Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture 2011 - 23rd Annual Conference Proceedings

CONFIDENCE INTERVALS FOR VARIANCE COMPONENTS USING NON-NORMAL DISTRIBUTIONS

Brent D. Burch

Follow this and additional works at: https://newprairiepress.org/agstatconference

 Part of the Agriculture Commons, and the Applied Statistics Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation

Burch, Brent D. (2011). "CONFIDENCE INTERVALS FOR VARIANCE COMPONENTS USING NON-NORMAL DISTRIBUTIONS," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1045

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

CONFIDENCE INTERVALS FOR VARIANCE COMPONENTS USING NON-NORMAL DISTRIBUTIONS

Brent D. Burch
Department of Mathematics and Statistics
Northern Arizona University
Flagstaff, Arizona 86011-5717
email: brent.burch@nau.edu

Abstract

Simulation studies are conducted to evaluate the performance of confidence intervals for variance components under non-normal distribution assumptions. Confidence intervals based on the pivotal quantity (PQ) method and the large-sample properties of the restricted maximum likelihood (REML) estimator are considered. Of particular interest is the actual coverage value of nominal 95% confidence intervals for a ratio of variance components. In the context of unbalanced one-way random effects models, simulation results and an empirical example involving arsenic concentrations in oyster tissue suggest that the REML-based confidence interval is preferred.

1. Introduction

Confidence intervals provide information about plausible values of a parameter under study. If the parameter is a variance component, then an interval estimate of the variability associated with an underlying source of variation can be obtained. Furthermore, if measurements depend on multiple sources of variation, then a confidence interval for a function of variance components can be used to quantify the impact of a particular source of variation on the measurements. Specifically, if σ_1^2 and σ_2^2 denote variances associated with two sources of variation, then $\theta = \sigma_1^2/\sigma_2^2$ or $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$ serve as ways to measure the importance of one source compared to the other source.

The vast majority of the statistical literature on confidence intervals for variance components presumes that the sources of variation are normally distributed. For instance, the confidence intervals developed by Wald (1940), Thomas and Hultquist (1978), Harville and Fenech (1985), and Burch and Iyer (1997) depend on normal distribution theory. In addition, the commonly used method to construct confidence intervals for variance components under normal theory relies on the pivotal quantity (PQ) approach. More recently, Jiang (1996, 2005) described the large-sample properties of restricted maximum likelihood (REML) estimators of θ or ρ. Using this approach, one can build asymptotic confidence intervals for functions of variance components.

The objective of this paper is to compare the PQ and REML generated confidence intervals in terms of their realized coverage probabilities. Burch (2011a) presented confidence interval results for θ or ρ for the balanced one-way random effects model without assuming normality. The work of Burch (2011a) was extended to confidence intervals for variance components in non-normal unbalanced one-way random effects models. See Burch (2011b)

for details. The current paper focuses on the PQ and REML methods to build confidence intervals, as described by Burch (2011b), and provides further details about the application involving arsenic in oyster tissue.

This paper is organized as follows. Section 2 provides background information about the arsenic concentrations in oyster tissue application. Section 3 presents an overview of the unbalanced one-way random effects model. Properties of the quadratic forms used to estimate variance components are stated without relying on normal distribution assumptions. Descriptions of the PQ and REML procedures used to construct confidence intervals for θ or ρ are also given. Section 4 provides confidence interval simulation results for a variety of scenarios and distributions. Also included are confidence interval results for the arsenic in oyster tissue application. Summary comments are given in Section 5.

2. Arsenic Concentration in Oyster Tissue

The National Oceanic and Atmospheric Administration (NOAA) National Status and Trends Program conducted ongoing comparisons of laboratory measurements of trace metals in marine sediments and biological tissues. The original purpose of these annual exercises was to assess the capabilities of laboratories to analyze the sediments and tissues for trace metals. In the eighth round of these interlaboratory comparisons, Willie and Berman (1995) reported that a total of 56 laboratories received materials to be analyzed. The materials included freeze dried marine sediment collected from Nova Scotia, freeze dried Pacific oyster tissue, and reference materials known as NRC CRM BCSS-1 and NIST SRM 1566a. The data considered in this paper were collected from the National Institute of Standards and Technology (NIST) Standard Reference Material (SRM) 1566a.

Altogether, 13 NIST SRM 1566a elements including arsenic, mercury, and lead were under investigation. Each laboratory was requested to perform four replicate measurements for each element. Not all laboratories provided results for all of the elements. In fact, 46 sets of results were received. In addition, six laboratories did not submit data for marine sediments and four laboratories did not submit data for biological tissues. In this paper we considered the NIST SRM 1566a arsenic concentration measurements.

Samples of the oyster material originated from a commercial source that had been ground and freeze dried. The oyster tissue material was then processed at NIST and bottled. 31 laboratories measured the arsenic concentrations (mg/kg) in the oyster tissue samples. Out of the 31 laboratories, 28 laboratories made four replicate arsenic concentration measurements, one laboratory made two replicate arsenic concentration measurements, and two laboratories made one arsenic concentration measurement. Figure 1 displays side-by-side dotplots of arsenic measurements the for the 31 laboratories.

With continued exposure over time, there is concern about arsenic's effect on chromosomes and increased risk of cancer. To help protect consumers, the FDA and EPA provide safety guidelines in regards to fish and fishery products. The published level represents the concentration of a toxic element at which point the agency will take legal action to remove the product from the market. For oysters, the level is 68 ppm (or 68 mg/kg) of arsenic. Table 1 displays the analysis of variance table for the arsenic concentration measurements.

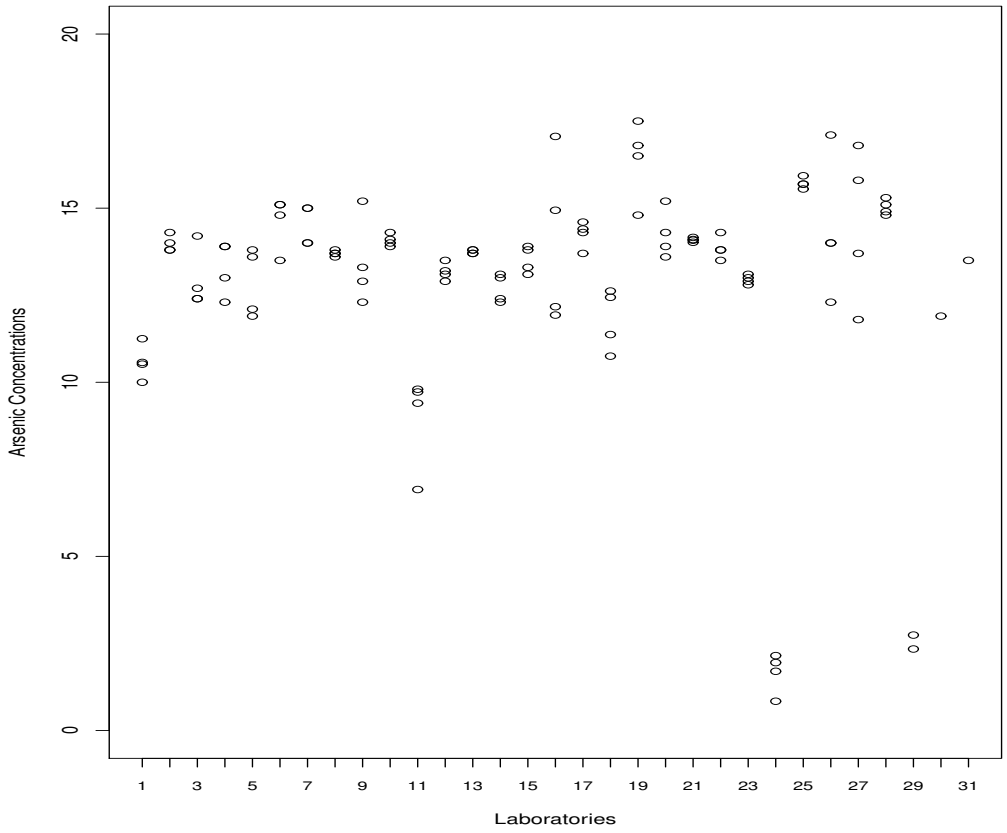

Figure 1: Dotplots of the arsenic concentration measurements for each of the 31 laboratories.

If σ_1^2 denotes the variance between laboratories and σ_2^2 denotes the variance within laboratories, then $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$ is the proportion of variation in arsenic concentrations attributed to the laboratories. An unbalanced one-way random effects model is used for this data and confidence intervals for θ or ρ are of primary concern. In the next Section we discuss the details underlying the unbalanced one-way random effects model.

3. Notation and the Unbalanced One-Way Random Effects Models

Let Y_{ij} denote the j^{th} arsenic concentration measurement from the i^{th} laboratory, where in general $i = 1, ..., a$, $j = 1, ..., b_i$, and the total sample size is $n = \sum_{i=1}^{a} b_i$. Since the b_i's are not all equal, the model is unbalanced. The corresponding one-way random effects model is

$$Y_{ij} = \mu + A_i + e_{ij}, \tag{1}$$

where the a laboratories in the study represent a random sample of laboratories from a population of laboratories. Furthermore, from within the i^{th} laboratory, a random sample of size b_i is selected. The random effect A_i in model (1) represents the randomness in the measurements due to the different laboratories and the e_{ij}'s are often referred to as random errors and may be interpreted as the deviations of the measurements within laboratories. μ

Table 1: ANOVA table for arsenic concentrations in oyster tissues

Source of Variation	df	Sum of Squares	Mean Square
Between Labs	30	997.06	33.24
Within Labs	85	76.96	0.91
Total	115		

is the overall mean of the measurements and in this paper is a nuisance parameter since the parameter under study involves variances.

Since A_i and e_{ij} are random in model (1), it is assumed that $A_i \overset{iid}{\sim}(0, \sigma_1^2)$ and $e_{ij} \overset{iid}{\sim}(0, \sigma_2^2)$ where the A_i's and e_{ij}'s are mutually independent. However, suppose that the underlying distribution is unknown. With this framework, the measurements from the same laboratory are correlated since $Cov(Y_{ij}, Y_{ij'}) = \sigma_1^2$ for $j \neq j'$ and the measurements from different laboratories not correlated. In addition, $E(Y_{ij}) = \mu$, $Var(Y_{ij}) = \sigma_1^2 + \sigma_2^2$, and the third and fourth central moments of Y_{ij} are $E[(Y_{ij} - \mu)^3]$ and $E[(Y_{ij} - \mu)^4]$, respectively. The kurtosis of A_i and e_{ij} are defined as $\kappa_1 = E({A_i}^4)/\sigma_1^4 - 3$ and $\kappa_2 = E({e_{ij}}^4)/\sigma_2^4 - 3$, respectively. If, for example, one imposes normality assumptions on the A_i's and e_{ij}'s, then $E[(Y_{ij} - \mu)^3] = 0$, $E[(Y_{ij} - \mu)^4] = 3(\sigma_1^2 + \sigma_2^2)^2$, and $\kappa_1 = \kappa_2 = 0$.

Recall that the ratio of variance components is $\theta = \sigma_1^2/\sigma_2^2$ and that $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. ρ is often referred to as the intraclass correlation coefficient since it is the correlation between two measurements within the same class (or laboratory). By definition, $0 \leq \theta < \infty$, $0 \leq \rho < 1$, and there is a one-to-one correspondence between θ and ρ since $\rho = \theta/(\theta + 1)$. In an effort to construct confidence intervals for θ or ρ, we first review the derivation of and then describe the properties of important quadratic forms involving the Y_{ij}'s.

The derivation of the aforementioned quadratic forms was provided by Burch and Iyer (1997) and Burch and Harris (2001) in linear mixed models having two sources of variation. Burch (2011b) presented the properties of the quadratic forms. We now summarize the above work in the current paper. Let the $n \times 1$ vector $\mathbf{Y}$ denote the sample so that (1) can viewed in matrix notation as $\mathbf{Y} = \mathbf{1}\mu + \mathbf{ZA} + \mathbf{e}$, where $\mathbf{1}$ is an $n \times 1$ vector of ones, $\mathbf{Z}$ is an $n \times a$ matrix whose elements in the i^{th} column are ones for the b_i observations in laboratory i, and $\mathbf{e}$ is the $n \times 1$ error vector. Let $\mathbf{H}$ be an $n \times (n-1)$ matrix whose columns span the space orthogonal to the space spanned by the column vector of ones, and satisfies $\mathbf{H'H} = \mathbf{I}$ where $\mathbf{I}$ is an $(n-1) \times (n-1)$ identity matrix.

The $(n-1) \times 1$ random vector $\mathbf{H'Y}$ has mean vector zero and variance-covariance matrix $\sigma_2^2\mathbf{I} + \sigma_1^2\mathbf{H'ZZ'H}$ where $\mathbf{I}$ is an $(n-1) \times (n-1)$ identity matrix. Let $0 = \Delta_1 < ... < \Delta_d$ be the distinct eigenvalues of $\mathbf{H'ZZ'H}$ having multiplicities $r_1, ..., r_d$, respectively. There exists an $(n-1) \times (n-1)$ orthogonal matrix $\mathbf{P}$ such that $\mathbf{P'}(\mathbf{H'ZZ'H})\mathbf{P}$ is a diagonal matrix with entries $\Delta_1, .., \Delta_1, ..., \Delta_d, ..., \Delta_d$, where each Δ_k is repeated r_k times, $k = 1, ..., d$. Note that $\mathbf{P} = [\mathbf{P}_1, ..., \mathbf{P}_d]$ where $\mathbf{P}_k$ corresponding to Δ_k is of size $(n-1) \times r_k$, and consider the quadratic forms $Q_k = \mathbf{Y'}(\mathbf{HP}_k{\mathbf{P'}}_k\mathbf{H'})\mathbf{Y}$, $k = 1, ..., d$.

The vector of quadratic forms, denoted by $\mathbf{Q} = (Q_1, ..., Q_d)$, where $d < n$, form a set of minimal sufficient statistics associated with the model devoid of the parameter μ. These quadratic forms can be used to estimate functions of variance components. The number of quadratic forms and their corresponding distributions depend on the underlying model structure. For example, in the arsenic concentration in oyster tissue application, $n = 116$, $d = 5$, $\Delta_1 = 0.00, \Delta_2 = 1.00, \Delta_3 = 1.05, \Delta_4 = 2.04, \Delta_5 = 4.00$, and $r_1 = 85$, $r_2 = r_3 = r_4 = 1, r_5 = 27$.

For one-way random effects models in general, $\Delta_1 = 0$, $r_1 = n - a$ (if at least one $b_i > 1$) and the zero eigenvalue signifies that there is replication in the experiment (multiple observations per laboratory). An equation that relates the eigenvalues and their replications to the sample size and number of observations per laboratory is

$$\sum_{k=2}^{d} r_k \Delta_k = n - \frac{\sum_{i=1}^{a} b_i^2}{n}$$

so that

$$\overline{\Delta} = \frac{\sum_{k=2}^{d} r_k \Delta_k}{\sum_{k=2}^{d} r_k} = \frac{n - \sum_{i=1}^{a} b_i^2/n}{a-1}$$

is the average of the non-zero eigenvalues. In the arsenic concentration in oyster tissue application, $\overline{\Delta} = 3.74$. It is interesting to note that the sum of squares within laboratories $\sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \overline{Y}_{i.}\right)^2 = Q_1$ and the sum of squares between laboratories $\sum_{i=1}^{a} b_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 = Q_2 + ... + Q_d$, where $\overline{Y}_{i.}$ and $\overline{Y}_{..}$ denote the i^{th} laboratory mean and overall mean, respectively.

Hammersley (1949) showed that the expectations, variances, and covariances involving Q_1 and $Q_2 + ... + Q_d$ are

$$E(Q_1) = (n-a)\sigma_2^2,$$

$$Var(Q_1) = V_1 \sigma_2^4,$$

$$E(Q_2 + ... + Q_d) = (a-1)\sigma_2^2 + \left(n - \frac{\sum_{i=1}^{a} b_i^2}{n}\right)\sigma_1^2,$$

$$Var(Q_2 + ... + Q_d) = V_2 \sigma_1^4 + 4V_3 \sigma_1^2 \sigma_2^2 + V_4 \sigma_2^4,$$

and

$$Cov(Q_1, Q_2 + ... + Q_d) = V_5 \sigma_2^4 \kappa_2,$$

where

$$V_1 = 2(n-a) + \kappa_2 \left(\sum_{i=1}^{a} \frac{1}{b_i} + n - 2a\right),$$

$$V_2 = 2\left[\sum_{i=1}^{a} b_i^2 - 2\frac{\sum_{i=1}^{a} b_i^3}{n} + \frac{\left(\sum_{i=1}^{a} b_i^2\right)^2}{n^2}\right] + \kappa_1\left[\sum_{i=1}^{a} b_i^2 - 2\frac{\sum_{i=1}^{a} b_i^3}{n} + \frac{\sum_{i=1}^{a} b_i^4}{n^2}\right],$$

$$V_3 = n - \frac{\sum_{i=1}^{a} b_i^2}{n},$$

$$V_4 = 2(a-1) + \kappa_2\left(\sum_{i=1}^{a}\frac{1}{b_i} + \frac{1-2a}{n}\right),$$

and

$$V_5 = (a-1) + \frac{a}{n} - \sum_{i=1}^{a}\frac{1}{b_i}.$$

Also see Shoukri et al. (1990) and Singh et al. (2002) for more information.

The properties of the quadratic forms stated here do not depend on normal distribution theory. However, note that the variances and covariances do depend on κ_1 and κ_2 (the kurtosis of A_i and e_{ij}, respectively). Kurtosis measures the peakedness or tail weight of a density function compared to the normal density function. For illustrative purposes, Figure 2 displays that densities and kurtosis values for the Uniform, Normal, t(5), and $\chi^2(1)$ distributions where centering and scaling was performed so that the four distributions have the same mean and variance. For reference, the normal distribution has $\kappa = 0$. Distributions with $\kappa < 0$ are often referred to as platykurtic whereas distributions with $\kappa > 0$ are referred to as leptokurtic.

3.1 Confidence Intervals for θ or ρ: The PQ Method

Under normal theory, $Q_k \overset{ind}{\sim}(\sigma_2^2 + \Delta_k\sigma_1^2)\chi^2(r_k)$, $k = 1, ..., d$ and $\kappa_1 = \kappa_2 = 0$. It follows that $Var(Q_1) = 2(n-a)\sigma_2^4$, $Var(Q_2 + ... + Q_d) = 2\sum_{k=2}^{d}\left(\sigma_2^2 + \Delta_k\sigma_1^2\right)^2 r_k$ and $Cov(Q_1, Q_2 + ... + Q_d) = 0$. A well-known PQ for θ or ρ is

$$\frac{Q_1/(n-a)}{\sum_{k=2}^{d}\frac{Q_k}{1+\Delta_k\theta}/(a-1)} \sim F(n-a, a-1), \tag{2}$$

where $F(n-a, a-1)$ is the F-distribution with numerator and denominator degrees of freedom $n-a$ and $a-1$, respectively. While this approach yields exact confidence intervals, the endpoints of the interval are not available in closed-form since the PQ is not a simple function of the θ. However, by replacing $\Delta_2, ..., \Delta_d$ in (2) with $\overline{\Delta}$,

$$\frac{(1+\overline{\Delta}\theta)Q_1/(n-a)}{\sum_{k=2}^{d} Q_k/(a-1)} \overset{approx}{\sim} F(n-a, a-1),$$

and an approximate $100(1-\alpha)\%$ confidence interval for θ is obtained since

$$P\left(\frac{1}{\overline{\Delta}}\left[\frac{MSA}{MSE}F_{\alpha/2} - 1\right] \le \theta \le \frac{1}{\overline{\Delta}}\left[\frac{MSA}{MSE}F_{1-\alpha/2} - 1\right]\right) \approx 1-\alpha, \tag{3}$$

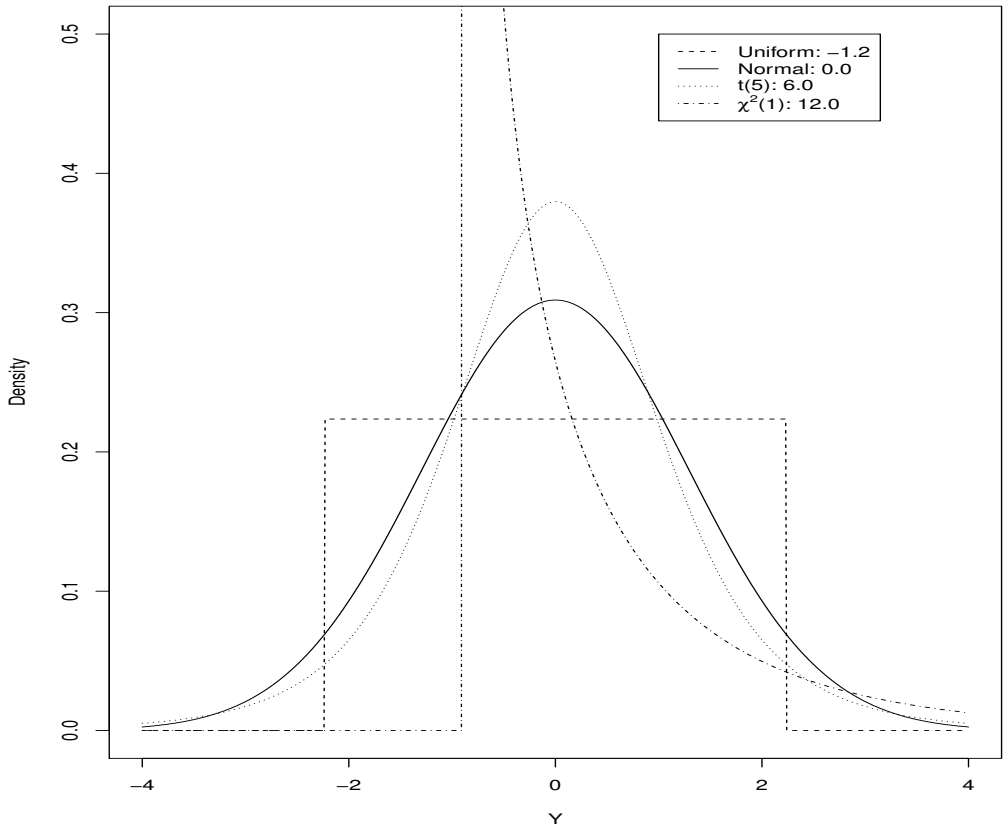

Figure 2: Density functions and kurtosis values for the Uniform, Normal, t(5), and $\chi^2(1)$ distributions all having a common mean and variance.

where $MSA = \sum_{k=2}^{d} Q_k/(a-1)$, $MSE = Q_1/(n-a)$, and $F_{\alpha/2}$ and $F_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the $F(n-a, a-1)$ distribution, respectively.

The interval in (3) was presented by Harville and Fenech (1985) and is similar to the method proposed by Thomas and Hultquist (1978). Note that if (L, U) denotes a confidence interval for θ, then the corresponding confidence interval for ρ is $(L/(L+1), U/(U+1))$.

3.2 Confidence Intervals for θ or ρ: The REML Method

In this Section we use the REML method to find a point estimator of θ. The large-sample properties of the REML estimator of θ are then used to construct an approximate confidence interval for θ. Burch and Harris (2001) developed closed-form approximations to REML estimators in linear mixed models having two variance components. See Burch (2011a) and Burch (2011b) for the development of REML-based confidence intervals for θ or ρ in balanced and unbalanced one-way random effects models, respectively. We will summarize the development of this approach in the following discussion.

The REML estimators of variance components depend on linear combinations of the Y_{ij}'s whose distribution is devoid of the nuisance parameter μ. Under normal theory, the

restricted log-likelihood function with respect to θ and σ_2^2 can be written as

$$L_{Restricted} \propto -\left[(n-1)ln\sigma_2^2 + \sum_{k=2}^{d} r_k ln(1+\Delta_k\theta)\right] - \frac{1}{\sigma_2^2}\left[Q_1 + \sum_{k=2}^{d}\frac{Q_k}{1+\Delta_k\theta}\right].$$

In unbalanced random effects models, $L_{Restricted}$ is not a simple function of the parameters and closed-form REML estimators of θ (and σ_2^2) are not available. An approximate restricted log-likelihood function is obtained by replacing Δ_k with $\overline{\Delta}$ for $k = 2, ..., d$. This simplification yields

$$L^*_{Restricted} \propto -\left[(n-1)ln\sigma_2^2 + (a-1)ln(1+\overline{\Delta}\theta)\right] - \frac{1}{\sigma_2^2}\left[Q_1 + \frac{1}{1+\overline{\Delta}\theta}\sum_{k=2}^{d} Q_k\right].$$

Differentiating $L^*_{Restricted}$ and maximizing over the parameter space yields an approximate REML estimator of θ, denoted by $\widehat{\theta}$, which is

$$\widehat{\theta} = \frac{1}{\overline{\Delta}}\left(\frac{MSA}{MSE} - 1\right), \text{ if } MSA \geq MSE$$

and zero otherwise. The corresponding approximate REML estimator of ρ is $\widehat{\theta}/(\widehat{\theta}+1)$.

Using $L^*_{Restricted}$ as a surrogate for $L_{Restricted}$ and assuming regularity conditions hold, Jiang (1996) established asymptotic normality of REML estimators in non-normal applications. Westfall (1987) and Burch and Harris (2001) showed that the regularity conditions for one-way random effects models relate to the number of laboratories, a, approaching infinity while the b_i's have a finite upper bound. Jiang (1996, 2005) also derived the asymptotic covariance matrix of REML estimators and combining these results we obtain

$$\widehat{\theta} \overset{asymp}{\sim} N\left(\theta, Var(\theta)\right) \tag{5}$$

where

$$Var(\widehat{\theta}) = \frac{(1+\overline{\Delta}\theta)^2}{\overline{\Delta}^2\sigma_2^4}\left(\frac{Var(SSE)}{(n-a)^2} + \frac{Var(SSA)}{(a-1)^2(1+\overline{\Delta}\theta)^2} - \frac{2Cov(SSA, SSE)}{(n-a)(a-1)(1+\overline{\Delta}\theta)}\right)$$

$$= \frac{(1+\overline{\Delta}\theta)^2}{\overline{\Delta}^2}\left(\frac{V_1}{(n-a)^2} + \frac{\theta^2V_2 + 4\theta V_3 + V_4}{(a-1)^2(1+\overline{\Delta}\theta)^2} - \frac{2\kappa_2 V_5}{(n-a)(a-1)(1+\overline{\Delta}\theta)}\right).$$

The distribution of $\widehat{\theta}$ tends to be positively skewed for finite sample sizes so employing the normal approximation in (5) may not produce acceptable results. The logarithmic transformation is applied to the REML estimator of θ using Slutzsky's theorem, and it follows that

$$ln\left(1+\overline{\Delta}\widehat{\theta}\right) \overset{asymp}{\sim} N\left(ln(1+\overline{\Delta}\theta), Var(ln(1+\overline{\Delta}\widehat{\theta}))\right).$$

$Var(ln(1+\overline{\Delta}\widehat{\theta}))$ may be written as

$$Var(ln(1+\overline{\Delta}\widehat{\theta})) = 2\left(\frac{1}{n-a} + \frac{c\theta^2 + 2(a-1)\overline{\Delta}\theta + (a-1)}{(a-1)^2(1+\overline{\Delta}\theta)^2}\right) + \kappa_1\theta^2\frac{c_1}{(a-1)^2(1+\overline{\Delta}\theta)^2}$$

$$+\kappa_2\left(\frac{c_{21}}{(n-a)^2}+\frac{c_{22}}{(a-1)^2(1+\overline{\Delta}\theta)^2}-\frac{2c_{23}}{(n-a)(a-1)(1+\overline{\Delta}\theta)}\right) \tag{6}$$

where c, c_1, c_{21}, c_{22}, and c_{23} are constants that depend on n, a and $b_i, i = 1, ..., d$. See Burch (2011b) for details.

In equation (6), $\kappa_1\theta^2 = E({A_i}^4)/\sigma_2^4 - 3\theta^2$, $\kappa_2 = E({e_{ij}}^4)/\sigma_2^4 - 3$, and to construct the endpoints of a REML-based confidence interval for θ, $\kappa_1\theta^2$ and κ_2 must be estimated. Simple plug-in estimators are given by

$$\widehat{\kappa_1\theta^2} = \frac{\frac{1}{n}\sum_{i=1}^{a} b_i(\overline{Y}_{i.}-\overline{Y}_{..})^4}{MSE^2} - 3\widehat{\theta}^2$$

if $MSA \geq MSE$, where $\widehat{\kappa_1\theta^2} = 0$ otherwise, and

$$\widehat{\kappa_2} = \frac{\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b_i}(Y_{ij}-\overline{Y}_{i.})^4}{MSE^2} - 3.$$

These estimators, however, do not correct for bias. Empirical results suggest that an estimator based on $\widehat{\kappa_1\theta^2}$ can drastically overestimate $\kappa_1\theta^2$ whereas using $\widehat{\kappa_2}$ can severely underestimate κ_2 for leptokurtic distributions. An and Ahmed (2008) and Burch (2011a) discuss in detail the use of empirical bias-corrected estimators of kurtosis in related applications. Furthermore, Burch (2011b) provides suggestions to modify $\widehat{\kappa_1\theta^2}$ and $\widehat{\kappa_2}$ which yield improved performance. Using this approach, the REML-based confidence interval for θ is

$$\left(\frac{1}{\overline{\Delta}}\left[\left(1+\overline{\Delta}\widehat{\theta}\right)e^{-Z_{1-\alpha/2}\sqrt{\widehat{Var}\left(ln\left(1+\overline{\Delta}\widehat{\theta}\right)\right)}}-1\right],\frac{1}{\overline{\Delta}}\left[\left(1+\overline{\Delta}\widehat{\theta}\right)e^{Z_{1-\alpha/2}\sqrt{\widehat{Var}\left(ln\left(1+\overline{\Delta}\widehat{\theta}\right)\right)}}-1\right]\right). \tag{7}$$

where $\widehat{Var}\left(ln\left(1+\overline{\Delta}\widehat{\theta}\right)\right)$ is the estimator of (6) and $Z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. The corresponding REML-based confidence interval for ρ is readily available.

4. Confidence Intervals for ρ

4.1 Simulation Study

The confidence intervals for θ using the PQ and REML methods are converted to confidence intervals for ρ and the simulated coverage probabilities are computed for a variety of scenarios. The values of ρ considered are 0.1, 0.5, and 0.9. Furthermore, since asymptotic results hold as a increases for bounded b_i where $i = 1, ..., a$, let $a = 5, 10, 50$ and 100, and consider the b_i-patterns listed in Table 2. For $a = 5$, each pattern has a total sample size of 24 but the patterns exhibit varying degrees of imbalance. Ahrens and Pincus (1981) used Φ to measure imbalance in one-way random effects models where $\Phi = (a/\sum_{i=1}^{a} b_i)(a/\sum_{i=1}^{a} 1/b_i)$, $0 < \Phi \leq 1$, and Φ is equal to 1 if and only if the model is balanced. Note that for $a = 10, 50$ and 100, the patterns listed in Table 2 are simply replicated so that, for instance, pattern 1

Table 2: b_i-patterns for $a = 5$

Pattern	b_1	b_2	b_3	b_4	b_5	Φ
1	5	5	5	5	4	0.99
2	10	5	5	2	2	0.69
3	10	10	2	1	1	0.39
4	20	1	1	1	1	0.26

for $a = 10$ results in ten classes with individual class sizes of 5, 5, 5, 5, 4, 5, 5, 5, 5, and 4. The overall sample sizes ($n = \sum_{i=1}^{a} b_i$) considered in this study range from 24 to 480.

Table 3 displays the kurtosis values for a variety of distributions based on a random variable X having mean μ and variance σ^2. Both symmetric and asymmetric distributions are included. It is interesting to note that the smallest possible value of kurtosis is -2, which occurs when X is distributed Bernoulli(1/2). At the other extreme, the kurtosis for $X \sim$ t(4) is infinite.

Table 3: Kurtosis values for various distributions

	Distribution	$\kappa = E[(X-\mu)^4]/\sigma^4 - 3$
Symmetric	Uniform(0,1)	-1.2
	N(0,1)	0.0
	t(5)	6.0
Asymmetric	Beta(0.4,0.6)	-1.33
	$\chi^2(1)$	12.0

Without loss of generality, let $\mu = 100$. Furthermore, assume that $A_i \overset{iid}{\sim}(0, \sigma_1^2)$, $e_{ij} \overset{iid}{\sim}(0, \sigma_2^2)$, A_i and e_{ij} are mutually independent, and these random variables depend on the distributions listed in Table 3. These distributions have varying degrees of skewness and kurtosis. If the distribution considered does not have the appropriate mean and variance for a particular simulation, then the distribution is simply located and scaled so that it does. For instance, if one considers the random variable $X_i \sim$ Uniform(0,1) for factor A_i, then let $A_i = \sigma_1\sqrt{12}(X_i - 0.5)$, where $i = 1, ..., a$. It follows that $E(A_i) = 0$ and $Var(A_i) = \sigma_1^2$. For this simulation study we assume that A_i and e_{ij} are built from the same underlying distribution.

For each combination of ρ, number of laboratories (a), observations per laboratory ($b_i, i = 1, ..., a$) and distribution, we generated 10,000 Monte Carlo random samples and simulated the coverage probabilities of the PQ and REML methods. The simulated coverage probabilities are summarized using boxplots for each interval procedure. For example, simu-

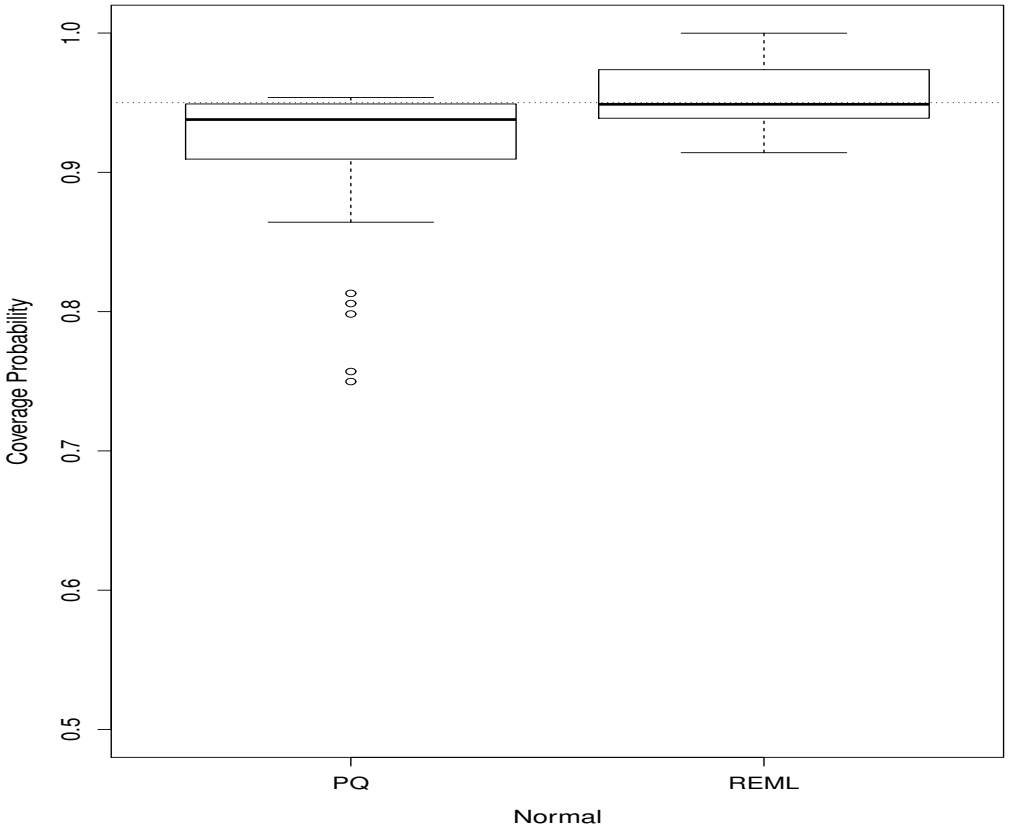

Figure 3: Simulated coverage probabilities of 95% nominal confidence intervals for $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. Results are for the PQ and REML procedures assuming normality.

lated coverage probabilities for nominal 95% confidence intervals for ρ are displayed in Figure 3 assuming A_i and e_{ij} are normally distributed. Each boxplot is comprised of coverage values for the forty eight combinations of ρ, a, and b_i-pattern. The nominal coverage probability of 0.95 is depicted by the horizontal dotted line. Even under normal distribution assumptions, Figure 3 indicates that the PQ procedure may not result in confidence intervals that have adequate coverage probabilities. This is primarily due to the imbalance in the models under consideration. The greater the degree of imbalance, the greater the degradation is in the coverage probability. The REML procedure is more likely to maintain coverage values that are close to the nominal level of 0.95.

Figure 4 summarizes the simulated coverage probability results for each of the Beta(0.4,0.6), Uniform(0,1), t(5), and $\chi^2(1)$ distributions. The PQ method performs adequately for the platykurtic distributions ($\kappa < 0$) but exhibits severe undercoverage for the leptokurtic distributions ($\kappa > 0$). In many cases the low coverage probability values in the boxplots are also associated with models having a large degree of imbalance. However, when considering a specific distribution in Figure 4, the least desirable coverage probabilities for both the

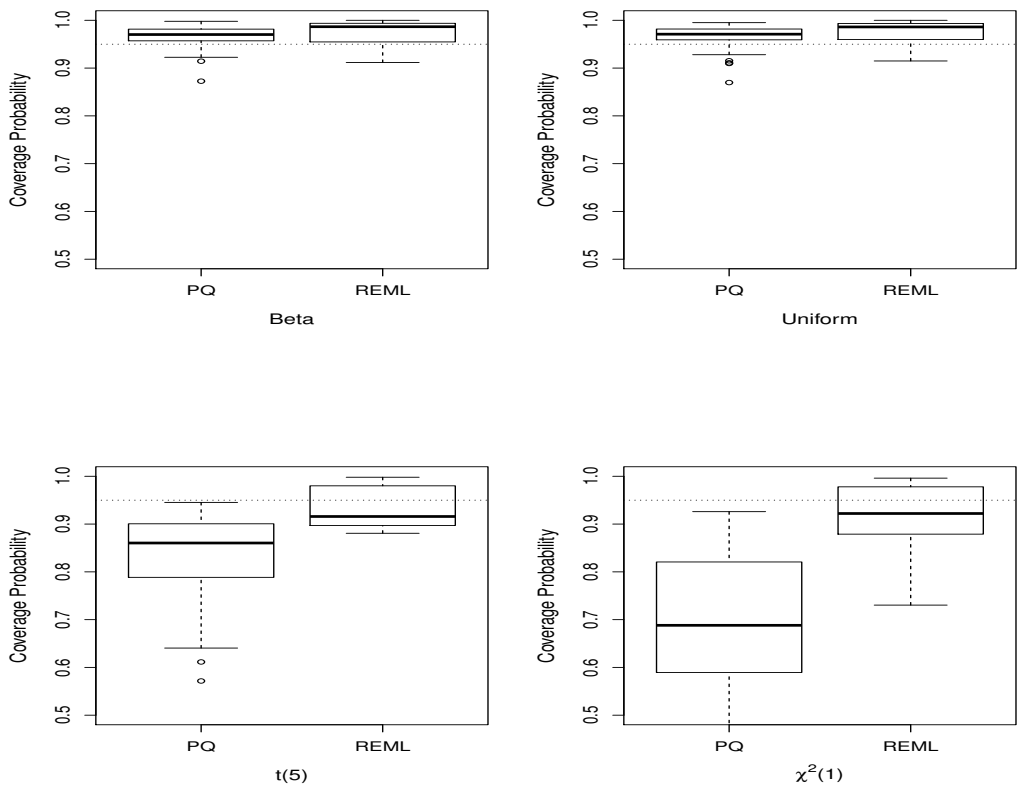

Figure 4: Simulated coverage probabilities of 95% nominal confidence intervals for $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. Results are for the PQ and REML procedures for the Beta(0.4,0.6), Uniform(0,1), t(5), and $\chi^2(1)$ distributions.

PQ and REML methods are primarily associated with large values of ρ. See Burch (2011b) for more details. Although not perfect, the REML procedure is more apt to have coverage values that are closer to the nominal level of 0.95.

From a different vantage point, Figure 5 summarizes the simulated coverage probability results for the four b_i-patterns listed in Table 2. Each boxplot is comprised of coverage values for the sixty combinations of ρ, a, and underlying distribution. It is clear that imbalance has a more detrimental effect on the coverage probabilities of the PQ method than it does on the coverage probabilities of the REML method. However, it is not obvious why the REML method's performance improves when the data is more unbalanced.

Although not explicitly depicted in the previous figures, the PQ method's performance can deteriorate as the sample increases since the PQ itself is not based on large-sample theory. For example, using $\rho = 0.5$, pattern 1 in Table 2, and the $\chi^2(1)$ distribution, the simulated coverage probability of 0.8342 for $a = 5$ decreases to 0.6497 for $a = 100$. See Burch (2011b) for more details. The REML procedure's performance, on the other hand, continues to improve as a increases due to the large-sample properties of the REML estimator. In general, the REML confidence intervals tend to be wider than the PQ confidence intervals

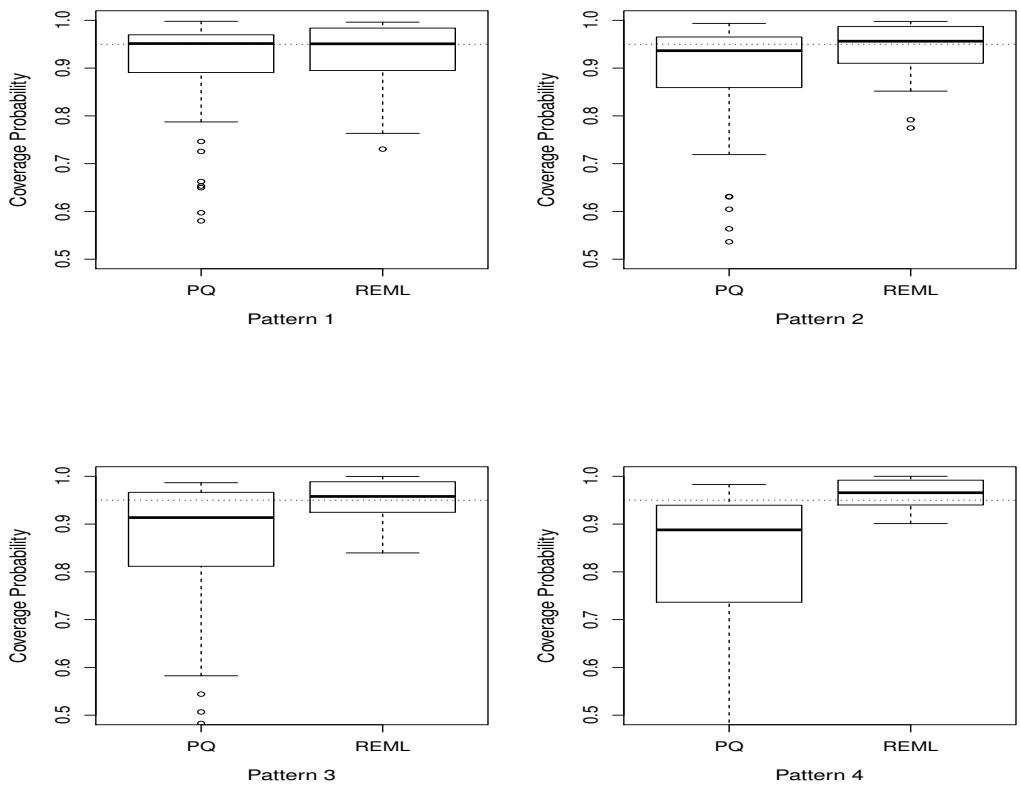

Figure 5: Simulated coverage probabilities of 95% nominal confidence intervals for $\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. Results are for the PQ and REML procedures for the four degrees of model imbalance.

for the cases considered. However, for distributions having a large kurtosis the REML procedure has coverage probabilities that are close to the nominal level of 0.95 and thus wider intervals are warranted.

4.2 Arsenic in Oyster Tissue

In the arsenic concentration in oyster tissue application, $n = 116$, $a = 31$, $b_1 = ... = b_{28} = 4$, $b_{29} = 2$, and $b_{30} = b_{31} = 1$. In addition, $\Phi = 0.87$, $\overline{\Delta} = 3.74$, $\widehat{\theta} = 9.56$, and the REML estimate for ρ is 0.91. 95% confidence interval estimates for ρ using the PQ and REML procedures are $(0.84, 0.95)$ and $(0.79, 0.96)$, respectively. The PQ interval for ρ was obtained by the appropriate transformation of the endpoints of the interval in (3). Similarly, the REML interval for ρ was obtained by the appropriate transformation of the endpoints of the interval in (7).

Conventional confidence intervals for ρ assume that A_i and e_{ij} are normally distributed. While one cannot directly determine whether or not the normal distribution assumptions are valid, in this example we computed $\widehat{A}_i = \overline{Y}_{i.} - \overline{Y}_{..}$, $i = 1, ..., 31$ and $\widehat{e}_{ij} = Y_{ij} - \overline{Y}_{i.}$, $i = 1, ..., 31$, $j = 1, ..., b_i$, and considered their distributions in lieu of the distributions of A_i and e_{ij}. Specifically, density histograms of the distributions of $\widehat{A}_i$ and $\widehat{e}_{ij}$ indicate that

the distribution of $\widehat{A}_i$ is asymmetric and the the distribution of $\widehat{e}_{ij}$ is more peaked than the normal distribution. Shapiro-Wilk tests for normality indicate that the distributions of $\widehat{A}_i$ and $\widehat{e}_{ij}$ are not normal and thus it is unlikely that the confidence level of the PQ procedure is actually 95%. In this example, the REML confidence interval procedure is much more likely to achieve the stated coverage probability.

5. Summary

This paper examines confidence intervals for variance components under non-normal distribution assumptions. In particular, confidence intervals for a ratio of variances ($\theta = \sigma_1^2/\sigma_2^2$) or the intraclass correlation coefficient ($\rho = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$) are under study. Whereas Burch (2011a) presented confidence intervals for functions of variances under non-normality using balanced one-way random effects models, the present paper provides results for unbalanced one-way random effects models. Additional details are provided by Burch (2011b). The combination of non-normality and imbalance can negatively impact the quality of commonly used confidence interval procedures.

The PQ procedure relies on the F-distribution to construct the endpoints of the interval. This method provides exact confidence results for the balanced one-way random effects model under normal distribution assumptions. However, even under normality, simulated coverage probabilities for unbalanced models may not be acceptable. Furthermore, the PQ method is not based on large-sample theory and simulated coverage probabilities may actually diverge from the nominal value as the sample size increases. The REML procedure incorporates estimators of kurtosis in the interval construction process and performed adequately for the majority of the cases considered. As the sample size increases, the variability of the simulated coverage probabilities using the REML confidence interval procedure tends to decrease and one is more likely to achieve a coverage probability that is close to the nominal value.

Table 4 provides a summary of the simulation study which considered 240 combinations of ρ, a, b_i-patterns, and distributions. The REML confidence interval procedure's mean coverage probability matched the nominal level of 0.95. To achieve the prescribed value, these intervals where, on average, wider than the PQ generated intervals. The PQ method fell short of the nominal coverage probability of 0.95.

Table 4: Comparing the PQ and REML Confidence Intervals

Interval Method	Mean Coverage Probability	Mean Length
PQ	0.8737	0.3327
REML	0.9508	0.4112

Acknowledgements

The author thanks the anonymous referee for the valuable suggestions which improved the quality of the article.

References

Ahrens, H., Pincus, R., 1981. On two measures of unbalancedness in a one-way model and their relation to efficiency. Biom. J. 23, 227–235.

An, L., Ahmed, S. E., 2008. Improving the performance of kurtosis estimator. Comput. Statist. Data Anal. 52, 2669–2681.

Burch, B. D., 2011a. Assessing the performance of normal-based and REML-based confidence intervals for the intraclass correlation coefficient. Comput. Statist. Data Anal. 55, 1018–1028.

Burch, B. D., 2011b. Confidence intervals for variance components in unbalanced one-way random effects models using non-normal distributions. J. Statist. Plan. Inference 141, 3793–3807.

Burch, B. D., Harris, I. R., 2001. Closed-form approximations to the REML estimator of a variance ratio (or heritability) in a mixed linear model. Biometrics 57, 1148–1156.

Burch, B. D., Iyer, H. K., 1997. Exact confidence intervals for a variance ratio (or heritability) in a mixed linear model. Biometrics 53, 1318–1333.

Hammersley, J. M., 1949. The unbiased estimate and standard error of the intraclass variance. Metron 15, 189–205.

Harville, D. A., Fenech, A. P., 1985. Confidence intervals for a variance ratio, or heritability, in an unbalanced mixed linear model. Biometrics 41, 137–152.

Jiang, J., 1996. REML estimation: Asymptotic behavior and related topics. Ann. Statist. 24, 255–286.

Jiang, J., 2005. Partially observed information and inference about non-gaussian mixed linear models. Ann. Statist. 33, 2695–2731.

Shoukri, M. M., Tracy, D. S., Main, I. U. H., 1990. The effect of kurtosis in estimation of the parameters of the one-way random effects model from familial data. Comput. Statist. Data Anal. 10, 339–345.

Singh, H. P., Shukla, S. K., Singh, S., 2002. The utilization of kurtosis in the estimation of the parameters of the one-way random effects model. Biom. J. 44, 1028–1040.

Thomas, J. D., Hultquist, R. A., 1978. Interval estimation for the unbalanced case of the one-way random effects model. Ann. Statist. 11, 582–587.

Wald, A., 1940. A note on the analysis of variance with unequal class frequencies. Ann. Math. Stat. 11, 96–100.

Westfall, P. H., 1987. A comparison of variance component estimates for arbitrary underlying distributions. J. Amer. Statist. Assoc. 82, 866–874.

Willie, S., Berman, S., 1995. NOAA national status and trends program eighth round intercomparison exercise results for trace metals in marine sediments and biological tissues. NOAA Tech. Memorandum NOS ORCA 83.