

Nonnegative variance component estimation for mixed-effects models

Jaesung Choi^{1,a}

^aDepartment of Statistics, Keimyung University, Korea

Abstract

This paper suggests three available methods for finding nonnegative estimates of variance components of the random effects in mixed models. The three proposed methods based on the concepts of projections are called projection method I, II, and III. Each method derives sums of squares uniquely based on its own method of projections. All the sums of squares in quadratic forms are calculated as the squared lengths of projections of an observation vector; therefore, there is discussion on the decomposition of the observation vector into the sum of orthogonal projections for establishing a projection model. The projection model in matrix form is constructed by ascertaining the orthogonal projections defined on vector subspaces. Nonnegative estimates are then obtained by the projection model where all the coefficient matrices of the effects in the model are orthogonal to each other. Each method provides its own system of linear equations in a different way for the estimation of variance components; however, the estimates are given as the same regardless of the methods, whichever is used. Hartley's synthesis is used as a method for finding the coefficients of variance components.

Keywords: mixed model, projection, quadratic form, random effect, synthesis

1. Introduction

Much literature has been devoted to the estimation of variance components in random-effects or mixed-effects models. A variance component should always be nonnegative by definition; however, we sometimes get it as negative. Searle (1971) and Searle and Gruber (2016) illustrated this with the simple hypothetical data of a one-way classification having three observations in two classes and insisted that there is nothing intrinsic in the analysis of variance method to prevent it. Nevertheless, when a negative estimate happens, it is not easy to handle this situation in interpretation and action. Therefore, many papers have been contributed to strategies to deal with the negative values as estimates of variance components. Nelder (1954) suggested that negative estimates of variance components can occur in certain designs such as split-plot and randomized block designs by randomization. Thompson (1961, 1963) discussed the interpretation of the negative estimate and suggests an alternative method when the analysis of variance method yields negative estimates. Milliken and Johnson (1984) also suggested a procedure to eliminate negative estimates of variance components in random-effects models.

The analysis of the variance method is almost exclusively applied to balanced data to estimate variance components. However, there are multiple methods for unbalanced data. Therefore, it is necessary to identify the types of data before choosing a method. Balanced data have the same numbers

¹ Department of Statistics, Keimyung University, 1095 Dalgubeol-Daero, Dalseo-Gu, Daegu 42601, Korea.
E-mail: jschoi@kmu.ac.kr

of observations in each cell; however, unbalanced data have unequal numbers of observations in the subclasses made by the levels of classification factors. Depending on the types of data, many methods can be applied to the estimation of variance components in a vector space. Representing data as vectors, the vector space of an observation vector can be partitioned in many ways, depending on the data structure. The vector space can always be partitioned into orthogonal vector subspaces according to the sources of variation for balanced data, but it is not true for unbalanced data. This is the main difference between balanced and unbalanced data from the view point of a vector space.

A random effect is a random variable representing the effect of a randomly chosen level from a population of levels that a random factor can assume, while a fixed effect is an unknown constant denoting the effect of a predetermined level of a factor. A linear model with these two types of effects is called a mixed-effects model. The primary concern with the model in this paper is in the nonnegative estimation of variance components of random effects. A negative estimate can happen in any method that contributes to the estimation. Therefore, many papers have investigated strategies for interpretation and alternatives. Such strategies are seen in Searle and Fawcett (1970), Hill (1965, 1967), Searle *et al.* (2009) and Harville (1969). However, it is necessary to have a method that yields nonnegative estimates.

Henderson (1953) suggested a method that uses reductions in sums of squares due to fitting both the full model and different sub-models of it to estimate variance components of random effects in mixed models. This method is called the fitting constants method or Henderson's Method 3. Even though it has been used extensively for the estimation of variance components in mixed models, it still has some defects producing negative estimates. Hartley's (1967) synthesis is also used to calculate the coefficients of variance components in the method. However, we should recognize whether quadratic forms for variance components are in the right form or not despite the usefulness of this method. Otherwise, expectations of the quadratic forms can be different from the real ones. This is going to be discussed in detail in projection model building.

This paper suggests three methods to produce nonnegative estimates for variance components in mixed models. They are based on the concept of projection defined on a vector space. The definition of a projection and its related concepts are discussed in Graybill (1983) and Johnson and Wichern (2014). Quadratic forms in the observations can be obtained as squared lengths of projections defined on proper vector subspaces. Each method requires that all vector subspaces for projections should be orthogonal to each other at the stage of fitting sub-models serially. It is possible to get nonnegative estimates when the orthogonality is satisfied with vector subspaces. Therefore, we also discuss how to construct orthogonal vector subspaces from a given mixed model. Quadratic forms as sums of squares due to random effects are then used to evaluate the expected values. Hereafter, equating quadratic forms to their expected values represents available equations for the estimates. For calculating the coefficients of variance components, Hartley's synthesis is applied but in a different manner, which will be discussed.

2. Mixed models

Mixed models are used to describe data from experimental situations where some factors are fixed and others are random. When two types of factors are considered in experiments, one is interested in both parts, that is, the fixed-effects part and the random-effects part, in models. Let α be a vector of all the fixed effects except μ in a mixed model and let δ_i denote a set of random effects for random factor i for $i = 1, 2, \dots, k$. Then, δ_i could be interaction effects or nested-factor effects when they are simply regarded as effects from random factors. The matrix notation of the mixed model for an observation

vector \mathbf{y} is

$$\begin{aligned}\mathbf{y} &= \mathbf{j}\mu + \mathbf{X}_F\boldsymbol{\alpha}_F + \mathbf{X}_R\boldsymbol{\delta}_R + \boldsymbol{\epsilon} \\ &= \mathbf{j}\mu + \mathbf{X}_F\boldsymbol{\alpha}_F + \sum_{i=1}^k \mathbf{X}_i\boldsymbol{\delta}_i + \boldsymbol{\epsilon},\end{aligned}\quad (2.1)$$

where \mathbf{j} is an $n \times 1$ vector of ones, μ is the general mean, \mathbf{X}_F is an $n \times a$ matrix, $\boldsymbol{\alpha}_F$ is the $a \times 1$ vector of fixed effects, the \mathbf{X}_i 's are the partitions of an $n \times r$ matrix \mathbf{X}_R , each being an $n \times r_i$ matrix for $r = \sum_{i=1}^k r_i$, $\boldsymbol{\delta}_i$ is an $r_i \times 1$ vector of random effects, $\mathbf{j}\mu + \mathbf{X}_F\boldsymbol{\alpha}_F$ is the fixed part of the model and $\mathbf{X}_R\boldsymbol{\delta}_R + \boldsymbol{\epsilon}$ is the random part of the model. $\boldsymbol{\delta}_i$'s are assumed to be independent and identically distributed as $N(\mathbf{0}, \sigma_{\delta_i}^2 \mathbf{I}_{r_i})$, and $\boldsymbol{\epsilon}$ is assumed to be distributed as $N(\mathbf{0}, \sigma_{\epsilon}^2 \mathbf{I}_n)$ and independent of $\boldsymbol{\delta}_i$. The mean and variance of \mathbf{y} from (2.1) is

$$\begin{aligned}E(\mathbf{y}) &= \mathbf{j}\mu + \mathbf{X}_F\boldsymbol{\alpha}_F, \\ \boldsymbol{\Sigma} = \text{var}(\mathbf{y}) &= \sum_{i=1}^k \mathbf{X}_i \text{var}(\boldsymbol{\delta}_i) \mathbf{X}_i^T + \sigma_{\epsilon}^2 \mathbf{I}_n.\end{aligned}\quad (2.2)$$

The expected value of the quadratic form $\mathbf{y}^T \mathbf{Q} \mathbf{y}$ is

$$E(\mathbf{y}^T \mathbf{Q} \mathbf{y}) = \text{tr}(\mathbf{Q}\boldsymbol{\Sigma}) + E(\mathbf{y})^T \mathbf{Q} E(\mathbf{y}).\quad (2.3)$$

Substituting the terms of (2.2) for (2.3) is

$$E(\mathbf{y}^T \mathbf{Q} \mathbf{y}) = \sum_{i=1}^r \sigma_{\delta_i}^2 \text{tr}(\mathbf{Q} \mathbf{X}_i \mathbf{X}_i^T) + \sigma_{\epsilon}^2 \text{tr}(\mathbf{Q}) + E(\mathbf{j}\mu + \mathbf{X}_F\boldsymbol{\alpha}_F)^T \mathbf{Q} E(\mathbf{j}\mu + \mathbf{X}_F\boldsymbol{\alpha}_F).\quad (2.4)$$

The expectation of any quadratic form in the observations of a vector \mathbf{y} is represented as a function of variance components and fixed effects. The variance components of the full model can be estimated by the fitting constants method of using reductions in the sums of squares due to fitting both the full model and the sub-model. This method has been widely used for the estimation of variance components for unbalanced data because it provides unbiased estimators of the variance components that do not depend on any fixed effects in the model. However, it still has an unsolved problem of having negative solutions as estimates. As an alternative, a method which is based on the concepts of projections is suggested. To discuss it, we consider the model (2.1) as representative. We naturally divide the model into a fixed part and a random part since there are two parts in the model. The random part of the model consists of random effects and errors:

$$\begin{aligned}\mathbf{y} &= \mathbf{j}\mu + \mathbf{X}_F\boldsymbol{\alpha}_F + \boldsymbol{\epsilon}_R \\ &= (\mathbf{j}, \mathbf{X}_F)(\mu, \boldsymbol{\alpha}_F)^T + \boldsymbol{\epsilon}_R,\end{aligned}\quad (2.5)$$

where $\boldsymbol{\epsilon}_R = \sum_{i=1}^k \mathbf{X}_i\boldsymbol{\delta}_i + \boldsymbol{\epsilon}$. The general mean μ and fixed effects $\boldsymbol{\alpha}_F$ of (2.5) can be estimated from normal equations. Regarding \mathbf{y} as an observation vector in the n -dimensional vector space, it can be decomposed into two component vectors orthogonal to each other. The decomposition of \mathbf{y} is done by projecting \mathbf{y} onto the vector subspace generated by $(\mathbf{j}, \mathbf{X}_F)$.

3. Projection method

For a mixed model such as (2.5), we can decompose \mathbf{y} into two components by means of projections referred to as a projection method. Denoting $(\mathbf{j}, \mathbf{X}_F)$ and $(\boldsymbol{\mu}, \boldsymbol{\alpha}_F)^T$ by \mathbf{X}_M and $\boldsymbol{\alpha}_M$, respectively, the projection of \mathbf{y} onto the vector subspace spanned by \mathbf{X}_M is $\mathbf{X}_M \mathbf{X}_M^- \mathbf{y}$, where \mathbf{X}_M^- denotes a Moore-Penrose generalized inverse of \mathbf{X}_M . Then, \mathbf{y} can be decomposed into two vectors, that is, $\mathbf{X}_M \mathbf{X}_M^- \mathbf{y}$ and $(\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{y}$, which are orthogonal. Instead of the fitting constants method, the projection method is used to try to estimate the nonnegative estimates of the variance components in a mixed model. To explain the method simply, suppose there are two factors A and B for a two-way cross-classified unbalanced data where A is fixed with a levels and B is random with b levels. The model for this is

$$\begin{aligned} \mathbf{y} &= \mathbf{j}\boldsymbol{\mu} + \mathbf{X}_F \boldsymbol{\alpha}_F + \mathbf{X}_\beta \boldsymbol{\delta}_\beta + \mathbf{X}_{\alpha\beta} \boldsymbol{\delta}_{\alpha\beta} + \boldsymbol{\epsilon} \\ &= \mathbf{X}_M \boldsymbol{\alpha}_M + \boldsymbol{\epsilon}_M, \end{aligned} \quad (3.1)$$

where \mathbf{y} is an observation vector in the n dimensional vector space, $\boldsymbol{\alpha}_F$ is a vector of fixed effects of A , $\boldsymbol{\delta}_\beta$ and $\boldsymbol{\delta}_{\alpha\beta}$ represent vectors of random effects of B and AB interaction respectively, and $\mathbf{X}_M = (\mathbf{j}, \mathbf{X}_F)$, $\boldsymbol{\alpha}_M = (\boldsymbol{\mu}, \boldsymbol{\alpha}_F)^T$ and $\boldsymbol{\epsilon}_M = \mathbf{X}_\beta \boldsymbol{\delta}_\beta + \mathbf{X}_{\alpha\beta} \boldsymbol{\delta}_{\alpha\beta} + \boldsymbol{\epsilon}$. The second expression of (3.1) represents the fixed-effects part and the random part. The random part $\boldsymbol{\epsilon}_M$ is obtained by the projection of \mathbf{y} onto a vector subspace generated by the \mathbf{X}_M , which is $(\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{y}$. So, \mathbf{y} is represented as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}_M \mathbf{X}_M^- \mathbf{y} + (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{y} \\ &= \mathbf{y}_M + \boldsymbol{e}_M, \end{aligned} \quad (3.2)$$

where $\mathbf{y}_M = \mathbf{X}_M \mathbf{X}_M^- \mathbf{y}$ satisfies the two conditions for being the projection of \mathbf{y} onto a vector subspace spanned by the columns of \mathbf{X}_M . The projection should be obtained by the orthogonal projection to the subspace and denoted as a linear combination of the column vectors of \mathbf{X}_M . $\mathbf{X}_M \mathbf{X}_M^- \mathbf{y}$ of (3.2) satisfies the conditions. The random part $\boldsymbol{e}_M = (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{y}$ is not affected by the fixed effects since \mathbf{y}_M is orthogonal to \boldsymbol{e}_M ; in addition, it has all the information about the variance components and random error variance. We can use \boldsymbol{e}_M for finding the related variance components since there are two random effects and random error terms in the model of (3.1). The model for the estimation of σ_β^2 defined as the variance component of $\boldsymbol{\delta}_\beta$ is

$$\begin{aligned} \boldsymbol{e}_M &= (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{y} \\ &= \mathbf{X}_B \boldsymbol{\delta}_\beta + \boldsymbol{\epsilon}_\beta, \end{aligned} \quad (3.3)$$

where $\mathbf{X}_B = (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{X}_\beta$ and $\boldsymbol{\epsilon}_\beta = (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) (\mathbf{X}_{\alpha\beta} \boldsymbol{\delta}_{\alpha\beta} + \boldsymbol{\epsilon})$. The projection of \boldsymbol{e}_M onto the subspace spanned by \mathbf{X}_B is $\mathbf{X}_B \mathbf{X}_B^- \boldsymbol{e}_M$, which is $[(\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{X}_\beta] [(\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{X}_\beta]^- \boldsymbol{e}_M$. Then,

$$\begin{aligned} \boldsymbol{e}_M &= \mathbf{X}_B \mathbf{X}_B^- \boldsymbol{e}_M + (\mathbf{I} - \mathbf{X}_B \mathbf{X}_B^-) \boldsymbol{e}_M \\ &= \mathbf{y}_B + \boldsymbol{e}_B, \end{aligned} \quad (3.4)$$

where $\mathbf{y}_B = \mathbf{X}_B \mathbf{X}_B^- \boldsymbol{e}_M$ is the projection of \boldsymbol{e}_M onto the column space of \mathbf{X}_B . \mathbf{y}_B and \boldsymbol{e}_B are orthogonal each other. Therefore, \boldsymbol{e}_B is not affected by the random effects $\boldsymbol{\delta}_\beta$. Therefore, \boldsymbol{e}_B is used for finding the subspace that has information about $\sigma_{\alpha\beta}^2$ defined as the variance component of $\boldsymbol{\delta}_{\alpha\beta}$. The model for this is

$$\boldsymbol{e}_B = (\mathbf{I} - \mathbf{X}_B \mathbf{X}_B^-) \boldsymbol{e}_M = \mathbf{X}_{AB} \boldsymbol{\delta}_{\alpha\beta} + \boldsymbol{\epsilon}_{\alpha\beta}, \quad (3.5)$$

where $X_{AB} = (\mathbf{I} - X_M X_M^- - X_B X_B^-) X_{\alpha\beta}$ and $\epsilon_{\alpha\beta} = (\mathbf{I} - X_M X_M^- - X_B X_B^-) \epsilon$. Therefore, the projection of e_B onto the subspace generated by X_{AB} is $y_{AB} = X_{AB} X_{AB}^- e_B$. Then,

$$\begin{aligned} e_B &= X_{AB} X_{AB}^- e_B + (\mathbf{I} - X_{AB} X_{AB}^-) e_B \\ &= y_{AB} + e_{AB}, \end{aligned} \quad (3.6)$$

where e_{AB} is $(\mathbf{I} - X_{AB} X_{AB}^-) e_B$. Finally, we can use e_{AB} for finding the coefficient matrix of the random error vector which generates the error space orthogonal to all the other spaces.

$$\begin{aligned} e_{AB} &= (\mathbf{I} - X_{AB} X_{AB}^-) e_B \\ &= (\mathbf{I} - X_M X_M^- - X_B X_B^- - X_{AB} X_{AB}^-) \epsilon. \end{aligned} \quad (3.7)$$

Thus, we can know that e_{AB} has all the information about σ_ϵ^2 of the random error vector ϵ . Denoting \mathbf{y} as the sum of orthogonal projections and error part,

$$\begin{aligned} \mathbf{y} &= y_M + y_B + y_{AB} + e_{AB} \\ &= X_M X_M^- \mathbf{y} + X_B X_B^- e_M + X_{AB} X_{AB}^- e_B + (\mathbf{I} - X_{AB} X_{AB}^-) e_B. \end{aligned} \quad (3.8)$$

Each term of (3.8) can be used to calculate the sums of squares that are quadratic forms in the observations. Since \mathbf{y} is partitioned as four terms, there are four available sums of squares. We denote them SS_M , SS_B , SS_{AB} , and SS_E where subscripts are corresponding factors. They are defined as

$$\begin{aligned} SS_M &= \mathbf{y}^T X_M X_M^- \mathbf{y}, \\ SS_B &= \mathbf{y}^T (\mathbf{I} - X_M X_M^-) X_B X_B^- (\mathbf{I} - X_M X_M^-) \mathbf{y}, \\ SS_{AB} &= \mathbf{y}^T (\mathbf{I} - X_M X_M^- - X_B X_B^-) X_{AB} X_{AB}^- (\mathbf{I} - X_M X_M^- - X_B X_B^-) \mathbf{y}, \\ SS_E &= \mathbf{y}^T X_E \mathbf{y}, \end{aligned} \quad (3.9)$$

where each SS term is given as the squared length of the projection of \mathbf{y} onto its own vector subspace, and $X_E = (\mathbf{I} - X_M X_M^- - X_B X_B^- - X_{AB} X_{AB}^-)$. All the sums of squares are evaluated by using the eigenvalues and eigenvectors of the projection matrices associated with the quadratic forms in \mathbf{y} . We can identify the coefficient matrices that are orthogonal to each other through the procedure for finding projections on subspaces.

4. Projection model

Since \mathbf{y} is composed of the sum of mutual orthogonal projections such as (3.8), \mathbf{y} can be represented by the orthogonal coefficients matrices of the effects of the assumed model (3.1). Temporarily, we denote \mathbf{y} as y_p to differentiate the model based on projections from the classical model (3.1). Then, the model for y_p is

$$y_p = X_M \alpha_M + X_B \delta_B + X_{AB} \delta_{\alpha\beta} + X_E \epsilon, \quad (4.1)$$

where $y_p = \mathbf{y}$. Since each coefficient matrix of the effects is derived from the corresponding orthogonal projection, the equation of (4.1) defines a projection model that is different from a classical two-way linear mixed model (3.1). It is useful for evaluating the coefficients of the variance components in the expectations of the quadratic form of an observation vector y_p . All the coefficient matrices

are orthogonal in the model. δ_β , $\delta_{\alpha\beta}$, and ϵ are assumed to be $N(\mathbf{0}, \sigma_\beta^2 \mathbf{I}_b)$, $N(\mathbf{0}, \sigma_{\alpha\beta}^2 \mathbf{I}_{ab})$, and $N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n)$ respectively. The expectation and the covariance matrix of \mathbf{y}_p of the projection model (4.1) is

$$\begin{aligned} E(\mathbf{y}_p) &= \mathbf{X}_M \boldsymbol{\alpha}_M, \\ \boldsymbol{\Sigma} &= \sigma_\beta^2 \mathbf{X}_B \mathbf{X}_B^T + \sigma_{\alpha\beta}^2 \mathbf{X}_{AB} \mathbf{X}_{AB}^T + \sigma_\epsilon^2 \mathbf{X}_E \mathbf{X}_E^T. \end{aligned} \quad (4.2)$$

Expectations of the SS terms except SS_M of (3.9) are

$$E(SS_B) = \sigma_\beta^2 \text{tr}(\mathbf{X}_B^T \mathbf{X}_B), \quad E(SS_{AB}) = \sigma_{\alpha\beta}^2 \text{tr}(\mathbf{X}_{AB}^T \mathbf{X}_{AB}), \quad E(SS_E) = \sigma_\epsilon^2 \text{tr}(\mathbf{X}_E). \quad (4.3)$$

Equating the three sums of squares, SS_B , SS_{AB} , and SS_E of (3.9) to their corresponding expectations leads to linear equations in the variance components, the solutions to which are taken as the estimators of those components. Now, the equations are

$$SS_B = \hat{\sigma}_\beta^2 \text{tr}(\mathbf{X}_B^T \mathbf{X}_B), \quad SS_{AB} = \hat{\sigma}_{\alpha\beta}^2 \text{tr}(\mathbf{X}_{AB}^T \mathbf{X}_{AB}), \quad SS_E = \hat{\sigma}_\epsilon^2 \text{tr}(\mathbf{X}_E). \quad (4.4)$$

Solutions from the linear equations (4.4) are nonnegative estimates of the variance components. We will identify them as projection method I, II, and III since there are three different ways of getting sums of squares by means of projections. The procedure using the system of linear equations like (4.4) is called projection method I. We can use \mathbf{e}_M , \mathbf{e}_B , and \mathbf{e}_{AB} since the projection method II uses residual vectors after projecting \mathbf{y} onto orthogonal subspaces. Then,

$$\begin{aligned} \mathbf{e}_M &= (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) \mathbf{y}_p \\ &= (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^-) (\mathbf{X}_M \boldsymbol{\alpha}_M + \mathbf{X}_B \delta_\beta + \mathbf{X}_{AB} \delta_{\alpha\beta} + \mathbf{X}_E \boldsymbol{\epsilon}). \end{aligned} \quad (4.5)$$

Since \mathbf{e}_M has three random components, $\mathbf{e}_M^T \mathbf{e}_M$ in the quadratic form of \mathbf{y}_p in which the coefficients matrices of the projection model are orthogonal is available for estimating their variance components. Denoting $\mathbf{e}_M^T \mathbf{e}_M$ as RSS_M ,

$$RSS_M = \mathbf{e}_M^T \mathbf{e}_M, \quad (4.6)$$

where RSS_M measures the variation due to the three random effects, and thus, the quantity is used for the estimation of three variance components σ_β^2 , $\sigma_{\alpha\beta}^2$, and σ_ϵ^2 . Representing the residual random vector \mathbf{e}_B as \mathbf{y}_p , has two random components as follows.

$$\begin{aligned} \mathbf{e}_B &= (\mathbf{I} - \mathbf{X}_B \mathbf{X}_B^-) \mathbf{e}_M \\ &= (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^- - \mathbf{X}_B \mathbf{X}_B^-) (\mathbf{X}_{AB} \delta_{\alpha\beta} + \mathbf{X}_E \boldsymbol{\epsilon}). \end{aligned} \quad (4.7)$$

Therefore, $\mathbf{e}_B^T \mathbf{e}_B$ is used as a variation quantity for two random effects vectors. Denoting $\mathbf{e}_B^T \mathbf{e}_B$ as RSS_B ,

$$\begin{aligned} RSS_B &= \mathbf{e}_B^T \mathbf{e}_B \\ &= \mathbf{y}_p^T (\mathbf{I} - \mathbf{X}_M \mathbf{X}_M^- - \mathbf{X}_B \mathbf{X}_B^-) \mathbf{y}_p, \end{aligned} \quad (4.8)$$

where RSS_B is used for estimating the two variance components $\sigma_{\alpha\beta}^2$ and σ_ϵ^2 since \mathbf{e}_B has just two random effects. Finally, expressing \mathbf{e}_{AB} as \mathbf{y}_p ,

$$\mathbf{e}_{AB} = (\mathbf{I} - \mathbf{X}_{AB} \mathbf{X}_{AB}^-) \mathbf{e}_B = \mathbf{X}_E \boldsymbol{\epsilon}, \quad (4.9)$$

which has just one random component ϵ . Therefore, $e_{AB}^T e_{AB}$ shows the variation due to the random error vector only, and this quantity is used for estimating the variance component σ_ϵ^2 . Denoting $e_{AB}^T e_{AB}$ as RSS_{AB} ,

$$RSS_{AB} = e_{AB}^T e_{AB}. \quad (4.10)$$

Therefore, RSS_M , RSS_B , and RSS_{AB} are another set of sums of squares for estimating variance components instead of using sums of squares derived from the projections as an alternative method. RSS_M , RSS_B , and RSS_{AB} are also evaluated by the eigenvalues and eigenvectors of the projection matrices associated with the quadratic forms in y_p . Now, the expected values of the RSS 's are

$$\begin{aligned} E(RSS_M) &= \text{tr}((I - X_M X_M^-) \Sigma) \\ &= \sigma_\beta^2 r_{M_B} + \sigma_{\alpha\beta}^2 r_{M_{AB}} + \sigma_\epsilon^2 r_{M_E}; \\ E(RSS_B) &= \text{tr}((I - X_B X_B^-)(I - X_M X_M^-) \Sigma) \\ &= \sigma_{\alpha\beta}^2 r_{B_{AB}} + \sigma_\epsilon^2 r_{B_E}; \\ E(RSS_{AB}) &= \text{tr}((I - X_{AB} X_{AB}^-)(I - X_B X_B^-)(I - X_M X_M^-) \Sigma) \\ &= \sigma_\epsilon^2 r_{AB_E}. \end{aligned} \quad (4.11)$$

Then, the linear equations of variance components are obtained by equating the RSS 's to their expected values, the solutions for which always produce nonlinear estimates. That is,

$$\begin{aligned} RSS_M &= \hat{\sigma}_\beta^2 r_{M_B} + \hat{\sigma}_{\alpha\beta}^2 r_{M_{AB}} + \hat{\sigma}_\epsilon^2 r_{M_E}; \\ RSS_B &= \hat{\sigma}_{\alpha\beta}^2 r_{B_{AB}} + \hat{\sigma}_\epsilon^2 r_{B_E}; \\ RSS_{AB} &= \hat{\sigma}_\epsilon^2 r_{AB_E}. \end{aligned} \quad (4.12)$$

Even though two systems of linear equations are not the same, either system will produce the same estimates of the variance components that are nonnegative. As another method, projection method III is also available for the estimation of variance components. This method is done as follows. For the model of (3.1), $y = X\theta + \epsilon$, where $X = (j, X_F, X_\beta, X_{\alpha\beta})$ and $\theta = (\mu, \alpha_F, \delta_\beta, \delta_{\alpha\beta})^T$. This method splits the vector space of an observation vector into two subspaces, one for the projection part and the other for the error part at each step. Then, the projection of y onto the subspace spanned by XX^- is given by XX^-y , and the error vector in the error vector space is $(I - XX^-)y$. Therefore, the coefficient matrix of ϵ is derived as $(I - XX^-)$ from it. The quadratic form $y'(I - XX^-)y$ denoted by BSS_0 is the sum of squares due to random error only, which has all the information about σ_ϵ^2 . For information about both $\sigma_{\alpha\beta}^2$ and σ_ϵ^2 , the vector space of the observation vector can be decomposed into two parts one for the projection part and the other for the error part. For this, the model to be fitted is $y = X_1\theta_1 + \epsilon_1$, where $X_1 = (j, X_F, X_\beta)$, $\theta_1 = (\mu, \alpha_F, \delta_\beta)^T$, and $\epsilon_1 = X_{\alpha\beta}\delta_{\alpha\beta} + \epsilon$. Then, the projection of y onto the subspace spanned by $X_1X_1^-$ is given by $X_1X_1^-y$, and the error vector in the error vector space is $(I - X_1X_1^-)y$. The quadratic form $y^T(I - X_1X_1^-)y$ denoted by BSS_1 has information about $\sigma_{\alpha\beta}^2$ and σ_ϵ^2 . Now, the error vector is represented by

$$\begin{aligned} (I - X_1X_1^-)y &= (I - X_1X_1^-)(X_1\theta_1 + \epsilon_1) \\ &= (I - X_1X_1^-)X_{\alpha\beta}\delta_{\alpha\beta} + (I - X_1X_1^-)\epsilon. \end{aligned} \quad (4.13)$$

Therefore, the coefficient matrix of $\delta_{\alpha\beta}$ is given by $(I - X_1X_1^-)X_{\alpha\beta}$. For information about three variance components σ_ϵ^2 , $\sigma_{\alpha\beta}^2$, and σ_β^2 , the vector space can be divided into two subspaces considering

the model matrix of the equation, $\mathbf{y} = \mathbf{X}_2\boldsymbol{\theta}_2 + \boldsymbol{\epsilon}_2$, where $\mathbf{X}_2 = (\mathbf{j}, \mathbf{X}_F)$, $\boldsymbol{\theta}_2 = (\mu, \boldsymbol{\alpha}_F)^T$, and $\boldsymbol{\epsilon}_2 = \mathbf{X}_\beta\boldsymbol{\delta}_\beta + \mathbf{X}_{\alpha\beta}\boldsymbol{\delta}_{\alpha\beta} + \boldsymbol{\epsilon}$. Then, the projection of \mathbf{y} onto the subspace spanned by $\mathbf{X}_2\mathbf{X}_2^-$ is given by $\mathbf{X}_2\mathbf{X}_2^-\mathbf{y}$, and the error vector in the error vector space is $(\mathbf{I} - \mathbf{X}_2\mathbf{X}_2^-)\mathbf{y}$. The quadratic form $\mathbf{y}'(\mathbf{I} - \mathbf{X}_2\mathbf{X}_2^-)\mathbf{y}$ denoted by \mathbf{BSS}_2 has information about σ_ϵ^2 , $\sigma_{\alpha\beta}^2$, and σ_β^2 . Now, the error vector is represented by

$$\begin{aligned} (\mathbf{I} - \mathbf{X}_2\mathbf{X}_2^-)\mathbf{y} &= (\mathbf{I} - \mathbf{X}_2\mathbf{X}_2^-)(\mathbf{X}_2\boldsymbol{\theta}_2 + \boldsymbol{\epsilon}_2) \\ &= (\mathbf{I} - \mathbf{X}_2\mathbf{X}_2^-)(\mathbf{X}_\beta\boldsymbol{\delta}_\beta + \mathbf{X}_{\alpha\beta}\boldsymbol{\delta}_{\alpha\beta} + \boldsymbol{\epsilon}). \end{aligned} \quad (4.14)$$

Therefore, the coefficient matrix of $\boldsymbol{\delta}_\beta$ is given by $(\mathbf{I} - \mathbf{X}_2\mathbf{X}_2^-)\mathbf{X}_\beta$. It is necessary to evaluate the expected values of the quadratic forms for constructing the equations for the variance components. They are

$$\begin{aligned} E(\mathbf{BSS}_2) &= \sigma_\beta^2 c_{2\beta} + \sigma_{\alpha\beta}^2 c_{2\alpha\beta} + \sigma_\epsilon^2 c_{2\epsilon}; \\ E(\mathbf{BSS}_1) &= \sigma_{\alpha\beta}^2 c_{2\alpha\beta} + \sigma_\epsilon^2 c_{2\epsilon}; \\ E(\mathbf{BSS}_0) &= \sigma_\epsilon^2 c_{2\epsilon}. \end{aligned} \quad (4.15)$$

The nonnegative estimates of variance components are given as solutions of linear equations of $\hat{\sigma}_\beta^2$, $\hat{\sigma}_{\alpha\beta}^2$, and $\hat{\sigma}_\epsilon^2$. The above equations are summarized as:

$$\begin{aligned} \mathbf{BSS}_2 &= \hat{\sigma}_\beta^2 c_{2\beta} + \hat{\sigma}_{\alpha\beta}^2 c_{2\alpha\beta} + \hat{\sigma}_\epsilon^2 c_{2\epsilon}; \\ \mathbf{BSS}_1 &= \hat{\sigma}_{\alpha\beta}^2 c_{2\alpha\beta} + \hat{\sigma}_\epsilon^2 c_{2\epsilon}; \\ \mathbf{BSS}_0 &= \hat{\sigma}_\epsilon^2 c_{2\epsilon}, \end{aligned} \quad (4.16)$$

where c_{ij} 's are coefficients of variance components of expected values of quadratic forms of (4.15).

5. Examples

Montgomery (2013)'s data are illustrated as a first example of nonnegative estimates of random effects for a two-way mixed model. The data are from an experiment for a gauge capability study where parts are randomly selected and three operators are fixed. An instrument or gauge is used to measure a critical dimension on a part. Twenty parts have been selected from the production process, and only three operators are assumed to use the gauge. The assumed model for the data in Table 1 is $y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \epsilon_{ijk}$, where the α_i ($i = 1, 2, 3$) are fixed effects such that $\sum_{i=1}^3 \alpha_i = 0$ and γ_j ($j = 1, 2, \dots, 20$), $(\alpha\gamma)_{ij}$, and ϵ_{ijk} are uncorrelated random variables having zero means and variances $\text{Var}(\gamma_j) = \sigma_\gamma^2$, $\text{Var}((\alpha\gamma)_{ij}) = \sigma_{\alpha\gamma}^2$, and $\text{Var}(\epsilon_{ijk}) = \sigma_\epsilon^2$. Under the assumed unrestricted model, estimated variance components are $\hat{\sigma}_\gamma^2 = 10.2798$, $\hat{\sigma}_{\alpha\gamma}^2 = -0.1399$, and $\hat{\sigma}_\epsilon^2 = 0.9917$. Applying the projection method I to the data, the linear equations of variance components are given as:

$$\begin{aligned} \mathbf{SS}_{\text{part}} &= 1185.425 = 114\hat{\sigma}_\gamma^2, \\ \mathbf{SS}_{\text{part}\times\text{operator}} &= 27.05 = 76\hat{\sigma}_{\alpha\gamma}^2, \\ \mathbf{SS}_{\text{error}} &= 59.5 = 60\hat{\sigma}_\epsilon^2. \end{aligned} \quad (5.1)$$

The solutions of the equations are $\hat{\sigma}_\gamma^2 = 10.3985$, $\hat{\sigma}_{\alpha\gamma}^2 = 0.3559$, and $\hat{\sigma}_\epsilon^2 = 0.9917$. All the variance components are estimated nonnegatively. When we apply projection method II to the same data, we

Table 1: Data for a measurement systems capability study from Montgomery (2013)

Part number	Operator 1	Operator 2	Operator 3
1	21, 20	20, 20	19, 21
2	24, 23	24, 24	23, 24
3	20, 21	19, 21	20, 22
4	27, 27	28, 26	27, 28
5	19, 18	19, 18	18, 21
6	23, 21	24, 21	23, 22
7	22, 21	22, 24	22, 20
8	19, 17	18, 20	19, 28
9	24, 23	25, 23	24, 24
10	25, 23	26, 25	24, 25
11	21, 20	20, 20	21, 20
12	18, 19	17, 19	18, 19
13	23, 25	25, 25	25, 25
14	24, 24	23, 25	24, 25
15	29, 30	30, 28	31, 30
16	26, 26	25, 26	25, 27
17	20, 20	19, 20	20, 20
18	19, 21	19, 19	21, 23
19	25, 26	25, 24	25, 25
20	19, 19	18, 17	19, 17

Table 2: Hypothetical data of a one-way classification from Searle and Gruber (2016)

Class	Observations
1	19, 17, 15
2	25, 5, 15

get

$$\begin{aligned}
 \mathbf{RSS}_{\text{fixed}} &= 114\hat{\sigma}_\gamma^2 + 76\hat{\sigma}_{\alpha\gamma}^2 + 60\hat{\sigma}_\epsilon^2, \\
 \mathbf{RSS}_{\text{part}} &= 76\hat{\sigma}_{\alpha\gamma}^2 + 60\hat{\sigma}_\epsilon^2, \\
 \mathbf{RSS}_{\text{part} \times \text{operator}} &= 60\hat{\sigma}_\epsilon^2,
 \end{aligned}
 \tag{5.2}$$

where $\mathbf{RSS}_{\text{fixed}} = 1271.975$, $\mathbf{RSS}_{\text{part}} = 86.55$, and $\mathbf{RSS}_{\text{part} \times \text{operator}} = 59.5$. The solutions for the equations are $\hat{\sigma}_\gamma^2 = 10.3985$, $\hat{\sigma}_{\alpha\gamma}^2 = 0.3559$, and $\hat{\sigma}_\epsilon^2 = 0.9917$, which are the same as the previous solutions. Therefore, either one of the projection methods can be used for the nonnegative estimation of variance components of random effects in a mixed model. Projection method III also gives the same result as projection methods I and II for the data.

As a second example, Searle and Gruber (2016)'s hypothetical data are illustrated. Searle explains why a negative estimate can occur in the estimation of variance component of random effects in a random model. Table 2 shows the data where the one-way random-effects model is assumed since class is a random factor. The assumed model is $y_{ij} = \mu + \delta_i + \epsilon_{ij}$, where the δ_i ($i = 1, 2$) are random effects and ϵ_{ij} are uncorrelated random errors having zero means and variances $\text{Var}(\delta_i) = \sigma_\delta^2$ and $\text{Var}(\epsilon_{ij}) = \sigma^2$. As a result of the analysis of variance, the estimates of variance components are given as $\hat{\sigma}_\delta^2 = -15.33$ and $\hat{\sigma}_\epsilon^2 = 52$. Searle demonstrated how negative estimates could come from the analysis of variance and insisted that there would be nothing intrinsic in the method to prevent it. However, the projection methods yield the same nonnegative estimates as $\hat{\sigma}_\delta^2 = 2$ and $\hat{\sigma}_\epsilon^2 = 52$ in any method.

6. Discussion

Variance should be a nonnegative quantity as a measure of variation in data by its definition. In this work, it shows that orthogonal projections are useful for defining a projection model for nonnegative variance estimation. There have been many unsuccessful attempts in literature to fix the problem of negative estimates for variance components over decades; however, the proposed methods in this paper always produce nonnegative estimates of variance components of the random effects in a mixed model. The two most important findings are checked and discussed for the estimation of the nonnegative variance component. One is that a projection model should be derived from an assumed mixed-effects model. The other is that expectations of quadratic forms associated with the random effects should be evaluated from the projection model.

This paper introduces terms such as projection method I, II, and III related to the methods, and the projection model for emphasizing projection rather than model fitting. The three methods are applied differently in the application even though they are based on the same assumed model. Each method uses projections in a different way, but summing up all orthogonal projections come to the observation vector. Depending on the types of projections, each method produces three different sets of equations for the evaluation of quadratic forms. Nonetheless, all of them show the same nonnegative estimates for variance components. It also shows that projection methods can be used for estimating variance components of the random effects in either random model or mixed model through examples. It should be noted that all the matrices associated with the quadratic forms come from the projection model and not from the assumed model. Hartley's synthesis can yield correct coefficients of variance components with the orthogonal coefficient matrices.

Acknowledgements

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant No.2018R1D1A1B07043021.

References

- Graybill FA (1983). *Matrices with Applications in Statistics*, Wadsworth, California.
- Hartley HO (1967). Expectations, variances and covariances of ANOVA mean squares by "synthesis", *Biometrics*, **23**, 105–114.
- Harville DA (1969). Variance component estimation for the unbalanced one-way random classification—a critique, *Aerospace Research Laboratories*, **ARL-69(0180)**.
- Henderson CR (1953). Estimation of variance and covariance components, *Biometrics*, **9**, 226–252.
- Hill BM (1965). Inference about variance components in the one-way model, *Journal of the American Statistical Association*, **60**, 806–825.
- Hill BM (1967). Correlated errors in the random model, *Journal of the American Statistical Association*, **62**, 1387–1400.
- Johnson RA and Wichern DW (2014). *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey.
- Milliken GA and Johnson DE (1984). *Analysis of Messy Data*, Van Nostrand Reinhold, New York.
- Montgomery DC (2013). *Design and Analysis of Experiments*, John Wiley & Sons, New York.
- Nelder JA (1954). The interpretation of negative components of variance, *Biometrika*, **41**, 544–548.
- Searle SR (1971). *Linear Models*, John Wiley & Sons, New York.
- Searle SR, Casella G, and McCulloch CE (2009). *Variance Components*, John Wiley & Sons, New

York.

Searle SR and Fawcett RF (1970). expected mean squares in variance components models having finite populations, *Biometrics*, **26**, 243–254.

Searle SR and Gruber MHG (2016). *Linear Models*, John Wiley & Sons, New York.

Thompson Jr WA (1961). Negative estimates of variance components: an introduction, *Bulletin, International Institute of Statistics*, **34**, 1–4.

Thompson Jr WA (1963). Non-negative estimates of variance components, *Technometrics*, **5**, 441–450.

Received April 20, 2020; Revised July 1, 2020; Accepted July 26, 2020