research
reports

No. 454 February 2005
2005 / 02 / 28

Ole F. Christensen,
Morten Frydenberg,
Jens Ledet Jensen and
Jørgen Granfeldt Pedersen

Tests and Confidence Intervals for an Extended Variance Component Using the Modified Likelihood Ratio Statistic

department of
theoretical
statistics
university of
aarhus

ISSN 0902-9893

Tests and confidence intervals for an extended variance component using the modified likelihood ratio statistic

Ole F. Christensen[*] Morten Frydenberg[†]
Jens Ledet Jensen[‡] Jørgen Granfeldt Pedersen[§]

Abstract

The large deviation modified likelihood ratio statistic is studied for testing a variance component equal to a specified value. Formulas are presented in the general balanced case, whereas in the unbalanced case only the one-way random effects model is studied. Simulation studies are presented, showing that the normal approximation to the large deviation modified likelihood ratio statistic gives confidence intervals for variance components with coverage probabilities very close to the nominal confidence coefficient.

Keywords: maximum likelihood; mixed models; random effects.

1 Introduction

In this paper we consider the use of the large deviation modified likelihood ratio test statistic for constructing confidence intervals for variance components. As the title suggests, we consider extended variance component models where the variance component is a covariance and as such can take negative values, see section 2.1. Formally, the test statistic can be described as follows. We consider a test of the hypothesis $\psi = \psi_0$, where ψ is one-dimensional and

[*]Bioinformatics Research Center, University of Aarhus, olefc@birc.au.dk
[†]Department of Biostatistics, University of Aarhus, morten@biostat.au.dk
[‡]Department of Theoretical Statistics, University of Aarhus, jlj@imf.au.dk
[§]Department of Theoretical Statistics, University of Aarhus, jqrgen@imf.au.dk

where we have a nuisance parameter η. With $l(\psi, \eta)$ denoting the log likelihood function, the signed likelihood ratio statistic is

$$r = \operatorname{sign}(\tilde{\psi} - \psi_0)\sqrt{2\left\{l(\tilde{\psi}, \tilde{\eta}) - l(\psi_0, \hat{\eta})\right\}},$$

where $(\tilde{\psi}, \tilde{\eta})$ is the maximum likelihood estimate under the full model and $\hat{\eta}$ is the maximum likelihood estimate under the hypothesis $\psi = \psi_0$. The *large deviation modified* likelihood ratio statistic takes the form

$$r_L = r + \frac{1}{r}\log\left(\frac{u}{r}\right),$$

where u denotes a statistic. The statistic u cannot be described in simple terms and can be quite difficult to derive. However, when the full model corresponds to a full exponential family it is possible to give an explicit formula for u. The modified statistic r_L has been studied during the last two decades and it has been realized that the modification to r not only improves the normal approximation in the center of the distribution, but also that the normal approximation has a small relative error in the tail of the distribution (Barndorff-Nielsen, 1986; Jensen, 1995, 1997). In Jensen (1995) formulas for calculating r_L have been given for a number of tests and here we supplement the list with classical test situations for variance component models. Since we consider test situations where the distribution of the likelihood ratio statistics depends on a nuisance parameter direct simulation of the distribution is not so obvious and even less so when looking for confidence intervals.

Variance components are of great interest in many applied areas. Good references both for the history and the state of the art of the subject are Searle *et al.* (1992) and Cox and Solomon (2003). For a detailed presentation of confidence intervals for variance components we refer to Burdick and Graybill (1992). The aim in this article is to construct a confidence interval for a given variance component, or equivalently to make a test for the hypothesis that it has a given value. For the general balanced model the method by Ting *et al.* (1990) is recommended by Burdick and Graybill (1992). For unbalanced designs, a comparison of four different methods is made in Lee and Khuri (2002), where it is demonstrated that two of the methods, the method in Thomas and Hultquist (1978) and the method in Khuri (2002) in general are superior to the others, and giving nearly identical results. For certain ratios of variance components exact confidence intervals can be calculated (see Burdick and Graybill, 1992), but ratios of variance components will not be considered here.

In section 2 we consider the balanced case and state the formula for r_L in a general setup. The derivation of the formula is given in section 3.

The unbalanced case is more difficult because there is no low-dimensional sufficient statistic. To handle this case we introduce a conditioning argument in section 4, thereby reducing the problem to one of the cases already treated.

In section 5 we present some simulation studies showing that for variance components, tests based on the modified statistics perform very well. For comparison, we study the methods suggested in Ting *et al.* (1990) and Thomas and Hultquist (1978).

2 Balanced random models

Tue Tjur (see Tjur, 1984) has identified a wide class of random effects models with k variance components, $\sigma_i^2, i = 1, \ldots, k$, where the inference for the variance components is based on a transformation of the data into k squared norms SS_i. These are independent and distributed as $\gamma_i \chi^2(f_i), i = 1, \ldots, k$, where the f_is are the degrees of freedom, and the parameters γ_i are called canonical variance components and are linear functions of the k variance components. The simplest such case is the balanced one-way random model which we treat first.

2.1 The balanced one-way random model

The traditional random effects model formulation for the jth observation in the ith sample, Y_{ij}, is

$$Y_{ij} = \mu + U_i + E_{ij}, \quad i = 1, \ldots, k, \; j = 1, \ldots, m, \tag{1}$$

where $U_i \sim N(0, \delta)$, $E_{ij} \sim N(0, \sigma^2)$, and all these variables are independent. The variance of Y_{ij} is $\delta + \sigma^2$ which explains the name variance components for δ and σ^2. The aim is to construct a confidence interval for δ or, equivalently, to make a test of the hypothesis of a specified value of δ. We will base the inference on the two independent statistics

$$SS_1 = \sum_{i,j} (y_{ij} - \bar{y}_{i\cdot})^2 \sim \gamma_1 \chi^2(f_1), \tag{2}$$

$$SS_2 = \sum_{i} m(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 \sim \gamma_2 \chi^2(f_2), \tag{3}$$

with

$$\gamma_1 = \sigma^2, \quad \gamma_2 = \sigma^2 + m\delta, \quad f_1 = mk - k, \quad f_2 = k - 1, \tag{4}$$

$$\text{and} \quad \bar{y}_{i\cdot} = \frac{1}{m} \sum_j y_{ij}, \quad \bar{y}_{\cdot\cdot} = \frac{1}{mk} \sum_{ij} y_{ij}.$$

The ANOVA estimates of σ^2 and δ are obtained by equating SS_1 and SS_2 to their means and solving for σ^2 and δ to give

$$\tilde{\sigma}^2 = \tilde{\gamma}_1 = \frac{SS_1}{f_1}, \quad \tilde{\delta} = \frac{1}{m}(\tilde{\gamma}_2 - \tilde{\gamma}_1) = \frac{1}{m}\left(\frac{SS_2}{f_2} - \frac{SS_1}{f_1}\right).$$

Clearly, the ANOVA estimate $\tilde{\delta}$ of δ may be negative which is inconsistent with the interpretation of δ as a variance. In order to circumvent this problem we will model the variance component δ as a covariance, i.e. we will consider the extended model where Y_{ij} is normal with mean μ and covariance

$$Cov(Y_{ij}, Y_{rs}) = \begin{cases} \delta + \sigma^2 & \text{if} \quad i = r,\ j = s \\ \delta & \text{if} \quad i = r,\ j \neq s \\ 0 & \text{if} \quad i \neq r \end{cases}. \tag{5}$$

It is the same covariance structure as in model (1), except that in (5) δ is not restricted to be positive. Consequently, SS_1 and SS_2 are independent and, furthermore, distributed as described by (2), (3) and (4) but now the domain of variation of (γ_1, γ_2) is $\mathbb{R}_+^2$.

Under the hypothesis $\delta = \delta_0$ the estimate $\hat{\gamma}_1$ is found by maximising the function

$$h(\gamma_1) = -\frac{SS_1}{\gamma_1} - \frac{SS_2}{\gamma_1 + m\delta_0} - f_1 \log(\gamma_1) - f_2 \log(\gamma_1 + m\delta_0),$$

which is twice the log likelihood function for γ_1 based on SS_1 and SS_2 and with δ fixed at δ_0. The function $h(\gamma_1)$ is of the form in appendix A with $d_1 = f_1$, $d_2 = f_2$, $s_1 = SS_1$, $s_2 = SS_2$, $a = 1$, and $c = m\delta_0$. The estimate of γ_2 under the hypothesis $\delta = \delta_0$ is $\hat{\gamma}_2 = \hat{\gamma}_1 + m\delta_0$. The signed square root r of the log likelihood ratio statistic can now be written

$$\begin{aligned} r = {} & \operatorname{sign}\left(\tilde{\delta} - \delta_0\right) \\ & \times \left\{ f_1\left(\frac{\tilde{\gamma}_1}{\hat{\gamma}_1} - 1\right) + f_1 \log\left(\frac{\hat{\gamma}_1}{\tilde{\gamma}_1}\right) + f_2\left(\frac{\tilde{\gamma}_2}{\hat{\gamma}_2} - 1\right) + f_2 \log\left(\frac{\hat{\gamma}_2}{\tilde{\gamma}_2}\right) \right\}^{1/2}. \end{aligned}$$

The new test statistic is

$$r_L = r + \frac{1}{r} \log \frac{u}{r}$$

with

$$\begin{aligned} u = {} & \operatorname{sign}\left(\tilde{\delta} - \delta_0\right) 2\sqrt{f_1 f_2}\tilde{\gamma}_1\tilde{\gamma}_2 \left| \left(\frac{1}{2\tilde{\gamma}_2} - \frac{1}{2\hat{\gamma}_2}\right)\hat{\gamma}_2^2 - \left(\frac{1}{2\tilde{\gamma}_1} - \frac{1}{2\hat{\gamma}_1}\right)\hat{\gamma}_1^2 \right| \\ & \times \left\{ 2f_1\hat{\gamma}_1(2\tilde{\gamma}_1 - \hat{\gamma}_1)\hat{\gamma}_2^4 + 2f_2\hat{\gamma}_2(2\tilde{\gamma}_2 - \hat{\gamma}_2)\hat{\gamma}_1^4 \right\}^{-1/2}, \end{aligned} \tag{6}$$

and r_L is approximately standard normal distributed. The derivation of u is given in section 3.

2.2 The general balanced case

In the general balanced case we have k independent statistics

$$SS_i \sim \gamma_i \chi^2(f_i), \quad i = 1, \dots, k.$$

The domain of variation of $(\gamma_1, \dots, \gamma_k)$ is $\mathbb{R}_+^k$ and the parameter of interest is

$$\delta = c_1\gamma_1 + c_2\gamma_2 + \cdots + c_k\gamma_k$$

with all the coefficients being non-zero. Note that the balanced random one-way model is the special case with $k = 2$, $c_1 = -1/m$ and $c_2 = 1/m$. An example where $k > 2$ is considered in section 5 — a three-way random effects model.

The estimates of the parameters γ_i in the full model are $\tilde{\gamma}_i = SS_i/f_i$, and the estimate of δ is

$$\tilde{\delta} = \sum_{i=1}^{k} c_i \tilde{\gamma}_i = \sum_{i=1}^{k} c_i SS_i / f_i.$$

Under the hypothesis $\delta = \delta_0$ we want to maximise the log-likelihood function, which amounts to maximising the function

$$h(\gamma_1, \dots, \gamma_k) = -\sum_{i=1}^{k} SS_i/\gamma_i - \sum_{i=1}^{k} f_i \log(\gamma_i), \quad \text{subject to} \quad \sum_{i=1}^{k} c_i\gamma_i = \delta_0.$$

A method for obtaining the estimates $\hat{\gamma}_1, \dots, \hat{\gamma}_k$ in this model is described in appendix B.

The signed square root r of the log likelihood ratio statistic becomes

$$r = \operatorname{sign}(\tilde{\delta} - \delta_0) \left\{ \sum_{i=1}^{k} \left(f_i \left(\frac{\tilde{\gamma}_i}{\hat{\gamma}_i} - 1 \right) + f_i \log \left(\frac{\hat{\gamma}_i}{\tilde{\gamma}_i} \right) \right) \right\}^{1/2}.$$

The new test statistic is

$$r_L = r + \frac{1}{r} \log \frac{u}{r}$$

where

$$u = \operatorname{sign}\,(\tilde{\delta} - \delta_0) \left(\prod_{i=1}^{k} \sqrt{2f_i}\tilde{\gamma}_i \right) \left| \sum_{i=1}^{k} \left(\frac{1}{2\tilde{\gamma}_i} - \frac{1}{2\hat{\gamma}_i} \right) c_i \hat{\gamma}_i^2 \right|$$
$$\times \left\{ \left(\prod_{i=1}^{k} d_i \right) \left(\sum_{i=1}^{k} \frac{c_i^2 \hat{\gamma}_i^4}{d_i} \right) \right\}^{-1/2} \qquad (7)$$

with

$$d_i = 2f_i\hat{\gamma}_i(2\tilde{\gamma}_i - \hat{\gamma}_i).$$

The derivation of u is given in section 3.

3 The r_L statistic for a full exponential family

In this section we describe the large deviation modified likelihood ratio statistic r_L for the case of a full exponential family. We then specialise to the case of independent chi-squared distributed variables to obtain the results of section 2. We use the formulation in Jensen (1997).

We consider an exponential family with minimal representation

$$\frac{dP_\theta}{d\nu}(x) = \exp\{\theta \cdot t(x) - \kappa(\theta)\}, \tag{8}$$

where θ and $t(x)$ are k-dimensional, $\theta \in \Theta$ with Θ an open subset of $\mathbb{R}^k$. In the following all vectors are row vectors. The mean value mapping is $\tau(\theta) = \frac{\partial \kappa}{\partial \theta}(\theta)$ and the variance function is $V(\theta) = \frac{\partial^2 \kappa}{\partial \theta \partial \theta^{\mathrm{T}}}(\theta)$. Under the full model the estimate $\tilde{\theta}$ of θ is the solution to

$$\tau(\theta) = t(x).$$

Let $\psi(\theta)$ be a real valued function defined on Θ. We consider the hypothesis $\psi(\theta) = \psi_0$ for some fixed value ψ_0. Let $\eta(\theta) = \frac{\partial \psi}{\partial \theta}(\theta)$ and assume that $\eta(\theta)_k > 0$ for all θ. The estimate of θ under the hypothesis $\psi(\theta) = \psi_0$ is denoted $\hat{\theta}$ and satisfies the equations

$$t(x)_i - \tau(\theta)_i - [t(x)_k - \tau(\theta)_k]\frac{\eta(\theta)_i}{\eta(\theta)_k} = 0, \quad i = 1, \ldots, k-1.$$

The signed square root r of the log likelihood ratio statistic is

$$r = \sqrt{2}\operatorname{sign}(\psi(\tilde{\theta}) - \psi_0)\left\{(\tilde{\theta} - \hat{\theta}) \cdot t(x) - \kappa(\tilde{\theta}) + \kappa(\hat{\theta})\right\}^{1/2}.$$

The large deviation modified likelihood ratio statistic is

$$r_L = r + \frac{1}{r}\log\left(\frac{u}{r}\right),$$

where from Jensen (1997) the statistic u is given by

$$u = \operatorname{sign}(\psi(\tilde{\theta}) - \psi_0)\left|V(\tilde{\theta})\right|^{1/2} |D|\,|A|^{-1/2}, \tag{9}$$

with

$$D = (\tilde{\theta}_k - \hat{\theta}_k) + \sum_{i=1}^{k-1}(\tilde{\theta}_i - \hat{\theta}_i)\frac{\eta(\hat{\theta})_i}{\eta(\hat{\theta})_k}, \tag{10}$$

and A is the $(k-1)\times(k-1)$ matrix with entries

$$A_{ij} = V(\hat{\theta})_{ij} - V(\hat{\theta})_{ik}\frac{\eta(\hat{\theta})_j}{\eta(\hat{\theta})_k} - \left(V(\hat{\theta})_{kj} - V(\hat{\theta})_{kk}\frac{\eta(\hat{\theta})_j}{\eta(\hat{\theta})_k}\right)\frac{\eta(\hat{\theta})_i}{\eta(\hat{\theta})_k} \tag{11}$$
$$+ (t(x)_k - \tau(\hat{\theta})_k)\left(\frac{\frac{\partial\eta_i}{\partial\theta_j}(\hat{\theta}) - \frac{\partial\eta_i}{\partial\theta_k}(\hat{\theta})\frac{\eta(\hat{\theta})_j}{\eta(\hat{\theta})_k}}{\eta(\theta)_k} - \frac{\eta(\hat{\theta})_i\left(\frac{\partial\eta_k}{\partial\theta_j}(\hat{\theta}) - \frac{\partial\eta_k}{\partial\theta_k}(\hat{\theta})\frac{\eta(\hat{\theta})_j}{\eta(\hat{\theta})_k}\right)}{\eta(\hat{\theta})_k^2}\right).$$

An alternative formulation is obtained if instead of $\psi(\theta) = \psi_0$ we specify the hypothesis as $\theta_k = \phi(\theta_1, \ldots, \theta_{k-1}, \psi_0)$ for some function ϕ. Using the relation

$$\psi(\theta_1, \ldots, \theta_{k-1}, \phi(\theta_1, \ldots, \theta_{k-1}, \psi_0)) = \psi_0$$

we find

$$\frac{\partial\phi}{\partial\theta_i}(\hat{\theta}, \psi_0) = -\frac{\frac{\partial\psi}{\partial\theta_i}(\hat{\theta})}{\frac{\partial\psi}{\partial\theta_k}(\hat{\theta})},$$

and

$$\frac{\partial^2\phi}{\partial\theta_i\partial\theta_j}(\hat{\theta}, \psi_0)$$
$$= \frac{-\frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}(\hat{\theta}) + \frac{\partial^2\psi}{\partial\theta_i\partial\theta_k}(\hat{\theta})\frac{\frac{\partial\psi}{\partial\theta_i}(\hat{\theta})}{\frac{\partial\psi}{\partial\theta_k}(\hat{\theta})}}{\frac{\partial\psi}{\partial\theta_k}(\hat{\theta})} + \frac{\frac{\partial\psi}{\partial\theta_i}(\hat{\theta})\left(\frac{\partial^2\psi}{\partial\theta_j\partial\theta_k}(\hat{\theta}) - \frac{\partial^2\psi}{\partial\theta_k^2}(\hat{\theta})\frac{\frac{\partial\psi}{\partial\theta_j}(\hat{\theta})}{\frac{\partial\psi}{\partial\theta_k}(\hat{\theta})}\right)}{\left(\frac{\partial\psi}{\partial\theta_k}(\hat{\theta})\right)^2}.$$

Thus (10) can alternatively be written as

$$D = (\tilde{\theta}_k - \hat{\theta}_k) - \sum_{i=1}^{k-1}(\tilde{\theta}_i - \hat{\theta}_i)\frac{\partial\phi}{\partial\theta_i}(\hat{\theta}, \psi_0), \tag{12}$$

and the matrix A from (11) is minus the second derivative of the log likelihood function with respect to $(\theta_1, \ldots, \theta_{k-1})$ under the hypothesis with $\theta_k = \phi(\theta_1, \ldots, \theta_{k-1}, \psi_0)$ and is given by

$$A_{ij} = V(\hat{\theta})_{ij} + V(\hat{\theta})_{ik}\frac{\partial\phi}{\partial\theta_j}(\hat{\theta}, \psi_0) + \left(V(\hat{\theta})_{kj} + V(\hat{\theta})_{kk}\frac{\partial\phi}{\partial\theta_j}(\hat{\theta}, \psi_0)\right)\frac{\partial\phi}{\partial\theta_i}(\hat{\theta}, \psi_0)$$
$$- (t(x)_k - \tau(\hat{\theta})_k)\frac{\partial^2\phi}{\partial\theta_i\partial\theta_j}(\hat{\theta}, \psi_0). \tag{13}$$

The formulas given here are for distributions having a density with respect to Lebesgue measure. If $k = 1$ we have $D = \tilde{\theta} - \theta$ and the term $|A|^{-1/2}$ is not present in the formula for u.

Typically, the calculation of r_L from the above formula will be difficult when r is very close to zero. However, since we want to use r_L in relation to testing a hypothesis we are mainly interested in the case of large values of $|r_L|$. Thus, if for example $|r| < 0.1$ the actual value of r_L will not be important, since r_L is close to r. In our implementation we defined $r_L = r$ when $|r| < 0.1$. As mentioned in the introduction r_L has a standard normal distribution to a high accuracy.

We now describe two examples that will give the formulas used in the previous section for variance component models.

Example 3.1 Let s_i, $i = 1, \ldots, k$ be independent with s_i having a scaled chi-squared distribution with f_i degrees of freedom. The joint density is

$$\left\{ \prod_{i=1}^{k} \frac{s_i^{f_i/2-1}}{\Gamma(f_i/2)} \right\} \exp \left\{ -\sum_{i=1}^{k} \theta_i s_i + \sum_{i=1}^{k} \frac{f_i}{2} \log(\theta_i) \right\}.$$

This corresponds to the situation in (8) with $\kappa(\theta) = -\sum_{i=1}^{k} \frac{f_i}{2} \log(\theta_i)$. The mean and variance functions are

$$\tau(\theta)_i = -\frac{f_i}{2\theta_i} \quad \text{and} \quad V(\theta)_{ij} = \begin{cases} \frac{f_i}{2\theta_i^2} & i = j \\ 0 & i \neq j. \end{cases}$$

Under the full model the estimate $\tilde{\theta}$ of θ is

$$\tilde{\theta}_i = \frac{f_i}{2s_i}, \quad i = 1, \ldots, k.$$

We define

$$\psi(\theta) = \sum_{i=1}^{k} \frac{c_i}{2\theta_i},$$

where all the c_is are non-zero, and consider the hypothesis $\psi(\theta) = \psi_0$. With this choice of $\psi(\theta)$ we have

$$\eta(\theta)_i = \frac{\partial \psi}{\partial \theta_i}(\theta) = -\frac{c_i}{2\theta_i^2} \quad \text{and} \quad \frac{\partial \eta_i}{\partial \theta_j}(\theta) = \begin{cases} c_i/\theta_i^3 & i = j \\ 0 & i \neq j. \end{cases}$$

Under the hypothesis the estimate $\hat{\theta}$ of θ is the solution to

$$-s_i + \frac{f_i}{2\theta_i} - \left(-s_k + \frac{f_k}{2\theta_k} \right) \frac{c_i \theta_k^2}{c_k \theta_i^2} = 0, \quad i = 1, \ldots, k-1, \quad \psi(\theta) = \psi_0. \tag{14}$$

The terms entering u in (9) are

$$|V(\tilde{\theta})| = \prod_{i=1}^{k} \frac{2s_i^2}{f_i},$$

$$D = (\frac{f_k}{2s_k} - \hat{\theta}_k) + \sum_{i=1}^{k-1} (\frac{f_i}{2s_i} - \hat{\theta}_i) \frac{c_i \hat{\theta}_k^2}{c_k \hat{\theta}_i^2} = \frac{1}{b_k} \sum_{i=1}^{k} \left(\frac{f_i}{2s_i} - \hat{\theta}_i \right) b_i, \tag{15}$$

where $b_i = c_i / \hat{\theta}_i^2$, and defining

$$d_i = \frac{f_i}{2\hat{\theta}_i^2} + \left(s_k - \frac{f_k}{2\hat{\theta}_k} \right) \frac{2c_i \hat{\theta}_k^2}{c_k \hat{\theta}_i^3} = \frac{f_i}{2\hat{\theta}_i^2} + \frac{2}{\hat{\theta}_i} \left(s_i - \frac{f_i}{2\hat{\theta}_i} \right) = \frac{1}{\hat{\theta}_i} \left(2s_i - \frac{f_i}{2\hat{\theta}_i} \right)$$

and

$$a_k = \frac{f_k \hat{\theta}_k^2}{2c_k^2} + \left(s_k - \frac{f_k}{2\hat{\theta}_k} \right) \frac{2\hat{\theta}_k^3}{c_k^2} = \frac{d_k}{b_k^2},$$

we find

$$A_{ij} = \begin{cases} a_k b_i b_j & i \neq j \\ d_i + a_k b_i^2 & i = j. \end{cases} \tag{16}$$

The determinant of A is then found to be

$$|A| = \left(1 + a_k \sum_{i=1}^{k-1} \frac{b_i^2}{d_i} \right) \prod_{i=1}^{k-1} d_i = \left(\sum_{i=1}^{k} \frac{b_i^2}{d_i b_k^2} \right) \prod_{i=1}^{k} d_i. \tag{17}$$

Writing $\hat{\gamma}_i = 1/(2\hat{\theta}_i)$ we obtain the formula (7) for the general balanced model in section 2.

Example 3.2 In the special case of Example 3.1 with $k = 2$ the estimate $\hat{\theta}$ is the solution to

$$-s_1 + \frac{f_1}{2\theta_1} - \left(-s_2 + \frac{f_2}{2\theta_2} \right) \frac{c_1 \theta_2^2}{c_2 \theta_1^2} = 0, \quad \frac{c_1}{2\theta_1} + \frac{c_2}{2\theta_2} = \psi_0.$$

From (15) and (17) we find the expression for u from (9) to be

$$u = \text{sign} \left(\frac{c_1 s_1}{f_1} + \frac{c_2 s_2}{f_2} - \psi_0 \right) \frac{2s_1 s_2}{\sqrt{f_1 f_2}} \left| \left(\frac{f_2}{2s_2} - \hat{\theta}_2 \right) \frac{c_2}{\hat{\theta}_2^2} + \left(\frac{f_1}{2s_1} - \hat{\theta}_1 \right) \frac{c_1}{\hat{\theta}_1^2} \right|$$
$$\times \left\{ \frac{1}{\hat{\theta}_1} \left(2s_1 - \frac{f_1}{2\hat{\theta}_1} \right) \frac{c_2^2}{\hat{\theta}_2^4} + \frac{1}{\hat{\theta}_2} \left(2s_2 - \frac{f_2}{2\hat{\theta}_2} \right) \frac{c_1^2}{\hat{\theta}_1^4} \right\}^{-1/2} \tag{18}$$

Taking $c_2 = -c_1 = 1/m$ and writing $\delta_0 = \psi_0$, $\hat{\gamma}_1 = 1/(2\hat{\theta}_1)$, and $\hat{\gamma}_2 = 1/(2\hat{\theta}_2) = \hat{\gamma}_1 + m\delta_0$ we get from (18) the expression (6) for the balanced one-way analysis of variance.

4 The unbalanced one-way random model

In many applications the number of observations in each group are not equal. The model we consider is as in (1) except now $j = 1, \dots, n_i$. The inference on the variance parameters σ^2 and δ should be based on

$$SS_1 = \sum_{i,j} (y_{ij} - \bar{y}_{i\cdot})^2 \sim \sigma^2 \chi^2(n-k), \tag{19}$$

and on

$$v_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}, \quad i = 1, \dots, k,$$

where $n = n_1 + \cdots + n_k$, $\bar{y}_{i\cdot} = \sum_j y_{ij}/n_i$, and $\bar{y}_{\cdot\cdot} = \sum_{ij} y_{ij}/n = \sum_i n_i \bar{y}_{i\cdot}/n$, and where SS_1 and $(v_1, \dots, v_k)$ are independent. Contrary to the balanced case we do not, however, have a sufficient reduction of the statistics $v_1, \dots, v_k$ to the squared norm

$$SS_2 = \sum_i n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_i n_i v_i^2,$$

and the latter is not chi-squared distributed. Transforming the density of $(\bar{y}_{1\cdot}, \dots, \bar{y}_{k\cdot})$ to $(v_1, \dots, v_{k-1}, \bar{y})$ and integrating with respect to $\bar{y}_{\cdot\cdot}$ we find that the marginal density of $(v_1, \dots, v_{k-1})$ is

$$\sqrt{\frac{2\pi}{\sum_{i=1}^k 1/\omega_i}} \frac{n}{n_k} \left(\prod_{i=1}^k \frac{1}{\sqrt{2\pi\omega_i}} \right) \exp\left\{ -\frac{1}{2} \sum_{i=1}^k v_i^2/\omega_i + \frac{1}{2} \frac{(\sum_{i=1}^k v_i/\omega_i)^2}{\sum_{i=1}^k 1/\omega_i} \right\}, \tag{20}$$

where

$$\omega_i = \delta + \sigma^2/n_i.$$

This is an exponential family where the order of the family increases with the number k of groups. The methodology based on the large deviation modified likelihood ratio statistic outlined in section 3 is therefore not directly applicable.

To get back to the situation of section 3 we condition on the statistics

$$z_i = v_i/\sqrt{SS_2}, \quad i = 1, \dots, k-1.$$

These are not exactly ancillary, but we take here a pragmatic view and condition on the z_is in order to reduce the dimension of the problem. If we transform the density (20) to the density of $(z_1, \dots, z_{k-1}, SS_2)$ we find that the conditional density of SS_2 given $(z_1, \dots, z_{k-1})$ is

$$\frac{(Q/2)^{(k-1)/2}}{\Gamma((k-1)/2)} SS_2^{(k-3)/2} \exp\{-Q SS_2/2\}, \tag{21}$$

with

$$Q = \sum_1^k z_i^2/\omega_i - \left(\sum_1^k z_i/\omega_i\right)^2 \left(\sum_1^k 1/\omega_i\right)^{-1}.$$

The independence of SS_1 and $(v_1, \ldots, v_k)$ implies that SS_1 and SS_2 are independent given the z_is, and from (19) and (21) we see that they both have a scaled chi-squared distribution. Thus we can use the general results of section 3. Since $\sum_1^k n_i z_i = 0$ and $\sum_1^k n_i z_i^2 = 1$ we find that $Q = (\sigma^2 + m\delta)^{-1}$ if $n_i = m$ for all i, and we are back to the situation from the balanced case as seen from (2) and (3). Similarly, if we consider the limit $n_i \to \infty$ with $n_i/m \to 1$ for all i then we again have $Q \sim (\sigma^2 + m\delta)^{-1}$. If we rewrite Q as $Q = \sum_1^k (1/\omega_i)(z_i - q/t)^2$ with $q = \sum_1^k z_i/\omega_i$ and $t = \sum_1^k 1/\omega_i$ we find that $\partial Q/\partial\delta = -\sum_1^k (1/\omega_i^2)(z_i - q/t)^2$. This shows that for a fixed value of σ^2, Q is a monotonic function of δ making it possible to base the inference on the conditional distribution given the z_is. Also, it is possible to show that the variance of the maximum likelihood estimate under the full model $\hat{\delta}$ is of order $1/(k-1)$ irrespectively of the value of the z vector. The argument is as follows. For a fixed value of σ^2 the estimate $\hat{\delta}$ is a function of $Q_0 SS_2/(k-1)$, where $Q_0 = Q(\delta_0)$ with δ_0 the true value of δ. The latter statistic has variance $1/(k-1)$ and using the general delta method we find that the variance of $\hat{\delta}$ is of order $[Q_0/\frac{\partial Q}{\partial \delta}(\delta_0)]^2/(k-1)$. Finally, from the formulae for Q and $\frac{\partial Q}{\partial \delta}$ we find trivially that there exist constants $0 < c_1 < c_2 < \infty$ such that $c_1 \leq |Q/\frac{\partial Q}{\partial \delta}| \leq c_2$ for $(\delta, \sigma 2)$ in a neighbourhood of their true value. Together, the above mentioned properties show that the conditional approach makes the situation very similar to the balanced case.

To use the results from section 3 we define

$$f_1 = n - k, \quad f_2 = k - 1, \quad \theta_1 = \frac{1}{2\sigma^2}, \quad \theta_2 = \frac{Q}{2},$$

and

$$\phi(\theta_1, \delta) = \frac{Q}{2} = \frac{1}{2}\sum_{i=1}^k \frac{1}{\omega_i}\left(z_i - \frac{q}{t}\right)^2,$$

where

$$\omega_i = \delta + \frac{1}{2n_i\theta_1}, \quad q = \sum_{i=1}^k \frac{z_i}{\omega_i} \text{ and } t = \sum_{i=1}^k \frac{1}{\omega_i}.$$

The maximum likelihood estimate in the full model is

$$\tilde{\theta}_1 = \frac{f_1}{2SS_1}, \quad \tilde{\theta}_2 = \frac{f_2}{2SS_2},$$

and the estimate $\hat{\theta}_1$ under the hypothesis $\delta = \delta_0$ or $\theta_2 = \phi(\theta_1, \delta_0)$ is the solution to the likelihood equation

$$-SS_1 + \frac{f_1}{2\theta_1} + \left(-SS_2 + \frac{f_2}{2\phi(\theta_1, \delta_0)}\right) \frac{\partial \phi}{\partial \theta_1}(\theta_1, \delta_0) = 0. \tag{22}$$

The estimate $\tilde{\delta}$ of δ in the full model is the solution to $\phi(\tilde{\theta}_1, \delta) = f_2/(2SS_2)$ and must be found by numerical methods. We note from (24) below that since $\phi(\tilde{\theta}_1, \delta)$ is a decreasing function of δ the sign of $\tilde{\delta} - \delta_0$ is the same as the sign of $\phi(\tilde{\theta}_1, \delta_0) - f_2/(2SS_2)$.

Using the notation $\hat{\theta}_2 = \phi(\hat{\theta}_1, \delta_0)$ the signed square root r of the log likelihood ratio statistic becomes

$$r = \operatorname{sign}\left(\phi(\tilde{\theta}_1, \delta_0) - \tilde{\theta}_2\right)$$
$$\times \left\{ f_1 \log\left(\frac{\tilde{\theta}_1}{\hat{\theta}_1}\right) + f_1\left(\frac{\hat{\theta}_1}{\tilde{\theta}_1} - 1\right) + f_2 \log\left(\frac{\tilde{\theta}_2}{\hat{\theta}_2}\right) + f_2\left(\frac{\hat{\theta}_2}{\tilde{\theta}_2} - 1\right)\right\}^{1/2}.$$

The new test statistic is

$$r_L = r + \frac{1}{r} \log \frac{u}{r}$$

where u is calculated from (9), (12), and (13),

$$u = \operatorname{sign}\left(\phi(\tilde{\theta}_1, \delta_0) - \tilde{\theta}_2\right) \frac{\sqrt{f_1 f_2}}{2\tilde{\theta}_1\tilde{\theta}_2} \left| \tilde{\theta}_2 - \hat{\theta}_2 - \left(\tilde{\theta}_1 - \hat{\theta}_1\right) \frac{\partial \phi}{\partial \theta_1}(\hat{\theta}_1, \delta_0) \right|$$
$$\times \left\{ \frac{f_1}{2\hat{\theta}_1^2} + \frac{f_2}{2\hat{\theta}_2^2} \frac{\partial \phi}{\partial \theta_1}(\hat{\theta}_1, \delta_0)^2 - \left(\frac{f_2}{2\hat{\theta}_2^2} - SS_2\right) \frac{\partial^2 \phi}{\partial \theta_1^2}(\hat{\theta}_1, \delta_0) \right\}^{-1/2}.$$

The derivatives of $\phi(\theta_1, \delta)$ are

$$\frac{\partial \phi}{\partial \theta_1}(\theta_1, \delta) = \sum_{i=1}^{k} \frac{1}{4 n_i \theta_1 2\omega_i^2} \left(z_i - \frac{q}{t}\right)^2, \tag{23}$$

$$\frac{\partial \phi}{\partial \delta}(\theta_1, \delta) = -\sum_{i=1}^{k} \frac{1}{2\omega_i^2} \left(z_i - \frac{q}{t}\right)^2, \tag{24}$$

and

$$\frac{\partial^2 \phi}{\partial \theta_1^2}(\theta_1, \delta) = -\sum_{i=1}^{k} \frac{1}{2 n_i \theta_1^3 \omega_i^3} \left(z_i - \frac{q}{t}\right)^2 - \left\{\sum_{i=1}^{k} \frac{1}{2 n_i \theta_1^2 \omega_i^2} \left(z_i - \frac{q}{t}\right)\right\}, \tag{25}$$

and where (23) and (25) are used in the calculation of u.

The solution $\hat{\theta}_1$ to (22) must be found by a numerical search. We start the iterative search at $\tilde{\theta}_1$ and use the Newton-Raphson method if the observed information is positive and the Fisher scoring method otherwise.

5 Simulation studies

In this section we will report the main results of some simulations studies.

Recall first the connection between tests and confidence intervals. Given a test of a simple hypothesis for a parameter, a confidence interval for that parameter with a nominal confidence coefficient of 1-α is obtained as the set of values of the parameter that would be accepted if used as null hypothesis and tested on level α. Conversely a procedure for constructing a 1-α confidence interval leads to a level α test which rejects the null hypothesis of a fixed value of the parameter if that value does not belong to the confidence interval.

In the simulations we have decided to formulate the results in terms of the tests, i.e. we give the observed levels of the tests when the nominal level is 5%. Since the signed likelihood ratio test r as well as its modification r_L are supposed to be symmetric we have recorded the frequency with which the tests reject the hypothesis in the upper and in the lower tail of the approximating normal distribution.

The results are easily translated to confidence intervals. The frequency with which the test rejects the null hypothesis in the lower tail is the frequency with which the confidence interval is entirely below the true value, and similarly the frequency with which the test rejects the null hypothesis in the upper tail is the frequency with which the confidence interval is entirely above the true value. The nominal confidence coefficient of a confidence interval is one minus the nominal level of the corresponding test and similarly the coverage probability of a confidence interval is one minus the observed level of the corresponding test. We are only investigating coverage of confidence intervals, and other measures such as the average length of intervals have not been considered.

For each simulated data set we have calculated the likelihood ratio statistics r and the modified statistics r_L for the hypothesis that δ equals the true value. For comparison, we have calculated an approximate 95% confidence interval for δ using methods suggested in Ting *et al.* (1990) for balanced random effects models and Thomas and Hultquist (1978) for the unbalanced one-way random model. We have then observed how often r and r_L were *above* 1.96 and how often the lower limit of the confidence interval was *above* the true value of δ, i.e. how often the three methods will reject the true hypothesis because the estimate is large. Similarly we observed how often r and r_L were *below* -1.96 and how often the upper limit of the confidence interval was *below* the true value of δ, i.e. how often the three methods will reject the true hypothesis because the estimate is small. Finally, we observed how often r and r_L were *outside* the interval -1.96 to 1.96 and how often the true value of δ was *outside* the confidence interval. Thus we have determined how often the three tests will reject the true hypothesis, in other words we have determined the levels of the three tests.

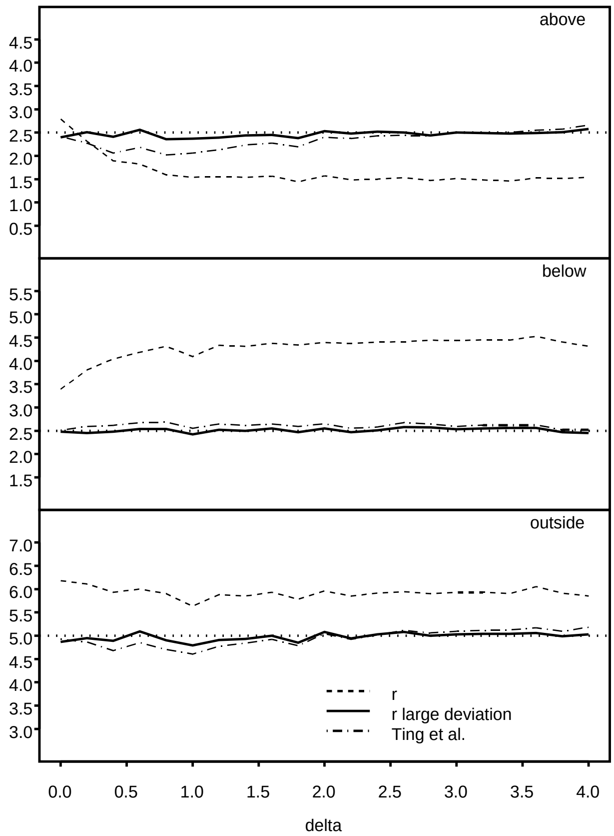

Figure 1: Balanced one-way random effects model with 5 groups and 2 replications within groups. Comparison of r, r_L, and the test based on the confidence intervals calculated according to Ting *et al.* (1990) based on 100000 simulations. Top plot: percentage of simulations with r above 1.96, r_L above 1.96, and lower limit of the confidence interval above the true value of δ. Middle plot: percentage of simulations with r below -1.96, r_L below -1.96, and upper limit of the confidence interval below the true value of δ. Bottom plot: percentage of simulations with r outside ± 1.96, r_L outside ± 1.96, and the true value of δ outside the confidence interval. In all three plots the dotted line shows the stated level of the tests.

Figure 1 shows the results for a balanced one-way random model with 5 groups and 2 replications within groups. The within group variance was set to 1 and the parameter of interest, the between group variance component, varies in the range from 0 to 4. The confidence interval was calculated using the method suggested in Ting *et al.* (1990). It is seen that the likelihood ratio statistics r is too often below -1.96 and too seldom above 1.96 compared to the stipulated level of 2.5%. Combining the tails one finds that the level of r is close to 6% rather than the stipulated 5%. The only exception is when δ is close to 0. The modified likelihood ratio statistics has both one- and two-sided levels very close to the stated levels. The test based on the confidence interval has observed one-sided and two-sided levels a bit further from the nominal levels than r_L but much closer to the nominal levels than r.

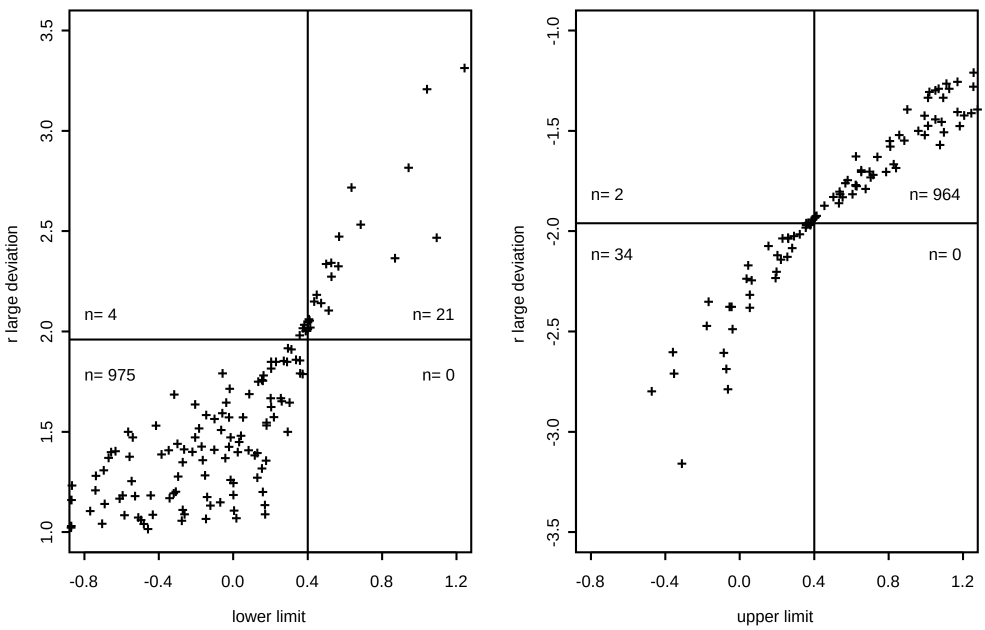

Figure 2: Balanced one-way random effects models with 5 groups and 2 replications within groups, $\delta = 0.4$. Left plot: r_L and lower limit of the confidence interval. Right plot: r_L and upper limit of the confidence interval. Confidence intervals calculated according to Ting *et al.* (1990). Based on 1000 simulations.

As the curves for r_L and the test based on the confidence interval follow each other one could suspect that the two methods give almost comparable results for a specific data set. To investigate this we have in Figure 2 plotted the values of r_L against the lower and upper limits of the confidence interval, respectively ($\delta = 0.4$, 1000 simulations). In the left plot it is seen that in 21 out of the 1000 simulations both methods rejected the true value, based on a one-sided test, while they only disagreed on 4 data sets. Similar results are found when considering the upper limit and small values of r_L, see the right plot. Based on this and other simulation studies for varying values of δ as well as varying number of groups and observations within groups the conclusion is that r_L and the method suggested by Ting *et al.* (1990) do in practice give comparable conclusions.

Figure 3 shows the results with 10 groups and 10 replications within groups. We see that with this increased sample size the performance of r is still poor, while r_L and the confidence interval have true levels extremely close to the stated levels. The method suggested by Ting *et al.* (1990) here gives results indistinguishable from r_L.

As an example of a general balanced model we have considered the three-way model involving P persons, D drugs and T timepoints

$$Y_{pdt} = \mu_{dt} + U_p + V_{pd} + W_{pt} + E_{pdt} \quad p = 1, \ldots, P; d = 1, \ldots, D; t = 1, \ldots, T, \tag{26}$$

where U_p, V_{pd}, W_{pt} and E_{pdt} are independent normal random components with means zero and variances $\delta, \sigma_V^2, \sigma_W^2$ and σ_E^2, respectively. The associated sums of squares are

$$\begin{aligned}
SS_1 &= \sum_{pdt} \left(y_{pdt} - \bar{y}_{pd\cdot} - \bar{y}_{p\cdot t} - \bar{y}_{\cdot dt} + \bar{y}_{p\cdot\cdot} + \bar{y}_{\cdot d\cdot} + \bar{y}_{\cdot\cdot t} - \bar{y}_{\cdots}\right)^2, \\
SS_2 &= \sum_{pd} T \left(\bar{y}_{pd\cdot} - \bar{y}_{p\cdot\cdot} - \bar{y}_{\cdot d\cdot} + \bar{y}_{\cdots}\right)^2, \\
SS_3 &= \sum_{pt} D \left(\bar{y}_{p\cdot t} - \bar{y}_{p\cdot\cdot} - \bar{y}_{\cdot\cdot t} + \bar{y}_{\cdots}\right)^2, \\
SS_4 &= \sum_{p} DT \left(\bar{y}_{p\cdot\cdot} - \bar{y}_{\cdots}\right)^2.
\end{aligned}$$

These are independent and $\gamma_i \chi^2 (f_i)$-distributed, where

$$\begin{aligned}
&\gamma_1 = \sigma_E^2, && f_1 = (P-1)(D-1)(T-1), \\
&\gamma_2 = \sigma_E^2 + T\sigma_V^2, && f_2 = (P-1)(D-1), \\
&\gamma_3 = \sigma_E^2 + D\sigma_W^2, && f_3 = (P-1)(T-1), \\
&\gamma_4 = \sigma_E^2 + T\sigma_V^2 + D\sigma_W^2 + DT\delta, && f_4 = (P-1).
\end{aligned}$$

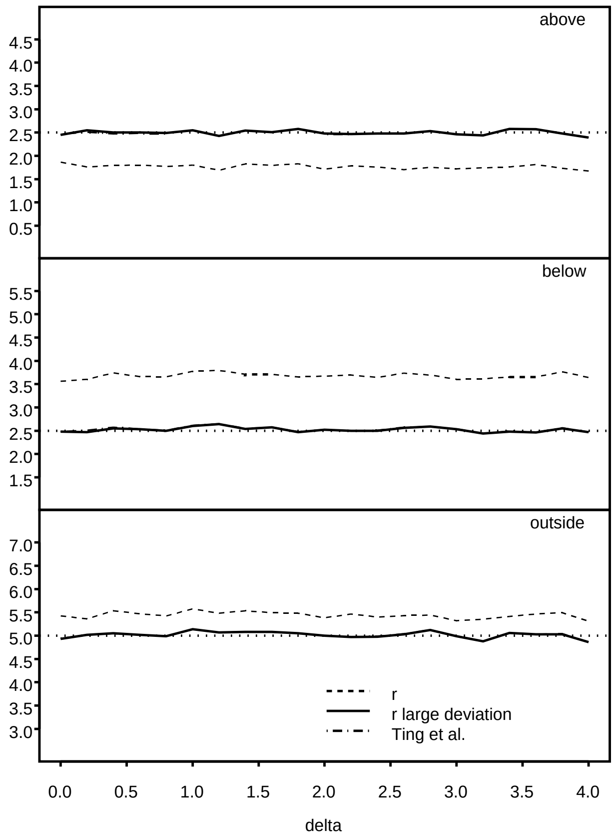

Figure 3: Balanced one-way random effects model with 10 groups and 10 replications within groups. Comparison of r, r_L and the test based on the confidence intervals calculated according to Ting *et al.* (1990) based on 100000 simulations. The results for the method suggested by Ting *et al.* (1990) are indistinguishable from the results for r_L and are therefore not visible in the plots. See Figure 1 for an explanation of the three plots.

Note that the random effect formulation in (26) imply that $0 < \gamma_1 < \gamma_2 < \gamma_4$ and $0 < \gamma_1 < \gamma_3 < \gamma_4$. But as in section 2.1 we consider the extended model where $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ vary freely in $\mathbb{R}_+^4$. Hence the parameter of interest δ is given as the following linear combination of the canonical variance components:

$$\delta = [\gamma_4 - \gamma_2 - \gamma_3 + \gamma_1] \, / \, (DT) \, .$$

In the simulations the parameters σ_V^2, σ_W^2 and σ_E^2 were set to 1.

Figure 4 illustrates the result when $P = 4$, $D = 2$ and $T = 8$. Again the performance of r is poor, while r_L has true levels very close to the stated levels. The test based on the confidence interval has observed one-sided and two-sided levels a bit further from the nominal levels than r_L but much closer to the nominal levels than r except when $\delta < 0.5$. In particular the observed upper level of the test based on the confidence interval is too small compared to the nominal level. Other choices of P, D and T gave comparable results.

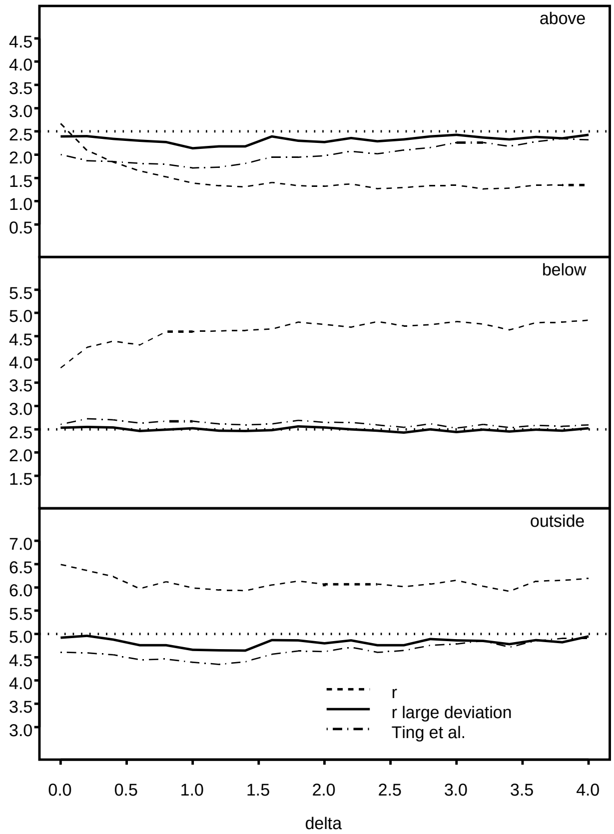

Figure 4: Three-way random effects model with $P = 4$, $D = 2$ and $T = 8$. Comparison of r, r_L and the test based on the confidence intervals calculated according to Ting *et al.* (1990) based on 100000 simulations. See Figure 1 for an explanation of the three plots.

Figure 5 shows the results for an unbalanced one-way random model with

6 groups and group sizes $2, 5, 5, 7, 7$, and 9, respectively. Here we have used the method from Thomas and Hultquist (1978) for the confidence intervals.

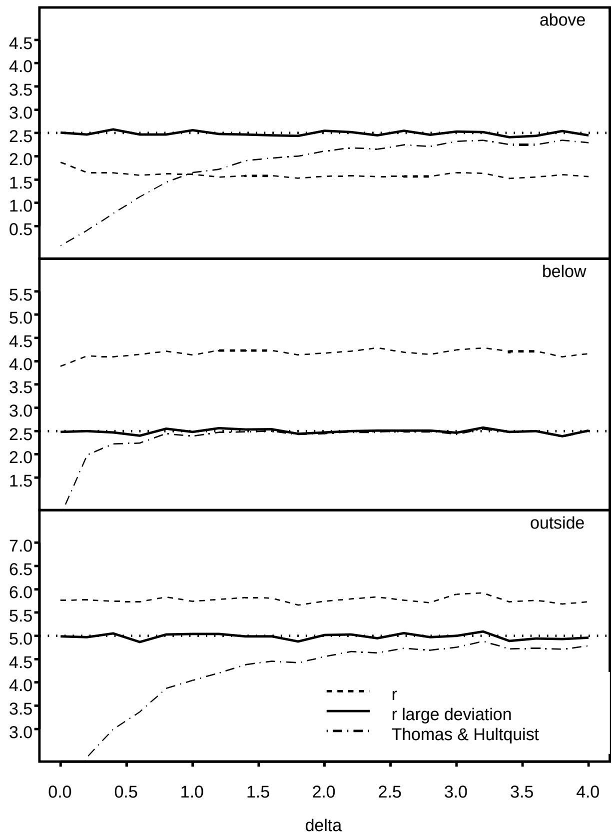

Figure 5: Unbalanced one-way random effects model with group sizes $2, 5, 5, 7, 7$ and 9. Comparison of r, r_L and the test based on the confidence intervals calculated according to Thomas and Hultquist (1978) based on 100000 simulations. See Figure 1 for an explanation of the three plots.

Figure 6 summarises the result for the situation with 3 groups and group sizes $2, 10$, and 100, respectively.

The likelihood ratio statistic r has again observed levels rather far from the stated level. The r_L statistic have observed levels very close to the stated level, one-sided as well as two-sided. The upper and lower bounds from the Thomas and Hultquist-method have observed coverage probabilities very close to the nominal confidence coefficients, when δ_0 is not too close to 0.

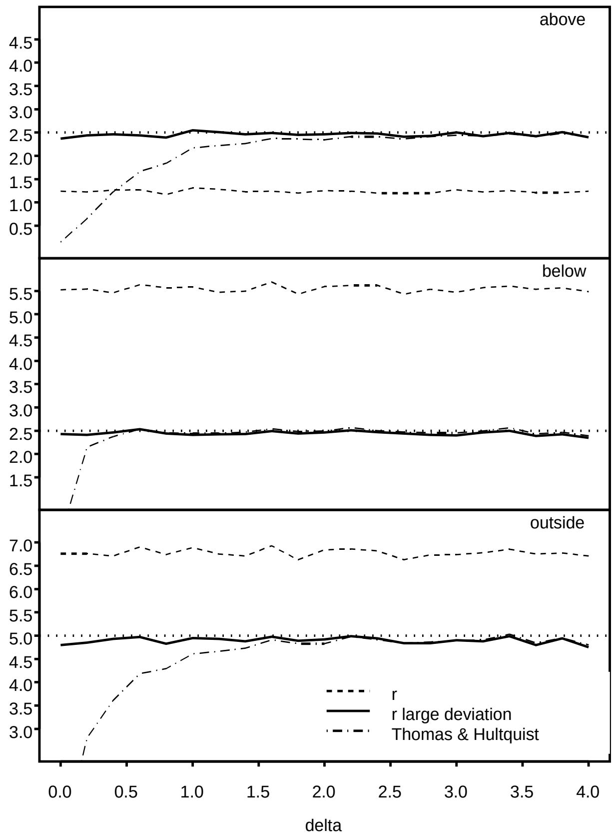

Figure 6: Comparison of r, r_L and the test based on the confidence intervals calculated according to Thomas and Hultquist (1978) based on 100000 simulations. See Figure 1 for an explanation of the three plots.

But when δ_0 is close to 0 this method has observed coverage probabilities much larger than the nominal confidence coefficients. For the Thomas and Hultquist-method similar observations were made in Lee and Khuri (2002).

6 Conclusion

Based on the simulations presented above and other simulation studies it is evident that for small samples inference concerning variance components should not be based on the signed likelihood ratio statistics. On the other hand, our simulations show that the large deviation modified likelihood ratio statistic produces tests and hence 95%-confidence limits with true levels very close to the stated levels. For balanced models, confidence intervals based on Ting *et al.* (1990) performed almost as well. For unbalanced one-way random effects models, confidence intervals based on r_L clearly performs better than the confidence intervals based on Thomas and Hultquist (1978) for small values for the variance component of interest. Furthermore, the formulas for the large deviation modified likelihood ratio statistic presented here are easily implemented; some **R** functions for the general balanced case are available on request.

The balanced case was handled in general using that the sum of squares statistics are independent with a scaled chi-squared distribution. In the unbalanced case we have only considered the one-way random model, where one of the sum of squares statistics is no longer chi-squared distributed. To overcome this problem a conditional argument is used: conditioning on a set of standardised group means the distribution of the sum of squares statistic is turned into a scaled chi-squared distribution. This conditioning idea may also be useful for constructing formulas in more general unbalanced cases, but we have not pursued this idea here.

Appendices

Appendix A

Let $\psi(\gamma) = a\gamma + c$ with the range of γ being

$$\begin{cases} \gamma > 0 & \text{if } a > 0,\ c \geq 0 \\ \gamma > -c/a & \text{if } a > 0,\ c < 0 \\ 0 < \gamma < -c/a & \text{if } a < 0,\ c > 0. \end{cases}$$

We want to maximise

$$l(\gamma) = -\frac{s_1}{\gamma} - d_1 \log(\gamma) - \frac{s_2}{\psi(\gamma)} - d_2 \log(\psi(\gamma)),$$

where $s_1 > 0,\ s_2 > 0,\ d_1 > 0$, and $d_2 > 0$. Let us note that the function tends to minus infinity when γ tends to the boundary of its range, so that the maximum is always attained. To find the maximum we want to solve

$$0 = \frac{s_1}{\gamma^2} - \frac{d_1}{\gamma} + \left[\frac{s_2}{\psi(\gamma)^2} - \frac{d_2}{\psi(\gamma)}\right] a,$$

which is equivalent to solving the cubic equation

$$-[d_1 + d_2]a^2\gamma^3 + [s_1 a^2 + s_2 a - (2d_1 + d_2)ac]\gamma^2 + [2s_1 ac - d_1 c^2]\gamma + s_1 c^2 = 0.$$

The real solutions to the cubic equations must be checked for consistency with the definition of the range of γ and then $\hat{\gamma}$ is found by evaluating l at the solutions.

Appendix B

Here we consider the maximisation of

$$\sum_{i=1}^{k} \left\{ -\frac{SS_i}{\gamma_i} - f_i \log(\gamma_i) \right\} \quad \text{subject to} \quad \sum_{i=1}^{k} c_i \gamma_i = \delta_0, \tag{27}$$

where δ_0 is fixed, all c_is are non-zero, and all SS_is are positive. Using a Lagrange multiplier we consider the function

$$\sum_{i=1}^{k} \left\{ -\frac{SS_i}{\gamma_i} - f_i \log(\gamma_i) \right\} + \lambda(\sum_{i=1}^{k} c_i \gamma_i - \delta_0).$$

Setting the derivative with respect to γ_i equal to zero we find

$$\gamma_i(\lambda) = \begin{cases} \gamma_i^-(\lambda) & \text{if } \lambda c_i < 0, \\ \gamma_i^-(\lambda) \text{ or } \gamma_i^+(\lambda) & \text{if } \lambda c_i > 0, \end{cases}$$

where

$$\gamma_i^-(\lambda) = \frac{f_i - \sqrt{f_i^2 - 4\lambda c_i SS_i}}{2\lambda c_i}, \quad \gamma_i^+(\lambda) = \frac{f_i + \sqrt{f_i^2 - 4\lambda c_i SS_i}}{2\lambda c_i}.$$

To study these solutions we define $\omega_i = 4c_i SS_i / f_i^2$, $i = 1, \dots, k$, and assume that these values have been ordered so that $\omega_1 > \omega_2 > \cdots > \omega_\nu > 0 > \omega_{\nu+1} > \cdots > \omega_k$, with $\nu = k$ when $c_i > 0$ for all i. The range of λ in the solutions $\gamma_i(\lambda)$ and $\gamma_i^+(\lambda)$ is given by the requirements $f_i^2 - 4\lambda c_i SS_i = f_i^2(1 - \lambda\omega_i) > 0$ for all i. With the ordering of ω_i this gives $\lambda_2 < \lambda < \lambda_1$, where $\lambda_1 = \omega_1^{-1}$. If $k > \nu$ the value of λ_2 is ω_k^{-1}, and if $k = \nu$ the value of λ_2 is $-\infty$. Since

$$c_i(\gamma_i^-)'(\lambda) = \frac{c_i^2}{\sqrt{f_i^2(1 - \lambda\omega_i)}}(\gamma_i^-)^2, \quad \text{and} \quad c_i(\gamma_i^+)'(\lambda) = \frac{-c_i^2}{\sqrt{f_i^2(1 - \lambda\omega_i)}}(\gamma_i^+)^2,$$

we see that $c_i\gamma_i^-(\lambda)$ is increasing and $c_i\gamma_i^+(\lambda)$ is decreasing. Furthermore,

$$\gamma_i^+(\lambda) \to \infty \ \text{ for } \lambda \to 0 \text{ and } \lambda c_i > 0.$$

Define now

$$\delta^-(\lambda) = \sum_{i=1}^{k} c_i \gamma_i^-(\lambda),$$

$$\delta_1(\lambda) = c_1 \gamma_1^+(\lambda) + \sum_{i=2}^{k} c_i \gamma_i^-(\lambda),$$

$$\delta_2(\lambda) = \begin{cases} \sum_{i=1}^{k-1} c_i \gamma_i^-(\lambda) + c_k \gamma_k^+(\lambda) & \text{if } k > \nu \\ \delta^-(\lambda) & \text{if } k = \nu. \end{cases}$$

Then δ^- is an increasing function,

$$\delta^-(\lambda_1) = \delta_1(\lambda_1) \text{ and } \delta_1(\lambda) \to \infty \text{ for } \lambda \searrow 0,$$

and if $k > \nu$

$$\delta^-(\lambda_2) = \delta_2(\lambda_2) \text{ and } \delta_2(\lambda) \to -\infty \text{ for } \lambda \nearrow 0,$$

and finally if $k = \nu$

$$\delta^-(\lambda) \to 0 \text{ for } \lambda \to -\infty.$$

These observations show that we can always find a value $\hat{\lambda}$ of λ satisfying

$$\begin{cases} \delta^-(\lambda) = \delta_0 & \text{if } \delta^-(\lambda_2) \le \delta_0 \le \delta^-(\lambda_1) \\ \delta_1(\lambda) = \delta_0, \ \lambda > 0 & \text{if } \delta_0 > \delta^-(\lambda_1) \\ \delta_2(\lambda) = \delta_0, \ \lambda < 0 & \text{if } \delta_0 < \delta^-(\lambda_2). \end{cases}$$

The solution to any one of these equations can be found simply doing a bisection search. Let $\hat{\gamma}_i^0$ be the corresponding value of γ_i used in deriving $\hat{\lambda}$.

Test cases show that in the majority of cases the above procedure will give the maximum likelihood estimates. To capture the remaining cases we use $\hat{\gamma}^0$ as the initial estimate for a second search procedure. If we fix all but two of the variables in (27) the maximisation problem is as in appendix A. We therefore iteratively fix all but variables i and j, $1 \le i \ne j \le k$, and maximise with respect to γ_i, γ_j. We continue with this procedure until convergence. When using appendix A we take

$$d_1 = f_i, \quad d_2 = f_j, \quad s_1 = SS_i, \quad s_2 = SS_j,$$

$$a = -\frac{c_i}{c_j} \text{ and } c = \frac{\delta_0 - \sum_{l \notin \{i,j\}} c_l \gamma_l}{c_j},$$

where γ_l is the present value in the search.

To check that we have reached a maximum we must show that the $(k-1)\times(k-1)$ matrix A given by (16) is positive definite. This is the case if and only if

$$\left(1+a_k\sum_{i=1}^{r}\frac{b_i^2}{d_i}\right)\prod_{i=1}^{r}d_i>0,\quad r=1,\ldots,k-1.$$

References

Barndorff-Nielsen, O. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–322.

Burdick, R. and Graybill, F. (1992). *Confidence intervals on variance components*. Marcel Decker, Inc., New York.

Cox, D. and Solomon, P. (2003). *Components of variance*. Chapman and Hall/CRC, Boca Raton.

Jensen, J. (1995). *Saddlepoint Approximations*. Clarendon Press, Oxford.

Jensen, J. (1997). A simple derivation of r^* for curved exponential families. *Scand. J. Statist.* **24**, 33–46.

Khuri, A. (2002). Graphical evaluation of the adequacy of the method of unweighted means. *J. Appl. Statist.* **29**, 1107–1119.

Lee, J. and Khuri, A. (2002). Comparison of confidence intervals on the among-group variance component for the unbalanced one-way random model. *Commun. Statist. Simula.* **31**, 35–47.

Searle, S., Casella, G. and McCulloch, C. (1992). *Variance components*. Wiley, New York.

Thomas, J. and Hultquist, R. (1978). Interval estimation for the unbalanced case of the one-way random effects model. *Ann. Statist.* **6**, 582–587.

Ting, N., Burdick, R., Graybill, F., Jeyaratnam, S. and Lu, T.-F. (1990). Confidence intervals on linear combinations of variance components that are unrestricted in sign. *J. Statist. Comput. Simul.* **35**, 135–143.

Tjur, T. (1984). Analysis of variance models in orthogonal designs. *International Statistical Review* **52**, 33–65.