

METHODOLOGY

Open Access



Exact distributions of statistics for making inferences on mixed models under the default covariance structure

Samaradasa Weerahandi^{1*} and Ching-Ray Yu²

*Correspondence:

Weerahandi@aol.com

¹X-Techniques, 23 Chestnut St,
08817 Edison, NJ, USA

Full list of author information is
available at the end of the article

Abstract

At this juncture when mixed models are heavily employed in applications ranging from clinical research to business analytics, the purpose of this article is to extend the exact distributional result of Wald (Ann. Math. Stat. 18: 586–589, 1947) to handle models involving a number of variance components.

Due to the unavailability of exact distributional results for underlying statistics, currently available methods provide small group/sample inference only for balanced ANOVA models or simple regression models. The exact distributional results developed in this article should prove useful in making inferences by such methods as parametric bootstrap, fiducial, and generalized p-value approach, when there are a number of variance components to deal with.

Keywords: Variance components, Compound symmetric covariance structure, REML, BLUP, Profile likelihood test, Wald test

1 Introduction

Ever since Wald (1947) published his celebrated paper, basically there has been only relaxation of the assumptions (cf. Seely and El-Bassiouni (1983)) made in that paper involving a simple mixed model. Therefore, the purpose of this article is to extend Wald's result to develop additional exact distribution theory to tackle various inference problems in mixed models having the compound symmetric covariance structure.

The variance components in mixed models (cf. Searle et al. (2006)) play an important role in such applications as clinical analytics, business analytics, and industrial process control (cf. Hamada and Weerahandi (2000), and Wu and Hamada (2009)), a whole area that classical approach to inference fails to provide small sample inference. However, the solutions they provided are valid only for ANOVA type problems.

Mixed-effects models first became popular due to the ease of handling covariance structures associated with longitudinal data, and then due to the well known advantage (cf. Henderson (1975) and Robinson (1991)) of BLUPs (Best linear unbiased predictors) over the LSEs (Least squares estimators) when a practitioner has to make inferences on a number of groups (e.g. markets, patient groups, and treatment levels) in a grouping structure.

In fact, in Corporate America the BLUP has replaced the LSE as the most widely used statistical technique in advanced business analytics.

In mixed-effects regression models, the MLE (Maximum likelihood estimator) based methods, are basically the only class of methods available for making inferences on models involving a number of random effects. But such solutions are asymptotic methods requiring, not only the sample sizes to be large, but also the number of levels in a grouping structure to be large, an unreasonable assumption. Therefore, in this article we will also briefly develop tests and interval estimates for variance components and their functions.

2 The model and problems

To describe the underlying problems more precisely, consider a regression model in a mixed model setting with a number of random effects in a grouping structure,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \cdots + \mathbf{Z}_J\mathbf{u}_j + \boldsymbol{\epsilon}, \quad (1)$$

where \mathbf{X} is an $N \times k$ matrix formed by a set of covariates corresponding to k fixed effects $\boldsymbol{\beta}$, \mathbf{Z}_j are $N \times k_j$ matrices formed by covariates corresponding to random effects \mathbf{u}_j of j^{th} factor with k_j levels. We make the usual assumption that $\boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \mathbf{I}_N)$. Also using the most widely used compound symmetric covariance structure, which is also the default setting in statistical software packages, we also make the assumption $\mathbf{u}_j \sim N(\mathbf{0}, \sigma_j^2 \mathbf{I}_{k_j})$, $j = 1, \dots, J$, and that they are distributed independently of each other and the error terms. It should be noted that the vector of fixed effects $\boldsymbol{\beta}$ contain the overall mean effect of each of the factors having random effects, which we denote as β_{zj} , $j = 1, \dots, J$.

The model can also be written in familiar compact form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad (2)$$

where $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_J)$ and $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_J)'$. Henderson (1975) provided the formulas for the BLUEs (Best linear unbiased estimators) of the fixed effects and the BLUPs of random effects when the variances are known. Robinson (1991) showed that various derivations available in the literature for these quantities all lead to Henderson's formulas

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{Z}\tilde{\mathbf{u}} = \mathbf{X}'\mathbf{y}, \quad (3)$$

$$\mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\beta}} + (\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Sigma}^{-1})\tilde{\mathbf{u}} = \mathbf{Z}'\mathbf{y},$$

where $\boldsymbol{\Sigma} = \text{Var}(\mathbf{u})$ and $\tilde{\mathbf{u}}$ is the BLUP of \mathbf{u} . Under the compound symmetric covariance structure, the default in statistical software packages, $\boldsymbol{\Sigma}_\rho = \boldsymbol{\Sigma}/\sigma_\epsilon^2 = \text{diag}(\rho_1 \mathbf{I}_1, \rho_2 \mathbf{I}_2, \dots, \rho_J \mathbf{I}_J)$, where $\rho_j = \sigma_j^2/\sigma_\epsilon^2$. The equations in (3) can be solved explicitly for $\hat{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$. It is evident from (2) and was formally proved by Henderson et al. (1975) that $\hat{\boldsymbol{\beta}}$ is actually the GLSE (Generalized least squares estimator)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{y}, \text{ where } \boldsymbol{\Gamma} = \mathbf{I}_N + \mathbf{Z}\boldsymbol{\Sigma}_\rho\mathbf{Z}'. \quad (4)$$

Then, we can easily obtain the explicit solution for the random part of the predictor as

$$\tilde{\mathbf{u}} = (\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Sigma}_\rho^{-1})^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (5)$$

2.1 The issue with available results

When we go about making inferences in applications involving a number of variance components, we encounter the fact that the variance components are unknown parameters,

and hence need to be tackled before we can carry out any kind of inference. The most widely used method is to estimate them by the REML (Restricted maximum likelihood) or the ML (Maximum likelihood), and then carry out tests and interval estimates using their asymptotic distributions. It is well known that (cf. Weerahandi (2004) and Gamage et al. (2013)), due to using such asymptotic assumptions, which require not only the sample sizes to be large, but also the number of factor levels in a grouping structure to be large, an unreasonable assumption in practice.

Therefore, the purpose of this article is to develop inference methods without relying on asymptotic distributional results. Since variance components are important in their own merit in a variety of applications we will undertake one application in the area of industrial process control.

3 Exact distributions of LSEs

To solve problems involving variance components, regardless one takes the GPQ (Generalized pivotal quantity) based approach, parametric bootstrap approach, or generalized fiducial approach, we need to tackle the variance components, which are typically unknown. The inferences based on MLE and asymptotic results are highly inaccurate in that they seriously miss the intended Type I error in testing of hypotheses and the intended probability coverage in interval estimation

3.1 Distributions for inferences on variance components

By considering the ordinary least squares regression of $\tilde{\mathbf{X}} = (\mathbf{X}, \mathbf{Z})$ on \mathbf{y} , Wald (1947) showed, albeit in non-matrix notation, that the distribution of the residual sum of squares, $S_e = \mathbf{y}' (\mathbf{I}_n - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}) \mathbf{y}$ is related to the familiar chi-squared distribution

$$U = \frac{S_e}{\sigma_e^2} \sim \chi_e^2, \text{ where } e = N - \text{rank}(\tilde{\mathbf{X}}). \quad (6)$$

Moreover, U was shown to be distributed independently of \mathbf{u} . Typically $\tilde{\mathbf{X}}$ is not of full rank as Wald (1947) implicitly assumed, but the assumption can be maintained by redefining $\tilde{\mathbf{X}}$ as the non-singular sub-matrix that one gets if the regression is run blindly. Gamage et al. (2013) showed how one variance component can be tackled by transforming $\tilde{\mathbf{X}}$ by an orthonormal basis for the vector space spanned by its columns. Actually, any number of variance components in a compound symmetric structure can be handled more conveniently and efficiently based on the LSE results, because underlying LSE results based on the QR algorithm is highly computationally efficient and fool proof in handling non-singular design matrices. This is accomplished by extending Wald's argument for any number of random effects, and properly redefining the $\tilde{\mathbf{X}}$ matrix.

From the theory of regression and extended Wald results, it is known that $\hat{\mathbf{u}}|\mathbf{u} \sim N(\mathbf{u}, \sigma_e^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1})$, where $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$ is the sub-matrix of $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$ corresponding to the \mathbf{u} part of the LSE regression. Therefore, the unconditional as well as the conditional distribution of $(\hat{\mathbf{u}} - \mathbf{u})$ given \mathbf{u} is given by

$$(\hat{\mathbf{u}} - \mathbf{u})|\mathbf{u} \sim N(\mathbf{0}, \sigma_e^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}),$$

a distribution free of \mathbf{u} . Hence the unconditional distribution of

$$(\hat{\mathbf{u}} - \mathbf{u}) \sim N(\mathbf{0}, \sigma_e^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}),$$

is also the same and distributed independently of \mathbf{u} . Hence, the LSE, as opposed to Henderson’s predictor $\tilde{\mathbf{u}}$, of \mathbf{u} is distributed as

$$\hat{\mathbf{u}} = (\hat{\mathbf{u}} - \mathbf{u}) + \mathbf{u} \sim N\left(\mathbf{0}, \sigma_e^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} + \Sigma_u\right), \tag{7}$$

where Σ_u is the covariance matrix of \mathbf{u} . In particular, individual variance components and variance ratios of practical importance can be tackled based on the distributional result

$$\hat{\mathbf{u}}_j \sim N\left(\mathbf{0}, \sigma_e^2\left((\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} + \rho_j I_j\right)\right), \text{ where } \rho_j = \frac{\sigma_j^2}{\sigma_e^2}. \tag{8}$$

In computing $\hat{\mathbf{u}}_j$ s by LSE, perhaps it is convenient to run the unconstrained regression, and then center each $\hat{\mathbf{u}}_j$ by subtracting the mean, the mean of estimated β_{zj} .

3.2 Distributions for functions of variance components

The distributional result (8) allows us to handle any variance component by the distribution of the corresponding component of $\hat{\mathbf{u}}$. In solving some of the inference problems in quality assurance and in applications of the BLUP, however, we also need to tackle multiple variance components simultaneously. To develop necessary distributional results for such applications, consider two factors of interests, and let $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ be the least squares estimators of their random effects \mathbf{u}_1 and \mathbf{u}_2 , each of which corresponds to non-singular sub-matrices of $\tilde{\mathbf{X}}$.

Then, it follows from (7) that

$$\begin{pmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0}_1 \\ \mathbf{0}_2 \end{pmatrix}, \sigma_e^2 \Sigma_{x\rho}\right), \tag{9}$$

where

$$\Sigma_{x\rho} = \Sigma_x(\rho_1, \rho_2) = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}_{12} + \begin{pmatrix} \rho_1 \mathbf{I}_1 & \mathbf{0}_1 \\ \mathbf{0}_2 & \rho_2 \mathbf{I}_2 \end{pmatrix}.$$

Let $\Sigma_{x\rho}^{-1}$ be the inverse of the matrix $\Sigma_{x\rho}$, found by Eigen value decomposition or otherwise. Then, from (9) we can obtain two independent random vectors as

$$\begin{pmatrix} \hat{\mathbf{v}}_1 \\ \hat{\mathbf{v}}_2 \end{pmatrix} = \Sigma_{x\rho}^{-1} \begin{pmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0}_1 \\ \mathbf{0}_2 \end{pmatrix}, \sigma_e^2 \mathbf{I}\right).$$

Now it is evident that the two variance components can be tackled using the two independent chi-squared random variables

$$V_l = \frac{S_l(\rho_1, \rho_2)}{\sigma_e^2} \sim \chi_{a_l}^2, \text{ where } S_l(\rho_1, \rho_2) = \hat{\mathbf{v}}_l' \hat{\mathbf{v}}_l, \text{ for } l = 1, 2, \tag{10}$$

where $a_l = \text{rank}(\tilde{\mathbf{X}}_l)$, which in many cases equal to $k_l - 1$.

As shown by Weerahandi and Gamage (2016), an alternative way of making inferences in problems involving two parameters can be accomplished by a marginal distribution of one statistic and the conditional distribution of the other statistic. In our application we readily have the marginals given by (8), the conditional distributions can be found using the multivariate distribution theory on normal distributions. For example, the conditional distribution of $\hat{\mathbf{u}}_2$ given $\hat{\mathbf{u}}_1$ is given by

$$\hat{\mathbf{u}}_2 | \hat{\mathbf{u}}_1 \sim N(\delta_{21}, \Sigma_{2,1}), \tag{11}$$

where

$$\delta_{21} = \Sigma_{22}\Sigma_{22}^{-1}\hat{\mathbf{u}}_1, \Sigma_{2,1} = \Sigma_{11} - \Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21},$$

and sub-matrices of $\Sigma_{x\rho}$ appearing in the above equation are obtained by the partition

$$\Sigma_x(\rho_1, \rho_2) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

It is now evident that we can tackle the two variance components of interest using the chi-squared statistics,

$$V_1 = \frac{S_1(\rho_1)}{\sigma_e^2} \sim \chi_{a_1}^2, \text{ where } S_1(\rho_1) = \hat{\mathbf{u}}_1' \left((\tilde{\mathbf{X}}'\tilde{\mathbf{X}})_1^{-1} + \rho_1 I_1 \right)^{-1} \hat{\mathbf{u}}_1 \quad (12)$$

$$V_{21} = \frac{S_{21}(\rho_1, \rho_2)}{\sigma_e^2} | \hat{\mathbf{u}}_1 \sim \chi_{a_2}^2, \text{ where } S_{21}(\rho_1, \rho_2) = (\hat{\mathbf{u}}_1 - \delta_{21})' \Sigma_{2,1}^{-1} (\hat{\mathbf{u}}_1 - \delta_{21}).$$

However, it is inconvenient to work with conditional distributions in deriving inference methods. We can avoid the above conditional distribution by applying the transformation suggested by Weerahandi and Gamage (2016). In our problem this is accomplished by considering the distribution of the CDF (Cumulative distribution function) of V_{21} , which is distributed as

$$U_{21} = F_{V_{21}} = C(\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2; \rho_1, \rho_2, \sigma_e^2) | \hat{\mathbf{U}}_1 = \hat{\mathbf{u}}_1 \sim U(0, 1), \quad (13)$$

where uppercase variables represent the observable random quantities, and C is the CDF of the chi-squared distribution with a_2 degrees of freedom, which is readily available from statistical software packages. But the above uniform distribution is free of $\hat{\mathbf{u}}_1$. Therefore, the random variable U_{21} is distributed as

$$U_{21} = C(\hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2; \rho_1, \rho_2, \sigma_e^2) \sim U(0, 1) \quad (14)$$

independently of $\hat{\mathbf{U}}_1$. Obviously the above argument can be extended for two sets of random effects to enable handling of any number of variance components in an iterative manner.

4 Illustration: Inference on functions of variance components

In this section we illustrate the application of foregoing results by undertaking a function of variance components that arise in practice of statistical quality assurance and in some applications of BLUPs. For a variety of problems associated with Gauge, ‘‘Repeatability’’ and ‘‘Reproducibility’’, and the reader is referred to Wu and Hamada (2009) for a detailed discussion of these notions and for various sums and ratios of variance component that are important in quality assurance measurement systems.

We will illustrate the approach that one can take in such applications, by considering the problems of making inferences on a variance component and a sum of one variance component σ_j^2 of and the error variance σ_e^2 . Making inferences about the variance ratio ρ_j itself is a trivial task because, it follows chi-squared random variate

$$V_j = \frac{S(\rho_j)}{\sigma_e^2} \sim \chi_a^2, \text{ where } S(\rho_j) = \hat{\mathbf{u}}_j' \left((\tilde{\mathbf{X}}'\tilde{\mathbf{X}})_j^{-1} + \rho_j I_j \right)^{-1} \hat{\mathbf{u}}_j, \quad (15)$$

and $a = k_j - 1$ and $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})_j^{-1}$ is the sub-matrix of $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$ corresponding to the \mathbf{u}_j part of the regression. In some applications it is more convenient to work with the alternate form of (15),

$$V_j = \tilde{S}(\sigma_e^2, \sigma_j^2) \sim \chi_a^2, \text{ where } \tilde{S}(\sigma_e^2, \sigma_j^2) = \hat{\mathbf{u}}_j' \left(\sigma_e^2 (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})_j^{-1} + \sigma_j^2 I_j \right)^{-1} \hat{\mathbf{u}}_j, \quad (16)$$

If needed, the computation of $S(\rho_j)$ can be simplified by diagonalizing the $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})_j^{-1}$ matrix by means of an orthogonal matrix, say \mathbf{P} , as

$$S(\rho_j) = \tilde{\mathbf{u}}_j' (\mathbf{D}_j + \rho_j \mathbf{I}_j)^{-1} \tilde{\mathbf{u}}_j = \sum_l \frac{u_{lj}^2}{(\rho_j + \lambda_{lj})}$$

where \mathbf{D}_j is a diagonal matrix formed by Eigen values λ_j of $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})_j^{-1}$ and $\tilde{\mathbf{u}}_j = \mathbf{P}\hat{\mathbf{u}}_j$.

It follows from (15) and (8) that

$$W = \frac{aS_e}{eS(\rho_j)} \sim F_{e,a}, \quad (17)$$

a test statistic involving no nuisance parameters. For example, it is easily seen that an upper $100\gamma\%$ confidence bound for ρ_j is given by

$$\rho_j \leq \max \left(0, s_j^{-1} \left(\frac{aS_e}{eq_\gamma} \right) \right),$$

where q_γ is the γ^{th} quantile of the F-distribution with a and e degrees of freedom, and s_j^{-1} is the inverse function of $s(\rho_j)$.

Making inferences on individual variance components, and sums and some ratios of variance components is not a trivial task. In fact, classical approach to inference fails to provide small sample inference even in the simple balanced mixed models (cf. Weerahandi (1992)). Whereas inferences on any function of variance components is easily accomplished by taking the generalized approach to inference. It should be emphasized that one can obtain equivalent results by taking the generalized fiducial inference approach introduced by Hannig et al. (2006). For example, most of the the problems undertaken by Cisewski and Hannig (2012) can be tackled using exact distributional results without relying on any asymptotic or approximate results. Moreover, now that we have exact distributional results, one can also take Parametric Bootstrap approach to make inferences involving exact probability statements, just like Generalized Inference does.

4.1 Inference on individual variance components

Perhaps the easiest way to make inferences on individual variance components as well as functions of variance components, is by replacing each nuisance parameter by their GPQ (Generalized pivotal quantity) that reduces to the parameter at the observed sample point, and then showing the resulting quantity leads to a well defined extremal region in the problem of testing of the parameter of interest. While the error variance, σ_e^2 is easily handled by the obvious GPQ suggested by (6)

$$G_{\sigma_e^2} = \frac{\sigma_e^2 s_e}{S_e} = \frac{s_e}{U}, \quad (18)$$

which is basically the same as the classical PQ (Pivotal quantity). The GPQs of other variance components is not that easy to derive, but by applying what is known as the substitution method (cf. Weerahandi (2004)) to Eqs. (6) and (15), we can obtain a GPQ that reduces to σ_j^2 at the observed sample point as

$$G_{\sigma_j^2} = \frac{\sigma_e^2 s_e}{S_e} s^{-1} \left(\frac{s_e S(\sigma_j^2 / \sigma_e^2)}{S_e} \right) = \frac{s_e}{U} s^{-1} \left(\frac{s_e V_j}{U} \right), \quad (19)$$

where $s()^{-1}$ is the inverse function of $s()$ unless it is negative, in which case the GPQ is set to 0. Obviously expression (19) provides a GPQ for σ_j^2 , because it has the required properties:

- The distribution of $G_{\sigma_j^2}$ is free of unknown parameters
- At observed sample points, $G_{\sigma_j^2}$ depends only on the parameter of interest, σ_j^2 .

In fact, at observed sample points, $G_{\sigma_j^2}$ reduces to σ_j^2 .

Nevertheless, the inverse function $s^{-1}()$ of $s()$ can be avoided, by first considering the problem of testing null hypotheses of the form

$$H_0 : \sigma_j^2 \geq \sigma_0^2.$$

A GTV (Generalized test variable), a notion introduced by Tsui and Weerahandi (1989), for testing the above hypotheses can also be obtained by the substitution method in alternative equivalent forms as

$$\begin{aligned} T &= \frac{s_e S \left(\sigma_j^2 / \sigma_e^2 \right)}{S_e s \left(\frac{\sigma_j^2 S_e}{\sigma_e^2 s_e} \right)} = \frac{V_j}{U} \frac{s_e}{s \left(\frac{U \sigma_j^2}{s_e} \right)}, \text{ or as} \\ T &= \frac{\tilde{S}(\sigma_e^2, \sigma_j^2)}{\tilde{s} \left(\frac{\sigma_e^2 s_e}{S_e}, \sigma_j^2 \right)} = \frac{V_j}{\tilde{s} \left(\frac{s_e}{U}, \sigma_j^2 \right)}. \end{aligned} \quad (20)$$

Notice that the distribution of the above test variable is free of nuisance parameters, because the expression of T in (20) involves only σ_j^2 , the parameter of interest. In addition, from its first expression, it is clear that the test variable reduces to 1 at the observed values of the random variables, because at the observed values S_e become equal to s_e , and S becomes equal to s , thus leading to a well defined point on the boundary of extreme regions based on T . Moreover, the random variable T tends to be stochastically increasing in σ_j^2 , because it follows from (16) that $\tilde{S}()$ is a stochastically decreasing function of σ_j^2 . Hence, the generalized p-value for testing H_0 can be obtained as

$$\begin{aligned} p &= \max \Pr(T \leq 1) = \Pr \left(V_j \leq \tilde{s} \left(\frac{s_e}{U}, \sigma_0^2 \right) \right) \\ &= EG_a \left(\tilde{s} \left(\frac{s_e}{U}, \sigma_0^2 \right) \right), \end{aligned} \quad (21)$$

where G_a is the CDF of the chi-squared distribution with a degrees of freedom and the expectation is taken with respect to the random variable $U \sim \chi_e^2$.

The generalized confidence intervals for the variance component σ_j^2 can be derived from the GPQ (19) or deduced directly from the p-value in (21). In the latter approach, if σ_1^2 and σ_2^2 are chosen such that

$$EG_a \left(\tilde{s} \left(\frac{s_e}{U}, \sigma_1^2 \right) \right) = \frac{1 - \gamma}{2}$$

and

$$EG_a \left(\tilde{s} \left(\frac{s_e}{U}, \sigma_2^2 \right) \right) = \frac{1 + \gamma}{2},$$

then $[\sigma_1^2, \sigma_2^2]$ is a $100\gamma\%$ generalized confidence interval for σ_j^2 .

4.2 Inference on sums of variances

To illustrate the approach one can take in dealing with various sums and ratios of variance components that arise in quality assurance applications (cf. Wu and Hamada (2009)) consider the particular problem of making inferences about quantities such as $\sigma_1^2 + \sigma_2^2$. Unlike simple in variance ratios, the problem is non-trivial even if one of the two variances is the error variance σ_e^2 . So, in order to illustrate the approach that one can take in dealing with sums of variance components, consider the particular problem of constructing interval estimates for the sum $\sigma_s^2 = \sigma_e^2 + \sigma_j^2$

It follows from (18) and (19) that

$$G_s = G_{\sigma_j^2} + G_{\sigma_e^2} = \frac{s_e}{U} \left(1 + s^{-1} \left(\frac{s_e V_j}{U} \right) \right)$$

is a potential GPQ for the sum. It is indeed a GPQ for σ_s^2 , because it has the required properties:

- The distribution of G_s is free of unknown parameters
- At observed sample points, G_s depends only on the parameter of interest.

Moreover, at observed sample points G_s reduces to σ_s^2 . Therefore, 100% generalized confidence intervals on σ_s^2 can be constructed by first finding quantiles of the GPQ

$$Pr(G_s \leq q_\gamma) = Pr \left(\frac{s_e}{U} \left(1 + s^{-1} \left(\frac{s_e V_j}{U} \right) \right) \leq q_\gamma \right),$$

and obtaining one-sided interval estimates as $\sigma_s^2 \leq q_\gamma$. The quantiles can be easily computed by exact numerical integration. The generalized confidence can be computed more conveniently using Monte Carlo integration. For example, 95% generalized confidence intervals for σ_s^2 can be obtained in the following steps:

- Generate a large set of random numbers from the independent chi-squared random variables (U_i, V_i) , say $i = 1, 2, \dots, 100000$ pairs,
- For each pair, compute $G_i = G_s(U_i, V_i)$, $i = 1, 2, \dots, 100000$, from $U_i \sim \chi_e^2$ and $V_i \sim \chi_a^2$,
- Sort the G_i samples and obtain the order statistic as $G_{(1)}, G_{(2)}, \dots, G_{(100000)}$,
- Obtain the 95% lower confidence bound for σ_s^2 as $G_{(5000)}$,
- Obtain the 95% upper confidence bound for σ_s^2 as $G_{(95000)}$,
- Obtain the 95% equal-tail confidence interval for σ_s^2 as $(G_{(2500)}, G_{(97500)})$.

Testing of hypotheses can also be carried out using the above interval estimates. For example, the null hypotheses $H_0 : \sigma_s^2 = \delta$ is rejected at 0.05 level of significance if the interval estimate, $(G_{(2500)}, G_{(97500)})$ does not contain the hypothesized value δ .

4.3 Inference on fixed effects

Next consider the problem of making inferences on some or all fixed effects in the presence of random effects. Confidence intervals and tests available from widely used software packages such as SAS PROC MIXED and R lme are based on asymptotic distributions of the REML and the ML of variance components. Unlike inferences directly on variance components, these methods are not too bad. But since fixed effects also depend on variance components to some extent, the purpose of this section to show how one can make inferences based on exact probability statements.

If variance components were known parameters, then we can make inferences about β based on its GLSE (Generalized Least Squares Estimator) specified by (4), which is distributed as

$$\hat{\beta} \sim N(\beta, \sigma_e^2 \Gamma), \text{ where } \Gamma = \mathbf{I}_N + \mathbf{Z}\Sigma_\rho\mathbf{Z}'. \quad (22)$$

But the issue of concern is that, while it is trivial to handle the error variance σ_e^2 , the classical approach fails to provide satisfactory inferences when the among group variances are unknown, as usually the case. The inference problem is not simple because $\hat{\beta}$ is itself a function of nuisance parameters, namely

$$\hat{\beta} = (\mathbf{X}'\Gamma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Gamma^{-1}\mathbf{y}.$$

There is of course no issue in point estimation. In fact, simple LSE based point estimates are unbiased estimates, which are as good as point estimates provided by the REML and the ML methods. The problem is making inferences about components of β beyond the point estimation is not trivial. So we will next develop generalized inference methods to tackle β .

To show how the generalized inferences based on exact probability statements can be developed based on (22), let us consider the case where we have one factor between group variance σ^2 . In this case the covariance matrix Γ appearing in (4) reduces to $\Gamma = \mathbf{I}_N + \rho\mathbf{Z}\mathbf{Z}'$, where $\rho = \sigma^2/\sigma_e^2$. Consider the problem of testing individual components of β , results of which can be easily extended to test a vector of fixed effects. In generalized approach to inference such nuisance parameters as ρ can be tackled by its GPQ, the random quantity

$$R = R(U, V) = \max\left(0, s^{-1}\left(\frac{s_e S(\rho)\sigma_e^2}{\sigma_e^2 S_e}\right)\right) = \max\left(0, s^{-1}\left(\frac{s_e V}{U}\right)\right), \quad (23)$$

which reduces to ρ at the observed set of sample points. In Section 4 we saw that this GPQ yields that same inferences on ρ as the classical PQ, but as we see below, it has greater potential when treated as a GPQ. Now we can make inferences on the β vector and its components based on the result (23) using the resulting GPQ

$$\beta_R = (\mathbf{X}'\Gamma_R^{-1}\mathbf{X})^{-1}\mathbf{X}'\Gamma_R^{-1}\mathbf{y},$$

where $\Gamma_R = \mathbf{I}_N + R\mathbf{Z}\mathbf{Z}'$.

To illustrate the steps in making such generalized inference, consider the specific problem of testing hypotheses of the form

$$H_{01} : \beta_i \geq \beta_0, \text{ or } H_{02} : \beta_i \leq \beta_0, \quad (24)$$

where β_0 is a specified constant with $\beta_0 = 0$ in (24) being the most widely used situation of testing the significance of a variable of special interest such as the efficacy of a drug. If there were no random effects, we would have used the t-test suggested by the classical regression theory to perform the test. To develop the counterpart in the presence of a random effects, let us use the distributional result

$$Z = \frac{\hat{\beta}_i(\rho) - \beta_i}{\sigma_e \sqrt{(\mathbf{X}'\Gamma^{-1}\mathbf{X})_i}} \sim N(0, 1), \quad (25)$$

where $\hat{\beta}_i(\rho)$ is the i^{th} component of $\beta(\rho)$ corresponding to the parameter β_i being tested. This result along with (6) implies that

$$T = \frac{\hat{\beta}_i(\rho) - \beta_i}{S_e \sqrt{(\mathbf{X}'(\mathbf{I}_N + \rho \mathbf{Z}\mathbf{Z}')^{-1} \mathbf{X})_i}} \sim t_e,$$

a result we could have used if the variance ratio ρ were known. Moreover, recall that we can tackle the ρ parameter using the distribution results (6) and (15), namely

$$U = \frac{S_e}{\sigma_e^2} \sim \chi_e^2, \text{ where } e = N - \text{rank}(\tilde{\mathbf{X}}) \text{ and}$$

$$V = \frac{S(\rho)}{\sigma_e^2} \sim \chi_a^2, \text{ where } S(\rho) = \hat{\mathbf{u}}' \left((\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} + \rho \mathbf{I}_l \right)^{-1} \hat{\mathbf{u}}_j.$$

in the case of one grouping structure with l groups.

Since ρ is unknown in practice, the above result leads to an approximate t-test only. However, we can derive a generalized test based on exact probability statement using forgoing distributional results. In view of the form of the T statistic in the known ρ case, it is evident that the unknown variance ratio can be tackled by using the potential generalized test variable

$$\begin{aligned} \tilde{T} &= \frac{\sigma_e}{\sqrt{S_e(\hat{\beta}_i(R) - \beta_i)}} \frac{(\hat{\beta}_i(\rho) - \beta_i)}{\sigma_e \sqrt{(\mathbf{X}'(\mathbf{I}_N + \rho \mathbf{Z}\mathbf{Z}')^{-1} \mathbf{X})_i}} \sqrt{(\mathbf{X}'(\mathbf{I}_N + R\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{X})_i} \\ &= \frac{Z}{\sqrt{U}(\hat{\beta}_i(R) - \beta_i)} \sqrt{(\mathbf{X}'(\mathbf{I}_N + R\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{X})_i}, \end{aligned} \tag{26}$$

where R is a function of U and V , and $Z \sim N(0, 1)$, $U \sim \chi_e^2$, and $V \sim \chi_a^2$ are mutually independent random variables.

It is clear from the second expression of (26) that the distribution of \tilde{T} is free of unknown parameter, and from the first expression it is clear that at the observed set of sample points, the test variable reduces to $1/\sqrt{s_e}$. Moreover the distribution of \tilde{T} is stochastically monotonic in β_i . Therefore, the hypotheses of the form (24) can be tested based on the generalized p-value

$$\begin{aligned} p &= Pr \left(\frac{Z}{\sqrt{U}} \sqrt{(\mathbf{X}'(\mathbf{I}_N + R\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{X})_i} \leq \frac{(\hat{\beta}_i(R) - \beta_0)}{\sqrt{s_e}} \right) \\ &= E \left[\Phi \left(\frac{(\hat{\beta}_i(R) - \beta_0) \sqrt{U}}{\sqrt{s_e(\mathbf{X}'(\mathbf{I}_N + R\mathbf{Z}\mathbf{Z}')^{-1} \mathbf{X})_i}} \right) \right]. \end{aligned} \tag{27}$$

or $1 - p$, where the expectation in (27) is taken with respect to the chi-squared random variable U and V , and $S(\rho)$ is given by (15).

5 Simulation study

The purpose of this section is to conduct a Monte-Carlo simulation study to study the performance of the proposed method compared with the widely used MLE based methods that are available from such software packages as SAS Proc Mixed and R lme. Weerahandi (2004, Chapter 4) provides simulated results showing serious Type I Error issues of ML and REML based tests in the context of random effects ANOVA. Whereas, in our simulation study concerning the regression case, we considered 3,5 and 10 groups and error variance is fixed at 1. Though not necessary, to minimize the extent of the simulation, sample size of each group is assumed to be equal and set at values $n = 10$ and

100. In each of the studies below, we also included one X variable with fixed effect 10 is included when X takes on normal random numbers with mean 0 and variance 0.1. Moreover, except when more than one random effect is necessary, the Z is set to be taken as normal random numbers with mean 1 and variance 1. In typical applications where mixed effects models add value over fixed effects models, the factor variance tends to be smaller than the error variance, and so we fixed the latter was set at 1 and the former was varied to have standard deviations at 0.01, 0.1, and 1.

5.1 Testing the variance component

The literature provides comparison of tests available for testing variance components basically for ANOVA type models only. Here we compare the performance of the proposed generalized test in Section 4 against the two methods available from the R software package. One widely used test is an REML based method available from the R lme function with the “interval” procedure, which is an improved REML based Wald-type confidence interval. The other competing method is the PL (Profile likelihood) based confidence interval available from R lmer with the “confint” procedure.

Table 1 below shows the Type-I error of alternative tests for testing hypotheses of the form $H_0 : \sigma_0^2 \leq \sigma_A^2$. When the sample size and the factor variance are both large, REML method fails due to non-convergence, and so such cases are indicated as NA. It is evident from Table 1 that the proposed test outperforms both the competing methods under all scenarios. The REML method seriously misses the intended size of tests when the factor variance is small and become conservative when the factor variance is large, whereas the PL method is too conservative in all cases. The proposed method also becomes somewhat conservative when the factor variance is also large. The performance of likelihood based methods somewhat improve as the number of factor levels increases.

Table 1 Type I error of competing tests of $H_0 : \sigma_A^2 \geq \sigma_0^2$ with intended size 0.05

n_g	n	σ_0^2	REML	PL	Proposed
3	10	.01	.254	.014	.043
3	10	.1	.082	.020	.046
3	10	1	.006	.010	.046
3	100	.01	.168	.012	.051
3	100	.1	.011	.008	.054
3	100	1	NA	.006	.051
5	10	.01	.281	.021	.042
5	10	.1	.076	.020	.043
5	10	1	.013	.016	.044
5	100	.01	.177	.019	.044
5	100	.1	.022	.017	.044
5	100	1	NA	.016	.048
10	10	.01	.300	.031	.047
10	10	.1	.095	.023	.050
10	10	1	.022	.022	.050
10	100	.01	.208	.026	.050
10	100	.1	.034	.028	.050
10	100	1	NA	.023	.048

5.2 Testing fixed effects

The literature does not seem to provide any comparison of tests available for testing fixed effects of mixed effects regression models, perhaps thinking that their optimality is covered by the classical theory. Although the situation is not as serious as that of variance components, the assumption is valid only if the variance components are known. Therefore, here we compare the performance of the proposed generalized test in Section 5 against the two methods available from software packages such as SAS and R, namely the REML and ML based tests. Again since multiple covariates add no value in understanding the comparative performance competing tests, we consider one covariate X , whose fixed effect is denoted as β . The performance was found to vary somewhat with choice of x , and therefore, to understand the overall performance of competing tests for $H_0 : \beta = 0$, we generated X as a 50% mixture of X used in the previous section and one taking on equally spaced values from -1 to 1.

Table 2 below shows the performance of competing methods as the number of groups take on typical values 3, 5, and 10, whereas the sample size from each group takes on value 10 and 100. In each case, the performance of competing tests were studied as the factor variance σ_A vary from 0.1 to 1.0. It is evident from Table 2 that when number of groups is small, type I error of MLE based methods tends to be somewhat larger than the intended size of 0.05, whereas the propose method stay closer to the intended size. The ML test is worse than the REML test and could be as large as .084, though not as bad as tests on variance components. When the sample size is small, Type I error of each test tends to go up with the factor variance, whereas when the sample sizes are large Type I errors of RMEL and PL methods tend to go down. When the number of groups and the sample sizes are both large, all three methods behave practically the same.

Table 2 Type I error of competing tests of $H_0 : \beta = 0$ with intended size 0.05

n_g	n	σ_{A0}	REML	ML	Proposed
3	10	.1	.061	.060	.051
3	10	.5	.068	.079	.054
3	10	1	.072	.084	.060
5	10	.1	.058	.057	.048
5	10	.5	.060	.061	.051
5	10	1	.060	.063	.054
10	10	.1	.053	.055	.048
10	10	.5	.053	.056	.050
10	10	1	.056	.056	.055
3	100	.1	.071	.080	.058
3	100	.5	.056	.062	.052
3	100	1	.054	.055	.052
5	100	.1	.070	.077	.065
5	100	.5	.055	.055	.050
5	100	1	.049	.049	.049
10	100	.1	.055	.058	.054
10	100	.5	.054	.055	.054
10	100	1	.054	.054	.052

6 Discussion

In this article we have shown how to make inferences in mixed effects models based on exact distributions of underlying statistics when the covariance structure is compound symmetric. The simulation studies undertaken in this article have shown the potential in developing tests and confidence intervals that are better than currently available asymptotic methods. Our results also show that there is potential for further improvement of inferences, especially when factor variances are too large or too small compared to the error variance.

Our tests and interval estimates are based on the generalized inference approach. Researchers are encouraged to develop alternative inferences by taking other promising approaches as Parametric Bootstrap and Fiducial inference with a view towards further improving the performance of tests, especially when factor variances are large compared to the error variance. Also desirable is to extend results to widely used and practically useful other covariance structures.

Gamage et al. (2013) developed generalized prediction intervals for BLUPs of mixed models. Further research necessary to extend such results to compare two BLUPS in the same model or BLUPs of two independent models. The results presented in this article should pave the way for such developments.

Abbreviations

ANOVA: Analysis of variance; BLUP: Best linear unbiased predictor; BLUE: Best linear unbiased estimator; CDF: Cumulative distribution function; GLSE: Generalized least squares estimator; GPQ: Generalized pivotal quantity; GTV: Generalized test variable; LSE: Least squares estimator; ML: Maximum likelihood; MLE: Maximum likelihood estimator; PQ: Pivotal quantity; QR: Q and R are the notations used for the two matrices in QR matrix decomposition; REML: Restricted maximum likelihood

Acknowledgements

The authors are grateful to the editors and referees who made very constructive comments and suggestions. Addressing them led to a substantially improved version of the manuscript. The authors also thank Professor Thomas Mathew of UMBC for reading the manuscript and making constructive suggestions.

Authors' contributions

Weerahandi developed the initial distributions after joint discussions with the co-author Yu. Co-author Yu, heavily contributed to the simulation studies to evaluate the Type I error performance of proposed solutions in this article compared with other solutions in the literature. All authors read and approved the final manuscript.

Authors' information

The first author of this article, Weerahandi, is the president of X-Techniques located at 23 Chestnut St, Edison, NJ 08817. The second author is a director of Pfizer located at 235 East 42nd Street, New York, NY 10017.

Funding

This research was not funded by any institution.

Availability of data and materials

This article dealing with a novel Distribution Theory does not have any Data to report.

Competing interests

Authors of this article do not have Competing interests of any kind, financially or otherwise.

Author details

¹X-Techniques, 23 Chestnut St, 08817 Edison, NJ, USA. ²Pfizer, 235 East 42nd St, New York, USA.

Received: 5 March 2020 Accepted: 31 July 2020

Published online: 17 August 2020

References

- Cisewski, J., Hannig, J.: Generalized fiducial inference for normal linear mixed models. *Ann. Stat.* **40**, 2102–2127 (2012)
- Gamage, J., Mathew, T., Weerahandi, S.: Generalized prediction intervals for BLUPs in mixed models. *J. Multivar. Anal.* **220**, 226–233 (2013)
- Hannig, J., Iyer, H., Patterson, P.: Fiducial generalized confidence intervals. *J. Am. Stat. Assoc.* **101**, 254–269 (2006)
- Hamada, M., Weerahandi, S.: Measurement system assessment via generalized inference. *J. Qual. Technol.* **32**, 241–253 (2000)

- Henderson, C. R.: Best linear unbiased estimation and prediction under a selection model. *Biometrics*. **31**, 423–447 (1975)
- Robinson, G. K.: That BLUP is a good thing: the Estimation of random effects. *Stat. Sci.* **6**, 15–32 (1991)
- Searle, S. R., Casella, G., McCulloch, C. E.: *Variance Components*. Wiley, Hoboken (2006)
- Seely, J. F., El-Bassiouni: Applying Wald's variance component test. *Ann. Stat.* **11**, 197–201 (1983)
- Tsui, K., Weerahandi, S.: Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *J. Am. Stat. Assoc.* **84**, 602–607 (1989)
- Wald, A.: A note on regression analysis. *Ann. Math. Stat.* **18**, 586–589 (1947)
- Weerahandi, S.: Testing variance components in mixed models with generalized p-values. *J. Am. Stat. Assoc.* **86**, 151–153 (1992)
- Weerahandi, S.: Generalized confidence intervals. *J. Am. Stat. Assoc.* **88**, 899–905 (1993)
- Weerahandi, S.: *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models*. Wiley, Hoboken (2004)
- Weerahandi, S., Gamage, J. A general method of inference for two-parameter continuous distributions. *Communications in Statistics (T&M)*. **45**, 2612–2625 (2016)
- Wu, J., Hamada, M. S.: *Experiments: Planning, Analysis, and Optimization*. Wiley, Hoboken (2009)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
