On testing variance components in ANOVA models*

Joachim Hartung and Guido Knapp

Department of Statistics, University of Dortmund, D–44221 Dortmund, Germany

Abstract: In this paper we derive asymptotic χ^2–tests for general linear hypotheses on variance components using repeated variance components models. In two examples, the two–way nested classification model and the two–way crossed classification model with interaction, we explicitly investigate the properties of the asymptotic tests in small sample sizes.

Key Words: Wald– and likelihood ratio test statistic, repeated variance components model, linear hypotheses on variance components

1 Introduction

In this paper we consider linear hypotheses on variance components as

$$H_0: \ K\sigma = d \quad , \tag{1}$$

where $\sigma = (\sigma_1^2, \ldots, \sigma_m^2)^T$ denotes a vector of unknown variance components, K is a known $(p \times m)$–matrix with rk $(K) = p \leq m$ and $d \in I\!R^p$ a known constant. For special linear combinations of variance components exact F– and χ^2–tests can be derived and in El–Bassiouni and Seely (1980) it is shown that under certain circumstances these tests are uniformly most powerful unbiased. However, no exact tests for example are known for testing that the variance of a certain factor is equal to a given $d_1 > 0$ or that the difference between two variance components equals a certain value. Here, we develop asymptotic χ^2–tests for such hypotheses.

* *To appear in:* $STATISTICS$

In section 2 we consider the class of variance components models of commutative quadratic type (see e. g. Seely (1971), Humak (1984), Rao and Kleffe (1988), Elpelt (1989), Hartung (1981, sec. 5), Hartung, Elpelt, and Voet (1997)) and introduce repeated variance components models (cf. Brown (1976)). Then, Wald and likelihood ratio test statistics are considered in section 3 using the approach of a repeated variance components model, where the asymptotic results refer to a 'large' number of observations resp. degrees of freedom in the experimental designs which can be interpreted as several independent observations from a reduced design. In practice, however, we deal with non–repeated models. Thus, in section 4 we consider, as examples, the two–way nested classification model and the two–way crossed classification model with interaction, where we explicitly study hypotheses about the differences of two variance components. In both models we can directly give the Wald test statistics for the hypotheses as well as the estimators of maximum likelihood type. In the two–way nested classification model, however, we use a numerical algorithm to maximize the likelihood function under the hypothesis, whereas in the two–way crossed classification model with interaction we give in addition an explicit approximation of the likelihood ratio test statistic which does not need a numerical algorithm. In a simulation study we examine the finite properties of the derived tests, especially in situations where the sample sizes are really 'small' and show that the asymptotic works satisfactorily in these cases. Hereby, a clear preference of the likelihood ratio test can be stated, on the whole.

Throughout this paper we use the following notation. For a real matrix A let A^T denote the transposed, A^+ the Moore–Penrose–inverse, rk (A) the rank, tr (A) the trace, and $\mathcal{R}(A)$ the range of A. Further we denote by $\otimes$ the Kronecker product, by I_r the $(r \times r)$–identity matrix, by 1_r the vector of r ones, by $J_r = 1_r 1_r^T$ the $(r \times r)$–matrix of ones, and by $0_{s \times t}$ the $(s \times t)$–matrix of zeros.

2 The repeated variance components model

We consider a q–dimensional observable random vector, say Y, that follows the general linear variance components model

$$\mathrm{E}\,(Y) = X\beta \qquad \text{and} \qquad \mathrm{Cov}\,(Y) = \sum_{i=1}^{m} \sigma_i^2\, U_i\ , \tag{2}$$

where the $(q \times l)$–matrix X and the symmetric positive semidefinite $(q \times q)$–matrices $U_1, \dots, U_m$ are known, whereas the parameter vector β varies in $I\!R^l$ and the parameter vector $\sigma = (\sigma_1^2, \dots, \sigma_m^2)^T$ varies in Ω, a subset of $I\!R^m_{(+)}$, the nonnegative orthant of $I\!R^m$. The variance component σ_m^2 is assumed to be strictly positive and U_m is positive definite to ensure the positive definiteness of the variance–covariance matrix.

In this model we make the following assumptions.

Assumption 1: The random vector Y has a q–dimensional normal distribution.
Assumption 2: $\Psi = span\{XX^+, U_1, \dots, U_m\}$ forms a $(m+1)$–dimensional commutative quadratic subspace of all real symmetric $(q \times q)$–matrices, i. e. Ψ is a subspace and $A, B \in \Psi$ implies $A^2 \in \Psi$ and $AB = BA$.

By lemma 6 in Seely (1971) there exists a basis $P_0, P_1, \dots, P_m$ of Ψ with $P_0 = XX^+$, where P_i, $i = 0, 1, \dots, m$, is idempotent and $P_iP_j = 0_{q\times q}$, $i \neq j$. Then there is a nonsingular $((m+1) \times (m+1))$–matrix

$$\Lambda^* = ((\lambda_{ij})_{i,j=0,1,\dots,m}) \tag{3}$$

so that

$$U_i = \sum_{j=0}^{m} \lambda_{ij} P_j\ ,\ i = 0, 1, \dots, m, \quad \text{with} \quad U_0 = P_0 = XX^+\ . \tag{4}$$

Assumption 3: For all $j = 1, \dots, m$ and $\sigma \in \Omega$ it holds

$$\tau_j = \sum_{i=1}^{m} \lambda_{ij}\, \sigma_i^2 > 0\ . \tag{5}$$

The vector $\tau = (\tau_1, \ldots, \tau_m)^T$ can be expressed as

$$\tau = \Lambda^T \sigma \ , \tag{6}$$

where the nonsingular $(m \times m)$–matrix Λ is the submatrix of Λ^*, which results from deleting the first row and the first column of Λ^*.

The three assumptions are in most cases fulfilled in balanced variance components models if, as it is usually assumed, the residual variance is strictly positive; note that for example in Humak (1984, p. 284) a balanced variance components model is given which does not fulfill these assumptions.

Let us consider the quadratic forms

$$T_j = Y^T P_j Y / f_j \ , \ j = 1, \ldots, m, \tag{7}$$

where tr $(P_j) = f_j$, $j = 1, \ldots, m$. In model (2) it holds that these quadratic forms are stochastically independent and that $f_j \cdot T_j / \tau_j$, $j = 1, \ldots, m$, is central χ^2–distributed with f_j degrees of freedom. It follows that the expectation vector of the random vector $T = (T_1, \ldots, T_m)^T$ is given by the vector τ and the variance–covariance matrix of T is the diagonal $(m \times m)$–matrix

$$D(\sigma) = 2 \cdot \mathrm{diag}\,(\tau_1^2, \ldots, \tau_m^2) \ . \tag{8}$$

Now the model (2) is ν–times statistically independently repeated, i. e. we observe independent random vectors Y_ρ, $\rho = 1, \ldots, \nu$, which have the same distributional properties as Y from (2). Thus, we get the following model

$$\begin{gathered} \tilde{Y} = (Y_1^T, Y_2^T, \ldots, Y_\nu^T)^T \\ \text{with} \quad \mathrm{E}\,(\tilde{Y}) = (1_\nu \otimes X)\beta \ , \ \mathrm{Cov}\,(\tilde{Y}) = \sum_{i=1}^m \sigma_i^2 \,(I_\nu \otimes U_i) \ . \end{gathered} \tag{9}$$

Due to (4) the variance–covariance matrix of $\tilde{Y}$ can be expressed as

$$\mathrm{Cov}\,(\tilde{Y}) = \sum_{i=1}^m \sigma_i^2 \,(I_\nu \otimes U_i) = \sum_{j=0}^m \tau_j \,(I_\nu \otimes P_j) \ , \tag{10}$$

where τ_0 is a linear combination of $\tau_1, \ldots, \tau_m$.

Note that in the repeated model (9) with $(1_\nu \otimes X)(1_v \otimes X)^+ = (1/\nu J_\nu \otimes XX^+)$ the set $\Psi_\nu = span\{1/\nu J_\nu \otimes XX^+, I_\nu \otimes U_1, \dots, I_\nu \otimes U_m\}$ does not form a $(m+1)$–dimensional commutative quadratic subspace of all real symmetric $(q \times q)$–matrices as in the corresponding non–repeated model (2), which can be seen from the number of minimal sufficient statistics in the following lemma.

Lemma 1:

In model (9) it holds

i) the $m+1$ matrices $(I_\nu \otimes P_i)$, $i = 0, 1, \dots, m$, are idempotent and mutually orthogonal matrices;

ii) the $m+2$ statistics

$$(1_\nu \otimes P_0)\tilde{Y} \ , \ \tilde{Y}^T((I_\nu - \frac{1}{\nu}J_\nu) \otimes P_0)\tilde{Y}, \ , \ \tilde{Y}^T(I_\nu \otimes P_j)\tilde{Y} \ , \ j = 1, \dots, m \ , \tag{11}$$

are minimal sufficient statistics for this model.

Proof:

i) It is $(I_\nu \otimes P_i)^2 = I_\nu \otimes P_i^2 = I_\nu \otimes P_i$, $i = 0, 1, \dots, m$,
and $(I_\nu \otimes P_i)(I_\nu \otimes P_j) = I_\nu \otimes P_i P_j = I_n \otimes 0_{q \times q}$, $i \neq j$,
because the $m+1$ matrices P_i, $i = 0, 1, \dots, m$, are idempotent and mutually orthogonal.

ii) The result is given following the lines of the proof of Theorem 1 in Seifert (1979).

Lemma 2:

An quadratic unbiased estimator of τ_j, $j = 1, \dots, m$, in model (9) is given by

$$\hat{\tau}_j = \frac{1}{\text{tr } (I_\nu \otimes P_j)} \tilde{Y}^T(I_\nu \otimes P_j)\tilde{Y} \ , \ j = 1, \dots, m.$$

Proof:

The expected value of $\tilde{Y}^T(I_\nu \otimes P_j)\tilde{Y}$, $j = 1, \dots, m$, is given by
$\text{E}\,(\tilde{Y}^T(I_\nu \otimes P_j)\tilde{Y}) = \text{tr}\,(I_\nu \otimes P_j)\text{Cov}\,(\tilde{Y}) = \tau_j \cdot \text{tr}\,(I_\nu \otimes P_j)$,
because $P_j X = 0_{q \times l}$, $j = 1, \dots, m$.

Due to (6) we have a unique relation between τ and σ. Hence, we use the quadratic forms $\tilde{Y}^T(I_\nu \otimes P_j)\tilde{Y}$, $j = 1, \dots, m$, to make inference about the unknown vector of variance components.

3 Derivation of the test statistics

In model (9) we consider the quadratic forms $\tilde{Y}^T(I_\nu \otimes P_j)\tilde{Y}$, $j = 1, \ldots, m$, and define for all $j = 1, \ldots, m$

$$T_j^\nu = \tilde{Y}^T(I_\nu \otimes P_j)\tilde{Y}/(\nu \cdot f_j) = \frac{1}{\nu}\sum_{\rho=1}^{\nu} Y_\rho^T P_j Y_\rho / f_j = \frac{1}{\nu}\sum_{\rho=1}^{\nu} T_j^\rho \ , \tag{12}$$

where $T_j^\rho = Y_\rho^T P_j Y_\rho / f_j$, $\rho = 1, \ldots, \nu$, $j = 1, \ldots, m$.

Let us denote $T^\nu = (T_1^\nu, \ldots, T_m^\nu)^T$ then it holds

$$\mathrm{E}\,(T^\nu) = \Lambda^T \sigma = \tau \quad , \tag{13}$$

and the variance–covariance matrix of T^ν is a diagonal $(m \times m)$–matrix given by

$$D^\nu(\sigma) = 2 \cdot \mathrm{diag}\,(\tau_1^2/(\nu \cdot f_1), \ldots, \tau_m^2/(\nu \cdot f_m)) = D(\sigma)/v \ , \tag{14}$$

and $D(\sigma)$ is the variance–covariance matrix of T from (8) in the corresponding non–repeated model (2).

For each $\rho = 1, \ldots, \nu$ the random variables $f_j \cdot T_j^\rho/\tau_j$, $j = 1, \ldots, m$, are independent χ^2–distributed random variables with f_j degrees of freedom. Thus, we consider the likelihood function

$$L(\sigma) = \prod_{\rho=1}^{\nu}\prod_{i=1}^{m}(C_i)^{-1}\left(\frac{f_i}{\tau_i}\right)^{f_i/2}(T_i^\rho)^{(f_i-2)/2}\ \exp\left\{-\frac{1}{2}\,\frac{f_i \cdot T_i^\rho}{\tau_i}\right\} \ , \tag{15}$$

where $(C_i)^{-1} = 2^{f_i/2}\,\Gamma(f_i/2)$, $i = 1, \ldots m$, and $\Gamma(x)$ denotes the gamma function.
So, the log–likelihood function reads

$$l(\sigma) = \sum_{\rho=1}^{\nu}\sum_{i=1}^{m}\left\{\ln(C_i)^{-1} + \frac{f_i}{2}\ln\left(\frac{f_i}{\tau_i}\right) + \left(\frac{f_i-2}{2}\right)\ln T_i^\rho - \frac{1}{2}\,\frac{f_i \cdot T_i^\rho}{\tau_i}\right\} \ . \tag{16}$$

For the first derivatives of the log–likelihood function (16) we get

$$\begin{aligned}\frac{\partial l(\sigma)}{\partial \sigma_j^2} &= \sum_{\rho=1}^{\nu}\sum_{i=1}^{m}\left(\frac{f_i}{2\tau_i^2}\cdot\lambda_{ij}\cdot T_i^\rho - \frac{f_i}{2\tau_i}\cdot\lambda_{ij}\right) \\ &= \sum_{i=1}^{m}\left(\frac{\nu \cdot f_i}{2\tau_i^2}\cdot\lambda_{ij}\cdot T_i^\nu - \frac{\nu \cdot f_i}{2\tau_i}\cdot\lambda_{ij}\right) \ , \ j = 1, \ldots, m,\end{aligned} \tag{17}$$

so that

$$\frac{\partial l(\sigma)}{\partial \sigma} = \Lambda(D^{\nu}(\sigma))^{-1}(T^{\nu} - \Lambda^T \sigma) \quad . \tag{18}$$

Due to (18) the maximum likelihood estimator of σ has the form

$$\hat{\sigma} = (\Lambda^T)^{-1} \cdot T^{\nu} \quad , \tag{19}$$

and thus, the maximum likelihood estimator of τ is given by

$$\hat{\tau} = T^{\nu} \quad . \tag{20}$$

The maximum likelihood estimator in (19) coincides with the usual ANOVA–estimator and asymptotically yields nonnegative estimates of the variance components (cf. Brown (1976)).

For the second derivatives of the log–likelihood function (16) we obtain

$$\begin{aligned} \frac{\partial^2 l(\sigma)}{\partial \sigma_j^2 \partial \sigma_k^2} &= \sum_{\rho=1}^{\nu} \sum_{i=1}^{m} \Big(-\frac{f_i}{\tau_i^3} \lambda_{ij}\, \lambda_{ik}\, T_i^{\rho} + \frac{f_i}{2\tau_i^2} \lambda_{ij}\, \lambda_{ik} \Big) \\ &= \sum_{i=1}^{m} \Big(-\frac{\nu \cdot f_i}{\tau_i^3} \lambda_{ij}\, \lambda_{ik}\, T_i^{\nu} + \frac{\nu \cdot f_i}{2\tau_i^2} \lambda_{ij}\, \lambda_{ik} \Big) \ , \ j,k = 1, \ldots, m, \end{aligned} \tag{21}$$

so that the mean values of these derivatives are

$$\mathrm{E}\left(\frac{\partial^2 l(\sigma)}{\partial \sigma_j^2 \partial \sigma_k^2} \right) = -\sum_{i=1}^{m} \frac{\nu \cdot f_i}{2\tau_i^2} \lambda_{ij}\, \lambda_{ik} \quad , \ j,k = 1, \ldots, m \, . \tag{22}$$

Thus, the information matrix is given by

$$\begin{aligned} I^{\nu}(\sigma) = \mathrm{E}\left(-\frac{\partial^2 l(\sigma)}{\partial \sigma \partial \sigma^T} \right) &= \Lambda(D^{\nu}(\sigma))^{-1} \Lambda^T \\ &= \nu \cdot \Lambda(D(\sigma))^{-1} \Lambda^T = \nu \cdot I(\sigma) \quad , \end{aligned} \tag{23}$$

where $I(\sigma)$ is the information matrix in the corresponding non–repeated model (2).

Due to the results of Anderson (1973) and Brown (1976), respectively, cf. also Schmidt and Thrum (1981), we can state the following theorem.

Theorem 1:

In model (9) it holds that $\sqrt{\nu}\,(\hat{\sigma}-\sigma)$ is asymptotically normally distributed with mean vector 0 and variance–covariance matrix $(\Lambda(D(\sigma))^{-1}\Lambda^T)^{-1}$ for $\nu \to \infty$, and under the hypothesis $H_0 : K\sigma = d$, $\sqrt{\nu}\,(K\hat{\sigma}-d)$ is asymptotically normally distributed with mean vector 0 and variance–covariance matrix $K(\Lambda(D(\sigma))^{-1}\Lambda^T)^{-1}K^T$ for $\nu \to \infty$.

Thus, the Wald–type test statistic for testing the general linear hypothesis (1) is given by

$$W = (K\hat{\sigma}-d)^T(K(\Lambda(D^{\nu}(\sigma))^{-1}\Lambda^T)^{-1}K^T)^{-1}(K\hat{\sigma}-d)\ , \tag{24}$$

which is under H_0 asymptotically χ^2–distributed with rk (K) degrees of freedom (cf. Rao (1973), p. 188)). For an application of the Wald test a consistent estimator of σ, usually the maximum likelihood estimator $\hat{\sigma}$, has to be replaced in $D^{\nu}(\sigma)$.

Note that a Wald test statistic using iterated MINQUE is given by Schmidt and Thrum (1981), Kleffe and Seifert (1988), cf. also Khuri, Mathew and Sinha (1998, p. 164).

An asymptotically equivalent test to the Wald test is given by the likelihood ratio test. Thus, we consider the ratio

$$\max_{\sigma\,:\;K\sigma=d} L(\sigma)\Big/\max_{\sigma} L(\sigma). \tag{25}$$

Considering the Lagrangian function

$$\mathcal{L}(\sigma,\lambda) = l(\sigma) - \lambda^T(K\sigma-d), \tag{26}$$

the maximum likelihood estimator of σ under H_0, say $\bar{\sigma} = (\bar{\sigma}_1^2,\ldots,\bar{\sigma}_m^2)^T$, is a solution of

$$\begin{aligned}(T^{\nu}-\tau) - D^{\nu}(\sigma)\Lambda^{-1}K^T\lambda &= 0\\ K\sigma &= d\,,\end{aligned} \tag{27}$$

where $\lambda \in I\!R^p$ is a vector of Lagrange multipliers.

Theorem 2:

The test statistic

$$LR = 2\big(l(\hat{\sigma}) - l(\bar{\sigma})\big) \tag{28}$$

is under $H_0 : K\sigma = d$ asymptotically χ^2–distributed with rk (K) degrees of freedom.

Proof: We note that the likelihood function (15) is built of independent identically distributed random vectors $T^\rho = (T_1^\rho, \ldots, T_m^\rho)^T$, $\rho = 1, \ldots \nu$, so that the proof is given using standard arguments of maximum likelihood theory (see e. g. Rao (1973), p. 418–419).

Using the representation of the log–likelihood function from (16) the likelihood ratio test statistic (28) can also be expressed as

$$LR = \sum_{i=1}^{m} \nu \cdot f_i \left\{ \frac{T_i^\nu}{\sum_{j=1}^m \lambda_{ij}\bar{\sigma}_i^2} - \ln\left(\frac{T_i^\nu}{\sum_{j=1}^m \lambda_{ij}\bar{\sigma}_i^2} \right) - 1 \right\} \tag{29}$$

with $T_i^\nu = \sum_{i=1}^m \lambda_{ij}\hat{\sigma}_i^2$, $i = 1 \ldots, m$.

4 Two Examples

4.1 Two–way nested classification model

We consider the balanced two–way nested classification model with random effects given by

$$\begin{gathered} y_{ijk} = \mu + a_i + b_{ij} + e_{ijk} \\ i = 1, \ldots, r;\ j = 1, \ldots, s;\ k = 1, \ldots, t;\ n = rst, \end{gathered} \tag{30}$$

where $\mu \in I\!R$ is a fixed effect and $a_1, \ldots, a_r$, $b_{11}, \ldots, b_{rs}$, $e_{111}, \ldots, e_{rst}$ are independent normally distributed random effects with $\mathrm{E}\,(a_i) = \mathrm{E}\,(b_{ij}) = \mathrm{E}\,(e_{ijk}) = 0$ and $\mathrm{Var}\,(a_i) = \sigma_a^2$, $\mathrm{Var}\,(b_{ij}) = \sigma_b^2$, $\mathrm{Var}\,(e_{ijk}) = \sigma_e^2 > 0$ for all i, j, and k, so $\sigma = (\sigma_a^2, \sigma_b^2, \sigma_e^2)^T$.

The unique basis of projection matrices in this model is given by

$$\begin{array}{lll} P_0 = \frac{1}{n} J_n & , & \\ P_a = (I_r - \frac{1}{r} J_r) \otimes \frac{1}{st} J_{st} & , & \mathrm{tr}\,(P_a) = r-1 \; , \\ P_b = I_r \otimes (I_s - \frac{1}{s} J_s) \otimes \frac{1}{t} J_t & , & \mathrm{tr}\,(P_b) = r(s-1) \; , \\ P_e = I_{rs} \otimes (I_t - \frac{1}{t} J_t) & , & \mathrm{tr}\,(P_e) = rs(t-1) \; , \end{array} \tag{31}$$

and the matrix Λ has the form

$$\Lambda = \begin{pmatrix} st & 0 & 0 \\ t & t & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad . \tag{32}$$

With $y = (y_{111}, y_{112}, \dots, y_{rst})^T$ let us denote the mean sum of squares of the random effects as

$$\begin{aligned} M_1 &= y^T P_a y/(r-1) \; , \\ M_2 &= y^T P_b y/(r(s-1)) \; , \\ M_3 &= y^T P_e y/(rs(t-1)) \; . \end{aligned} \tag{33}$$

For an application of the Wald test statistic we have to replace $D(\sigma)$ in (24) by a consistent estimator. Using a result from Hartung and Voet (1986) the best invariant unbiased estimator for $D(\sigma)$ is given by

$$\widehat{D(\sigma)} = 2 \cdot \mathrm{diag}\left(\frac{M_1^2}{r+1}, \frac{M_2^2}{r(s-1)+2}, \frac{M_3^2}{rs(t-1)+2} \right) . \tag{34}$$

For testing the hypothesis $H_0 : \sigma_a^2 = \sigma_b^2$ with $K = (1, -1, 0)^T$ and $d = 0$ the Wald test statistic has the form

$$W_1 = \frac{\left(\frac{1}{st}(M_1 - M_2) - \frac{1}{t}(M_2 - M_3) \right)^2}{\frac{1}{s^2 t^2} \left(\frac{2\, M_1^2}{r+1} + \frac{2\,(s+1)^2\, M_2^2}{r(s-1)+2} + \frac{2\, s^2\, M_3^2}{rs(t-1)+2} \right)}, \tag{35}$$

which can also be expressed in terms of the maximum likelihood estimators $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$ as

$$W_1 = \frac{(\hat{\sigma}_a^2 - \hat{\sigma}_b^2)^2}{\widehat{\mathrm{Var}}\,(\hat{\sigma}_a^2 - \hat{\sigma}_b^2)}. \tag{36}$$

We reject H_0 at level α if $W_1 > \chi^2_{1;1-\alpha}$, where $\chi^2_{1;\gamma}$ denotes the γ–quantile of the χ^2–distribution with one degree of freedom.

In order to apply the likelihood ratio test statistic for testing $H_0 : \sigma_a^2 = \sigma_b^2$ we have to make use of a numerical algorithm to maximize the log-likelihood under H_0. In the following simulation studies, which have been carried out in SAS 6.12 using PROC IML, we use the Newton–Raphson ridge optimization algorithm to obtain the maximum likelihood estimator under H_0.

In the first simulation study we investigate the behaviour of the significance level and the power of both tests where we focus our attention on 'small' degrees of freedom of the mean sum of squares. Due to the fact that a given two–way nested classification model can possibly be interpreted as a replication of a reduced design, which depends on the number r of levels of the A–factor, we only consider sample sizes with increasing r and three pairs of sample sizes (s, t) to make the simulations not too complex. For the error variance σ_e^2 we always choose the value one. The variance component σ_a^2 is generated as a random number from a uniform distribution over the interval $(0, 10)$ in each run, and the variance component σ_b^2 is set equal to the generated value of σ_a^2. In table 1 the results for the Wald and the likelihood ratio test concerning the estimated size of the tests given the nominal level of $\alpha = 0.01$ and $\alpha = 0.05$, respectively, are presented based on 10,000 replications of the model.

We observe that the estimated significance levels of the likelihood ratio test are nearly independent from the chosen sample sizes and exceed the nominal significance levels, but in a compatible manner; for small r the largest estimated sizes are observed, e. g. near 7 % for $\alpha = 0.05$, and with increasing r the estimated sizes go towards the corresponding nominal ones. The estimated significance levels of the Wald test, however, do not show such a homogeneous behaviour as the estimated sizes of the likelihood ratio test. For all r with $s = t = 3$ the estimated sizes of the Wald test mostly fall below the nominal significance level, but for $r \geq 10$ the simulation indicates that the actual size of the test attains the nominal size. For the other two considered sample sizes of s and t the estimated significance levels are considerably larger than the nominal significance levels, and the larger s and t the larger the estimated sizes. But for increasing r the estimated significance levels of the Wald test becomes smaller and the case $r = 20$ indicates that for large r the actual size of the test may go towards the nominal significance level.

Consequently, the likelihood ratio test seems to be preferable to the Wald test in small sample sizes.

Yet, we generate another data to compare the power of both tests and we restrict to all sample sizes r with $s = 5$ and $t = 6$. Due to the fact that the Wald test is rather liberal in these situations, in the power comparison we used as critical values of the likelihood ratio and the Wald test the simulated empirical 95 %–quantiles of the distributions of the corresponding test statistics under H_0. As possible alternative hypotheses we consider the cases $\delta = \sigma_b^2 - \sigma_a^2 = 0.5, 1, 2, 5, 10, 20, 50,$ and 100. The results of this simulation study are given in table 2, where again every estimated point of the power function is based on 10,000 replications. We observe that in all considered cases the estimated power function of the Wald test lies above the estimated power function of the likelihood ratio test, so one would recommend the Wald test if the size of the likelihood ratio and the Wald test are nearly the same.

4.2 Two–way crossed classification model with interaction

Let us consider the balanced two-way crossed classification random model with interaction given by

$$\begin{gathered} y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk} \,, \\ i = 1, \ldots, r;\; j = 1, \ldots, s;\; k = 1, \ldots, t;\; n = rst, \end{gathered} \tag{37}$$

where $\mu \in I\!R$ is a fixed effect and $a_1, \ldots, a_r,\ b_1, \ldots, b_s,\ (ab)_{11}, \ldots, (ab)_{rs},\ e_{111}, \ldots, e_{rst}$ are independent normally distributed random effects with $\mathrm{E}\,(a_i) = \mathrm{E}\,(b_j) = \mathrm{E}\,((ab)_{ij}) = \mathrm{E}\,(e_{ijk}) = 0$ and $\mathrm{Var}\,(a_i) = \sigma_a^2$, $\mathrm{Var}\,(b_j) = \sigma_b^2$, $\mathrm{Var}\,((ab)_{ij}) = \sigma_{ab}^2$, $\mathrm{Var}\,(e_{ijk}) = \sigma_e^2 > 0$ for all i, j and k, so $\sigma = (\sigma_a^2, \sigma_b^2, \sigma_{ab}^2, \sigma_e^2)^T$.
Let M_1, M_2, M_3 and M_4 represent the A–factor, B–factor, AB–interaction and residual error mean squares, then it holds

$$\mathrm{E}(M) = \Lambda^T \sigma, \quad M = (M_1, M_2, M_3, M_4)^T, \tag{38}$$

and

$$\Lambda = \begin{pmatrix} st & 0 & 0 & 0 \\ 0 & rt & 0 & 0 \\ t & t & t & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} , \tag{39}$$

where the unique basis of projection matrices is given by

$$\begin{aligned}
&P_0 = \tfrac{1}{n} J_n &&, \\
&P_a = \left(I_r - \tfrac{1}{r} J_r\right) \otimes \tfrac{1}{st} J_{st} &&, \quad \mathrm{tr}\,(P_a) = r-1\,, \\
&P_b = \tfrac{1}{r} J_r \otimes \left(I_s - \tfrac{1}{s} J_s\right) \otimes \tfrac{1}{t} J_t &&, \quad \mathrm{tr}\,(P_b) = s-1\,, \\
&P_{ab} = \left(I_r - \tfrac{1}{r} J_r\right) \otimes \left(I_s - \tfrac{1}{s} J_s\right) \otimes \tfrac{1}{t} J_t &&, \quad \mathrm{tr}\,(P_{ab}) = (r-1)(s-1)\,, \\
&P_e = I_{rs} \otimes (I_t - \tfrac{1}{t} J_t) &&, \quad \mathrm{tr}\,(P_e) = rs(t-1)\,.
\end{aligned} \tag{40}$$

The best invariant unbiased estimator of the covariance matrix $D(\sigma)$ (cf. Hartung and Voet (1986)) has the form

$$\widehat{D(\sigma)} = 2 \cdot \mathrm{diag}\left(\frac{M_1^2}{r+1}, \frac{M_2^2}{s+1}, \frac{M_3^2}{(r-1)(s-1)+2}, \frac{M_4^2}{rs(t-1)+2}\right) . \tag{41}$$

Thus, the Wald test statistic for testing the hypothesis $H_0 : \sigma_a^2 = \sigma_b^2$, with $K = (1,-1,0,0)^T$ and $d = 0$, can be described as

$$W_1 = \frac{\Big(r(M_1 - M_3) - s(M_2 - M_3)\Big)^2}{\dfrac{2r^2M_1^2}{r+1} + \dfrac{2s^2M_2^2}{s+1} + \dfrac{2(r-s)^2M_3^3}{(r-1)(s-1)+2}} \quad , \tag{42}$$

and we reject H_0 at level α if $W_1 > \chi^2_{1;1-\alpha}$.

In this model we explicitly consider the equations (27) which has to be solved by the maximum likelihood estimator under H_0. Here, (27) has the form

$$\begin{aligned}
(M_1 - \tau_1) + 2\lambda \frac{\tau_1^2}{st(r-1)} &= 0 \\
(M_2 - \tau_2) - 2\lambda \frac{\tau_2^2}{rt(s-1)} &= 0 \\
(M_3 - \tau_3) + 2\lambda \left(\frac{1}{rt} - \frac{1}{st}\right) \frac{\tau_3^2}{(r-1)(s-1)} &= 0 \\
(M_4 - \tau_4) &= 0 \\
\frac{1}{s}\tau_1 - \frac{1}{r}\tau_2 + \left(\frac{1}{r} - \frac{1}{s}\right)\tau_3 &= 0
\end{aligned} \tag{43}$$

If both main effects have the same number of levels, i. e. $r = s$, we get the following solution of (43)

$$\bar{\tau}_4 = M_4 \ , \ \bar{\tau}_3 = M_3 \ , \text{ and } \bar{\tau}_2 = \bar{\tau}_1 = (M_1 + M_2)/2 \quad . \tag{44}$$

So the test statistic (29) can be written as

$$LR_1 = (r-1)\left\{\ln\left(\frac{M_1+M_2}{2M_1}\right) + \ln\left(\frac{M_1+M_2}{2M_2}\right)\right\} \ . \tag{45}$$

Under H_0, the mean value of LR_1 is given by

$$\mathrm{E}\,(LR_1) = 2(r-1)\Big\{\mathrm{E}\ \ln\left(\chi^2_{2(r-1)}\right) - \mathrm{E}\ \ln\left(\chi^2_{r-1}\right) - \ln 2\Big\} \, . \tag{46}$$

In Bartlett and Kendall (1946) it is shown that the mean value of the logarithm of a χ^2–distributed random variable with f degrees of freedom is given by

$$\mathrm{E}\ \ln(\chi^2_f) = \ln 2 + \psi(f/2), \tag{47}$$

where $\psi(x) = \mathrm{d}\,\ln\Gamma(x)/\mathrm{d}x$ is the psi function. With an approximation of the psi function given in Abramowitz and Stegun (1964, p. 259) we get the following approximation of the mean value

$$\mathrm{E}\ \ln(\chi^2_f) \approx \ln f - \frac{1}{f} - \frac{1}{3f^2} + 2\frac{1}{15f^4} - \cdots \ , \tag{48}$$

and so it holds for the likelihood ratio test statistic from (45)

$$\mathrm{E}\,(LR_1) \approx 1 + \frac{1}{2(r-1)} \quad . \tag{49}$$

Therefore, we reject the hypothesis $H_0 : \sigma_a^2 = \sigma_b^2$ at the significance level α if

$$LR_1^* = \frac{LR_1}{1 + \dfrac{1}{2(r-1)}} > \chi^2_{1;1-\alpha} \, . \tag{50}$$

If the number of levels of the A– and B–factor are different, i. e. $r \neq s$, we get the following solution of (43)

$$\bar{\tau}_4 = M_4 \quad , \quad \bar{\tau}_3 = (s\bar{\tau}_2 - r\bar{\tau}_1)/(s-r) \tag{51}$$

and $\bar{\tau}_1$ and $\bar{\tau}_2$ are solutions of

$$\begin{aligned}(M_1-\tau_1)(s\tau_2-r\tau_1)^2+\tau_1^2(s\tau_2-r\tau_1-(s-r)M_3)r(s-1)=0\\(M_2-\tau_2)(s\tau_2-r\tau_1)^2+\tau_2^2(r\tau_1-s\tau_2-(r-s)M_3)s(r-1)=0\end{aligned} \tag{52}$$

Instead of using a numerical algorithm for computing a solution of (52) we use the following approximation. It holds

$$D(\sigma)\Lambda^{-1}K=\begin{bmatrix}2\sigma_a^4 st/(r-1)\\-2\sigma_b^4 rt/(s-1)\\0\\0\end{bmatrix}+O(r,s)\ , \tag{53}$$

where $O(r,s)\in I\!R^4$ and $\lim O(r,s)=0$ for $r\to\infty$ and $s\to\infty$.

Thus, instead of solving (43) we consider the following system of equations, where we omit the term $O(r,s)$,

$$\begin{aligned}(M_1-\tau_1)+2\lambda\sigma_a^4 st/(r-1) &= 0\\(M_2-\tau_2)-2\lambda\sigma_b^4 rt/(s-1) &= 0\\(M_3-\tau_3) &= 0\\(M_4-\tau_4) &= 0\\\frac{1}{s}\tau_1-\frac{1}{r}\tau_2+\left(\frac{1}{r}-\frac{1}{s}\right)\tau_3 &= 0\end{aligned} \tag{54}$$

The estimators for τ_3 and τ_4 are now given by

$$\bar{\tau}_3=M_3 \quad \text{and} \quad \bar{\tau}_4=M_4\ . \tag{55}$$

With

$$2\lambda=\frac{s-1}{rt}\frac{M_2+\tau_2}{\sigma_b^4} \tag{56}$$

the first equation in (54) can be written as

$$(M_1-\tau_1)+\frac{st(s-1)}{rt(r-1)}\frac{\sigma_a^4}{\sigma_b^4}(M_2-\tau_2)=0\ , \tag{57}$$

which under $H_0:\sigma_a^2=\sigma_b^2$ reduces to

$$(M_1-\tau_1)+\frac{s(s-1)}{r(r-1)}(M_2-\tau_2)=0\ . \tag{58}$$

Finally, we yield for τ_1 and τ_2 the estimators

$$\begin{aligned}\bar{\tau}_1 &= \frac{1}{(r-1)+(s-1)}\left((r-1)M_1+(s-1)\left(\frac{s}{r}M_2+(1-\frac{s}{r})M_3\right)\right)\\ \bar{\tau}_2 &= \frac{1}{(r-1)+(s-1)}\left((s-1)M_2+(r-1)\left(\frac{r}{s}M_1+(1-\frac{r}{s})M_3\right)\right)\end{aligned} \tag{59}$$

We note that the approximate solution of (54) coincides with the exact solution of (43), if the numbers of levels of the A– and B–factor are identical.

Now, we consider the expected value of the 'approximate' likelihood ratio test statistic using the estimators from (55) and (59) and observe that the last term in both equations on the right hand side of (59) is $O(r,s)$ so that estimators from (59) can be written as

$$\bar{\tau}_1 = \bar{\tau}_2 \approx \frac{1}{((r-1)+(s-1)}\left((r-1)M_1+(s-1)M_2\right). \tag{60}$$

Using (60) the likelihood ratio test statistic LR can be approximated as

$$LR^{\dagger} = (r-1)\ln\left(\frac{(r-1)M_1+(s-1)M_2}{((r-1)+(s-1))M_1}\right)+(s-1)\ln\left(\frac{(r-1)M_1+(s-1)M_2}{((r-1)+(s-1))M_2}\right), \tag{61}$$

which under H_0 is approximately

$$\begin{aligned}LR^{\dagger} &\approx ((r-1)+(s-1))\ln\chi^2_{(r-1)-(s-1)}-(r-1)\ln\chi^2_{r-1}-(s-1)\ln\chi^2_{s-1}\\ &-((r-1)+(s-1))\ln((r-1)+(s-1))+(r-1)\ln(r-1)+(s-1)\ln(s-1).\end{aligned} \tag{62}$$

With the approximation formula (48) we yield for the expected value of the likelihood ratio test statistic

$$\mathrm{E}\left(LR^{\dagger}\right) \approx 1+\frac{1}{3(r-1)}+\frac{1}{3(s-1)}-\frac{1}{3((r-1)+(s-1))}. \tag{63}$$

Using the inequality

$$\frac{1}{(r-1)+(s-1)} \leq \frac{1}{4}\left(\frac{1}{r-1}+\frac{1}{s-1}\right) \tag{64}$$

the expected value of (61) can also be approximated by

$$\mathrm{E}\left(LR^{\dagger}\right) \approx 1+\frac{1}{4(r-1)}+\frac{1}{4(s-1)}. \tag{65}$$

Therefore, we reject the hypothesis $H_0 : \sigma_a^2 = \sigma_b^2$, if

$$LR^{\ddagger} = LR^{\dagger}/c > \chi^2_{1;1-\alpha} \quad \text{with } c \text{ chosen from (65).} \tag{66}$$

We note that (50) and (66) are identical if $r = s$.

In a simulation study, which has been carried out in similar way like the one in example 4.1, we study the sizes of the proposed tests. We consider the likelihood ratio test using the Newton-Raphson ridge optimization algorithm to maximize the likelihood function under H_0, the 'approximate' likelihood ratio test from (66), and the Wald test from (42). The results are based in each case on 10,000 replications and the variance components σ_{ab}^2 and σ_e^2 have been set equal to one. For the variance components σ_a^2 and σ_b^2 we used the same process of generation as in the previous example. The results of this simulation study concerning the sizes of the tests are presented in table 3, where only the results with increasing r and three pairs of (s, t) are reported, because the results with increasing s and different pairs (r, t) are quite similar.

The estimated sizes of the likelihood ratio statistic (LR) exceed for all sample sizes the nominal significance level, but in an acceptable manner. Moreover, the estimated sizes do not depend on the sample sizes on the whole, they are rather homogeneous. Regarding the 'approximate' likelihood statistic (LR‡) we see that the consideration of a factor, which corrects for the expected value of the likelihood ratio test statistic in small sample sizes, has an important impact on the estimated sizes. In all considered cases, the variation of the estimated significance levels about the nominal significance level is rather small. The Wald test mainly produces very conservative results. In our simulation study for $\alpha = 0.01$ the Wald test never rejects the hypothesis, except for $r = 20$. For $\alpha = 0.05$ the estimated sizes are also very small, but sometimes we observe estimated sizes, which seriously exceed 0.05, e. g. the case $r = 3$, $s = 10$, $t = 15$. Thus, the test statistics LR and LR‡, respectively, are more appropriate for testing the hypothesis $H_0 : \sigma_a^2 = \sigma_b^2$ than the Wald test.

Finally, in table 4 the estimated power function of the 'approximate' likelihood statistic LR‡ and the Wald test are put together for fixed $s = 10$ and $t = 15$ with increasing r. Again, as critical values we used the corresponding simulated empirical 95 %–quantiles of the distributions of the test statistics under H_0. For $r > 10$ the estimated power function of the Wald test always lies above the estimated power function of the 'approximate' likelihood ratio test. The estimated power function are identical for $r = s = 10$ and for $r = 20$ the 'approximate' likelihood test $\mathrm{LR}^{\ddagger}$ has greater power than the Wald test. So, even if both tests have equal size the more powerful test depends on the sample sizes. But, for the 'approximate' likelihood ratio test we find out that with increasing r the estimated power functions monotonously grows for each δ with fixed s and t.

5 Conclusions

Our simulation studies show that the approach using the likelihood ratio test statistic yields better results concerning the nominal significance level in small sample sizes than the Wald test, on the whole. Especially, if one can deduce a factor, which corrects for the expected value of the likelihood ratio test statistic in finite samples, the modified likelihood ratio test nearly attains the prescribed size of the test. The Wald test is in general easier to compute in the considered class of variance components models, but the distribution of the Wald test statistic is badly approximated by a χ^2–distribution in small sample sizes. This fact is already known in tests on a single variance component to be equal to zero (cf. Kleffe and Seifert (1988)). But, if the sizes of both tests are similar our simulation study indicates that the power of the Wald test is often larger in these situations.

Acknowlegdement

We would like to thank two referees for their valuable comments on the first version of this manuscript. Furthermore, the support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Abramowitz, M. and I. A. Stegun (1964): *Handbook of Mathematical Functions.* Dover Publications, New York.

Anderson, T. W. (1973): Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics* **1**, 135–141.

Bartlett, M. S. and D. G. Kendall (1946): The statistical analysis of variance–heterogeneity and the logarithmic transformation. *Journal of the Royal Statistical Society B* **8**, 128–138.

Brown, K. G. (1976): Asymptotic behavior of MINQUE–type estimators of variance components. *The Annals of Statistics* **4**, 746–754.

El–Bassiouni, M. Y. and J. Seely (1980): Optimal tests for certain functions of the parameters in a covariance matrix with linear structure. *Sankhya* A **42**, 64–77.

Elpelt, B. (1989): On linear statistical models of commutative quadratic type. *Communications in Statistics – Theory and Methods* **18**, 3407–3450.

Graybill, F. A. (1976): *Theory and Application of the Linear Model.* Duxbury Press, North Scituate, Massachusetts.

Hartung, J. (1981): Nonnegative minimum biased invariant estimation in variance component models. *The Annals of Statistics* **9**, 278–292.

Hartung, J. and B. Voet (1986): Best invariant unbiased estimators for the mean squared error of variance component estimators. *Journal of the American Statistical Association* **81**, 689–691.

Hartung, J., Elpelt, B. and B. Voet (1997): *Modellkatalog Varianzanalyse.* Oldenbourg, München.

Humak, K. M. S. (1984): *Statistische Methoden der Modellbildung III. Statistische Inferenz für Kovarianzparameter.* Akademie–Verlag, Berlin.

Khuri, A. I., Mathew, T. and B. K. Sinha (1998): *Statistical Tests for Mixed Linear Models.* Wiley, New York.

Kleffe, J. and B. Seifert (1988): On the role of MINQUE in testing of hypotheses under mixed linear models. *Communications in Statistics – Theory and Methods* **17**, 1287–1309.

Rao, C. R. (1973): *Linear Statistical Inference and Its Applications.* Second Edition, Wiley & Sons, New York.

Rao, C. R. and J. Kleffe (1988): *Estimation of Variance Components and Applications.* North–Holland, Amsterdam.

Schmidt, W. H. and Thrum, R. (1981): Contributions to asymptotic theory in regression models with linear covariance structure. *Mathematische Operationsforschung Statistik, Series Statistics* **12**, 243–269.

Seely, J. (1971): Quadratic subspaces and completeness. *Annals of Mathematical Statistics* **42**, 710–721.

Seifert, B. (1979): Optimal testing for fixed effects in general balanced mixed classification models. *Mathematische Operationsforschung Statistik, Series Statistics* **10**, 237–255.

Table 1: Estimated size (in %) of the likelihood ratio (LR) and the Wald test for different sample sizes r, s, t ($\alpha = 0.01$ and $\alpha = 0.05$) in model (30) for testing $H_0 : \sigma_a^2 = \sigma_b^2$

			$\alpha = 0.01$		$\alpha = 0.05$	
r	s	t	LR	Wald	LR	Wald
3	3	3	1.8	0.1	7.0	0.2
3	5	6	1.5	0.4	6.9	15.4
3	8	10	1.6	14.3	6.7	23.3
4	3	3	1.4	0.0	6.6	2.8
4	5	6	1.5	3.5	6.2	13.2
4	8	10	1.4	11.5	6.5	19.4
5	3	3	1.4	0.0	5.9	3.5
5	5	6	1.3	3.8	5.8	12.1
5	8	10	1.3	10.2	5.9	16.9
6	3	3	1.1	0.0	5.6	4.0
6	5	6	1.2	4.0	5.6	10.7
6	8	10	1.2	8.6	5.7	15.1
7	3	3	1.4	0.1	6.1	4.6
7	5	6	1.3	3.7	5.7	9.7
7	8	10	1.2	8.0	5.6	13.8
8	3	3	1.4	0.3	5.6	4.3
8	5	6	1.1	3.8	5.5	9.6
8	8	10	1.3	7.4	5.8	13.2
9	3	3	1.1	0.4	5.6	4.3
9	5	6	1.2	3.4	5.2	8.4
9	8	10	1.0	6.7	5.4	12.5
10	3	3	1.2	0.6	5.9	4.9
10	5	6	1.1	3.7	5.6	8.9
10	8	10	1.1	6.0	5.4	11.1
20	3	3	1.1	0.9	5.5	5.1
20	5	6	1.2	2.7	5.5	7.0
20	8	10	1.0	3.9	5.2	8.3

Table 2: Estimated power (in %) of the likelihood ratio (LR) and the Wald test for different values $\delta = \sigma_b^2 - \sigma_a^2$, different sample sizes r, and, $s = 5$, $t = 6$ in model (30) for testing $H_0 : \sigma_a^2 = \sigma_b^2$ and critical values determined as empirical 95–quantiles of the 10,000 generated test statistics under H_0

		$\delta = \sigma_b^2 - \sigma_a^2$								
r	Test	0	0.5	1	2	5	10	20	50	100
3	LR	5	5	6	7	8	10	12	15	17
	Wald	5	6	7	9	13	17	22	28	32
4	LR	5	5	6	7	10	13	17	24	28
	Wald	5	6	8	10	16	21	28	37	43
5	LR	5	6	7	9	13	18	24	34	39
	Wald	5	7	9	12	19	26	35	47	55
6	LR	5	6	7	9	14	21	30	42	50
	Wald	5	7	10	13	21	31	42	58	66
7	LR	5	6	8	10	16	24	35	50	58
	Wald	5	8	11	15	25	36	49	66	74
8	LR	5	7	8	11	19	29	41	59	68
	Wald	5	8	11	16	27	39	54	72	80
9	LR	5	7	9	13	21	33	48	67	76
	Wald	5	9	12	18	30	44	61	80	87
10	LR	5	7	9	13	23	36	52	70	80
	Wald	5	9	12	18	31	46	63	81	89
20	LR	5	9	13	21	40	61	82	96	99
	Wald	5	12	17	27	49	70	88	98	100

Table 3: Estimated size (in %) of the likelihood ratio (LR), the approximate likelihood ratio (LR‡), and the Wald test for different sample sizes r, s, t ($\alpha = 0.01$ and $\alpha = 0.05$) in model (37) for testing $H_0 : \sigma_a^2 = \sigma_b^2$

			$\alpha = 0.01$			$\alpha = 0.05$		
r	s	t	LR	LR‡	Wald	LR	LR‡	Wald
3	5	6	1.5	1.1	0	7.0	5.2	0.0
3	8	10	1.5	1.4	0	6.8	5.6	2.9
3	10	15	1.4	1.4	0	6.7	5.7	9.1
4	5	6	1.6	1.0	0	6.9	5.1	0.0
4	8	10	1.6	1.3	0	6.3	5.0	1.6
4	10	15	1.1	1.1	0	5.8	4.9	4.9
5	5	6	1.5	1.0	0	6.9	5.4	0.0
5	8	10	1.3	1.1	0	6.0	4.9	0.8
5	10	15	1.6	1.2	0	6.3	5.4	3.3
6	5	6	1.2	0.8	0	6.1	4.8	0.0
6	8	10	1.3	0.9	0	6.0	5.9	0.4
6	10	15	1.3	1.1	0	6.2	5.3	2.1
7	5	6	1.4	1.0	0	6.0	4.9	0.1
7	8	10	1.2	1.0	0	5.8	4.9	0.3
7	10	15	1.4	1.2	0	6.0	5.2	1.4
8	5	6	1.3	0.9	0	6.1	4.8	0.8
8	8	10	1.1	1.0	0	5.7	4.8	0.2
8	10	15	1.6	1.2	0	6.1	5.3	1.2
9	5	6	1.3	1.0	0	5.8	4.8	1.8
9	8	10	1.2	1.0	0	6.0	5.2	0.6
9	10	15	1.5	1.2	0	6.3	5.7	1.2
10	5	6	1.3	1.0	0	6.0	5.0	3.1
10	8	10	1.2	1.0	0	5.7	4.9	1.1
10	10	15	1.2	0.9	0	5.7	5.2	1.0
20	5	6	1.1	1.2	3.3	5.8	5.3	11.3
20	8	10	1.1	0.9	1.0	5.5	5.0	6.0
20	10	15	1.1	0.9	0.5	5.0	4.6	4.4

Table 4: Estimated power (in %) of the approximate likelihood ratio (LR‡) and the Wald test for different values $\delta = \sigma_b^2 - \sigma_a^2$, different sample sizes r, and, $s = 10$, $t = 15$ in model (37) for testing $H_0 : \sigma_a^2 = \sigma_b^2$ and critical values determined as empirical 95–quantiles of the 10,000 generated test statistics under H_0

		$\delta = \sigma_b^2 - \sigma_a^2$								
r	Test	0	0.5	1	2	5	10	20	50	100
3	LR‡	5	5	5	6	9	14	21	37	53
	Wald	5	6	7	9	18	27	41	64	81
4	LR‡	5	5	6	8	14	22	35	59	78
	Wald	5	6	8	12	22	34	50	75	90
5	LR‡	5	6	7	9	17	27	43	69	86
	Wald	5	7	9	14	24	38	56	81	94
6	LR‡	5	6	7	11	20	32	50	77	92
	Wald	5	8	10	16	28	43	62	86	96
7	LR‡	5	6	8	12	22	36	55	82	95
	Wald	5	8	10	16	29	44	65	88	97
8	LR‡	5	6	8	12	23	38	58	85	96
	Wald	5	7	10	15	28	44	64	89	97
9	LR‡	5	6	9	13	24	39	60	87	97
	Wald	5	7	10	15	27	42	63	89	97
10	LR‡	5	6	9	14	26	42	64	89	98
	Wald	5	6	9	14	26	42	64	89	98
20	LR‡	5	8	12	18	34	54	78	96	99
	Wald	5	5	5	7	15	27	48	82	96