VERSITA

Mathematica Slovaca

DOI: 10.2478/s12175-011-0079-1
Math. Slovaca **62** (2012), No. 1, 143–155

A NOTE ON REDUCTION OF THE NUMBER OF PARAMETERS IN LINEAR STATISTICAL MODELS

Lubomír Kubáček

(*Communicated by Gejza Wimmer*)

ABSTRACT. In certain settings the mean response is modeled by a linear model using a large number of parameters. Sometimes it is desirable to reduce the number of parameters prior to conducting the experiment and prior to the actual statistical analysis. Essentially, it means to formulate a simpler approximate model to the original "ideal" one. The goal is to find conditions (on the model matrix and covariance matrix) under which the reduction does not influence essentially the data fit. Here we try to develop such conditions in regular linear model without and with linear restraints. We emphasize that these conditions are independent of observed data.

1. Introduction

A choice of proper number of parameters in a regression model depeds on actual conditions. Physical, biological, economical, etc., theory of the investigated object can be well elaborated and thus the number of parameters of a regression model is known in advance (cf. the beginning of section 4). Here it can occur that some parameters are nuisance, or they are not important for the observer. Thus it is reasonable to investigate under which conditions these parameters can be neglected.

Another situation sets in, when the theory of the investigated object is not developed sufficiently, e.g. the regression function must be chosen as a polynomial of the unknown order. However, only a small subset, several first terms of the model can be interpreted properly within the discipline (e.g. in medicine). Thus

2010 Mathematics Subject Classification: Primary 62J05.
Keywords: linear statistical model, reduction of number of parameters.
Supported by the Council of the Czech Governement MSM 6 198 959 214.

also in this case it is reasonable to solve the problem under which conditions the model can be reduced.

The problem of the reduction of the number of parameters can arise in other situations as well. The theoretical model (physical, biological, etc.) can be too complicated and in an actual situation some parameters are not important, however the useful parameters must be estimated. Problems of this type are investigated and solution suggested in case of regular models in, e.g., [1], [3], [4]. Similar problems are solved in [6].

Another problem sets in, when nuisance parameters have to be eliminated by a transformation of the data, however no loss of information on useful parameters can occur (in more detail cf. [2]).

The aim of the paper is to contribute to a solution of the problem in regular linear models without/with constraints on parameters.

2. Notation and preliminaries

The notation

$$\boldsymbol{Y} \sim N_n\left[(\mathbf{X}, \mathbf{S})\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}, \boldsymbol{\Sigma}\right] \tag{1}$$

means that $\boldsymbol{Y}$ is assumed to be normally distributed n-dimensional random vector with the mean value $E(\boldsymbol{Y})$ equal to $(\mathbf{X}, \mathbf{S})\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}$ and with the covariance matrix $\mathrm{Var}(\boldsymbol{Y})$ equal to $\boldsymbol{\Sigma}$. The covariance matrix is assumed to be known and positive definite. Its structure can be $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$, where $\sigma^2 \in (0, \infty)$ and $\mathbf{V}$ is given.

$\boldsymbol{\beta}$	...	useful vector parameter,
$\boldsymbol{\gamma}$	...	nuisance vector parameter,
$\chi^2_f(0)$	...	central chi-squared random variable with f degrees of freedom,
$\chi^2_f(\delta)$	...	non-centeral chi-squared random variable with the non-centrality parameter equal to δ and with f degrees of freedom,
$\chi^2_f(0; 1-\alpha)$	...	$(1-\alpha)$-quantile of the central chi-squared distribution with f degrees of freedom,
$\widehat{\boldsymbol{\beta}}_{\text{full}}$	...	BLUE (best linear unbiased estimator) of the vector parameter $\boldsymbol{\beta}$ in the full model,
$\widehat{\widehat{\boldsymbol{\beta}}}_{\text{full}}$	...	BLUE of the vector parameter $\boldsymbol{\beta}$ in the full model with constraints,

$\widehat{\widehat{\boldsymbol{\beta}}}_{\mathrm{red}}$ ... BLUE of the vector parameter $\boldsymbol{\beta}$ in the reduced model with constraints.

In the following text it is assumed that the model is regular, i.e. the rank of the matrix $(\mathbf{X}, \mathbf{S})$ is $r(\mathbf{X}, \mathbf{S}) = k + l \leq n$ and $\boldsymbol{\Sigma}$ is positive definite. The model (1) is assumed to be the full model.

Lemma 1. *In the model* (1) *the BLUE of the vector parameter* $\begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix}$ *is*

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_{\mathrm{full}} \\ \widehat{\gamma}_{\mathrm{full}} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{1,2}, & \mathbf{A}_{1,2} \\ \mathbf{A}_{2,1}, & \mathbf{A}_{2,2} \end{pmatrix} \begin{pmatrix} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \\ \mathbf{S}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{A}_{1,1} &= \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{S}\big[\mathbf{S}'(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^{+}\mathbf{S}\big]^{-1}\mathbf{S}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{C}^{-1}, \\
\mathbf{A}_{1,2} &= -\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{S}\big[\mathbf{S}'(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^{+}\mathbf{S}\big]^{-1} = \mathbf{A}'_{2,1}, \\
\mathbf{A}_{2,2} &= \big[\mathbf{S}'(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^{+}\mathbf{S}\big]^{-1}, \qquad \mathbf{C} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}, \\
\mathbf{M}_X &= \mathbf{I} - \mathbf{X}\mathbf{X}^{+}, \qquad (\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^{+} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}
\end{aligned}$$

($^{+}$ *means the Moore–Penrose generalized inverse of the matrix; in more detail cf.* [8]).

P r o o f. It is sufficient to take into account the relationship

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\gamma} \end{pmatrix} = \left[\begin{pmatrix} \mathbf{X}' \\ \mathbf{S}' \end{pmatrix} \boldsymbol{\Sigma}^{-1}(\mathbf{X}, \mathbf{S})\right]^{-1} \begin{pmatrix} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \\ \mathbf{S}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \end{pmatrix},$$

the Rohde equality (cf. also [1, p. 344])

$$\begin{aligned}
&\begin{pmatrix} \mathbf{A}, & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1}, & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}^{-1}, & (\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix},
\end{aligned}$$

which is valid for any positive definite matrix $\begin{pmatrix} \mathbf{A}, & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{pmatrix}$ and the equality

$$(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^{+} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$$

(cf. [1, p. 346]). □

Lemma 2. *In the model* (1) *it is valid that*

$$Q(\widehat{\boldsymbol{\beta}}, \widehat{\gamma}) = \left[\boldsymbol{Y} - (\mathbf{X}, \mathbf{S})\begin{pmatrix} \widehat{\boldsymbol{\beta}}_{\mathrm{true}} \\ \widehat{\gamma}_{\mathrm{true}} \end{pmatrix}\right]' \boldsymbol{\Sigma}^{-1} \left[\boldsymbol{Y} - (\mathbf{X}, \mathbf{S})\begin{pmatrix} \widehat{\boldsymbol{\beta}}_{\mathrm{true}} \\ \widehat{\gamma}_{\mathrm{true}} \end{pmatrix}\right] \sim \chi^2_{n-k-l}(0).$$

Proof is well known and therefore it is omitted (cf. [7, pp. 223–225]).

The inequality $Q(\widehat{\boldsymbol{\beta}}, \widehat{\gamma}) < \chi^2_{n-k-l}(0; 1-\alpha)$, for sufficiently small $\alpha > 0$, means that hypothesis "the model is true for the data" cannot be rejected.

3. Criteria for a reduction of the number of parameters

Let the reduced model be of the form

$$\boldsymbol{Y} \sim N_n(\mathbf{X}\boldsymbol{\kappa}, \boldsymbol{\Sigma}), \tag{2}$$

where symbols used have the same meaning as in the model (1), i.e. the matrices $\mathbf{X}$ and $\boldsymbol{\Sigma}$ in (1) and 2) are the same. The new vector $\mathbf{X}\boldsymbol{\kappa}$ should be an "approximation" of $\mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\gamma}$.

In view of Lemma 2 the agreement of the reduced model (2) with the data $\boldsymbol{Y}$ can be judged by the help of the random variable

$$(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}}), \tag{3}$$

where $\widehat{\boldsymbol{\kappa}} = \mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$.

Lemma 3. *The random variable* (3) *is distributed in the model* (1) *as* $\chi^2_{n-k}(\delta)$, *where* δ *is given as*

$$\delta = \boldsymbol{\gamma}'\mathbf{S}(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+\mathbf{S}\boldsymbol{\gamma}.$$

P r o o f. Regarding the Ogasawara-Takahashi theorem (cf. [7, p. 222]), it is valid that

$$(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}}) = \boldsymbol{Y}'(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+\boldsymbol{Y} \sim \chi^2_{r(M_X)}(\delta),$$

where

$$\begin{aligned}\delta &= E(\boldsymbol{Y}')(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+E(\boldsymbol{Y})\\ &= (\boldsymbol{\beta}', \boldsymbol{\gamma}')\begin{pmatrix}\mathbf{X}'\\ \mathbf{S}'\end{pmatrix}(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+(\mathbf{X}, \mathbf{S})\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}\\ &= \boldsymbol{\gamma}'\mathbf{S}'(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+\mathbf{S}\boldsymbol{\gamma}.\end{aligned}$$

□

One way how to decide whether the model (1) can be approximated by the reduced one (2) is to judge whether the inequality

$$P\left\{\frac{(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})}{n-k} \le \frac{\chi^2_{n-k-l}(0; 1-\alpha)}{n-k-l}\right\} \ge 1 - \alpha - \varepsilon$$

is satisfied for sufficiently small real number $\varepsilon > 0$; e.g. $\alpha = 0.05$, $\varepsilon = 0.04$. Here ε is admissible (with respect to the opinion of a statistician) enlargement of the level of the test on agreement between data and the reduced model.

This inequality can be rewritten as follows. Since the statistic

$$(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})$$

is distributed as $\chi^2_{n-k}(\delta)$, we can write the inequality in the form

$$P\left\{\frac{\chi^2_{n-k}(\delta)}{n-k} \le \frac{\chi^2_{n-k-l}(0;1-\alpha)}{n-k-l}\right\} \ge 1-\alpha-\varepsilon.$$

A simple rule how to decide whether this condition is satisfied is given by the following theorem.

Theorem 1. *If in the model* (1) *the parameter* $\boldsymbol{\gamma}$ *satisfies the inequality*

$$\boldsymbol{\gamma}'\mathbf{S}(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+\mathbf{S}\boldsymbol{\gamma} \le \delta_{\max},$$

then the condition is satisfied and thus the agreement of the reduced model (2) *with the model* (1) *is sufficient. Here* $\delta_{\max}$ *is a solution of the equation*

$$P\left\{c\chi^2_h(0) \le \frac{n-k}{n-k-l}\chi^2_{n-k-l}(0;1-\alpha)\right\} = 1-\alpha-\varepsilon,$$

$$c = \frac{n-k+2\delta_{\max}}{n-k+\delta_{\max}}, \qquad h = \frac{(n-k+\delta_{\max})^2}{n-k+2\delta_{\max}}.$$

P r o o f. Since non-central chi-squared random variable $\chi^2_f(\delta)$ with f degrees of freedom and with the non-centrality parameter δ can be approximated by $c\chi^2_h(0)$ (cf. [5]), where

$$c = \frac{f+2\delta}{f+\delta}, \qquad h = \frac{(f+\delta)^2}{f+2\delta},$$

it is valid that

$$P\left\{\frac{\chi^2_{n-k}(\delta)}{n-k} \le \frac{\chi^2_{n-k-l}(0;1-\alpha)}{n-k-l}\right\} \approx P\left\{c\chi^2_h(0) \le \frac{n-k}{n-k-l}\chi^2_{n-k-l}(0;1-\alpha)\right\}.$$

According Lemma 3 $\delta = \boldsymbol{\gamma}'\mathbf{S}(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+\mathbf{S}\boldsymbol{\gamma}$ and $\delta \le \delta_{\max}$ implies

$$\begin{aligned}
&P\left\{\frac{\chi^2_{n-k}\left[\boldsymbol{\gamma}'\mathbf{S}'(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+\mathbf{S}\boldsymbol{\gamma}\right]}{n-k} \le \frac{\chi^2_{n-k-l}(0;1-\alpha)}{n-k-l}\right\}\\
&\ge P\left\{\chi^2_{n-k}(\delta_{\max}) \le \frac{n-k}{n-k-l}\chi^2_{n-k-l}(0;1-\alpha)\right\}\\
&= 1-\alpha-\varepsilon.
\end{aligned}$$

Here the implication

$$\delta_1 < \delta_2 \implies P\{\chi^2_f(\delta_1) \le c^2\} > P\{\chi^2_f(\delta_2) \le c^2\},$$

which is valid for any δ_i, $i = 1, 2$, $c^2 > 0$ and any $f \ge 1$, is taken into account. $\square$

Remark 1. The solution $\delta_{\max}$ of the equation

$$P\{\chi_f^2(\delta_{\max}) \le \chi_f^2(0; 1-\alpha)\} = 1 - \alpha - \varepsilon,$$

for given α and ε, can be obtained either as a solution of the equation

$$\int_0^{\chi_f^2(0;1-\alpha)} \exp[-(y+\delta_{\max})] \sum_{r=0}^{\infty} \frac{1}{r!}\left(\frac{\delta_{\max}}{2}\right)^2 \frac{y^{r+f/2-1}}{2^{r+f/2}\Gamma\left(r+\frac{f}{2}\right)}\,\mathrm{d}y = 1-\alpha-\varepsilon,$$

or iteratively from the mentioned approximation, i.e.

$$\begin{aligned}
&P\left\{\chi^2_{n-k}(\delta_{\max}) \le \frac{n-k}{n-k-l}\chi^2_{n-k-l}(0;1-\alpha)\right\}\\
&\approx P\left\{\frac{n-k+2\delta_{\max}}{n-k+\delta_{\max}}\chi^2_{\frac{(n-k+\delta_{\max})^2}{n-k+2\delta_{\max}}}(0) \le \frac{n-k}{n-k-l}\chi^2_{n-k-l}(0;1-\alpha)\right\}\\
&= 1-\alpha-\varepsilon
\end{aligned}$$

(by the help of a statistical tables of chi-squared distribution). In the equation

$$P\left\{\chi^2_{\frac{(n-k+\delta_{\max})^2}{n-k+2\delta_{\max}}}(0) \le \frac{n-k+\delta_{\max}}{n-k+2\delta_{\max}}\,\frac{n-k}{n-k-l}\chi^2_{n-k-l}(0;1-\alpha)\right\} = 1-\alpha-\varepsilon$$

in the first step $\delta^{(1)}_{\max} = 1$ can be chosen. Let

$$P\left\{\chi^2_{\frac{(n-k+\delta^{(1)}_{\max})^2}{n-k+2\delta^{1)}_{\max}}}(0) \le \frac{n-k+\delta^{(1)}_{\max}}{n-k+2\delta^{(1)}_{\max}}\,\frac{n-k}{n-k-l}\chi^2_{n-k-l}(0;1-\alpha)\right\} = \omega(\delta^{(1)}_{\max}).$$

If $\omega(\delta^{(1)}_{\max}) < 1-\alpha-\varepsilon$, then $\delta^{(2)}_{\max}$ can be chosen as 2; in the case $\omega(\delta^{(1)}_{\max}) > 1-\alpha-\varepsilon$, then $\delta^{(2)}_{\max}$ can be chosen as 0.5. Now by a simple linear interpolation on the basis of the values $\omega(\delta^{(1)}_{\max})$, $\omega(\delta^{(2)}_{\max})$, $\delta^{(1)}_{\max}$, $\delta^{(2)}_{\max}$, the resulting

$$\delta_{\max} = \delta^{(1)}_{\max} + \frac{(\delta^{(2)}_{\max} - \delta^{(1)}_{\max})(1-\alpha-\varepsilon-\omega(\delta^{(1)}_{\max})}{\omega(\delta^{(2)}_{\max}) - \omega(\delta^{(1)}_{\max})}$$

can be obtained. If $|\omega(\delta_{\max}) - (1-\alpha-\varepsilon)|$ is not sufficiently small, then the next step of the iteration is used.

Remark 2. In view of Theorem 1 the set

$$\mathcal{A} = \left\{\boldsymbol{u}:\ \boldsymbol{u} \in R^l,\ \ \boldsymbol{u}'\mathbf{S}'(\mathbf{M}_X\mathbf{\Sigma}\mathbf{M}_X)^{+}\mathbf{S}\boldsymbol{u} \le \delta_{\max}\right\}$$

enables us to decide, whether the model (1) can be reduced, i.e. $\boldsymbol{\gamma} \in \mathcal{A}$ implies this possibility.

A utilization of the rule given in Theorem 1 indicates the statistical significance of a difference between the models (1) and (2). Even the difference is statistically significant it need not be of an importance from the practical viewpoint. In this case more moderate criterion for a neglecting parameters can be used. It is based on the difference between the mean value $E(\tilde{\sigma}^2)$ of the estimator of σ^2 in the reduced model and σ^2; in detail cf. the following text.

The matrix $\mathbf{\Sigma}$ can be considered in the form $\mathbf{\Sigma} = \sigma^2\mathbf{V}$. If in the reduced model (2) the mean value of the estimator $\tilde{\sigma}^2$, i.e. $E(\tilde{\sigma}^2)$ is smaller than $\sigma^2 + d^2$, where d^2 is suitable small for further consideration, the reduction is admissible.

The choice of d^2 depends on the actual situation. As an example let us consider a time course $f(\cdot, \boldsymbol{\beta}, \boldsymbol{\gamma})$ of the blood pressure of a patient. Usually the following relationship

$$\begin{aligned} f(t, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \beta_0 &+ \sum_{i=1}^{k_1} \left[\beta_{2i-1} \cos\left(\frac{2\pi}{T_i} t\right) + \beta_{2i+1} \sin\left(\frac{2\pi}{T_i} t\right)\right] \\ &+ \sum_{j=k_1+1}^{k_1+k_2} \left[\gamma_{2j-1} \cos\left(\frac{2\pi}{T_j} t\right) + \gamma_{2j+1} \sin\left(\frac{2\pi}{T_j} t\right)\right], \qquad t > 0, \end{aligned}$$

is used. The values T_i, $i = 1, \ldots, k_1$, T_j, $j = k_1 + 1, \ldots, k_1 + k_2$, are determined by periodogram. The impact, which is to be determined for the patient is

$$I = \int_{t_1}^{t_2} f(t, \boldsymbol{\beta}, \boldsymbol{\gamma}) \, \mathrm{d}t.$$

Variance of the estimator $\widehat{I} = \int_{t_1}^{t_2} f(t, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}) \, \mathrm{d}t$ of the impact depends on σ^2. It can occur that the needed variance of $\widehat{I}$ is considerable larger than $\mathrm{Var}(\widehat{I})$, i.e. the needed variance $\gg \mathrm{Var}(\widehat{I})$. In this case it is possible to choose d^2 in such a way that the needed variance of the impact is attained by the help of $\sigma^2 + d^2 = E(\tilde{\sigma}^2)$ instead of σ^2. In more detail

$$\mathrm{Var}(\widehat{I}) = \mathrm{Var}\left(\int_{t_1}^{t_2} f(t, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}) \, \mathrm{d}t\right) = \sigma^2 K << \text{ needed variance},$$

$$\mathrm{Var}(\tilde{I}) = \mathrm{Var}\left(\int_{t_1}^{t_2} f(t, \widehat{\boldsymbol{\kappa}}, \mathbf{0}) \, \mathrm{d}t\right) = (\sigma^2 + d^2)L = \text{ needed variance}.$$

The constants K and L can be easily determined in the full model by the help of the function $f(\cdot, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and in the reduced model by the help of the function $f(\cdot, \boldsymbol{\kappa}, \mathbf{0})$.

Lemma 4. *In the model* (1) *with* $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$ *the mean value of the statistic*

$$\tilde{\sigma}^2 = \frac{(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})'\mathbf{V}^{-1}(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})}{n-k}$$

is

$$E(\tilde{\sigma}^2) = \sigma^2 + \frac{\boldsymbol{\gamma}'\mathbf{S}'(\mathbf{M}_X\mathbf{V}\mathbf{M}_x)^+\mathbf{S}\boldsymbol{\gamma}}{n-k}.$$

P r o o f. It is valid that

$$\begin{aligned} &E\left(\frac{(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})'\mathbf{V}^{-1}(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\kappa}})}{n-k}\right) \\ &= E\left(\frac{\boldsymbol{Y}'(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+\boldsymbol{Y}}{n-k}\right) \\ &= \frac{1}{n-k}E(\boldsymbol{Y}')(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+E(\boldsymbol{Y}) + \frac{1}{n-k}\operatorname{Tr}\left[(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+\operatorname{Var}(\boldsymbol{Y})\right] \\ &= \frac{1}{n-k}\boldsymbol{\gamma}'\mathbf{S}'(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+\mathbf{S}\boldsymbol{\gamma} + \sigma^2, \end{aligned}$$

since

$$\begin{aligned} &\operatorname{Tr}\left[(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+\sigma^2\mathbf{V}\right] \\ &= \sigma^2\operatorname{Tr}\left[\mathbf{I} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\right] \\ &= \sigma^2\operatorname{Tr}(\mathbf{I}_{n,n}) - \sigma^2\operatorname{Tr}\left[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\right] = \sigma^2(n-k) \end{aligned}$$

and

$$\begin{aligned} E(\boldsymbol{Y}')(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+E(\boldsymbol{Y}) &= (\boldsymbol{\beta}', \boldsymbol{\gamma}')\begin{pmatrix}\mathbf{X}'\\ \mathbf{S}'\end{pmatrix}(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+(\mathbf{X}, \mathbf{S})\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} \\ &= \boldsymbol{\gamma}'\mathbf{S}'(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+\mathbf{S}\boldsymbol{\gamma}. \end{aligned}$$

Thus the statement is proved. □

Remark 3. If the mean value of the statistic $\frac{\boldsymbol{Y}'(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+\boldsymbol{Y}}{n-k}$ in the model (1) may attain the value $\sigma^2 + d^2$, then obviously the condition

$$\frac{\boldsymbol{\gamma}'\mathbf{S}'(\mathbf{M}_X\mathbf{V}\mathbf{M}_X)^+\mathbf{S}\boldsymbol{\gamma}}{n-k} \le d^2$$

is sufficient for it. The value d^2 is chosen by the statistician and the choice depends on the actual situation as it was mentioned in the consideration before Lemma 4. Sometimes the reduction of the number of parameters is more important than the best fit to the data.

In an actual situation we want to know whether the condition for reducing a model is satisfied. Thus we need to know, at least approximately, the actaul value $\boldsymbol{\gamma}^*$ of the parameter $\boldsymbol{\gamma}$, i.e. an estimator and confidence region for it is necessary. If all elements of the confidence region satisfy the condition for reduction, then the reduction is admissible.

4. Model with constraints on $\boldsymbol{\beta}$

In a measurement of a gravity field of the Earth so-called relative gravity-meters have been used. They make possible to measure a difference of the intensity of the gravity field between two different points, however the result of the measurement is of the form

$$y = \beta + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 + \gamma_4 t^4 + \varepsilon,$$

where β is true value of the difference, $\gamma_1, \dots, \gamma_4$ are nuisance parameters, ε is a measurement error and t is a time interval between the measurement of the first and the other point. In more detail the behaviour of the device can be characterized as follows. If it is located on some point at the Earth surface and stays here during the whole working day, its registrations are not constant in the time. They are slowly changing and their time course can be described by a polynomial of the fourth order. Parameters of the polynomial are useless for an observer, they are nuisance parameters.

If the device is placed on the point P_1 at the time t_1 and on the point P_2 at the time t_2, then the difference of these two registrations gives the difference $\beta_2 - \beta_1$ between intensities of the gravity field of the points P_1, P_2, however it is influenced by the value $(t_2 - t_1)\gamma_1 + (t_2 - t_1)^2\gamma_2 + t_2 - t_1)^3\gamma_3 + (t_2 - t_1)^4\gamma_4$.

Therefore a design of measurement is prepared in such a way that the differencies $\beta_i - \beta_j$ of intensities of the gravity field and also the parameters $\gamma_1, \dots, \gamma_4$ can be estimated for the working day. The design can be of a different forms and it can be characterized as free polygon with loops, closed polygon, several polygons creating a network, etc. If, e.g. a closed polygon with loops is considered, then sums of diferences must be zero. The number of constraints depends on the number of loops and on the fact that the polygon is closed. In different designs different constraints occur.

In general it leads to the model

$$\boldsymbol{Y} \sim N_n \left[(\mathbf{X}, \mathbf{S}) \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix}, \boldsymbol{\Sigma} \right], \quad \boldsymbol{b} + \mathbf{B}\boldsymbol{\beta} = \mathbf{0}. \tag{4}$$

Lemma 5. *Let the model* (4) *be regular, i.e.* $r(\mathbf{X},\mathbf{S}) = k+l \le n$, $r(\mathbf{B}) = q < k$, $\boldsymbol{\Sigma}$ *p.d. Then*

$$\left[\boldsymbol{Y} - (\mathbf{X},\mathbf{S})\begin{pmatrix}\hat{\hat{\boldsymbol{\beta}}}_{\text{full}}\\ \hat{\hat{\boldsymbol{\gamma}}}_{\text{full}}\end{pmatrix}\right]'\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{Y} - (\mathbf{X},\mathbf{S})\begin{pmatrix}\hat{\hat{\boldsymbol{\beta}}}_{\text{full}}\\ \hat{\hat{\boldsymbol{\gamma}}}_{\text{full}}\end{pmatrix}\right] \sim \chi^2_{n+q-k-l}(0)$$

and

$$(\boldsymbol{Y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_{\text{red}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_{\text{red}}) \sim \chi^2_{n+q-k}(\delta),$$

where

$$\begin{aligned}\delta &= E(\boldsymbol{Y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_{\text{red}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_{\text{red}})\\ &= \boldsymbol{\gamma}'\mathbf{S}'\left\{(\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^+ + \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\right\}\mathbf{S}\boldsymbol{\gamma}.\end{aligned}$$

P r o o f. Here the notation

$$\begin{aligned}\mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'} &= \mathbf{X}\mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1},\\ &= \mathbf{X}\mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1},\\ \mathbf{M}^{\Sigma^{-1}}_{XC^{-1}B'} &= \mathbf{I} - \mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'},\\ \mathbf{P}^{C}_{C^{-1}B'} &= \mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{C}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{C} = \mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B},\\ \mathbf{M}^{C}_{C^{-1}B'} &= \mathbf{I} - \mathbf{P}^{C}_{C^{-1}B'},\\ \mathbf{P}^{\Sigma^{-1}}_{X} &= \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}, \qquad \mathbf{M}^{\Sigma^{-1}}_{X} = \mathbf{I} - \mathbf{P}^{\Sigma^{-1}}_{X},\\ \mathbf{P}_{B'} &= \mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1}\mathbf{B}, \qquad \mathbf{M}_{B'} = \mathbf{I} - \mathbf{P}_{B'}\end{aligned}$$

will be used.

The BLUE $\hat{\hat{\boldsymbol{\beta}}}_{\text{red}}$ in the reduced model

$$\boldsymbol{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \qquad \boldsymbol{b} + \mathbf{B}\boldsymbol{\beta} = \mathbf{0}$$

is

$$\hat{\hat{\boldsymbol{\beta}}}_{\text{red}} = \mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} - \mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}(\mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} + \boldsymbol{b})$$

and

$$\text{Var}(\hat{\hat{\boldsymbol{\beta}}}_{\text{red}}) = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1} = (\mathbf{M}_{B'}\mathbf{C}\mathbf{M}_{C'})^+.$$

The bias of $\hat{\hat{\boldsymbol{\beta}}}_{\text{red}}$ in the model (4) is given by the relationship

$$\begin{aligned}&E(\hat{\hat{\boldsymbol{\beta}}}_{\text{red}}) - \boldsymbol{\beta}\\ &= \mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\gamma} - \mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}(\mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\gamma} + \boldsymbol{b})\\ &= \left[\mathbf{I} - \mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\right]\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\gamma} = \mathbf{M}^{C}_{C^{-1}B'}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\gamma}\end{aligned}$$

and thus

$$
\begin{aligned}
E(\boldsymbol{Y} - \mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}}_{\text{red}}) &= \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\gamma - \mathbf{X}\mathbf{M}^{C}_{C^{-1}B'}\mathbf{C}^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{S}\gamma - \mathbf{X}\boldsymbol{\beta}\\
&= (\mathbf{I} - \mathbf{X}\mathbf{M}^{C}_{C^{-1}B'}\mathbf{C}^{-1}\mathbf{X}'\Sigma^{-1})\mathbf{S}\gamma\\
&= (\mathbf{M}^{\Sigma^{-1}}_{X} + \mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'})\mathbf{S}\gamma,
\end{aligned}
$$

$$
\begin{aligned}
&E(\boldsymbol{Y} - \mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}}_{\text{red}})'\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Y} - \mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}}_{\text{red}})\\
&= \boldsymbol{\gamma}'\mathbf{S}'(\mathbf{M}^{\Sigma^{-1}}_{X} + \mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'})'\boldsymbol{\Sigma}^{-1}(\mathbf{M}^{\Sigma^{-1}}_{X} + \mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'})\mathbf{S}\boldsymbol{\gamma}\\
&= \boldsymbol{\gamma}'\mathbf{S}'\left[\left(\mathbf{M}^{\Sigma^{-1}}_{X}\right)'\boldsymbol{\Sigma}^{-1}\mathbf{M}^{\Sigma^{-1}}_{X} + \left(\mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'}\right)'\boldsymbol{\Sigma}^{-1}\mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'}\right]\mathbf{S}\gamma,
\end{aligned}
$$

since $\left(\mathbf{M}^{\Sigma^{-1}}_{X}\right)'\boldsymbol{\Sigma}^{-1}\mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'} = \mathbf{0}$. Now it is sufficient to take into account the following relationships

$$
\begin{aligned}
\left(\mathbf{M}^{\Sigma^{-1}}_{X}\right)'\boldsymbol{\Sigma}^{-1}\mathbf{M}^{\Sigma^{-1}}_{X} &= (\mathbf{M}_X\boldsymbol{\Sigma}\mathbf{M}_X)^{+},\\
\left(\mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'}\right)'\boldsymbol{\Sigma}^{-1}\mathbf{P}^{\Sigma^{-1}}_{XC^{-1}B'} &= \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}.
\end{aligned}
$$

□

Now we can continue similarly as in Theorem 1. In this way we obtain the following statement.

Theorem 2. *If in the true model* (4) *the nuisance vector parameter* $\boldsymbol{\gamma}$ *is in the set*

$$\left\{\gamma:\ \boldsymbol{\gamma}'\mathbf{S}'\left[(\mathbf{M}_C\boldsymbol{\Sigma}\mathbf{M}_X)^{+} + \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\right]\mathbf{S}\gamma \le \delta_{\max}\right\}$$

where $\delta_{\max}$ *is a solution of the equation*

$$P\left\{\chi^2_{n+q-k}(\delta) \le \chi^2_{n+q-k-l}(0; 1-\alpha)\right\} = 1 - \alpha - \varepsilon,$$

then the true model (4) *can be substituted by the reduced model without an essential deterioration of the agreement between data and the model.*

5. Numerical example

Let values of a quadratic polynomial be measured at the points

$$x \in \mathcal{S} = \{-0.8; -0.6; -0.4; -0.2; 0; 0.2; 0.4; 0.6; 0.8\}$$

The observation vector is

$$\boldsymbol{Y} \sim N_9 \left[(\boldsymbol{\varphi}_0, \boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \gamma \end{pmatrix}, \sigma^2 \mathbf{I} \right],$$

i.e. $n = 9$, $k = 2$, $l = 1$. Here

$$\boldsymbol{\varphi}_j = \left[\varphi_j(x_1), \dots, \varphi_j(x_9) \right]', \qquad j = 0, 1, 2,$$

and $\varphi_0(\cdot)$, $\varphi_1(\cdot)$, $\varphi_2(\cdot)$ are the Chebyshev polynomials, i.e.

$$\varphi_0(x) = 1, \qquad x \in \mathcal{S},$$

$$\varphi_1(x) = x - \overline{x}, \qquad x \in \mathcal{S}, \quad \overline{x} = \frac{1}{9} \sum_{i=1}^{9} x_i,$$

$$\varphi_2(x) = x^2 - \frac{\sum\limits_{i=1}^{9} x_i^2 (x_i - \overline{x})}{\sum\limits_{i=1}^{9} (x_i - \overline{x})^2} (x - \overline{x}) - \frac{1}{9} \sum_{i=1}^{9} x_i^2, \qquad x \in \mathcal{S}.$$

In this case the solution of the equation

$$P \left\{ \frac{7 + 2\delta}{7 + \delta} \chi^2_{\frac{(7+\delta)^2}{7+2\delta}}(0) \le \frac{7}{6} \chi_6^2(0; 0.95) \right\} = 1 - 0.05 - 0.04$$

is $\delta_{\max} = 1.24$. Thus the boundary for γ is given by the relationships

$$\gamma' \mathbf{S}' (\mathbf{M}_X \sigma^2 \mathbf{I} \mathbf{M}_X)^+ \mathbf{S} \gamma = \sigma^{-2} \gamma' \mathbf{S}' \mathbf{M}_X \mathbf{S} \gamma \le 1.24,$$

i.e. $|\gamma| \le 1.92\sigma$. Here $\mathbf{S}' \mathbf{M}_X \mathbf{S} = \sum\limits_{i=1}^{9} \varphi_2^2(x_i) = 0.336275$.

Another solution is implied by the requirement

$$\frac{\gamma' \mathbf{S}' (\mathbf{M}_X \mathbf{V} \mathbf{M}_X)^+ \mathbf{S} \gamma}{9 - 2} \le d^2,$$

where $\mathbf{V} = \mathbf{I}$. Thus

$$|\gamma| \le \frac{d\sqrt{9-2}}{\sqrt{\sum\limits_{i=1}^{9} \varphi_2^2(x_i)}} = 4.562d.$$

The choice $d = \sigma \frac{1.920}{4.562} = \sigma 0.421$ gives the same restriction for the parameter γ as in the preceding consideration.

The standard deviation of the estimator $\widehat{\gamma}_{\text{true}}$ in the true model is

$$\sqrt{\operatorname{Var}(\widehat{\gamma}_{\text{true}})} = \frac{\sigma}{\sqrt{\sum\limits_{i=1}^{9} \varphi_2^2(x_i)}} = \frac{\sigma}{\sqrt{0.336275}} = 1.724\sigma.$$

Ackonwlegement. The author is deeply indebted to the reviewer for his/her corrections, comments and a help in formulations.

REFERENCES

[1] FIŠEROVÁ, E.—KUBÁČEK, L.—KUNDEROVÁ, P.: *Linear Statistical Models; Regularity and Singularities*, Academia, Praha, 2007.
[2] KUBÁČEK, L.—KUBÁČKOVÁ, L.—VOLAUFOVÁ, J.: *Statistical Models with Linear Structures*, Veda, Bratislava, 1995.
[3] KUBÁČEK, L.: *Underparametrized regression models*, Tatra Mt. Math. Publ. **14** (2006), 241–254.
[4] KUBÁČEK, L.: *Admissible regions for neglecting parameters in linear statistical models.* Acta Comment. Univ. Tartu. Math. **14** (2010), 15–25.
[5] PATNAIK, P. B.: *The non-central* χ^2 *and* F*-distributions and their applications*, Biometrika **36** (1949), 202–232.
[6] PUNTANEN, S: *Some further results related to reduced singular linear models*, Comm. Statist. Theory Methods **26** (1997), 375–385.
[7] RAO, C. R.: *Linear Statistical Inference and Its Applications*, Academia. Praha, 1978 (Czech).
[8] RAO, C. R.—MITRA, S. K.: *Generalized Inverse of Matrices and its Applications*, John Wiley & Sons, New York-London-Sydney-Toronto, 1971.

Received 9. 7. 2009
Accepted 26. 11. 2009

Department of Mathematical Analysis
and Applications of Mathematics
Faculty of Sciences
17. listopadu 12
CZ–771 46 Olomouc
CZECH REPUBLIC
E-mail: lubomir.kubacek@upol.cz