Published in final edited form as:

J Stat Plan Inference. 2009 December 1; 139(12): 3962–3973. doi:10.1016/j.jspi.2009.03.014.

Estimation of Variance Components in the Mixed-Effects Models: A Comparison Between Analysis of Variance and Spectral Decomposition

Mi-Xia Wu^{a,b,*}, Kai-Fun Yu^b, and Ai-Yi Liu^b

^aCollege of Applied Sciences, Beijing University of Technology, Beijing 100022, P.R. China

^bDivision of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH/DHHS, 6100 Executive Boulevard, Rockville, MD 20852, U.S.A.

Abstract

The mixed-effects models with two variance components are often used to analyze longitudinal data. For these models, we compare two approaches to estimating the variance components, the analysis of variance approach and the spectral decomposition approach. We establish a necessary and sufficient condition for the two approaches to yield identical estimates, and some sufficient conditions for the superiority of one approach over the other, under the mean squared error criterion. Applications of the methods to circular models and longitudinal data are discussed. Furthermore, simulation results indicate that better estimates of variance components do not necessarily imply higher power of the tests or shorter confidence intervals.

Keywords

Analysis of variance; Best linear unbiased estimate; Least squares estimate; Mixed-effects model; Spectral decomposition

1 Introduction

In the past two decades, the mixed-effects linear model has received considerable attention from both theoretical and practical points of view due to its extensive applications, for example, in analyzing longitudinal data arising from the health sciences, computer graphics and mechanical engineering. For a comprehensive overview of this model, see Davidian and Giltinan (1996), Diggle, *et al.* (2002) and Demidenko (2004).

Consider the mixed-effects linear model with two variance components

$$y_{it} = x'_{it}\beta + u_i + \varepsilon_{it}, i = 1, \dots, N, t = 1, \dots, T,$$

© 2009 Elsevier B.V. All rights reserved.

*Corresponding author. E-mail address: wumixia@bjut.edu.cn.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

where β is $p \times 1$ vector of fixed effects, u_i is random effect, u_i and ε_{it} are assumed to be independent and normally distributed with mean 0 and variance σ_u^2 and σ_ε^2 , respectively. Writing this model in matrix form, we have

$$y = X\beta + Zu + \varepsilon, \quad (1)$$

where y and X are of dimensions $n \times 1$ and $n \times p$, respectively, $n = NT$, $Z = I_N \otimes \mathbf{1}_T$, $\otimes$ denotes the Kronecker product, I_N is the identity matrix of order N , and $\mathbf{1}_T$ is a vector of ones of dimension T , $u = (u_1, u_2, \dots, u_N)'$ and $\varepsilon = (\varepsilon_{11}, \dots, \varepsilon_{1T}, \dots, \varepsilon_{N1}, \dots, \varepsilon_{NT})'$. Thus the covariance matrix of y is given by

$$\text{Cov}(y) = V(\sigma^2) = \sigma_u^2 Z Z' + \sigma_\varepsilon^2 I_n. \quad (2)$$

Analysis of variance is one of the popular methods in the statistical literature to estimate the variance components σ_u^2 and σ_ε^2 . Compared to other estimates such as the maximum likelihood estimate (MLE), restricted maximum likelihood estimate (REMLE), minimum norm quadratic unbiased estimate (MINQUE), the analysis of variance estimate (ANOVAE) has simple closed forms that make it feasible to study analytically its small sample behavior and construct exact tests and confidence intervals for the variance components based on the estimates; see, e.g. Wald (1940), Searle, *et al.* (1992), and Burch and Iyer (1997). In contrast, when the parameter lies in the boundary of the parametric space, the asymptotic distributions of the likelihood-based tests, such as the likelihood ratio tests, do not follow chi-square distributions, as showed in Stram and Lee (1994) and Crainiceanu and Ruppert (2004). Moreover, solutions based on maximum likelihood (ML) estimates and restricted maximum likelihood (REML) estimates usually have very poor performance in mixed models with relatively small sample size; see Burdick and Larsen (1997).

Exact tests and confidence intervals are often required in practice. For the two-variance-component mixed-effects model (1), a new approach, referred to as the spectral decomposition, to estimating the variance components, was recently developed by Wang and Yin (2002). The resulting estimate, the spectral decomposition estimate (SDE), possesses similar advantages as the ANOVAE, having a simple closed form and allowing the construction of exact tests and confidence intervals.

The present paper serves as the first in the literature to compare analytically the small sample properties of ANOVAE and SDE under model (1). In Section 2 we introduce the ANOVAE and SDE of the variance components. A necessary and sufficient condition is derived in Section 3 under which the two methods yield identical estimates. In Section 4, a relationship between the two estimates is established, and some conditions are found under which one estimate is superior to the other in terms of mean squared error (MSE). Exact test and confidence intervals based on the two estimates are constructed in Section 5. Applications to two special models are given in Section 6. Section 7 presents some simulation results and some discussions are given in Sections 8.

2 ANOVA and spectral decomposition

Denote by A^- a generalized inverse of a matrix A , and write $P_A = A(A'A)^-A'$. Then P_A is the orthogonal projector onto $\mathcal{M}(A)$, the column space of a matrix A . Let $Q_A = I - P_A$. Using Henderson Method III (Henderson, 1953 or Searle, *et al.*, 1992), we can obtain the ANOVAE of σ_ε^2 and σ_u^2 as

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n - r_0} y'(I - P_{(X:Z)})y, \quad (3)$$

$$\begin{aligned} \hat{\sigma}_u^2 = & \{y'(P_{(X:Z)} \\ & - P_X)y \\ & - (r_0 - \text{rk}(X))\hat{\sigma}_\varepsilon^2\} \\ & / \{T\text{tr}(Q_X P_Z)\}, \end{aligned} \quad (4)$$

where $(X : Z)$ is a $n \times (p + N)$ matrix consisting of the column vectors of the matrices X and Z , $\text{rk}(A)$ and $\text{tr}(A)$ stand for the rank and trace of a matrix A , respectively, and $r_0 = \text{rk}(X : Z)$.

Note that the covariance matrix of y can be decomposed as

$$\text{Cov}(y) = \lambda P_Z + \sigma_\varepsilon^2 Q_Z,$$

where $\lambda = T\sigma_u^2 + \sigma_\varepsilon^2$. Left-multiplying model (1) by P_Z and Q_Z , respectively, gives

$$\begin{aligned} P_Z y &= P_Z X\beta + \epsilon_1, \epsilon_1 \sim N(0, \lambda P_Z), \\ Q_Z y &= Q_Z X\beta + \epsilon_2, \epsilon_2 \sim N(0, \sigma_\varepsilon^2 Q_Z), \end{aligned}$$

both being singular linear models. From the unified theory of least squares (Rao, 1973), we can use

$$\tilde{\sigma}_\varepsilon^2 = y'(Q_Z - Q_Z X(X'Q_Z X)^{-1}X'Q_Z)y/b, \quad (5)$$

$$\tilde{\lambda} = y'(P_Z - P_Z X(X'P_Z X)^{-1}X'P_Z)y/m, \quad (6)$$

where $m = N - \text{rk}(P_Z X)$ and $b = \text{rk}(Q_Z) - \text{rk}(Q_Z X)$, to estimate σ_ε^2 and λ , respectively. An estimate of σ_u^2 is then given by

$$\tilde{\sigma}_u^2 = (\tilde{\lambda} - \tilde{\sigma}_\varepsilon^2)/T. \quad (7)$$

Wang and Yin (2002) derived the two estimates $\tilde{\sigma}_\varepsilon^2$ and $\tilde{\sigma}_u^2$ and termed them as the spectral decomposition estimates of σ_ε^2 and σ_u^2 , respectively. It is easy to verify (Craig's Theorem, see Rao, 1973) that the two quadratic forms, $y'(P_Z - P_Z X(X'P_Z X)^{-1}X'P_Z)y$ and $y'(Q_Z - Q_Z X(X'Q_Z X)^{-1}X'Q_Z)y$, are independent, and distributed as $\lambda\chi_m^2$ and $\sigma_\varepsilon^2\chi_b^2$, respectively, where χ_k^2 denotes a χ^2 variable with k degrees of freedom. Using this we can construct exact tests and confidence intervals for σ_ε^2 or σ_u^2 based on these quadratic forms; see Section 4 for details.

3 Conditions for equality

We first notice that the ANOVA and spectral decomposition approach lead to the same estimate for σ_ε^2 , as stated below.

Theorem 3.1

Under model (1) the ANOVAE and SDE of σ_ε^2 are identical, that is,

$$\widehat{\sigma}_\varepsilon^2 = \tilde{\sigma}_\varepsilon^2.$$

Proof—It follows from $\mathcal{M}(A : B) = \mathcal{M}(A : Q_A B)$ and $A'Q_A B = 0$ that

$$P_{(A:B)} = P_A + Q_A B(B'Q_A B)^- B'Q_A, \quad (8)$$

and

$$\text{rk}(A:B) = \text{rk}(A) + \text{rk}(Q_A B)$$

for any matrices A and B with the same number of rows. Thus

$$\begin{aligned} Q_Z - Q_Z X(X'Q_Z X)^- X'Q_Z &= I - (P_Z + Q_Z X(X'Q_Z X)^- X'Q_Z) = I - P_{(X:Z)}, \\ b = \text{rk}(Q_Z) - \text{rk}(Q_Z X) &= n - r_0. \end{aligned}$$

Therefore $\widehat{\sigma}_\varepsilon^2$ and $\tilde{\sigma}_\varepsilon^2$

However, the estimates of σ_u^2 from the ANOVA and spectral decomposition can be quite different. We derive below conditions under which the two estimates, $\tilde{\sigma}_u^2$ and $\widehat{\sigma}_u^2$, are identical. Write $v = y - X\beta$,

$$\begin{aligned} A_0 &= \left\{ (P_{(X:Z)} - P_X) - \frac{(r_0 - \text{rk}(X))(I - P_{(X:Z)})}{n - r_0} \right\} \Bigg/ \{T \text{tr}(Q_X P_Z)\}, \\ A_1 &= \left\{ \frac{P_Z - P_{ZX}(X'P_Z X)^- X'P_Z}{N - \text{rk}(P_Z X)} - \frac{Q_Z - Q_Z X(X'Q_Z X)^- X'Q_Z}{n - r_0} \right\} \Bigg/ T. \end{aligned}$$

Clearly, both $\tilde{\sigma}_u^2$ and $\widehat{\sigma}_u^2$ are unbiased estimators of σ_u^2 , and can be reexpressed as $\widehat{\sigma}_u^2 = v'A_0 v$, $\tilde{\sigma}_u^2 = v'A_1 v$, where $v \sim N(0, V(\sigma^2))$. Using the fact (Wang and Chow, 1994) that for any known symmetric matrix A , $\text{Var}(v'A v) = 2\text{tr}(A V A V)$, we have

$$\begin{aligned} \text{Var}(\widehat{\sigma}_u^2) &= 2 \left\{ \sigma_\varepsilon^4 \frac{(n - \text{rk}(X))(r_0 - \text{rk}(X))}{T^2 [\text{tr}(Q_X P_Z)]^2 (n - r_0)} + \sigma_\varepsilon^2 \sigma_u^2 \frac{2}{T \text{tr}(Q_X P_Z)} \right. \\ &\quad \left. + \sigma_u^4 \frac{\text{tr}(Q_X P_Z)^2}{[\text{tr}(Q_X P_Z)]^2} \right\}, \end{aligned} \quad (9)$$

and

$$\text{Var}(\tilde{\sigma}_u^2) = 2 \left\{ \sigma_\varepsilon^4 \frac{(n+N-r_0-\text{rk}(P_Z X))}{T^2 [N-\text{rk}(P_Z X)](n-r_0)} + \sigma_\varepsilon^2 \sigma_u^2 \frac{2}{T [N-\text{rk}(P_Z X)]} + \sigma_u^4 \frac{1}{N-\text{rk}(P_Z X)} \right\}. \quad (10)$$

We will need the following two lemmas to prove the main results.

Lemma 3.1

(Wang and Chow, 1994). Let $P = P_A P_B$. Then

- i. P is an orthogonal projection matrix if and only if $P_A P_B = P_B P_A$.
- ii. If $P_A P_B = P_B P_A$, then P is the orthogonal projection matrix onto $\mathcal{M}(A) \cap \mathcal{M}(B)$.

Lemma 3.2

The following statements are equivalent.

- i. $P_A P_B = P_B P_A$
- ii. $\mathcal{M}(A) \cap \mathcal{M}(B) = \mathcal{M}(P_B A)$
- iii. $\text{rk}(P_B A) = \dim(\mathcal{M}(A) \cap \mathcal{M}(B))$
- iv. $P_B A (A' P_B A)^- A' P_B = P_A P_B$

where $\dim(\cdot)$ denotes the dimension of a space.

Proof—Note that for any vector $\mathbf{c} \in \mathcal{M}(A) \cap \mathcal{M}(B)$, there exist vectors α and γ such that $\mathbf{c} = A\alpha = B\gamma$. It follows that $P_B P_A \mathbf{c} = P_B P_A A\alpha = P_B A\alpha = P_B B\gamma = B\gamma = \mathbf{c}$, yielding

$$\mathcal{M}(A) \cap \mathcal{M}(B) \subseteq \mathcal{M}(P_B P_A) \subseteq \mathcal{M}(P_B A). \quad (11)$$

If (i) holds, then $P_A (P_B A) = P_B P_A A = P_B A$, which is equivalent to $\mathcal{M}(P_B A) \subseteq \mathcal{M}(A)$. Note that $\mathcal{M}(P_B A) \subseteq \mathcal{M}(B)$. Thus $\mathcal{M}(P_B A) \subseteq \mathcal{M}(A) \cap \mathcal{M}(B)$. This together with (11) yield (ii).

Conversely, if (ii) is true, then by (11) we have $\mathcal{M}(A) \cap \mathcal{M}(B) = \mathcal{M}(P_B P_A)$, and $P_A (P_B P_A) = P_B P_A$. Thus $P_B P_A$ is symmetric, and (i) holds.

That (ii) and (iii) are equivalent is obvious from (11) and the fact that $\dim(\mathcal{M}(P_B A)) = \text{rk}(P_B A)$, and that (iv) and (i) are equivalent is a direct consequence of Lemma 3.1.

The two lemmas lead to the following theorem.

Theorem 3.2

$\text{Var}(\tilde{\sigma}_u^2) = \text{Var}(\tilde{\sigma}_u^2)$ if and only if $P_X P_Z$ is symmetric, i.e. $P_X P_Z = P_Z P_X$.

$$\sigma_u^2 \geq 0 \text{ and } \sigma_\varepsilon^2 > 0$$

Proof—Note that for any $\sigma_u^2 \geq 0$ and $\sigma_\varepsilon^2 > 0$, $\text{Var}(\tilde{\sigma}_u^2) = \text{Var}(\tilde{\sigma}_u^2)$ if and only if the following three equalities all hold:

- a.
$$\frac{(n - \text{rk}(X))(r_0 - \text{rk}(X))}{[\text{tr}(Q_X P_Z)]^2} = \frac{n+N - r_0 - \text{rk}(P_Z X)}{N - \text{rk}(P_Z X)},$$
- b. $\text{tr}(Q_X P_Z) = N - \text{tr}(P_X P_Z) = N - \text{rk}(P_Z X)$
- c. $[\text{tr}(Q_X P_Z)]^2 / [\text{tr}(Q_X P_Z)]^2 = 1 / (N - \text{rk}(P_Z X)).$

Write $r_1 = \dim(\mathcal{M}(X) \cap \mathcal{M}(Z))$. Then $r_0 = \text{rk}(X) + \text{rk}(Z) - r_1$. Since $\text{rk}(Z) = N$, we have

$$\begin{aligned} n+N - r_0 - \text{rk}(P_Z X) \\ = n - \text{rk}(X) \\ + r_1 - \text{rk}(P_Z X). \end{aligned}$$

Thus the three conditions (a)–(c) are equivalent to

$$\begin{aligned} [\text{tr}(Q_X P_Z)]^2 \\ = \text{tr}(Q_X P_Z) = N - \text{rk}(P_Z X) = N - r_1. \end{aligned} \quad (12)$$

Note that the last equality above is equivalent to $\text{rk}(P_Z X) = r_1$. Hence by Lemma 3.2 the last equality of (12) holds if and only if $P_X P_Z = P_Z P_X$. On the other hand, if $P_X P_Z = P_Z P_X$, then the first two equalities of (12) hold. Theorem 3.2 follows.

We now give the main results.

Theorem 3.3

$\hat{\sigma}_u^2 = \hat{\sigma}_{u'}^2$ if and only if $P_X P_Z = P_Z P_X$.

Proof—Note that $P_X P_Z = P_Z P_X$ if and only if $Q_X P_Z = P_Z Q_X$. By Lemma 3.2, if $P_X P_Z = P_Z P_X$, then

$$\begin{aligned} P_Z X(X' P_Z X)^- X' P_Z \\ = P_X P_Z, Q_X Z(Z' Q_X Z)^- Z' Q_X = Q_X P_Z. \end{aligned}$$

Using (8) we have

$$P_{(X:Z)} - P_X = Q_X P_Z = P_Z - P_Z X(X' P_Z X)^- X' P_Z.$$

Thus $\hat{\sigma}_u^2 = \hat{\sigma}_{u'}^2$

On the other hand, if $\hat{\sigma}_u^2 = \hat{\sigma}_{u'}^2$, then $\text{Var}(\hat{\sigma}_u^2) = \text{Var}(\hat{\sigma}_{u'}^2)$. By Theorem 3.2, we have $P_X P_Z = P_Z P_X$. This completes the proof.

Remark 3.1

Both $\hat{\sigma}_u^2$ and $\hat{\sigma}_{u'}^2$ are uniformly minimum variance unbiased (UMVU) estimates of σ_u^2 if $P_X P_Z = P_Z P_X$, see Wu and Wang (2004).

We exemplify the main results with two popular mixed-effects models, both satisfying the identity condition $P_X P_Z = P_Z P_X$, and thus the ANOVA and spectral decomposition yield the same estimates for σ_u^2 .

Example 3.1

The one-way classification model has the form

$$y_{ij} = \mu + u_i + \varepsilon_{ij}, i=1, \dots, a, j=1, \dots, b,$$

where μ is a fixed parameter, u_i is a random effect, u_i and ε_{ij} are independent with $u_i \sim N(0, \sigma_u^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. Here $X = \mathbf{1}_{ab}$, $Z = I_a \otimes \mathbf{1}_b$. Thus $P_X P_Z = P_Z P_X = P_X = J_{ab}$, where $\bar{J}_t = \mathbf{1}_t, \mathbf{1}_t' / t$.

Example 3.2

The two-way classification model is given by

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \\ i=1, 2, \dots, a, j=1, 2, \dots, b,$$

where α_i is a random effect, β_j is a fixed effect. We assume that $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, and all α_i s and ε_{ij} s are mutually independent.

In matrix form, the model can be expressed as

$$y = X\gamma + Z\alpha + \varepsilon,$$

where $\gamma = (\mu, \beta_1, \dots, \beta_b)'$, $\alpha = (\alpha_1, \dots, \alpha_a)'$, and $X = (\mathbf{1}_a \otimes \mathbf{1}_b : \mathbf{1}_a \otimes I_b)$, $Z = I_a \otimes \mathbf{1}_b$. It is easy to verify that

$$P_X P_Z = P_Z P_X = (\bar{J}_a \otimes I_b) \cdot (I_a \otimes \bar{J}_b) \\ = \bar{J}_a \otimes \bar{J}_b.$$

4 Conditions for superiority

An important question to be answered concerning ANOVAE and SDE is, “If $P_X P_Z \neq P_Z P_X$, then which estimator is better and in what sense?”. In this section, we obtain conditions under which one estimator has smaller mean squared error than the other. Since both estimators are unbiased, their mean squared error is identical to their variances. To this end, let us express the spectral decomposition of the matrix $Q_X Z Z' Q_X$ as

$$Q_X Z Z' Q_X = \sum_{i=1}^g d_i M_i, \quad (13)$$

where $d_1, \dots, d_g$ are the different nonzero eigenvalues of the matrix $Q_X Z Z' Q_X$, M_i is the orthogonal projector onto the corresponding eigenvalue subspace. Note that $g > 1$ because $P_X P_Z \neq P_Z P_X$.

Using the facts that $M_i^2 = M_i$, $M_i M_j = 0 (i \neq j)$,

$$\begin{aligned} P_{(X:Z)} - P_X &= Q_X Z (Z' Q_X Z)^{-1} Z' Q_X \\ &= P_{Q_X Z} = \sum_{i=1}^g M_i, \end{aligned}$$

we obtain

$$y'(P_{(X:Z)} - P_X)y = \sum_{i=1}^g y'M_i y, \quad (14)$$

with $y'M_i y \sim (d_i \sigma_u^2 + \sigma_\varepsilon^2) \chi_{m_i}^2$, where $m_i = \text{tr}(M_i)$, and $y'M_1 y, \dots, y'M_g y$ are independent.

Since

$$\begin{aligned} Q_X Z Z' Q_X (P_Z \\ - P_Z X (X' P_Z X)^{-1} X' P_Z) &= T (P_Z \\ - P_Z X (X' P_Z X)^{-1} X' P_Z), \end{aligned}$$

T is a nonzero eigenvalue of $Q_X Z Z' Q_X$ and $P_Z - P_Z X (X' P_Z X)^{-1} X' P_Z$ is the orthogonal projector onto the eigenvalue subspace corresponding to T . Without loss of generality, we assume that

$$\begin{aligned} M_1 &= P_Z - P_Z X (X' P_Z X)^{-1} X' P_Z \\ d_1 &= T, m_1 = m = N - \text{rk}(P_Z X), \end{aligned}$$

and

$$\begin{aligned} \hat{\sigma}_u^2(i) &= (y'M_i y \\ &\quad / (m_i - \hat{\sigma}_\varepsilon^2)) / d_i, \quad i=1, \dots, g. \end{aligned}$$

It follows straightforwardly that $\hat{\sigma}_u^2(1) = \hat{\sigma}_u^2$ and each $\hat{\sigma}_u^2(i)$ is an unbiased estimate of σ_u^2 . Combining (4) and (14), we obtain the following theorem.

Theorem 4.1

The ANOVA of σ_u^2 is a weighted average of $\{\hat{\sigma}_u^2(i): i=1, \dots, g\}$, that is

$$\hat{\sigma}_u^2 = T m_1 \hat{\sigma}_u^2 / \Delta + \sum_{i=2}^g d_i m_i \hat{\sigma}_u^2(i) / \Delta, \quad (15)$$

where $\Delta = T \text{tr}(Q_X P_Z) = \sum_{i=1}^g d_i m_i$.

Thus the SDE $\hat{\sigma}_u^2$ under model (1), in fact, is an unbiased estimate deduced by only one quadratic form of $y'M_1 y, \dots, y'M_g y$.

In the following, we will compare the variance of $\widehat{\sigma}_u^2$, the weighted average of $\{\widehat{\sigma}_u^2(i): i=1, \dots, g\}$, with that of $\tilde{\sigma}_u^2$ when $P_X P_Z \neq P_Z P_X$. Recall Eq. (9) and Eq. (10), we have

Theorem 4.2

The corresponding coefficients of $\sigma_u^2 \sigma_\varepsilon^2$ and σ^4 in $\text{Var}(\tilde{\sigma}_u^2)$ are larger than their counterparts in $\text{Var}(\widehat{\sigma}_u^2)$ if $P_X P_Z \neq P_Z P_X$.

Proof—Using the facts that

$$\begin{aligned} \text{rk}(P_Z X) &= \text{tr}(P_Z X (X^\top P_Z X)^{-1} X^\top P_Z) \geq \text{tr}(P_Z P_X P_Z) \\ &= \text{tr}(P_X P_Z), \end{aligned}$$

and

$$\begin{aligned} [\text{tr}(Q_X P_Z)]^2 &= \text{tr}(Q_X P_Z Q_X P_Z Q_X) \leq \text{tr}(Q_X P_Z Q_X) = \text{tr}(Q_X P_Z) \\ &= N - \text{tr}(P_X P_Z), \end{aligned}$$

we have

$$\frac{[\text{tr}(Q_X P_Z)]^2}{[\text{tr}(Q_X)]^2} \leq \frac{1}{\text{tr}(Q_X P_Z)} \leq \frac{1}{N - \text{rk}(P_Z X)},$$

and the equalities above hold if and only if $P_X P_Z = P_Z P_X$ by Theorem 3.2. Theorem 4.2 thus follows.

Note that

$$\begin{aligned} r_0 - \text{rk}(X) &= N - \dim(\mathcal{M}(X) \cap \mathcal{M}(Z)), \\ n + N - r_0 - \text{rk}(P_Z X) &= n - \text{rk}(P_Z X : Q_Z X), \end{aligned}$$

we obtain the ratio of the coefficient of σ_ε^4 in $\text{Var}(\widehat{\sigma}_u^2)$ to that in $\text{Var}(\tilde{\sigma}_u^2)$:

$$\begin{aligned} & \frac{\frac{(n - \text{rk}(X))(r_0 - \text{rk}(X))}{[\text{tr}(Q_X P_Z)]^2} / \frac{(n + N - r_0 - \text{rk}(P_Z X))}{N - \text{rk}(P_Z X)}}{\frac{n - \text{rk}(X)}{n - \text{rk}(P_Z X : Q_Z X)} \frac{N - \dim(\mathcal{M}(X) \cap \mathcal{M}(Z))}{N - \text{tr}(P_X P_Z)} \frac{N - \text{rk}(P_Z X)}{N - \text{tr}(P_X P_Z)}} \\ &= \frac{(n - \text{rk}(X))(r_0 - \text{rk}(X))}{[\text{tr}(Q_X P_Z)]^2} \frac{N - \dim(\mathcal{M}(X) \cap \mathcal{M}(Z))}{N - \text{tr}(P_X P_Z)} \frac{N - \text{rk}(P_Z X)}{N - \text{tr}(P_X P_Z)}. \end{aligned} \quad (16)$$

Because

$$\text{rk}(P_Z X) \geq \text{tr}(P_X P_Z) \geq \dim(\mathcal{M}(X) \cap \mathcal{M}(Z)) \geq 0,$$

and

$$n - \text{rk}(P_Z X : Q_Z X) \leq n - \text{rk}(X),$$

it follows that the first two fractions in the right-hand side of (16) are larger than or equal to 1, and the third fraction is no larger than 1. Combining these with Theorem 4.2 we have

Corollary 4.1

If the right-hand side of (16) is less than 1, then $\text{Var}(\hat{\sigma}_u^2) < \text{Var}(\tilde{\sigma}_u^2)$.

Remark 4.1

The ANOVAE $\hat{\sigma}_u^2$ is not uniformly superior to the SDE $\tilde{\sigma}_u^2$. If $(\sigma_e^2/T) \gg \sigma_u^2$ and the right-hand side of (16) > 1 , then $\text{Var}(\hat{\sigma}_u^2) > \text{Var}(\tilde{\sigma}_u^2)$.

Example 6.2 in Section 6 illustrates that (16) is not always less than 1. Note that

$(\text{Var}(\hat{\sigma}_u^2) - \text{Var}(\tilde{\sigma}_u^2)) / \sigma_e^2$ is a quadratic function of $\tau = T\sigma_u^2 / \sigma_e^2$ with the negative coefficients for the terms τ^2 and τ , and the common term is obtained by subtracting the denominator from the numerator of (16). Thus when the right-hand side of (16) is larger than 1, $\text{Var}(\hat{\sigma}_u^2) > \text{Var}(\tilde{\sigma}_u^2)$ for small τ and $\text{Var}(\hat{\sigma}_u^2) > \text{Var}(\tilde{\sigma}_u^2)$ for large τ ; see Example 6.2. That is, the SD estimate is preferred when the predictor varies considerably across subjects and the random effect variance and the number of repeated observations for each subject, T , are small.

5 Exact tests and confidence intervals

A common interest in model (1) is testing hypotheses and constructing confidence intervals concerning the relative magnitude, $\theta = \sigma_u^2 / \sigma_e^2$, of the variation due to the random effects and the random errors. To compare the two approaches under investigation, we assume that $P_X P_Z \neq P_Z P_X$ so the two methods yield different estimates of σ_u^2 .

Consider the hypothesis

$$H_0: \theta = \theta_0 \leftrightarrow H_1: \theta > \theta_0,$$

where $\theta_0 \geq 0$. Since the spectral estimates $\tilde{\lambda}$ and $\tilde{\sigma}_e^2$ in (5) and (6) are independent, with $m_1 \tilde{\lambda} / \lambda \sim \chi_{m_1}^2$, $b \tilde{\sigma}_e^2 / \sigma_e^2 \sim \chi_b^2$, we can construct an exact test statistic

$$\begin{aligned} F(\theta_0) &= \frac{T\tilde{\theta} + 1}{T\theta_0 + 1} \\ &= \frac{\tilde{\lambda}}{(T\theta_0 + 1)\tilde{\sigma}_e^2} \\ &= \frac{y' M_1 y}{(T\theta_0 + 1)m_1 \tilde{\sigma}_e^2}, \end{aligned} \quad (17)$$

which follows an F-distribution with degrees of freedom m_1 and b under the null hypothesis $H_0: \theta = \theta_0$. Here, $\tilde{\theta} = \tilde{\sigma}_u^2 / \tilde{\sigma}_e^2$, is the spectral decomposition estimate of θ .

We reject the null hypothesis H_0 if $F(\theta_0) > F_{a,b}(\alpha)$. The power of the test is given by

$$\begin{aligned}
 &P(F(\theta_0) > F_{m_1, b}(\alpha)) \\
 &= P\left\{ \frac{T\theta + 1}{T\theta_0 + 1} F_{m_1, b} > F_{m_1, b}(\alpha) \right\},
 \end{aligned} \tag{18}$$

where $F_{m_1, b} = b\chi_{m_1}^2 / m_1\chi_b^2$ is an F-statistic with degrees of freedom m_1 and b , and $F_{m_1, b}(\alpha)$ is the corresponding upper $100\alpha\%$ quartile.

In contrast the quadratic forms $y'(P_{(X:Z)} - P_X)y$ and $y'(I - P_{(X:Z)})y$ in the ANOVA method can not be directly used to construct an exact test for H_0 except when $\theta_0 = 0$, though the two quadratic forms are independent for any $\theta \geq 0$. From (14) we have

$y'(P_{(X:Z)} - P_X)y / \sigma_\varepsilon^2 \sim \sum_{i=1}^g (d_i\theta + 1)\chi_{m_i}^2$, which is not a χ^2 variable, but rather a linear combination of two or more independently distributed χ^2 random variables if $\theta \neq 0$.

In the case of $\theta_0 = 0$, the exact test statistic based on the quadratic forms in the ANOVA method can be given by

$$F_A = \frac{y'(P_{(X:Z)} - P_X)y/a}{y'(I - P_{(X:Z)})y/b}, \tag{19}$$

where b is defined as in (5), $a = \text{rk}(Q_X Z) = \sum_{i=1}^g m_i$, and m_i is defined in (14). Under H_0 the statistic follows an F -distribution with degrees of freedom a and b .

If $\theta_0 > 0$, F_A is intractable because of its complicated distribution. As an alternative instead, Wald's test statistic

$$W(\theta_0) = \frac{\{\sum_{i=1}^g y' M_i y / (d_i \theta_0 + 1)\} / a}{y'(I - P_{(X:Z)})y/b} \tag{20}$$

is often used in the literature, see Wald (1940). Clearly, $W(0) = F_A$.

To compare the power of the two tests based on $F(\theta_0)$ and $W(\theta_0)$, we define

$$F_i(\theta_0) = \frac{y' M_i y}{(d_i \theta_0 + 1) m_i \widehat{\sigma_\varepsilon^2}}, i=2, \dots, g.$$

Then $W(\theta_0)$, in fact, is a weighted sum of these F -statistics, that is

$$W(\theta_0) = \frac{m_1}{a} F(\theta_0) + \sum_{i=2}^g \frac{m_i}{a} F_i(\theta_0). \tag{21}$$

It is easy to see that $W(\theta_0)$ follows an F -distribution with degrees of freedom a and b under the null hypothesis $H_0 : \theta = \theta_0$, and the power is given by

$$P\{W(\theta_0) > F_{a,b}(\alpha)\} = P\left\{ \frac{\sum_{i=1}^g (d_i \theta + 1) \chi_{m_i}^2 / (d_i \theta_0 + 1)}{a \chi_b^2 / b} > F_{a,b}(\alpha) \right\}. \quad (22)$$

Intuitively the power of the two tests based on $F(\theta_0)$ and $W(\theta_0)$ should be close when m_1/a approaches to 1; see simulation results in Section 7. In this case, the former test is more appealing in practice due to its simplicity.

Note that the pivotal quantities $F(\theta) \sim F_{m_1,b}$ and $W(\theta) \sim F_{a,b}$ are decreasing functions of θ . Thus two exact $100(1 - \alpha)\%$ confidence intervals for θ , based on pivotal quantities $F(\theta)$ and $W(\theta)$, respectively, can be given by the sets

$$\{\theta \in [0, +\infty): F_{m_1,b}(\alpha/2) \leq F_0(\theta) \leq F_{m_1,b}(1 - \alpha/2)\} \quad (23)$$

and

$$\{\theta \in [0, +\infty): F_{a,b}(\alpha/2) \leq W(\theta) \leq F_{a,b}(1 - \alpha/2)\}. \quad (24)$$

Furthermore, (23) can be simplified as

$$\left\{ \theta \in [0, +\infty): \frac{y' M_1 y}{T m_1 \hat{\sigma}_\varepsilon^2 F_{m_1,b}(1 - \alpha/2)} - \frac{1}{T} \leq \theta \leq \frac{y' M_1 y}{T m_1 \hat{\sigma}_\varepsilon^2 F_{m_1,b}(\alpha/2)} - \frac{1}{T} \right\}.$$

In Section 7, we will compare the expected lengths of these confidence intervals via Monte Carlo simulation.

6 Applications

Section 3 describes two simple examples in which the identity condition $P_X P_Z = P_Z P_X$ always holds and thus the ANOVAE and SDE are identical. In this section we examine two more popular models and discuss when the identity condition holds.

6.1 Circular data

The circular feature of a machined part is one of the most basic geometric primitives, and can be described easily by its center and radius. Due to imperfections introduced in manufacturing, the desired feature may not be truly circular. In order to control the production, we need to estimate the geometric parameters (center and radius), which requires data on machined parts along their circumferences, and a corresponding statistical model. In practice the data can be obtained using a computer-controlled coordinate measuring machine, while one of the models

adopted for such data is the mixed-effects model provided by Wang and Lam (1997), which takes into consideration of the variability in center location of different machined parts.

Let (x_{ij}, y_{ij}) be the j th coordinated measurement on the circumference of the i th machined part, $i = 1, \dots, m, j = 1, \dots, n$, where m is the number of machined parts, and n is the number of measurements taken from the circumference. Denote by $\tau_{i(j)}$ the angle of the j th measurement of the i th part. In general the phase change or angular difference $\tau_{i(j+1)} - \tau_{i(j)}$ can be assumed to be known. Write

$$\tau_{i(j)} = \theta_{i0} + \theta_j, j = 1, 2, \dots, n,$$

$\alpha_i = \rho_i \cos \theta_{i0}$ and $\beta_i = \rho_i \sin \theta_{i0}$, where θ_j is known and θ_{i0} is fixed but unknown. Then for the data points (x_{ij}, y_{ij}) , we have the following model (see Wang and Lam, 1997)

$$\begin{cases} x_{ij} = \xi + \alpha_i \cos \theta_j - \beta_i \sin \theta_j + u_{1i} + \epsilon_{1ij}, \\ y_{ij} = \eta + \alpha_i \sin \theta_j + \beta_i \cos \theta_j + u_{2i} + \epsilon_{2ij}, \end{cases} \quad (25)$$

where $(\xi, \eta)'$ is the designed location of the center of the part, $(u_{1i}, u_{2i})'$ is the random variable in the location of the center of the i th machined part, $\epsilon_{kij} \sim N(0, \sigma^2)$, $u_{ki} \sim N(0, \sigma_0^2)$, $k = 1, 2, i = 1, \dots, m, j = 1, \dots, n$, and all ϵ_{kij} s and u_{ki} s are independent.

Put $z = (z_1, \dots, z_m)'$, $z_i = (x'_i, y'_i)'$, $x_i = (x_{i1}, \dots, x_{in})'$, $y_i = (y_{i1}, \dots, y_{in})'$, $\gamma = (\xi, \eta, \alpha_1, \beta_1, \dots, \alpha_m, \beta_m)'$, $u = (u_{11}, u_{21}, u_{12}, u_{22}, \dots, u_{1m}, u_{2m})'$ and

$$\Phi = \begin{pmatrix} \phi_1 & -\phi_2 \\ \phi_2 & \phi_1 \end{pmatrix}$$

with $\phi_1 = (\cos \theta_1, \dots, \cos \theta_n)'$, $\phi_2 = (\sin \theta_1, \dots, \sin \theta_n)'$. Then model (25) can be rewritten in matrix form as

$$z = X\gamma + Z_u + \epsilon,$$

where $X = (\mathbf{1}_m \otimes I_2 \otimes \mathbf{1}_n : I_m \otimes \Phi)$, $Z = I_{2m} \otimes \mathbf{1}_n$. Clearly, the covariance matrix of z is

$$\text{Cov}(z) = \sigma^2 I_{2mn} + \sigma_0^2 (I_{2m} \otimes \mathbf{1}_n \mathbf{1}_n').$$

It is easy to verify that $P_X P_Z = P_Z P_X$ under Model (25) is equivalent to

$$\begin{cases} \bar{c} = \frac{1}{n} \sum_{j=1}^n \cos \theta_j = 0, \\ \bar{s} = \frac{1}{n} \sum_{j=1}^n \sin \theta_j = 0. \end{cases} \quad (26)$$

According to Theorems 3.1 and 3.2, the SDE and ANOVA of the variance components (σ_0^2, σ^2) are equal if (26) holds. And the condition $\bar{c} = \bar{s} = 0$ can be easily satisfied by sampling n measurements equally spaced around the circumference of the circular feature.

6.2 Longitudinal data

Consider the popular model for longitudinal data:

$$y_{it} = \beta_0 + x_{it}\beta_1 + u_i + \varepsilon_{it}, i=1, 2, \dots, N, t=1, 2, \dots, T, \quad (27)$$

where u_i is the random effect, u_i and ε_{it} have the same assumption as model (1), see Diggle, *et al.* (2002). In fact, Model (27) is a special case with design matrix $X = (\mathbf{1}_n : x)$ where $x = (x'_1, x'_2, \dots, x'_N)'$, $x_i = (x_{i1}, x_{i2}, \dots, x_{iT})'$. Without loss of generality, we assume that $\text{rk}(X) = 2$. That is, $x \neq a\mathbf{1}_n$ for any scalar quantity a .

Note that

$$\text{tr}(Q_X P_Z) = N - 1 - s_b(x), [\text{tr}(Q_X P_Z)]^2 = N - 1 - 2s_b(x) + s_b^2(x),$$

where

$$s_b(x) = T \sum_{i=1}^N (\bar{x}_i - \bar{x}_\cdot)^2 / \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_\cdot)^2,$$

$\bar{x}_i = \sum_{t=1}^T x_{it}/T$ and $\bar{x}_\cdot = \sum_{i=1}^N \sum_{t=1}^T x_{it}/(TN)$. Thus $\text{tr}(P_X P_Z) = [\text{tr}(P_X P_Z)]^2$ if and only if $s_b(x) = 0$ or $s_b(x) = 1$. Combining with the following facts

- i. $0 \leq s_b(x) \leq 1$
- ii. $s_b(x) = 0$ if and only if $\bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_N$.
- iii. $s_b(x) = 1$ if and only if $x_i = c_i \mathbf{1}_T$, $i = 1, \dots, N$

, we can show that $P_Z P_X = P_X P_Z$ in Model (27) is equivalent to $s_b(x) = 0$ or $s_b(x) = 1$. In fact, if $P_Z P_X = P_X P_Z$ then $\text{tr}(P_X P_Z) = [\text{tr}(P_X P_Z)]^2$, thus we prove $s_b(x) = 0$ or $s_b(x) = 1$. On the other hand, we have $P_Z P_X = J_n$ if $s_b(x) = 0$; $P_Z P_X = P(\mathbf{1}; \mathbf{c}) \otimes J_T$ if $s_b(x) = 1$, where $\mathbf{c} = (c_1, \dots, c_N)'$. Thus we obtain $P_Z P_X = P_X P_Z$.

It follows from Theorem 3.3 that the necessary and sufficient condition for the identity of the SDE and ANOVAE of σ_u^2 under model (27) is $\bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_N$, or $x_i = c_i \mathbf{1}_T$, $i = 1, \dots, N$. However, the two conditions are not usually satisfied in practice.

In the following, we compare the the SDE and ANOVAE of σ_u^2 under general case $0 < s_b(x) < 1$. For convenience, we simply write $s_b = s_b(x)$.

Clearly, if $0 < s_b < 1$, then $\text{rk}(P_Z X) = 2$, $\dim(\mathcal{M}(X) \cap \mathcal{M}(Z)) = 1$, and the ratio of the coefficients of σ_ε^4 in $\text{Var}(\hat{\sigma}_u^2)$ and $\text{Var}(\tilde{\sigma}_u^2)$ by (16) is equal to

$$f(s_b) = \frac{(n-2)(N-1)(N-2)}{(n-3)(N-1-s_b)^2},$$

which is a continuous and increasing function of s_b in the interval $(0, 1)$, and for any $N > 2$ and $T > 1$, it has

$$\lim_{s_b \rightarrow 0_+} f(s_b) = \frac{(n-2)(N-2)}{(n-3)(N-1)} = 1 - \frac{N(T-1)-1}{(n-3)(N-1)} < 1,$$

$$\lim_{s_b \rightarrow 1_-} f(s_b) = \frac{(n-2)(N-1)}{(n-3)(N-2)} > 1.$$

So there exists $s_{b0} \in (0, 1)$ such that $f(s_b) < 1$ if $s_b < s_{b0}$ and $f(s_b) > 1$ if $s_b > s_{b0}$. In fact, s_{b0} can be given by

$$s_{b0} = N - 1 - \sqrt{\frac{(N-1)(N-2)(n-2)}{(n-3)}}. \quad (28)$$

From Corollary 4.1, it follows that the ANOVAE of σ_u^2 is superior to the SDE if $0 < s_b < s_{b0}$.

Because

$$0.5 - \frac{1}{2T} < s_{b0} < 0.5 + \frac{1}{2N-3},$$

it follows that (i) $f(s_b) < 1$ if $s_b < 0.5 - 1/(2T)$; (ii) $f(s_b) > 1$ if $s_b > 0.5 + 1/(2N-3)$.

Combining with Theorem 4.2, we conclude that for any $\sigma_\varepsilon^2 > 0$, $\sigma_u^2 \geq 0$ the ANOVAE of σ_u^2 is superior to the SDE if $s_b < 0.5 - 1/(2T)$. On the other hand, the ANOVAE of σ_u^2 is not always superior to its SDE. Let $v(\tau) = \{\text{Var}(\hat{\sigma}_u^2) - \text{Var}(\tilde{\sigma}_u^2)\} / \sigma_\varepsilon^4$ which is the quadratic function of $\tau = T\sigma_u^2 / \sigma_\varepsilon^2$, and the coefficients of τ^2 and τ are negative. Using the fact that $v(0) > 0$ if $s_b > 0.5 + 1/(2N-3)$, we can show that there exist two different solutions for the equation $v(\tau) = 0$ in this case, denoted by τ_1, τ_2 ($\tau_1 < \tau_2$). Clearly, $\tau_1 < 0 < \tau_2$. Thus $v(\tau) > 0$ for any $\tau \in (0, \tau_2)$ if

$$\text{Var}(\hat{\sigma}_u^2) > \text{Var}(\tilde{\sigma}_u^2) \text{ if } s_b > 0.5 + 1/(2N$$

$s_b > 0.5 + 1/(2N-3)$. That is, $-3)$ and $T\sigma_u^2 / \sigma_\varepsilon^2 < \tau_2$.

7 Simulation

Under model (27), T and Ts_b both are distinct nonzero eigenvalues of the matrix $Q_X Z Z' Q_X$, and their multiplicities are $N-2$ and 1 , respectively. We fix $\sigma_\varepsilon^2 = 1$, and compare via Monte Carlo simulation the power of $F_0(\theta_0)$ and $W(\theta_0)$. Results are displayed for two different choices $x_{it} = t + e_{it} + \alpha_i$ and $x_{it} = \log(t+1) + e_{it} + 2\alpha_i$, and three different sample sizes (N, T) : $(12, 3)$, $(20, 4)$, $(5, 10)$, with 10000 replicates in each case, where $\alpha_i \sim N(0, 1)$ and $e_{it} \sim N(0, 1)$ are independent. Note that for each choice of sample size (N, T) , x_{it} s are fixed. Here s_{b0} can be calculated by (28) for the three sample sizes of (N, T) , that is, 0.354, 0.387 and 0.499, respectively.

We first obtain two sets of simulated values, $(0.345, 0.280, 0.039)$ and $(0.732, 0.741, 0.431)$, of s_b under the three sample sizes of (N, T) , respectively, for $x_{it} = t + e_{it} + \alpha_i$ and $x_{it} = \log(t+1) + e_{it} + 2\alpha_i$. Clearly, the three values of s_b in the first set and the last value in the second set are smaller than the corresponding s_{b0} , and the first two values of s_b in the second set are larger than the corresponding s_{b0} . Using the results in example 6.2, we can obtain the following results

under three sample sizes above. For the first choice $x_{it} = t + e_{it} + a_i$ the ANOVAE of σ_u^2 is superior to its SDE for any $\sigma_u^2 \geq 0$ and $\sigma_e^2 > 0$. For the second choice $x_{it} = \log(t+1) + e_{it} + 2a_i$, the SDE of σ_u^2 is superior to its ANOVAE only if $\sigma_u^2 \leq 0.227$ for $(N, T) = (12, 3)$, or only if $\sigma_u^2 \leq 0.159$ for $(N, T) = (20, 4)$.

Secondly, we take random numbers from the independent χ^2 distributions $\chi_{m_1}^2, \chi_1^2$ and χ_b^2 with 10000 replicates for each of the six cases above. Using formula (18) and (22), we can obtain simulated test powers of $F(\theta_0)$ and $W(\theta_0)$, see Table 1, Table 2.

Table 1 and Table 2 show that the test based on $W(\theta_0)$ is more powerful than that based on $F(\theta_0)$ for the null hypothesis $H_0: \theta = \theta_0$ under sample size $(N, T) = (5, 10)$, a scenario with relatively large number of repeated observations for each subject but small number of subjects. There is no substantial difference between the power of the two tests above under the sample size $(N, T) = (12, 3)$ or $(20, 4)$, a scenario with relatively small number of repeated observations for each subject but large number of subjects, regardless of whether the ANOVAE is superior to the SDE. As a subsequent observation, the result indicates that better estimates of variance components do not necessarily imply higher power of the tests. A similar phenomenon is observed in the expected lengths of confidence interval (24) and (23); see Figure 1 below.

Thus we can adopt the simpler test statistic $F(\theta_0)$ and confidence interval (23) in practice when T , the number of repeated observations for each subject, is relatively small. However they are not good choices due to the low power of the test and the long length of the confidence interval when T is relatively large, especially, when T is larger than N , the number of subjects.

8 Discussion

In this paper, we study the finite sample properties of ANOVAE and SDE under model (1). We establish the necessary and sufficient condition for the equality of the two estimates, and some sufficient conditions for the superiority of one estimate over the other under the mean squared error criterion. Furthermore, we consider exact tests and confidence intervals based on the two estimates, and demonstrate via simulations that better estimates of variance components do not necessarily imply higher power of the tests or shorter confidence intervals.

As comparison, the likelihood method has poor small sample size performance in mixed models (see Burdick, 1997). Moreover, when the parameter lies in the boundary of the parametric space, the likelihood-based tests such as Wald and likelihood ratio tests, do not asymptotically follow chi-square distributions as demonstrated in Stram and Lee (1994) and Crainiceanu and Ruppert (2004). For the two models considered in Section 7 with the same settings, Table 3 presents the coverage probabilities of the confidence intervals based on the maximum likelihood estimate. The results in the table clearly indicate that with relatively small sample sizes, the coverage probabilities are substantially below the nominal level of 95%.

As noticed in the literature, both ANOVAE and SDE may take negative values. This is not a problem for Genetic field in which variance component is permitted to be negative, see Burton *et al.* (1999) and Hazelton and Gurrin (2003). However, estimates of variance components are often required to be non-negative in many other cases. One remedy is to estimate the two variance components using $\max\{0, \hat{\sigma}_u^2\}$ and $\max\{0, \hat{\sigma}_e^2\}$. However further research is needed to find conditions for the superiority among the two truncated estimates. In this regard Theorem 4.1 in this paper is a potentially useful tool for such investigation.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institute of Child Health and Human Development, National Institutes of Health. Wu's research was also partially supported by Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality PHR (IHLB) and National Natural Science Foundation of China (NSFC). The opinions expressed in the article are not necessarily of the National Institutes of Health. The authors wish to thank two referees for their valuable suggestions which considerably improved the paper.

References

- Burch BD, Iyer HK. Exact confidence intervals for a variance ratio (or Heritability) in a mixed linear model. *Biometrics* 1997;53:1318–1333.
- Burdick RK, Larsen GA. Confidence intervals on measures of variability in repeatability and reproducibility study. *J. Quality Tech* 1997;29:261–273.
- Burton P, Tiller K, Gurrin L, Cookson W, Musk A, Palmer L. Genetic variance components analysis for binary phenotypes using generalized linear mixed models (GLMMs) and Gibbs sampling. *Genetic Epidemiology* 1999;17:118–140. [PubMed: 10414556]
- Crainiceanu CM, Ruppert D. Likelihood ratio tests in Linear Mixed Models with One Variance Component. *J. Royal Stat. Soc. B* 2004;66:165–185.
- Davidian, M.; Giltinan, DM. *Nonlinear Models for Repeated Measurement Data*. London: Chapman and Hall; 1996.
- Demidenko, E. *Mixed Models: Theory and Applications*. New Jersey: John Wiley & Sons; 2004.
- Diggle, PJ.; Liang, KY.; Zeger, SL. *Analysis of Longitudinal Data*. 2nd ed.. New York: Oxford Science; 2002.
- Hazelton ML, Gurrin LC. A note on genetic variance components in mixed models. *Genetic Epidemiology* 2003;24:297–301. [PubMed: 12687647]
- Henderson CR. Estimation of variance and covariance components. *Biometrics* 1953;9:226–253.
- Rao, CR. *Linear Statistical Inference and Its Applications*. 2nd ed.. New York: Wiley; 1973.
- Searle, SR.; Casella, G.; McCulloch, CE. *Variance Components*. New York: Wiley; 1992.
- Stram DO, Lee JW. Variance components testing in the longitudinal mixed-effects model. *Biometrics* 1994;50:1171–1177. [PubMed: 7786999]
- Wald A. A note on the analysis of variance with unequal class frequencies. *Ann. Math. Statist* 1940;11:96–100.
- Wang CM, Lam CT. A mixed-effects models for the analysis of circular measurements. *Technometrics* 1997;39:119–126.
- Wang, SG.; Chow, SC. *Advanced Linear Models*. New York: Marcel Dekker Inc.; 1994.
- Wang SG, Yin SJ. A new estimate of the parameters in linear mixed models. *Science in China (Series A)* 2002;45:1301–1311.
- Wu MX, Wang SG. Simultaneous optimal estimates of fixed effects and variance components in the mixed model. *Science in China (Series A)* 2004;47:787–799.

Figure 1.
Expected lengths of exact confidence intervals based on $F(\theta_0)$ (solid line) and $W(\theta_0)$ (dash line).

Table 1

Power of tests for hypothesis $\theta = \theta_0$ under Model (27) with $x_{it} = t + e_{it} + \alpha_i$, where all $\alpha_i \sim N(0, 1)$, $e_{it} \sim N(0, 1)$ are independent (runs: 10000)

θ_0	θ	$N = 5$		$T = 10$		$N = 12$		$T = 3$		$N = 20$		$T = 4$	
		$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$
0	0.1	0.2490	0.2815	0.1345	0.1351	0.2483	0.2518	0.2483	0.2518	0.2483	0.2518	0.2483	0.2518
	0.3	0.5519	0.6239	0.3689	0.3750	0.6993	0.7067	0.6993	0.7067	0.6993	0.7067	0.6993	0.7067
	0.5	0.6964	0.7797	0.5799	0.5874	0.8987	0.9044	0.8987	0.9044	0.8987	0.9044	0.8987	0.9044
	0.7	0.7808	0.8550	0.7220	0.7344	0.9685	0.9714	0.9685	0.9714	0.9685	0.9714	0.9685	0.9714
	0.9	0.8288	0.8982	0.8206	0.8311	0.9876	0.9891	0.9876	0.9891	0.9876	0.9891	0.9876	0.9891
0.5	1	0.8479	0.9138	0.8508	0.8638	0.9922	0.9929	0.9922	0.9929	0.9922	0.9929	0.9922	0.9929
	1.5	0.2131	0.2436	0.2518	0.2582	0.4200	0.4221	0.4200	0.4221	0.4200	0.4221	0.4200	0.4221
	2	0.3708	0.4314	0.4854	0.4988	0.7472	0.7598	0.7472	0.7598	0.7472	0.7598	0.7472	0.7598
	3	0.4909	0.5652	0.7750	0.7941	0.8987	0.9062	0.8987	0.9062	0.8987	0.9062	0.8987	0.9062
1.5	2	0.6468	0.7306	0.8987	0.9132	0.983	0.9864	0.983	0.9864	0.983	0.9864	0.983	0.9864
	3	0.1053	0.1145	0.1249	0.1262	0.1794	0.1818	0.1794	0.1818	0.1794	0.1818	0.1794	0.1818
	4	0.2352	0.2672	0.3381	0.3515	0.5343	0.5472	0.5343	0.5472	0.5343	0.5472	0.5343	0.5472
	5	0.3526	0.4093	0.5383	0.5557	0.7780	0.7929	0.7780	0.7929	0.7780	0.7929	0.7780	0.7929
	6	0.4504	0.5213	0.6824	0.7059	0.8987	0.9070	0.8987	0.9070	0.8987	0.9070	0.8987	0.9070
		0.5300	0.6055	0.7850	0.8072	0.9577	0.9628	0.9577	0.9628	0.9577	0.9628	0.9577	0.9628

* Significance level $\alpha = 0.05$.

Table 2

Power of tests for hypothesis $\theta = \theta_0$ under Model (27) with $x_{it} = \log(t + 1) + e_{it} + 2\alpha_i$, where all $\alpha_i \sim N(0, 1)$, $e_{it} \sim N(0, 1)$ are independent (runs:10000)

θ_0	θ	$N = 5$		$T = 10$		$N = 12$		$T = 3$		$N = 20$		$T = 4$	
		$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$	$F(\theta_0)$	$W(\theta_0)$
0	0.1	0.2490	0.2562	0.1345	0.1315	0.2483	0.2441						
	0.3	0.5519	0.5884	0.3689	0.362	0.6993	0.6922						
	0.5	0.6964	0.7467	0.5799	0.5708	0.8987	0.8961						
	0.7	0.7808	0.8288	0.7220	0.7141	0.9685	0.9681						
	0.9	0.8288	0.8780	0.8206	0.8151	0.9876	0.9877						
	1	0.8479	0.8963	0.8508	0.8500	0.9922	0.9918						
0.5	1	0.2131	0.236	0.2518	0.2487	0.4200	0.4165						
	1.5	0.3708	0.4211	0.4854	0.4833	0.7472	0.7492						
	2	0.4909	0.5568	0.6585	0.6623	0.8987	0.9004						
	3	0.6468	0.7229	0.8508	0.8572	0.983	0.9849						
1.5	2	0.1053	0.1131	0.1249	0.1249	0.1794	0.1796						
	3	0.2352	0.2646	0.3381	0.3442	0.5343	0.5393						
	4	0.3526	0.4053	0.5383	0.5480	0.7780	0.7872						
	5	0.4504	0.5176	0.6824	0.6948	0.8987	0.9041						
	6	0.5300	0.6025	0.7850	0.7968	0.9577	0.9610						

* Significance level $\alpha = 0.05$.

Table 3

Estimated coverage probabilities * the confidence interval for θ based on MLE under Model (27) (runs: 10000). Case I: $x_{it} = t + e_{it} + \alpha_i$. Case II: $x_{it} = \log(t + 1) + e_{it} + 2\alpha_i$.

θ	$N = 5$		$N = 10$		$N = 12$		$N = 20$		$T = 4$	
	Case I	Case II	Case I	Case II	Case I	Case II	Case I	Case II	Case I	Case II
0.5	0.7122	0.6944	0.8597	0.8408	0.8922	0.8745				
1	0.7235	0.7088	0.8597	0.8317	0.8829	0.8692				
1.5	0.7205	0.7127	0.8509	0.8292	0.8798	0.8669				

* Nominal confidence level $1 - \alpha = 0.95$.