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}}\{\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{\cdot,i}.$$

R e m a r k 3.1. If the bias of the linear estimator brought to the square is smaller than the term $|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}|$ for the above mentioned $\delta\boldsymbol{\beta}$, then the quadratic correction is of no use. Thus the following rule for the first orientation can be used.

If

$$(6)\qquad \left(\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}\right)^2(\{\mathbf{C}^{-1}\mathbf{B}_{0,i}^3\mathbf{C}^{-1}\}_{i,i})^2$$
$$\leqslant 4\{\mathbf{C}^{-1}\mathbf{B}_{0,i}^2\mathbf{C}^{-1}\}_{i,i}\sqrt{\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}},$$

the linear estimator is to be preferred. If the opposite inequality occurs, then some more detailed investigation should be performed.

Remark 3.2. If $\mathbf{C}^{-1} = \sigma^2\mathbf{C}_0^{-1}$, then the last inequality can be written in the form

$$(7)\qquad \sigma^4\left(\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}_0^{-1}\mathbf{B}_{0,i}\mathbf{C}_0\mathbf{B}_{0,i}\mathbf{C}_0^{-1}\}_{i,i}}\right)^2(\{\mathbf{C}_0^{-1}\mathbf{B}_{0,i}^3\mathbf{C}_0^{-1}\}_{i,i})^2$$
$$\leqslant 4\{\mathbf{C}_0^{-1}\mathbf{B}_{0,i}^2\mathbf{C}_0^{-1}\}_{i,i}\sqrt{\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}_0^{-1}\mathbf{B}_{0,i}\mathbf{C}_0\mathbf{B}_{0,i}\mathbf{C}_0^{-1}\}_{i,i}}}.$$

The role played by the parameter σ is now quite obvious.

Remark 3.3. The correction term in the case $h(\boldsymbol{\beta}) = \beta_i$ is

$$\tau_{0,i} = -\delta\hat{\boldsymbol{\beta}}'\mathbf{B}_{0,i}\delta\hat{\boldsymbol{\beta}} + \mathrm{Tr}(\mathbf{B}_{0,i}\mathbf{C}^{-1}).$$

If $\mathbf{B}_{0,i} = \overline{\mathbf{B}}_i$, then according to [10], [13], [15] it can be approximed by the random variable

$$\tau_{0,i} \approx -c_i^2\chi_{f_i}^2(0) + \mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1}),$$

where

$$c_i^2 = V_i/E_i, \quad f_i = E_i^2/V_i,$$
$$E_i = \delta\boldsymbol{\beta}'\overline{\mathbf{B}}_i\delta\boldsymbol{\beta} + 2\mathbf{b}'\overline{\mathbf{B}}_i\delta\boldsymbol{\beta} + \mathbf{b}'\overline{\mathbf{B}}_i\mathbf{b},$$
$$V_i = 2\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1}\overline{\mathbf{B}}_i\mathbf{C}^{-1}) + 4(\delta\boldsymbol{\beta}+\mathbf{b})'\overline{\mathbf{B}}_i\mathbf{C}^{-1}\overline{\mathbf{B}}_i(\delta\boldsymbol{\beta}+\mathbf{b}).$$

If $\mathbf{B}_{0,i}$ is not positive semidefinite, then the distribution function of $\tau_{0,i}$ cannot be obtained so simply (cf. [2]).

Remark 3.4. If a function $h(\boldsymbol{\beta}) = \mathbf{h}'\boldsymbol{\beta}$, $\boldsymbol{\beta} \in \mathbb{R}^k$, is under consideration, then the matrix $\mathbf{B}_{0,h}$ is used instead of $\mathbf{B}_{0,i}$. In the case of a function

$$h(\boldsymbol{\beta}) = h_0 + \mathbf{h}'\delta\boldsymbol{\beta} + \frac{1}{2}\delta\boldsymbol{\beta}'\mathbf{H}_1\delta\boldsymbol{\beta},$$

the matrix $\mathbf{A}_h$ must be used.

4. Example

In the first step let us investigate the bias and the MSE of the linear and of the quadratic estimator, respectively.

The Michaelis-Menten model is under consideration; i.e.

$$f(x;\beta_1,\beta_2)=\frac{\beta_1 x}{\beta_2+x}.$$

Let $(\beta_{1,0},\beta_{2,0})'=(5,1)'$ and

x	1	2	3	4	5	6
$f(x;\beta_{1,0},\beta_{2,0})$	2.5	3.33	3.75	4	4.17	4.29

If the observation vector is $\mathbf{Y}\sim N_6(\mathbf{f}(\cdot,\boldsymbol{\beta}),\sigma^2\mathbf{I})$ and $\sigma=0.1$, then

$$K^{(\mathrm{int})}(\boldsymbol{\beta}_0)=0.025443,\quad K^{(\mathrm{par})}(\boldsymbol{\beta}_0)=0.090940.$$

Thus the intrinsic nonlinearity can be neglected (in detail cf. [14]) and the linearization regions with respect to the bias of the whole vector $\boldsymbol{\beta}$ and the single parameters β_1 and β_2 are given in Figs. 4.1–4.3, respectively.

In the figures seven numbers are given; five white numbers are connected with the white ellipse (linearization region). They give the maximum coordinates of the ellipse points, a size of the raster rectangulars and the step in the first coordinate used for the construction of the ellipse. Two dark numbers have the analogous meaning for the dark ellipse (0.95-confidence ellipse).

In this situation the linearization for the functions $h_1(\boldsymbol{\beta})=\beta_1$ or $h_2(\boldsymbol{\beta})=\beta_2$, respectively, is possible, even if the confidence ellipse for the vector parameter is not essentially smaller than the linearization region (cf. Fig. 4.1). Thus it is interesting whether the quadratic estimator is not better.

As far as the parameter β_1 is concerned the shift $\delta\boldsymbol{\beta}=(0,\delta\beta_2)'$ is dangerous (cf. Fig. 4.2). Let $\delta\boldsymbol{\beta}$ be chosen on the boundary of the 0.95-confidence ellipse, i.e. $\delta\beta_2=0.1076$. Thus we obtain

$$b(\hat{\beta}_1)=-0.004469\quad\text{and}\quad b(\hat{\beta}_2)=-0.007042,$$

i.e. $\mathbf{b}=(-0.004469,-0.007042)'$. The bias $b(\tilde{\tilde{\beta}}_1)$ is

$$b(\tilde{\tilde{\beta}}_1)=-2\mathbf{b}'\mathbf{B}_{0,1}\delta\beta+\mathbf{b}'\mathbf{B}_{0,1}\mathbf{b}=-0.000604,$$

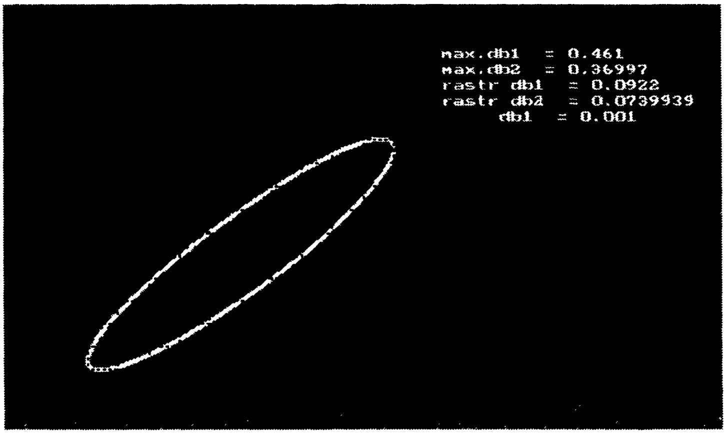

Figure 4.1 Linearization region for the whole vector $\boldsymbol{\beta}$.

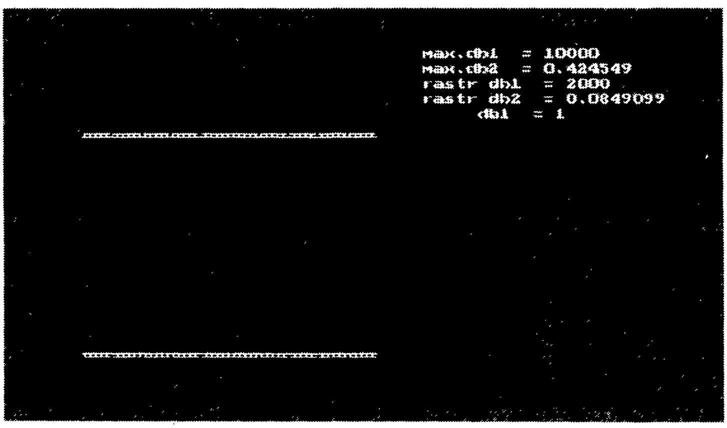

Figure 4.2. Linearization region for β_1.

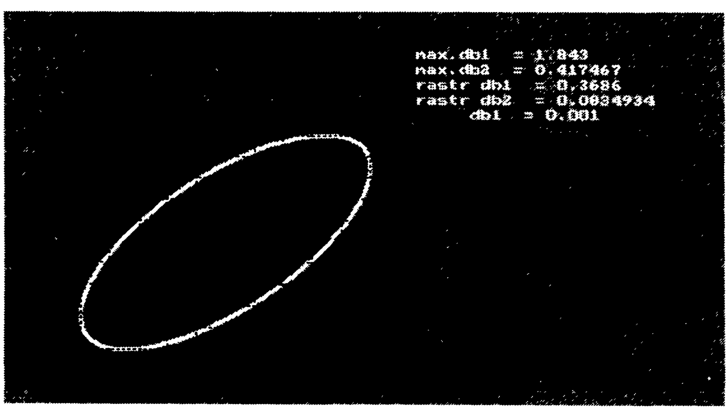

Figure 4.3. Linearization region for β_2.

which is significantly smaller (in the absolute value) than $b(\hat{\beta}_1) = -0.004469$. Even for $\delta\boldsymbol{\beta} = 3 \times (0, 0.1076)'$ the situation is analogous:

$$\mathbf{b}(\hat{\boldsymbol{\beta}}) = (-0,040224, -0.063374)', \quad b(\tilde{\tilde{\beta}}_1) = -0.017342.$$

(The values $\sqrt{\mathrm{var}(\hat{\beta}_1)}$ and $\sqrt{\mathrm{var}(\hat{\beta}_2)}$ are 0.139 and 0.125, respectively.)

The bias $b(\tilde{\tilde{\beta}}_2)$ due to the shift $\delta\boldsymbol{\beta} = (0, 0.1076)'$ and $\delta\boldsymbol{\beta} = 3 \times (0, 0.1076)'$ is -0.000849 and -0.24219, respectively.

Further we have ($\delta\boldsymbol{\beta} = (0, 0.1076)'$)

$$\begin{aligned} \mathrm{var}(\hat{\beta}_1) &= 0.01936, & \mathrm{var}(\tilde{\tilde{\beta}}_1) &= 0.02170, \\ \mathrm{var}(\hat{\beta}_2) &= 0.01245, & \mathrm{var}(\tilde{\tilde{\beta}}_2) &= 0.01511, \end{aligned}$$

and

$$\begin{aligned} \mathrm{MSE}(\hat{\beta}_1) &= 0.01938, & \mathrm{MSE}(\tilde{\tilde{\beta}}_1) &= 0.02170, \\ \mathrm{MSE}(\hat{\beta}_2) &= 0.01250, & \mathrm{MSE}(\tilde{\tilde{\beta}}_2) &= 0.01511. \end{aligned}$$

If $\delta\boldsymbol{\beta} = 3 \times (0,\ 0.1076)'$, then

$$\begin{aligned} \mathrm{MSE}(\hat{\beta}_1) &= 0.02098, & \mathrm{MSE}(\tilde{\tilde{\beta}}_1) &= 0.02622, \\ \mathrm{MSE}(\hat{\beta}_2) &= 0.01646, & \mathrm{MSE}(\tilde{\tilde{\beta}}_2) &= 0.02005. \end{aligned}$$

If we denote $\delta\boldsymbol{\beta} = \begin{pmatrix} 0 \\ x \end{pmatrix}$, we obtain

$$\begin{aligned} \mathrm{MSE}(\hat{\beta}_1) &= 0.01936 + 0.148981x^4, \\ \mathrm{MSE}(\tilde{\tilde{\beta}}_1) &= 0.019406 + 0.022028x - 0.005977x^2 - 0.009023x^3 \\ &\qquad + 0.002744x^4 + (0.469452x^3 + 0.142744x^4)^2, \\ \mathrm{MSE}(\hat{\beta}_2) &= 0.012448 + 0.369822x^4, \\ \mathrm{MSE}(\tilde{\tilde{\beta}}_2) &= 0.012493 + 0.024572x - 0.000777x^2 - 0.012996x^3 \\ &\qquad + 0.003449x^4 + (0.662448x^3 + 0.177955x^4)^2, \end{aligned}$$

cf. Table 1.

x	MSE($\hat{\beta}_1$)	MSE($\tilde{\tilde{\beta}}_1$)	MSE($\hat{\beta}_2$)	MSE($\tilde{\tilde{\beta}}_2$)
0.03	0.01936	0.02006	0.01245	0.01323
0.1	0.01938	0.02154	0.01249	0.01493
0.2	0.01960	0.02352	0.01304	0.01727
0.3	0.02057	0.02545	0.01544	0.01984
0.4	0.02317	0.02789	0.02192	0.02366
0.5	0.02867	0.03254	0.03556	0.03200
0.6	0.03867	0.04325	0.06038	0.05220
0.7	0.05513	0.06760	0.10124	0.09855

Table 1.

It is obvious that the quadratic corrections in this case have no sense.
The inequalities from Theorem 3.1 for the parameter β_1 (β_2) are

$$
\begin{aligned}
-4\mathbf{e}_1'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta} &= 0.00237\ (0.00264),\\
4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} &= 0.00654\ (0.00524),\\
4\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} &= 0.00654\ (0.00598),\\
\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta} &= -\,0.00447\ (-0.00704),\\
\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} &= 0.02878\ (0.02872),\\
\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= 0.02878\ (0.03279),\\
4\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,1}\delta\boldsymbol{\beta} &= 0.00009\ (0.00014),\\
4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} &= 0.00055\ (0.00055),\\
4\,\mathrm{Tr}[\overline{\mathbf{B}}_i\mathbf{C}^{-1})^2]\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= 0.00055\ (0.00072),\\
-4\mathbf{e}_1'\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b} &= -\,0.00015\ (-0.00015),\\
2\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= 0.00073\ (0.00058),\\
2\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= 0.00073\ (0.00067),\\
-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta} &= -\,0.00058\ (-0.00082),\\
\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2} &= 0.00641\ (0.00639),\\
\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})K^{(\mathrm{par})}(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2} &= 0.00641\ (0.00730),
\end{aligned}
$$

$$8\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta} = -\,0.00001\ (-0.00002),$$
$$4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}K^{\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2} = 0.00012\ (0.00012),$$
$$4\,\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2} = 0.00012\ (0.00016),$$
$$\mathbf{b}'\mathbf{B}_{0,i}\mathbf{b} = -\,0.00002\ (-0.00002),$$
$$\frac{1}{4}\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2 = 0.00036\ (0.00036),$$
$$\frac{1}{4}\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2 = 0.00036\ (0.00041),$$
$$4\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b} = 0.00000\ (0.00000),$$
$$\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2 = 0.00001\ (0.00001),$$
$$\mathrm{Tr}[(\overline{\mathbf{B}}_i\mathbf{C}^{-1})^2][K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2 = 0.00001\ (0.00001).$$

Here the index i on the left-hand side means 1 for the first number on the right-hand side and 2 for the second (in the bracket).

In many cases the upper bound is significantly larger than the actual value (by virtue of the Schwarz inequality). Nevertheless, some information on the individual terms can be obtained in this way:

$$\mathbf{C}_0 = \begin{pmatrix} 3.3261, & -3.8124 \\ -3.8124, & 5.1731 \end{pmatrix},$$
$$\mathbf{C}_0^{-1} = \begin{pmatrix} 1.936, & 1.427 \\ 1.427, & 1.246 \end{pmatrix},$$
$$\mathbf{B}_{0,2} = \begin{pmatrix} 0, & 0.10000 \\ 0.10000, & -0.60813 \end{pmatrix},$$

$$\left(\frac{\chi_2^2(0;0.95)}{\{\mathbf{C}_0^{-1}\mathbf{B}_{0,2}\mathbf{C}_0\mathbf{B}_{0,2}\mathbf{C}_0^{-1}\}_{i,i}}\right)^2(\{\mathbf{C}_0^{-1}\mathbf{B}_{0,2}^3\mathbf{C}_0^{-1}\}_{2,2})^2 = 0.3214,$$
$$4\{\mathbf{C}_0^{-1}\mathbf{B}_{0,2}^2\mathbf{C}_0^{-1}\}_{2,2}\sqrt{\frac{\chi_2^2(0;0.95)}{\{\mathbf{C}_0^{-1}\mathbf{B}_{0,2i}\mathbf{C}_0\mathbf{B}_{0,2}\mathbf{C}_0^{-1}\}_{2,2}}} = 2.3918.$$

The left-hand side of (7) is smaller than the right-hand side of (7) even for $\sigma = 1.65$. In this case
$$K^{(\mathrm{int})}(\boldsymbol{\beta}_0) = 0.4198, \quad K^{(\mathrm{par})}(\boldsymbol{\beta}_0) = 1.5005.$$

An experiment characterized by these values would be extremely badly planned. Thus the Michaelis-Menten model can be linearized at the considered point $\boldsymbol{\beta}_0 = (5,1)'$ under a sufficiently small value of σ only. The quadratic corrections are of no use (cf. also Tab. 1). If a sufficiently small σ cannot be attained, then the methods given in [11], [12] must be used.

References

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

[2] *J. P. Imhof*: Computing the distribution of quadratic forms in normal variables. Biometrika *48* (1961), 419–426.

[3] *A. Jenčová*: A comparison of linearization and quadratization domains. Appl. Math. *42* (1997), 279–291.

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

[5] *L. Kubáček, L. Kubáčková, J. Volaufová*: Statistical Models with Linear Structures. Veda, Bratislava, 1995.

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

[7] *L. Kubáček*: Quadratic regression models. Math. Slovaca *46* (1996), 111–126.

[8] *L. Kubáček*: Corrections of estimators in linearized models. Acta Univ. Palack. Olomuc., Fac. Rerum Math. *37* (1998), 69–80.

[9] *L. Kubáček, L. Kubáčková*: Regression models with a weak nonlinearity. Technical Reports. University of Stuttgart.

[10] *P. B. Patnaik*: The non-central χ^2 and F-distributions and their applications. Biometrika *36* (1949), 202–232.

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

[12] *R. Potocký, To Van Ban*: Confidence regions in nonlinear regression models. Appl. Math. *37* (1992), 29–39.

[13] *F. E. Satterthwaite*: An approximate distribution of estimates of variance components. Biometrics Bulletin *2* (1946), 110-114.

[14] *E. Tesaříková, L. Kubáček*: How to deal with regression models with a weak nonlinearity. Discuss. Math. Probab. Stat. *21* (2001), 21–48.

[15] *B. L. Welch*: The generalization of Student's problem when several different population variances are involved. Biometrika *34* (1947), 28–35.

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

LINEAR VERSUS QUADRATIC ESTIMATORS IN LINEARIZED MODELS*

Lubomír Kubáček, Olomouc

(Received January 30, 2001)

Abstract. In nonlinear regression models an approximate value of an unknown parameter is frequently at our disposal. Then the linearization of the model is used and a linear estimate of the parameter can be calculated. Some criteria how to recognize whether a linearization is possible are developed. In the case that they are not satisfied, it is necessary to take into account either some quadratic corrections or to use the nonlinear least squares method. The aim of the paper is to find some criteria for an ordering linear and quadratic estimators.

Keywords: nonlinear regression model, linearization, quadratization

MSC 2000: 62J05, 62F10

1. Introduction and notation

How to proceed in estimation of parameters in nonlinear models is a frequently occurring problem. There are several possibilities; to linearize the model, to use the nonlinear least squares method, the maximum likelihood principle, a polynomial estimator, etc.

The aim of the paper is to find out some simple rules how to recognize whether the linearization is sufficient for the solution of the problem or whether it is necessary to use some quadratic corrections in order to obtain an estimator with smaller MSE (mean square error) than MSE of the linear estimator. Since criteria for recognizing the possibility of linearizing the model have been already developed (cf. [4], [6], [9]), the problem is what to do when such a criterion is not satisfied.

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

The notation $\mathbf{Y} \sim N_n(\mathbf{f}(\boldsymbol{\beta}), \boldsymbol{\Sigma})$, $\boldsymbol{\beta} \in \mathbb{R}^k$, 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 the known analytical form with continuous second derivatives, the k-dimensional parameter $\boldsymbol{\beta}$ is unknown and can be any element of the k-dimensional Euclidean space $\mathbb{R}^k$. The covariance matrix $\operatorname{var}(\mathbf{Y})$ of the vector $\mathbf{Y}$ is given and is denoted by $\boldsymbol{\Sigma}$.

The linearized and quadratized approximations of this model, i.e.

$$\mathbf{Y} = \mathbf{f}_0 + \mathbf{F}\delta\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

and

$$\mathbf{Y} = \mathbf{f}_0 + \mathbf{F}\delta\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\kappa}_f(\delta\boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \tag{2}$$

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

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

Remark 1.1. The vector $\boldsymbol{\beta}_0$ should be chosen in such a way that the inequality

$$\mathbf{v}_0'\boldsymbol{\Sigma}^{-1}\mathbf{v}_0 \leqslant \chi_k^2(1-\alpha),$$

be satisfied for a sufficiently small α. Here

$$\mathbf{v}_0 = \mathbf{F}(\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}[\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}_0)]$$

($\mathbf{y}$ is a realization of $\mathbf{Y}$), $\chi_k^2(1-\alpha)$ is the $(1-\alpha)$ quantile of the central chi-square distribution with k degrees of freedom. By virtue of Proposition 2.6.1 in [11] this ensures that $\boldsymbol{\beta}_0$ is an element of the confidence region for the parameter $\boldsymbol{\beta}$ with probability $1-\alpha$.

The task is to estimate a function $h(\cdot)$ of the form

$$h(\boldsymbol{\beta}) = h_0 + \mathbf{h}'\delta\boldsymbol{\beta} + \frac{1}{2}\delta\boldsymbol{\beta}'\mathbf{H}_1\delta\boldsymbol{\beta},$$

where h_0 is a known number, $\mathbf{h}$ is a known k-dimensional vector and $\mathbf{H}_1$ is a known $k \times k$ symmetric matrix.

2. Preliminaries

In what follows the following well known statement will be useful.

Lemma 2.1. *Let* $\boldsymbol{\eta} \sim N_k(\boldsymbol{\mu}, \mathbf{V})$, $\mathbf{h} \in \mathbb{R}^k$ *and let* $\mathbf{A}$ *be an* $k \times k$ *symmetric matrix. Then*

$$\begin{aligned} \operatorname{var}(\mathbf{h}'\boldsymbol{\eta}) &= \mathbf{h}'\mathbf{V}\mathbf{h}, \\ \operatorname{cov}(\mathbf{h}'\boldsymbol{\eta}, \boldsymbol{\eta}'\mathbf{A}\boldsymbol{\eta}) &= 2\mathbf{h}'\mathbf{V}\mathbf{A}\boldsymbol{\mu}, \\ E(\boldsymbol{\eta}'\mathbf{A}\boldsymbol{\eta}) &= \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + \operatorname{Tr}(\mathbf{A}\mathbf{V}), \\ \operatorname{var}(\boldsymbol{\eta}'\mathbf{A}\boldsymbol{\eta}) &= 2\operatorname{Tr}(\mathbf{A}\mathbf{V}\mathbf{A}\mathbf{V}) + 4\boldsymbol{\mu}'\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu}. \end{aligned}$$

P r o o f   cf., e.g., in [5]. □

In [7] a simple quadratic estimator in the model (2) is given in the form

$$\tilde{\tilde{\boldsymbol{\beta}}} = \hat{\boldsymbol{\beta}} - \mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\frac{1}{2}\boldsymbol{\kappa}_f(\delta\hat{\boldsymbol{\beta}}) + \frac{1}{2}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}(\operatorname{Tr}(\mathbf{C}^{-1}\mathbf{F}_1), \dots, \operatorname{Tr}(\mathbf{C}^{-1}\mathbf{F}_n))',$$

where

$$\begin{aligned} \tilde{\tilde{\boldsymbol{\beta}}} &= \boldsymbol{\beta}_0 + \delta\tilde{\tilde{\boldsymbol{\beta}}}, \\ \delta\hat{\boldsymbol{\beta}} &= \mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{f}_0), \\ \hat{\boldsymbol{\beta}} &= \boldsymbol{\beta}_0 + \delta\hat{\boldsymbol{\beta}}, \\ \mathbf{C} &= \mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F}. \end{aligned}$$

We suppose that the rank of the matrix $\mathbf{F}$ is $r(\mathbf{F}) = k < n$ and the matrix $\boldsymbol{\Sigma}$ is positive definite.

Let

$$\mathbf{b}(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}, \quad \mathbf{P}_F^{\Sigma^{-1}} = \mathbf{F}\mathbf{C}^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}, \quad \mathbf{M}_F^{\Sigma^{-1}} = \mathbf{I} - \mathbf{P}_F^{\Sigma^{-1}}$$

and

$$K^{(\mathrm{int})}(\boldsymbol{\beta}_0) = \sup\left\{\frac{\sqrt{\boldsymbol{\kappa}_f'(\mathbf{u})\boldsymbol{\Sigma}^{-1}\mathbf{M}_F^{\Sigma^{-1}}\boldsymbol{\kappa}_f(\mathbf{u})}}{\mathbf{u}'\mathbf{C}\mathbf{u}} \colon \mathbf{u} \in \mathbb{R}^k\right\},$$

$$K^{(\mathrm{par})}(\boldsymbol{\beta}_0) = \sup\left\{\frac{\sqrt{\boldsymbol{\kappa}_f'(\mathbf{u})\boldsymbol{\Sigma}^{-1}\mathbf{P}_F^{\Sigma^{-1}}\boldsymbol{\kappa}_f(\mathbf{u})}}{\mathbf{u}'\mathbf{C}\mathbf{u}} \colon \mathbf{u} \in \mathbb{R}^k\right\}$$

(cf. [1]) and let $\chi_k^2(1-\alpha)$ be the $(1-\alpha)$-quantile of the chi-square central distribution with k degrees of freedom.

If the linearization region (with respect to bias) (cf. e.g. [4]) for the estimator $\hat{\boldsymbol\beta}$, i.e.

$$\left\{\boldsymbol\beta_0 + \delta\boldsymbol\beta\colon\ \delta\boldsymbol\beta'\mathbf{C}\delta\boldsymbol\beta \leqslant c_b\sqrt{\chi_k^2(1-\alpha)}/K^{(\mathrm{par})}(\boldsymbol\beta_0)\right\}$$
$$\left(\Rightarrow \forall\{\mathbf{h}' \in \mathbb{R}^k\}|\mathbf{h}'\mathbf{u}(\boldsymbol\beta)| \leqslant c_b\sqrt{\chi_k^2(1-\alpha)}\sqrt{\mathbf{h}'\mathbf{C}^{-1}\mathbf{h}}\right),$$

covers the actual value $\boldsymbol\beta^*$ of the parameter $\boldsymbol\beta$ with sufficiently high probability, there is no use to correct the linear estimator $\hat{\boldsymbol\beta} = \boldsymbol\beta_0 + \delta\hat{\boldsymbol\beta}$. If this situation does not occur, then it may be useful to use a correction of the linear estimator. The decision whether to realize this correction or not depends, e.g., on the mean square error (MSE) of the linear estimator and on the MSE of the quadratic estimator.

In the following the aim is to find a simple rule for the above mentioned decision (another investigation of this situation is given in [3]).

In order to be a little more general, the problem is formulated as follows.

Let (2) be under consideration. The function

$$h(\boldsymbol\beta) = h_0 + \mathbf{h}'\delta\boldsymbol\beta + \frac{1}{2}\delta\boldsymbol\beta'\mathbf{H}_1\delta\boldsymbol\beta$$

is to be estimated.

In the linearized situation, i.e. (1) and $h(\boldsymbol\beta) = h_0 + \mathbf{h}'\delta\boldsymbol\beta$, the estimator is

$$h(\hat{\boldsymbol\beta}) = h_0 + \mathbf{h}'\delta\hat{\boldsymbol\beta}.$$

Its bias is

$$b_h(\hat{\boldsymbol\beta}) = h_0 + \mathbf{h}'E(\delta\hat{\boldsymbol\beta}) - h(\boldsymbol\beta) = \frac{1}{2}\mathbf{L}_h'\boldsymbol\kappa_f(\delta\boldsymbol\beta) - \frac{1}{2}\delta\boldsymbol\beta'\mathbf{H}_1\delta\boldsymbol\beta = \delta\boldsymbol\beta'\mathbf{A}_h\delta\boldsymbol\beta,$$

where

$$\mathbf{L}_h' = \mathbf{h}'\mathbf{C}^{-1}\mathbf{F}'\boldsymbol\Sigma^{-1},$$
$$\mathbf{B}_h = \sum_{i=1}^{n}\frac{1}{2}\{\mathbf{L}_h'\}_i\mathbf{F}_i,$$
$$\mathbf{A}_h = \mathbf{B}_h - \frac{1}{2}\mathbf{H}_1.$$

Thus the simplest quadratic unbiased (as far as the second order terms of $\delta\boldsymbol\beta$ are concerned) estimator of the function $h(\cdot)$ is

$$h(\tilde{\tilde{\boldsymbol\beta}}) = h_0 + \mathbf{h}'\delta\hat{\boldsymbol\beta} - \delta\hat{\boldsymbol\beta}'\mathbf{A}_h\delta\hat{\boldsymbol\beta} + \mathrm{Tr}(\mathbf{C}^{-1}\mathbf{A}_h).$$

Remark 2.1. If $\mathbf{H}_1 = \sum_{i=1}^{n} \{\mathbf{L}_h'\}_i \mathbf{F}_i$, then it is sufficient to use the linear estimator $h(\tilde{\hat{\boldsymbol\beta}}) = h(\boldsymbol\beta_0) + \mathbf{h}'\delta\hat{\boldsymbol\beta}$.

Let $\mathbf{B}_{0,i} = \sum_{j=1}^{n} \{\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{F}'\boldsymbol\Sigma^{-1}\}_j \frac{1}{2}\mathbf{F}_j$ and $\mathbf{B}_{0,h} = \sum_{i=1}^{k} h_i \mathbf{B}_{0,i}$, where $\mathbf{e}_i \in \mathbb{R}^k$, $\{\mathbf{e}_i\}_j = \delta_{i,j}$ (Kronecker delta).

In the following the notation $\overline{\mathbf{B}}_i$ means $\sum_{i=1}^{k} |\lambda_i| \mathbf{f}_i\mathbf{f}_i'$, where $\mathbf{B}_{0,i} = \sum_{i=1}^{k} \lambda_i \mathbf{f}_i\mathbf{f}_i'$ is the spectral decomposition of the matrix $\mathbf{B}_{0,i}$.

The bias of the simple quadratic estimator is

$$\begin{aligned} b_h(\tilde{\hat{\boldsymbol\beta}}) &= E_\beta[h_0 + \mathbf{h}'\widehat{\delta\boldsymbol\beta} - \widehat{\delta\boldsymbol\beta}\mathbf{A}_h\widehat{\delta\boldsymbol\beta} + \mathrm{Tr}(\mathbf{C}^{-1}\mathbf{A}_h)] \\ &\quad - h_0 - \mathbf{h}'\delta\boldsymbol\beta - \frac{1}{2}\delta\beta'\mathbf{H}_1\delta\boldsymbol\beta \\ &= h_0 + \mathbf{h}'\delta\boldsymbol\beta + \delta\boldsymbol\beta'\mathbf{B}_{0,h}\delta\boldsymbol\beta \\ &\quad - (\delta\boldsymbol\beta'\mathbf{A}_h\delta\boldsymbol\beta + 2\mathbf{b}'(\hat{\boldsymbol\beta})\mathbf{A}_h\delta\boldsymbol\beta + \mathbf{b}'(\hat{\boldsymbol\beta})\mathbf{A}_h\mathbf{b}(\hat{\boldsymbol\beta})) \\ &\quad - h_0 - \mathbf{h}'\delta\boldsymbol\beta - \frac{1}{2}\delta\boldsymbol\beta'\mathbf{H}_1\delta\boldsymbol\beta \\ &= -2\mathbf{b}'(\hat{\boldsymbol\beta})\mathbf{A}_h\delta\boldsymbol\beta - \mathbf{b}'(\hat{\boldsymbol\beta})\mathbf{A}_h\mathbf{b}(\hat{\boldsymbol\beta}), \end{aligned}$$

where

$$\mathbf{u}(\hat{\boldsymbol\beta}) = (\delta\boldsymbol\beta'\mathbf{B}_{0,1}\delta\boldsymbol\beta, \ldots, \delta\boldsymbol\beta'\mathbf{B}_{0,k}\delta\boldsymbol\beta)'.$$

Its variance is

(3) $$\begin{aligned} \mathrm{var}(h(\tilde{\hat{\boldsymbol\beta}})) &= \mathbf{h}'\mathbf{C}^{-1}\mathbf{h} + 2\,\mathrm{Tr}(\mathbf{A}_h\mathbf{C}^{-1}\mathbf{A}_h\mathbf{C}^{-1}) \\ &\quad + 4E(\delta\hat{\boldsymbol\beta})'\mathbf{A}_h\mathbf{C}^{-1}\mathbf{A}_h E(\delta\hat{\boldsymbol\beta}) - 4\mathbf{h}'\mathbf{C}^{-1}\mathbf{A}_h E(\delta\hat{\boldsymbol\beta}) \end{aligned}$$

(cf. Lemma 2.1).

The MSE of the linear estimator is

(4) $$\mathrm{MSE}(\mathbf{h}'\hat{\boldsymbol\beta}) = \mathrm{var}(\mathbf{h}'\delta\hat{\boldsymbol\beta}) + b_h^2(\hat{\boldsymbol\beta}) = \mathbf{h}'\mathbf{C}^{-1}\mathbf{h} + (\delta\boldsymbol\beta'\mathbf{A}_h\delta\boldsymbol\beta)^2.$$

The MSE of the quadratic estimator is

(5) $$\begin{aligned} \mathrm{MSE}(\mathbf{h}'\tilde{\hat{\boldsymbol\beta}}) &= \mathrm{var}(h(\tilde{\hat{\boldsymbol\beta}})) + b_h^2(\tilde{\hat{\boldsymbol\beta}}) \\ &= \mathbf{h}'\mathbf{C}^{-1}\mathbf{h} + 2\,\mathrm{Tr}(\mathbf{A}_h\mathbf{C}^{-1}\mathbf{A}_h\mathbf{C}^{-1}) \\ &\quad + 4E(\delta\hat{\boldsymbol\beta})'\mathbf{A}_h\mathbf{C}^{-1}\mathbf{A}_h E(\delta\hat{\boldsymbol\beta}) - 4\mathbf{h}'\mathbf{C}^{-1}\mathbf{A}_h E(\delta\hat{\boldsymbol\beta}) \\ &\quad + (2\mathbf{b}'\mathbf{A}_h\delta\boldsymbol\beta + \mathbf{b}'\mathbf{A}_h\mathbf{b})^2. \end{aligned}$$

To find a region of shifts $\delta\boldsymbol\beta$ where (5) $\leqslant$ (4) is tedious for a large k (dimension of $\boldsymbol\beta$). A rule for the first orientation is given in (6).

3. Main results

If the dimension k of the parameter $\boldsymbol{\beta}$ is large, then the investigation of the quantities $\mathbf{b}(\hat{\boldsymbol{\beta}})$, $\mathbf{b}(\tilde{\hat{\boldsymbol{\beta}}})$, $\mathrm{MSE}(\hat{\beta}_i)$, $i = 1, \ldots, k$, and $\mathrm{MSE}(\tilde{\hat{\beta}}_i)$, $i = 1, \ldots, k$, in this way is tedious, since the number of different directions of $\delta\boldsymbol{\beta}$ necessary for an investigation may be huge. Thus it seems that the following theorem can be useful in practice.

Theorem 3.1. *The following inequalities are valid:*

$$
\begin{aligned}
|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}},\\
|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}},\\
|\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant \sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant \mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|4\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|4\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\,\mathrm{Tr}[(\overline{\mathbf{B}}_i\mathbf{C}^{-1})^2]\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b}| &\leqslant 2\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b}| &\leqslant 2\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\\
|-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant \sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{(3/2)},\\
|-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant \mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{(3/2)},\\
|8\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{(3/2)},\\
|8\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &\leqslant 4\,\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{(3/2)},\\
|\mathbf{b}'\mathbf{B}_{0,i}\mathbf{b}| &\leqslant \frac{1}{4}\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2,\\
|\mathbf{b}'\mathbf{B}_{0,i}\mathbf{b}| &\leqslant \frac{1}{4}\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2,\\
|4\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b}| &\leqslant \sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2,\\
|4\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b}| &\leqslant \mathrm{Tr}[(\overline{\mathbf{B}}_i\mathbf{C}^{-1})^2][K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2.
\end{aligned}
$$

Obviously the right-hand sides depend on the quantity independent of the direction of the shift $\delta\boldsymbol{\beta}$. They depend on the Mahalanobis distance $\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}}$ only.

P r o o f. The Schwarz inequality and the relation implied by the definition of the quantity $K^{(\mathrm{par})}(\boldsymbol{\beta}_0)$, i.e.

$$\mathbf{b}'\mathbf{C}\mathbf{b} \leqslant \frac{1}{4}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},$$

will be used. We have

$$\begin{aligned}|\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &= |\delta\boldsymbol{\beta}'\mathbf{C}^{1/2}\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \\ &\leqslant \delta\boldsymbol{\beta}'\mathbf{C}^{1/2}\sum_{i=1}^{k}|\lambda_i|\mathbf{f}_i\mathbf{f}_i'\mathbf{C}^{1/2}\delta\boldsymbol{\beta} = \sum_{i=1}^{k}|\lambda_i|(\mathbf{f}'\mathbf{C}^{1/2}\delta\boldsymbol{\beta})^2 \\ &\leqslant \sum_{i=1}^{k}|\lambda_i|\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} = \operatorname{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\end{aligned}$$

where $\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2} = \sum_{i=1}^{k}\lambda_i\mathbf{f}_i\mathbf{f}_i'$ is the spectral decomposition of the matrix $\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}$.

Another procedure is

$$\begin{aligned}|\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &= |\delta\boldsymbol{\beta}\mathbf{C}^{1/2}\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \\ &= |\operatorname{Tr}(\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\delta\boldsymbol{\beta}'\mathbf{C}^{1/2})| \\ &\leqslant \sqrt{\operatorname{Tr}(\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2})}\sqrt{\operatorname{Tr}[(\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\delta\boldsymbol{\beta}'\mathbf{C}^{1/2})^2]} \\ &= \sqrt{\operatorname{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta},\end{aligned}$$

$$\begin{aligned}|-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &= 2|\mathbf{b}'\mathbf{C}^{1/2}\sum_{i=1}^{k}\lambda_i\mathbf{f}_i\mathbf{f}_i'\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \leqslant 2\sum_{i=1}^{k}|\lambda_i|\,|\mathbf{b}'\mathbf{C}^{1/2}\mathbf{f}_i|\,|\mathbf{f}_i'\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \\ &\leqslant 2\sum_{i=1}^{k}|\lambda_i|\sqrt{\mathbf{b}'\mathbf{C}\mathbf{b}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} = 2\operatorname{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\mathbf{b}'\mathbf{C}\mathbf{b}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} \\ &\leqslant \operatorname{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2}.\end{aligned}$$

Another procedure is

$$\begin{aligned}|-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta}| &= 2|\mathbf{b}'\mathbf{C}^{1/2}\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\delta\boldsymbol{\beta}| \\ &= 2|\operatorname{Tr}(\mathbf{C}^{-1/2}\mathbf{B}_{0,i}\mathbf{C}^{-1/2}\mathbf{C}^{1/2}\mathbf{b}'\delta\boldsymbol{\beta}\mathbf{C}^{1/2})| \\ &\leqslant 2\sqrt{\operatorname{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\operatorname{Tr}(\mathbf{C}^{1/2}\delta\boldsymbol{\beta}\mathbf{b}'\mathbf{C}^{1/2}\mathbf{C}^{1/2}\mathbf{b}'\delta\boldsymbol{\beta}\mathbf{C}^{1/2})} \\ &= 2\sqrt{\operatorname{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\mathbf{b}'\mathbf{C}\mathbf{b}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} \\ &\leqslant 2\sqrt{\operatorname{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\frac{1}{2}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2}.\end{aligned}$$

The other inequalities can be proved analogously. □

The term $-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}$ is dominant when $\mathrm{MSE}(\hat{\beta}_i)$ and $\mathrm{MSE}(\tilde{\tilde{\beta}}_i)$ is compared in the small neighbourhood of the point $\boldsymbol{\beta}_0$.

Lemma 3.1.
(i)

$$\max\{|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}|\colon\ \delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}=c^2\}=\left|-4c\frac{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}{\sqrt{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{i,i}\mathbf{C}^{-1}\}_{i,i}}}\right|.$$

(ii) *The equality*

$$|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}|=(\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta})^2$$

in the direction of the vector $\{\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{\cdot,i}$ *is attained for*

$$\delta\boldsymbol{\beta}=\frac{c}{\sqrt{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}}\{\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{\cdot,i},$$

where

$$c^3=4\frac{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}^2\mathbf{C}^{-1}\}_{i,i}}{[\{\mathbf{C}^{-1}\mathbf{B}_{0,i}^3\mathbf{C}^{-1}\}_{i,i}]^2}\sqrt{[\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}]^3}.$$

P r o o f. The gradient of the function

$$f(\delta\boldsymbol{\beta})=-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta},\quad \delta\boldsymbol{\beta}\in\mathbb{R}^k,$$

is $-4\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{e}_i$ at the point $\boldsymbol{\beta}=\boldsymbol{\beta}_0$. If the vector $\delta\boldsymbol{\beta}$, satisfying the equality $c^2=\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}$, is directed as the gradient, then obviously the function $f(\cdot)$ attains its maximum. Further procedure is evident. □

The distance c from the last lemma can be compared with the value

$$\sqrt{\chi_k^2(1-\alpha)}$$

(the boundary of the confidence ellipsoid). The shift $\delta\boldsymbol{\beta}$ in the direction of the gradient which attains the boundary of the confidence ellipsoid is

$$\delta\boldsymbol{\beta}=\sqrt{\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}}\{\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{\cdot,i}.$$

R e m a r k 3.1. If the bias of the linear estimator brought to the square is smaller than the term $|-4\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta}|$ for the above mentioned $\delta\boldsymbol{\beta}$, then the quadratic correction is of no use. Thus the following rule for the first orientation can be used.

If

$$\text{(6)} \qquad \left(\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}\right)^2(\{\mathbf{C}^{-1}\mathbf{B}_{0,i}^3\mathbf{C}^{-1}\}_{i,i})^2 \leqslant 4\{\mathbf{C}^{-1}\mathbf{B}_{0,i}^2\mathbf{C}^{-1}\}_{i,i}\sqrt{\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{C}\mathbf{B}_{0,i}\mathbf{C}^{-1}\}_{i,i}}},$$

the linear estimator is to be preferred. If the opposite inequality occurs, then some more detailed investigation should be performed.

Remark 3.2. If $\mathbf{C}^{-1} = \sigma^2\mathbf{C}_0^{-1}$, then the last inequality can be written in the form

$$\text{(7)} \qquad \sigma^4\left(\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}_0^{-1}\mathbf{B}_{0,i}\mathbf{C}_0\mathbf{B}_{0,i}\mathbf{C}_0^{-1}\}_{i,i}}\right)^2(\{\mathbf{C}_0^{-1}\mathbf{B}_{0,i}^3\mathbf{C}_0^{-1}\}_{i,i})^2 \leqslant 4\{\mathbf{C}_0^{-1}\mathbf{B}_{0,i}^2\mathbf{C}_0^{-1}\}_{i,i}\sqrt{\frac{\chi_k^2(1-\alpha)}{\{\mathbf{C}_0^{-1}\mathbf{B}_{0,i}\mathbf{C}_0\mathbf{B}_{0,i}\mathbf{C}_0^{-1}\}_{i,i}}}.$$

The role played by the parameter σ is now quite obvious.

Remark 3.3. The correction term in the case $h(\boldsymbol{\beta}) = \beta_i$ is

$$\tau_{0,i} = -\delta\hat{\boldsymbol{\beta}}'\mathbf{B}_{0,i}\delta\hat{\boldsymbol{\beta}} + \operatorname{Tr}(\mathbf{B}_{0,i}\mathbf{C}^{-1}).$$

If $\mathbf{B}_{0,i} = \overline{\mathbf{B}}_i$, then according to [10], [13], [15] it can be approximated by the random variable

$$\tau_{0,i} \approx -c_i^2\chi_{f_i}^2(0) + \operatorname{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1}),$$

where

$$\begin{aligned} c_i^2 &= V_i/E_i, \quad f_i = E_i^2/V_i, \\ E_i &= \delta\boldsymbol{\beta}'\overline{\mathbf{B}}_i\delta\boldsymbol{\beta} + 2\mathbf{b}'\overline{\mathbf{B}}_i\delta\boldsymbol{\beta} + \mathbf{b}'\overline{\mathbf{B}}_i\mathbf{b}, \\ V_i &= 2\operatorname{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1}\overline{\mathbf{B}}_i\mathbf{C}^{-1}) + 4(\delta\boldsymbol{\beta}+\mathbf{b})'\overline{\mathbf{B}}_i\mathbf{C}^{-1}\overline{\mathbf{B}}_i(\delta\boldsymbol{\beta}+\mathbf{b}). \end{aligned}$$

If $\mathbf{B}_{0,i}$ is not positive semidefinite, then the distribution function of $\tau_{0,i}$ cannot be obtained so simply (cf. [2]).

Remark 3.4. If a function $h(\boldsymbol{\beta}) = \mathbf{h}'\boldsymbol{\beta}$, $\boldsymbol{\beta} \in \mathbb{R}^k$, is under consideration, then the matrix $\mathbf{B}_{0,h}$ is used instead of $\mathbf{B}_{0,i}$. In the case of a function

$$h(\boldsymbol{\beta}) = h_0 + \mathbf{h}'\delta\boldsymbol{\beta} + \frac{1}{2}\delta\boldsymbol{\beta}'\mathbf{H}_1\delta\boldsymbol{\beta},$$

the matrix $\mathbf{A}_h$ must be used.

4. Example

In the first step let us investigate the bias and the MSE of the linear and of the quadratic estimator, respectively.

The Michaelis-Menten model is under consideration; i.e.

$$f(x;\beta_1,\beta_2) = \frac{\beta_1 x}{\beta_2 + x}.$$

Let $(\beta_{1,0},\beta_{2,0})' = (5,1)'$ and

x	1	2	3	4	5	6
$f(x;\beta_{1,0},\beta_{2,0})$	2.5	3.33	3.75	4	4.17	4.29

If the observation vector is $\mathbf{Y} \sim N_6(\mathbf{f}(\cdot,\boldsymbol{\beta}),\sigma^2\mathbf{I})$ and $\sigma = 0.1$, then

$$K^{(\mathrm{int})}(\boldsymbol{\beta}_0) = 0.025443, \quad K^{(\mathrm{par})}(\boldsymbol{\beta}_0) = 0.090940.$$

Thus the intrinsic nonlinearity can be neglected (in detail cf. [14]) and the linearization regions with respect to the bias of the whole vector $\boldsymbol{\beta}$ and the single parameters β_1 and β_2 are given in Figs. 4.1–4.3, respectively.

In the figures seven numbers are given; five white numbers are connected with the white ellipse (linearization region). They give the maximum coordinates of the ellipse points, a size of the raster rectangulars and the step in the first coordinate used for the construction of the ellipse. Two dark numbers have the analogous meaning for the dark ellipse (0.95-confidence ellipse).

In this situation the linearization for the functions $h_1(\boldsymbol{\beta}) = \beta_1$ or $h_2(\boldsymbol{\beta}) = \beta_2$, respectively, is possible, even if the confidence ellipse for the vector parameter is not essentially smaller than the linearization region (cf. Fig. 4.1). Thus it is interesting whether the quadratic estimator is not better.

As far as the parameter β_1 is concerned the shift $\delta\boldsymbol{\beta} = (0,\delta\beta_2)'$ is dangerous (cf. Fig. 4.2). Let $\delta\boldsymbol{\beta}$ be chosen on the boundary of the 0.95-confidence ellipse, i.e. $\delta\beta_2 = 0.1076$. Thus we obtain

$$b(\hat{\beta}_1) = -0.004469 \quad \text{and} \quad b(\hat{\beta}_2) = -0.007042,$$

i.e. $\mathbf{b} = (-0.004469, -0.007042)'$. The bias $b(\tilde{\tilde{\beta}}_1)$ is

$$b(\tilde{\tilde{\beta}}_1) = -2\mathbf{b}'\mathbf{B}_{0,1}\delta\beta + \mathbf{b}'\mathbf{B}_{0,1}\mathbf{b} = -0.000604,$$

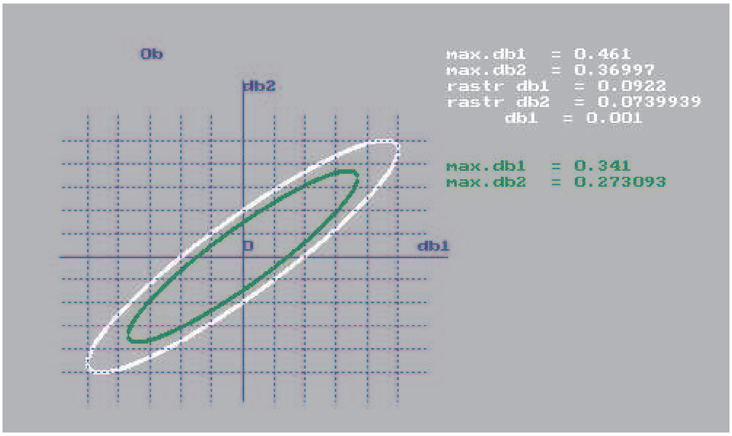

Figure 4.1 Linearization region for the whole vector $\boldsymbol{\beta}$.

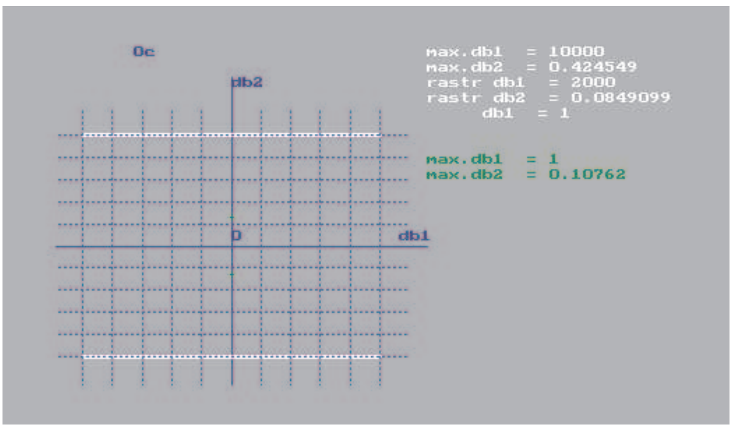

Figure 4.2. Linearization region for β_1.

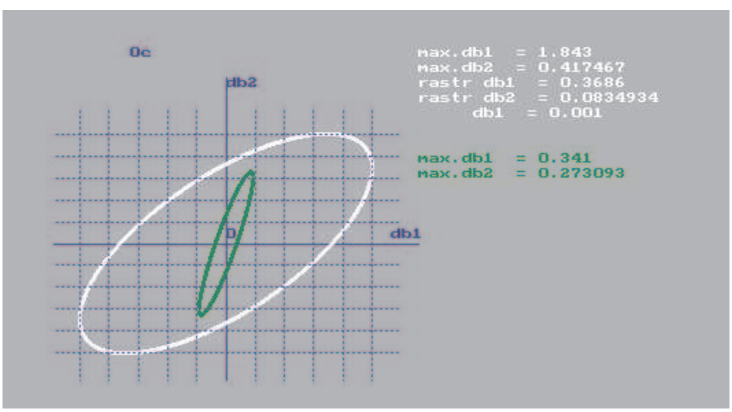

Figure 4.3. Linearization region for β_2.

which is significantly smaller (in the absolute value) than $b(\hat{\beta}_1) = -0.004469$. Even for $\delta\boldsymbol{\beta} = 3 \times (0, 0.1076)'$ the situation is analogous:

$$\mathbf{b}(\hat{\boldsymbol{\beta}}) = (-0,040224, -0.063374)', \quad b(\tilde{\tilde{\beta}}_1) = -0.017342.$$

(The values $\sqrt{\text{var}(\hat{\beta}_1)}$ and $\sqrt{\text{var}(\hat{\beta}_2)}$ are 0.139 and 0.125, respectively.)

The bias $b(\tilde{\tilde{\beta}}_2)$ due to the shift $\delta\boldsymbol{\beta} = (0, 0.1076)'$ and $\delta\boldsymbol{\beta} = 3 \times (0, 0.1076)'$ is -0.000849 and -0.24219, respectively.

Further we have ($\delta\boldsymbol{\beta} = (0, 0.1076)'$)

$$\text{var}(\hat{\beta}_1) = 0.01936, \quad \text{var}(\tilde{\tilde{\beta}}_1) = 0.02170,$$
$$\text{var}(\hat{\beta}_2) = 0.01245, \quad \text{var}(\tilde{\tilde{\beta}}_2) = 0.01511,$$

and

$$\text{MSE}(\hat{\beta}_1) = 0.01938, \quad \text{MSE}(\tilde{\tilde{\beta}}_1) = 0.02170,$$
$$\text{MSE}(\hat{\beta}_2) = 0.01250, \quad \text{MSE}(\tilde{\tilde{\beta}}_2) = 0.01511.$$

If $\delta\boldsymbol{\beta} = 3 \times (0,\ 0.1076)'$, then

$$\text{MSE}(\hat{\beta}_1) = 0.02098, \quad \text{MSE}(\tilde{\tilde{\beta}}_1) = 0.02622,$$
$$\text{MSE}(\hat{\beta}_2) = 0.01646, \quad \text{MSE}(\tilde{\tilde{\beta}}_2) = 0.02005.$$

If we denote $\delta\boldsymbol{\beta} = \begin{pmatrix} 0 \\ x \end{pmatrix}$, we obtain

$$\begin{aligned}
\text{MSE}(\hat{\beta}_1) &= 0.01936 + 0.148981x^4, \\
\text{MSE}(\tilde{\tilde{\beta}}_1) &= 0.019406 + 0.022028x - 0.005977x^2 - 0.009023x^3 \\
&\quad + 0.002744x^4 + (0.469452x^3 + 0.142744x^4)^2, \\
\text{MSE}(\hat{\beta}_2) &= 0.012448 + 0.369822x^4, \\
\text{MSE}(\tilde{\tilde{\beta}}_2) &= 0.012493 + 0.024572x - 0.000777x^2 - 0.012996x^3 \\
&\quad + 0.003449x^4 + (0.662448x^3 + 0.177955x^4)^2,
\end{aligned}$$

cf. Table 1.

x	MSE($\hat{\beta}_1$)	MSE($\tilde{\tilde{\beta}}_1$)	MSE($\hat{\beta}_2$)	MSE($\tilde{\tilde{\beta}}_2$)
0.03	0.01936	0.02006	0.01245	0.01323
0.1	0.01938	0.02154	0.01249	0.01493
0.2	0.01960	0.02352	0.01304	0.01727
0.3	0.02057	0.02545	0.01544	0.01984
0.4	0.02317	0.02789	0.02192	0.02366
0.5	0.02867	0.03254	0.03556	0.03200
0.6	0.03867	0.04325	0.06038	0.05220
0.7	0.05513	0.06760	0.10124	0.09855

Table 1.

It is obvious that the quadratic corrections in this case have no sense.
The inequalities from Theorem 3.1 for the parameter β_1 (β_2) are

$$\begin{aligned}
-4\mathbf{e}_1'\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta} &= \quad 0.00237\ (0.00264),\\
4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} &= \quad 0.00654\ (0.00524),\\
4\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}\sqrt{\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta}} &= \quad 0.00654\ (0.00598),
\end{aligned}$$

$$\begin{aligned}
\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta} &= -\,0.00447\ (-0.00704),\\
\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= \quad 0.02878\ (0.02872),\\
\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= \quad 0.02878\ (0.03279),
\end{aligned}$$

$$\begin{aligned}
4\delta\boldsymbol{\beta}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,1}\delta\boldsymbol{\beta} &= \quad 0.00009\ (0.00014),\\
4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= \quad 0.00055\ (0.00055),\\
4\,\mathrm{Tr}[\overline{\mathbf{B}}_i\mathbf{C}^{-1})^2]\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= \quad 0.00055\ (0.00072),
\end{aligned}$$

$$\begin{aligned}
-4\mathbf{e}_1'\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b} &= -\,0.00015\ (-0.00015),\\
2\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= \quad 0.00073\ (0.00058),\\
2\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})\sqrt{\{\mathbf{C}^{-1}\}_{i,i}}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta} &= \quad 0.00073\ (0.00067),
\end{aligned}$$

$$\begin{aligned}
-2\mathbf{b}'\mathbf{B}_{0,i}\delta\boldsymbol{\beta} &= -\,0.00058\ (-0.00082),\\
\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2} &= \quad 0.00641\ (0.00639),\\
\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})K^{(\mathrm{par})}(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2} &= \quad 0.00641\ (0.00730),
\end{aligned}$$

$$8\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\delta\boldsymbol{\beta} = -\,0.00001\ (-0.00002),$$
$$4\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}K^{\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2} = 0.00012\ (0.00012),$$
$$4\,\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]K^{(\mathrm{par})}(\boldsymbol{\beta}_0)(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^{3/2} = 0.00012\ (0.00016),$$
$$\mathbf{b}'\mathbf{B}_{0,i}\mathbf{b} = -\,0.00002\ (-0.00002),$$
$$\frac{1}{4}\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^2]}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2 = 0.00036\ (0.00036),$$
$$\frac{1}{4}\,\mathrm{Tr}(\overline{\mathbf{B}}_i\mathbf{C}^{-1})[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2 = 0.00036\ (0.00041),$$
$$4\mathbf{b}'\mathbf{B}_{0,i}\mathbf{C}^{-1}\mathbf{B}_{0,i}\mathbf{b} = 0.00000\ (0.00000),$$
$$\sqrt{\mathrm{Tr}[(\mathbf{B}_{0,i}\mathbf{C}^{-1})^4]}[K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2 = 0.00001\ (0.00001),$$
$$\mathrm{Tr}[(\overline{\mathbf{B}}_i\mathbf{C}^{-1})^2][K^{(\mathrm{par})}(\boldsymbol{\beta}_0)]^2(\delta\boldsymbol{\beta}'\mathbf{C}\delta\boldsymbol{\beta})^2 = 0.00001\ (0.00001).$$

Here the index i on the left-hand side means 1 for the first number on the right-hand side and 2 for the second (in the bracket).

In many cases the upper bound is significantly larger than the actual value (by virtue of the Schwarz inequality). Nevertheless, some information on the individual terms can be obtained in this way:

$$\mathbf{C}_0 = \begin{pmatrix} 3.326\,1, & -3.812\,4 \\ -3.812\,4, & 5.173\,1 \end{pmatrix},$$
$$\mathbf{C}_0^{-1} = \begin{pmatrix} 1.936, & 1.427 \\ 1.427, & 1.246 \end{pmatrix},$$
$$\mathbf{B}_{0,2} = \begin{pmatrix} 0, & 0.100\,00 \\ 0.100\,00, & -0.608\,13 \end{pmatrix},$$

$$\left(\frac{\chi_2^2(0;0.95)}{\{\mathbf{C}_0^{-1}\mathbf{B}_{0,2}\mathbf{C}_0\mathbf{B}_{0,2}\mathbf{C}_0^{-1}\}_{i,i}}\right)^2(\{\mathbf{C}_0^{-1}\mathbf{B}_{0,2}^3\mathbf{C}_0^{-1}\}_{2,2})^2 = 0.3214,$$
$$4\{\mathbf{C}_0^{-1}\mathbf{B}_{0,2}^2\mathbf{C}_0^{-1}\}_{2,2}\sqrt{\frac{\chi_2^2(0;0.95)}{\{\mathbf{C}_0^{-1}\mathbf{B}_{0,2i}\mathbf{C}_0\mathbf{B}_{0,2}\mathbf{C}_0^{-1}\}_{2,2}}} = 2.3918.$$

The left-hand side of (7) is smaller than the right-hand side of (7) even for $\sigma = 1.65$. In this case
$$K^{(\mathrm{int})}(\boldsymbol{\beta}_0) = 0.4198, \quad K^{(\mathrm{par})}(\boldsymbol{\beta}_0) = 1.5005.$$

An experiment characterized by these values would be extremely badly planned. Thus the Michaelis-Menten model can be linearized at the considered point $\boldsymbol{\beta}_0 = (5,1)'$ under a sufficiently small value of σ only. The quadratic corrections are of no use (cf. also Tab. 1). If a sufficiently small σ cannot be attained, then the methods given in [11], [12] must be used.

References

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

[2] *J. P. Imhof*: Computing the distribution of quadratic forms in normal variables. Biometrika *48* (1961), 419–426.

[3] *A. Jenčová*: A comparison of linearization and quadratization domains. Appl. Math. *42* (1997), 279–291.

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

[5] *L. Kubáček, L. Kubáčková, J. Volaufová*: Statistical Models with Linear Structures. Veda, Bratislava, 1995.

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

[7] *L. Kubáček*: Quadratic regression models. Math. Slovaca *46* (1996), 111–126.

[8] *L. Kubáček*: Corrections of estimators in linearized models. Acta Univ. Palack. Olomuc., Fac. Rerum Math. *37* (1998), 69–80.

[9] *L. Kubáček, L. Kubáčková*: Regression models with a weak nonlinearity. Technical Reports. University of Stuttgart.

[10] *P. B. Patnaik*: The non-central χ^2 and F-distributions and their applications. Biometrika *36* (1949), 202–232.

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

[12] *R. Potocký, To Van Ban*: Confidence regions in nonlinear regression models. Appl. Math. *37* (1992), 29–39.

[13] *F. E. Satterthwaite*: An approximate distribution of estimates of variance components. Biometrics Bulletin *2* (1946), 110-114.

[14] *E. Tesaříková, L. Kubáček*: How to deal with regression models with a weak nonlinearity. Discuss. Math. Probab. Stat. *21* (2001), 21–48.

[15] *B. L. Welch*: The generalization of Student's problem when several different population variances are involved. Biometrika *34* (1947), 28–35.

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