Likelihood inference for small variance components

Steven E. STERN and A. H. WELSH

ABSTRACT

The authors explore likelihood-based methods for making inferences about the components of variance in a general normal mixed linear model. In particular, they use local asymptotic approximations to construct confidence intervals for the components of variance when the components are close to the boundary of the parameter space. In the process, they explore the question of how to profile the restricted likelihood (REML). Also, they show that general REML estimates are less likely to fall on the boundary of the parameter space than maximum likelihood estimates and that the likelihood ratio test based on the local asymptotic approximation has higher power than the likelihood ratio test based on the usual chi-squared approximation. They examine the finite sample properties of the proposed intervals by means of a simulation study.

RÉSUMÉ

Les auteurs explorent l'emploi de méthodes fondées sur la vraisemblance pour l'inférence concernant les composantes de la variance dans le cadre d'un modèle linéaire général mixte sous le postulat de normalité. Ils utilisent notamment des approximations asymptotiques locales pour construire des intervalles de confiance pour les composantes de la variance lorsque celles-ci sont proches de la frontière de l'espace des paramètres. Ce faisant, ils s'interrogent sur la façon optimale de profiler la vraisemblance restreinte (VRAR). Ainsi montrent-ils que les estimations VRAR sont généralement moins susceptibles de se trouver sur la frontière de l'espace que celles obtenues par vraisemblance maximale et que le test du rapport des vraisemblances fondé sur l'approximation asymptotique locale est plus puissant que celui qui s'appuie sur l'approximation du khi-deux usuelle. Des simulations illustrent les propriétés des intervalles proposés dans de petits échantillons.

1. INTRODUCTION

Linear models with multiple components of variance are widely used for modelling variability in structured data such as arises from sample surveys and designed experiments. The components of variance are often regarded as nuisance parameters with interest directed towards the form of the mean function in the model. However, in some applications, such as in genetics, the components of variance are of

1

direct interest and it is important to be able to make inference about them. In this paper, we explore general likelihood-based methods for making such inferences and, in particular, constructing confidence intervals for the components of variance.

Formally, suppose $y = (y_1, \ldots, y_n)'$ is an observation on the linear mixed model:

$$y = X\alpha + \sum_{r=1}^{p} Z_r u_r + e, \tag{1}$$

where X and Z_r are known $n \times k$ and $n \times c_r$ design matrices, α is a k-vector of fixed effects, the u_r are c_r-vectors of random effects, and e is an n-vector of unobserved errors. We assume e and the u_r's are all independent of one another. Moreover, we assume the elements of each u_r are independent and normally distributed with mean zero and variance σ_r^2, while the elements of e are independent and normally distributed with mean zero and variance σ_{p+1}^2. Thus

$$E(y) = X\alpha \quad \text{and} \quad \text{Var}(y) = V = \sum_{r=1}^{p+1} \sigma_r^2 J_r,$$

where $J_r = Z_r Z_r'$ for $r = 1, \ldots, p$ and J_{p+1} is the $n \times n$ identity matrix. The components of the vector $\sigma^2 = (\sigma_1^2, \ldots, \sigma_{p+1}^2)'$ are restricted to the non-negative half-line and are referred to as components of variance or variance components.

Likelihood-based inferences about the entire variance vector σ^2 in model (1) are commonly made using the profile log-likelihood, L_p:

$$L_p(\sigma^2) = L\{\sigma^2, \tilde{\alpha}(\sigma^2)\} = -\frac{1}{2}y'Py - \frac{1}{2}\log|V|,$$

where $\tilde{\alpha}(\sigma^2) = (X'V^{-1}X)^{-1}X'V^{-1}y$ is the constrained maximum likelihood estimator of α for fixed σ^2 and $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$. However, the profile score function, ∇L_p, has $E(\nabla L_p) = O(1)$ and $\text{Var}(\nabla L_p) + E\{\nabla'(\nabla L_p)\} = O(1)$; in other words, $L_p(\sigma^2)$ is score and information biased. An adjusted version of the profile log-likelihood can be constructed by adding $\beta(\sigma^2) = -\frac{1}{2}\log|X'V^{-1}X|$ to $L_p(\sigma^2)$. The resulting adjusted profile log-likelihood, known as the restricted or residual log-likelihood (REML) (Patterson & Thompson 1971, 1974), is given by

$$L_R(\sigma^2) = -\frac{1}{2}y'Py - \frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X|, \tag{2}$$

which is score and information unbiased. For scale inference in normal mixed models, this REML log-likelihood coincides with the conditional profile likelihood of Cox & Reid (1987) and the modified profile likelihood of Barndorff-Nielsen (1983).

In general, the entirety of σ^2 is not of interest and inference is desired for only one or some of its components. In such situations, the REML log-likelihood cannot be used directly for inference since it depends on nuisance parameters, and a further profiling must be employed. How we should do this and what, if any, adjustments should be incorporated into the resulting profile log-likelihood to make it score and/or information unbiased is explored in Section 2. There we show that profiling the REML log-likelihood with "constrained REML" estimates rather than the usual constrained maximum likelihood estimates is the preferred method.

We consider setting confidence intervals for the components of variance themselves. The close connection between confidence intervals and tests means that our results are also relevant to hypothesis testing. Also, our methods can be adapted to

obtain confidence intervals for various functions of the variance components. However, it is often possible to construct exact or nearly exact inference procedures specific to the function of the variance components under study. We focus on the variance components themselves, where exact inferences are rarely possible.

For the balanced one-way classification model, where $p = 1$, X is an n-vector of ones, and J_1 is a block-diagonal matrix of $nc_1^{-1} \times nc_1^{-1}$ matrices of ones, exact confidence intervals exist for σ_2^2, the intra-class correlation coefficient $\sigma_1^2/(\sigma_1^2 + \sigma_2^2)$ and certain other functions of the variance components; see Searle *et al.* (1992). Exact intervals for σ_1^2 are not available in this case, but an approximate interval was obtained by Williams (1962). For a review of intervals for the unbalanced one-way classification, see Burdick & Graybill (1988).

More generally, inferences for model (1) are based on a normal approximation to the distribution of the maximum likelihood or REML estimators. For convenience, let $\tilde{\sigma}^2$ denote the maximum likelihood estimator, $\bar{\sigma}^2$ the REML estimator, and let $\hat{\sigma}^2$ denote either estimator as appropriate when we discuss them together. Similarly, let $\ell(\sigma^2)$ denote either the profile log-likelihood, $L_P(\sigma^2)$, or the REML log-likelihood, $L_R(\sigma^2)$. The Fisher information, $I = -E[\nabla'\{\nabla\ell(\sigma^2)\}]$, is the $(p+1) \times (p+1)$ matrix with (r, s)-th element

$$I_{rs} = \begin{cases} \frac{1}{2}\mathrm{tr}(PJ_rPJ_s) & \text{for REML;} \\ \frac{1}{2}\mathrm{tr}(V^{-1}J_rV^{-1}J_s) & \text{for maximum likelihood.} \end{cases} \tag{3}$$

Under mild conditions, the REML and maximum likelihood estimators are asymptotically equivalent (cf. Cressie & Lahiri 1993, Richardson & Welsh 1994) and the two expressions for I, when scaled by n^{-1}, have the same limit. However, better finite sample approximations to the distributions of the estimators are obtained using the appropriate expression from (3). Thus, an approximate $100(1 - \alpha)\%$ confidence interval for σ_1^2 is given by

$$\left[\hat{\sigma}_1^2 - \Phi^{-1}(1 - \alpha/2)\sqrt{\hat{I}^{11}}, \hat{\sigma}_1^2 + \Phi^{-1}(1 - \alpha/2)\sqrt{\hat{I}^{11}}\right], \tag{4}$$

where Φ denotes the standard normal cumulative distribution function, $\hat{I}^{rs}$ denotes the (r, s)-th component of $\hat{I}^{-1}$, and $\hat{I}$ is the matrix $I = I(\sigma^2)$ evaluated at $\hat{\sigma}^2$.

The coverage accuracy of (4) is often poor in small samples, particularly when some variance components are near zero. In such situations, (4) often includes negative values. Truncation of (4) at zero solves the problem of negative values, but does not improve coverage accuracy. The poor coverage properties of (4) and similar intervals are, in part, a consequence of the fact that if σ^2 lies on the boundary, the asymptotic distribution of $\hat{\sigma}_1^2$ is not normal, but rather a mixture of normal and point mass distributions (Moran 1971). So, for the case of small variance components, we would like to set confidence intervals based on an asymptotic approximation to the distribution of maximum likelihood and REML estimators which "interpolates" between the boundary and non-boundary cases. Such an approximation should be a mixture distribution which becomes more normal as σ^2 moves away from the boundary. Moran (1971) obtained a result of this type for evaluating the local power of a Wald test of the hypothesis that one variance component was equal to zero; see also Self & Liang (1987).

Typically, inferences based on approximations to the distribution of the likelihood ratio prove superior to inferences based on approximations to the distribution of the parameter estimators. Therefore, we consider constructing confidence intervals by inverting a likelihood ratio test based on a version of the profile REML

3

log-likelihood. There has been some recent work on the problem of testing hypotheses about the fixed effect parameters α when the variance components have been estimated by REML (cf., e.g., Richardson & Welsh 1996, Welham & Thompson 1997, Kenward & Roger 1997) but there is little on using REML to make inferences about the variance components of a general mixed model. The same boundary issues that arise in obtaining asymptotic approximations for the estimators arise for likelihood ratios as well. Chernoff (1954) showed that when the parameter is on the boundary, the asymptotic distribution of the likelihood ratio is a mixture of χ^2 distributions. In Section 3, we obtain asymptotic approximations for the distributions of both estimators and likelihood ratios which apply when parameters are near the boundary and which therefore interpolate between the boundary and non-boundary cases. These results apply equally to the case of REML estimators and REML likelihood ratios.

In Section 4, we discuss the construction of confidence intervals using various methods based on standard asymptotics as well as the local asymptotics of Section 3. Section 5 reports on a small simulation study evaluating the finite sample properties of these intervals for the one-way classification model.

2. PROFILING THE LOG-LIKELIHOOD

In making likelihood-based inferences about a subset of q of the variance components in the model (1), we need to profile the likelihood over both α and the other $p - q + 1$ variance components. To avoid proliferation of subscripts and to simplify the presentation, let $\tau = (\tau_1, \ldots, \tau_{p-q+1})$ denote variance components over which we need to profile, let $\theta = (\theta_1, \ldots, \theta_q) = (\sigma_1^2, \ldots, \sigma_q^2)$ the remaining variance components, and reorder the variance components if necessary, so that $\sigma^2 = (\theta, \tau)$. We consider several possible methods of profiling over both α and τ and compare them on the basis of their score and information biases. We then specialise the methods to the one-way classification model to gain additional insight into the approaches and to simplify subsequent implementation in this particular case.

2.1. Profile Log-Likelihood.

The simplest approach is to replace both α and τ in the log-likelihood by their respective constrained maximum likelihood estimators. In this context, the constrained maximum likelihood estimator of α is given by $\tilde{\alpha}\{\theta, \tilde{\tau}(\theta)\}$, while the constrained maximum likelihood estimator $\tilde{\tau}(\theta)$ of τ satisfies the system

$$y'P\{\theta, \tilde{\tau}(\theta)\}J_r P\{\theta, \tilde{\tau}(\theta)\}y = \text{tr}[V\{\theta, \tilde{\tau}(\theta)\}^{-1}J_r], \quad r = q+1, \ldots, p+1. \tag{5}$$

So, the profile log-likelihood for θ is given by

$$L_P(\theta) = L_p\{\theta, \tilde{\tau}(\theta)\} = -\frac{1}{2}y'\tilde{P}y - \frac{1}{2}\log|\tilde{V}|, \tag{6}$$

where $\tilde{P}$ and $\tilde{V}$ are the matrices $P = P(\theta, \tau)$ and $V = V(\theta, \tau)$ evaluated at $\tilde{\tau}(\theta)$. Not surprisingly, this approach is neither score nor information unbiased. This suggests we adjust the profile log-likelihood to reduce these biases.

2.2. β-Adjusted Profile Log-Likelihood.

If estimation of α is the main source of score and information bias in the profile log-likelihood for θ, we may be able to reduce these biases by incorporating the

4

REML adjustment function, $\beta(\theta,\tau) = -\frac{1}{2}\log|X'V^{-1}X|$, evaluated at the constrained maximum likelihood estimator, $\tilde{\tau}(\theta)$, into the profile log-likelihood (6); which amounts to simply evaluating the REML log-likelihood (2) at the constrained maximum likelihood estimator. Thus, the β-adjusted profile log-likelihood is

$$L_{RP}(\theta) = L_R\{\theta, \tilde{\tau}(\theta)\} = -\frac{1}{2}y'\tilde{P}y - \frac{1}{2}\log|\tilde{V}| - \frac{1}{2}\log|X'\tilde{V}^{-1}X|. \qquad (7)$$

The score bias associated with this method is of order $O(n^{-1})$, since the quantity $\operatorname{tr}(P - V^{-1})$ is of order $O(1)$, even though $\operatorname{tr}(V^{-1})$ and $\operatorname{tr}(P)$ are generally of order $O(n)$. However, L_{RP} is not generally information unbiased.

2.3. B-Adjusted Profile Log-Likelihood.

The score bias reduction of the β-adjusted profile log-likelihood is fortuitous because the adjustment function $\beta\{\theta, \tilde{\tau}(\theta)\}$ is not specifically constructed to adjust for the effect of profiling over both α and τ together. This suggests that we should be able to construct a better adjustment function.

In quite general circumstances, Stern (1997) showed that a score-unbiased adjusted profile log-likelihood can be constructed using an appropriate adjustment of the profile log-likelihood. In the case that $q = p$, this adjustment function is:

$$\begin{aligned}
B(\theta) &= \frac{1}{4}\operatorname{tr}(\tilde{V}^{-2}J_{p+1}) \sum_{r=1}^{p}\sum_{s=1}^{p} \tilde{I}^{rs}\{y'\tilde{P}J_r\tilde{P}y - (\tilde{V}^{-1}J_r)\} \\
&\quad \times \quad [\operatorname{tr}\{(X'\tilde{V}^{-1}X)^{-1}X'\tilde{V}^{-1}J_{p+1}\tilde{V}^{-1}X\}\operatorname{tr}(\tilde{V}^{-2}J_s) \\
&\quad - \quad \operatorname{tr}\{(X'\tilde{V}^{-1}X)^{-1}X'\tilde{V}^{-}J_s\tilde{V}^{-1}X\}\operatorname{tr}(\tilde{V}^{-2}J_{p+1})]
\end{aligned}$$

with $\tilde{I}^{rs}$ the (r,s)-th component of the inverse Fisher information matrix for the maximum likelihood estimator, $I^{-1}(\sigma^2) = I^{-1}(\theta, \tau)$ given in (3), evaluated at $\tilde{\tau}(\theta)$. The formula for the general case is easily derived, but unwieldy and thus not presented here. The adjusted profile log-likelihood is then just $L_{PA}(\theta) = L_P(\theta) + B(\theta)$. Many other authors have worked on adjustments to reduce score bias, including Bartlett (1955), Barndorff-Nielsen (1983, 1994), Barndorff-Nielsen & Cox (1984), Cox & Reid (1987, 1992), Liang (1987), Levin & Kong (1990), McCullagh & Tibshirani (1990), Barndorff-Nielsen & Chamberlin (1992), DiCiccio & Stern (1993), and Ghosh & Mukerjee (1994). There has also been some work on bias-reduction of the estimators themselves, including Firth (1993) and Kuk (1995).

Stern (1997) also showed how to construct an adjustment functions designed to reduce both score and information biases. Several other authors including Godambe (1960), DiCiccio et $al.$ (1996) and McCullagh & Tibshirani (1990), have also worked on reducing information bias. However, the effect of adjusting for information bias after adjusting for score bias is often small in practice, so explicit additional adjustments to reduce information bias are generally omitted. Nonetheless, information bias reduction provides a useful criterion for choosing between potential score bias adjustment functions and between profiling methods.

2.4. Profiled REML Log-Likelihood.

Rather than employing adjustments to the profile log-likelihood, we might try profiling over τ in a different way. Since the REML log-likelihood has been adjusted for the effect of profiling over α, it makes intuitive sense to use the "constrained

5

REML" estimate of τ, $\bar{\tau}(\theta)$ to profile the REML log-likelihood, rather than using the constrained maximum likelihood estimate as was done in (7). The constrained REML estimate $\bar{\tau}(\theta)$ satisfies the system

$$y'P\{\theta, \bar{\tau}(\theta)\}J_r P\{\theta, \bar{\tau}(\theta)\}y = \text{tr}[P\{\theta, \bar{\tau}(\theta)\}J_r], \quad r = q+1, \ldots, p+1$$

and the profiled REML log-likelihood is given by

$$L_{RR}(\theta) = L_R\{\theta, \bar{\tau}(\theta)\} = -\frac{1}{2}y'\bar{P}y - \frac{1}{2}\log|\bar{V}| - \frac{1}{2}\log|X'\bar{V}^{-1}X|,$$

where $\bar{P}$ and $\bar{V}$ are the matrices $P = P(\theta, \tau)$ and $V = V(\theta, \tau)$ evaluated at $\bar{\tau}(\theta)$.

The profiled REML log-likelihood $L_{RR}(\theta)$ is both score and information unbiased. This assertion is a consequence of the fact that $L_{RR}(\theta)$ can be viewed as the profile log-likelihood for data from a normal distribution with zero mean vector and variance matrix P^{-1}. Since θ is a simple function of the canonical parameter, this profile log-likelihood is score and information unbiased. The fact that L_{RP} is score but not information unbiased is a consequence of the fact that $\text{tr}(P - V^{-1}) = O(1)$ implies $\bar{\tau}(\theta) - \tilde{\tau}(\theta) = O(n^{-1})$, meaning the the score functions of L_{RP} and L_{RR} are equivalent to first order (preserving score unbiasedness) but not to second order.

2.5. The One-way Classification Model.

The one-way classification model corresponds to the case $p = 1$ and $X = 1_n$. We write $\theta = \sigma_1^2$ and $\tau = \sigma_2^2$ and suppose we are interested in θ. Furthermore, we suppose that the matrix Z_1 corresponds to a nested design having c_1 random effect levels with m_i observations within the i-th level, so that $\sum_{i=1}^{c_1} m_i = n$. Then, $J = Z_1 Z_1'$ is a $n \times n$ block-diagonal matrix with i-th diagonal matrix J_{m_i}, a $m_i \times m_i$ matrix of ones and V^{-1} is a block-diagonal matrix with i-th diagonal component

$$V_i^{-1} = \frac{1}{\tau}I_{m_i} - \frac{\theta}{\tau(\tau + m_i\theta)}J_{m_i},$$

where I_{m_i} is the $m_i \times m_i$ identity matrix. In this case, the profile log-likelihood is:

$$L_P(\theta) = -\frac{1}{2}\sum_{i=1}^{c_1}\sum_{j=1}^{m_i}\left\{\frac{(y_{ij} - \bar{y}_i)^2}{\tilde{\tau}(\theta)} - \frac{(\bar{y}_i - \bar{y})^2}{\tilde{\tau}(\theta) + m_i\theta}\right\} - \frac{1}{2}\sum_{i=1}^{c_1}\log[\tilde{\tau}(\theta)^{m_i-1}\{\tilde{\tau}(\theta) + m_i\theta\}],$$

where $\bar{y}_i = m_i^{-1}\sum_{j=1}^{m_i}y_{ij}$, $\bar{y} = \sum_{i=1}^{c_1}\frac{m_i\bar{y}_i}{\tau + m_i\theta} / \sum_{i=1}^{c_1}\frac{m_i}{\tau + m_i\theta}$, and $\tilde{\tau}(\theta)$ solves:

$$0 = \frac{1}{\tau^2}\sum_{i=1}^{c_1}\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{c_1}\frac{m_i(\bar{y}_i - \bar{y})^2}{(\tau + m_i\theta)^2} - \sum_{i=1}^{c_1}\frac{1}{\tau + m_i\theta} - \frac{n - c_1}{\tau}. \quad (8)$$

The β-adjusted profile log-likelihood in this case is:

$$L_{RP}(\theta) = L_P(\theta) - \frac{1}{2}\log\left\{\sum_{i=1}^{c_1}\frac{m_i}{\tilde{\tau}(\theta) + m_i\theta}\right\},$$

where the constrained maximum likelihood estimator $\tilde{\tau}(\theta)$ is defined by (8). Moreover, using the fact that $-\log(x) \approx 1 - x$ as $x \to 1$, Stern's (1997) adjustment function can be written in the form:

$$B(\theta) = -\frac{1}{2}\log\left[\left\{\sum_{i=1}^{g}\frac{m_i}{\tilde{\tau}(\theta) + m_i\theta}\right\}^{-1}\sum_{i=1}^{g}\frac{m_i^2(\bar{y}_i - \bar{y})^2}{\{\tilde{\tau}(\theta) + m_i\theta\}^2}\right].$$

This form of $B(\theta)$ allows for more direct comparison with the REML adjustment functions. The profiled REML log-likelihood, $L_{RR}(\theta)$, is of the same form as $L_{RP}(\theta)$ but replaces $\tilde{\tau}(\theta)$ by the constrained REML estimator $\bar{\tau}(\theta)$, which solves:

$$\frac{1}{\tau^2}\sum_{i=1}^{c_1}\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_i)^2+\sum_{i=1}^{c_1}\frac{m_i(\bar{y}_i-\bar{y})^2}{(\tau+m_i\theta)^2}-\sum_{i=1}^{c_1}\frac{1}{\tau+m_i\theta}-\frac{n-c_1}{\tau}+\left\{\frac{\sum_{i=1}^{c_1}\frac{m_i}{(\tau+m_i\theta)^2}}{\sum_{i=1}^{c_1}\frac{m_i}{\tau+m_i\theta}}\right\}.$$

Calculations for the one-way classification model (not reported) confirm that, in addition to the theoretical advantages, the $\bar{\tau}(\theta)$-profiled REML log-likelihood generally yields the most accurate intervals of those we have considered.

3. LOCAL ASYMPTOTIC RESULTS

Asymptotic approximations to the distribution of estimators for range-restricted parameters are available for the case that the parameter of interest is interior to the parameter space and the case that it is on the boundary. In the latter case, the probability that the estimator takes a boundary value is important, so we begin by comparing the frequency of this occurrence for the maximum likelihood and REML estimators. We then develop local approximations to the distributions of these estimators similar to those of Moran (1971). These results allow interpolation between the two cases and provide approximations which can be used for inference.

3.1. The Probability of a Zero Estimate.

Consider first the balanced one-way classification model. In this case, exact calculation of the probability of a zero estimate is possible, and it is well known that the REML estimator has lower probability of equalling zero than the maximum likelihood estimator; see Searle (1992). The difference between the two probabilities is most marked for small c_1, since the estimators are asymptotically equivalent and $\lim_{c_1\to\infty}Pr(\tilde{\theta}=0)-Pr(\bar{\theta}=0)=0$.

In general, exact calculation of the probability of a zero estimate is not possible. Nonetheless, a general result is possible. Here, we must distinguish between the two forms of I given in (3), so let $\lambda_{rs}=\frac{1}{2}\mathrm{tr}(PJ_rPJ_s)$ denote the Fisher information for the REML estimator, and $\nu_{rs}=\frac{1}{2}\mathrm{tr}(V^{-1}J_rV^{-1}J_s)$ denote the Fisher information for the maximum likelihood estimator.

LEMMA 1. *Provided the REML adjustment function $\beta(\sigma^2)=-\frac{1}{2}\log|X'V(\sigma^2)^{-1}X|$ is concave, λ_{rs} and ν_{rs} are of order $O(n)$ and $\beta_{rs}=\frac{\partial^2}{\partial\sigma_r^2\partial\sigma_s^2}\beta(\sigma^2)=O(1)$ for $r,s=1,\ldots,p+1$, then $Pr(\bar{\sigma}_r^2=0)<Pr(\tilde{\sigma}_r^2=0)\{1+O(n^{-1/2})\}$.*

The proof relies on the standard result that $Pr(\hat{\sigma}_r^2=0)=\Phi\{-(I^{rr})^{-1/2}\sigma_r^2\}\{1+O(n^{-1/2})\}$. Since $L_R(\sigma^2)=L_P(\sigma^2)+\beta(\sigma^2)$, it is clear that $\lambda_{rs}=\nu_{rs}-\beta_{rs}$, and thus $\lambda^{rr}=\nu^{rr}+\frac{1}{2}\sum_{s=1}^{p+1}\sum_{t=1}^{p+1}\nu^{rs}\nu^{rt}\beta_{st}+O(n^{-3})$. The matrix β_{rs} is positive semi-definite because $\beta(\sigma^2)$ is assumed concave. Thus, $\nu^{rr}<\lambda^{rr}+O(n^{-3})$, which implies that $(\lambda^{rr})^{-1/2}<(\nu^{rr})^{-1/2}+O(n^{-3/2})$ and the result then follows.

3.2. Asymptotic Distribution of the Maximum Likelihood and REML Estimators.

Suppose that σ^2 is partitioned (and reordered, if necessary) as $\sigma^2=(\phi_1,\phi_2)$, where ϕ_1 is the single component of σ^2 which is close to the boundary and ϕ_2 denotes

7

the remaining p components of σ^2. The asymptotic distribution of either the maximum likelihood or REML estimators when ϕ_1 is near the boundary is given in the following theorem, which is essentially Theorem IV of Moran (1971).

THEOREM 1. *Let $\sigma^2 = (\phi_1, \phi_2)$ with $\phi_1 = n^{-1/2}a, a \in [0, a_0)$ and the components of ϕ_2 interior to the parameter space. Also let $C = \lim_{n\to\infty} n^{-1}I$. Then, for all $t = (t_1, t_2)$, with $t_1 \geq -a$ and $t_2 \in \mathbb{R}^p$, uniformly in $a \in [0, a_0)$, $Pr\{n^{1/2}(\hat{\sigma}^2 - \sigma^2) \leq t\} \to \delta F_1(t_1, t_2) + (1 - \delta)F_2(t_1, t_2)$, where*

a) $1 - \delta = \Phi\{-\frac{a}{\sqrt{C^{11}}}\}$, *with C^{11} the $(1,1)$-th component of the matrix C^{-1};*

b) $F_1(t_1, t_2)$ *is the $(p+1)$-variate distribution of (z_1, z_2) on $\{t_1 > -a, t_2 \in \mathbb{R}^p\}$ which has the density of a $N(0, C^{-1})$ distribution conditional on $z_1 > -a$; and*

c) $F_2(t_1, t_2)$ *is the $(p+1)$-variate distribution of (z_1, z_2) on $\{t_1 = -a, t_2 \in \mathbb{R}^p\}$ which has the joint density of the distribution of $z_1 = -a$, and $z_2 = C_{22}^{-1}(y_2 + C_{21}a)$ where C_{11}, $C_{21} = C_{12}'$ and C_{22} are the 1×1, $p \times 1$ and $p \times p$ submatrices of C corresponding to the partitioning of $\sigma^2 = (\phi_1, \phi_2)$, and (y_1, y_2) has a $N(0, C)$ distribution conditional on $y_1 + C_{11}a - C_{12}C_{22}^{-1}(y_2 + C_{21}a) \leq 0$.*

The proof parallels that of Moran (1971) and is omitted.

3.3. Marginal Distribution When the Parameter of Interest is Near the Boundary.

Suppose that $\theta = \phi_1$ so that the parameter of interest is near the boundary. Then,

$$\delta F_1(t_1, t_2) = \frac{|C|^{1/2}}{(2\pi)^{(p+1)/2}} \int_{-a}^{t_1} \int_{-\infty}^{t_{21}} \cdots \int_{-\infty}^{t_{2p}} \exp\left(z'Cz/2\right) dz_1 dz_{21} \cdots dz_{2p},$$

where we have used the notation $t_2 = (t_{21}, \ldots, t_{2p})'$ and $z_2 = (z_{21}, \ldots, z_{2p})'$. Thus, the marginal cumulative distribution function of $z_1 = n^{1/2}(\hat{\theta} - \theta)$ is

$$
\begin{aligned}
\delta F_1(t_1, \infty) + 1 - \delta &= \Phi\{(C^{11})^{-1/2}t_1\} - \Phi\{(C^{11})^{-1/2}a\} + 1 - \delta \\
&= \Phi\{(C^{11})^{-1/2}t_1\}, \quad (9)
\end{aligned}
$$

for $t_1 > -a$. In other words, the approximation provided by Theorem 1 is a mixture of a point mass at the origin and a normal distribution. Moreover, the weighting in this mixture associated with the point mass is 0.5 when $a = 0$ and decreases as a increases, until, with $a \gg 3\sqrt{C^{11}}$, we essentially obtain the usual normal approximation. The approximation (9) is used as a basis for inference in Section 4.

3.4. Marginal Distribution When a Nuisance Parameter is Near the Boundary.

It is not surprising that the marginal asymptotic distribution of the estimator of the parameter of interest is affected by the proximity of the parameter of interest to the boundary. However, it is not as widely appreciated that the marginal asymptotic distribution of the estimator of the parameter of interest is also affected by the proximity of nuisance parameters to the boundary.

As an example of this phenomenon, suppose that $p = q = 1$, that $\theta = \phi_2$ is the parameter of interest and the nuisance parameter $\tau = \phi_1$ is near the boundary. For the first component of the limiting mixture distribution, we have

$$\delta F_1(\infty, t_2) = \frac{1}{\sqrt{2\pi C^{22}}^{1/2}} \int_{-\infty}^{t_2} \exp\left(\frac{z_2^2}{2C^{22}}\right) [1 - \Phi\{(C_{11})^{1/2}(C_{12}z_2/C_{11} - a)\}]dz_2.$$

8

and letting $a \to \infty$ yields $z_2 \sim N(0, C^{22})$, the usual Gaussian marginal distribution. For the second component of the limiting mixture distribution, we have

$$
\begin{aligned}
(1-\delta)f_2(z_2) \quad = \quad & \frac{(1-\delta)(C_{22})^{1/2}}{(2\pi)^{1/2}\rho} \exp\left\{ -\frac{C_{22}}{2C^{22}}(z_2 - aC_{12}/C_{22})^2 \right\} \\
& \times \Phi\left[\left\{ \frac{C^{12}C_{22}}{\sqrt{C^{11}}} + C_{12}\sqrt{C^{11}} \right\}(z_2 - aC_{12}/C_{22}) - \frac{a}{\sqrt{C^{11}}} \right].
\end{aligned}
$$

So, integrating over $-\infty < z_2 \le t_2$ yields the distribution function of the second component of the limiting mixture distribution.

This limiting distribution is quite awkward to use. Moreover, in the general multivariate case, the marginal distribution of the estimator of the parameter of interest depends on the number of nuisance parameters on or near the boundary in a very complicated way. Specifically, if p components allowed to be near the boundary, any subset of the p estimators of these components can equal zero, so there are 2^p components in the limiting mixture distribution.

3.5. Asymptotic Distribution of the Likelihood Ratio Statistic.

We again write $\sigma^2 = (\theta, \tau)$ with θ a single variance component. Let $m(\theta) = \ell\{\theta, \hat{\tau}(\theta)\}$ denote either the profile log-likelihood, in which case $\hat{\tau}(\theta) = \tilde{\tau}(\theta)$, or the REML log-likelihood, in which case $\hat{\tau}(\theta) = \bar{\tau}(\theta)$. Also, let $\breve{\theta}$ solve the score equation $m'(\theta) = 0$. Thus $\hat{\theta}$ maximises $m(\theta)$ over the parameter space while $\breve{\theta}$ maximises $m(\theta)$ over $\mathbb{R}$, even if $\breve{\theta}$ turns out to be negative. Then $\eta n^{1/2}(\breve{\theta} - \theta)$ is asymptotically standard normal, where $\eta^2 = -\lim_{n\to\infty} n^{-1}E\{m''(\theta)\}$. Note that η accounts for the variation associated with estimation of the nuisance parameters, as it is defined in terms of the profiled objective function. Standard results show $\eta = (C^{11})^{-1/2}$, with C^{11} as defined in Theorem 1.

THEOREM 2. *Let $\theta = n^{-1/2}a, a \in [0, a_0)$ with the remaining components interior to the parameter space. Also let $\eta^2 = -\lim_{n\to\infty} n^{-1}E\{m''(\theta)\}$ and $W(\theta) = 2\{m(\hat{\theta}) - m(\theta)\}$. Then for all $t > 0$,*

$$
P\{W(\theta) \le t\} \approx \begin{cases} \Phi(t^{1/2}), & \theta = 0 \\ \Phi(t^{1/2}) - \Phi(-t^{1/2})1_{\{t < \eta^2 a^2\}} - \Phi(-\frac{t}{2\eta a} - \frac{\eta a}{2})1_{\{t \ge \eta^2 a^2\}}, & \theta > 0. \end{cases}
$$

We provide a heuristic outline of the proof. Expanding $m(\theta)$ and $m(\hat{\theta})$ about $\breve{\theta}$ yields $m(\theta) \approx m(\breve{\theta}) - \frac{1}{2}\eta^2 n(\breve{\theta} - \theta)^2$ and $m(\hat{\theta}) \approx m(\breve{\theta}) - \frac{1}{2}\eta^2 n(\breve{\theta} - \hat{\theta})^2$, so we can write the likelihood ratio statistic as

$$
W(\theta) = 2\{m(\hat{\theta}) - m(\theta)\} \approx \eta^2 n\{(\breve{\theta} - \theta)^2 - (\breve{\theta} - \hat{\theta})^2\} = \begin{cases} \eta^2 n(\breve{\theta} - \theta)^2 & \text{if } \hat{\theta} > 0 \\ \eta^2 n\theta^2 - 2\eta^2 n\theta\breve{\theta} & \text{if } \hat{\theta} = 0. \end{cases}
$$

Next note that

$$
\begin{aligned}
Pr(\breve{\theta} \le t | \breve{\theta} > 0) \quad &= \quad P(\breve{\theta} \le t | \hat{\theta} > 0) = \delta^{-1}[\Phi\{\eta n^{1/2}(t - \theta)\} - (1 - \delta)], \text{ for } t > 0, \\
P(\breve{\theta} \ge t | \breve{\theta} \le 0) \quad &= \quad P(\breve{\theta} \ge t | \hat{\theta} = 0) = 1 - (1 - \delta)^{-1}\Phi\{\eta n^{1/2}(t - \theta)\}, \text{ for } t \le 0.
\end{aligned}
$$

Hence, recalling $\delta = P(\hat{\theta} > 0)$, we have $P\{W(0) \le t\} = \Phi(t^{1/2})$, while for $\theta > 0$,

$$
P\{W(\theta) \le t\} \quad \approx \quad \delta P\{\eta^2 n(\breve{\theta} - \theta)^2 \le t | \hat{\theta} > 0\}
$$

9

$$+(1-\delta)P\{\eta^2 n\theta^2 - 2\eta^2 n\theta\breve{\theta} \le t|\hat{\theta}=0\}$$
$$\approx \Phi(t^{1/2}) - (1-\delta) - \{\Phi(-t^{1/2}) - (1-\delta)\}1_{\{t<\eta^2 a^2\}}$$
$$+\left\{1-\delta-\Phi\left(-\frac{t}{2\eta a}-\frac{\eta a}{2}\right)\right\}1_{\{t\ge\eta^2 a^2\}}$$
$$= \Phi(t^{1/2}) - \Phi(-t^{1/2})1_{\{t<\eta^2 a^2\}} - \Phi\left(-\frac{t}{2\eta a}-\frac{\eta a}{2}\right)1_{\{t\ge\eta^2 a^2\}}.$$

REMARKS.

1. The cumulative distribution function of $W(\theta)$ is continuous even though the distribution of $\hat{\theta}$ has a jump. To see this, note that when $t = \eta^2 a^2$, we have $\Phi(-t^{1/2}) = \Phi(-\eta a)$ and $\Phi(-\frac{t}{2\eta a} - \frac{\eta a}{2}) = \Phi(-\frac{\eta a}{2} - \frac{\eta a}{2}) = \Phi(-\eta a)$.

2. The test of $H_0 : \theta = 0$ based on the local asymptotic approximation has higher power for local alternatives than the test based on the chi-squared approximation. If c is the $(1-\alpha)$-quantile of a χ_1^2 distribution, then $\Phi(c^{1/2}) - \Phi(-c^{1/2}) = 1 - \alpha$ which implies that $\Phi(c^{1/2}) = 1 - \alpha/2$. On the other hand, using the asymptotic distribution from Theorem 2, the critical value, c', satisfies $P(c') = Pr\{W(0) \le c'\} = 1 - \alpha$. Since $P(\cdot)$ is monotonically increasing, $P(c) = \Phi(c^{1/2}) = 1 - \frac{\alpha}{2} > 1 - \alpha = P(c')$ implies that $c' < c$ and hence $Pr_a\{W(0) \ge c\} < Pr_a\{W(0) \ge c'\}$, where Pr_a denotes probabilities calculated under a local alternative, $\theta = n^{-1/2}a$, $a > 0$.

3. Theorem 2 continues to hold for any adjusted profile log-likelihood function $m(\theta) = L_p\{\theta, \hat{\tau}(\theta)\} + b(\theta)$, provided the adjustment function $b(\theta)$ has derivatives which are of order $O_p(1)$. This condition is satisfied for all of the adjusted profile log-likelihoods mentioned in Section 2.

4. CONFIDENCE INTERVALS FOR NEAR BOUNDARY PARAMETERS

The simplest approximate $100(1 - \alpha)\%$ confidence interval for a variance component is the interval (4) based on the normal approximation. We now use the approximations derived in Section 3 to construct alternative confidence intervals.

4.1. Central Confidence Intervals.

The first method we consider is based on the estimators themselves and we assume that only the component of interest can be near the boundary. This is not an assumption we would be prepared to make in general, but it is reasonable for the two-component model and helps to show how the local asymptotic approximations generate alternatives to (4).

Suppose initially that we treat $C^{11} \approx nI^{11}$ as known. We see that $n^{1/2}(\hat{\theta} - \theta)$ is not a pivotal quantity because its distribution depends on $a = n^{1/2}\theta$. However, we can still construct a confidence interval by inverting appropriate probability statements. Doing so, we obtain the confidence interval $(\hat{\theta}_L, \hat{\theta}_U)$, where $\hat{\theta}_U = \hat{\theta} + (I^{11})^{1/2}\Phi^{-1}(1 - \alpha/2)$ and

$$\hat{\theta}_L = \begin{cases} 0 & \text{for } \hat{\theta} \le (I^{11})^{1/2}\Phi^{-1}(1-\alpha) \\ \hat{\theta} - (I^{11})^{1/2}\Phi^{-1}(1-\alpha/2) & \text{for } \hat{\theta} > (I^{11})^{1/2}\{\Phi^{-1}(1-\alpha) - \Phi^{-1}(\alpha/2)\} \\ \hat{\theta} - (I^{11})^{1/2}\Phi^{-1}(1-\alpha) & \text{otherwise.} \end{cases} \quad (10)$$

Of course, I is not known so we replace it by $\hat{I}$ to obtain a usable interval. We recover the approximate confidence interval (4) when $\hat{\theta}$ is sufficiently large.

10

4.2. Wald-test Confidence Intervals.

Implementing interval (10) by replacing the unknown σ^2 in I^{11} by $\hat{\sigma}^2 = (\hat{\theta}, \hat{\tau})$ ignores the fact that the quantity $\sqrt{n}(\hat{\theta} - \theta)$ is not truly pivotal. In particular, its standard error depends on both θ and τ. Therefore, we may be able to obtain better intervals by "studentising" the quantity $\sqrt{n}(\hat{\theta} - \theta)$ using a more appropriate estimate of its standard error. One such approach is to obtain a confidence interval by inverting an appropriate Wald test.

Formally, an approximate $100(1 - \alpha)\%$ Wald-test confidence interval is

$$\{\theta \geq 0 : n\hat{\eta}^2(\hat{\theta} - \theta)^2 \leq c_\alpha(\theta)\}, \tag{11}$$

where $c_\alpha(\theta)$ is an approximate $(1 - \alpha)$-quantile of the random quantity $n\hat{\eta}^2(\hat{\theta} - \theta)^2$ and $\hat{\eta}^{-2} = nI^{11}\{\theta, \hat{\tau}(\theta)\}$ is an estimate of the variance of $\sqrt{n}\hat{\theta}$ which is allowed to depend on the value of θ. Typical choices for $\hat{\tau}(\theta)$, the estimate of the nuisance parameters, would be $\tilde{\tau}(\theta)$ or $\bar{\tau}(\theta)$. The critical value, $c_\alpha(\theta)$, can be obtained from the relationship: $P\{n\hat{\eta}^2(\hat{\theta} - \theta)^2 \leq t\} \approx \Phi(t^{1/2}) - \Phi(-t^{1/2})1_{\{t \leq n\hat{\eta}^2\theta^2\}}$. Note that the cumulative distribution function of the Wald-type test statistic in not continuous at $t = n^{1/2}\hat{\eta}\theta$, unlike the cumulative distribution function of $W(\theta)$.

To avoid the implicit truncation at zero in (11), we can construct intervals on the logarithmic scale and exponentiate the resultant endpoints. To construct Wald-test intervals similar to (11) for $\psi = \log\theta$, all that is required is an appropriate change in the Fisher information to account for the reparameterisation. However, in the present case, such an approach is complicated by the fact that the distribution of the estimator $\hat{\psi} = \log\hat{\theta}$ has a point mass at negative infinity (which means that such a transformation approach cannot by used to modify the interval of Section 4.1). Moreover, our focus on local asymptotics means that we must let ψ approach negative infinity, and it is not clear what the most appropriate rate for this convergence is. Nonetheless, this approach can be implemented, and the results (not reported) are no better, and often worse, than those for interval (11).

4.3. Likelihood-based Confidence Intervals.

An approximate $100(1 - \alpha)\%$ likelihood-based confidence interval is

$$\{\theta \geq 0 : W(\theta) \leq c_\alpha(\theta)\}, \tag{12}$$

where $c_\alpha(\theta)$ is an approximate $(1 - \alpha)$-quantile of the distribution of $W(\theta)$, and can be obtained using the approximation of Theorem 2.

An alternative interval is the set:

$$\{\theta \geq 0 : w(\theta) \leq \chi_1^2(1 - \alpha)\}, \tag{13}$$

where $\chi_1^2(1 - \alpha)$ is the $(1 - \alpha)$-quantile of the chi-squared distribution with one degree of freedom and $w(\theta) = 2\{m(\breve{\theta}) - m(\theta)\}$ is the likelihood ratio statistic calculated without restricting the estimator $\breve{\theta}$ to be positive. Such intervals were discussed in general by Feng & McCulloch (1992). The increased "regularity" of $w(\theta)$ may lead to better small sample coverage accuracy. One drawback of this method, however, is its lack of intuitive appeal, since it allows negative values for $\breve{\theta}$ (though negative values are not allowed in the confidence interval itself). This fact leads us to suspect that properties other than coverage accuracy (e.g., "shape") of such intervals may be less than admirable. Moreover, (13) may yield the null set.

5. SIMULATION RESULTS

A simulation study to explore the finite sample properties of the confidence intervals presented in the previous section was carried out. For simplicity, all simulations were carried out in the context of the one-way classification model, $y = 1_n\alpha + Zu + e$, described in Section 2.5. As the notation there suggests, $\theta = \sigma_1^2$, the variance of the elements of u, is the parameter of interest. None of the intervals performed particularly well, though it was generally the case that the coverage accuracy was best for the likelihood-based intervals (12) and (13) and worst for the central intervals (4) and (10). In addition, the coverage of intervals based on the adjusted profile log-likelihoods tended to be more accurate than those based on the usual profile log-likelihood, though not exclusively so. Among the adjusted profile log-likelihoods discussed in Section 2, the profiled REML log-likelihood was generally the most accurate and so we present results only for this adjustment of the profile log-likelihood here. Finally, the coverage of intervals based on the local asymptotics tended to be more accurate than those based on the standard asymptotics, though again not exclusively so.

TABLE 1: Coverage accuracy of likelihood-based intervals (100,000 simulations).

No. of Groups	Obs./ Group	(θ, τ)	Nominal Coverage	Standard Asymptotics ML	REML	Local Asymptotics ML	REML
5	10	(0.1,1)	90%	92.1	96.0	85.6	91.4
			95%	98.9	98.5	95.9	97.2
			99%	99.9	99.7	99.7	99.5
	10[a]	(0.1,1)	90%	92.3	96.1	85.2	91.7
			95%	99.0	98.5	96.0	97.1
			99%	99.9	99.7	99.7	99.4
	10[b]	(0.1,1)	90%	92.4	96.4	84.1	92.4
			95%	99.0	98.4	96.3	97.0
			99%	99.9	99.7	99.7	99.4
	20	(0.1,1)	90%	83.1	90.2	81.1	87.4
			95%	92.7	97.4	89.2	94.3
			99%	99.9	99.7	99.6	99.4
10	20	(0.1,1)	90%	86.6	89.3	86.5	89.1
			95%	92.6	94.5	92.2	94.0
			99%	98.6	99.3	97.8	98.7
	20	(0.1,3)	90%	91.3	93.2	87.6	89.8
			95%	98.1	98.0	95.1	95.9
			99%	99.8	99.7	99.6	99.3
	20[c]	(0.1,3)	90%	91.1	93.5	87.1	90.0
			95%	98.1	98.1	95.1	96.0
			99%	99.8	99.6	99.6	99.3
200	200	(0.1,1)	90%	89.8	90.0	89.8	90.0
			95%	94.9	95.0	94.9	95.0
			99%	99.0	99.0	99.0	99.0

[a] Group Sizes = (6,6,10,14,14); [b] Group Sizes = (2,6,10,14,18);
[c] Group Sizes = (10,10,15,15,20,20,25,25,30,30).

Some selected results on the coverage accuracy of the likelihood-based confidence intervals (12) are presented in Table 1, with the likelihood ratio statistic,

$W(\theta)$, based on either the unadjusted profile log-likelihood (ML) or the profiled REML log-likelihood (REML). The intervals based on local asymptotics tend to have lower coverage than the corresponding intervals based on the standard asymptotics, indicating that the local asymptotic approximation generally will not improve the coverage accuracy if the standard approximation produces undercoverage. Typically, however, in cases with small variance components, methods based on the standard asymptotics are seen to overcover, so that local asymptotic methods offer an improvement in coverage accuracy. For instance, at the nominal 90% level the ML intervals typically perform quite well with the standard asymptotics. On the other hand, both the likelihood and REML intervals are typically conservative at the nominal 95% and 99% levels so the coverage is improved by using the local asymptotic approximation. Generally, however, the differences in performance are not substantial. Furthermore, we note that results (not reported here) on the coverage accuracy of interval (13) based on the profiled REML log-likelihood show that its coverage accuracy is typically slightly superior to the more intuitively appealing intervals derived from (12). However, interval (13) does not perform substantially better, and does occasionally perform worse, than interval (12). Moreover, for the smaller sample sizes, (13) led to null intervals in an appreciable number of samples. In particular, for the balanced case of $c_1 = 5$ and $m = 10$ with $(\theta, \tau) = (0.1, 1)$, there were about 3% of samples which led to empty 95% confidence intervals, and nearly 5% of samples which led to empty 90% confidence intervals.

TABLE 2: Upper and lower non-coverage probability of central intervals
(100,000 simulations).

No. of Groups	Obs./ Group	(θ, τ)	Nom. Cov.	Tail	Standard Asymptotics		Local Asymptotics	
					ML	REML	ML	REML
5	10	(0.1,1)	90%	Upper	13.0	4.8	13.0	4.8
				Lower	0.0	0.0	0.0	0.0
			95%	Upper	4.1	1.2	4.1	1.2
				Lower	0.0	0.0	0.0	0.0
	10^a	(0.1,1)	90%	Upper	19.1	5.8	19.1	5.8
				Lower	0.0	0.0	0.0	0.0
			95%	Upper	7.1	1.5	7.1	1.5
				Lower	0.0	0.0	0.0	0.0
10	20	(0.1,1)	90%	Upper	23.4	17.0	23.4	17.0
				Lower	0.0	0.0	0.6	0.7
			95%	Upper	19.2	13.7	19.2	13.7
				Lower	0.0	0.0	0.0	0.0
	20^b	(0.1,3)	90%	Upper	7.3	1.9	7.3	1.9
				Lower	0.0	0.0	0.6	0.8
			95%	Upper	0.3	0.0	0.3	0.0
				Lower	0.0	0.0	0.0	0.0
200	200	(0.1,1)	90%	Upper	8.0	7.3	8.0	7.3
				Lower	2.7	2.7	3.0	3.0
			95%	Upper	4.9	4.4	4.9	4.4
				Lower	1.0	1.1	1.0	1.1

[a] Group Sizes = (6,6,10,14,14), [b] Group Sizes = (10,10,15,15,20,20,25,25,30,30).

The upper and lower tail coverage accuracy of central confidence intervals is presented in Table 2. The standard asymptotic approximation leads to the use of

the usual interval (4) while the local asymptotic approximation leads to the interval (10). These intervals exhibit extremely poor, and decidedly asymmetric, coverage accuracy, even for the "large sample" case of $c_1 = 200$ and $m = 200$. Moreover, the use of local asymptotic approximations has little impact on the coverage accuracy. This outcome is not altogether surprising. When the probability associated with point mass at the origin in the local asymptotic approximation exceeds 0.025, symmetric non-coverage is impossible and the only non-coverage occurs in the upper tail. In general, extremely large samples are needed before the asymptotics apply for these intervals. This is largely due to treating the standard error as constant in θ. Confidence intervals based on the Wald test use essentially the same construction but treat the standard error as a function of θ and, although we do not present the results here, generally perform nearly as well as the likelihood-based intervals presented in Table 1. The general conclusion from Table 2 is that straightforward application of asymptotics associated with parameter estimators is not a very reliable or accurate method for constructing confidence intervals.

Finally, the power of likelihood ratio and REML tests against the null hypothesis that $\theta = 0$ using the two asymptotic approximations is explored in Table 3.

TABLE 3: Power against $H_0 : \theta = 0$ of likelihood-based tests (100,000 simulations).

No. of Groups	Obs./ Group	(θ, τ)	Nominal Coverage	Standard Asymptotics		Local Asymptotics	
				ML	REML	ML	REML
5	10	(0.1,1)	90%	17.9	23.9	26.3	33.9
			95%	12.1	16.6	17.9	23.9
			99%	4.7	7.0	7.2	10.3
	10^a	(0.1,1)	90%	17.6	23.6	25.5	33.3
			95%	12.0	16.6	17.6	23.6
			99%	4.9	7.2	7.2	10.3
	10^b	(0.1,1)	90%	17.4	23.9	25.1	33.2
			95%	12.1	17.0	17.4	23.9
			99%	5.2	7.6	7.5	10.7
	20	(0.1,1)	90%	38.7	46.0	48.4	56.1
			95%	30.9	37.5	38.7	46.0
			99%	17.9	22.7	22.7	28.3
10	20	(0.1,1)	90%	68.0	73.8	76.5	80.7
			95%	59.9	64.9	68.0	72.8
			99%	42.8	47.9	49.7	55.0
	20	(0.1,3)	90%	23.2	28.5	33.4	39.7
			95%	16.0	20.1	23.2	28.5
			99%	6.6	8.8	9.7	12.7
	20^c	(0.1,3)	90%	23.3	28.8	33.3	39.8
			95%	16.3	20.6	23.3	28.8
			99%	6.8	9.1	9.9	13.0

[a] Group Sizes = (6,6,10,14,14), [b] Group Sizes = (2,6,10,14,18),
[c] Group Sizes = (10,10,15,15,20,20,25,25,30,30)

As noted in the remarks following Theorem 2, the tests based on the local asymptotic approximation have higher power than their analogues based on the standard asymptotic approximation. Also, the REML tests have higher power than the

likelihood ratio tests, so that the REML test based on the local asymptotic approximation has the highest power in these comparisons. The power is related to the length of the confidence intervals presented in Table 1 with higher power being associated with shorter intervals. Of course, there are other aspects of the interval methods which affect the power, but Table 3 does provide an indirect exploration of the lengths of the intervals.

The above results show that further research on the problem of setting confidence intervals for small components of variance is needed. In the meanwhile, our tentative suggestion is that intervals be constructed from the REML version of the likelihood ratio test with the local asymptotic approximation to its sampling distribution, as this method seems to have the best overall combination of coverage and power properties.

REFERENCES

O. E. Barndorff-Nielsen (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70, 343–365.

O. E. Barndorff-Nielsen (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society Series B*, 56, 125–140.

O. E. Barndorff-Nielsen & S. R. Chamberlin (1992). Stable and invariant adjusted likelihood roots. *Technical Report.* Department of Theoretical Statistics, Aarhus University, Aarhus, Denmark.

R. K. Burdick & F. A. Graybill (1988). The present status of confidence interval estimation on variance components in balanced and unbalanced random models. *Communications in Statistics: Theory and Methods* (Special Issue on the Analysis of the Unbalanced Mixed Model), 17, 1165–1195.

H. Chernoff (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25, 573–578.

D. R. Cox & N. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society Series B*, 49, 1–39.

N. Cressie & S. N. Lahiri (1993). The asymptotic distribution of REML estimators. *Journal of Multivariate Analysis*, 45, 217–233.

T. J. DiCiccio, M. A. Martin, S. E. Stern & G. A. Young (1996). Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society Series B*, 58, 198–203.

T. J. DiCiccio & S. E. Stern (1993). An adjustment to profile likelihood based on observed information. *Technical Report.* Department of Statistics, Stanford University, Stanford, CA.

Z. Feng & C. E. McCulloch (1992). Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space. *Statistics and Probability Letters*, 13, 325–332.

D. Firth (1993). Bias reduction of maximum likelihood estimators. *Biometrika*, 80, 27–38.

J. K. Ghosh & R. Mukerjee (1994). Adjusted *versus* conditional likelihood: power properties and Bartlett-type adjustments. *Journal of the Royal Statistical Society Series B*, 56, 185–188.

V. P. Godambe (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31, 1208–1211.

M. G. Kenward & J. H. Roger (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53, 983–997.

A. Y. C. Kuk (1995). Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society Series B*, 57, 395–407.

B. Levin & F. Kong (1990). Bartlett's bias correction to the profile score function is a saddlepoint correction. *Biometrika*, 77, 219–221.

K.-Y. Liang (1987). Estimating functions and approximate conditional likelihood. *Biometrika*, 74, 695–702.

P. McCullagh & R. J. Tibshirani (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society Series B*, 52, 325–344.

P. A. P. Moran (1971). Maximum likelihood estimation in non-standard conditions. *Proceedings of the Cambridge Philosophical Society*, 70, 441–450.

H. D. Patterson & R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545–554.

H. D. Patterson & R. Thompson (1974). Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometric Conference*, 197–207.

A. M. Richardson & A. H. Welsh (1994). Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *The Australian Journal of Statistics*, 36, 31–43.

A. M. Richardson & A. H. Welsh (1996). Covariate screening in mixed linear models. *Journal of Multivariate Analysis*, 58, 27–54.

S. R. Searle, G. Casella & C. E. McCulloch (1992). *Variance Components*. Wiley, New York.

S. G. Self & K.-Y. Liang (1987). Asymptotic properties of maximum likelihood ratio tests under non-standard conditions. *Journal of the American Statistical Association*, 82, 605–610.

S. E. Stern (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *Journal of the Royal Statistical Society Series B*, 59, 653–665.

S. J. Welham & R. Thompson (1997). A likelihood ratio test for fixed model terms using residual maximum likelihood. *Journal of the Royal Statistical Society Series B*, 59, 701–714.

J. S. Williams (1962). A confidence interval for variance components. *Biometrika*, 49, 278–281.

Steven E. STERN
Steven.Stern@anu.edu.au
Department of Statistics and Econometrics
The Australian National University
Canberra, ACT
Australia, 0200

A. H. WELSH
Alan.Welsh@anu.edu.au
Centre for Mathematics and its Applications
The Australian National University
Canberra, ACT
Australia, 0200