ON EXACT INFERENCE IN LINEAR MODELS WITH TWO VARIANCE-COVARIANCE COMPONENTS

JÚLIA VOLAUFOVÁ — VIKTOR WITKOVSKÝ

ABSTRACT. Linear models with variance-covariance components are used in a wide variety of applications. In most situations it is possible to partition the response vector into a set of independent subvectors, such as in longitudinal models where the response is observed repeatedly on a set of sampling units (see, e.g., Laird & Ware 1982). Often the objective of inference is either a test of linear hypotheses about the mean or both, the mean and the variance components. Confidence intervals for parameters of interest can be constructed as an alternative to a test. These questions have kept many statisticians busy for several decades. Even under the assumption that the response can be modeled by a multivariate normal distribution, it is not clear what test to recommend except in a few settings such as balanced or orthogonal designs. Here we investigate statistical properties, such as accuracy of p-values and powers of exact (Crainiceanu & Ruppert 2004) tests and compare with properties of approximate asymptotic tests. Simultaneous exact confidence regions for variance components and mean parameters are constructed as well.

1. Introduction

The maximum-likelihood-based computational procedures for inference in mixed models have been around since the seminal paper by H a r t l e y and R a o in 1967 (see [3]). Since the mid 1990s procedures have been available in statistical computing packages that make analysis of mixed models appear to be a routine.

The following notation will be used: for a matrix A, A' denotes the transpose of A, A^{-1} denotes the inverse of A if it exists, A^- and A^+ denote the generalized

inverse and the Moore-Penrose of A, respectively (see, e.g., [9]), $R(A)$ denotes the linear subspace spanned by the columns of A, and P_A denotes the orthogonal projection matrix onto $R(A)$ (see [4]). If A is square, $\mathrm{tr}(A)$ denotes the trace of A. $r(A)$ denotes the rank (the number of linearly independent columns) of A.

A linear model with two variance-covariance components can be formulated as $Y \sim N_n\big(X\beta, \sigma^2 V(\lambda)\big)$, where $\sigma^2 > 0$ and $\lambda \in \Lambda$ are the variance-covariance components, $\beta \in \mathbb{R}^k$ is the unknown mean parameter and X is a known model matrix, columns of which are population indicators or additional model co-variates. Here we shall consider models with $V(\lambda) = I + \lambda V$, where V is a given nonnegative definite matrix with $r(V) \geq 1$. The parameter space Λ is defined by the requirement that $V(\lambda)$ be a positive definite matrix. In other words, if $\xi_1 \geq \xi_2 \geq \cdots \geq \xi_{r(V)} > 0$ are the nonzero eigenvalues of V, then $\Lambda = \{\lambda \in \mathbb{R} : \lambda > -1/\xi_1\}$. However, the most frequently considered parameter space Λ is $\{\lambda \in \mathbb{R} : \lambda \geq 0\}$.

In applications, the variance-covariance components σ^2 and λ are not known, and most of the time the focus is inference on fixed effects vector parameters. The variance components in such case are nuisance parameters. Even in simple cases, when the design is unbalanced, it is not quite straightforward what the best method for inference is in a final sample setting. An alternative approach is to formulate the questions of interest simultaneously in terms of the function of the mean vector parameter and the function of variance components. The inference is often testing statistical hypotheses and/or construct confidence intervals on functions of model parameters. A suggestive method is the likelihood ratio approach (LR). Here we adopt the development presented by C r a i n i c e a n u and R u p p e r t in 2004 (see [2]) and try to extend/investigate their approach.

Primarily, our interest is in testing a simultaneous linear hypothesis

$$H_0 : H'\beta = H'\beta_0 \text{ and } \lambda = \lambda_0 \quad \text{vs.} \quad H_A : H'\beta \neq H'\beta_0 \text{ or } \lambda \neq \lambda_0, \qquad (1)$$

where H is a given matrix, for which $R(H) \subset R(X')$, i.e., the hypothesis is estimable (see, e.g., [10]). The vector $H'\beta_0 = \theta_0$ and the constant λ_0 are known.

Alternatively, we may be interested in simultaneous confidence regions for some combinations of the parameters θ, λ, and σ^2, where $\theta = H'\beta$. The construction of the exact confidence regions is based on inverting the exact (restricted) likelihood ratio tests of following null hypotheses:

$$H_0 : \theta = \theta_0, \qquad (2)$$

$$H_0 : \theta = \theta_0 \text{ and } \lambda = \lambda_0, \qquad (3)$$

$$H_0 : \theta = \theta_0 \text{ and } \lambda = \lambda_0 \text{ and } \sigma^2 = \sigma_0^2, \qquad (4)$$

$$H_0 : \lambda = \lambda_0, \qquad (5)$$

$$H_0 : \lambda = \lambda_0 \text{ and } \sigma^2 = \sigma_0^2. \qquad (6)$$

For developing the exact likelihood ratio test for testing hypothesis (1) (i.e., (3)) we start with investigating the profile likelihood $L_\lambda(\beta, \sigma^2 \mid Y)$ for fixed $\lambda \in \Lambda$. Then the $-2 \log L_\lambda$ takes the form

$$
\begin{aligned}
-2 \log L_\lambda(\beta, \sigma^2 \mid Y) = {} & n \log(2\pi) + n \log(\sigma^2) + \log |V(\lambda)| \\
& + (1/\sigma^2)(Y - X\beta)'V(\lambda)^{-1}(Y - X\beta).
\end{aligned}
\tag{7}
$$

Hence the profile maximum likelihood estimators of σ^2 is

$$
\hat{\sigma}^2_\lambda = (1/n)Y'\big(M_X V(\lambda)M_X\big)^+ Y,
\tag{8}
$$

where $M_X = I - P_X$.

Since $H'\beta$ is estimable, $H = X'G$, $R(G) \subset R(X)$, for some matrix G. Then if we denote $X^* = (I - P_G)X$, the affine set of expectation vectors of Y is a shifted linear subspace parametrized by an unconstrained vector parameter γ, i.e.,

$$
\{X\beta : H'\beta = H'\beta_0\} = X\beta_0 + \{X^*\gamma : \gamma \in \mathbb{R}^k\},
\tag{9}
$$

(see, e.g., [1] or [6]).

The MLE of σ^2 under H_0 is then $\widetilde{\sigma^2}_{\lambda_0} = (1/n)Y^{*\prime}\big(M_{X^*}V(\lambda_0)M_{X^*}\big)^+ Y^*$, where $Y^* = Y - X\beta_0$. Since $Y'\big(M_X V(\lambda)M_X\big)^+ Y = Y^{*\prime}\big(M_X V(\lambda)M_X\big)^+ Y^*$, we get the test statistic in the form

$$
\begin{aligned}
-2 \log LRT = {} & n \log \left(1 + \frac{Y^{*\prime}\big((M_{X^*}V(\lambda_0)M_{X^*})^+ - (M_X V(\lambda_0)M_X)^+\big)Y^*}{Y^{*\prime}\big(M_X V(\lambda_0)M_X\big)^+ Y^*} \right) \\
& + \sup_{\lambda \in \Lambda} \left\{ n \log \left(\frac{Y'\big(M_X V(\lambda_0)M_X\big)^+ Y}{Y'\big(M_X V(\lambda)M_X\big)^+ Y} \right) + \log \left(\frac{|V(\lambda_0)|}{|V(\lambda)|} \right) \right\}.
\end{aligned}
\tag{10}
$$

Let B be a $n \times (n-p)$- matrix, $n - p = \text{tr}(M_X)$, for which $BB' = M_X$, $B'B = I$. Let the spectral decomposition of $B'VB$ result in $B'VB = \sum_{i=0}^{r} \mu_i Q_i$. Here $0 = \mu_0 < \mu_1 < \cdots < \mu_r$ denote the distinct eigenvalues of $B'VB$ with the multiplicity ν_i for μ_i. Q_i are mutually orthogonal projection matrices. Following the lines in O l s e n, S e e l y, and B i r k e s (see [7]), we note that $Y'BQ_iB'Y$ form a set of $r + 1$ mutually independent random variables, $Y'BQ_iB'Y \overset{H_0}{\sim} (1+\lambda_0\mu_i)\sigma^2\chi^2(\nu_i)$. By $\chi^2(\nu_i)$ we denote a chi-square distributed random variable with ν_i degrees of freedom. Hence

$$
Y'\big(M_X V(\lambda)M_X\big)^+ Y \overset{H_0}{\sim} \sigma^2 \sum_{i=0}^{r} (1 + \lambda_0\mu_i)/(1 + \lambda\mu_i)\chi^2(\nu_i).
\tag{11}
$$

Further, since $R(G) \subset R(X)$, it is straightforward to derive that

$$Y^{*\prime}\left(\left(M_{X^*}V(\lambda_0)M_{X^*}\right)^+ - \left(M_X V(\lambda_0)M_X\right)^+\right)Y^* \stackrel{H_0}{\sim} \sigma^2\chi^2(h), \qquad (12)$$

where $h = r(H)$, $\chi^2(h)$ independent with all $\chi^2(\nu_i)$. Hence, under null hypothesis (1), the exact distribution of the LRT statistic is expressed as

$$-2\log LRT \stackrel{H_0}{\sim} n\log\left(1 + \frac{\chi^2(h)}{\displaystyle\sum_{i=0}^{r}\chi_i^2(\nu_i)}\right)$$

$$+ \sup_{\lambda \in \Lambda}\left\{n\log\left(1 + \frac{\displaystyle\sum_{i=0}^{r}\frac{(\lambda-\lambda_0)\mu_i}{1+\lambda\mu_i}\chi_i^2(\nu_i)}{\displaystyle\sum_{i=0}^{r}\frac{1+\lambda_0\mu_i}{1+\lambda\mu_i}\chi_i^2(\nu_i)}\right) + \sum_{j=1}^{r(V)}\log\left(\frac{1+\lambda_0\xi_j}{1+\lambda\xi_j}\right)\right\}$$

$$\sim \sup_{\lambda \in \Lambda}\left\{n\log\left(\frac{\displaystyle\sum_{i=0}^{r}\chi_i^2(\nu_i) + \chi^2(h)}{\displaystyle\sum_{i=0}^{r}\frac{1+\lambda_0\mu_i}{1+\lambda\mu_i}\chi_i^2(\nu_i)}\right) + \sum_{j=1}^{r(V)}\log\left(\frac{1+\lambda_0\xi_j}{1+\lambda\xi_j}\right)\right\}. \quad (13)$$

This result coincides with the one obtained by C r a i n i c e a n u and R u p p e r t in [2]. In their paper, they provide a detailed algorithm obtaining the LRT and its distribution. On the other hand, one may consider an asymptotic approach to approximate the distribution if the sample size is large enough. For the hypothesized λ_0 in the interior of Λ, under H_0, $-2\log LRT \stackrel{D}{\to} \chi^2(h+1)$. For λ_0 on the border of the parameter space Λ, S t r a m and L e e in [11] suggest that under not very restrictive conditions, which are usually met in the model we investigate here, $-2\log LRT \stackrel{D}{\to} 0.5\chi^2(h) : 0.5\chi^2(h+1)$. P i n h e i r o and B a t e s in [8] suggest the asymptotic distribution of the LRT to be $-2\log LRT \stackrel{D}{\to} 0.65\chi^2(h) : 0.35\chi^2(h+1)$. In this paper we compare the finite sample behavior of the approximate and exact distribution of the LRT by simulation specifically for small sample sizes in a simple longitudinal model.

2. Simulations

We set up a simulation study to investigate how well the asymptotic distribution approximates the exact one. We compare here the empirical quantiles of LRT compared to asymptotic χ^2 distribution under H_0. It is obvious

that the distribution depends on the hypothesized value of λ but not of β. Hence a secondary question that we want to address is the sensitivity (or insensitivity) of the LRT test distribution under H_0 with respect to λ_0.
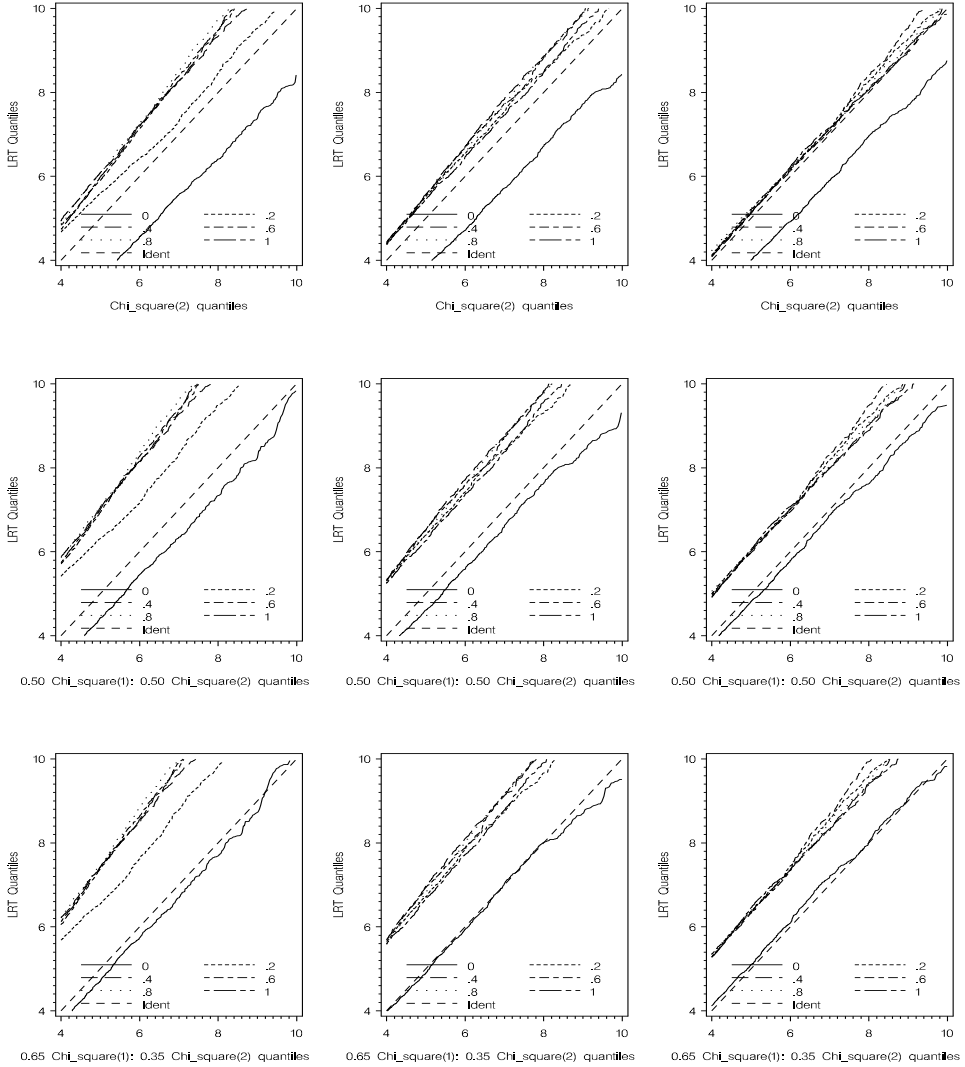

FIGURE 1. Quantile plots: left column $(5, 7)$, central column $(10, 15)$, right column $(30, 40)$, $r(H) = 1$.

JÚLIA VOLAUFOVÁ — VIKTOR WITKOVSKÝ

The simulation study is set up with independent samples from two populations. Three configurations with different number of sampling units (m_1, m_2) are investigated: $(5, 7), (10, 15),$ and $(30, 40)$. For each sampling unit unequal repeated observations are generated. First the number of observations on the ith sampling unit, T_i, is generated using Poisson distribution with parameter $\theta = 4$. Then "time" points are generated as discrete uniform with support being a subset of $\{1, 2, \ldots, 20\}$ of the size T_i that resulted from Poisson distribution. The response vector Y is modeled as $Y = (X_1, X_2)(\beta_1', \beta_2')'$, where Y is $n = \sum_{i=1}^{m} T_i$ dimensional, with $m = m_1 + m_2$, X_1 is a matrix containing covariates (functions of time) and the columns of X_2 are the dummy variables for the two populations. Three configurations for matrix X_1 were considered linear, linear+quadratic, and linear+quadratic+cubic polynomials of time. Hence the dimension of β_2 is 2 and the dimension of β_1 depends on the configuration: it is either a scalar, or two-, or three-dimensional vector. The response is modeled by a random intercept model, which results in a two-variance components model with the parameter space for λ, Λ, being the set of all nonnegative real numbers. The null hypothesis is formulated as $H_0 : \beta_1 = 0$ and $\lambda = \lambda_0$. The hypothesized values λ_0 are chosen to be the values 0, 0.2, 0.4, 0.6, 0.8, and 1.0. Notice that the hypothesized value $\lambda_0 = 0$ falls on the border of the parameter set Λ. Hence 54 different configurations were investigated. For each configuration 10,000 simulations were carried out, the empirical quantiles recorded and compared to asymptotic (approximate) distribution. The selected results are presented in the form of graphs. Figure 1 corresponds to one-column setting for matrix X_1, which means that the hypothesized parameter β_1 is a scalar. We notice that if the hypothesized λ_0 is in the interior of the parameter space, i.e., $\lambda_0 \neq 0$, with increasing sample size the $\chi^2(2)$ distribution provides a pretty good approximation to the exact distribution (the first row of panels). For all $\lambda_0 = 0$ the $\chi^2(2)$ approximation is not good even for higher sample sizes. However, the $0.65\chi^2(2) : 0.35\chi^2(3)$ mixture provides a very accurate approximation to probabilities (bottom row of panels).We observe the same phenomenon for higher rank of H, Figure 2 shows the results for $r(H) = 2$.

Based on our study, as a good approximation for the exact CDF at $\lambda_0 = 0$ we suggest to use a mixture distribution with estimated value of the optimum weight w based on small sample generated from the exact distribution $w\chi^2(h) : (1 - w)\chi^2(h + 1)$. For hypothesized values of λ_0 in the interior of the parameter space, the approximation does not seem to depend too strongly on λ_0 (all lines are almost on top of each other). That would suggest that the underlying exact distribution of the LRT is close to be invariant with respect to λ_0 (for sufficiently large values). In particular, for large values of λ_0 the exact distribution converges quickly to fixed distribution. For large λ_0 we suggest to use an approximation based on non-central chi-square distribution, $\chi^2(h + 1, \delta)$, with the noncentrality parameter δ, to be estimated based on small sample generated from the exact distribution, see Figure 3.
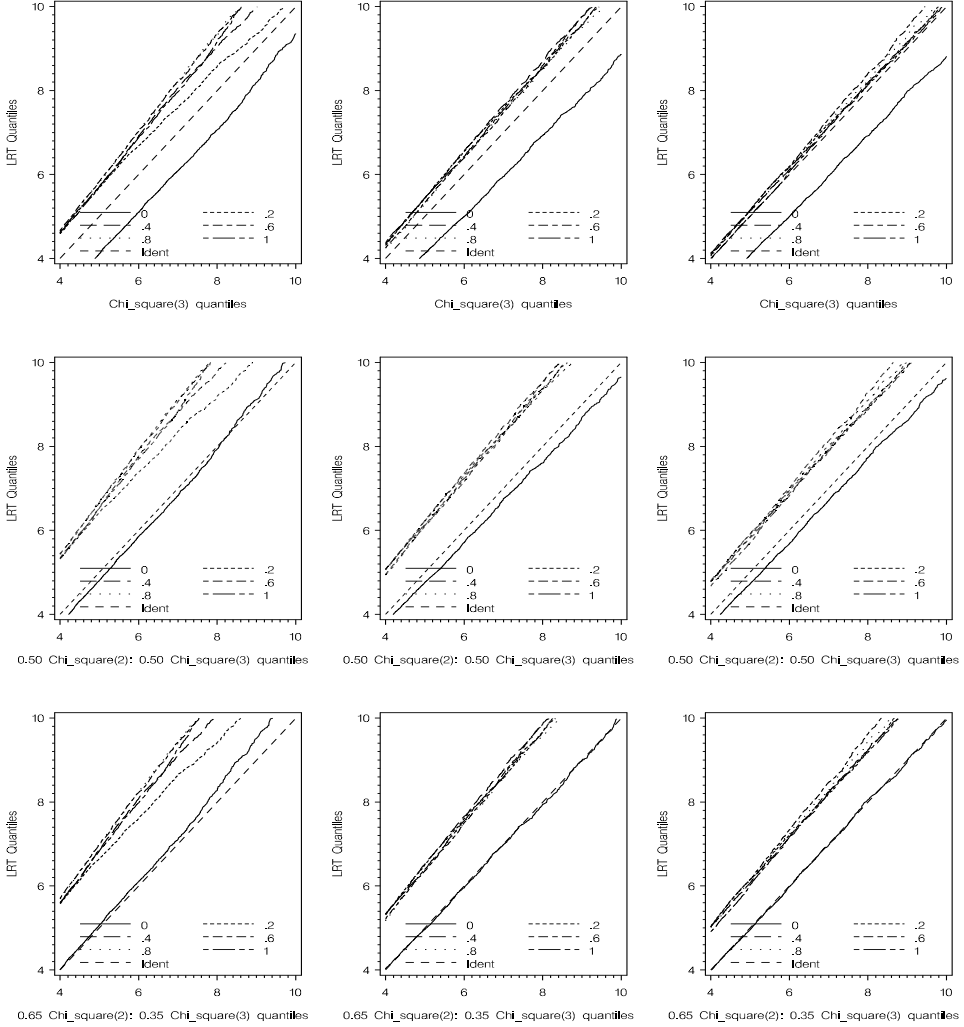
FIGURE 2. Quantile plots: left column $(5, 7)$, central column $(10, 15)$, right column $(30, 40)$, $r(H) = 2$.

3. Conclusion

The main goal in this paper was to investigate the approximate (asymptotic) characteristics of the distribution of the likelihood ratio test for small sample size setting in regard to exact distribution. We found that even for very small total number of observations, the asymptotic χ^2 approximation is adequate and complex exact calculations can be easily avoided.
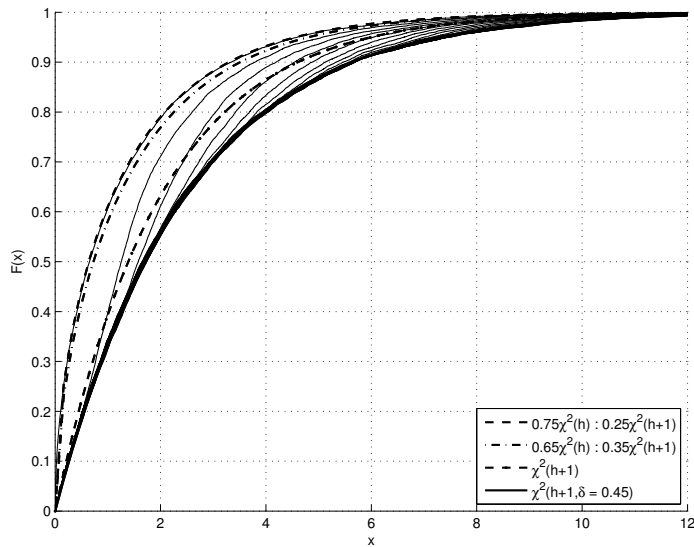
FIGURE 3. Exact and approximate CDFs of the LRT for model configuration with sample sizes $(5, 7)$ and $h = r(H) = 1$, plotted for different values of λ_0. The thin lines present exact CDFs for $\lambda_0 = 0 : 0.05 : 1$ (from left to right). The thick lines present approximations: (i) mixture distribution $w\chi^2(h) : (1-w)\chi^2(h+1)$, with the estimated weight $w = 0.75$, (ii) mixture distribution $0.65\chi^2(h) : 0.35\chi^2(h + 1)$, (iii) $\chi^2(h + 1)$, and (iv) noncentral chi-square distribution $\chi^2(h + 1, \delta)$ with the estimated non-centrality parameter $\delta = 0.45$.

REFERENCES

[1] CHRISTENSEN, R.: *Plane Answers to Complex Questions: The Theory of Linear Models.* Springer-Verlag, New York, 1987.

[2] CRAINICEANU, C. M.—RUPPERT, D.: *Likelihood ratio tests in linear mixed models with one variance component*, J. R. Stat. Soc. Ser. B Stat. Methodol. **66** (2004), 165–185.

[3] HARTLEY, H. O.—RAO, J. N. K.: *Maximum-likelihood estimation for the mixed alysis of variance model*, Biometrika **54** (1967), 93–108.

[4] HARVILLE, D. A.: *Matrix Algebra from a Statistician's Perspective.* Springer, New York, 1997.

[5] LAIRD, N. M.—WARE, J. H.: *Random-effects models for longitudinal data*, Biometrics, **38** (1982), 963–974.

[6] LAMOTTE, L. R.: *On non-estimable conditions and restricted least squares in linear regression models*, in: Proceedings of the Conference in Honor of Shayle R. Searle, Aug. 9–10, 1996, Biometrics Unit, Cornell University, 1998, pp. 145–160.

[7] OLSEN, A.—SEELY, J.—BIRKES, D.: *Invariant quadratic unbiased estimation for two variance components*, Ann. Statist. **4** (1976), 878–890.

 [8] PINHEIRO, J. C.—BATES, D. M.: *Mixed-effects Models in S and S-Plus*. Springer,
     New York, 2000.
 [9] RAO, C. R.—MITRA, S. K.: *Generalized Inverse of Matrices and Its Applications*.
     John Wiley & Sons, New York, 1971.
[10] SEELY, J.: *Estimability and linear hypotheses*, Amer. Statist. **31** (1977), 121–123.
[11] STRAM, D. O.— LEE, J. W.: *Variance components testing in the longitudinal mixed
     effects model*, Biometrics **50** (1994), 1171–1177.

Júlia Volaufová
School of Public Health
LSU Health Sciences Center
2020 Gravier Street
New Orleans, LA 70112
USA
E-mail: jvolau@lsuhsc.edu

Viktor Witkovský
Institute of Measurement Science
Slovak Academy of Sciences
Dúbravská cesta 9
SK–841-04 Bratislava
SLOVAKIA
E-mail: witkovsky@savba.sk