

Alternative Approaches to Mixed Linear Models

George W. Pasdirtz, Hazleton Laboratories

The inclusion of fixed effects and random effects (independent variables not under sampling control) in the linear model require special SAS* procedures (PROC VARCOMP or PROC MIXED) to estimate the variances of the random components. When these procedures are inconvenient or inappropriate for a particular application, alternative techniques are available. This paper presents a small-scale simulation study investigating the properties of two alternative approaches.

Introduction

The treatment of random components in ordinary least squares (OLS) is one of the "messy" problems (Milliken and Johnson, 1984) in data analysis. A first step in confronting mixed-model issues is an understanding of the sampling plan:

- If the values of a classification factor are under experimental control or include all possible levels of the population, the factor is a *fixed effect*.
- If the values of a classification factor result from a random process which is not under experimental control, the factor is a *random effect*.

The messiness is a result of added sources of variation introduced by the random components. The added variance can obscure the fixed effects unless it is removed from the error term of the model.

A standard example is an industrial experiment in which different classes of computing machines are being tested with different operators. The determination of whether machines and/or operators are random or fixed effects depends on the sampling plan and on the *inference space* to which the results are being projected (McLean, Sanders, and Stroup, 1991). If comparisons are restricted to the computing machines of only two manufacturers and if differences in manufacturing process are ignored, then the machines chosen are a fixed factor. If the results are to be generalized to all computing machines that have similar characteristics and the particular machines were only chosen for convenience, the variance introduced by the specific machines needs to be controlled in order to make the wider inference.

A similar analysis of the sampling plan for operators yields a more complicated picture. Assume that each operator can be classified by observable physical characteristics. If a random sample is drawn from the population of all operators with those characteristics (say male and female), operators are a fixed effect. If operators are not sampled but rather assigned to machines as they

arrive at the computer laboratory, the variance associated with each operator is an uncontrolled component of the study's overall variance. There may then be a general component of variance associated with these particular males and females and a specific component of variation associated with their use of a particular computer.

There are a number of difficult problems associated with the estimation of variance components:

- Hand calculation is required when using PROC GLM to compute method-of-moments variance estimates.
- The more elegant maximum likelihood (ML) techniques can produce estimates which are outside the parameter space (e.g., negative variances), are biased downwards in small samples, and can take excessive computing times.
- None of the techniques have good properties in unbalanced designs.

Milliken and Johnson (1984, Chapters 18-23) and Schwarz (1993) document the standard tactical difficulties of using mixed-model computer algorithms. There may be a number of additional strategic problems.

In an applied research setting it is never clear whether equal variances exist, even within fixed factors. For example, it may be that one of the computing machines is more sophisticated and thus produces greater operator variability. Levene (1960) developed a test statistic for homogeneity of variance that was generalized by Draper and Hunter (1969). The Levene statistic estimates the variance of each independent variable using the absolute values of the residuals from the initial model fit. Although the Levene test is not asymptotically distribution free (Miller, 1968), a Monte Carlo study conducted by Brown and Forsythe (1974) indicated that with at least ten observations per cell, asymptotic normality is not a problem. Furthermore, increasing percentages of trimming (from ten to fifty percent, which is the median) convert the Levene estimate into a very robust test.

After the results of a robust Levene test are available, how do we make use of the information? The mixed-model literature assumes that a variance component estimation should take place even without a prior test of homogeneity. However, Freedman and Peters (1984) argue that complicated estimators of variance may be of little value in small samples and may be worse than no estimate at all. For example, a number of textbook data sets used to illustrate variance component estimation contain under twenty observations with as few as two observations per cell.

This paper investigates two simple estimators that can be applied to mixed-model problems:

- The Levene estimate of variance followed by Weighted Least Squares (WLS).
- The simple average of OLS coefficients across replications.

Both estimators assume that the variance of each independent variable is open to question, possibly for different reasons. If no statistical differences in variances were detected using the Levene test, then an unweighted OLS estimate could be used regardless of the sampling plan.

A small-scale Monte Carlo study was conducted to investigate:

- Asymptotic bias (the convergence of parameter estimates as a function of increasing sample size).
- Misspecification (the effect of erroneously applying a WLS estimate to an equal-variance problem).

The SAS code necessary to apply each estimator in practice is included.

The Mixed Model

In the standard two-way linear model:

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \\ &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \text{NID}(0, \sigma^2) \end{aligned}$$

μ is the grand mean, α_i and β_j are fixed effects, γ_{ij} is the interaction term, and e_{ijk} is the error term such that the y_{ijk} are distributed independently and identically normal with zero mean and common variance, $\text{NID}(0, \sigma^2)$.

In the two-way mixed model:

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \\ &= \mu + \alpha_i + \text{NID}(0, \sigma_i^2) + \gamma_{ij} + \text{NID}(0, \sigma^2) \\ &= \mu + \alpha_i + \beta_j + \text{NID}(0, \sigma_{ij}^2) + \text{NID}(0, \sigma^2) \end{aligned}$$

μ is the grand mean, α_i is the fixed effect, β_j is the random effect, γ_{ij} is the interaction term, and e_{ijk} is the error term distributed $\text{NID}(0, \sigma^2)$. The distributions of the random components are also NID with variances depending on whether variance is introduced by the random factor, the random factor in combination with the fixed factor (e.g., different operator variances for each machine), or both. The mean of each random component

is conventionally assumed to be zero, but is not guaranteed to be zero by the OLS estimation itself.

The SAS code to generate data for a 2×3 design with k replications per cell is:

```
do rep = 1 to &k;
  do op = 1 to 3;
    do machine = 1 to 2;
      y = &mu + (machine-1) * &b * &sigma
          + op * &s * &sigma * rannor(0)
          + &sigma * rannor(0);
      output;
    end; end; end;
```

Notice that there are two random number streams and that, in this case, there is no interaction between operators and machines. Notice also that each effect is scaled by $\&b$ and $\&s$ relative to $\&sigma$. The scaling assumption will prove useful later in describing the WLS model. With minor modifications, this code was used in the Monte Carlo study presented below.

The full design matrix for one replication with main effects and interactions is:

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ F_1 \\ F_2 \\ R_1 \\ R_2 \\ FR_{11} \\ FR_{12} \\ FR_{21} \\ FR_{22} \\ e_{111} \\ e_{112} \\ e_{121} \\ e_{122} \\ e_{211} \\ e_{212} \\ e_{221} \\ e_{222} \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{122} \\ e_{211} \\ e_{212} \\ e_{221} \\ e_{222} \end{bmatrix}$$

The fixed machine effect (F) has two levels, the random operator effect (R) has three levels and the interaction (FR) has six levels.

Variance Estimation

In matrix terms, the linear model is:

$$y = Xb + e$$

where X is the $n \times m$ design matrix. The usual OLS assumptions are:

$$E(e) = 0, \quad V(e) = I\sigma^2, \quad \text{and} \quad e \sim \text{NID}(0, I\sigma^2)$$

To test the homogeneity-of-variance assumption, Levene's test reestimates the model:

$$|e| = |y - X\hat{b}| = Xc + u$$

using the absolute values of the residuals. The estimated coefficients, c , are a robust estimate of the *relative standard deviation* associated with each column of the design matrix. The advantage of using $|e|$ is that the distribution of e^2 can be markedly skewed (Carroll and Ruppert, 1988).

The estimated standard deviation of each observation is:

$$|\hat{e}| = Xc$$

where $|\hat{e}|$ is an $n \times 1$ vector of estimated standard deviations with the estimation error in u removed.

Weighted Least Squares

Weighted Least Squares (Draper and Smith, 1981: 108-116) drops the homogeneity of variance assumption for a full $n \times n$ variance-covariance matrix, V , where:

$$E(e) = 0, \quad V(e) = V\sigma^2, \quad \text{and} \quad e \sim \text{NID}(0, V\sigma^2)$$

If the covariances are 0, V becomes:

$$V\sigma^2 = \begin{bmatrix} w_1^{-1} & 0 & & \\ & w_2^{-1} & & \\ & & \ddots & \\ 0 & & & w_n^{-1} \end{bmatrix} \sigma^2$$

where the w 's are *weights* applied to the common variance, σ^2 . Thus, observations that have larger variances receive smaller weights.

The WLS estimator:

$$b = (X'V^{-1}X)X'V^{-1}y$$

is then computed by decomposing the weighting matrix:

$$V^{-1} = P^{-0.5}P^{-0.5}$$

and applying OLS to the reweighted X and y matrices.

$$P^{-0.5}y = P^{-0.5}Xb + P^{-0.5}e$$

As a result of reweighting, the WLS estimator has minimum variance. As long as the Levene weights are proportional to the true variances, the reweighted mixed-model estimates will also have minimum variance.

A disadvantage of the full WLS estimator is the $n \times n$ dimension of V and P . Laird and Ware (1982) have suggested an estimated weighted least squares (EWLS) procedure in which V is formed by iterative estimation of:

$$b = \left(\sum_{i=1}^k X_i' V^{-1} X_i \right)^{-1} \left(\sum_{i=1}^k X_i' V^{-1} y_i \right)$$

$$V = k^{-1} \sum_{i=1}^k (y_i - X_i \hat{b})(y_i - X_i \hat{b})'$$

where the dimension of V is now $k \times k$ and there are k replications of the basic design matrix. Lindstrom and Bates (1988) have compared EWLS to the standard ML estimators and found that convergence is slow. A natural extension of EWLS would be to apply the Levene weights to the estimate of V . Carroll and Ruppert (1988) suggest that as few as two iterations would then be needed for convergence.

Gumpertz and Pantula (1989) have studied a further simplification:

$$\bar{b} = k^{-1} \sum_{i=1}^k b_i$$

the mean of the individual regression coefficients across replications (BBAR). Each replication of the full design matrix and its associated regression curve is a random drawing from a larger population of response curves. By comparison, EWLS is a weighted mean of population estimates.

The coefficients in $\bar{b}$ are a random drawing from some k -variate population with:

$$V(b) = (k-1)^{-1} \sum_{i=1}^k (b_i - \bar{b})(b_i - \bar{b})'$$

$$\hat{\sigma}^2 = [k(n-m)]^{-1} \sum_{i=1}^k [y_i'y_i - b_i'X_i'y_i]$$

and are thus *random coefficients*.

In their Monte Carlo study, Gumpertz and Pantula found that BBAR had greater power in small samples than did EWLS. To cover a broad range of situations, however, they recommend 50 replicates. Carroll and Ruppert suggest, however, that effective WLS estimates can be obtained with far fewer observations.

SAS Coding

The WLS estimator can easily be coded in SAS using the WEIGHT command in PROC GLM.

```

proc reg data = a;      * OLS step; * by rep;
  model dvar = dmodel ;
  output out=b r=yr;
data c;
  set b;
  yr = abs(yr);
proc reg data = c;      * Levene Step;
  model yr = dmodel;
  output out=d p=vhat;
data e;
  set d;
  vhat = vhat**(-.5);
proc reg data = e outest = f; * WLS Step;
  * by rep;
  weight vhat;
  model dvar = dmodel;
  * proc means data = f;

```

Notice that negative or zero weights effectively eliminate the observation. To compute the BBAR estimator, comment out the weight statement, uncomment the by rep; statements, and uncomment the proc means statement. Similar coding can be developed for PROC GLM.

Monte Carlo Study

Simulation provides a way of comparing estimated parameters with known values to resolve mathematical and theoretical problems.

- The Levene estimate is not asymptotically distribution free (Miller, 1968) but might still produce relatively accurate estimates of weights in V.
- Since $(X'V^{-1}X)^{-1}$ is a concave function of V, WLS estimators may have a downward bias, especially when V is variable (Freedman and Peters, 1984).
- The WLS estimate is justified by asymptotic reasoning, meaning that the theoretical results hold as $n \rightarrow \infty$. Weighting with small sample size may be worse than no weighting at all.

A small-scale Monte Carlo study will generate information about the behavior of estimators in finite samples.

- Bias can be investigated by comparing observed parameter estimates with "true" values, $\text{bias}(\theta) = (\theta - \hat{\theta})$.
- Asymptotic convergence can be investigated by increasing sample size and observing the effect on bias.
- Misspecification can be investigated by applying WLS to data generated with homogeneous variances.

The design of the Monte Carlo study thus involved different settings for the macro variables &n, &b, &s, &mu, and &sigma in the SAS code above:

- Sample sizes [3 and 5 observations per cell].

- Scaling for random factor variances [0.25, 1.0] which were taken as a direct linear function of operator classification.
- Estimators [OLS, WLS, BBAR].
- True models [OLS, MIXED].

Fifty Monte Carlo simulations were conducted for each combination. A 2σ difference ($\&b = 2$) in machine performance was assumed in all simulations, with $\sigma = 1$ and $\mu = 50$. Three observations per cell provide adequate power ($p \geq 0.46$) at conventional levels ($p \leq 0.05$) to detect a 2σ difference, while $\&k = 5$ provides higher power ($p \geq 0.79$). Fifty Monte Carlo simulations provide adequate power ($p \geq 0.59$) to detect a 0.25σ difference between estimators. All power calculations were computed with JMP*.

The results are presented in Appendix 1. An analysis of covariance was conducted on the machine effect (d1) and the average of the non-zero interaction terms (FR). Controlling for &k and &s, there were no significant differences in estimators. There was a significant difference ($p \leq 0.01$) in root mean square error (RMSE) between OLS and WLS (not displayed), with WLS being closer to the true value. The BBAR estimator does not produce a RMSE in SAS for the full design matrix.

An inspection of the regression coefficients, however, indicates that there were numerical differences. The true value of the d1 effect in the simulation was -2. OLS came closest to the true value at -2.03, followed by BBAR at -1.88 and finally WLS at -2.43. Notice also from the interaction term (N*MODEL) that BBAR was least affected by sample size. Finally, the mixed model does indeed introduce bias (0.12), but the effect is not statistically significant given the Monte Carlo sample size used in the simulation.

In terms of the random components (FR) which were set at zero in the simulation, BBAR is the least biased (0.022), followed by WLS (0.053) and OLS (0.20761). Again, none of these biases were significant.

Discussion

This paper investigated a classical "messy" problem in statistical analysis: the mixed model. Mathematically, the presence of *variance components* introduces additional error that can obscure fixed effects. When comparing models estimated in OLS to WLS or EWLS, a reduction in RMSE and an increase in model sum of squares is often observed. Thus, a strong argument can be made both mathematically and practically that more complicated procedures must be applied to the mixed model.

Simple estimators such as WLS and BBAR did not introduce significant bias. Sophisticated ML techniques, on the other hand, can produce very biased estimates (Milliken and Johnson, 1984), especially in unbalanced models. Even though OLS does an acceptable job of estimating fixed effect parameters, the use of WLS or

BBAR is indicated in situations where large variance components relative to fixed effects are expected.

This study did not investigate the effect on p -values and statistical decision making. The reduction in RMSE observed for WLS suggests that it would be better than OLS. The BBAR estimates of $V(\bar{b})$ and σ^2 suggested by Gumpertz and Pantula (1989) would require further coding in PROC IML to produce p -values.

References

- Brown, M. B. and A. B. Forsythe (1974), "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, 69, 364-367.
- Carroll, R. J. and D. Ruppert (1988), *Transformation and Weighting in Regression*, New York: Chapman and Hall.
- Conover, W. J., M. E. Johnson, and M. M. Johnson (1981), "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data," *Technometrics*, 23, 351-361.
- Draper, N. R. and W. G. Hunter (1969), "Transformations: Some Examples Revisited," *Technometrics*, 11, 23-40.
- Draper, N. R. and H. Smith (1981), *Applied Regression Analysis*, Second Edition, New York: Wiley.
- Freedman, D. A. and S. C. Peters (1984), "Bootstrapping a Regression Equation: Some Empirical Results," *Journal of the American Statistical Association*, 79, 97-106.
- Gumpertz, M. and S. G. Pantula (1989), "A Simple Approach to Inference in Random Coefficient Models," *The American Statistician*, 43, 203-210.
- Laird, N. M. and J. H. Ware (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963-974.
- Levene, H. (1960), "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics*, (eds.) I. Olkin et. al., Ch. 25 (Stanford, CA: Stanford University Press) 278-292.
- Lindstrom, M. J. and D. M. Bates (1988), "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," *Journal of the American Statistical Association*, 83, 1014-1022.
- McLean, R. A., W. L. Sanders, and W. W. Stroup (1991), "A Unified Approach to Mixed Linear Models," *The American Statistician*, 45, 54-64.
- Miller, R. G. (1968), "Jackknifing Variances," *Annals of Mathematical Statistics*, 39, 567-582.
- Milliken, G. A. and D. E. Johnson (1984), *Analysis of Messy Data*, (New York: Van Nostrand Reinhold).
- Schwarz, C. J. (1993), "The Mixed-Model ANOVA: The Truth, the Computer Packages, the Books. Part I: Balanced Data," *The American Statistician*, 47, 48-59.
- *SAS® and JMP® are registered trademarks of the SAS Institute, Inc., Cary, NC, USA.

Author

George W. Pasdirtz, Ph.D.
Hazleton Laboratories
3301 Kinsman Boulevard
Madison, WI 53704

Appendix 1. Estimation Results.

Number of observations in data set = 2401

Dependent Variable: FR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	3.32861071	0.41607634	0.55	0.8219
Error	2392	1820.98611078	0.76128182		
Corrected Total	2400	1824.31472149			
R-Square C.V. Root MSE FR Mean					
	0.001825	-4214.967	0.8725147		-0.0207004

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	-.0058395879 B	-0.01	0.9947	0.87983568
S	-.0365280816	-0.77	0.4419	0.04749368
N	0.0029943217 B	0.13	0.8943	0.02252823
MODEL	OLS	0.2076187200 B	0.23	0.8172
	WLS	0.0531939033 B	0.06	0.9528
	BBAR	0.0215283188 B	0.02	0.9803
	TRUE	0.0000000000 B	.	0.87331409
N*MODEL	OLS	-.0423070234 B	-0.86	0.3890
	WLS	-.0062355856 B	-0.13	0.8990
	BBAR	0.0000000000 B	.	.
	TRUE	0.0000000000 B	.	.
TYPE	MIXED	-.0599499207 B	-1.68	0.03562026
	OLS	0.0000000000 B	.	.
	START	0.0000000000 B	.	.

Dependent Variable: D1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	26.96206074	3.37025759	0.69	0.6991
Error	2392	11652.75431146	4.87155281		
Corrected Total	2400	11679.71637220			
R-Square C.V. Root MSE D1 Mean					
	0.002308	-111.8785	2.2071594		-1.9728177

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	-1.736938645 B	-0.78	0.4352	2.22567911
S	0.173992998	1.45	0.1477	0.12014254
N	-0.061311921 B	-1.08	0.2821	0.05698861
MODEL	OLS	-0.294168402 B	-0.13	0.8970
	WLS	-0.687172558 B	-0.30	0.7624
	BBAR	-0.138551713 B	-0.06	0.9500
	TRUE	0.0000000000 B	.	2.20918175
N*MODEL	OLS	0.037574910 B	0.30	0.7623
	WLS	0.122466802 B	0.99	0.3242
	BBAR	0.0000000000 B	.	.
	TRUE	0.0000000000 B	.	.
TYPE	MIXED	0.120166233 B	1.33	0.09010691
	OLS	0.0000000000 B	.	.
	START	0.0000000000 B	.	.