WISHING AWAY RANDOM EFFECTS: WHAT DO WE LOSE?

BU-1481 -M **April, 2000**

Ranjani Natarajan
Dept. of Statistics
University of Florida
Gainesville, FL 32610

and

Charles E. McCulloch
Depts. of Biometrics and Statistical Sciences
Cornell University
Ithaca, NY 14853

Keywords: Best unbiased linear prediction, fixed effects, independence generalized estimating equation, maximum likelihood, mixed model, random effects.

Abstract

The choice between modeling a factor as fixed or random can at times be a difficult one. This choice is often affected by the lack of readily available software for fitting random effects models, especially those with non-normal responses. More important than software, however, is the influence of the widely-held belief that fixed effects models lead to more precise inference than do random effects models. This belief leads many to fit a fixed effects model when perhaps a random effects model would be more appropriate. In this paper we show the belief that fixed effects lead to more precise inference to be fallacious, and examine the consequences of choosing the fixed effects approach when the data truly originate from a random effects model. Models considered are the linear (normal theory) and binary mixed models. Results range from a loss of efficiency to complete breakdown of the point estimates (in the binary setting).

Wishing Away Random-Effects: What Do We Lose?

Ranjini Natarajan and Charles E. McCulloch

April 2000

Abstract

The choice between modeling a factor as fixed or random can at times be a difficult one. This choice is often affected by the lack of readily available software for fitting random effects models, especially those with non-normal responses. More important than software, however, is the influence of the widely-held belief that fixed effects models lead to more precise inference than do random effects models. This belief leads many to fit a fixed effects model when perhaps a random effects model would be more appropriate. In this paper we show the belief that fixed effects analyses lead to more precise inference to be fallacious, and examine the consequences of choosing the fixed effects approach when the data truly originate from a random effects model. Models considered are the linear (normal theory) and binary mixed models. Results range from a loss of efficiency to the complete breakdown of the point estimates (in the binary setting).

Key Words: Best unbiased linear prediction; Fixed effects; Independence generalized estimating equation; Maximum likelihood; Mixed model; Random effects;

Ranjini Natarajan is Assistant Professor, Department of Statistics and Division of Biostatistics, University of Florida, Gainesville, FL 32610. Charles E. McCulloch is Professor, Department of Statistical Science and Biometrics Unit, Cornell University, Ithaca NY 14853. This work was supported by NSF grants DMS-9625476 and DMS-9709091. We are grateful to Douglas Grove and Ramon Littell for several helpful conversations.

1 Introduction

Conventional wisdom holds that treating a factor as fixed rather than random reduces the scope of inference to the levels included in the experiment but increases the precision of estimates. This notion likely originated from the comparison of analysis of variance (ANOVA) results for normally distributed responses, such as the one for a balanced two-way layout with t treatments and b blocks depicted in Table 1. As this table shows, the sums of squares are identical under either model, however, under the random effects assumption it is the larger mean square due to interaction and not the mean square due to error which is the appropriate estimator of the variance of the treatment effect. Comparisons such as this have led to the widespread belief that random effects models lead to more conservative inferences than fixed effects models. In fact this thinking is so pervasive that it has resulted in a boycott of random effects models in some applications (see, Chinchilli 1988, pp. 362; Overall 1979, pp 63-86)

Regardless of any apparent gains in precision with a fixed effects model, in some situations it is difficult to justify its use. For example, consider the experiment of McLaren (1996) comparing the terminal height growth of understory Fir trees following an artificial browsing treatment. Data were collected on unmanipulated (control) and clipped trees from multiple sites in Isle Royale, MI. As tree growth can vary substantially across sites, a natural model to consider for these data is a two-way ANOVA model containing the main effects of "treatment" and "site" and their interaction "treatment $\times$ site". Since the specific sites in the data set are of interest, the conventional criteria would argue for treating this factor as fixed. This was the approach taken by McLaren. However, suppose the data within sites are correlated, as is often the case with clustered designs (e.g., data gathered from individuals in a longitudinal study or multiple clinics in a clinical trial). Then the fixed effects model would not be appropriate as it makes the assumption that the observations are independent. A better approach in this situation would be to treat "site" and "treatment $\times$ site" as random effects as an attempt to accommodate this correlation.

So the question is, by choosing the more appropriate model are we forsaking precision?

The answer to this apparent contradiction lies in recognizing that the conventional thinking is flawed in that a comparison is made between two different definitions of treatment effect. In the fixed effects model, the variance of the treatment effect is calculated conditionally on the blocks at hand. However, with the random effects model the variance is calculated unconditionally by averaging over all possible blocks. It is natural that the latter function should have larger standard errors since it has a much broader scope of inference (McLean, Sanders and Stroup 1991). Therefore this comparison is neither meaningful nor is it fair.

An interesting consequence of the balanced nature of the ANOVA design in Table 1 is that when comparisons are made using the same definition of the treatment effect under both models, then the standard errors coincide. However, this need not be true for other estimable functions. In fact, the results in this paper show that for a variety of other estimable functions, it is the fixed effects estimator which has the larger variance. This is an interesting realization as it advocates strongly for the use of random effects models in balanced settings where potential correlations are suspected. One loses nothing by modeling the factor as random as far as the treatment effect is concerned. However, substantial gains in efficiency may be obtained for other effects.

The remainder of the paper is organized as follows. We show in Section 2.1 that when the data arise from a linear mixed model, misspecification of the random effects by treating them as fixed preserves the unbiasedness of the resulting point estimates but leads to larger standard errors of these estimates. The only exception to this is the case described in Appendix 1 in which the estimates from the fixed and random effects models coincide. That the random effects standard errors should be smaller or at worst comparable to the fixed effects standard errors is not entirely surprising in this context since the linear mixed model technology for prediction is *best* prediction which by definition implies smallest variance among the class of linear unbiased estimators. However, to actually use these best predicted values, estimates must be inserted for the unknown variance components, potentially reducing their effectiveness. In Section 2.3 we examine the mean square error of these plug-in estimates through a simulation study on a two-factor ANOVA model. Our results indicate that despite the additional variability induced by the estimated variances, the mixed model predictions remain substantially more accurate. With the logit normal model for binary data, the consequences of misspecification can be more severe and our results in Section 3 reveal a complete breakdown of the point estimates themselves.

2 The Linear Mixed Model

We suppose the observed data is an $n \times 1$ vector $\mathbf{y}$ that arises from the linear mixed model:

$$\mathbf{y} \;=\; \mathbf{X}\beta + \mathbf{Z}\,\mathbf{u} + \epsilon, \tag{1}$$

where β is a $p \times 1$ vector of regression parameters, $\mathbf{X}\,(n \times p)$ is a design matrix of rank $r_{\mathrm{X}}\,(\leq p)$. $\mathbf{Z}\,(n \times q)$ is an incidence matrix corresponding to the unobserved vector of random-effects $\mathbf{u}$ and ϵ is a $n \times 1$ vector of random error. We let $\mathbf{H}\,(n \times (p+q))$ denote the horizontal concatenation of the matrices $\mathbf{X}$ and $\mathbf{Z}$, and $r_{\mathrm{H}} = \mathrm{rank}\,(\mathbf{H})$. The vectors $\mathbf{u}$ and ϵ are assumed to be statistically independent and to arise from the multivariate Normal (MVN) distributions MVN $(\mathbf{0}, \mathbf{D})$ and MVN $(\mathbf{0}, \sigma_e^2\,\mathbf{I})$ respectively. The inferential goal is prediction of a realization of the random variable $W = \boldsymbol{\ell}^{\mathrm{T}}\,(\mathbf{X}\beta + \mathbf{Z}\mathbf{u})$, where $\boldsymbol{\ell}$ is a known $n \times 1$ vector. The two-way ANOVA model is a special instance of the model in (1).

By way of terminology, a predictor $\widehat{W}$ is said to be unbiased if $E\left[\widehat{W} - W\right] = 0$, and its mean squared error (MSE) is the quantity $E\left[\left(\widehat{W} - W\right)^2\right]$, where the expectations are taken with respect to the joint distribution of W and $\mathbf{y}$. These expectations may be interpreted as the unconditional bias and MSE or as the average conditional bias and MSE by applying the usual definition of iterated expectations.

2.1 Ordinary Least Squares (OLS) Procedures

Suppose the investigator fits a fixed effects model corresponding to the conditional specification in (1). The best linear unbiased estimator (BLUE) of W under this model is the OLS estimator $\widehat{W}_{\mathrm{FE}} = \boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\widehat{\boldsymbol{\delta}}_{\mathrm{FE}}$, where $\widehat{\boldsymbol{\delta}}_{\mathrm{FE}} = \left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\mathbf{y}$ with the superscript minus denoting a generalized inverse. Then straightforward calculations reveal that (see Appendix 2 for proofs):

f1. $\widehat{W}_{\text{FE}}$ is an unbiased predictor of W, and

f2. The MSE of $\widehat{W}_{\text{FE}}$ is $\sigma_e^2 \boldsymbol{\ell}^{\text{T}}\mathbf{H}\left[\mathbf{H}^{\text{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\text{T}}\boldsymbol{\ell}$.

In some situations, the estimable function W of interest may only involve the fixed effects $\boldsymbol{\beta}$. In such cases, a parsimonious alternative to modeling the $\mathbf{u}$ as fixed unknown parameters may be to set them equal to their mean, zero. The BLUE of W under this model is the OLS estimator $\widehat{W}_{\text{IND}} = \boldsymbol{\ell}^{\text{T}}\mathbf{X}\widehat{\boldsymbol{\beta}}_{\text{IND}}$, where $\widehat{\boldsymbol{\beta}}_{\text{IND}} = \left[\mathbf{X}^{\text{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\text{T}}\mathbf{y}$. The predictor $\widehat{W}_{\text{IND}}$ can be recognized as the generalized estimating equation (GEE) solution for the marginal model $E\left[\mathbf{y}\right] = \mathbf{X}\boldsymbol{\beta}$ based on an independence working correlation matrix. Its properties are (see Appendix 2 for proofs):

g1. $\widehat{W}_{\text{IND}}$ is an unbiased predictor of W, and

g2. The MSE of $\widehat{W}_{\text{IND}}$ is $\boldsymbol{\ell}^{\text{T}}\mathbf{X}\left[\mathbf{X}^{\text{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\text{T}}\mathbf{V}\mathbf{X}\left[\mathbf{X}^{\text{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\text{T}}\boldsymbol{\ell}$, which is the population analogue of the robust variance estimate of $\widehat{W}_{\text{IND}}$ (Zeger, Liang and Albert, 1988).

2.2 Best Linear Unbiased Predictor (BLUP) and Empirical BLUP

If the variance components $\mathbf{D}$ and σ_e^2 were known, the mixed model strategy would be to estimate W by the BLUP $\widehat{W}_{\text{BLUP}} = \boldsymbol{\ell}^{\text{T}}\left(\mathbf{X}\widehat{\beta}_{\text{BLUP}} + \mathbf{Z}\widetilde{\mathbf{u}}_{\text{BLUP}}\right)$, where $\widehat{\beta}_{\text{BLUP}} = \left[\mathbf{X}^{\text{T}}\mathbf{V}^{-1}\mathbf{X}\right]^{-}\mathbf{X}^{\text{T}}\mathbf{V}^{-1}\mathbf{y}$ is the generalized least squares estimator, $\widetilde{\mathbf{u}}_{\text{BLUP}} = \mathbf{D}\mathbf{Z}^{\text{T}}\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\beta}_{\text{BLUP}}\right)$ and $\mathbf{V} = \text{Var}\left(\mathbf{y}\right) = \sigma_e^2\mathbf{I} + \mathbf{Z}\mathbf{D}\mathbf{Z}^{\text{T}}$. The predictor $\widehat{W}_{\text{BLUP}}$ is unbiased and its MSE is given by

$$\Psi\left(\mathbf{D}, \sigma_e^2\right) = \boldsymbol{\ell}^{\text{T}}\mathbf{H}\begin{bmatrix} \mathbf{X}^{\text{T}}\mathbf{X}/\sigma_e^2 & \mathbf{X}^{\text{T}}\mathbf{Z}/\sigma_e^2 \\ \mathbf{Z}^{\text{T}}\mathbf{X}/\sigma_e^2 & \mathbf{Z}^{\text{T}}\mathbf{Z}/\sigma_e^2 + \mathbf{D} \end{bmatrix}^{-}\mathbf{H}^{\text{T}}\boldsymbol{\ell},$$

which is best (smallest) by definition, within the class of linear unbiased predictors.

In practice, $\mathbf{D}$ and σ_e^2 are not known, hence an Empirical BLUP of W (denoted by $\widehat{W}_{\text{EBLUP}}$) is used, which replaces the variance components with their estimates. Kackar and Harville (1984) have shown that for even-valued, translation invariant estimators of $\mathbf{D}$ and

σ_e^2, $\widehat{W}_{\mathrm{EBLUP}}$ is unbiased. The MSE of $\widehat{W}_{\mathrm{EBLUP}}$ is naturally larger than that of $\widehat{W}_{\mathrm{BLUP}}$, but is unavailable in closed-form owing to the complicated dependence of $\widehat{W}_{\mathrm{EBLUP}}$ on $\mathbf{y}$.

In summary, the point estimates of W are unbiased regardless of whether the $\mathbf{u}$ are modeled as fixed, random or even ignored. However, their MSEs take on different expressions depending on the assumption about $\mathbf{u}$; our objective is to compare these MSEs. Towards this goal, we carried out several simulation studies with designs covering a range of sample sizes and varying degrees of imbalance in two-factor mixed models. Six variations of the understory Fir design were constructed by altering the cell sizes and the number of sites (Table 2). Three of these six designs were balanced (labeled B1, B2, B3), and three unbalanced (labeled U1, U2, U3).

2.3 Simulation of Two-way Mixed Model with Interaction

This section describes simulations run to compare the MSEs of $\widehat{W}_{\mathrm{FE}}$, $\widehat{W}_{\mathrm{IND}}$ and $\widehat{W}_{\mathrm{EBLUP}}$. Observations were generated from the two-way ANOVA model

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + u_j + \alpha u_{ij} + \epsilon_{ijk}, \\ \epsilon_{ijk} &\sim \mathrm{N}\left(0, \sigma_e^2\right), \end{aligned}$$

$(i = 1, 2;\ j = 1, \cdots, b;\ \text{k}=1, \cdots, n_{ij})$, where y_{ijk} is the k^{th} observation from the i^{th} treatment and j^{th} block, μ is the overall mean, α_i and u_j are the effects due to treatment and block (respectively) and αu_{ij} denotes their interaction. We modeled u_j and αu_{ij} as statistically independent and normally distributed random variables with mean 0 and variances σ_u^2 and $\sigma_{\alpha u}^2$ respectively. The true values of μ, α_i's and σ_e^2 were chosen (without loss of generality) to be $\mu = 0$, $\alpha_i = 0\ \forall i$ and $\sigma_e^2 = 1$. Further, four different values were specified for $\sigma_u^2 = \sigma_{\alpha u}^2 = \sigma^2$: 0, 0.5, 1, 2.

For each of the twenty four combinations of design and σ^2 values, we generated 2000 independent data sets. For each data set, the predictors $\widehat{W}_{\mathrm{FE}}$, $\widehat{W}_{\mathrm{IND}}$ and $\widehat{W}_{\mathrm{EBLUP}}$ were calculated for the six predictable functions $(W_1, W_2, \cdots, W_6)$ displayed in Table 3. These functions cover a variety of effects that are typically of interest in comparative experimental settings.

The MSEs of the OLS estimates and the EBLUPs of these functions are summarized in Table 4; the OLS estimates reported are exact while those for EBLUPs are Monte Carlo estimates based on the 2000 simulated data sets. Overall, our results show that the MSE of $\widehat{W}_{\text{EBLUP}}$ is smaller than, or at worst comparable to, those of $\widehat{W}_{\text{FE}}$ or $\widehat{W}_{\text{IND}}$. The following points should be noted with regard to the behavior of the various procedures.

- The functions W_2 and W_4 are not estimable under the fixed-effects model, and W_1, W_3, W_5 and W_6 are not estimable under the model where $\mathbf{u}$ is set equal to zero. All six functions are however estimable under the mixed model.

- The OLS estimates of W_2 and W_4 coincide with their mixed model counterparts in the balanced designs, as one would expect given the results of Zyskind (1967). The OLS and EBLUP estimates of W_1 and W_3 also coincide and, as we show in Appendix 1, this will always be the case for predictors whose vector $\boldsymbol{\ell}$ lies in the column space of $\mathbf{X}$.

- Predictable functions that involve specific random effects (e.g., W_5 and W_6) are extremely sensitive to their misspecification as fixed effects and one observes a loss in efficiency by using OLS procedures for estimating them even in balanced designs.

- The performance of $\widehat{W}_{\text{IND}}$ worsens as σ^2 increases. However, the reverse is true for $\widehat{W}_{\text{FE}}$. This is intuitive because with increasing heterogeneity there is less to be gained by "borrowing strength" across the population. Thus, the EBLUP does not perform very differently from the fixed effects approach in this situation. This finding corroborates the Bayesian rationale for considering factors with infinite variance as fixed-effects.

- In addition to a loss in precision for small values of σ^2, the penalty on the fixed effects estimates of W_5 and W_6 also depends on the number of levels of the random effect. In particular, they are more variable relative to the EBLUP as the number of levels grow, due to the increase in the number of parameters to be estimated.

- The performance of the OLS procedures deteriorates with imbalance. A comparison of the balanced design B1 with the unbalanced design U2 (or B2 with U3) reveals a greater loss in efficiency due to OLS in the unbalanced designs, although they have the same overall sample sizes as the corresponding balanced designs.

3 Logit Normal Model

In order to investigate the consequences of misspecifying a random factor with non-normal data, we now suppose the data are distributed according to a Bernoulli distribution with mean

$$E[y_i|\mathbf{u}] = exp\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\mathbf{u}\right) / \left(1 + exp\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\mathbf{u}\right)\right); \tag{2}$$

the random effects $\mathbf{u}$ are still assumed to arise from a MVN $(\mathbf{0}, \mathbf{D})$ distribution. This model is referred to as the logit normal model as it corresponds to specifying that the logit of the conditional mean response is a linear function of the linear predictor $\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\mathbf{u}$. The formulation in equation (2) is a natural extension of the linear mixed model in equation (1).

3.1 Maximum Likelihood Estimation

The classical estimation method for the logit normal model is maximum likelihood. The likelihood function is given by

$$L(\boldsymbol{\beta}, \mathbf{D}; \mathbf{y}) = \int \prod_{i=1}^{n} exp\left(y_i\left\{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\mathbf{u}\right\}\right) / \left(1 + exp\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\mathbf{u}\right)\right) f(\mathbf{u}|\mathbf{D})\, d\mathbf{u},$$

which involves a potentially high-dimensional integration over the random effects distribution. As there is no closed-form solution for this integral, the MLEs are usually obtained by either numerically maximizing an estimate of the likelihood function (when the dimension of $\mathbf{u}$ is small), or by using an EM algorithm (for more complicated problems). In the example described in Section 3.3 we opt for the former approach as the likelihood function can be reduced to a series of one-dimensional integrals. These integrals are estimated numerically using the Gauss-Hermite quadrature method.

3.2 Ordinary Logistic Regression (OLR) Procedures

The models resulting from misspecifying the **u**'s as fixed effects or ignoring them completely belong to the class of ordinary logistic regression models. Again, the classical estimation strategy in these models is maximum likelihood which involves maximizing

$$L(\boldsymbol{\beta}, \mathbf{u}) \quad = \quad \prod_{i=1}^{n} exp\left(y_i \left\{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\mathbf{u}\right\}\right) / \left(1 + exp\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_i^{\mathrm{T}}\mathbf{u}\right)\right)$$

with respect to $\boldsymbol{\beta}$ and $\mathbf{u}$ when the $\mathbf{u}$ are treated as fixed unknown parameters and by maximizing

$$L(\boldsymbol{\beta}) \quad = \quad \prod_{i=1}^{n} exp\left(y_i \left\{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right\}\right) / \left(1 + exp\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right)$$

with respect to $\boldsymbol{\beta}$ when the $\mathbf{u}$ are set equal to zero. These calculation are much simpler than the one required for the logit normal model and can be done using any standard statistical software package.

3.3 Simulation of a Randomized Complete Blocks Design

We conducted a simulation study to compare the performance of the MLEs from the logit normal model with those from the OLR models using responses generated from a Bernoulli distribution with conditional mean

$$E\left[y_{ijk}|u_j\right] \quad = \quad exp\left(\mu + \alpha_i + u_j\right) / \left(1 + exp\left(\mu + \alpha_i + u_j\right)\right) \tag{3}$$

$(i = 1, \ldots, 4,\ j = 1, \ldots, b,\ k = 1, \ldots, n_{ij})$. The random effects u_j were assumed to follow a $N\left(0, \sigma_u^2\right)$ distribution. We suppose that n_{ij} equals 0 or 1, and that the levels of the fixed and random factor are arranged according to either randomized complete block designs with $b = 6$ or 12 blocks, or balanced incomplete block designs with $(b = 8, \sum_i n_{ij} = 3)$ or $(b = 12, \sum_i n_{ij} = 2)$. The true values of the fixed effects were taken to be $\mu = 0$, $\alpha_1 = -.5$, $\alpha_2 = -.25$, $\alpha_3 = .25$, $\alpha_4 = .5$, and three values were chosen for σ_u^2: 0, 1.5 and 3.

A complication that arises in the binary context is that the MLEs under a given model may not exist (i.e., be finite) for particular data configurations. The results in Table 5 display the percentage of cases (based on 2000 data sets generated from each of the 12 combinations of b and σ_u^2) with finite MLEs under the OLR model that treats u_j as a fixed factor. The MLEs from the mixed model and the OLR model with $u_j = 0$ were finite for all 2000 data sets. Three patterns emerge from these results. We observe that existence of the MLE under the fixed effects model is seriously negatively impacted either as (i) the number of levels of the random factor b increase, (ii) the imbalance worsens or (iii) the degree of heterogeneity σ_u^2 becomes larger. While the first two observations are consistent with the behavior of the fixed effects estimates in the linear mixed model, the third is not. Recall that for normally distributed data, increased heterogeneity actually resulted in improved performance of the fixed effects estimates relative to their EBLUP counterparts. Since the number of data sets with finite parameter estimates under the fixed effects model is too small to provide a reliable picture of the performance of these estimates (ranging from 50% down to 0%), we do not make any further comparisons with this method.

Table 6 compares the bias and MSE of the ML estimates of the estimable function $W = \alpha_1 - \alpha_4$ under the logit normal model and the OLR model with $\mathbf{u} = \mathbf{0}$. Although we report both the bias and MSE, our focus is primarily on the bias since it would be an important consideration for larger sample sizes. The numbers reported are the average across the 2000 simulated data sets. The OLR estimate of W is no longer unbiased on account of the lack of equivalence of marginal and conditional parameters in non-normal models. The bias can be seen to worsen substantially as σ_u^2 increases. As Zeger, Liang and Albert (1988) have noted, the bias may be alleviated by multiplying the prediction of W by $\sqrt{1 + 0.58\widehat{\sigma}_u^2}$, where $\widehat{\sigma}_u^2$ is an estimate of σ_u^2. However, in order to do this, one needs to fit a mixed model to get an estimate of the variance component. The mixed model prediction of W has comparatively negligible bias for the configurations considered here.

4 Conclusion

The main finding of this paper is that the use of a random effects model, when justified, can produce large gains in efficiency compared to the use of the corresponding fixed effects model, or a model that completely ignores the effect. This finding countermands the widely-held belief that treating a factor as random leads to more conservative inferences than does treating it as fixed. These results, as indicated, apply to the situation where the effect of the factor is indeed random; however, as argued by Grizzle (1987), such situations are the norm in practice rather than the exception. An important issue not discussed in this paper is the accuracy of estimation of the MSEs, particularly in the mixed model context. This is an important practical consideration and several simulation studies we conducted (not reported here) suggest that existing methods for the estimation of standard errors of prediction are not very accurate in small samples and when the heterogeneity is small. Finding adequate estimators for these situations is an area where further research is needed.

Appendix 1

First consider the fixed effects model. Suppose that $\boldsymbol{\ell}$ is in the column-space of $\mathbf{H} = (\mathbf{X}, \mathbf{Z})$, i.e., $\mathbf{H}\boldsymbol{\lambda} = \boldsymbol{\ell}$ for some vector $\boldsymbol{\lambda}$. Then

$$
\begin{aligned}
\widehat{W}_{\text{FE}} &= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\mathbf{y} \\
&= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\mathbf{y} \\
&= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{y}, \\
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{y}.
\end{aligned}
$$

In words, for linear combinations in the column space of $\mathbf{H}$, the linear combination of the predicted values is exactly equal to the same linear combination of the data.

For the mixed model there is a similar result. Suppose now that $\boldsymbol{\ell}$ is in the column space of $\mathbf{X}$. Then we have

$$
\begin{aligned}
\widehat{W}_{\text{BLUP}} &= \boldsymbol{\ell}^{\mathrm{T}}\left(\mathbf{X}\widehat{\boldsymbol{\beta}}_{\text{BLUP}} + \mathbf{Z}\widetilde{\mathbf{u}}_{\text{BLUP}}\right), \\
&= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\left(\mathbf{X}\left[\mathbf{X}\mathbf{V}^{-1}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y} + \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{y}\right),
\end{aligned}
$$

where $\mathrm{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}$ or $\mathbf{I} - \mathbf{V}\mathbf{P} = \mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}$. This gives

$$
\begin{aligned}
\widehat{W}_{\text{BLUP}} &= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{V}\mathbf{P} + \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}}\mathbf{P}\right)\mathbf{y}, \\
&= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\left(\mathbf{I} - \sigma_e^2\mathbf{P}\right)\mathbf{y}, && \text{since } \mathbf{V} = \sigma_e^2\mathbf{I} + \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}} \\
&= \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y}, && \text{since } \mathbf{X}^{\mathrm{T}}\mathbf{P} = \mathbf{0}, \\
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{y}.
\end{aligned}
$$

Thus, $\widehat{W}_{\text{BLUP}}$ and $\widehat{W}_{\text{FE}}$ coincide.

Appendix 2

f1. $E\left[\widehat{W}_{\text{FE}} - W\right] = 0.$

$$
E\left[\widehat{W}_{\text{FE}}\right] = \boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\,\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}E\left[\mathbf{y}\right],
$$

$$
\begin{aligned}
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\mathbf{H}\left(\beta^{\mathrm{T}},0\right)^{\mathrm{T}},\\
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left(\beta^{\mathrm{T}},0\right)^{\mathrm{T}}, \qquad \text{by definition of } \left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\\
&= E\left[W\right].
\end{aligned}
$$

f2. $E\left[\left(\widehat{W}_{\mathrm{FE}}-W\right)^{2}\right]=\sigma_e^2\boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\boldsymbol{\ell}.$

$$
\begin{aligned}
E\left[\left(\widehat{W}_{\mathrm{FE}}-W\right)^{2}\right] &= E\left[\left(\boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\mathbf{y}-\mathbf{h}^{\mathrm{T}}\mathbf{H}\left(\beta^{\mathrm{T}},\mathbf{u}^{\mathrm{T}}\right)^{\mathrm{T}}\right)^{2}\right],\\
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\left(\mathbf{V}+\mathbf{X}\beta\beta^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\right)\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\boldsymbol{\ell}\\
&\quad -\boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\mathbf{H}\begin{pmatrix}\beta\beta^{\mathrm{T}} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{pmatrix}\mathbf{H}^{\mathrm{T}}\boldsymbol{\ell},\\
&= \sigma_e^2\,\boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\boldsymbol{\ell}+\boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\mathbf{H}\begin{pmatrix}\beta\beta^{\mathrm{T}} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{pmatrix}\mathbf{H}^{\mathrm{T}}\boldsymbol{\ell}\\
&\quad -\boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\mathbf{H}\begin{pmatrix}\beta\beta^{\mathrm{T}} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{pmatrix}\mathbf{H}^{\mathrm{T}}\boldsymbol{\ell},\\
&= \sigma_e^2\boldsymbol{\ell}^{\mathrm{T}}\mathbf{H}\left[\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]^{-}\mathbf{H}^{\mathrm{T}}\boldsymbol{\ell}.
\end{aligned}
$$

g1. $E\left[\widehat{W}_{\mathrm{IND}}-W\right]=0.$

$$
\begin{aligned}
E\left[\widehat{W}_{\mathrm{IND}}\right] &= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{X}\mathbf{H}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}E\left[\mathbf{y}\right],\\
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\mathbf{X}\beta,\\
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{X}\beta, \qquad \text{by definition of } \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\\
&= E\left[W\right].
\end{aligned}
$$

g2. $E\left[\left(\widehat{W}_{\mathrm{IND}}-W\right)^{2}\right]=\boldsymbol{\ell}^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\boldsymbol{\ell}.$

$$
\begin{aligned}
E\left[\left(\widehat{W}_{\mathrm{IND}}-W\right)^{2}\right] &= E\left[\left(\boldsymbol{\ell}^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\mathbf{y}-\boldsymbol{\ell}^{\mathrm{T}}\mathbf{X}\beta\right)^{2}\right],\\
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\left(\mathbf{V}+\mathbf{X}\beta\beta^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\right)\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\mathbf{h}-\mathbf{h}^{\mathrm{T}}\mathbf{X}\beta\beta^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{L},\\
&= \boldsymbol{\ell}^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-}\mathbf{X}^{\mathrm{T}}\boldsymbol{\ell}.
\end{aligned}
$$

References

Chinchilli, V. M. (1988), "Clinical Efficacy Trials with Quantitative Data," in *Biopharmaceutical Statistics for Drug Development*, ed. Peace, K. E., New York: Marcel Dekker.

Grizzle, J. E., (1987), Letter to the Editor, *Controlled Clinical Trials*, 8, 392-393.

Kackar, 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.

McLaren, B. E. (1996), "Plant-Specific Response to Herbivory: Simulated Browsing of Suppressed Balsam Fir on Isle Royale", *Ecology*, 77, 228-235.

McLean, R. A., Sanders. W. L., and Stroup, W. W. (1991), "A Unified Approach to Mixed Linear Models", *American Statistician*, 45, 54-64.

Overall, JE (1979), "General Linear Model Analysis of Variance," in *Coordinating Clinical Trials in Psychopharmacology: Planning, Documentation and Analysis*, eds. Levine, J, Washington D.C.: U.S. Government Printing Office.

Zeger, S. L., Liang, K-Y. and Albert P. S. (1988), "Models for Longitudinal Data: A Generalized Estimating Equation Approach," *Biometrics*, 44, 1049-1060.

Zyskind, G. (1967), "On Canonical Forms, Non-negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models, *Annals of Mathematical Statistics*, 38, 1092-1109.

Table 1: Analysis of variance in the two-factor model $y_{ijk} = \mu + \alpha_i + u_j + \alpha u_{ij} + \epsilon_{ijk}$; $\epsilon_{ijk} \sim N\left(0, \sigma_e^2\right)$, $i = 1, \cdots, t$, $j = 1, \cdots, b$, $k = 1, \cdots, n$. In the fixed model u_j and αu_{ij} are modeled as unknown parameters. In the random model they are modeled as statistically independent random variables with $u_j \sim \mathrm{N}\left(0, \sigma_u^2\right)$ and $\alpha u_{ij} \sim \mathrm{N}\left(0, \sigma_{\alpha u}^2\right)$.

Source	Sums of Squares	Expected Mean Square	
		fixed	random
Treatment	n b $\sum_i \left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2$	$\sigma_e^2 + nb\frac{\sum_i \alpha_i^2}{t-1}$	$\sigma_e^2 + nb\frac{\sum_i \alpha_i^2}{t-1} + n\sigma_{\alpha u}^2$
Blocks	n t $\sum_j \left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2$	$\sigma_e^2 + na\frac{\sum_j u_j^2}{b-1}$	$\sigma_e^2 + na\sigma_u^2$
Interaction	$n \sum_i \sum_j \left(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\right)^2$	$\sigma_e^2 + n\frac{\sum_i \sum_j (\alpha u)_{ij}^2}{(t-1)(b-1)}$	$\sigma_e^2 + n\sigma_{\alpha u}^2$
Error	$\sum_i \sum_j \sum_k \left(Y_{ijk} - \bar{Y}_{ij.}\right)^2$	σ_e^2	σ_e^2

Table 2: Cell sizes n_{ij} and number of sites b for two-way mixed models

B1

	Site		
Treatment	1	2	3
1	2	2	2
2	2	2	2

U1

	Site		
Treatment	1	2	3
1	1	1	4
2	4	4	4

B2

	Site		
Treatment	1	2	3
1	4	4	4
2	4	4	4

U2

	Site		
Treatment	1	2	3
1	1	1	4
2	1	4	1

B3

	Site					
Treatment	1	2	3	4	5	6
1	4	4	4	4	4	4
2	4	4	4	4	4	4

U3

	Site					
Treatment	1	2	3	4	5	6
1	1	1	4	1	1	4
2	1	4	1	1	4	1

Table 3: Predictable functions for two-way mixed models

Function	Expression	Interpretation[1]
W_1	$\mu + \alpha_1 + \frac{1}{b}\sum_{j=1}^{b} u_j + \frac{1}{b}\sum_{j=1}^{b} \alpha u_{1j}$	Conditional mean response for control trees
W_2	$\mu + \alpha_1$	Marginal mean response for control trees
W_3	$\alpha_1 - \alpha_2 + \frac{1}{b}\sum_{j=1}^{b} \alpha u_{1j} - \frac{1}{b}\sum_{j=1}^{b} \alpha u_{2j}$	Difference in conditional mean response for control versus clipped trees
W_4	$\alpha_1 - \alpha_2$	Difference in marginal mean response for control versus clipped trees
W_5	$\mu + \alpha_1 + u_1 + \alpha u_{11}$	Conditional mean response for control trees in site 1
W_6	$u_1 - u_2 + \alpha u_{11} - \alpha u_{12}$	Difference in conditional mean response for site 1 versus 2 for control trees

[1]using terminology of the understory Fir data.

Table 4: A comparison of MSEs for two-way mixed models. The values for the OLS methods are exact; those for EBLUP are accurate to within ± .05. The dashes indicate lack of estimability under the particular model

	$\sigma^2 = 0$			$\sigma^2 = 0.5$			$\sigma^2 = 1$			$\sigma^2 = 2$		
	IND	FE	EBLUP	IND	FE	EBLUP	IND	FE	EBLUP	IND	FE	EBLUP
						Design B1						
W_1	–	.17	.17	–	.17	.17	–	.17	.17	–	.17	.17
W_2	.17	–	.17	.50	–	.50	.83	–	.83	1.50	–	1.50
W_3	–	.33	.33	–	.33	.33	–	.33	.33	–	.33	.33
W_4	.33	–	.33	.67	–	.67	1.00	–	1.00	1.67	–	1.67
W_5	–	.50	.25	–	.50	.41	–	.50	.46	–	.50	.51
W_6	–	1.00	.26	–	1.00	.79	–	1.00	.91	–	1.00	.99
						Design B2						
W_1	–	.08	.08	–	.08	.08	–	.08	.08	–	.08	.08
W_2	.08	–	.08	.42	–	.43	.75	–	.75	1.42	–	1.42
W_3	–	.17	.17	–	.17	.17	–	.17	.17	–	.17	.17
W_4	.17	–	.17	.50	–	.50	.83	–	.83	1.50	–	1.50
W_5	–	.25	.12	–	.25	.24	–	.25	.25	–	.25	.25
W_6	–	.50	.10	–	.50	.47	–	.50	.49	–	.50	.51
						Design B3						
W_1		.04	.04	–	.04	.04	–	.04	.04	–	.04	.04
W_2	.04	–	.04	.21	–	.21	.37	–	.37	.71	–	.71
W_3	–	.08	.08	–	.08	.08	–	.08	.08	–	.08	.08
W_4	.08	–	.08	.25	–	.25	.42	–	.42	.75	–	.75
W_5	–	.25	.06	–	.25	.22	–	.25	.24	–	.25	.24
W_6	–	.50	.05	–	.50	.43	–	.50	.47	–	.50	.49
						Design U1						
W_1	–	.25	.18	–	.25	.23	–	.25	.24	–	.25	.25
W_2	.17	–	.17	.67	–	.55	1.17	–	.93	2.17	–	1.61
W_3	–	.33	.27	–	.33	.31	–	.33	.32	–	.33	.34
W_4	.25	–	.26	.75	–	.64	1.25	–	.98	2.25	–	1.65
W_5	–	1.00	.25	–	1.00	.75	–	1.00	.85	–	1.00	.96
W_6	–	2.00	.17	–	2.00	1.25	–	2.00	1.55	–	2.00	1.84
						Design U2						
W_1	–	.25	.19	–	.25	.23	–	.25	.24	–	.25	.24
W_2	.17	–	.19	.67	–	.56	1.17	–	.89	2.17	–	1.61
W_3	–	.50	.39	–	.50	.50	–	.50	.48	–	.50	.49
W_4	.33	–	.38	1.08	–	.83	1.83	–	1.14	3.33	–	1.78
W_5	–	1.00	.35	–	1.00	.74	–	1.00	.88	–	1.00	.97
W_6	–	2.00	.31	–	2.00	1.29	–	2.00	1.68	–	2.00	1.94
						Design U3						
W_1	–	.12	.09	–	.12	.11	–	.12	.12	–	.12	.12
W_2	.08	–	.09	.33	–	.27	.58	–	.45	1.08	–	.79
W_3	–	.25	.18	–	.25	.24	–	.25	.25	–	.25	.24
W_4	.17	–	.18	.54	–	.38	.92	–	.56	1.67	–	.91
W_5	–	1.00	.17	–	1.00	.62	–	1.00	.79	–	1.00	.93
W_6	–	2.00	.14	–	2.00	1.19	–	2.00	1.51	–	2.00	1.77

Table 5: Percentage of 2000 data sets generated from model (2) for which the fixed effects MLE exists. The mixed model estimates and the OLR estimates from the model with $\mathbf{u} = \mathbf{0}$ exist in all 2000 cases

σ_u^2	b	$\sum_i n_{ij}$	%
0	6	4	49
	12	4	23
	8	3	11
	12	2	0
1.5	6	4	14
	12	4	1
	8	3	2
	12	2	0
3	6	4	6
	12	4	0
	8	3	0
	12	2	0

Table 6: A comparison of the bias and MSE of estimators of $W = \alpha_1 - \alpha_4$ in logit-normal models. The sampling standard error of the OLR and mixed model MSEs is less than .05 and .08 respectively.

			OLR		Mixed model	
σ_u^2	b	$\sum_i n_{ij}$	bias	MSE	bias	MSE
0	6	4	.10	1.37	.01	1.71
	12	4	-.09	.85	-.17	1.04
	8	3	.08	1.36	-.04	1.91
	12	2	.11	1.34	-.01	1.94
1.5	6	4	.26	1.24	.04	2.14
	12	4	.15	0.71	-.13	1.33
	8	3	.31	1.31	.10	2.24
	12	2	.29	1.32	.11	2.14
3	6	4	.34	1.11	.07	2.10
	12	4	.26	.64	-.11	1.38
	8	3	.37	1.23	.12	2.05
	12	2	.36	1.44	.18	2.14