Kybernetika

VOLUME *43* (2007), NUMBER 4

The Journal of the Czech Society for Cybernetics and Information Sciences

Published by:
Institute of Information Theory and Automation of the AS CR, v.v.i.

Editorial Office:
Pod Vodárenskou věží 4, 182 08 Praha 8

Kybernetika is a bi-monthly international journal dedicated for rapid publication of high-quality, peer-reviewed research articles in fields covered by its title.

Kybernetika traditionally publishes research results in the fields of Control Sciences, Information Sciences, System Sciences, Statistical Decision Making, Applied Probability Theory, Random Processes, Fuzziness and Uncertainty Theories, Operations Research and Theoretical Computer Science, as well as in the topics closely related to the above fields.

The Journal has been monitored in the Science Citation Index since 1977 and it is abstracted/indexed in databases of Mathematical Reviews, Current Mathematical Publications, Current Contents ISI Engineering and Computing Technology.

Kybernetika. Volume *43* (2007) ISSN 0023-5954, MK ČR E 4902.

Published bimonthly by the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. — Address of the Editor: P. O. Box 18, 182 08 Prague 8, e-mail: kybernetika@utia.cas.cz. — Printed by PV Press, Pod vrstevnicí 5, 140 00 Prague 4. — Orders and subscriptions should be placed with: MYRIS TRADE Ltd., P. O. Box 2, V Štíhlách 1311, 142 01 Prague 4, Czech Republic, e-mail: myris@myris.cz. — Sole agent for all "western" countries: Kubon & Sagner, P. O. Box 34 01 08, D-8 000 München 34, F.R.G.

Published in October 2007.

KYBERNETIKA — VOLUME 43 (2007), NUMBER 4, PAGES 463 – 470

TEST OF LINEAR HYPOTHESIS IN MULTIVARIATE MODELS

Lubomír Kubáček

In regular multivariate regression model a test of linear hypothesis is dependent on a structure and a knowledge of the covariance matrix. Several tests procedures are given for the cases that the covariance matrix is either totally unknown, or partially unknown (variance components), or totally known.

Keywords: multivariate model, linear hypothesis, variance components, insensitive region

AMS Subject Classification: 62J05

1. NOTATIONS AND AUXILIARY STATEMENTS

Let a model

$$\underline{\boldsymbol{Y}} \sim N_{nm}(\boldsymbol{XB}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}) \tag{1}$$

be under consideration. Here $\underline{\boldsymbol{Y}}$ is an $n \times m$ normally distributed matrix with the mean value matrix $\mathrm{E}(\underline{\boldsymbol{Y}})$ equal to $\boldsymbol{XB}$. The covariance matrix of the vector $\mathrm{vec}(\underline{\boldsymbol{Y}})$ (the vector composed of the columns of the matrix $\underline{\boldsymbol{Y}}$) is $\mathrm{Var}[\mathrm{vec}(\underline{\boldsymbol{Y}})] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}$ ($\boldsymbol{I}$ is the $n \times n$ identity matrix). The model is regular if the rank $r(\boldsymbol{X})$ of the matrix $\boldsymbol{X}$ is $r(\boldsymbol{X}) = k < n$ and the $m \times m$ matrix $\boldsymbol{\Sigma}$ is positive definite (p.d.).

The linear hypothesis of the unknown $k \times m$ parameter matrix $\boldsymbol{B}$ is considered in the form

$$H_0: \quad \boldsymbol{HB} + \boldsymbol{H}_0 = \boldsymbol{0}, \tag{2}$$

where $h \times k$ matrix $\boldsymbol{H}$ is assumed to be known. The $h \times m$ matrix $\boldsymbol{H}_0$ is also assumed to be known. The hypothesis is regular if $r(\boldsymbol{H}) = h < k$. The alternative hypothesis is

$$H_a: \quad \boldsymbol{HB} + \boldsymbol{H}_0 \neq \boldsymbol{0}.$$

Lemma 1.1. The best linear unbiased estimator of the matrix $\boldsymbol{B}$ is

$$\widehat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\underline{\boldsymbol{Y}} \sim N_{km}[\boldsymbol{B}, \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}].$$

Proof. Cf. [1]. □

Lemma 1.2. One of the test statistics for the regular hypothesis (2) in the case of the known matrix $\boldsymbol{\Sigma}$ is

$$T = \mathrm{Tr}\Big\{(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)\boldsymbol{\Sigma}^{-1}\Big\} \sim \chi^2_{mh}(\delta), \tag{3}$$

where

$$\delta = \mathrm{Tr}\Big\{(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0)\boldsymbol{\Sigma}^{-1}\Big\}.$$

The symbol $\chi^2_{mh}(\delta)$ means the noncentral chi-square random variable with mh degrees of freedom and with the parameter of noncentrality equal to δ, $\boldsymbol{B}^*$ means the actual value of the matrix $\boldsymbol{B}$.

Proof. The statement can be obtained from an univariate model $\mathrm{vec}(\underline{\boldsymbol{Y}}) \sim N_{nm}[(\boldsymbol{I} \otimes \boldsymbol{X})\mathrm{vec}(\boldsymbol{B}), \boldsymbol{\Sigma} \otimes \boldsymbol{I}]$ in a standard way by utilization of the relationship $\mathrm{vec}(\boldsymbol{X}\boldsymbol{B}) = (\boldsymbol{I} \otimes \boldsymbol{X})\mathrm{vec}(\boldsymbol{B})$. □

Lemma 1.3. The matrix $(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})$ is the $m \times m$ Wishart matrix with the $n-k$ degrees of freedom and with the covariance matrix $\boldsymbol{\Sigma}$, i.e. $(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}}) \sim W_m(n-k, \boldsymbol{\Sigma})$.

Proof. The matrix $\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}}$ is distributed as $N_{nm}(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{M}_X)$, where $\boldsymbol{M}_X = \boldsymbol{I} - \boldsymbol{P}_X$ and $\boldsymbol{P}_X$ is the Euclidean projector on the subspace $\mathcal{M}(\boldsymbol{X}) = \{\boldsymbol{X}\boldsymbol{u} : \boldsymbol{u} \in \mathbb{R}^k\}$. Thus for any generalized inverse (cf. [6]) $\boldsymbol{M}_X^-$ of the matrix $\boldsymbol{M}_X$ the matrix $(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'\boldsymbol{M}_X^-(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})$ has the Wishart distribution $W_m([r(\boldsymbol{M}_X), \boldsymbol{\Sigma}]$. One version of the matrix $\boldsymbol{M}_X^-$ is $\boldsymbol{I}$. □

Lemma 1.4. If $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{V}$ ($\boldsymbol{V}$ is p.d.), then the best estimator of σ^2 is

$$\widehat{\sigma}^2 = \frac{\mathrm{Tr}[(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})\boldsymbol{V}^{-1}]}{m(n-k)} \sim \sigma^2 \frac{\chi^2_{m(n-k)}(0)}{m(n-k)}.$$

This estimator is independent of the estimator $\widehat{\boldsymbol{B}}$.

Proof. The statement is a transcription of the well known statement from the theory of the univariate linear models (cf. e.g. [2]). □

Corollary 1.5. If $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{V}$, then one of the test statistics for the regular hypothesis (2) is

$$T = \frac{\mathrm{Tr}\Big\{(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)\boldsymbol{V}^{-1}\Big\}/(mh)}{\mathrm{Tr}[(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})\boldsymbol{V}^{-1}]/[m(n-k)]} \sim F_{mh,m(n-k)}(\delta),$$

where

$$\delta = \frac{\mathrm{Tr}\Big\{(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0)\boldsymbol{V}^{-1}\Big\}}{\sigma^2}$$

and $F_{mh,m(n-k)}(\delta)$ is the noncentral Fisher–Snedecor random variable with degrees of freedom equal to mh and $m(n-k)$ and with the noncentrality parameter equal to δ.

2. DIFFERENT STRUCTURES OF THE MATRIX $\boldsymbol{\Sigma}$

Let $\boldsymbol{\Sigma}$ be given. Then

$$(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0) = \boldsymbol{Q}_1 \sim W_m(h, \boldsymbol{\Sigma})$$

(possibly noncentral) and therefore, under the null hypothesis, for any nonzero $\boldsymbol{f} \in \mathbb{R}^m$ it is valid

$$\boldsymbol{f}'\boldsymbol{Q}_1\boldsymbol{f}/(\boldsymbol{f}'\boldsymbol{\Sigma}\boldsymbol{f}) \sim \chi_h^2(0).$$

Let $\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0 \neq \boldsymbol{0}$ ($\boldsymbol{B}^*$ is the actual value of the matrix $\boldsymbol{B}$) and let $\lambda_{\max}$ be the maximum solution of the equation

$$\det\left\{(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0) - \lambda\boldsymbol{\Sigma}\right\} = 0$$

and let $\boldsymbol{f}_{\max}$ satisfy the relationship

$$\left\{(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0) - \lambda_{\max}\boldsymbol{\Sigma}\right\}\boldsymbol{f}_{\max} = \boldsymbol{0}.$$

Then

$$\delta = \boldsymbol{f}'_{\max}(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\boldsymbol{B}^* + \boldsymbol{H}_0)\boldsymbol{f}_{\max}\Big/\boldsymbol{f}'_{\max}\boldsymbol{\Sigma}\boldsymbol{f}_{\max},$$

i. e. the parameter of noncentrality of the statistic

$$\chi_h^2(\delta) = \boldsymbol{f}'_{\max}\boldsymbol{Q}_1\boldsymbol{f}_{\max}/\boldsymbol{f}'_{\max}\boldsymbol{\Sigma}\boldsymbol{f}_{\max} \tag{4}$$

is for this vector $\boldsymbol{f}_{\max}$ maximum and therefore the chance to detect that H_0 is not true is also maximum.

It is of some importance to compare the power functions of the statistics (3) and (4).

Let

$$\underline{\boldsymbol{Y}} = \begin{pmatrix} -2, & 1, & 4 \\ -1, & 2, & 2 \\ 0, & 4, & -4 \\ 1, & 2, & 2 \\ 2, & 1, & 4 \end{pmatrix} \boldsymbol{B}_{3,3} + \boldsymbol{\varepsilon}_{5,3}, \quad \mathrm{Var}[\mathrm{vec}(\underline{\boldsymbol{Y}})] = \begin{pmatrix} 1^2, & 0, & 0 \\ 0, & 2^2, & 0 \\ 0, & 0, & 3^2 \end{pmatrix} \otimes \boldsymbol{I}_{5,5}$$

and the null hypothesis be $\begin{pmatrix} 1, & 1, & 1 \\ 0, & 1, & 1 \end{pmatrix}\boldsymbol{B} = \boldsymbol{0}$. It means $h = 2$, $m = 3$, $n = 5$, $k = 3$. If

$$\begin{pmatrix} 1, & 1, & 1 \\ 0, & 1, & 1 \end{pmatrix}\boldsymbol{B} = \begin{pmatrix} 0.5, & -0.5, & 1.0 \\ 0, & 0.5, & -0.5 \end{pmatrix},$$

then $\boldsymbol{f}'_{\max}\boldsymbol{Q}_1\boldsymbol{f}_{\max}\Big/\boldsymbol{f}'_{\max}\boldsymbol{\Sigma}\boldsymbol{f}_{\max} \sim \chi_2^2(\delta_1), \delta_1 = 2.994$ and $T \sim \chi_6^2(\delta_2), \delta_2 = 6.603$ (cf. Lemma 1.2).

If $\chi_f^2(\delta)$ is approximated by $\frac{f+2\delta}{f+\delta}\chi^2_{\frac{(f+\delta)^2}{f+2\delta}}(0)$, then we obtain for $\alpha = 0.05$ $\mathrm{P}\{\chi_2^2(2.994) \geq 5.99\} = 21\,\%$ and $\mathrm{P}\{\chi_6^2(6.603) \geq 12.6\} = 44\,\%$. It shows a prevalence of the test (3) versus (4). However it can be utilized only in the case of the known matrix $\boldsymbol{\Sigma}$, or if its estimator is very precise.

If the matrix $\boldsymbol{\Sigma}$ is unknown and (2) is true, then the relationships

$$\begin{aligned}\boldsymbol{Q}_1 &= (\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0) \sim W_m(h, \boldsymbol{\Sigma}),\\ \boldsymbol{Q}_2 &= (\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}}) \sim W_m(n-k, \boldsymbol{\Sigma})\end{aligned}$$

(it is to be remarked that $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are independent) can be utilized for a construction of different tests for the hypothesis (2). As and example can serve the statistic $\boldsymbol{g}'\boldsymbol{Q}_1\boldsymbol{g}/\boldsymbol{g}'\boldsymbol{Q}_2\boldsymbol{g} \sim F_{h,n-k}$, where

$$\frac{\boldsymbol{g}'\boldsymbol{Q}_1\boldsymbol{g}}{\boldsymbol{g}'\boldsymbol{Q}_2\boldsymbol{g}} = \max\left\{\frac{\boldsymbol{u}'\boldsymbol{Q}_1\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{Q}_2\boldsymbol{u}} : \boldsymbol{u} \in \mathbb{R}^m\right\}.$$

This statistic has the Fisher–Snedecor distribution $F_{h,n-k}(0)$ if the hypothesis H_0 is true and the distribution is independent of $\boldsymbol{g}$. However if H_0 is not true then the statistics has the largest realization and thus there is the greatest chance to recognize that H_0 is not true.

If $n-k$ tends to infinity, then $\widehat{\boldsymbol{\Sigma}} = (\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})/(n-k)$ tends to $\boldsymbol{\Sigma}$ in probability and thus $\mathrm{Tr}\left\{(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)\widehat{\boldsymbol{\Sigma}}^{-1}\right\}$ tends in distribution to χ^2_{mh}. This fact can be also utilized mainly in connection to a consideration at the beginning of this section. Other tests based on the matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, respectively, are analyzed in [4] and therefore they are omitted here.

Lemma 2.1. Let $\boldsymbol{\Sigma} = \sum_{i=1}^p \vartheta_i \boldsymbol{V}_i$, where ϑ_i, $i = 1, \dots, p$, are unknown parameters, $\boldsymbol{\vartheta} \in \underline{\vartheta} \subset R^p$, and $\boldsymbol{V}_1, \dots, \boldsymbol{V}_p$, are known symmetric matrices. The set $\underline{\vartheta}$ is open and it is valid $\boldsymbol{\vartheta} \in \underline{\vartheta} \Rightarrow \sum_{i=1}^p \vartheta_i \boldsymbol{V}_i$ is p.d. Let the matrix $\boldsymbol{S}_{\Sigma_0^{-1}}$ be regular. Here

$$\left\{\boldsymbol{S}_{\Sigma_0^{-1}}\right\}_{i,j} = \mathrm{Tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{V}_i\boldsymbol{\Sigma}_0^{-1}\boldsymbol{V}_j), \quad i,j = 1, \dots, p,$$

and $\boldsymbol{\Sigma}_0 = \sum_{i=1}^p \vartheta_i^{(0)}\boldsymbol{V}_i, \boldsymbol{\vartheta}^{(0)} = (\vartheta_1^{(0)}, \dots, \vartheta_p^{(0)})'$ is an approximate value of he unknown parameter $\boldsymbol{\vartheta}$. Then the unbiased $\boldsymbol{\vartheta}^{(0)}$-locally minimum variance quadratic invariant estimator of the parameter $\boldsymbol{\vartheta}$ is

$$\widehat{\boldsymbol{\vartheta}} = \frac{1}{n-k}\boldsymbol{S}^{-1}_{\Sigma_0^{-1}}\begin{pmatrix}\mathrm{Tr}(\underline{\boldsymbol{Y}}'\boldsymbol{M}_X\underline{\boldsymbol{Y}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{V}_1\boldsymbol{\Sigma}_0^{-1})\\ \vdots \\ \mathrm{Tr}(\underline{\boldsymbol{Y}}'\boldsymbol{M}_X\underline{\boldsymbol{Y}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{V}_p\boldsymbol{\Sigma}_0^{-1})\end{pmatrix}, \quad \mathrm{Var}_{\vartheta_0}(\widehat{\boldsymbol{\vartheta}}) = \frac{2}{n-k}\boldsymbol{S}^{-1}_{\Sigma_0^{-1}}.$$

P r o o f. Cf. [5]. □

Now the problem arises whether the matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\vartheta}}) = \sum_{i=1}^p \widehat{\vartheta}_i\boldsymbol{V}_i$ can be used instead the matrix $\boldsymbol{\Sigma}$ in the statistic (3) without any essential deterioration of the inference.

In the following text a procedure for a construction of an insensitivity region is described. For the sake of simplicity only a problem of the risk α of the test is analyzed and problems of construction of the insensitivity region for the power function of the test is omitted.

Lemma 2.2. Let

$$T(\boldsymbol{\vartheta}) = \mathrm{Tr}\Big\{(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\vartheta})\Big\}.$$

Then

$$\frac{\partial T(\boldsymbol{\vartheta})}{\partial \vartheta_i} = -\mathrm{Tr}\Big\{(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\vartheta})\boldsymbol{V}_i\boldsymbol{\Sigma}^{-1}(\boldsymbol{\vartheta})\Big\},$$

thus $T(\boldsymbol{\vartheta} + \delta\boldsymbol{\vartheta}) \approx T(\boldsymbol{\vartheta}) + \sum_{i=1}^{p} \frac{\partial T(\boldsymbol{\vartheta})}{\partial \vartheta_i}\delta\vartheta_i = T(\boldsymbol{\vartheta}) + \xi$ and

$$\xi \sim_1 (-h\boldsymbol{a}'\delta\boldsymbol{\vartheta}, 2h\delta\boldsymbol{\vartheta}'\boldsymbol{S}_{\Sigma^{-1}}\delta\boldsymbol{\vartheta}),$$

where $\boldsymbol{a}' = [\mathrm{Tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}^{-1}), \ldots, \mathrm{Tr}(\boldsymbol{V}_p\boldsymbol{\Sigma}^{-1})]$.

P r o o f. Since under the null hypothesis (2)

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial T(\boldsymbol{\vartheta})}{\partial \vartheta_i}\right) &= -\mathrm{E}\Big([\mathrm{vec}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)]'\Big\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_i\boldsymbol{\Sigma}^{-1}) \otimes [\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\Big\} \\
&\times \mathrm{vec}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)\Big) = -\mathrm{Tr}\Big(((\boldsymbol{I} \otimes \boldsymbol{H})[\boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}](\boldsymbol{I} \otimes \boldsymbol{H}')\Big\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_i\boldsymbol{\Sigma}^{-1}) \\
&\otimes [\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\Big\}\Big) = -\mathrm{Tr}\Big((\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_i\boldsymbol{\Sigma}^{-1}) \otimes \Big\{\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}' \\
&\times [\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\Big\}\Big) = -h\mathrm{Tr}(\boldsymbol{V}_i\boldsymbol{\Sigma}^{-1}),
\end{aligned}$$

we have $\mathrm{E}\left(\sum_{i=1}^{p} \frac{\partial T(\boldsymbol{\vartheta})}{\partial \vartheta_i}\delta\vartheta_i\right) = -h\boldsymbol{a}'\delta\boldsymbol{\vartheta}$.

Further

$$\begin{aligned}
\mathrm{cov}\left(\frac{\partial T(\boldsymbol{\vartheta})}{\partial \vartheta_i}, \frac{\partial T(\boldsymbol{\vartheta})}{\partial \vartheta_j}\right) &= 2\,\mathrm{Tr}\Big((\boldsymbol{I} \otimes \boldsymbol{H})[\boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}](\boldsymbol{I} \otimes \boldsymbol{H}')\Big\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_i\boldsymbol{\Sigma}^{-1}) \\
&\otimes [(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\Big\}(\boldsymbol{I} \otimes \boldsymbol{H})[\boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}](\boldsymbol{I} \otimes \boldsymbol{H}')\Big\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_j\boldsymbol{\Sigma}^{-1}) \\
&\otimes [(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}\Big\}\Big) = 2\mathrm{Tr}\Big[(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{V}_j) \otimes \boldsymbol{I}_{h,h}\Big] = 2h\,\{\boldsymbol{S}_{\Sigma^{-1}}\}_{i,j}\,, \\
&\qquad i,j = 1, \ldots, p.
\end{aligned}$$

□

Theorem 2.3. If H_0 is true and $\delta\boldsymbol{\vartheta} \in \mathcal{N}_{\vartheta_0}$, where an insensitivity region is

$$\begin{aligned}
\mathcal{N}_{\vartheta_0} &= \left\{\delta\boldsymbol{\vartheta} : (\delta\boldsymbol{\vartheta} - \boldsymbol{u}_0)'\boldsymbol{A}_0(\delta\boldsymbol{\vartheta} - \boldsymbol{u}_0) \le c^2\right\}, \boldsymbol{u}_0 = \boldsymbol{A}_0^{-1}h\delta_{\max}\boldsymbol{a}_0, \\
\boldsymbol{A}_0 &= 2t^2h\boldsymbol{S}_{\Sigma_0^{-1}} - h^2\boldsymbol{a}_0\boldsymbol{a}_0', \quad c^2 = \delta_{\max}^2 + h^2\delta_{\max}^2\boldsymbol{a}_0'\boldsymbol{A}_0^{-1}\boldsymbol{a}_0, \\
\boldsymbol{a}_0' &= [\mathrm{Tr}(\boldsymbol{V}_1\boldsymbol{\Sigma}_0^{-1}), \ldots, \mathrm{Tr}(\boldsymbol{V}_p\boldsymbol{\Sigma}_0^{-1})],
\end{aligned}$$

then $\mathrm{P}_{H_0}\Big\{T(\boldsymbol{\vartheta}_0+\delta\boldsymbol{\vartheta}) \geq \chi^2_{mh}(0;1-\alpha)\Big\} \leq \alpha+\varepsilon$. Here $\delta_{\max}$ is a solution of the equation $\mathrm{P}\left\{\chi^2_{mh}(0)+\delta \geq \chi^2_{mh}(0;1-\alpha)\right\} = \alpha+\varepsilon$ and t is sufficiently large real number.

P r o o f. If H_0 is true, then for a given $\delta\boldsymbol{\vartheta}$ and sufficiently large t the inequality

$$\xi < -h\boldsymbol{a}_0'\delta\boldsymbol{\vartheta} + t\sqrt{2h\delta\boldsymbol{\vartheta}'\boldsymbol{S}_{\Sigma_0^{-1}}\delta\boldsymbol{\vartheta}} \tag{5}$$

occurs with probability near to one. If

$$-h\boldsymbol{a}_0'\delta\boldsymbol{\vartheta} + t\sqrt{2h\delta\boldsymbol{\vartheta}'\boldsymbol{S}_{\Sigma_0^{-1}}\delta\boldsymbol{\vartheta}} < \delta_{\max}, \tag{6}$$

then $\mathrm{P}\left\{\chi^2_{mh}(0)+\xi \geq \chi^2_{mh}(0;1-\alpha)\right\} \leq \alpha+\varepsilon$. The inequality (5) is implied by the inequality $(\delta\boldsymbol{\vartheta}-\boldsymbol{u}_0)'\boldsymbol{A}_0(\delta\boldsymbol{\vartheta}-\boldsymbol{u}_0) \leq c^2$. □

Remark 2.4. The value t need not be larger than 4. In [3] an optimum choice of t was studied for some cases and it was found that the value $t=3$ can be sufficient large.

Corollary 2.5 If $p=1$, i. e. $\boldsymbol{\Sigma}=\sigma^2\boldsymbol{V}$, then the inequality (6) can be rewritten as

$$-h\frac{m}{\vartheta}\delta\vartheta + t\sqrt{2hm\frac{(\delta\vartheta)^2}{\vartheta^2}} < \delta_{\max}.$$

Since $\delta\vartheta$ can be negative in this case, it must satisfy the inequality $\left|\frac{\delta\vartheta}{\vartheta}\right| < \frac{\delta_{\max}}{hm+t\sqrt{2hm}}$, what can be approximated as $\left|\frac{\delta\sigma}{\sigma}\right| < \frac{1}{2}\frac{\delta_{\max}}{hm+t\sqrt{2hm}}$, where $\vartheta=\sigma^2$. From Lemma 2.1 we obtain $\sqrt{\mathrm{Var}(\widehat{\sigma})} = \frac{0.707\sigma}{\sqrt{m(n-k)}}$. In this case the value $\widehat{\vartheta}$, i. e. the matrix $\widehat{\boldsymbol{\Sigma}}=\widehat{\vartheta}\boldsymbol{V}$ can be used in the test (3) instead the actual value if the following inequality

$$\frac{1}{2}\frac{\delta_{\max}}{hm+t\sqrt{2hm}} \gg t\frac{0.707}{\sqrt{m(n-k)}}$$

is satisfied. If $\alpha=0.05$, $\varepsilon=0.05$, $m=5$, $h=4$, $t=3$, then $n-k \gg 617$. It is quite clear that a requirement on the accuracy of the estimator $\widehat{\vartheta}$ can be rigorous.

In the case $p=1$ obviously the test from Corollary 1.5 must be used. The example is given only for a demonstration how large the necessary number of observations can be.

Remark 2.6. If the matrix $2t^2h\boldsymbol{S}_{\Sigma_0^{-1}} - h^2\boldsymbol{a}_0\boldsymbol{a}_0'$ is not p.d., then from the practical purposes in the spectral decomposition $2t^2h\boldsymbol{S}_{\Sigma_0^{-1}} - h^2\boldsymbol{a}_0\boldsymbol{a}_0' = \sum_{i=1}^m \lambda_i\boldsymbol{f}_i\boldsymbol{f}_i'$ the negative eigenvalues λ_i are substituted by their absolute values $|\lambda_i|$. In this way the shape of the insensitivity region $\mathcal{N}_{\vartheta_0}$ is always ellipsoid.

Remark 2.7. If $p \geq 2$, and only $\widehat{\mathbf{\Sigma}} = \sum_{i=1}^{p} \widehat{\vartheta}_i \boldsymbol{V}_i$ is at our disposal, the matrix $\widehat{\mathbf{\Sigma}}$ can be used in the test (3) in such case only that $\widehat{\delta\boldsymbol{\vartheta}} \in \mathcal{N}_{\vartheta_0}$ with certainty. Thus a consideration on the basis of $\mathrm{Var}(\widehat{\boldsymbol{\vartheta}})$ from Lemma 2.1 must be made.

If the estimator $\widehat{\mathbf{\Sigma}} = \frac{1}{n-k}(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})$ is at our disposal only and the test (3) is to be used, the analogous consideration as in Theorem 2.3 can be made.

Let $\boldsymbol{A}*\boldsymbol{B}$ means the Hadamard product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, i. e. $\{\boldsymbol{A}*\boldsymbol{B}\}_{i,j} = A_{i,j}B_{i,j}$ and $\mathrm{diag}(\mathbf{\Sigma})$ means the vector composed of the entries of the diagonal of the matrix $\mathbf{\Sigma}$.

If $\boldsymbol{W} \sim W_m(n-k, \mathbf{\Sigma})$, then

$$\boldsymbol{K} = \frac{1}{n-k}\left\{\mathrm{diag}(\mathbf{\Sigma})[\mathrm{diag}(\mathbf{\Sigma})]' + \mathbf{\Sigma} * \mathbf{\Sigma}\right\} \tag{7}$$

is the matrix with the following property. Its (i,j)th entry is the dispersion of $\widehat{\sigma}_{i,j} = \{\boldsymbol{W}\}_{i,j}/(n-k)$.

If $\delta\mathbf{\Sigma}$ is a matrix of infinitesimal shifts of the entries of the matrix $\mathbf{\Sigma}$, it is valid under the null hypothesis H_0:

$$T(\mathbf{\Sigma} + \delta\mathbf{\Sigma}) \approx \mathrm{Tr}\Big\{(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0) \times(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\delta\mathbf{\Sigma}\mathbf{\Sigma}^{-1})\Big\} = \chi^2_{mh}(0) + \xi,$$

where

$$\xi = -\mathrm{Tr}\Big\{(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)'[\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}']^{-1}(\boldsymbol{H}\widehat{\boldsymbol{B}} + \boldsymbol{H}_0)\mathbf{\Sigma}^{-1}\delta\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\Big\}.$$

Further

$$\xi \sim_1 \Big[-h\mathrm{Tr}(\mathbf{\Sigma}^{-1}\delta\mathbf{\Sigma}), 2h\mathrm{Tr}(\mathbf{\Sigma}^{-1}\delta\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\delta\mathbf{\Sigma})\Big].$$

Theorem 2.8. If H_0 is true and $\delta\mathbf{\Sigma} \in \mathcal{N}_{\Sigma_0}$, where

$$\begin{aligned}
\mathcal{N}_{\Sigma_0} &= \Big\{\delta\mathbf{\Sigma} : \big[\mathrm{vec}(\delta\mathbf{\Sigma}) - \boldsymbol{u}_0\big]'\boldsymbol{A}_0\big[\mathrm{vec}(\delta\mathbf{\Sigma}) - \boldsymbol{u}_0\big] \leq c^2\Big\},\\
\boldsymbol{u}_0 &= h\delta_{\max}\boldsymbol{A}_0^{-1}\mathrm{vec}(\mathbf{\Sigma}_0^{-1}),\\
\boldsymbol{A}_0 &= 2t^2h(\mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0) - h^2\mathrm{vec}(\mathbf{\Sigma}_0^{-1})[\mathrm{vec}(\mathbf{\Sigma}_0^{-1})]',\\
c^2 &= \delta^2_{\max} + h^2\delta^2_{\max}[\mathrm{vec}(\mathbf{\Sigma}_0^{-1})]'\boldsymbol{A}_0^{-1}[\mathrm{vec}(\mathbf{\Sigma}_0^{-1})],
\end{aligned}$$

$$\mathrm{P}\left\{\chi^2_{mh}(0) + \delta_{\max} \geq \chi^2_{mh}(0; 1-\alpha)\right\} = \alpha + \varepsilon,$$

then

$$\mathrm{P}\{T(\mathbf{\Sigma}_0 + \delta\mathbf{\Sigma}) \geq \chi^2_{mh}(0; 1-\alpha)\} \leq \alpha + \varepsilon.$$

P r o o f is analogous as in Theorem 2.3. □

Remark 2.9. Let $\boldsymbol{k} = \mathrm{vec}(\boldsymbol{K})$ from (7) and $\sqrt{\{\boldsymbol{k}\}_i} = \{\boldsymbol{l}\}_i,\ i = 1, \ldots, m^2$. The vector $\boldsymbol{l}$ is composed of the standard deviations $\sqrt{\mathrm{Var}(\widehat{\sigma}_{i,j})} = l_{i,j}$ of the estimators $\frac{1}{n-k}\{(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})'(\underline{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}})\}_{i,j}$ of $\{\boldsymbol{\Sigma}\}_{i,j} = \sigma_{i,j}$. The vector $\boldsymbol{l}$ generates the class of 2^{m^2} vectors which have the same absolute values of their coordinates, however different signs, e. g.

$$\boldsymbol{r} = (+l_{1,1}, -l_{1,2}, \ldots, +l_{1,m}, \ldots, +l_{2,1}, \ldots, +l_{2,m}, \ldots, -l_{m,1}, \ldots, -l_{m,m})'.$$

Now if the vectors $\boldsymbol{r}$ are sufficiently small with respect to the set $\mathcal{N}_{\Sigma_0}$, i. e.

$$-h[\mathrm{vec}(\boldsymbol{\Sigma}_0^{-1})]'\boldsymbol{r} + t\sqrt{2h\boldsymbol{r}'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\boldsymbol{r}} \ll \delta_{\max},$$

then the estimator of $\boldsymbol{\Sigma}$ can be used in the test (3). This check is rather rough, nevertheless for the first orientation is sufficient.

ACKNOWLEDGEMENT

This work was partially supported by the Ministry of Education, Youth and Sports of the Czech Republic under research project MSM 6 198 959 214.

(Received November 23, 2005.)

REFERENCES

[1] T. W. Anderson: Introduction to Multivariate Statistical Analysis. Wiley, New York 1958.

[2] L. Kubáček, L. Kubáčková, and J. Volaufová: Statistical Models with Linear Structures. Veda (Publishing House of Slovak Academy of Sciences), Bratislava 1995.

[3] E. Lešanská: Optimization of the size of nonsensitiveness regions. Appl. Math. *47* (2002), 9–23.

[4] C. R. Rao: Linear Statistical Inference and Its Applications. Second edition. Wiley, New York 1973.

[5] C. R. Rao and J. Kleffe: Estimation of Variance Components and Applications. North–Holland, Amsterdam 1988.

[6] C. R. Rao and S. K. Mitra: Generalized Inverse of Matrices and Its Applications. Wiley, New York 1971.

Lubomír Kubáček, Department of Mathematical Analysis and Applied Mathematics, Faculty of Science, Palacký University, Tomkova 40, 779 00 Olomouc. Czech Republic.
e-mail: kubacekl@aix.upol.cz