

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1993 - 5th Annual Conference Proceedings

THE ESTIMATION OF FIXED EFFECTS IN A MIXED LINEAR MODEL

F. Nabugoomu

O. B. Allen

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



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Nabugoomu, F. and Allen, O. B. (1993). "THE ESTIMATION OF FIXED EFFECTS IN A MIXED LINEAR MODEL," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1376>

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

THE ESTIMATION OF FIXED EFFECTS IN A MIXED LINEAR MODEL

F. Nabugoomu and O.B. Allen
University of Guelph

Abstract

The estimation of fixed effects is considered for small, unbalanced, mixed linear models. The two-stage estimator, in which the variance components are first estimated by ML or REML, is compared to the intra-block (IB) estimator, the ordinary least squares (OLS) estimator (ignoring the random effects) and the Gauss-Markov (GM) estimator. Comparison is made, based on 100 simulated data sets each, for 6 designs (3 BIBD's and 3 unbalanced designs). In comparing loss of information, relative to the GM lower bound, the two-stage procedures (using either ML or REML) are recommended for all but the smallest and least balanced design. The study also compared estimates of the variance of the two-stage estimators, using either the GM lower bound or the Kackar-Harville (KH) approximation. Estimators of the variance using REML estimates of the variance components are recommended, since estimators using ML estimates were seriously biased downward for all designs considered.

Key Words: mixed model, variance components, REML, ML, two-stage estimator, recovery of inter-block information, Kackar-Harville variance approximation.

1. Introduction

Consider the classical mixed linear model

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{u} + \mathbf{e} \quad (1)$$

where $E(\mathbf{y}) = \mathbf{X} \boldsymbol{\beta}$, $\text{Cov}(\mathbf{y}) = \mathbf{V} = \phi_0 \mathbf{V}_0 + \phi_1 \mathbf{V}_1$, \mathbf{X} and \mathbf{Z} are taken to be of full column rank, $\mathbf{V}_0 = \mathbf{I}$ and $\mathbf{V}_1 = \mathbf{Z} \mathbf{Z}'$. If the variance components, ϕ_0 and ϕ_1 , are known, the minimum variance linear unbiased (Gauss-Markov) estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{\text{GM}} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \quad (2)$$

and the variance-covariance matrix of this estimator is

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{GM}}) = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \quad (3)$$

The usual two-stage procedure, when the variance components are unknown, is to retain the same form for the

estimator of β , but to replace the variance components with estimates. In the context of incomplete block designs, this procedure is referred to as the recovery of inter-block information. Historically, the variance components have been estimated by equating selected quadratic forms to their expectations (Kempthorne, 1952). More recently, maximum likelihood (ML) and residual maximum likelihood (REML), usually assuming a multivariate normal likelihood, have been used to estimate the variance components (Harville, 1977).

It is easily shown that the conditional expectation (conditioning on the estimated variance components) of the two-stage estimator of β remains unbiased, regardless of what the variance component estimates are, but the estimator will no longer be minimum variance. The unconditional expectation is also unbiased, provided the estimators of the variance components are even functions of $\mathbf{y} - \mathbf{X}\beta$ (Kackar and Harville, 1981).

When the variance components are accurately estimated, the above two-stage estimator will be very close to the optimal estimator. However, it is not obvious that the two-stage estimator will perform well for small data sets where the variance components may be poorly estimated. Historically, it was recommended not to recover inter-block information if the design has fewer than 10 degrees of freedom for estimating the block mean square (Cochran and Cox, 1957).

For small incomplete block designs, the recommended procedure is to condition on the block totals (ie. treat the block effects as fixed), obtaining an estimator of β which utilizes only intra-block information. A third alternative is to ignore the blocks altogether and simply estimate β using the unadjusted treatment means. Both of these estimators are also unconditionally unbiased.

In this paper the behavior of the two-stage estimators (using ML or REML) is compared with the Gauss-Markov (GM) estimator (assuming the variance ratio is known), the intra-block (IB) estimator

$$\hat{\beta}_{IB} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Y} \quad (4)$$

where $\mathbf{W} = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{X}$, and the least squares estimator ignoring blocks (OLS)

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (5)$$

for a series of small designs (both balanced and unbalanced).

Since all estimators are unconditionally unbiased, the estimators will be compared based on their variances.

For the latter two estimators, the variances are

$$\text{Var}(\hat{\beta}_{\text{IB}}) = \phi_0 (\mathbf{W}'\mathbf{W})^{-1} \quad (6)$$

$$\text{Var}(\hat{\beta}_{\text{OLS}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (7)$$

For the two stage estimators, the variance cannot be written simply as a matrix expression. However, we know that it is bounded below by $\text{Var}(\hat{\beta}_{\text{GM}})$.

The variance of $\hat{\beta}_{\text{ML}}$ and $\hat{\beta}_{\text{REML}}$ is commonly taken to be (3), with the variance components replaced by their estimates. The second question addressed by this paper is how well does this estimator of the variance of the two stage estimators behave. Kackar and Harville (1984) showed that

$$\text{Var}(\mathbf{a}'\hat{\beta}_{\text{TS}}) = \mathbf{a}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{a} + E\{\mathbf{a}'(\hat{\beta}_{\text{TS}} - \hat{\beta}_{\text{GM}})(\hat{\beta}_{\text{TS}} - \hat{\beta}_{\text{GM}})'\mathbf{a}\} \quad (8)$$

They approximated the second term by $\text{trace}(\mathbf{A}(\phi) \mathbf{B}(\phi))$ where

$$\mathbf{A}(\phi) = \text{Var}\left(\frac{\partial}{\partial \phi} \mathbf{a}' \hat{\beta}_{\text{GM}}\right)$$

and $\mathbf{B}(\phi)$ is the mean square error matrix of $\hat{\phi} = (\hat{\phi}_0, \hat{\phi}_1)'$. Details of the evaluation of these expressions are given in Nabugoomu (1988).

The numerical assessment of the loss of information due to estimation of the ratio of the variance components, has been carried out for incomplete block designs (t is the number of treatments, k the block size and r the number of replicates) in some special cases. Kempthorne (1952) found, for a lattice (t=25, k=5, r=3), that the average percent increase in average variance of treatment differences, due to estimation of the ratio of the variance components (ie. the percent loss of information) was never larger than 2.7%. Yates (1940) reported that the percent loss of information for four lattice designs (t=25, k=5, r=2; t=16, k=4, r=3; t=27, k=3, r=3; 5x5 lattice square, r=3) ranged from 1.7% to 4.6%. Graybill and Deal (1959) provide an upper bound for the loss of information for balanced incomplete block designs for which the error degrees exceed 4; for the three cases tabulated (t=6, k=3, r=10; t=6, k=4, r=10; t=10, k=3, r=9) the upper bounds fall between 7.8 and 9.8%. Khatri and Shah (1975) presented methods for evaluating the exact variance of the two-stage estimator, when the variance components are estimated by the method of moments. They evaluated four BIB designs (the smallest being t=10, k=5, r=6) and 4 PBIB designs and in no case did the loss of information exceed 7.3%. Tong and Cornelius (1989) compared estimators of the slope of a simple linear regression, under a one-fold nested error structure. They compared two stage estimators using the method

of moments or ML to estimate the variance components, with the Gauss-Markov estimator, the ordinary least squares estimator and the intra-block estimator. All of the results above, with the exception of Tong and Cornelius, assume the variance components are estimated by equating certain quadratic forms to their expectations.

2. Simulation Details

In order to examine in more detail the loss of information due to the estimation of the ratio of variance components, especially for small and unbalanced designs, a simulation study was conducted. The simulated data were generated according to the following model

$$y_{ij} = \beta_i + u_j + e_{ij}$$

where the e_{ij} were independently sampled from a Normal population with mean 0 and variance $\phi_0 = 40$ and the u_j independently from a Normal population with mean 0 and variance ϕ_1 , independently of e_{ij} . For each of the designs considered, 100 data sets were generated. This was done for $\lambda = \phi_1/\phi_0 = 0.0, 0.25, 0.5, 1, 2, 4, 8, 12$, but the same random numbers were used for each of the ratios. This was done deliberately in order to improve the precision of the comparison of the effect of the ratio.

We report here on the results from 6 "small" designs, three BIB designs and three unbalanced designs. The following BIB's were considered:

t	k	r	b	ℓ	E	plan
4	2	3	6	1	.67	11.1
5	3	6	10	3	.83	11.1a
7	4	4	7	2	.88	11.8

where b is the number of blocks, ℓ is the number of times treatments are blockmates, E is the efficiency factor and plan refers to the plan number in Cochran and Cox (1957). The three unbalanced designs considered are given below:

Blocks	DESIGN 4				DESIGN 5				DESIGN 6					
	<u>TREATMENTS</u>				<u>TREATMENTS</u>				<u>TREATMENTS</u>					
	1	2	3	4	1	2	3	4	1	2	3	4	5	6
1	x		x		x				x	x	x	x	x	
2	x		x	x			x	x	x	x				
3			x	x	x						x	x	x	
4			x		x		x		x	x				
5	x	x	x				x		x	x	x	x	x	x
6	x	x	x	x			x				x	x		
7													x	x
8									x	x	x			
9											x	x	x	x
10									x	x				

REML and ML estimates were obtained for each data set, using BMDP program P3V. In those data sets where the likelihood (for either ML or REML) was maximized by a negative value of ϕ_1 , the estimate of ϕ_1 was taken to be 0.

3. Results

Table 1 below gives the loss in information of the estimated difference between treatments 1 and 2, based on the 100 simulated data sets, relative to the Gauss-Markov estimator (2). The loss in information is defined as the variance of the estimator minus the Gauss-Markov variance expressed as a percent of the Gauss Markov variance. The loss of information is given for the two-stage estimator using either ML or REML, the ordinary least squares estimator (5) and the intra-block estimator (6). The variances of the two-stage estimators were determined empirically from the 100 simulated data sets, but the variance of the ordinary least squares estimator and intra-block estimator are true values.

For $\phi_1 = 0$, the Gauss-Markov estimator is equivalent to the ordinary least squares estimator. When the ratio is large, the Gauss-Markov estimator is equivalent to the intra-block estimator. It is clear from Table 1 that the ordinary least squares estimator performs poorly when λ is large and that the intra-block estimator performs poorly when λ is near 0.

For all of the designs considered, except design 5, the two-stage procedures perform well, revealing only a modest loss of information. In fact, for designs 2, 3, 4 and 6 the estimated loss in information is negative (the empirical variance of the two-stage estimator is actually less than the Gauss-Markov lower bound) although the 80% confidence interval on the true loss in information for these procedures includes 0 in all cases. Because the same simulated random variables were used for all values of the ratio, the results tend to be very consistent as the ratio is varied. Design 5 is too small and unbalanced to effectively recover inter-block information. For this design the intra-block estimator should be used unless one is confident that the ratio is less than 1.0, in which case the ordinary least squares estimator is recommended.

The two-stage estimators generally exhibit quite similar loss in information. ML generally performs slightly better for large values of the ratio but REML performs slightly better for small values of the ratio. On this basis, either is recommended, except for design 5.

Table 1 also compares estimates of the variance of the two-stage estimators of the difference between treatments 1 and 2. The lower bound (3) is evaluated at the true value of ϕ and at the ML and REML estimates. The Kackar-Harville approximation (8) is similarly evaluated at ϕ , $\hat{\phi}_{ML}$ and $\hat{\phi}_{REML}$. Each entry is the average of the evaluations for the 100 data sets, expressed as a percent loss of information relative to the Gauss-Markov lower bound (3). These can be compared to the empirical variance of the two-stage estimators (columns 2 and 3), which are unbiased.

The estimated variance (either the lower bound or the Kackar-Harville approximation) tends to be smaller than the quantity it is estimating, using either ML or REML. ML, however, is dramatically biased downward whereas REML generally exhibits only a modest to negligible downward bias. For small unbalanced designs, the Kackar-Harville approximation evaluated at the true variance parameters appears to underestimate the true variance, at least for large values of the ratio. Thus, whether one wishes to estimate the Gauss-Markov lower bound or the Kackar-Harville approximation, REML is recommended over ML.

4. Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to OBA and by a Commonwealth Scholarship to FN.

5. References

- Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs*, second edition. John Wiley and Sons, New York.
- Graybill, F.A. and Deal, R.B. (1959). Combining unbiased estimators. *Biometrics* 15, 543-550.
- Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320-340.
- Kacker, R.N. and Harville, D.A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics - Theory and Methods*, Series a 10, 1249-1261.
- Kacker, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* 79, 853-862.
- Kempthorne, O. (1952). *The Design and Analysis of Experiments*. John Wiley and Sons, New York.
- Khatri, C.G. and Shah, K.R. (1975). Exact variance of combined inter and intra-block estimates in incomplete block designs. *Journal of the American Statistical Association* 70, 402-436.
- Nabugoomu, F. 1988. The analysis of fixed effects in mixed linear models. M.Sc. Thesis, University of Guelph.
- Tong, L. and Cornelius, P.L. 1989. Studies on the estimation of the slope parameter in the simple linear regression model with one-fold nested error structure. *Communications in Statistics -Simulation and Computation* 18, 201-225.

Table 1: Loss of information and estimation of variance for two-stage procedures for estimation of fixed effects

Column 1 gives the ratio of the variance components, ϕ_1/ϕ_0 . All the remaining columns give the difference between that particular variance and the GM lower bound, expressed as a percent of the GM lower bound. All variances refer to the estimated difference between treatments 1 and 2. Columns 2 and 3 represent the empirical variance (based on 100 data sets) of the two-stage estimator (with the variance components estimated by REML and ML, respectively). Below each is the 80% confidence interval for the true loss of information. Columns 4 and 5 give the true variance for the ordinary least squares and intra-block estimator, respectively. The last 5 columns deal with the average (based on 100 data sets) of the estimated variances of the two-stage estimators. Columns 6 and 7 give the estimates of the GM lower bound using, respectively, REML and ML. Columns 9 and 10 give the estimates of Kackar-Harville approximation, using REML and ML. Column 8 gives the KH approximation, evaluated at the true values of the variance components.

DESIGN 1

RATIO	<u>TWO STAGE</u>		OLS	IB	<u>LOWER BOUND</u>		<u>KACKAR-HARVILLE</u>		
	REML	ML			REML	ML	EXACT	REML	ML
12	7.0 (-10,30)	3.7 (-13,26)	512	2.0	-8.1	-52	2.1	-3.2	-51
8	9.3 (-8,33)	7.1 (-10,30)	335	2.9	-9.3	-54	3.1	-2.5	-53
4	14.4 (-4,39)	11.0 (-6,35)	158	5.6	-13	-52	5.7	-1.7	-49
2	12.1 (-5,36)	11.2 (-6,35)	71	10	-15	-53	9.8	-2.8	-49
1	9.3 (-8,33)	9.1 (-8,33)	30	17	-22	-53	15	-3.9	-48
.5	8.2 (-9,32)	9.5 (-8,33)	11	25	-25	-54	21	-6.1	-48
.25	7.6 (-9,31)	9.6 (-8,33)	3.7	33	-27	-54	26	-7.0	-49
0	9.0 (-8,33)	11.8 (-6,36)	0	50	-26	-53	33	-5.4	-47

Applied Statistics in Agriculture

DESIGN 2

RATIO	<u>TWO STAGE</u>		OLS	IB	<u>LOWER BOUND</u>		<u>KACKAR-HARVILLE</u>		
	REML	ML			REML	ML	EXACT	REML	ML
12	-3.0 (-18,18)	-3.1 (-18,18)	486	.5	7.3	-14	.2	7.6	-14
8	-2.8 (-18,18)	-3.0 (-18,18)	320	.8	7.1	-14	.3	7.6	-14
4	-2.4 (-18,19)	-2.5 (-18,18)	154	1.5	6.6	-14	.5	7.5	-14
2	-2.2 (-18,19)	-2.3 (-18,19)	71	2.9	5.6	-14	1.0	7.1	-13
1	-2.9 (-18,18)	-3.0 (-18,18)	31	5.0	4.0	-15	1.6	6.5	-14
.5	-3.0 (-18,18)	-3.0 (-18,18)	13	8.0	2.5	-16	2.5	5.8	-14
.25	-3.9 (-19,17)	-3.8 (-19,17)	4.5	11	1.0	-17	3.5	5.0	-14
0	-2.7 (-18,18)	-2.5 (-18,18)	0	20	-1.1	-18	5.6	3.8	-15

DESIGN 3

RATIO	<u>TWO STAGE</u>		OLS	IB	<u>LOWER BOUND</u>		<u>KACKAR-HARVILLE</u>		
	REML	ML			REML	ML	EXACT	REML	ML
12	-4.1 (-19,17)	-4.1 (-19,17)	514	0.3	1.8	-33	0.1	2.1	-33
8	-3.9 (-19,17)	-4.0 (-19,17)	339	0.4	1.6	-33	0.2	2.0	-33
4	-3.5 (-19,17)	-3.6 (-19,17)	165	0.8	1.3	-33	0.4	1.9	-33
2	-2.8 (-18,18)	-2.7 (-18,18)	78	1.6	0.9	-33	0.7	2.0	-32
1	-1.7 (-17,20)	-1.9 (-17,19)	35	2.9	0.2	-27	1.3	2.3	-26
.5	-1.9 (-17,19)	-1.9 (-17,19)	15	4.8	-0.7	-27	2.2	2.5	-26
.25	-2.9 (-18,18)	-2.8 (-18,18)	5.5	7.1	-1.9	-28	3.3	2.3	-26
0	-3.5 (-19,17)	-3.0 (-18,18)	0	14	-3.9	-28	6.7	1.6	-26

DESIGN 4

RATIO	<u>TWO STAGE</u>		OLS	IB	<u>LOWER BOUND</u>		<u>KACKAR-HARVILLE</u>		
	REML	ML			REML	ML	EXACT	REML	ML
12	-12 (-26,7)	-13 (-26,6)	449	1.4	-9	-28	.7	.3	-27
8	-10 (-24,9)	-11 (-25,8)	294	2.0	-1.4	-28	1.0	.4	-27
4	-7.1 (-22,13)	-8.1 (-23,12)	140	3.4	-2.4	-28	1.8	.4	-26
2	-3.6 (-19,17)	-.2 (-16,21)	64	5.9	-4.1	-34	3.0	.3	-32
1	1.3 (-15,23)	.4 (-15,22)	27	9.6	-5.6	-29	4.5	-.6	-27
.5	5.0 (-11,28)	4.7 (-12,27)	10	14	-7.4	-30	5.8	-1.6	-27
.25	5.6 (-11,28)	5.6 (-11,28)	3.6	19	-9.1	-31	6.8	-2.9	-28
0	1.2 (-15,23)	1.4 (-15,23)	0	29	-13	-34	8.0	-6.7	-30

DESIGN 5

RATIO	<u>TWO STAGE</u>		OLS	IB	<u>LOWER BOUND</u>		<u>KACKAR-HARVILLE</u>		
	REML	ML			REML	ML	EXACT	REML	ML
12	68 (42,104)	56 (32,90)	309	12	51	-35	14	85	-30
8	58 (33,92)	50 (27,83)	201	15	27	-42	19	57	-38
4	50 (27,83)	47 (24,79)	93	23	-2.0	-54	26	22	-51
2	49 (25,81)	50 (26,82)	41	34	-16	-60	30	5.6	-56
1	47 (24,79)	56 (32,90)	16	48	-20	-62	31	.4	-58
.5	51 (27,83)	63 (37,98)	5.8	62	-22	-61	29	-2.0	-58
.25	50 (27,83)	65 (39,101)	1.9	73	-23	-60	27	-3.1	-57
0	57 (32,91)	63 (38,99)	0	91	-24	-57	25	-4.8	-54

DESIGN 6

RATIO	<u>TWO STAGE</u>		OLS	IB	<u>LOWER BOUND</u>		<u>KACKAR-HARVILLE</u>		
	REML	ML			REML	ML	EXACT	REML	ML
12	-10 (-24,9)	-10 (-24,9)	155	.3	-7.1	-30	.1	-7.0	-30
8	-10 (-24,9)	-10 (-24,9)	101	.4	-7.2	-30	.1	-7.0	-30
4	-10 (-24,9)	-10 (-24,9)	48	.8	-7.3	-30	.3	-7.0	-30
2	-11 (-25,9)	-10 (-24,9)	22	1.4	-7.4	-29	.4	-6.9	-29
1	-11 (-25,8)	-11 (-25,8)	9.3	2.3	-7.1	-29	.7	-6.4	-28
.5	-11 (-25,8)	-11 (-25,8)	3.6	3.4	-7.4	-28	.9	-6.6	-28
.25	-11 (-25,8)	-11 (-25,8)	1.2	4.5	-8.2	-28	1.1	-7.2	-28
0	-12 (-26,6)	-12 (-26,7)	0	6.9	-12	-30	1.2	-11	-30