University of Groningen

On Drawbacks of Least Squares Lehmann-Scheffé Estimation of Variance Components

Žežula, Ivan; Klein, Daniel

Published in:
Metron

DOI:
[10.1007/s40300-021-00196-8](https://doi.org/10.1007/s40300-021-00196-8)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Žežula, I., & Klein, D. (2021). On Drawbacks of Least Squares Lehmann-Scheffé Estimation of Variance Components. *Metron*, 79(1), 109-119. <https://doi.org/10.1007/s40300-021-00196-8>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

On drawbacks of least squares Lehmann–Scheffé estimation of variance components

Ivan Žežula¹ · Daniel Klein¹

Received: 13 June 2019 / Accepted: 4 January 2021 / Published online: 19 January 2021
© Sapienza Università di Roma 2021

Abstract

Estimation of variance components is one of the basic problems in linear models with mixed effects, and a vast literature exists on the subject. Unfortunately, there are only few situations in which uniformly best estimators exist, which usually results into need of using an iterative estimation procedure. A new non-iterative method, called least squares Lehmann–Scheffé method, was proposed and its superiority over commonly accepted methods was claimed. Since there was no scientific response to these claims, we decided to analyze it thoroughly.

Keywords Least squares estimation · Mixed linear model · Variance components

1 Introduction

Estimation of variance components is one of the basic problems in linear models with mixed effects (LMM—linear mixed models), and a vast literature exists on the subject. Brief summary of the model and methods used is contained in encyclopedic reference entry [6]. More readable overview with some applications can be found in Chapter XII of textbook [1]. One of the best reviews of history, development, and methods of the subject is the monograph [7]. One can find extensive literature review therein. Another, more application oriented book is [5]. Most recent review of modern methods and computational tools in the area of LMM is book [2].

General form of mixed linear model can be written as

$$Y = X\beta + \sum_{i=1}^r Z_i\alpha_i + \varepsilon, \quad (1)$$

This work was supported by the Slovak Research and Development Agency under the contract no. APVV-17-0568.

✉ Ivan Žežula
ivan.zezula@upjs.sk

¹ Institute of Mathematics, University of P. J. Šafárik, Košice, Slovakia

where X and $Z_1, \dots, Z_r$ are known design matrices, β are fixed effects, $\alpha_1, \dots, \alpha_r$ random effects, and ε random error. Usual assumptions are

$$\begin{aligned} E\varepsilon &= 0, & \text{var } \varepsilon &= \sigma_e^2 I_n, \\ E\alpha_i &= 0 \quad \forall i, & \text{var } \alpha_i &= \sigma_i^2 I_{r_i} \quad \forall i, \\ \text{cov}(\alpha_i, \alpha_j) &= 0 \quad \forall i \neq j, & \text{cov}(\alpha_i, \varepsilon) &= 0 \quad \forall i. \end{aligned}$$

Thus,

$$EY = X\beta, \quad \text{var } Y = \sum_{i=1}^r \sigma_i^2 Z_i Z_i' + \sigma_e^2 I_n.$$

The second order parameters σ_e^2 and $\sigma_1^2, \dots, \sigma_r^2$ are called variance components.

Example Blood pressure studies were made at 15 different, randomly chosen clinics throughout New York City, with 5 patients on each of 4 treatments (placebo and 3 medications) at each clinic. In this case a suitable model equation for the k th patient on treatment i at clinic j would be

$$Ey_{ijk} = \mu + a_i + b_j + c_{ij}, \quad i = 1, \dots, 4, \quad j = 1, \dots, 15, \quad k = 1, \dots, 5.$$

The research interest is the efficacy of treatments— a_i are fixed effects. Since there are much more clinics in New York City, we do not observe the whole population, clinic and interaction effects (b_j, c_{ij}) are random. For technical details, see Example 7, p. 14, in [7].

The main research interest is inference on the mean value, and prediction of the random effects. The problem is that for any inference (estimation, confidence intervals, testing) we need inversion of the variance matrix. Thus, unlike in simple regression model, we cannot avoid estimating variance components.

Several methods of variance components estimation have been developed. The best known is naturally maximum likelihood (ML) method and restricted maximum likelihood (REML) method. They have assured asymptotic optimality properties, but are iterative and face many problems of practical application (e.g. problem of initial values and problem of possible negative estimates). The main advantage of MINQE method (see [6]) is some degree of optimality even in non-normal case, but it requires choosing some prior weights. Iterated MINQE converges to REML estimators. Henderson's Method 3 is very simple to compute, but it is ambiguous in more complicated designs and has no assured optimality properties.

If we are not lucky enough to have a very nice and simple design, where uniformly best unbiased estimators exists (e.g. $r = 1, Z_1 = I$), then we have two options: either to use a general method of estimation such as ML, REML, MINQE, or Henderson method 3, or to use some locally best estimators, where we have to choose prior values of the parameters. That means, with the exceptions of Henderson Method 3, that we always have to choose some starting (prior) values of estimated parameters and/or go through an iterative estimation procedure.

Un-noticed or ignored in recent texts on the subject is the method of [8]. This is a relatively new method of estimation of variance components, which eliminates both above mentioned problems: no prior values of parameters are needed and the method is non-iterative. It can be applied in any linear model with mixed effects. According to the author, it is based on the Lehmann–Scheffé theorem. Since it uses least-squares principle, the proposed method is called the least-squares Lehmann–Scheffé method (LSLS method).

Slanger [8] also compared his method with Henderson's Method 3 (see [3]) in one simple model, and in [9] compared the method with REML in six simulated designs. Superiority

of the new method was claimed in all cases. If it were true, it would really simplify life to many applied statisticians. Since there seems to be no response to the method in the current literature, we decided to analyze the method more thoroughly.

2 Principle of the method

Let us consider the simplest case of mixed linear model with just one random effect:

$$Y = X\beta + Z\alpha + \varepsilon, \quad (2)$$

so that

$$\begin{aligned} E(Y) &= X\beta, \\ \text{var}(Y) &= \sigma_\alpha^2 Z Z' + \sigma_\varepsilon^2 I. \end{aligned} \quad (3)$$

Let P_X and $M_X = I - P_X$ be the matrices of orthogonal projection onto the subspace generated by the columns of matrix X and onto its orthogonal complement, respectively. Slanger [8] proposed to estimate the variance components by functions

$$\hat{\sigma}_\varepsilon^2 = Y' M_X Q_\varepsilon M_X Y \quad \text{and} \quad \hat{\sigma}_\alpha^2 = Y' M_X Q_\alpha M_X Y \quad (4)$$

(matrices Q_α and Q_ε are considered symmetric w.l.o.g.). Since

$$\begin{aligned} E(Y' M_X Q M_X Y) &= \text{Tr} [Q M_X \text{var}(Y) M_X] \\ &= \sigma_\alpha^2 \text{Tr} [Q M_X Z Z' M_X] + \sigma_\varepsilon^2 \text{Tr} [Q M_X] \\ &= \sigma_\alpha^2 [\text{vec}(M_X Z Z' M_X)]' \text{vec}(Q) + \sigma_\varepsilon^2 [\text{vec}(M_X)]' \text{vec}(Q) \end{aligned} \quad (5)$$

(the $\text{vec}()$ operator working column-wise), unbiased estimating matrices Q_ε and Q_α must fulfill

$$\begin{aligned} [\text{vec}(M_X)]' \text{vec}(Q_\varepsilon) &= 1 \quad \text{and} \quad [\text{vec}(M_X Z Z' M_X)]' \text{vec}(Q_\varepsilon) = 0, \\ [\text{vec}(M_X)]' \text{vec}(Q_\alpha) &= 0 \quad \text{and} \quad [\text{vec}(M_X Z Z' M_X)]' \text{vec}(Q_\alpha) = 1. \end{aligned}$$

According to Lehmann–Scheffé theorem (see [4]), unbiased estimator T of some statistic $h(\theta)$ is uniformly minimum variance unbiased estimator (UMVUE) in a class $\mathcal{T}$, if it is uncorrelated with every statistic $Z \in \mathcal{T}$ with zero mean value and finite variance. The author considered only statistics Z of the form $Y' M_X S M_X Y$, where S is some appropriate matrix. Because of zero mean value, matrix S must fulfill

$$\left(\begin{bmatrix} [\text{vec}(M_X)]' \\ [\text{vec}(M_X Z Z' M_X)]' \end{bmatrix} \right) \text{vec}(S) \stackrel{\text{df}}{=} B \text{vec}(S) = 0, \quad (6)$$

i.e., $\text{vec}(S)$ must lie in the nullspace of B . Let $\text{vec}(C_1), \dots, \text{vec}(C_m)$ be a basis of this nullspace, then $S = \sum_{i=1}^m k_i C_i$, where k_i are some real numbers. Under normality, the covariance between $Y' M_X Q M_X Y$ and $Y' M_X S M_X Y$ is

$$\begin{aligned} \text{cov}(Y' M_X Q M_X Y, Y' M_X S M_X Y) &= 2\sigma_\alpha^4 [\text{vec}(M_X Z Z' M_X S M_X Z Z' M_X)]' \text{vec}(Q) \\ &\quad + 2\sigma_\alpha^2 \sigma_\varepsilon^2 [\text{vec}(M_X Z Z' M_X S M_X + M_X S M_X Z Z' M_X)]' \text{vec}(Q) \\ &\quad + 2\sigma_\varepsilon^4 [\text{vec}(M_X S M_X)]' \text{vec}(Q). \end{aligned}$$

Then, LSLS-estimator is uncorrelated with all statistics Z if it holds $\forall i$

$$\begin{aligned} [\text{vec}(\mathbf{M}_X \mathbf{Z} \mathbf{Z}' \mathbf{M}_X \mathbf{C}_i \mathbf{M}_X \mathbf{Z} \mathbf{Z}' \mathbf{M}_X)]' \text{vec}(\mathbf{Q}) &= 0, \\ [\text{vec}(\mathbf{M}_X \mathbf{Z} \mathbf{Z}' \mathbf{M}_X \mathbf{C}_i \mathbf{M}_X + \mathbf{M}_X \mathbf{C}_i \mathbf{M}_X \mathbf{Z} \mathbf{Z}' \mathbf{M}_X)]' \text{vec}(\mathbf{Q}) &= 0, \\ [\text{vec}(\mathbf{M}_X \mathbf{C}_i \mathbf{M}_X)]' \text{vec}(\mathbf{Q}) &= 0. \end{aligned} \quad (7)$$

If we denote $\mathbf{C}_v = (\text{vec}(\mathbf{C}_1), \dots, \text{vec}(\mathbf{C}_m))$, then Eq. (7) can be rewritten in the form

$$\begin{pmatrix} \mathbf{C}'_v (\mathbf{M}_X \mathbf{Z} \mathbf{Z}' \mathbf{M}_X \otimes \mathbf{M}_X \mathbf{Z} \mathbf{Z}' \mathbf{M}_X) \\ \mathbf{C}'_v (\mathbf{M}_X \mathbf{Z} \mathbf{Z}' \mathbf{M}_X \otimes \mathbf{M}_X + \mathbf{M}_X \otimes \mathbf{M}_X \mathbf{Z} \mathbf{Z}' \mathbf{M}_X) \\ \mathbf{C}'_v (\mathbf{M}_X \otimes \mathbf{M}_X) \end{pmatrix} \times \text{vec}(\mathbf{Q}) \stackrel{\text{df}}{=} \mathbf{R}_0 \text{vec}(\mathbf{Q}) = 0,$$

where $\otimes$ is the Kronecker product. Let $\mathbf{e}_i$ denote the vector of appropriate dimension containing 1 at i -th place and zeros otherwise. Negative i denotes position counted from the end (-1 being the last position). Thus, LSLS-estimator must satisfy

$$\begin{pmatrix} \mathbf{B} \\ \mathbf{R}_0 \end{pmatrix} \text{vec}(\mathbf{Q}) \stackrel{\text{df}}{=} \mathbf{R} \text{vec}(\mathbf{Q}) = \mathbf{h}, \quad (8)$$

where $\mathbf{h} = \mathbf{e}_2$ for $\mathbf{Q}_\alpha$, and $\mathbf{h} = \mathbf{e}_1$ for $\mathbf{Q}_\varepsilon$. A least-squares solution is

$$\text{vec}(\hat{\mathbf{Q}}) = (\mathbf{R}' \mathbf{R})^{-1} \mathbf{R}' \mathbf{h}.$$

If $\mathbf{R} \text{vec}(\hat{\mathbf{Q}}) - \mathbf{h} = 0$, UMVUE is claimed to be found. If $\mathbf{R} \text{vec}(\hat{\mathbf{Q}}) - \mathbf{h} \neq 0$, MP-inverse for $\text{vec}(\hat{\mathbf{Q}})$ is used. This solution is called LSLS-estimator in both cases (also denoted by LSLS_b because of possible bias). LSLS-estimator is claimed to be the estimator which is as close as analytically possible to satisfying the Lehmann–Scheffé criterion.

Because exact solution of system (8) exists only in special cases, the author proposed to use Lagrange multipliers method to assure unbiasedness of the estimator together with almost zero covariance property. This estimator, denoted by LSLS_u , is the solution of system

$$\begin{pmatrix} \mathbf{R}'_0 \mathbf{R}_0 & \mathbf{B}' \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \text{vec}(\mathbf{Q}) \\ \lambda \end{pmatrix} = \mathbf{h}, \quad (9)$$

where $\mathbf{h} = \mathbf{e}_{-1}$ for $\mathbf{Q}_\alpha$ and $\mathbf{h} = \mathbf{e}_{-2}$ for $\mathbf{Q}_\varepsilon$.

Generalization of the method to more variance components is simple. Every row of the matrix $\mathbf{B}$ in (6) must correspond to one variance component, there will be more restrictions to zero covariance in (7), and more right-hand-side vectors $\mathbf{h} = \mathbf{e}_i$ have to be used.

3 Analysis of the method

Although promised features of the method seem to be nice, and simulations done by the author look like confirming it, the reality is rather different. The method has some serious drawbacks, which will be analyzed in detail. Main problems are:

1. no reasonable properties of the estimators are assured;
2. mean value of an estimator can be a function of true values of all variance components;
3. the method tries to provide uniformly best estimators even if they do not exist;
4. the estimators are not unique.

Optimality properties of LSLS estimators

Author's hope for the good/optimality properties of the estimators are based on Lehmann–Scheffé theorem. It is true, that it suffices to consider quadratic estimators in the case of normally distributed vector of observations $\mathbf{Y}$. However, the theorem guarantees the optimality in the given family of estimators only in the case, when the assumptions are met (i.e. zero covariance with all estimators of zero), and does not say anything about the other cases. As a consequence, no optimality properties are guaranteed in the case when approximate data-driven solution (“almost zero covariance”) is used. Metric in the space of distributions can be quite different than that in data space. This holds for both LSLS_b and LSLS_u estimators.

The author chose to use matrix $\mathbf{M}_X$ to transform the values of $\mathbf{Y}$. In fact, such projection leads to translation-invariant estimators. But this choice of the projector does not take into consideration the information about covariance between the elements of $\mathbf{Y}$. If we want invariance, it would be better to use $\mathbf{M}_X^{V^{-1}} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$ with $\mathbf{V} = \text{var}(\mathbf{Y})$, instead. But to use that, we need to know prior values of estimated parameters (or their first stage estimates), and it was not wished. This is how the problem of prior values was avoided, but not solved.

However, there is another—and bigger—problem with this transformation. The method is not applicable in models, where invariant unbiased estimators of variance components do not exist. If we consider model (1) with mean and variance structure (3), general theory of variance components models (see e.g. [6]) says that function $\gamma = f_1'\sigma_\alpha^2 + f_2'\sigma_\varepsilon^2$ is unbiasedly estimable if and only if $\mathbf{f} = (f_1, f_2)' \in \mathcal{R}(\mathbf{H})$, where

$$\mathbf{H} = \begin{pmatrix} \text{Tr}(\mathbf{Z}\mathbf{Z}'\mathbf{Z}\mathbf{Z}' - \mathbf{P}_X\mathbf{Z}\mathbf{Z}'\mathbf{P}_X\mathbf{Z}\mathbf{Z}') & \text{Tr}(\mathbf{M}_X\mathbf{Z}\mathbf{Z}') \\ \text{Tr}(\mathbf{M}_X\mathbf{Z}\mathbf{Z}') & \text{Tr}(\mathbf{M}_X) \end{pmatrix},$$

and $\mathcal{R}(\cdot)$ denotes the range of a matrix. Function $\gamma = \mathbf{f}'\boldsymbol{\theta}$ is unbiasedly and invariantly (with respect to the mean shift) estimable if and only if $\mathbf{f} \in \mathcal{R}(\mathbf{H}_I)$, where

$$\mathbf{H}_I = \begin{pmatrix} \text{Tr}(\mathbf{M}_X\mathbf{Z}\mathbf{Z}') & \text{Tr}(\mathbf{M}_X\mathbf{Z}\mathbf{Z}') \\ \text{Tr}(\mathbf{M}_X\mathbf{Z}\mathbf{Z}') & \text{Tr}(\mathbf{M}_X) \end{pmatrix}.$$

Let us now consider the case, when a function γ is unbiasedly estimable, but not invariantly estimable, i. e. $\mathbf{f} \in \mathcal{R}(\mathbf{H})$ & $\mathbf{f} \notin \mathcal{R}(\mathbf{H}_I)$.

Using LSLS method, variance components are estimated in the model

$$\mathbf{Y}^* = \mathbf{M}_X\mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}^*,$$

where $\mathbf{Y}^* = \mathbf{M}_X\mathbf{Y}$, and $\boldsymbol{\varepsilon}^* = \mathbf{M}_X\boldsymbol{\varepsilon}$. In this model holds

$$\mathbf{E}(\mathbf{Y}^*) = \mathbf{0} \quad \text{and} \quad \text{var}(\mathbf{Y}^*) = \sigma_\alpha^2\mathbf{M}_X\mathbf{Z}\mathbf{Z}'\mathbf{M}_X + \sigma_\varepsilon^2\mathbf{M}_X.$$

Function γ is unbiasedly estimable if and only if $\mathbf{f} \in \mathcal{R}(\mathbf{H}^*)$, where $\mathbf{H}^*$ is constructed as above. But in this model $\mathbf{X}^* = \mathbf{0}$, so that $\mathbf{H}^* = \mathbf{H}_I$. Thus, no unbiased invariant estimator of γ exists, but LSLS-method still claims to provide such a one. The problem is that pre-multiplying of $\mathbf{Y}$ by $\mathbf{M}_X$ automatically eliminates all non-invariant estimators.

Example Let us consider the model (1) with

$$\mathbf{X} = \mathbf{I}_2 \otimes \mathbf{1}_2, \mathbf{Z} = \mathbf{1}_2 \otimes \mathbf{I}_2,$$

where $\mathbf{1}_d$ is the vector of all 1's of appropriate dimension. It is easy to find that

$$\mathbf{H} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{H}_I = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}.$$

Thus, both variance components are unbiasedly estimable, but not invariantly and unbiasedly estimable. In fact, only sum of the variance components—up to a multiple—is invariantly unbiasedly estimable in this model.

Mean value of LSLS estimators

Since the system (8) need not be solved exactly and the mean of the estimators is given by (5), LSLS-estimator need not be unbiased. In fact, the mean value of the estimator in many situations is a linear combination of all variance components and the bias depends on the true values of all components. This can lead to severe bias, if the variance components differ substantially. Moreover, the mean depends in such a case also on the choice of the basis, see the paragraph on uniqueness.

Uniform optimality property

As it was already stated, in most models there exist only locally best estimators, and uniformly best estimators do not exist. The method does not take into consideration such a fact and it tries to come as close as possible to such non-existing estimators. That is a very hard task. Thus, relationship to alternative estimators is unclear.

Uniqueness of LSLS estimators

The LSLS estimator is obtained by solving the system (8). This system is not solvable for unbalanced models, as the author stated. But there exist also balanced models, for which the system (8) is not solvable. Example of such a model is the following one: $Y = X\beta + Z\alpha + \epsilon$, where

$$X = (\mathbf{1}_2, I_2) \otimes \mathbf{1}_6, Z = I_3 \otimes \mathbf{1}_4.$$

Therefore, the system can be “solved” in many cases only approximately. The following objections can be raised:

The author proposed the least-squares method, which leads to infinitely many Q matrices. One of the solutions is then chosen using Moore–Penrose g-inverse, which minimizes the Euclidean norm of the residual vector. This approach does not take into account variances of the observations; a variance-based norm could be more appropriate.

There is more to the last argument. Zero correlation estimators are denoted by T' and they are of the form $Y'M_X S M_X Y$, where S is the matrix determined from (6). But such a matrix is not unique. Therefore, the basis of the nullspace of the matrix B is found in order to cover all suitable matrices S . The problem is, that there are infinitely many bases of the nullspace of matrix B and it is not clear, which basis should be chosen. For example let us consider model (1), where

$$X = \mathbf{1}_4 \otimes \mathbf{1}_2, Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \otimes \mathbf{1}_2,$$

and let $\text{var}(\alpha) = \sigma_\alpha^2 I_3$ and $\text{var}(\epsilon) = \sigma_\epsilon^2 I_8$. Now, let us consider two different bases of nullspace of matrix B , C_v^1 and C_v^2 . The first one was the default basis given by Maple

ColumnSpace() procedure, the second one using GaussianElimination() procedure.

Using them, we obtain (exact form) LSLS_u estimators ${}_1\hat{\sigma}_\alpha^2$ and ${}_1\hat{\sigma}_\varepsilon^2$ or ${}_2\hat{\sigma}_\alpha^2$ and ${}_2\hat{\sigma}_\varepsilon^2$, where ${}_i\hat{\sigma}^2$ is LSLS_u estimator of variance component using basis $\mathbf{C}_v^i$. Then, we get two different pairs of estimators, and the question is which one to choose?

The first criterion can be the mean value. Using (5), we can see all estimators produced are unbiased. (Beware: LSLS_b produces both pairs biased, and different. Thus, the mean in this case also depends on the choice of the basis.)

The second criterion can be the variance or MSE in the biased case. Natural choice is the estimator with smaller variance/MSE. It is easy to derive that an estimator of the form $\mathbf{Y}'\mathbf{M}_X\mathbf{Q}\mathbf{M}_X\mathbf{Y}$ has under normality the variance

$$\begin{aligned} & 2\sigma_\alpha^4 \operatorname{Tr}(\mathbf{Q}\mathbf{M}_X\mathbf{Z}\mathbf{Z}'\mathbf{M}_X\mathbf{Q}\mathbf{M}_X\mathbf{Z}\mathbf{Z}'\mathbf{M}_X) \\ & + 4\sigma_\alpha^2\sigma_\varepsilon^2 \operatorname{Tr}(\mathbf{Q}\mathbf{M}_X\mathbf{Z}\mathbf{Z}'\mathbf{M}_X\mathbf{Q}\mathbf{M}_X) \\ & + 2\sigma_\varepsilon^4 \operatorname{Tr}(\mathbf{Q}\mathbf{M}_X\mathbf{Q}\mathbf{M}_X). \end{aligned}$$

Comparing the two variances, we get

$$\operatorname{var}({}_2\hat{\sigma}_\alpha^2) - \operatorname{var}({}_1\hat{\sigma}_\alpha^2) \doteq 0.1584179\sigma_\alpha^4 + 0.1794101\sigma_\alpha^2\sigma_\varepsilon^2 + 0.0975839\sigma_\varepsilon^4, \quad (10)$$

$$\operatorname{var}({}_2\hat{\sigma}_\varepsilon^2) - \operatorname{var}({}_1\hat{\sigma}_\varepsilon^2) \doteq -0.0000037\sigma_\alpha^4 + 0.0001242\sigma_\alpha^2\sigma_\varepsilon^2 + 0.2506588\sigma_\varepsilon^4, \quad (11)$$

where σ_α^2 and σ_ε^2 are true parameter values. The difference (10) is always positive, and (11) forms in 3D a saddle-shaped surface taking both signs in corresponding regions. Thus, basis $\mathbf{C}_v^1$ produces always better estimator of σ_α^2 than $\mathbf{C}_v^2$, and many times better estimator of σ_ε^2 . Maybe there is another basis which is uniformly better than the two ones chosen. But there is no clue how to optimize. Such a dependence on the choice of the basis is very unfortunate.

Last but not least, the method requires solving very large linear systems, unless the number of observation is small. The matrix of the system (8) has dimensions $(3n^2 - 4) \times n^2$, and $(n^2 + 2) \times (n^2 + 2)$ in the case of (9). That can make the method unmanageable with standard software even for middle-sized data (several hundreds of observations, or more).

Knowing that, the author suggested to reduce the number of equations in system (8). He suggested to use the vech operator in matrix $\mathbf{B}$ instead of vec operator. Vech operator makes a vector from any symmetric matrix stacking its columns one under another taking into consideration only elements on and under the diagonal. In such a way the number of columns in matrix $\mathbf{B}$ will reduce from original n^2 to $\frac{1}{2}n(n+1)$. The number of elements of basis of the nullspace of this matrix will reduce, too. But again, in the same model the vec-based estimators and vech-based estimators differ, and have different variances. It is not clear, which way gives estimators with smaller variance. There is no rule of thumb, since practical experience shows that both can be better.

4 Numerical example and simulations

To verify theoretical claims, we used data provided by [1, Example 12.7]. Interpretation (not provided with the data) can be similar to Example 9 from [7]: maternal ability of mice strains. Let y_{ijk} be the weight of k -th 5-weeks-old litter from mouse j of strain i . Appropriate model can be

$$y_{ijk} = \mu + \beta_i + \alpha_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1, 2, \quad j = 1, 2, 3, \quad k = 1, \dots, n_{ij}, \quad (12)$$

Table 1 Model (12), i.e. with interactions

Method	Mathematica		R				
	LSLS _b	LSLS _u	LSLS _b	LSLS _u	Henderson3	ML	REML
$\hat{\sigma}_\varepsilon^2$	71.73	71.90	72.44	72.60	71.66	64.49	67.55
$\hat{\sigma}_\alpha^2$	431.64	1125.82	644.75	1133.93	599.48	739.08	1112.78
$\hat{\sigma}_\gamma^2$	-8.83	-17.13	-11.47	-13.09	-11.24	0	0

Table 2 Model (13), i.e. without interaction

Method	Mathematica		R				
	LSLS _b	LSLS _u	LSLS _b	LSLS _u	Henderson3	ML	REML
$\hat{\sigma}_\varepsilon^2$	67.60	67.70	67.80	67.90	67.55	64.49	67.55
$\hat{\sigma}_\alpha^2$	623.80	1123.52	815.47	1114.06	1110.22	739.08	1112.78

$n_{ij} \in [3, 5, 5], [6, 3, 3]$, where β_i are fixed effects of the strains, α_j are random effects of the animals, and γ_{ij} are possible interaction effects. The true values of parameters are known to be $\mu + \beta_1 = 200$, $\mu + \beta_2 = 150$, $\sigma_\varepsilon^2 = 64$, $\sigma_\alpha^2 = 784$, and $\sigma_\gamma^2 = 25$.

An alternative model can be model without interactions:

$$y_{ijk} = \mu + \beta_i + \alpha_j + \varepsilon_{ijk}, \quad i = 1, 2, \quad j = 1, 2, 3, \quad k = 1, \dots, n_{ij}, \quad (13)$$

with the same n_{ij} .

The resulting parameter estimates using different methods for both models are summarized in Tables 1 and 2. For LSLS method, there are two sets of results: one computed in Wolfram Mathematica, and one computed in R. Both use default way of choosing the vector space basis in given environment. We can see that change of the basis can produce huge changes in the estimate values.

Observing large differences in estimates inspired us to conduct a small simulation study. We have considered two models:

- mixed effects two-way model (12) with fixed β and random α and γ , and
- random effects model, all as before but with random β .

Both models are very similar to that used for simulations in [9], but with slightly smaller number of categories and overall number of observations.

All random effects were simulated independent with normal distribution. 5000 estimates of unknown parameters were computed using LSLS_b-method, LSLS_u-method, ML, REML, and Henderson method 3 in all cases.

Summarized results from the mixed effects model are in Tables 3, 4 and 5. Because of the problem with negative values acquired by some estimators, we give both overall sample means and sample means of positive values only.

The same summary of the random effects model results is in Tables 6, 7, 8 and 9.

Our simulation did not confirm results of [9] that Slanger's method is superior to REML. According to our results, all methods have approximately the same precision for estimating σ_ε^2 . LSLS_u worked very badly in both models, estimates having hugely different values than those from the other methods (and the true parameter values). LSLS_u-estimates of σ_ε^2 were

Table 3 Mixed effects model, estimation of $\sigma_\alpha^2 = 784$

Method	Henderson3	LSLS _b	LSLS _u	ML	REML
Average	773.18	439.44	23.45	513.74	772.41
St. error	11.35	6.44	0.33	7.55	11.32
Average of > 0	796.78	451.54	23.71	529.95	794.82
St. error of > 0	11.51	6.52	0.33	7.67	11.49
Percentage of > 0	97.12%	97.40%	98.94%	96.94%	97.18%

Table 4 Mixed effects model, estimation of $\sigma_\gamma^2 = 25$

Method	Henderson3	LSLS _b	LSLS _u	ML	REML
Average	25.47	21.65	52.40	15.52	28.11
St. error	0.60	0.52	0.72	0.35	0.55
Average of > 0	42.85	36.81	59.11	28.25	42.03
St. error of > 0	0.73	0.63	0.73	0.52	0.71
Percentage of > 0	67.02%	66.56%	89.88%	54.94%	66.88%

Table 5 Mixed effects model, estimation of $\sigma_\varepsilon^2 = 64$

Method	Henderson3	LSLS _b	LSLS _u	ML	REML
Average	64.15	64.06	63.48	61.86	62.95
St. error	0.29	0.29	0.29	0.29	0.28
Average of > 0	64.15	64.06	63.48	61.86	62.95
St. error of > 0	0.29	0.29	0.29	0.29	0.28
Percentage of > 0	100%	100%	100%	100%	100%

Table 6 Random effects model, estimation of $\sigma_\beta^2 = 160$

Method	Henderson3	LSLS _b	LSLS _u	ML	REML
Average	171.39	91.59	29.19	107.29	169.06
St. error	4.41	2.21	0.32	2.20	3.50
Average of > 0	278.56	140.94	29.19	164.70	213.24
St. error of > 0	5.74	2.82	0.32	2.92	4.14
Percentage of > 0	66.92%	69.00%	100%	65.14%	79.28%

Table 7 Random effects model, estimation of $\sigma_\alpha^2 = 784$

Method	Henderson3	LSLS _b	LSLS _u	ML	REML
Average	762.07	579.83	157.18	527.33	762.84
St. error	11.05	8.44	2.03	7.52	11.05
Average of > 0	782.89	597.88	157.18	548.38	782.40
St. error of > 0	11.19	8.56	2.03	7.66	11.20
Percentage of > 0	97.42%	97.06%	100%	96.16%	97.50%

Table 8 Random effects model, estimation of $\sigma_v^2 = 25$

Method	Henderson3	LSLS _b	LSLS _u	ML	REML
Average	25.82	46.07	141.77	31.02	25.94
St. error	0.61	0.59	1.53	0.61	0.52
Average of > 0	42.64	49.65	141.77	44.30	38.96
St. error of > 0	0.74	0.60	1.53	0.77	0.67
Percentage of > 0	67.80%	93.56%	100%	70.02%	66.60%

Table 9 Random effects model, estimation of $\sigma_e^2 = 64$

Method	Henderson3	LSLS _b	LSLS _u	ML	REML
Average	64.13	64.61	173.01	63.07	62.89
St. error	0.29	0.29	1.65	0.29	0.28
Average of > 0	64.13	64.61	173.01	63.07	62.89
At. error of > 0	0.29	0.29	1.65	0.29	0.28
Percentage of > 0	100%	100%	100%	100%	100%

heavily biased even in the random effects model. The only advantage of the method seemed to be lesser number of negative estimates when compared with the other methods. LSLS_b-method was very unreliable, sometimes giving reasonable values and sometimes fairly off.

5 Summary

LSLS method has some advantages compared to widely used methods. It is not iterative and no prior values of unknown parameters are needed. But the method has also substantial drawbacks, which are the consequence of several theoretical mistakes. In fact, in many situations no optimality properties are assured, the estimators can be substantially biased, and they are not uniquely determined. Thus, estimators with quite different properties can be obtained from the same data. Moreover, our simulation results show that they are quite often severely biased. As a result, we cannot recommend this method for practical use.

Acknowledgements The authors are grateful to prof. D. R. Cox, University of Oxford, for proposing the problem and for his comments on early drafts of the paper.

References

1. Christensen, R.: Plane Answers to Complex Question (The Theory of Linear Models), 4th edn. Springer, New York (2011)
2. Gałecki, A., Burzykowski, T.: Linear Mixed-Effects Models Using R: A Step-by-Step Approach. Springer, New York (2013)
3. Henderson, C.R.: Estimation of variance and covariance components. *Biometrics* **9**(2), 226–252 (1953)
4. Lehmann, E.L., Scheffé, H.: Completeness, similar regions, and unbiased estimation—part I. *Sankhya* **10**, 305–340 (1950)
5. Pinheiro, J.C., Bates, D.M.: Mixed effects models in S and S-PLUS. Springer, New York (2000)

6. Rao, R.C., Kleffe, J.: Estimation of variance components. In: Krishnaiah, P.R. (ed.) *Handbook of Statistics I*, pp. 1–40. North Holland, Amsterdam (1980)
7. Searle, S.S., Casella, G., McCulloch, C.E.: *Variance Components*, 2nd edn. Wiley, New York (2006)
8. Slinger, W.D.: Least squares Lehmann–Scheffé estimation of variances and covariances with mixed linear models. *J Anim Sci* **74**, 2577–2585 (1996)
9. Slinger, W.D., Carlson, J.K.: A comparison via simulation of least squares Lehmann–Scheffé estimators of two variances and heritability with those of restricted maximum likelihood. *J Anim Sci* **81**, 1950–1958 (2003)

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