Robust Estimation in Linear Mixed Models – Revisited

Steven W. Enck

Equitable Life Assurance Society

New York, NY 10104

Charles E. McCulloch

Department of Statistical Science, Cornell University

Ithaca, NY 14853

BU-1448-M

Abstract

Fellner (1986)'s robust estimation method in linear mixed models is found to produce biased parameter estimates. A revised method is proposed in which the random effects and error terms are inflated by the REML factors prior to being bounded. In addition, the robust modifications to the mixed model equations and REML variance equations are made to be compatible, which improves the stability of the algorithm. Simulations confirm the necessity and effectiveness of the new method. Revised parameter estimates are provided for Fellner's metallic oxide data set.

KEY WORDS: Restricted maximum likelihood, Winsorize.

1. INTRODUCTION

Fellner (1986) presents a robust estimation method applicable to linear mixed models. The algorithm, in effect, solves for the parameters using modified observations. These "pseudo-observations" are created by bounding the *outlying* random components of the model and then inflating *all* random components slightly to adjust for the bias caused by bounding. We suggest two improvements to Fellner's method.

The mixed model equations (MMEs), which are solved to obtain the fixed and random effects and error terms, give rise to shrinkage estimators. In other words, the variation in the resulting random effects and error terms is smaller than the variances they are intended to estimate. Since these terms are bounded in the robust procedure, it is necessary to inflate the terms by the REML factors prior to bounding. In Fellner (1986)'s method, the terms are bounded first and then inflated to adjust for the smaller spread. In effect, only extreme outliers are bounded. Biased estimates result, as all terms are subsequently reinflated by a bias adjustment factor which is derived assuming all outliers are bounded.

Linear mixed models are usually estimated in an iterative procedure which alternates between revising the fixed and random effect estimates and the variance estimates until convergence is achieved. To assure a stable algorithm, the mixed model equations and REML variance equations should be consistent. The variation in the two vectors Fellner introduces to solve the MMEs is inconsistent with the variances being solved for in the REML variance equations. This can cause biased estimates and a potentially unstable algorithm.

The following revised algorithm corrects the above errors. Simulation results are presented which compare the two methods. In addition, revised estimates are provided for Fellner's metallic oxide data set.

2. LINEAR MIXED MODELS

Our notation for a linear mixed model is:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{u}_1 + \ldots + \mathbf{Z}_c\mathbf{u}_c + \mathbf{e}$$

where:

$\tilde{\mathbf{y}}$ is an n × 1 vector of observations

$\mathbf{X}$ is an n × p known design matrix for the fixed effects

$\mathbf{X}$ is an unknown p × 1 vector of fixed effects

$\mathbf{Z}_i$ is an n × q_i known design matrix for random effect i, i=1,...,c

$\tilde{\mathbf{0}}$ is a q_i × 1 unknown vector of random effects, i=1,...,c

$\mathbf{e}$ is an n × 1 unknown vector of random errors

Let q = $q_1 + q_2 + \ldots + q_n$. Define $\mathbf{Z}$ as an n x q partitioned matrix and $\mathbf{u}'$ as a q x 1 partitioned vector as follows:

$$\mathbf{Z} = [\mathbf{Z}_1 \cdots \mathbf{Z}_c] \quad \text{and} \quad \mathbf{u}' = [\mathbf{u}_1' \;\; \mathbf{u}_2' \cdots \;\; \mathbf{u}_c']$$

The general equation for a linear mixed model reduces to:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

The error structure of $\mathbf{y}$ satisfies, for i≠j and i=1,...,c, $\mathrm{E}(\mathbf{u}_i) = \mathbf{0}$, $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, and $\mathrm{Cov}(\mathbf{u}_i, \mathbf{e}) = \mathbf{0}$. In addition, $\mathrm{Var}(\mathbf{u}_i) = \mathbf{D}_i = \sigma_i^2 \cdot \mathbf{I}_{q_i}$ and $\mathrm{Var}(\mathbf{e}) = \mathbf{R} = \sigma_{c+1}^2 \cdot \mathbf{I}_n$. We simplify and write $\mathrm{Var}(\mathbf{u}) = \mathbf{D} = \mathrm{diag}[\mathbf{D}_1, \ldots, \mathbf{D}_c]$.

It is apparent from the above description that:

$\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{Var}(\mathbf{y}) = \mathbf{R} + \mathbf{ZDZ}'$.

3. ESTIMATION

Fellner (1986) presented the following iterative procedure to produce estimates of the parameters in a linear mixed model.

1. Obtain starting values for the variance components, and set $\tilde{\mathbf{y}} = \mathbf{y}$ and $\tilde{\mathbf{0}} = \mathbf{0}$. Note that in the robust method, $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{0}}$ will be modified after each iteration prior to step 7.

2. Using the current variance estimates, $\tilde{\mathbf{y}}$, and $\tilde{\mathbf{0}}$, solve the mixed model equations for $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$, and obtain $\hat{\mathbf{e}}$.

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{D^{-1} + Z'R^{-1}Z} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}}\tilde{\mathbf{y}} \\ \mathbf{Z'R^{-1}}\tilde{\mathbf{y}} + \mathbf{D^{-1}}\tilde{\mathbf{0}} \end{bmatrix}$$

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}}$$

3. For Restricted Maximum Likelihood, form **T** as the last q rows and columns of $(\mathbf{C'C})^{-1}$, partitioned similarly to $\mathbf{D}$:

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} \cdots \mathbf{T}_{1c} \\ \vdots \quad \vdots \\ \mathbf{T}_{c1} \cdots \mathbf{T}_{cc} \end{bmatrix} \qquad \text{where} \qquad \mathbf{C} = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{X} & \mathbf{R}^{-1/2}\mathbf{Z} \\ \mathbf{0} & \mathbf{D}^{-1/2} \end{bmatrix}$$

4. For each variance i in $\mathbf{D}$, define two quantities:

q_i = # of random effects with variance i

$v_i = \text{trace}(T_{ii})/\sigma_i^2$

5. The REML variance estimates satisfy:

$$\hat{\sigma}_i^2 = \frac{\|\hat{\mathbf{u}}_i\|^2}{(q_i - v_i)} \qquad \hat{\sigma}_{c+1}^2 = \frac{\|\hat{\mathbf{e}}\|^2}{n - p - \sum_i (q_i - v_i)}$$

6. Using the new variance estimates, form the revised variance-covariance matrices, $\mathbf{R}$ and $\mathbf{D}$.

7. Repeat the process beginning with step 2.

4. FELLNER'S ROBUST METHOD

The following definitions are needed. Note that r is the bound in standard errors and P_3 is the CDF of the χ_3^2 distribution.

$$\psi(x) = \begin{cases} -r & x < -r \\ x & |x| \le r \\ r & x > r \end{cases}$$

$$h = P_3(r^2) + 2r^2[F(-r)]$$

The function $\psi(x)$ is a bound applied to the random effects and error terms after they have been standardized. The term h is the bias adjustment term applied to the variances to adjust for the bias caused by bounding. Its derivation is described in Fellner (1986).

Fellner's robust analysis makes some slight modifications to the procedure defined in section 3.

1. The error terms and random effects solved for in the MMEs are bounded before being used in the REML equations. The resulting variance estimates are then divided by the bias adjustment factor h. Fellner writes the revised REML equations as:

$$\hat{\sigma}_i^2 = \hat{\sigma}_i^2 \cdot \frac{\left\| \psi(\hat{\sigma}_i^{-1} \cdot \hat{\mathbf{u}}_i) \right\|^2}{h \cdot (q_i - v_i)} \qquad \hat{\sigma}_{c+1}^2 = \frac{\hat{\sigma}_{c+1}^2 \cdot \left\| \psi(\hat{\sigma}_{c+1}^{-1} \cdot \hat{\mathbf{e}}) \right\|^2}{h \cdot \left\{ (n-p) - \sum_i (q_i - v_i) \right\}}$$

   where the variances on the right side of the equations are the estimates from the previous iteration.

2. Before step 7 in each iteration, two vectors are defined and used in solving the mixed model equations. These vectors, in effect, modify the MMEs so that they are solved using the bounded random effects and error terms. The vectors are defined as:

$$\tilde{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}} + \mathbf{R}^{1/2}\psi(\mathbf{R}^{-1/2}\hat{\mathbf{e}})$$

$$\tilde{\mathbf{0}} = \hat{\mathbf{u}} - \mathbf{D}^{1/2}\psi(\mathbf{D}^{-1/2}\hat{\mathbf{u}})$$

5. PROPOSED REVISIONS

5.1 REML Variance Equations

The mixed model equations are known as shrinkage estimators because they shrink the spread of the error terms and random effects. The REML variance equations adjust for this by dividing the square of the norm of the vectors by a factor smaller than the number of observations or random effects. When the robust procedure is used, it is important that the error terms and random effects be inflated by the appropriate REML factors prior to being bounded. In Fellner's method, the bound is applied prior to inflating the error terms and random effects by this factor. As a result, the bound affects only extreme outliers and quite possibly is not used at all. However, the variances are subsequently inflated by the bias adjustment term h, which is derived assuming that all outliers are bounded.

While the equations presented by Fellner are written very compactly, it is more intuitive in deriving the robust analysis to write the REML equations as follows:

$$\hat{\sigma}_i^2 = \frac{\left\| \hat{\mathbf{u}}_i \cdot \sqrt{\frac{q_i}{(q_i - v_i)}} \right\|^2}{q_i} \qquad \hat{\sigma}_{c+1}^2 = \frac{\left\| \hat{\mathbf{e}} \cdot \sqrt{\frac{n}{\left\{ (n-p) - \sum_i (q_i - v_i) \right\}}} \right\|^2}{n}$$

It is important to note that the variation in each of the vectors within the norm brackets is consistent with the variance solved for in the REML equation.

The robust analysis requires only one change. Define the following vectors which are used in place of $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{e}}$ in the above equations:

$$\hat{\mathbf{u}}_i' = \frac{1}{\sqrt{h}} \cdot \hat{\sigma}_i \cdot \sqrt{\frac{(q_i - v_i)}{q_i}} \cdot \psi\left(\frac{1}{\hat{\sigma}_i} \cdot \hat{\mathbf{u}}_i \cdot \sqrt{\frac{q_i}{(q_i - v_i)}} \right)$$

$$\hat{\mathbf{e}}' = \frac{1}{\sqrt{h}} \cdot \hat{\sigma}_{c+1} \cdot \sqrt{\frac{\left\{ (n-p) - \sum_i (q_i - v_i) \right\}}{n}} \cdot \psi\left(\frac{1}{\hat{\sigma}_{c+1}} \cdot \hat{\mathbf{e}} \cdot \sqrt{\frac{n}{\left\{ (n-p) - \sum_i (q_i - v_i) \right\}}} \right)$$

The vectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{e}}$ are first inflated by the REML factors so that their variation reflects the variances they are intended to estimate. They are then standardized by dividing by the appropriate standard errors and bounded. The resulting vectors are reinflated by both the appropriate standard error and REML factor. Finally, the terms are divided by the square root of the bias adjustment factor h to adjust for the bias caused by bounding.

The revised vectors, $\hat{\mathbf{u}}_i'$ and $\hat{\mathbf{e}}'$, have a number of properties. The outlying terms have been pulled into the bound, while all terms have been pushed out slightly to adjust for the bias caused by bounding. For large, normally distributed data sets, the bias which results from bounding is adjusted for by the bias adjustment factor h. In other words, the variance estimates should not change substantially between the nonrobust and robust methods. However, since the revised vectors are also used in the MMEs, the influence of outlying observations on the fixed and random effects is reduced in the robust method.

The above methodology can be written more succinctly as follows:

$$\hat{\sigma}_i^2 = \frac{\left\| \hat{\sigma}_i \cdot \psi\left(\hat{\sigma}_i^{-1} \cdot \hat{\mathbf{u}}_i \cdot \sqrt{\frac{q_i}{(q_i - v_i)}} \right) \right\|^2}{h \cdot q_i}$$

$$\hat{\sigma}_{c+1}^2 = \frac{\left\| \hat{\sigma}_{c+1} \cdot \psi\left(\hat{\sigma}_{c+1}^{-1} \cdot \hat{\mathbf{e}} \cdot \sqrt{\frac{n}{\left\{ (n-p) - \sum_i (q_i - v_i) \right\}}} \right) \right\|^2}{h \cdot n}$$

Note the differences between Fellner's equations and those written above. Fellner does not inflate the terms to reflect the REML factors prior to bounding.

5.2 Mixed Model Equations

To provide a stable algorithm, the MMEs must be compatible with the REML variance equations. However, if we examine Fellner's equations for the vectors $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{0}}$, it is apparent that they are incompatible. In the equations for $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{0}}$, the bounded error terms and random effects are not inflated by the bias adjustment factor, h. However, in the REML variance equations, the terms are inflated by the bias adjustment factor.

Using the notation from section 5.1, the equations for $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{0}}$ should be revised as follows:

$$\tilde{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}} + \hat{\mathbf{e}}'$$

$$\tilde{\mathbf{0}}_i = \hat{\mathbf{u}}_i - \hat{\mathbf{u}}_i' \qquad \tilde{\mathbf{0}} = \left[\tilde{\mathbf{0}}_1' \;\; \tilde{\mathbf{0}}_2' \cdots \;\; \tilde{\mathbf{0}}_c'\right]$$

It is apparent in the above equations that the revised MMEs and REML equations are consistent. At each iteration, modified error terms and random effects are computed and used in solving both the MMEs and REML equations.

5.3 A Note on the Likelihood Function being Maximized.

The REML equations were derived by maximizing the likelihood of a function of the data. However, there has been no mention as to which likelihood should be maximized in the robust analysis. Note that if the actual data is used in calculating the likelihood and we begin the robust procedure using the nonrobust REML estimates, the robust estimates will be equal to the nonrobust estimates.

Recall that the robust procedure, in effect, solves for the parameters using modified observations. Outlying error terms and random effects are pulled into a bound, while all error terms and random effects are pushed out slightly. It therefore seems logical to calculate the likelihood using the vector of pseudo-observations defined as follows:

$$\tilde{\tilde{\mathbf{y}}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}' + \hat{\mathbf{e}}'$$

The vector of pseudo-observations, $\tilde{\tilde{\mathbf{y}}}$, should be redefined prior to solving the REML equations in each iteration. Further, the new variance estimates at each iteration should increase the likelihood of these pseudo-observations. If the likelihood is not increased by the new estimates, the change in the variance estimates should be halved until the likelihood does increase.

6. SIMULATION RESULTS

To illustrate the necessity of this revision and the effectiveness of the procedure, simulations were run using generated data to see if the parameters used to generate the data could be retrieved by the methods. One hundred replications were run. For each replication, 200 normally distributed observations were generated. Two fixed effects were fit with 100 observations each. Fifty random effects were fit, each having 4 observations. The design matrix, $\mathbf{X}$, can be written according to Kronecker products as $\mathbf{I}_2 \otimes \mathbf{1}_{100}$. Similarly, the design matrix $\mathbf{Z}$ can be written as $\mathbf{I}_{50} \otimes \mathbf{1}_4$. The simulation results are presented in Table 1.

The revised method reproduces all values used to generate the data within 2 standard errors, while Fellner's method overestimates the error variance. Both fixed effect estimates and the error variance estimate are closer to their true values in the revised method, while only the random effect variance estimate is slightly further away.

7. FELLNER'S DATA SET

Fellner applied his method to data gathered from a sampling study designed to explore the effects of process and measurement variation on properties of lots of metallic oxide

(Bennett 1954). The revised procedure has been applied to the same data set and new estimates are provided in Table 2.

As expected, the revised variance estimates are smaller than Fellner's. The bound in the revised method is applied to all outliers, while Fellner's bound was applied only to the most extreme.

8. CONCLUSION

The revised method will provide estimates which are robust to the influence of outlying observations. The simulations provide evidence that the method produces more nearly unbiased parameter estimates for large, normally distributed data sets.

REFERENCES

Bennett, C.A. (1954), "Effect of Measurement Error on Chemical Process Control," *Industrial Quality Control*, 11, 17-20.

Fellner, William H. (1986), "Robust Estimation of Variance Components," *Technometrics*, 28, 51-60.

Jennrich, Robert I., and Schluchter, Mark D. (1986), "Unbalanced Repeated-Measures Models with Structured Covariance Matrices," *Biometrics*, 42, 805-820.

Searle, Shayle R., Casella, George, and McCulloch, Charles E. (1992), *Variance Components*, New York : John Wiley.

TABLES

Table 1 - Simulation Results

Parameter	Actual	Fitted Value: Non-Robust	Fellner	Revised
Fixed Effect #1	10.0000	10.0348	10.0376	10.0366
Fixed Effect #2	20.0000	20.0175	20.0164	20.0140
Random Effect Variance	2.0000	1.9702	2.1067	1.8718
Error Variance	5.0000	5.0035	5.1384	4.9480

Parameter	Standard Error (Estimate): Non-Robust	Fellner	Revised
Fixed Effect #1	0.0367	0.0365	0.0363
Fixed Effect #2	0.0375	0.0371	0.0367
Random Effect Variance	0.0703	0.0730	0.0768
Error Variance	0.0547	0.0594	0.0601

Parameter	Standard Units from Actual: Non-Robust	Fellner	Revised
Fixed Effect #1	0.9482	1.0301	1.0083
Fixed Effect #2	0.4667	0.4420	0.3815
Random Effect Variance	-0.4239	1.4616	-1.6693
Error Variance	0.0640	2.3300	-0.8652

Table 2 - Revised Metallic Oxide Estimates

		Robust	
Parameter	Non-Robust	Fellner	Revised
Type I Mean	3.856	3.862	3.873
Type II Mean	3.063	3.342	3.356
Lot Variance	0.607	0.176	0.136
Sample Variance	0.043	0.037	0.022
Chemist Variance	0.032	0.034	0.021
Analysis Variance	0.043	0.037	0.028