Sensitivity Analysis for the Decomposition of Mixed Partitioned Multivariate Models into Two Seemingly Unrelated Submodels

Eva Fišerová
Palacký University Olomouc

Lubomír Kubáček
Palacký University Olomouc

Abstract

The paper is focused on the decomposition of mixed partitioned multivariate models into two seemingly unrelated submodels in order to obtain more efficient estimators. The multiresponses are independently normally distributed with the same covariance matrix. The partitioned multivariate model is considered either with, or without an intercept. The elimination transformation of the intercept that preserves the BLUEs of parameter matrices and the MINQUE of the variance components in multivariate models with and without an intercept is stated. Procedures on testing the decomposition of the partitioned model are presented. The properties of plug-in test statistics as functions of variance components are investigated by sensitivity analysis and insensitivity regions for the significance level are proposed. The insensitivity region is a safe region in the parameter space of the variance components where the approximation of the variance components can be used without any essential deterioration of the significance level of the plug-in test statistic. The behavior of plug-in test statistics and insensitivity regions is studied by simulations.

Keywords: multivariate model, decomposition, plug-in statistic, joint test, variance components, insensitivity region.

1. Introduction

A multivariate approach to modeling (see, e.g., Anderson 1958; Kshirsagar 1972; Kubáček 2008; Seber 2004) has several advantages in comparison with a series of univariate models. Specifically, multivariate models respect the association between outcomes, and thus, in general, procedures are more efficient. Further, they can evaluate the joint influence of predictors on all outcomes and avoid the issue of multiple testing. On the other hand, there are situations when the multivariate model can be decomposed to a series of simpler models, univariate or multivariate, depending on the issue. Moreover, from a practical point of view, collecting data is usually easier in decomposed models.

The paper deals with a special case of a decomposition of a partitioned multivariate model with independent multiresponses with the same covariance matrix into two seemingly unrelated multivariate submodels (Zellner 1962) in order to obtain more efficient estimators. Namely, the multiresponse variables in the model are partitioned into two sets $\underline{\mathbf{Y}}^1$ and $\underline{\mathbf{Y}}^2$. Similarly,

the set of predictors is partitioned into two sets $\mathbf{X}_1$ and $\mathbf{X}_2$. As an example, let us consider the nutrigenomic study in the mouse. The response variable might be expressions of chosen genes ($\underline{\mathbf{Y}}_{i\cdot}^1$) and concentrations of hepatic fatty acids ($\underline{\mathbf{Y}}_{i\cdot}^2$) measured on subjects. The predictors might be genotype ($\mathbf{X}_1$) and type of diet ($\mathbf{X}_2$). The problem is to decide, roughly speaking, if it is possible to explain separately expressions of genes by genotype and hepatic fatty acids concentrations by diet or not. Fišerová and Kubáček (2012) proposed tests for the verification of the significance of a model decomposition under normality of random errors in the case when the covariance matrix is known or completely unknown. Further, Fišerová and Kubáček (2013) shown that the proposed tests may be used in models without an intercept, as well as in models with an intercept. These tasks are summarized in Section 2. Here, we will focused on the situation when the covariance matrix includes unknown variance components.

If variance components can be estimated via the maximum likelihood method, the technique of Kenward and Roger (1996) is useful for testing hypotheses about the decomposition of the model. Nevertheless, the maximum likelihood approach is suitable for replicated models or models with large number of observations. In the paper we consider a model without replications when the minimum norm quadratic unbiased estimators (MINQUE) based on Rao's procedure (Rao and Kleffe 1988) are used instead. This approach is valid even for models with small number of observations. The MINQU estimators are derived in Section 3. Estimated values of variance components can be plugged into the test statistic for a known covariance matrix. The investigation of statistical properties of a plug-in test statistic is rather difficult and therefore we will study the quality of a plug-in test statistic as a function of the variance components by sensitivity analysis. The sensitivity approach provides the so-called insensitivity regions (Kubáček 1996) in the space of variance components where the approximation of variance components do not cause any essential damage of the chosen statistical characteristic. Namely, we propose the insensitivity region for the significance level (Kubáček 2007b) as it is shown in Section 4. If we know that the true value of the variance components is with sufficiently high probability within the insensitivity region for the significance level, then the significance level of the plug-in test statistic does not exceed the chosen tolerable value. The sensitivity approach is investigated mostly in univariate models, e.g., Kubáček (1996); Fišerová and Kubáček (2003, 2004, 2006); Kubáček and Fišerová (2003); Lešanská (2002a,b). Some results for multivariate models are presented in Kubáček (2006, 2007a,b) and Fišerová and Kubáček (2009). The behavior of plug-in statistics and insensitivity regions is studied by simulations in Section 5.

2. Tests of the decomposition in case of a known covariance matrix

Let us consider the multivariate model in a partitioned form

$$
\left(\underset{(n\times p_1)}{\underline{\mathbf{Y}}^1} , \underset{(n\times p_2)}{\underline{\mathbf{Y}}^2} \right) = \left(\underset{(n\times k_1)}{\mathbf{X}_1} , \underset{(n\times k_2)}{\mathbf{X}_2} \right) \begin{pmatrix} \underset{(k_1\times p_1)}{\mathbf{B}_{11}} , & \underset{(k_1\times p_2)}{\mathbf{B}_{12}} \\ \underset{(k_2\times p_1)}{\mathbf{B}_{21}} , & \underset{(k_2\times p_2)}{\mathbf{B}_{22}} \end{pmatrix} + \left(\underset{(n\times p_1)}{\underline{\boldsymbol{\varepsilon}}_1} , \underset{(n\times p_2)}{\underline{\boldsymbol{\varepsilon}}_2} \right). \tag{1}
$$

Here $\underline{\mathbf{Y}} = (\underline{\mathbf{Y}}^1, \underline{\mathbf{Y}}^2)$ is a random matrix (observation matrix), $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ is a known design matrix, $\mathbf{B}_{11}$, $\mathbf{B}_{12}$, $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$ are matrices of unknown parameters and $\underline{\boldsymbol{\varepsilon}} = (\underline{\boldsymbol{\varepsilon}}_1, \underline{\boldsymbol{\varepsilon}}_2)$ is a random error matrix. We will assume that the matrix $\mathbf{X}$ is of full column rank, the multiresponses are independent with the same positive definite covariance matrix $\boldsymbol{\Sigma}$ and the random errors are normally distributed. The covariance matrix $\boldsymbol{\Sigma}$ of the multiresponse $\underline{\mathbf{Y}}_{i\cdot} = (Y_{i1}, Y_{i2}, \ldots, Y_{ip})'$ is partitioned in the same way, i.e.,

$$
\mathrm{var}\begin{pmatrix} \underline{\mathbf{Y}}_{i\cdot}^1 \\ \underline{\mathbf{Y}}_{i\cdot}^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}, & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21}, & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \ i = 1, 2, \ldots, \mathrm{n}. \tag{2}
$$

Further, let us consider a system of two seemingly unrelated (Zellner 1962) multivariate submodels

$$\underset{(n \times p_1)}{\underline{\mathbf{Y}}^1} = \underset{(n \times k_1)}{\mathbf{X}_1} \underset{(k_1 \times p_1)}{\mathbf{B}_1} + \underset{(n \times p_1)}{\underline{\boldsymbol{\varepsilon}}_1} \, , \qquad \underset{(n \times p_2)}{\underline{\mathbf{Y}}^2} = \underset{(n \times k_2)}{\mathbf{X}_2} \underset{(k_2 \times p_2)}{\mathbf{B}_2} + \underset{(n \times p_2)}{\underline{\boldsymbol{\varepsilon}}_2} \qquad (3)$$

with the covariance matrix $\boldsymbol{\Sigma}$ of the multiresponse $\underline{\mathbf{Y}}_{i\cdot}$ in the form (2). Note that models in (3) are seemingly unrelated because there is a link between them described by $\mathrm{cov}(\underline{\mathbf{Y}}_{i\cdot}^1, \underline{\mathbf{Y}}_{i\cdot}^2) = \boldsymbol{\Sigma}_{12}$. If $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, the models in (3) are independent. The problem is to decide which of the models (1) and (3) should be chosen for modeling in order to obtain more efficient estimators.

The issue with a decomposition of model (1) into (3) leads to testing the hypothesis that "the system of two seemingly unrelated multivariate submodels (3) is a true model", i.e., to test $\mathbf{B}_{12} = \mathbf{0}$ and $\mathbf{B}_{21} = \mathbf{0}$ simultaneously. If the covariance matrix $\boldsymbol{\Sigma}$ is known, Fišerová and Kubáček (2012) proposed the test statistics

$$T_{21} = \mathrm{Tr}\big[(\underline{\mathbf{Y}}^1)' \mathbf{M}_{X_1} \mathbf{X}_2 (\mathbf{X}_2' \mathbf{M}_{X_1} \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{M}_{X_1} \underline{\mathbf{Y}}^1 \boldsymbol{\Sigma}_{11}^{-1}\big] \sim \chi^2_{p_1 k_2} \text{ under } \mathbf{B}_{21} = \mathbf{0}, \qquad (4)$$

$$T_{12} = \mathrm{Tr}\big[(\underline{\mathbf{Y}}^2)' \mathbf{M}_{X_2} \mathbf{X}_1 (\mathbf{X}_1' \mathbf{M}_{X_2} \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{M}_{X_2} \underline{\mathbf{Y}}^2 \boldsymbol{\Sigma}_{22}^{-1}\big] \sim \chi^2_{p_2 k_1} \text{ under } \mathbf{B}_{12} = \mathbf{0}. \qquad (5)$$

The symbol $\mathrm{Tr}(\boldsymbol{\Sigma})$ denotes trace of the matrix $\boldsymbol{\Sigma}$ and $\mathbf{M}_{X_i} = \mathbf{I}_n - \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'$, $i = 1, 2$. To test the hypotheses $\mathbf{B}_{21} = \mathbf{0}$ and $\mathbf{B}_{12} = \mathbf{0}$ simultaneously, one can use, e.g., the Bonferroni correction in order to preserve the type I error rate α. More precisely, if $T_{21} \leq \chi^2_{p_1 k_2}(1 - \alpha/2)$ and $T_{12} \leq \chi^2_{p_2 k_1}(1 - \alpha/2)$, where $\chi^2_{p_1 k_2}(1 - \alpha/2)$ denotes the $(1 - \alpha/2)$-quantile of a $\chi^2_{p_1 k_2}$ distribution, neither of the hypotheses $\mathbf{B}_{21} = \mathbf{0}$, $\mathbf{B}_{12} = \mathbf{0}$ can be rejected on the significance level α.

Note that the decomposition of model (1) leads to two seemingly unrelated submodels. If the decomposition is significant, the prediction of $\underline{\mathbf{Y}}^1$ conditional on $\mathbf{X}_1$ is not improved also by regressing on $\mathbf{X}_2$. However the predictors $\mathbf{X}_2$ are necessary for the calculation of the prediction of $\underline{\mathbf{Y}}^1$. Analogous conclusions hold for the prediction of $\underline{\mathbf{Y}}^2$.

Until now we have considered only the model without an intercept. A partitioned form of the model with the intercept can be written as

$$\begin{pmatrix} \underset{(n \times p_1)}{\underline{\mathbf{Y}}^1}, & \underset{(n \times p_2)}{\underline{\mathbf{Y}}^2} \end{pmatrix} = \begin{pmatrix} \underset{(n \times 1)}{\mathbf{1}}, & \underset{(n \times k_1)}{\mathbf{X}_1}, & \underset{(n \times k_2)}{\mathbf{X}_2} \end{pmatrix} \begin{pmatrix} \underset{(1 \times p_1)}{\mathbf{b}_1}, & \underset{(1 \times p_2)}{\mathbf{b}_2} \\ \underset{(k_1 \times p_1)}{\mathbf{B}_{11}}, & \underset{(k_1 \times p_2)}{\mathbf{B}_{12}} \\ \underset{(k_2 \times p_1)}{\mathbf{B}_{21}}, & \underset{(k_2 \times p_2)}{\mathbf{B}_{22}} \end{pmatrix} + \begin{pmatrix} \underset{(n \times p_1)}{\underline{\boldsymbol{\varepsilon}}_1}, & \underset{(n \times p_2)}{\underline{\boldsymbol{\varepsilon}}_2} \end{pmatrix}, \qquad (6)$$

where $\mathbf{1}$ is a vector of ones. We will assume that the design matrix $(\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2)$ is of full column rank, and therefore all regression parameters are unbiasedly estimable. If the model includes also an intercept, then the question is, where the intercept should go in the decomposed model, in $\mathbf{X}_1$, in $\mathbf{X}_2$, or in both $\mathbf{X}_1$ and $\mathbf{X}_2$. Naturally, all cases are possible and results depend on particular tasks. To avoid this situation, Fišerová and Kubáček (2013) proposed the transformation for an elimination of the intercept that leads to the identical BLUEs of parameter matrices $\mathbf{B}_{11}$, $\mathbf{B}_{12}$, $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$ in model (6) with the intercept and model without the intercept in the form

$$\big(\mathbf{M}_1 \underline{\mathbf{Y}}^1, \mathbf{M}_1 \underline{\mathbf{Y}}^2\big) = \big(\mathbf{M}_1 \mathbf{X}_1, \mathbf{M}_1 \mathbf{X}_2\big) \begin{pmatrix} \mathbf{B}_{11}, & \mathbf{B}_{12} \\ \mathbf{B}_{21}, & \mathbf{B}_{22} \end{pmatrix} + \big(\mathbf{M}_1 \underline{\boldsymbol{\varepsilon}}_1, \mathbf{M}_1 \underline{\boldsymbol{\varepsilon}}_2\big). \qquad (7)$$

Here $\mathbf{M}_1 = \mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$. Therefore testing the decomposition of the partitioned model with the intercept can be done similarly as for the model without the intercept. The process is the following. First we transform model (6) to (7). Next we test hypotheses $\mathbf{B}_{12} = \mathbf{0}$ and $\mathbf{B}_{21} = \mathbf{0}$ simultaneously via test statistics T_{12}, T_{21} using the substitution

$$\underline{\mathbf{Y}}^j \rightarrow \mathbf{M}_1 \underline{\mathbf{Y}}^j, \ \mathbf{X}_j \rightarrow \mathbf{M}_1 \mathbf{X}_j, \ \boldsymbol{\Sigma} \rightarrow \mathbf{M}_1 \boldsymbol{\Sigma} \mathbf{M}_1, \ j = 1, 2.$$

Obviously, test statistics T_{12} and T_{21} have the same degrees of freedom in the case of the model without the intercept since the assumptions on full column rank of the design matrices imply that the ranks of the transformed design matrices $\mathbf{M}_1\mathbf{X}_j$ are equal to k_j, $j = 1, 2$, as well.

3. Tests of the decomposition in case of a covariance matrix with unknown variance components

Now we will consider the covariance matrix $\mathbf{\Sigma}$ of the structure $\mathbf{\Sigma} = \sum_{i=1}^s \vartheta_i \mathbf{V}_i$, where $\mathbf{V}_1, \ldots, \mathbf{V}_s$ are known $(p_1 + p_2) \times (p_1 + p_2)$ symmetric and positive semidefinite matrices and $\vartheta_i > 0, i = 1, \ldots, s$, are unknown parameters (variance components). Denote $\mathbf{\Sigma}_0 = \sum_{i=1}^s \vartheta_{0,i} \mathbf{V}_i$, where $\boldsymbol{\vartheta}_0 = (\vartheta_{0,1}, \ldots, \vartheta_{0,s})'$ is an approximate value of the vector parameter $\boldsymbol{\vartheta}$. The $\boldsymbol{\vartheta}_0$-locally minimum norm quadratic unbiased estimator ($\boldsymbol{\vartheta}_0$-LMINQUE) of $\boldsymbol{\vartheta}$ based on Rao's procedure (Rao and Kleffe 1988) is stated in the following lemma.

Lemma 1 The $\boldsymbol{\vartheta}_0$-locally MINQUE of $\boldsymbol{\vartheta}$ in the model (1) is

$$\widehat{\boldsymbol{\vartheta}} = \frac{1}{n - k_1 - k_2} \mathbf{S}_{\Sigma_0^{-1}}^{-1} \widehat{\boldsymbol{\gamma}},$$

where the ith component of the vector $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_s)'$ is

$$\widehat{\gamma}_i = \mathrm{Tr}\left[\left(\begin{array}{c} (\underline{\mathbf{Y}}^1)' \\ (\underline{\mathbf{Y}}^2)' \end{array} \right) \mathbf{M}_{(X_1, X_2)}(\underline{\mathbf{Y}}^1, \underline{\mathbf{Y}}^2) \mathbf{\Sigma}_0^{-1} \mathbf{V}_i \mathbf{\Sigma}_0^{-1} \right]$$

and the (i, j)th element of $(s \times s)$ matrix $\mathbf{S}_{\Sigma_0^{-1}}$ is

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

Under normality, the covariance matrix of the estimator $\widehat{\boldsymbol{\vartheta}}$ at the point $\boldsymbol{\vartheta}_0$ is

$$\mathrm{var}_{\vartheta_0}(\widehat{\boldsymbol{\vartheta}}) = \frac{2}{n - k_1 - k_2} \mathbf{S}_{\Sigma_0^{-1}}^{-1}. \tag{8}$$

Proof. For simplicity, the proof proceeds for the univariate form of model (1) which can be expressed as

$$\mathrm{vec}(\underline{\mathbf{Y}}^1, \underline{\mathbf{Y}}^2) \sim N_{n(p_1+p_2)}\left\{ \left[\mathbf{I}_{p_1+p_2} \otimes (\mathbf{X}_1, \mathbf{X}_2) \right] \mathrm{vec} \left(\begin{array}{cc} \mathbf{B}_{11}, & \mathbf{B}_{12} \\ \mathbf{B}_{21}, & \mathbf{B}_{22} \end{array} \right), \sum_{i=1}^s \vartheta_i \mathbf{V}_i \otimes \mathbf{I}_n \right\}. \tag{9}$$

Here, the symbol $\mathrm{vec}(\underline{\mathbf{Y}}^1)$ denotes the column vector composed of the columns of $\underline{\mathbf{Y}}^1$. The notation $\otimes$ means the Kronecker multiplication of matrices (Rao and Mitra 1988). Then, according to Rao and Kleffe (1988), the $\boldsymbol{\vartheta}_0$-locally MINQUE of the vector parameter $\boldsymbol{\vartheta}$ in model (9) is

$$\widehat{\boldsymbol{\vartheta}} = \mathbf{S}_{\mathbf{D}}^{-1} \widehat{\boldsymbol{\gamma}}, \quad \mathbf{D} = [\mathbf{M}_{[I \otimes (X_1, X_2)]}(\mathbf{\Sigma}_0 \otimes \mathbf{I}) \mathbf{M}_{[I \otimes (X_1, X_2)]}]^+,$$

where the ith component of the vector $\widehat{\boldsymbol{\gamma}}$ is given as

$$\widehat{\gamma}_i = \left[\mathrm{vec}(\underline{\mathbf{Y}}^1, \underline{\mathbf{Y}}^2) \right]' \mathbf{D}(\mathbf{V}_i \otimes \mathbf{I}) \mathbf{D} \mathrm{vec}(\underline{\mathbf{Y}}^1, \underline{\mathbf{Y}}^2), \ i = 1, \ldots, s.$$

Now it is sufficient to take into account the equality $\mathbf{D} = \mathbf{\Sigma}_0^{-1} \otimes \mathbf{M}_{(X_1, X_2)}$, which is substituted into the previous formulas, and to simplify each of the expressions. □

The estimator of the variance components and its covariance matrix depends on approximate values $\boldsymbol{\vartheta}_0$. To eliminate this dependency, it is necessary to use an iterative procedure. The calculated estimated values of the variance components are used in the next iteration as

approximate ones. The iterative procedure is very robust. It usually stops after two iterations for any initial value of the variance components even for distributions different from the Gaussian (Bognárová, Kubáček, and Volaufová 1996).

Note that the iterated MINQUE is practically the same as the maximum likelihood estimator in the case of Gaussian distribution. Moreover, the MINQUE procedure can be used even for negative variance components and symmetric matrices $\mathbf{V}_i$ for errors with normal distribution. The formulas for the estimators are the same as under the assumption of positive variance components and p.s.d. matrices $\mathbf{V}_i$.

If model (1) can be decomposed into (3), the variance components can be estimated on the basis of either $\underline{\mathbf{Y}}^1$ or $\underline{\mathbf{Y}}^2$ in model (3). The explicit formulas for $\boldsymbol{\vartheta}_0$-LMINQUE in the submodels follows directly from Lemma 1. Particularly, using the notation $\boldsymbol{\Sigma}_{jj,0}$ for an approximate value of matrix $\boldsymbol{\Sigma}_{jj}$, $j = 1, 2$, the expressions for $\boldsymbol{\vartheta}_0$-LMINQUE of $\boldsymbol{\vartheta}$ in submodels (3) are equal to

$$\widehat{\boldsymbol{\vartheta}} = \frac{1}{n - k_j} \mathbf{S}_{\boldsymbol{\Sigma}_{jj,0}^{-1}}^{-1} \widehat{\boldsymbol{\gamma}}, \quad \mathrm{var}_{\vartheta_0}(\widehat{\boldsymbol{\vartheta}}) = \frac{2}{n - k_j} \mathbf{S}_{\boldsymbol{\Sigma}_{jj,0}^{-1}}^{-1}, \quad j = 1, 2.$$

The ith component of the vector $\widehat{\boldsymbol{\gamma}}$ and the (p, q)th element of the matrix $\mathbf{S}_{\boldsymbol{\Sigma}_{jj,0}^{-1}}$ are given as

$$\widehat{\gamma}_i = \mathrm{Tr}\left[(\underline{\mathbf{Y}}^j)' \mathbf{M}_{X_j} \underline{\mathbf{Y}}^j \boldsymbol{\Sigma}_{jj,0}^{-1} \mathbf{V}_i \boldsymbol{\Sigma}_{jj,0}^{-1}\right], \quad \left\{\mathbf{S}_{\boldsymbol{\Sigma}_{jj,0}^{-1}}\right\}_{p,q} = \mathrm{Tr}(\boldsymbol{\Sigma}_{jj,0}^{-1} \mathbf{V}_p \boldsymbol{\Sigma}_{jj,0}^{-1} \mathbf{V}_q).$$

If the mixed partitioned model includes also an intercept, the elimination transformation (7) can be used. This transformation preserves not only the BLUEs of the regression parameters matrices, but also the estimates of the variance components, as it will be shown in the following theorem.

Theorem 1. The $\boldsymbol{\vartheta}_0$-locally MINQUE of $\boldsymbol{\vartheta}$ in models (6) and (7) are the same.

Proof. For the sake of simplicity, the proof proceeds for the univariate form of model (6). Let us denote

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \mathrm{vec}(\underline{\boldsymbol{\varepsilon}}_1) \\ \mathrm{vec}(\underline{\boldsymbol{\varepsilon}}_2) \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} \mathrm{vec}(\underline{\mathbf{Y}}^1) \\ \mathrm{vec}(\underline{\mathbf{Y}}^2) \end{pmatrix}, \quad \boldsymbol{\Sigma} \otimes \mathbf{I} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} \otimes \mathbf{I}, & \boldsymbol{\Sigma}_{12} \otimes \mathbf{I} \\ \boldsymbol{\Sigma}_{21} \otimes \mathbf{I}, & \boldsymbol{\Sigma}_{22} \otimes \mathbf{I} \end{pmatrix},$$

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{I} \otimes \mathbf{1}, & \mathbf{0} \\ \mathbf{0}, & \mathbf{I} \otimes \mathbf{1} \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} \mathbf{I} \otimes \mathbf{X}_1, & \mathbf{0}, & \mathbf{I} \otimes \mathbf{X}_2, & \mathbf{0} \\ \mathbf{0}, & \mathbf{I} \otimes \mathbf{X}_2, & \mathbf{0}, & \mathbf{I} \otimes \mathbf{X}_1 \end{pmatrix}$$

$$\boldsymbol{\beta}_1 = (\mathbf{b}_1, \mathbf{b}_2)', \quad \boldsymbol{\beta}_2 = \left(\mathrm{vec}(\mathbf{B}_{11})', \mathrm{vec}(\mathbf{B}_{22})', \mathrm{vec}(\mathbf{B}_{21})', \mathrm{vec}(\mathbf{B}_{12})'\right)'.$$

Then the model (6) can be rewritten as

$$\mathbf{Y} = (\mathbf{A}_1, \mathbf{A}_2) \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\varepsilon}. \tag{10}$$

By Rao and Kleffe (1988), the $\boldsymbol{\vartheta}_0$-locally MINQUE of $\boldsymbol{\vartheta}$ in model (10) is given as

$$\widehat{\boldsymbol{\vartheta}} = \mathbf{S}_G^{-1} \widehat{\boldsymbol{\gamma}}, \quad \mathbf{G} = [\mathbf{M}_{(A_1, A_2)}(\boldsymbol{\Sigma}_0 \otimes \mathbf{I})\mathbf{M}_{(A_1, A_2)}]^+,$$

where $\widehat{\gamma}_i = \mathbf{Y}' \mathbf{G}(\mathbf{V}_i \otimes \mathbf{I})\mathbf{G}\mathbf{Y}$, $i = 1, \ldots, s$. Further, using the relationships

$$\mathbf{P}_{A_1} = \mathbf{A}_1(\mathbf{A}_1' \mathbf{A}_1)^{-1} \mathbf{A}_1', \ \mathbf{P}_{(A_1, A_2)} = \mathbf{P}_{A_1} + \mathbf{P}_{M_{A_1} A_2},$$

$$\mathbf{P}_{(A_1, A_2)} \mathbf{P}_{A_1} = \mathbf{P}_{A_1}, \ \mathbf{M}_{(A_1, A_2)} = \mathbf{M}_{A_1} - \mathbf{P}_{M_{A_1} A_2},$$

we obtain

$$\mathbf{M}_{(A_1, A_2)} \mathbf{M}_{A_1} = \mathbf{M}_{(A_1, A_2)}, \quad \mathbf{M}_{(A_1, A_2)} \mathbf{M}_{M_{A_1} A_2} = \mathbf{M}_{A_1} \mathbf{M}_{M_{A_1} A_2}.$$

Thus, the matrix $\mathbf{G}$ can be rewritten as

$$\mathbf{G} = [\mathbf{M}_{(A_1, A_2)}(\boldsymbol{\Sigma}_0 \otimes \mathbf{I})\mathbf{M}_{(A_1, A_2)}]^+ = [\mathbf{M}_{M_{A_1} A_2}(\mathbf{M}_{A_1}(\boldsymbol{\Sigma}_0 \otimes \mathbf{I})\mathbf{M}_{A_1})\mathbf{M}_{M_{A_1} A_2}]^+ = \widetilde{\mathbf{G}}.$$

Therefore, the equalities

$$\mathbf{G}(\mathbf{V}_i \otimes \mathbf{I})\mathbf{G} = \widetilde{\mathbf{G}}\mathbf{M}_{A_1}(\mathbf{V}_i \otimes \mathbf{I})\mathbf{M}_{A_1}\widetilde{\mathbf{G}},$$

$$\widehat{\gamma}_i = \mathbf{Y}'\mathbf{G}(\mathbf{V}_i \otimes \mathbf{I})\mathbf{G}\mathbf{Y} = \mathbf{Y}'\mathbf{M}_{A_1}\widetilde{\mathbf{G}}\mathbf{M}_{A_1}(\mathbf{V}_i \otimes \mathbf{I})\mathbf{M}_{A_1}\widetilde{\mathbf{G}}\mathbf{M}_{A_1}\mathbf{Y} = \widetilde{\gamma}_i,$$

also holds. Summarizing the above results we obtain $\widehat{\boldsymbol{\vartheta}} = \mathbf{S}_G^{-1}\widehat{\boldsymbol{\gamma}} = \mathbf{S}_{\widetilde{G}}^{-1}\widetilde{\boldsymbol{\gamma}}$, i.e., the $\boldsymbol{\vartheta}_0$-LMINQUE of $\boldsymbol{\vartheta}$ in model (10) and in the model $\mathbf{M}_{A_1}\mathbf{Y} = \mathbf{M}_{A_1}\mathbf{A}_2\boldsymbol{\beta}_2 + \mathbf{M}_{A_1}\boldsymbol{\varepsilon}$ are the same. Using the relationship $\mathbf{M}_{A_1} = \mathrm{diag}\{\mathbf{I} \otimes \mathbf{M}_1, \mathbf{I} \otimes \mathbf{M}_1\}$ (diagonal matrix), the last model can be rewritten into multivariate form (7), and thus the proof is finished. $\square$

If the variance components are estimated, the hypothesis about the decomposition of model (1) can be tested by the plug-in statistics T_{21} and T_{12}, when the matrices $\widehat{\boldsymbol{\Sigma}}_{jj} = \sum_{i=1}^{s} \widehat{\vartheta}_i\mathbf{V}_{i,(jj)}$, where $\mathbf{V}_{i,(jj)}$ is the corresponding part of $\mathbf{V}_i$, $j = 1, 2$, are plugged into formulas (4). Testing the decomposition of the model with the intercept (6) can be done by the plug-in statistics T_{21} and T_{12} for the transformed model (7), since the transformation affects neither the BLUEs of the regression parameter matrices nor the MINQUE of the variance components.

Obviously, the substitution of the true values of the variance components by their estimated values influences the optimum quality of the estimators $\widehat{\mathbf{B}}_{21}$, $\widehat{\mathbf{B}}_{12}$ and, consequently, the significance level and the power of the test. The investigation of statistical properties of the plug-in test statistics T_{21} and T_{12} is rather difficult and therefore we will study the quality of the plug-in test statistic as a function of the variance components by sensitivity analysis as it is shown in the next section.

4. Sensitivity analysis for the significance level

The main idea of the sensitivity approach (Kubáček 1996) is to consider the plug-in statistic as a function of the variance components and to find a safe region in the parameter space of the variance components where the approximation of the variance components does not cause any essential damage of the significance level of the plug-in test statistic (Kubáček 2007b). The plug-in test statistic can have a higher significance level. Let $\varepsilon > 0$ be the maximum admissible increase of the significance level. The goal is to find a region in the parameter space of the variance components such that shifts $\delta\boldsymbol{\vartheta}$ around the true value $\boldsymbol{\vartheta}^*$ within this region cause the significance level of the plug-in test statistic T_{21} to be not greater than $\alpha/2 + \varepsilon/2$. (We consider the significance level $\alpha/2 + \varepsilon/2$ since the Bonferroni correction for multiple tests on $\mathbf{B}_{21} = \mathbf{0}$ and $\mathbf{B}_{12} = \mathbf{0}$ is used.) Such a region is called an *insensitivity region for the significance level* and will be denoted by $\mathcal{N}_{\varepsilon, T_{21}}$. More precisely, $\mathcal{N}_{\varepsilon, T_{21}}$ is a neighborhood of the vector $\boldsymbol{\vartheta}^*$ with the property

$$\boldsymbol{\vartheta} \in \mathcal{N}_{\varepsilon, T_{21}} \Rightarrow P\left\{T_{21}(\boldsymbol{\vartheta}) \leq \chi^2_{k_2p_1}(1 - \alpha/2)\right\} \geq 1 - \alpha/2 - \varepsilon/2.$$

The derivation of the insensitivity region for the significance level is based on an approximation of the plug-in test statistic T_{21} by $T_{12}(\boldsymbol{\vartheta}) = T_{21}(\boldsymbol{\vartheta}^*) + \delta T_{21}$. The variable $\delta T_{21} = \delta\boldsymbol{\vartheta}'\frac{\partial T_{21}}{\partial\boldsymbol{\vartheta}}$ characterizes the change of the statistic $T_{21}(\boldsymbol{\vartheta}^*)$ caused by the shift $\delta\boldsymbol{\vartheta}$ around $\boldsymbol{\vartheta}^*$. Obviously, the significance level of T_{21} increases with increasing δT_{21} and vice versa. Hence the problem is to find the upper limit for δT_{21} so that the significance level increased by a maximum tolerated value. Using the Chebyshev inequality it holds that

$$P\left\{|\delta T_{21} - E(\delta T_{21})| \geq t\sqrt{\mathrm{var}(\delta T_{21})}\right\} \leq \frac{1}{t^2}, \ t > 0. \tag{11}$$

The inequality (11) together with the probability statement for the tolerated significance level

$$P\left\{T_{21} + \delta T_{21} \geq \chi^2_{k_2p_1}(1 - \alpha/2)\right\} \leq \alpha/2 + \varepsilon/2,$$

implies that for a sufficiently large $t > 0$ such that

$$P\{\delta T_{21} < E(\delta T_{21}) + t\sqrt{\text{var}(\delta T_{21})}\} \approx 1, \tag{12}$$

the inequality $E(\delta T_{21}) + t\sqrt{\text{var}(\delta T_{21})} \leq \delta_{\varepsilon, T_{21}}$, where

$$\delta_{\varepsilon, T_{21}} = \chi^2_{k_2 p_1}(1 - \alpha/2) - \chi^2_{k_2 p_1}(1 - \alpha/2 - \varepsilon/2), \tag{13}$$

is a sufficient condition for the upper limit for δT_{21}. The explicit form of the insensitivity region $\mathcal{N}_{\varepsilon, T_{21}}$ is stated in Theorem 2.

Theorem 2 The insensitivity region $\mathcal{N}_{\varepsilon, T_{21}}$ for the significance level of the statistic T_{21} is

$$\mathcal{N}_{\varepsilon, T_{21}} = \left\{ \boldsymbol{\vartheta}^* + \delta\boldsymbol{\vartheta} : \left(\delta\boldsymbol{\vartheta} - \delta_{\varepsilon, T_{21}}\mathbf{A}^{-1}\mathbf{a}\right)' \mathbf{A} \left(\delta\boldsymbol{\vartheta} - \delta_{\varepsilon, T_{21}}\mathbf{A}^{-1}\mathbf{a}\right) \leq \delta^2_{\varepsilon, T_{21}}\left(1 + \mathbf{a}'\mathbf{A}^{-1}\mathbf{a}\right) \right\},$$

where $\delta_{\varepsilon, T_{21}}$ is given by (13),

$$\mathbf{a} = k_2\left[\text{Tr}(\boldsymbol{\Sigma}^{-1}_{11}\mathbf{V}_{1,(11)}), \ldots, \text{Tr}(\boldsymbol{\Sigma}^{-1}_{11}\mathbf{V}_{s,(11)})\right]', \quad \mathbf{A} = 2t^2 k_2 \mathbf{S}_{\boldsymbol{\Sigma}^{-1}_{11}} - \mathbf{aa}',$$

and $t > 0$ is a sufficiently large number such that the probability statement (12) holds.

Proof. It is necessary to determine the mean value and the variance of variable δT_{21} which characterizes a change of the statistic $T_{21}(\boldsymbol{\vartheta}^*)$ caused by the shift $\delta\boldsymbol{\vartheta}$. Let

$$\xi_i = \left.\frac{\partial T_{21}(\boldsymbol{\vartheta})}{\partial \vartheta_i}\right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*} = -\text{Tr}\left[\underline{\mathbf{Y}}'_1 \mathbf{P}_{M_{X_1} X_2} \underline{\mathbf{Y}}_1 \boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\mathbf{V}_{i,(11)}\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\right], \quad i = 1, \ldots, s.$$

The mean value of the variable ξ_i is

$$E_{\boldsymbol{\vartheta}^*}(\xi_i) = -\text{Tr}\left(\left\{\left[\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\mathbf{V}_{i,(11)}\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\right] \otimes \mathbf{P}_{M_{X_1} X_2}\right\} \left[\boldsymbol{\Sigma}_{11}(\boldsymbol{\vartheta}^*) \otimes \mathbf{I}\right]\right)$$
$$= -\text{Tr}(\mathbf{P}_{M_{X_1} X_2})\text{Tr}\left[\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\mathbf{V}_{i,(11)}\right] = -k_2\text{Tr}\left[\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\mathbf{V}_{i,(11)}\right].$$

Further we calculate the covariance between the variables ξ_i and ξ_j:

$$\text{cov}_{\boldsymbol{\vartheta}^*}(\xi_i, \xi_j) = 2\text{Tr}\left(\left\{\left[\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\mathbf{V}_{i,(11)}\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\right] \otimes \mathbf{P}_{M_{X_1} X_2}\right\} \left[\boldsymbol{\Sigma}_{11}(\boldsymbol{\vartheta}^*) \otimes \mathbf{I}\right]\right.$$
$$\left. \times \left\{\left[\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\mathbf{V}_{j,(11)}\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\right] \otimes \mathbf{P}_{M_{X_1} X_2}\right\} \left[\boldsymbol{\Sigma}_{11}(\boldsymbol{\vartheta}^*) \otimes \mathbf{I}\right]\right)$$
$$= 2\text{Tr}(\mathbf{P}_{M_{X_1} X_2})\text{Tr}\left[\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\mathbf{V}_{i,(11)}\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)\mathbf{V}_{j,(11)}\right] = 2k_2\left\{\mathbf{S}_{\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)}\right\}_{i,j}.$$

Now we are able to determine the upper limit for $\delta T_{21} = \delta\boldsymbol{\vartheta}'\boldsymbol{\xi}$. Let $t > 0$ be a sufficiently large number such that (12) holds, i.e., with probability sufficiently near to one it is true that

$$\delta\boldsymbol{\vartheta}'\boldsymbol{\xi} < E_{\boldsymbol{\vartheta}^*}(\boldsymbol{\xi}')\delta\boldsymbol{\vartheta} + t\sqrt{\delta\boldsymbol{\vartheta}'\text{var}_{\boldsymbol{\vartheta}^*}(\boldsymbol{\xi})\delta\boldsymbol{\vartheta}} \leq \delta_{\varepsilon, T_{21}}.$$

Substituting the mean value and the covariance matrix of the vector $\boldsymbol{\xi}$, and by a simple calculation we obtain the inequality

$$\delta\boldsymbol{\vartheta}'(2t^2 k_2 \mathbf{S}_{\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)} - \mathbf{aa}')\delta\boldsymbol{\vartheta} - 2\mathbf{a}'\delta\boldsymbol{\vartheta} \leq \delta^2_{\varepsilon, T_{21}},$$

which is equivalent with

$$\left[\delta\boldsymbol{\vartheta} - \delta_{\varepsilon, T_{21}}(2t^2 k_2 \mathbf{S}_{\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)} - \mathbf{aa}')^{-1}\mathbf{a}\right]'(2t^2 k_2 \mathbf{S}_{\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)} - \mathbf{aa}')$$
$$\times \left[\delta\boldsymbol{\vartheta} - \delta_{\varepsilon, T_{21}}(2t^2 k_2 \mathbf{S}_{\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)} - \mathbf{aa}')^{-1}\mathbf{a}\right] \leq \delta^2_{\varepsilon, T_{21}} + \delta^2_{\varepsilon, T_{21}}\mathbf{a}'(2t^2 k_2 \mathbf{S}_{\boldsymbol{\Sigma}^{-1}_{11}(\boldsymbol{\vartheta}^*)} - \mathbf{aa}')^{-1}\mathbf{a},$$

thereby the statement is proved. $\qquad\square$

Analogous considerations can be made for the plug-in test statistic T_{12}. In this case the explicit formula for the insensitivity region for the significance level results in

$$\mathcal{N}_{\varepsilon,T_{12}} = \left\{ \boldsymbol{\vartheta}^* + \delta\boldsymbol{\vartheta} : \left(\delta\boldsymbol{\vartheta} - \delta_{\varepsilon,T_{12}}\mathbf{B}^{-1}\mathbf{b}\right)' \mathbf{B} \left(\delta\boldsymbol{\vartheta} - \delta_{\varepsilon,T_{12}}\mathbf{B}^{-1}\mathbf{b}\right) \le \delta_{\varepsilon,T_{12}}^2 \left(1 + \mathbf{b}'\mathbf{B}^{-1}\mathbf{b}\right) \right\},$$

where

$$\delta_{\varepsilon,T_{12}} = \chi_{k_1 p_2}^2(1 - \alpha/2) - \chi_{k_1 p_2}^2(1 - \alpha/2 - \varepsilon/2),$$
$$\mathbf{b} = k_1 \left[\mathrm{Tr}(\boldsymbol{\Sigma}_{22}^{-1}\mathbf{V}_{1,(22)}), \ldots, \mathrm{Tr}(\boldsymbol{\Sigma}_{22}^{-1}\mathbf{V}_{p,(22)})\right]', \quad \mathbf{B} = 2t^2 k_1 \mathbf{S}_{\Sigma_{22}^{-1}} - \mathbf{b}\mathbf{b}'.$$

The size of the insensitivity region depends on the parameters ε and t chosen by the user. The parameter ε is related to the user's opinion that ε causes a tolerable increase of the significance level. The larger ε, the larger the insensitivity region, but also a higher significance level follows. The parameter t corresponds to the approximation of the plug-in test statistic, namely with the upper limit for the variable δT_{21} that describes the change of the statistic $T_{21}(\boldsymbol{\vartheta}^*)$ caused by the shift $\delta\boldsymbol{\vartheta}$. For $t = 5$, from the Chebyshev inequality (11) it follows that at least 96% of the data values of δT_{21} must be within 5 standard deviations of the mean or, equivalently, no more than 4% of the data values can be more than 5 standard deviations away from the mean. If δT_{21} is approximately normally distributed, at least 99.7% of the data values of δT_{21} must be within 3 standard deviations of the mean. Hence it is reasonable to choose the parameter t in the interval $\langle 3, 5 \rangle$. The smaller t, the larger the insensitivity region but also cases a higher tail probability. The procedure for the optimal choice of the parameter t that maximizes the size of the insensitivity region is derived in Lešanská (2002a).

Both insensitivity regions $\mathcal{N}_{\varepsilon,T_{12}}$ and $\mathcal{N}_{\varepsilon,T_{21}}$ are suitable for a justification of the utilization of plug-in joint tests T_{12}, T_{21} for a decomposition of model (1) into two seemingly unrelated submodels (3). The process is as follows. First, we determine estimates of the variance components. Then we compute the insensitivity regions $\mathcal{N}_{\varepsilon,T_{12}}$ and $\mathcal{N}_{\varepsilon,T_{21}}$ for the estimated values of the variance components and chosen values ε and t. Finally, we set the confidence domain for the variance components for a sufficiently high confidence level and check whether this confidence domain is embedded into the insensitivity regions. If this confidence domain is included into both insensitivity regions, plug-in joint tests are admissible and, moreover, the significance level of plug-in joint tests does not exceed the value of $\alpha + \varepsilon$. If the confidence domain is not embedded into both insensitivity regions, the experiment requires better design, other measurement devices, or more observations to be sure that the approximation of the variance components by their estimates do not cause an increase in significance level by more than a tolerable ε. The criterion is very demanding, in some cases the confidence domain is not embedded into the insensitivity regions, however, the estimated values of the variance components lie in the insensitivity regions what implies that the increase of the significance level is almost a tolerable one (see Section 5).

The determination of an exact confidence domain for the variance components is difficult since the distribution of the estimator $\widehat{\boldsymbol{\vartheta}}$ is unknown even for a normally distributed vector $\mathrm{vec}(\mathbf{Y})$. Some approximation can be derived using the Bonferroni inequality and the Chebyshev inequality which imply that

$$P\left\{\forall i = 1, \ldots, s : |\widehat{\vartheta}_i - E(\widehat{\vartheta}_i)| \le \sqrt{\frac{s}{\alpha}}\sqrt{\mathrm{var}(\widehat{\vartheta}_i)}\right\} \ge 1 - \alpha.$$

Hence at least a $(1 - \alpha)100\%$-confidence domain for the variance components is a set given by the Cartesian product

$$I_{1-\alpha}(\boldsymbol{\vartheta}) = \mathrm{X}_{i=1}^{s} \left\{ u : |u - \widehat{\vartheta}_i| \le \sqrt{\frac{s}{\alpha}}\sqrt{\mathrm{var}(\widehat{\vartheta}_i)} \right\}.$$

Recall that the covariance matrix of the variance components estimator is given by (8).

It can be easily proved, similarly as in Lešanská (2002b), that a k^2-multiple of $\boldsymbol{\vartheta}^*$ makes a homothetic change of the boundary of the insensitivity region for the significance level with the coefficient k^2 and the centre at the point 0.

5. Simulation study

By simulations we will study the behavior of the plug-in test statistics T_{12}, T_{21} for the decomposition of model (1) and the insensitivity regions for the significance level. We will consider different choices of the covariance matrix, parameter matrices, number of observations and the true model (model (1) or the system of seemingly unrelated submodels (3)).

We considered $n = 40$ and $n = 400$ observations, a multiresponse $\underline{\mathbf{Y}}_{i\cdot}^j$, $j = 1, 2$, with dimensions $p_1 = 3$ and $p_2 = 4$, and the number of regressors equal to $k_1 = 2$ and $k_2 = 2$. The parameter matrices were chosen as

$$\mathbf{B}_1 = \mathbf{B}_{11} = \begin{pmatrix} 3, & 2, & 2 \\ 2, & 3, & 3 \end{pmatrix}, \quad \mathbf{B}_2 = \mathbf{B}_{22} = \begin{pmatrix} 2, & 4, & 4, & 1.5 \\ 4, & 2, & 4, & 4 \end{pmatrix}, \tag{14}$$

$$\mathbf{B}_{12} = \begin{pmatrix} 1, & 3, & 1, & 15 \\ 2, & 7, & 8, & 3 \end{pmatrix}, \quad \mathbf{B}_{21} = \begin{pmatrix} 4, & 4, & 1 \\ 2, & 8, & 3 \end{pmatrix}. \tag{15}$$

The design matrices were considered in the form of $\mathbf{X} = \mathbf{1}_{10} \otimes \mathbf{T}$ and $\mathbf{X} = \mathbf{1}_{100} \otimes \mathbf{T}$, with the matrix $\mathbf{T}$

$$\mathbf{T} = \begin{pmatrix} 1, & 0, & 0, & 0 \\ 1, & 1, & 0, & 0 \\ 1, & 1, & 1, & 0 \\ 1, & 0, & 1, & 1 \end{pmatrix}.$$

The symbol $\mathbf{1}_{10}$ denotes the vector of 10 ones. Similar designs of experiments were used in the simulation study in Fišerová and Kubáček (2012).

The observation matrices $\underline{\mathbf{Y}}^1$ and $\underline{\mathbf{Y}}^2$ were generated in a natural way, a normally distributed error term was added to the true mean. The multiresponses were considered to be independent with the same covariance matrix $\boldsymbol{\Sigma}$ chosen in the following two forms: either $\mathbf{V}_1 = \mathrm{diag}\{1,1,1,0,0,0,0\}$ and $\mathbf{V}_2 = \mathrm{diag}\{0,0,0,1,1,1,1\}$ (the corresponding covariance matrix is denoted by $\boldsymbol{\Sigma}_1$), or $\mathbf{V}_1 = \mathrm{diag}\{1,1,0,0,0,0,1\}$ and $\mathbf{V}_2 = \mathrm{diag}\{0,0,1,1,1,1,0\}$ (covariance matrix $\boldsymbol{\Sigma}_2$). The variance components were chosen either $\vartheta_1 = 5$ and $\vartheta_2 = 3$, or $\vartheta_1 = 0.05$ and $\vartheta_2 = 0.03$.

Table 1: Empirical probabilities (in %) of rejecting the hypothesis "the true model is the system of two seemingly unrelated models (3)" on the significance level α. Data are simulated from model (3).

| parameter matrices | α | $\vartheta_1 = 5$ and $\vartheta_2 = 3$ | | | | $\vartheta_1 = 0.05$ and $\vartheta_2 = 0.03$ | | | |
| | | $n = 400$ | | $n = 40$ | | $n = 400$ | | $n = 40$ | |
		$\boldsymbol{\Sigma}_1$	$\boldsymbol{\Sigma}_2$	$\boldsymbol{\Sigma}_1$	$\boldsymbol{\Sigma}_2$	$\boldsymbol{\Sigma}_1$	$\boldsymbol{\Sigma}_2$	$\boldsymbol{\Sigma}_1$	$\boldsymbol{\Sigma}_2$
(14)	5	5.2	4.9	6.1	5.9	5.2	5.1	5.7	5.6
(14)	1	0.9	1.1	1.5	1.2	0.9	1.1	1.3	1.3
100*(14)	5	5.1	5.1	5.6	5.6	4.6	5.0	6.2	5.6
100*(14)	1	1.1	1.1	1.5	1.1	1.0	1.2	1.3	1.3

10 000 simulations were done for all cases. First, the data were simulated from the system of two seemingly unrelated submodels (3), i.e., for matrices $\mathbf{B}_1$, $\mathbf{B}_2$ given by (14) and for zero matrices $\mathbf{B}_{12}$, $\mathbf{B}_{21}$. The empirical probabilities of rejecting the hypothesis $\mathbf{B}_{12} = \mathbf{0}$ and $\mathbf{B}_{21} = \mathbf{0}$ simultaneously are presented in Table 1. We can see that the obtained empirical significance levels for plug-in test statistics T_{12} and T_{21} are essentially equal to the nominal
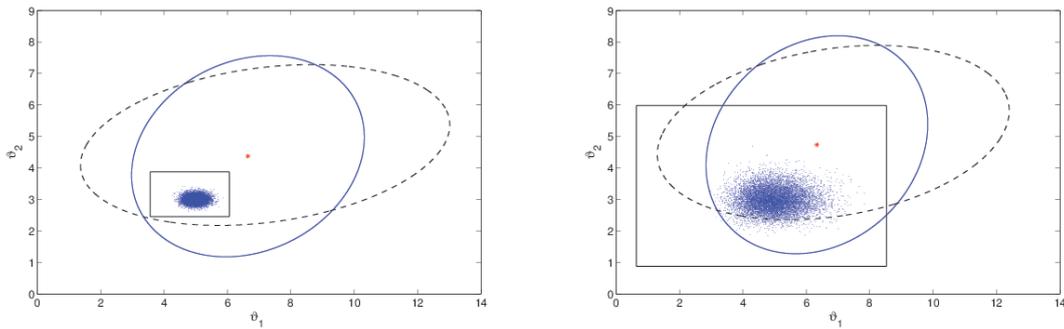
Figure 1: Insensitivity regions for the significance level $\alpha = 1\%$ together with 95%-confidence domain for the variance components for different sample sizes (left: $n = 400$; right: $n = 40$). $\mathcal{N}_{5,T_{21}}$ by solid line, $\mathcal{N}_{5,T_{12}}$ by dashed line. Data are simulated from model (3) for $\boldsymbol{\Sigma}_2$, $\vartheta_1 = 5$, $\vartheta_2 = 3$.

levels. Fišerová and Kubáček (2012) have shown in a simulation study that the joint test by T_{12} and T_{21} is conservative in case of a known covariance matrix, the obtained empirical significance level was equal to half of the nominal one. For data simulated from the model (1), i.e., for matrices $\mathbf{B}_{11}$, $\mathbf{B}_{22}$ and $\mathbf{B}_{12}$, $\mathbf{B}_{21}$ given by (14) and (15), respectively, the plug-in test statistics T_{21} and T_{12} rejected the decomposition of model (1) in all cases.

Finally we will investigate the insensitivity regions for the significance level. Let us assume a tolerable increase of the significance level $\alpha = 1\%$ or $\alpha = 5\%$ equal to $\varepsilon = 5\%$. It means, we are satisfied if the true type I error rate is $\alpha/2 + \varepsilon/2 = 5\%$ (3%) for a nominal significance level $\alpha = 5\%$ ($\alpha = 1\%$) for each of the plug-in test statistic T_{21} and T_{12}. Further, we assume that the parameter t equals 3, i.e., at least 89% of data values of δT_{21} and δT_{12} must be within 3 standard errors of the mean.

For $\alpha = 1\%$, the resulting insensitivity regions are displayed together with 95%-confidence domain for the variance components in Figure 1. The data were simulated from model (3) for the covariance matrix $\boldsymbol{\Sigma}_2$ and $\vartheta_1 = 5$, $\vartheta_2 = 3$. We can see that the insensitivity regions $\mathcal{N}_{5,T_{21}}$ and $\mathcal{N}_{5,T_{12}}$ are large enough. The statistic T_{12} allows greater shifts in direction of ϑ_1, and T_{21} in direction of ϑ_2. It means, the statistic T_{12} is more sensitive to changes in ϑ_2, and T_{21} in ϑ_1. Furthermore, we can notice, that the confidence domain for the variance components is embedded in both insensitivity regions for sample size $n = 400$ (left figure). For smaller sample size, the confidence domain increases more than the insensitivity ones (enlarged only slightly) and thus the confidence domain is not embedded into the insensitivity regions. Nevertheless, almost all variance components estimates lie within the insensitivity regions. Unfortunately, this is not generally true. The observed relative frequencies of the variance components estimates within the insensitivity regions are indicated in Table 2. The results are averages of a hundred times repeated 10 000 simulations. Obviously, the confidence region for the variance components is smaller for greater sample size and thus the relative frequencies are essentially 100%. However, for smaller sample size and significance level $\alpha = 5\%$, the relative frequencies are only 55-70%. Nevertheless, in this case the plug-in test is sufficiently good as it is shown in Table 1. Large differences between the relative frequencies are due to large differences in size of the insensitivity regions for significance levels $\alpha = 5\%$ and $\alpha = 1\%$. For example, in the case $\vartheta_1 = 5$ and $\vartheta_2 = 3$, the semiaxes of insensitivity region $\mathcal{N}_{5,T_{21}}$ are 1.75 and 1.28 for $\alpha = 5\%$, and 4.35 and 2.98 for $\alpha = 1\%$. This effect is related to the construction of the insensitivity regions, namely with the fact that a tolerable increase ε of the significance level α leads to larger $\delta_{\varepsilon,T_{21}}$, $\delta_{\varepsilon,T_{12}}$ for smaller α.

Note that the insensitivity regions for the significance level are shown as ellipses in Figure 1, although, in general, they should be open sets to the right hand corner. This relates to the fact that the p-values becomes less extreme with increasing variances. However, the construction

Table 2: Relative frequency (in %) of the variance components estimates within insensitivity regions for the significance level. Data are simulated from model (3) for $\mathbf{\Sigma}_2$.

| | | $\vartheta_1 = 5$ and $\vartheta_2 = 3$ | | | | | |
| | | $n = 400$ | | | $n = 40$ | | |
parameters	α	$\mathcal{N}_{5,T_{21}} \cap \mathcal{N}_{5,T_{12}}$	$\mathcal{N}_{5,T_{21}}$	$\mathcal{N}_{5,T_{12}}$	$\mathcal{N}_{5,T_{21}} \cap \mathcal{N}_{5,T_{12}}$	$\mathcal{N}_{5,T_{21}}$	$\mathcal{N}_{5,T_{12}}$
(14)	5	98.9	99.8	99.6	56.7	66.7	67.6
(14)	1	100	100	100	96.7	98.3	97.9
100*(14)	5	98.8	99.8	99.7	55.1	67.3	64.5
100*(14)	1	100	100	100	96.9	98.5	98.0

| | | $\vartheta_1 = 0.05$ and $\vartheta_2 = 0.03$ | | | | | |
| | | $n = 400$ | | | $n = 40$ | | |
parameters	α	$\mathcal{N}_{5,T_{21}} \cap \mathcal{N}_{5,T_{12}}$	$\mathcal{N}_{5,T_{21}}$	$\mathcal{N}_{5,T_{12}}$	$\mathcal{N}_{5,T_{21}} \cap \mathcal{N}_{5,T_{12}}$	$\mathcal{N}_{5,T_{21}}$	$\mathcal{N}_{5,T_{12}}$
(14)	5	98.68	99.35	99.22	54.79	65.61	65.54
(14)	1	100	100	100	97.04	98.47	98.17
100*(14)	5	98.91	99.84	99.78	58.35	68.85	68.30
100*(14)	1	100	100	100	95.94	98.22	97.21

of the insensitivity regions for estimators or the confidence level is different, and thus we proposed closed regions due to a uniform methodology and for easier handling.

The insensitivity regions for the significance level and the confidence domain for variance components result in intervals for the covariance matrix $\mathbf{\Sigma}_1$ since in this case the statistic T_{21} is a function of the parameter ϑ_1 only, and T_{12} is a function of ϑ_2. The obtained results are similar as in the case of the covariance matrix $\mathbf{\Sigma}_2$ and therefore they are omitted.

6. Conclusion

The proposed plug-in joint test seems to be a proper method for a decomposition of a mixed multivariate model (with independent responses with the same covariance matrix) into two seemingly unrelated submodels. The decomposition is advantageous at least from two viewpoints. The estimators of the regression parameters are more efficient, and data collection can be easier. The sensitivity approach is an appropriate technique for a justification that the estimated values of the variance components can be plugged in without any essential deterioration of the regression coefficients estimates and the inference. This is based on identifying safe regions in the space of the variance components where plug-in estimators cause only negligible changes of the optimum quality of estimators and test statistics. In particular, the proposed insensitivity region for the significance level guarantees that the true type I error rate does not exceed the chosen tolerable value.

The used methodology is general and suitable for more complex (mixed) models, e.g. models with restrictions on the regression parameters, singular models, and other statistical inference such as a confidence level or power of a test as well.

Acknowledgements

References

Anderson T (1958). *An Introduction to Multivariate Statistical Analysis.* J. Wiley, New York.

Bognárová M, Kubáček L, Volaufová J (1996). "Comparison of MINQUE and LMVQUIE by Simulation." *Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica*, **35**, 25–38.

Fišerová E, Kubáček L (2003). "Sensitivity Analysis in Singular Mixed Linear Models with Constraints." *Kybernetika*, **39**, 317–332.

Fišerová E, Kubáček L (2004). "Statistical Problems of Measurement in Triangle." *Folia Facultatis Scientiarium Naturalium Universitatis Masarykiana Brunesis, Mathematica*, **15**, 77–94.

Fišerová E, Kubáček L (2006). "Insensitivity Regions and Outliers in Mixed Models with Constraints." *Austrian Journal of Statistics*, **35**, 245–252.

Fišerová E, Kubáček L (2009). "Insensitivity Regions for Deformation Measurement on a Dam." *Environmetrics*, **20**, 776–789.

Fišerová E, Kubáček L (2012). "Decomposition of Multivariate Statistical Models." *Tatra Mountains Mathematical Publications*, **51**, 33–43.

Fišerová E, Kubáček L (2013). "Some remarks on decomposition of partitoned multivariate models into two seemingly unrelated submodels." In S Aivazian, P Filzmoser, Y Kharin (eds.), *Computer Data Analysis and Modeling: Theoretical and Applied Stochastics. Vol.1.*, pp. 36–41. Publishing center of BSU, Minsk.

Kenward M, Roger J (1996). "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood." *Biometrics*, **53**, 983–997.

Kshirsagar A (1972). *Multivariate Analysis.* Marcel Dekker, Inc., New York.

Kubáček L (1996). "Linear Model with Inaccurate Variance Components." *Applications of Mathematics*, **41**, 433–445.

Kubáček L (2006). "Statistical methods in environmetrics." In L Dušek, J Hřebíček, J Žižka (eds.), *Proceedings of the 2nd International Summer School on Computational Biology*, pp. 68–83. Masaryk University, Brno.

Kubáček L (2007a). "Multivariate Regression Model with Constraints." *Mathematica Slovaca*, **57**, 271–296.

Kubáček L (2007b). "Test of Linear Hypothesis in Multivariate Models." *Kybernetika*, **43**, 463–470.

Kubáček L (2008). *Multivariate Statistical Models Revisited.* Palacký University, Olomouc.

Kubáček L, Fišerová E (2003). "Problems of Sensitiveness and Linearization in a Determination of Isobestic Points." *Mathematica Slovaca*, **53**, 407–426.

Lešanská E (2002a). "Optimization of the Size of Nonsensitiveness Regions." *Applications of Mathematics*, **47**, 9–23.

Lešanská E (2002b). "Insensitivity Regions and their Properties." *Journal of Electrical Engineering*, **53**, 68–71.

Rao C, Kleffe J (1988). *Estimation of Variance Components and Applications.* North–Holland, Amsterdam, New York, Oxfords, Tokyo.

Rao C, Mitra S (1988). *Generalized Inverse of Matrices and Its Applications.* J. Wiley, New York.

Seber G (2004). *Multivariate Observations.* John Wiley & Sons, Inc., Hoboken, New Jersey.

Zellner A (1962). "An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias." *Journal of the American Statistical Association*, **57**, 348–368.

Affiliation:

Eva Fišerová
Mathematical Analysis and Applications of Mathematics
Faculty of Science
Palacký University
771 46 Olomouc, Czech Republic
E-mail: eva.fiserova@upol.cz
URL: http://mant.upol.cz/en/clen.asp?id=7

Lubomír Kubáček
Mathematical Analysis and Applications of Mathematics
Faculty of Science
Palacký University
771 46 Olomouc, Czech Republic
E-mail: lubomir.kubacek@upol.cz
URL: http://mant.upol.cz/en/clen.asp?id=3