Summary

Standard prediction of random effects under the mixed linear model takes an empirical Bayesian approach to produce estimates of their posterior mean given the data. While grounded in theory aimed at optimal mean square error, this approach naturally shrinks predictions toward a population mean and can give a misleading impression of the actual distribution of the random effects. In particular, the narrow spread characterizing the predictor distribution can have undesirable effects when the objective is to classify subjects relative to a threshold. The constrained Bayes methodology (Louis 1984 *JASA* 79, 393-398; Ghosh 1992 *JASA* 87, 533-540) provides a competitor for the common posterior mean method. In this paper, we examine this approach for predicting random effects under mixed linear models, with particular attention to the handling of covariates. Our study finds that the general methodology of Ghosh (1992) is flexible and compares well with a direct implementation of constraints, suggesting that this alternative prediction approach might be readily incorporated into common software for mixed linear models. We provide an example geared toward predicting CD4 cell counts among HIV-infected children at the time that they reach diagnosis of Class A disease status.
1. Introduction

The standard mixed linear model (e.g., Laird and Ware, 1982) remains an extremely popular practical tool for analyzing longitudinal, repeated measures, or otherwise correlated continuous data. In such analyses, the prediction of linear combinations of fixed and random effects can sometimes be of great interest. The typical approach implemented in commercial software is to obtain empirical best linear unbiased predictors (EBLUPs), which estimate the posterior mean of the linear combination given the response data (Littell et al., 2006). The general acceptance of these empirical Bayes-like predictions stems from their intuitive appeal and their theoretical underpinnings as minimal prediction mean square error estimates (Searle et al., 1992). They are also referred to as shrinkage estimators, given their familiar characteristic of pulling subject-specific predictions toward an appropriate population mean.

Due to the shrinkage phenomenon, EBLUPs stemming from linear or nonlinear mixed models exhibit distributions that can be much narrower than those assumed to characterize the random variables being predicted. Several authors (e.g., Efron and Morris, 1971; Louis, 1984; Ghosh, 1992) pointed out potential drawbacks to this feature and proposed alternative approaches that can be applied to reduce shrinkage and/or more closely match the predictor and underlying true distributions.

One effect of overshrinkage in certain applications is that it can lead to a lack of sensitivity for identifying “extreme” experimental units relative to a fixed threshold (i.e., the probability that an EBLUP lies beyond a threshold given that the true random variable does can be quite small). In order to improve sensitivity in such a context, Lyles and Xu (1999) proposed constrained Bayes predictors of random intercepts and slopes aimed to minimize prediction mean squared error given that the mean and variance of the predicted random variables matches that of the true ones. These criteria fully agree in spirit with those targeted by the general methodology
of Ghosh (1992), although to our knowledge the implementation of such constrained Bayes estimation has not been advocated for wide use in the mixed linear model context.

In this paper, we extend the models considered by Lyles and Xu (1999) to incorporate fixed and/or time-dependent covariates, and we compare their direct constrained Bayes strategy with that advocated by Ghosh (1992). Our ultimate goal is to exhibit the flexibility, performance, and convenient implementation of the Ghosh paradigm, and to illustrate its potential worth as an alternative prediction method that might be generally applicable in commercial mixed linear model software. We use simulations to evaluate and compare the approaches. Using data from a pediatric HIV study, we illustrate methods for predicting random intercepts and slopes in CD4 cell count as well as predicting a subject’s actual response at a time point of significant interest.

2. Methods

2.1 Models and Posterior Mean Predictions

We use two familiar normal-theory mixed linear models for illustration: the random intercept and random intercept/slope models, respectively. The random intercept (or one-way random effects ANOVA) model is given as follows (e.g., Searle et al., 1992):

\[
Y_{ij} = \mu + b_i + e_{ij}
\]  
(i = 1,2,…, k; j = 1,2,…, n_i), with i indexing subject and j indexing the observation. Typical normality assumptions dictate that \(b_i \sim N(0,\sigma_b^2)\) and \(e_{ij} \sim N(0,\sigma_w^2)\), with independence across subjects and between the random terms \(b_i\) and \(e_{ij}\).

Under model (1), a common objective is to predict the ith subject’s random subject-specific mean, i.e., \(\mu_i = \mu + b_i\) (i=1,…,k). The EBLUP, as provided by standard mixed model software, is an estimate of the posterior mean \(E(\mu_i \mid Y) = E(\mu_i \mid Y_i)\), where \(Y\) and \(Y_i\) denote the complete and ith subject-specific data vectors, respectively:
\[ \tilde{\mu}_i = E(\mu | Y_i = y_i) = \nu_i \bar{y}_i + (1 - \nu_i)\mu \quad , \]  

where \( \bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij} \) and \( \nu_i = [1 + \sigma_w^2/(n_i \sigma_b^2)]^{-1} \). From this form it is clear that the parameter \( \nu_i \) governs the extent to which the predicted value “shrinks” toward the population mean \( \mu \), with more excessive shrinkage occurring when \( \nu_i \) is small (i.e., when \( \sigma_w^2/(n_i \sigma_b^2) \) is large).

Technically, the BLUP is obtained by replacing \( \mu \) in (2) by its best linear unbiased estimate (Searle et al., 1992), whereas in practice the EBLUP also replaces the variance components in (2) by their estimates.

For a second illustration, consider the random intercept/slope model, also known as a randomized regression or linear growth curve model (e.g., Diggle et al., 1994):

\[ Y_{ij} = (\alpha + a_i) + (\beta + b_i) t_{ij} + e_{ij} \quad (3) \]

\((i = 1,2,\ldots, k; j = 1,2,\ldots, n_i)\), where \( t_{ij} \) denotes the time at which \( Y_{ij} \) is measured. Typically this model assumes independence across subjects and normally distributed random effects as follows:

\[ \begin{bmatrix} a_i \\ b_i \\ e_{ij} \end{bmatrix} \sim N_3 \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix} \right) , \]

with \( \sigma_1^2, \sigma_2^2, \sigma_{12}, \) and \( \sigma_e^2 \) denoting the variance of the random intercepts, the variance of the random slopes, their covariance, and the random error variance, respectively.

Under model (3), it is common to seek predictions of the ith subject’s random intercept \((\alpha_i = \alpha + a_i)\) and slope \((\beta_i = \beta + b_i)\). As with model (1) and most feasible mixed linear models, standard software provides EBLUPs for these quantities. In this case, they are estimates of the posterior means \( E(\alpha_i | Y_i) = E(\alpha_i | Y_i) \) and \( E(\beta_i | Y_i) = E(\beta_i | Y_i) \). The normality assumptions
accompanying model (3) yield
\[ \tilde{\beta}_i = E(\beta_i \mid Y_i = y_i) = \beta + [\sigma_{12} \mathbf{1}'_{n_i} + \sigma_2^2 \mathbf{t}'_i] \Sigma_i^{-1} (y_i - \alpha \mathbf{1}_{n_i} - \beta \mathbf{t}_i), \]  
(4)
where \( \Sigma_i = \text{Var}(Y_i) = Z_i \Delta Z'_i + \sigma^2 \mathbf{1}_{n_i} \), \( Z_i \) is the design matrix for the simple linear regression of \( Y_i \) on time \( (t_i) \) for subject \( i \), and \( \Delta = \text{Var}[(a_i, b_i)'] \). Assuming \( n_i \geq 2 \), Lyles and Xu (1999) showed that \( E(\beta_i \mid Y_i) \) takes an intuitively appealing form:
\[ \tilde{\beta}_i = E(\beta_i \mid Y_i) = \gamma_{i1} + \gamma_{i2} \hat{\alpha}_{i,ols} + \gamma_{i3} \hat{\beta}_{i,ols}, \]  
(5)
where \( \hat{\alpha}_{i,ols} \) and \( \hat{\beta}_{i,ols} \) represent the ordinary least squares (OLS) intercept and slope from regressing \( Y_i \) on \( t_i \). The coefficients in (5) are given by
\[ \gamma_{i2} = (\sigma_{12}\gamma_{\beta i} - \sigma_2^2\gamma_{a\beta}) / \delta_i, \quad \gamma_{i3} = (\sigma_2^2\gamma_{ai} - \sigma_{12}\gamma_{a\beta}) / \delta_i, \quad \text{and} \quad \gamma_{i1} = \beta(1 - \gamma_{i3}) - \sigma\gamma_{i2}, \]
with \( \delta_i = (\gamma_{ai}\gamma_{\beta i} - \sigma_{a\beta}^2), \quad \gamma_{ai} = \text{Var}(\hat{\alpha}_{i,ols}) = \sigma_i^2 + \sigma_2^2 \{1/n_i + \bar{t}_i^2/[(n_i-1)s_{ti}^2]\}, \quad \gamma_{\beta i} = \text{Var}(\hat{\beta}_{i,ols}) = \sigma_2^2 + \bar{t}_i^2\sigma^2/[(n_i-1)s_{ti}^2], \) where \( \bar{t}_i \) and \( s_{ti}^2 \) denote the sample mean and variance of the observation times \( t_i = (t_{i1}, \ldots, t_{in_i})' \).

Similarly, one can show that
\[ \tilde{\alpha}_i = E(\alpha_i \mid Y_i) = \tau_{i1} + \tau_{i2} \hat{\alpha}_{i,ols} + \tau_{i3} \hat{\beta}_{i,ols}, \]  
(6)
with \( \tau_{i2} = (\sigma_1^2\gamma_{\beta i} - \sigma_{12}\gamma_{a\beta}) / \delta_i, \quad \tau_{i3} = (\sigma_{12}\gamma_{ai} - \sigma_1^2\gamma_{a\beta}) / \delta_i, \quad \text{and} \quad \tau_{i1} = \alpha(1 - \tau_{i2}) - \beta\tau_{i3}. \)

Now, consider the problem of predicting the unknown response under model (3) for subject \( i \) at some clinically or otherwise significant time point \( (t_i^*) \). In other words, we seek to predict the value of \( Y_{it}^* = E(Y_{ij} \mid \alpha_i, \beta_i, t_{ij} = t_i^*) = \alpha_i + \beta_i t_i^* \). Clearly, the posterior mean of \( Y_{it}^* \) is
\[ \tilde{Y}_{it}^* = E(Y_{it}^* \mid Y_i) = \tilde{\alpha}_i + \tilde{\beta}_i t_i^*, \]  
(7)
where \( \tilde{\beta}_i \) and \( \tilde{\alpha}_i \) are as defined in (5) and (6), respectively, for \( n_i \geq 2 \). EBLUPs for \( \tilde{\beta}_i \) and \( \tilde{\alpha}_i \) are obtained by inserting parameter estimates into the expressions for \( E(\beta_i \mid Y_i) \) and \( E(\alpha_i \mid Y_i) \), where \( n_i=1 \) is permissible. The EBLUP for \( Y^*_i \) inserts the EBLUPs for \( \tilde{\beta}_i \) and \( \tilde{\alpha}_i \) into (7).

2.2 Constrained Bayes Predictions

The constrained Bayes (CB) approach (Louis, 1984) was extended by Ghosh (1992) into a flexible paradigm allowing minimization of a mean squared error criterion subject to matching the posterior expectation of the first two moments of a parameter distribution to the corresponding moments of the histogram of the set of estimates. This general idea provides a natural alternative to the EBLUP in the mixed linear models context when overshrinkage could detract from the desired application of predicted values.

Lyles and Xu (1999) applied a slight adaptation of the CB idea under models (1) and (3) by minimizing prediction mean squared error (MSEP) among unbiased candidates whose variances match that of the assumed random effects distribution. While this must result in some sacrifice in overall MSEP relative to the posterior mean, it provides a set of predictions that more faithfully reproduce the underlying distribution of interest and are less likely to under-represent the extremeness of experimental units in the tails.

Under model (1), the CB predictor for \( \mu_i \) recommended by Lyles and Xu is obtained directly by forcing the first two moments of the \( \tilde{\mu}_i \) and \( \mu_i \) distributions to match:

\[
\tilde{\mu}_{i, LX} = \sqrt{v_i} \bar{y}_i + (1 - \sqrt{v_i}) \mu .
\] (8)

The square root is indicative of the reduction in shrinkage relative to the posterior mean in (2).

Under model (3), use of a Lagrangian multiplier to enforce equality of the second moments while minimizing MSEP yields a constrained Bayes alternative to the posterior mean in (5):
The coefficients in (9) are defined as

\[ \gamma_{1i} = \beta(1 - \gamma_{13}) - \alpha \gamma_{12}, \quad \gamma_{12} = \pm \eta_i \left\{ \sigma_2^2 [v_{\beta i} + \eta_i (2 \epsilon_{\alpha \beta i} + \eta_i v_{\alpha i})] \right\}^{1/2}, \quad \text{and} \quad \gamma_{13} = \gamma_{12}/\eta_i, \]

where \( \eta_i = (v_{\beta i} \sigma_{12} - \sigma_{22}^2 \epsilon_{\alpha \beta i})(v_{\alpha i} \sigma_{22}^2 - \sigma_{12} \epsilon_{\alpha \beta i})^{-1}. \) The “±” sign in front of \( \gamma_{12} \) is needed because there are two roots, although the positive root is usually correct. We take the positive or negative root for \( \gamma_{12} \) depending on which yields the lower value of the MSEP criterion:

\[ \text{MSEP} = E(\tilde{\beta}_i - \beta_i)^2 = (\gamma_{12}^2 v_{\alpha i} + \gamma_{13}^2 \epsilon_{\beta i} + 2 \gamma_{12} \gamma_{13} \epsilon_{\alpha \beta i}) - 2(\gamma_{12} \sigma_{12} + \gamma_{13} \sigma_{22}) + \sigma_2^2. \] (10)

The definitions of \( \eta_i \) and \( \gamma_{12} \) given here serve to correct an error present in Lyles and Xu (1999).

The Appendix provides analogous constrained Bayes predictors for \( \alpha_i \) and for \( Y_{it}^* \). We obtain empirical constrained Bayes (ECB) predictions for practical use by replacing unknown parameters by their estimates in equations (8), (9), (A1), and (A3), and when calculating the MSEP criterion in (10).

In contrast to the preceding direct model-specific CB predictors, consider the general CB paradigm provided by Ghosh (1992). Using \( \beta_i \) under model (3) to illustrate, we first take \( \tilde{\beta}_{i,B} \) to indicate the posterior mean (or Bayes) predictor for subject \( i \), where an algebraic expression for \( \tilde{\beta}_{i,B} \) was given in (5). Ghosh’s approach defines the CB estimate \( (\tilde{\beta}_{i,G}) \) as follows:

\[ \tilde{\beta}_{i,G} = w \tilde{\beta}_{i,B} + (1 - w) \tilde{\beta}_B, \] (11)

where \( \tilde{\beta}_B = k^{-1} \sum_{j=1}^k \tilde{\beta}_{j,B}, \quad w = (1 + H_1/H_2)^{1/2}, \quad H_2 = \sum_{j=1}^k (\tilde{\beta}_{j,B} - \tilde{\beta}_B)^2, \) and

\[ H_1 = \text{tr}[\text{Var}(\beta - \tilde{\beta} I_k | Y)] = (1 - k^{-1}) \sum_{j=1}^k \text{Var}(\beta_j | Y_j), \] (12)
with \( \mathbf{\beta} \) representing the \( k \)-vector \((\beta_1, \beta_2, ..., \beta_k)'\).

We supply the latter equality in (12) as a result of assumed independence across experimental units for the class of mixed models under consideration here. Note that in addition to the posterior means, this paradigm requires only the corresponding posterior variances. Using our previous notation (see equation (4) and Appendix), we have:

\[
\text{Var}(\beta_i | \mathbf{Y}_i) = \sigma^2 - \left[ \sigma_{12} \mathbf{1}'_{n_i} + \sigma_2 \mathbf{t}_i' \right] \Sigma_i^{-1} \left[ \sigma_{12} \mathbf{1}'_{n_i} + \sigma_2 \mathbf{t}_i' \right]' ,
\]

(13)

\[
\text{Var}(\alpha_i | \mathbf{Y}_i) = \sigma_1^2 - \left[ \sigma_1^2 \mathbf{1}'_{n_i} + \sigma_1 \mathbf{t}_i' \right] \Sigma_i^{-1} \left[ \sigma_1^2 \mathbf{1}'_{n_i} + \sigma_1 \mathbf{t}_i' \right]' ,
\]

(14)

and

\[
\text{Var}(\mathbf{Y}_it^* | \mathbf{Y}_i) = \text{Var}(\mathbf{Y}_it^*) - \left[ \psi_{i1} \mathbf{1}'_{n_i} + \psi_{i2} \mathbf{t}_i' \right] \Sigma_i^{-1} \left[ \psi_{i1} \mathbf{1}'_{n_i} + \psi_{i2} \mathbf{t}_i' \right]' .
\]

(15)

ECB predictions for practical use could be obtained by replacing unknown parameters by their estimates when computing the posterior means and variances, and the building blocks for these calculations are already built into standard software for mixed linear models.

2.3 Incorporating fixed or time-dependent covariates

Consider the following extensions of models (1) and (3) to include a set of \( T \) covariates, some of which may be time-dependent:

\[
\mathbf{Y}_{ij} = \mathbf{\mu} + \mathbf{b}_i + \sum_{t=1}^{T} \theta_t \mathbf{c}_{ijt} + \mathbf{e}_{ij}
\]

(16)

and

\[
\mathbf{Y}_{ij} = (\alpha + \mathbf{a}_i) + (\mathbf{\beta} + \mathbf{b}_i)\mathbf{t}_{ij} + \sum_{t=1}^{T} \theta_t \mathbf{c}_{ijt} + \mathbf{e}_{ij} ,
\]

(17)

where \( \mathbf{c}_{ijt} \) represents the observed value of the \( t \)-th covariate for subject \( i \) at time point \( j \) \((t=1,..,T; i=1,..,k; j=1,..,n_i)\). Let \( \mathbf{c}_{ij}' = (\mathbf{c}_{ij1}, \mathbf{c}_{ij2}, ..., \mathbf{c}_{ijT})' \) and form the \( n_i \times t \) matrix \( \mathbf{C}_i \) by stacking the row vectors \( \mathbf{c}_{ij}' \) in order. Now, define the transformed observed data vector \( \mathbf{y}_i^* = \mathbf{y}_i - \mathbf{C}_i \mathbf{\theta} \), where \( \mathbf{\theta} = (\theta_1, \theta_2, ..., \theta_T)' \). The extension to the posterior mean formula in (2) is
\[ \tilde{\mu}_i = E(\mu_i \mid Y_i, C_i) = v_i \bar{y}_i^* + (1 - v_i) \mu, \]  

(18)

with \( \mu_i \) and \( v_i \) defined exactly as before and \( \bar{y}_i^* = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}^* \). In practice, one may be more likely to seek to predict \( \tilde{Y}_{ij} = E(Y_{ij} \mid b_i, C_i) = \mu_i + c_{ij}^t \theta \). Standard mixed linear model software typically provides the EBLUP for \( b_i \), from which EBLUPs for \( \mu_i \) and \( \tilde{Y}_{ij} \) are easily obtained.

Similarly, extensions to (4) and (5) under the randomized regression model (17) are

\[ \tilde{\beta}_i = E(\beta_i \mid Y_i, C_i) = \beta + [\sigma_{12} I_{n_i} + \sigma_2^2 t_i^2 \Sigma_i^{-1} (y_i^* - \alpha 1_{n_i} - \beta t_i)] \]

and

\[ \tilde{\beta}_i = E(\beta_i \mid Y_i, C_i) = \gamma_{i1} + \gamma_{i2} \hat{\alpha}_{i, ols} + \gamma_{i3} \hat{\beta}_{i, ols}, \]  

(19)

where \( \beta_i \), \( \gamma_{i1} \), \( \gamma_{i2} \), and \( \gamma_{i3} \) are defined as before, but with \( \hat{\alpha}_{i, ols} \) and \( \hat{\beta}_{i, ols} \) now representing the OLS intercept and slope from regressing \( y_i^* \) on \( t_i \). Again, the algebraic expression in (19) requires \( n_i \geq 2 \). Standard software typically provides EBLUPs for \( a_i \) and \( b_i \), from which EBLUPs for \( \alpha_i \) and \( \beta_i \) follow directly. In turn, the analogue to equation (7) becomes

\[ \tilde{Y}_{it}^* = E(Y_{it}^* \mid Y_i, C_i) = \tilde{\alpha}_i + \tilde{\beta}_i t_i^* + c_{i,t}^* \theta, \]  

(20)

which can arguably be defined only for non-time-dependent covariates unless the values of any time dependent ones are known at time \( t_i^* \) (as indicated by the notation \( c_{i,t}^* \)).

Extensions of the CB predictors \( \tilde{\mu}_{i, LX}, \tilde{\beta}_{i, LX}, \) and \( \tilde{\alpha}_{i, LX} \) in equations (8), (9), and (A1) with covariate adjustment according to models (16) and (17) require no changes to the coefficients already given, once the transformation \( y_i^* = y_i - C_i \theta \) is made. The same is true for \( \tilde{Y}_{it,LX}^* \) in equation (A3), except one adds the term \( c_{i,t}^* \theta \) as in (20). Corresponding ECB
predictions for practical use follow, once estimates of the mixed linear model parameters are inserted. More importantly, by adapting the paradigm of Ghosh (1992) as in (11) and (12), ECB predictions appear straightforward for a broad class of general linear mixed models because i) EBLUPs accounting for covariates come directly out of standard software, and ii) the required conditional variances [e.g., (13)-(15)] are unchanged by the addition of covariates. Note that in the case of $\widetilde{Y}_{it}^*$, Ghosh’s paradigm requires a separate application of posterior mean and variance calculations analogous to those in (11) and (12) for each unique value of $t_i^*$.

3. Example

As an example, we utilize longitudinal data on CD4 cell counts collected for the Pediatric Pulmonary and Cardiovascular Complications of Vertically Transmitted (P²C²) HIV Infection Study (The P²C² Study Group, 1996). This National Heart, Lung, and Blood Institute-funded study enrolled infants born to HIV-positive women during the years 1990-1993, and followed them prospectively during the first few years of life. Specifically, we analyze data on 59 vertically infected infants who contributed a total of 539 CD4 counts over time, with the number of measurements per child ranging from 3-19. Initial CD4 counts were typically observed at or within a few weeks of birth. The length of follow-up on children ranged from 1 to 6 years, with a median of 3.5 years. Also recorded for each child was the age at which he or she was determined to have reached Class A (mildly symptomatic) HIV status (Centers for Disease Control and Prevention, 1994). Across the 59 subjects, this age ranged from 0.4 to 16 months.

We fit a mixed linear model to these data, with age as the longitudinal metameter. While there was some indication of right skewness in the CD4 counts, standard transformations tended to overcorrect this and for the sake of a clear illustration we chose to analyze the untransformed CD4 counts. For an illustration with covariate adjustment, the child’s gender (1 for male, 0 for
female) and the concurrent CD8 cell count were accounted for via the following model:

$$CD4_{ij} = (\alpha + a_i) + (\beta + b_i) \text{AGE}_{ij} + \theta_1 \text{GENDER}_i + \theta_2 \text{CD8}_{ij} + e_{ij}. \quad (21)$$

The primary objective is to compare EBLUP and ECB predictions of the random intercepts ($\alpha_i = \alpha + a_i$) and random slopes ($\beta_i = \beta + b_i$). For this purpose, we investigate both the direct ECB approach patterned after Lyles and Xu (1999; ‘LX ECB’) and the ECB method following Ghosh (1992). Secondly, we also compare EBLUP and Ghosh ECB predictions of $Y_{it}^*$, where $Y_{it}^* = \alpha_i + \beta_i t_i^* + \theta_1 \text{GENDER}_i + \theta_2 \text{CD8}_i$ represents the unknown model-based CD4 count at time $t_i^*$. For this latter purpose, $t_i^*$ was defined as the age at which the child was diagnosed with Class A HIV disease, and model (21) was re-fit with the initial CD8 count (CD8$_i$) in place of the time-dependent version in light of the fact that CD8 was unrecorded at the times $t_i^*$. Table I provides the coefficient and variance component estimates from fitting both versions of model (21) by maximum likelihood via SAS PROC MIXED (SAS Institute, Inc., 2004). The table indicates a highly significant average decline of approximately 400 CD4 cells per year, little effect of gender, and a significant positive association with the CD8 count, regardless of whether it was measured only initially or treated as time-dependent.

In Figure 1A, we plot EBLUPs for the random intercepts $\alpha_i$ against the corresponding Ghosh ECB predictions, based on the model treating CD8 as time-dependent. The EBLUPs were obtained directly from the mixed linear model software, and the Ghosh ECBs were computed readily using the EBLUPs and posterior variance calculations with variance components replaced by their MLEs (e.g., eqns. 11-15). The reduction in shrinkage afforded by the CB method is evident in the characteristic ‘tilting’ in the pattern of plotted points.

Figure 1B plots the LX ECB predictions of $\alpha_i$ versus the Ghosh ECBs. To obtain the LX
ECBs, we inserted the MLEs for variance components into the formulae provided herein, with covariate adjustment as described in Section 2.3. With a few exceptions, the two approaches produce essentially identical results. The sample means of the 59 EBLUP, Ghosh ECB, and LX ECB predicted values were 1675.5, 1675.5, and 1675.3, respectively. The corresponding sample variances were 365470, 475026, and 473752. Comparing these to $\hat{\alpha} = 1675.5$ and $\hat{\sigma}_1^2 = 468832$ (Table I) highlights the moment matching characteristics of the CB approaches, as well as the overshrinkage of the EBLUP.

**Figure 2** is the counterpart to Figure 1, for the predicted random slopes ($\beta_i$). The tilting remains prominent in Figure 2A, while Figure 2B reveals somewhat more pronounced discrepancies between the Ghosh and LX ECB predictions than in the case of the intercepts. The sample means of the EBLUP, Ghosh ECB, and LX ECB predicted values were $-388.2$, $-388.2$, and $-395.3$, respectively, with sample variances of 27904, 48316, and 49401. Comparing these to $\hat{\beta} = -388.2$ and $\hat{\sigma}_2^2 = 47843$ (Table I) again highlights the ECB properties in action.

**Figure 3** illustrates the reduction in shrinkage of the Ghosh ECB predictions (open circles) of CD4 cell counts at the time of Class A disease ($Y_{it}^*$), relative to the EBLUPs (closed circles). Separate plots are presented for females and males, with overlays of the population average regression lines calculated at the overall mean of the 59 initial CD8 counts (1294.7 cells). The lines provide a relevant visual reference based on the fit of model (21) (Table I), although we do not expect the plotted points to directly follow these linear trends given that subjects with less rapidly declining CD4 counts theoretically reach Class A disease at later ages.

4. **Simulation Study**

While the close agreement of the sample means and variances of the ECB predictions to the
corresponding estimated moments ($\hat{\alpha}$ and $\hat{\sigma}_1^2$, $\hat{\beta}$ and $\hat{\sigma}_2^2$) in the real-data example is indicative, we conducted simulation studies to further assess the quality of the variance match and to compare the performance of the Ghosh and LX ECB methods. Several combinations of covariates and true parameter values were examined, with qualitatively similar results. Here, we summarize simulations designed to mimic the conditions observed in the example.

Specifically, we generated data according to model (21) for 20,000 “subjects”, with true parameter values equal to the estimates listed in the top half of Table I. The fabricated CD4 data were unbalanced with $n_i$ ranging randomly between 2 and 10, and measurements were unequally timed over approximate 2 month intervals. Simulated subjects were male or female with probability 0.5. For simplicity, time-varying CD8 counts were generated at each visit from a normal distribution mimicking the sample mean and variance of the initial CD8 counts in the actual example. To illustrate results for predicting $Y_{it}^*$, the same simulation exercise was repeated except with a time independent initial CD8 count in place of the time-varying version. The time point of interest ($t_i^*$) was taken to occur at 2 years for each simulated subject.

Table II summarizes the simulation results for predicting the $\alpha_i$’s and $\beta_i$’s, and Table III summarizes the results for the $Y_{it}^*$’s. In each case, the sample means of the BLUPs and the two CB predictors closely match the true mean of the random variable being predicted. The sample variances over 20,000 simulated subjects for both the LX and Ghosh CB methods are very close to the corresponding true variances in each case, while the overshrinkage of the BLUPs is evident by their notably tighter sampling distributions. As a final note, the empirical prediction MSEs of the LX and Ghosh methods are similar, though predictably somewhat larger than those for the corresponding EBLUPs. In every case, the Ghosh method achieves a small MSE advantage relative to the LX approach.
5. Discussion

Louis (1984) and Ghosh (1992) discuss the motivation and potential benefits of constrained Bayes estimation, which seeks to optimize a traditional MSE criterion subject to matching the posterior expectation of the first two moments of a parameter distribution to the corresponding true moments. In particular, the known overall MSE advantage of the traditional posterior mean approach (which underlies the BLUP in the mixed linear model setting) is sometimes worth sacrificing to obtain a set of predictions with a histogram more closely matching a true distribution of random effects.

Our purpose has been to outline and compare in some detail the application of a direct (‘LX’) CB approach considered by Lyles and Xu (1999) for certain mixed linear models, as opposed to the general method of Ghosh (1992). We explored both approaches in the presence of covariates (possibly time-dependent), and conclude based on simulations and a real-data example that both may be effectively applied to achieve the moment matching goals of the CB paradigm.

The LX approach, while presentable in closed form for the models considered here, relies upon a strict form for candidate predictors and could require cumbersome (if not infeasible) extensions to be applied to arbitrary mixed linear models. Fortunately, however, the general method developed by Ghosh (1992) appears remarkably flexible and consistent in its application for arbitrary fixed and random effects design matrices in the mixed linear model context. In practice, it requires only EBLUPs and estimates of the posterior variances of the random effects being predicted, with the latter readily obtainable under normal-theory mixed models. Further, our simulation studies summarized here (and others, unreported) consistently show the Ghosh approach to be as effective as the direct LX method at matching moments, and also suggest slight prediction MSE gains via its use for unbalanced data. Our conclusion is that implementation of the Ghosh (1992) paradigm for ECB predictions of random effects in
commercial software for mixed linear models (e.g., SAS PROC MIXED and similar procedures in other packages such as Splus, R, SPSS, STATA or BMDP) could comprise both a feasible and potentially useful advance. We would welcome this software advance, for the purpose of allowing practitioners the freedom to select a validated alternative to the traditional EBLUP when overshrinkage could run counter to the objective at hand.
ACKNOWLEDGEMENTS

The efforts of R.H.L. and A.K.M. were supported in part by an R01 from the National Institute of Environmental Health Sciences (ES012458). We thank the P²C²HIV study investigators and the National Heart, Lung and Blood Institute for use of their database, and appreciate support provided by the Biostatistics Core of the Emory Center for AIDS Research (P30 AI050409).

REFERENCES


APPENDIX

A constrained Bayes predictor for the $i$th subject’s random intercept ($\alpha_i$) may be obtained via calculations similar to those leading to $\tilde{\beta}_{i,	ext{LX}}$ in equation (8), as follows:

$$\tilde{\alpha}_{i,	ext{LX}} = \tau_{i1} + \tau_{i2}\hat{\alpha}_{i,\text{ols}} + \tau_{i3}\hat{\beta}_{i,\text{ols}}, \quad (A1)$$

where $$\tau_{i1} = \alpha (1 - \tau_{i2}) - \beta \tau_{i3}, \quad \tau_{i2} = \pm\{\sigma_1^2/[v_{ai} + \kappa_i(2c_{\alpha \beta i} + \kappa_i v_{\beta i})]\}^{1/2}, \text{ and } \tau_{i3} = \kappa_i \tau_{i2},$$

with $\kappa_i = (v_{ai}\sigma_{i2} - \sigma_1^2c_{\alpha \beta i})(v_{\beta i}\sigma_1^2 - \sigma_{i2}c_{\alpha \beta i})^{-1}$. Specifically, $\tilde{\alpha}_i$ defined in this way minimizes MSEP among predictors of the form (A1) subject to the constraints that $E(\tilde{\alpha}_i) = E(\alpha_i) = \alpha$ and $\text{Var}(\tilde{\alpha}_i) = \text{Var}(\alpha_i) = \sigma_1^2$, where the MSEP criterion is

$$E(\tilde{\alpha}_i - \alpha_i)^2 = (\tau_{i1}^2v_{ai} + \tau_{i2}^2v_{\beta i} + 2\tau_{i2}\tau_{i3}c_{\alpha \beta i}) - 2(\tau_{i2}\sigma_1^2 + \tau_{i3}\sigma_{i2}) + \sigma_1^2. \quad (A2)$$

In an analogous manner, we define a constrained Bayes predictor for $Y_{it}^*$ as

$$\tilde{Y}_{it,\text{LX}} = \varphi_{i1} + \varphi_{i2}\tilde{\alpha}_{i,\text{ols}} + \varphi_{i3}\hat{\beta}_{i,\text{ols}}, \quad (A3)$$

where

$$\varphi_{i1} = \alpha (1 - \varphi_{i2}) - \beta (\varphi_{i3} - t_i^*), \quad \varphi_{i2} = \pm\{\psi_{i3}/[v_{ai} + \omega_i(2c_{\alpha \beta i} + \omega_i v_{\beta i})]\}^{1/2}, \text{ and } \varphi_{i3} = \omega_i \varphi_{i2},$$

with $\omega_i = (v_{ai}\psi_{i2} - \psi_{i1}c_{\alpha \beta i})(v_{\beta i}\psi_{i1} - \psi_{i2}c_{\alpha \beta i})^{-1}$, $\psi_{i1} = \sigma_1^2 + t_i^*\sigma_{i2}$, $\psi_{i2} = \sigma_{i2} + t_i^*\sigma_2^2$, and

$$\psi_{i3} = \sigma_1^2 + t_i^*\sigma_2^2 + 2t_i^*\sigma_{i2}. \text{ This minimizes MSEP for predictors of the form (A3), subject to the constraints } E(\tilde{Y}_{it}^*) = E(Y_{it}^*) = \alpha + \beta t_i^* \text{ and } \text{Var}(\tilde{Y}_{it}^*) = \text{Var}(Y_{it}^*) = \varphi_{i1}^2v_{ai} + \varphi_{i3}^2v_{\beta i} + 2\varphi_{i2}\varphi_{i3}c_{\alpha \beta i}.$$

As with $\gamma_{i2}$ in equation (9), technically the choice of the positive or negative root to define $\tau_{i2}$ and $\varphi_{i2}$ should be based on which minimizes the corresponding MSEP criterion.

However, in our experience the negative roots have never applied except in the case of $\gamma_{i2}$.  

17
Table I. Summary of mixed linear models fit to CD4 cell count data *

<table>
<thead>
<tr>
<th>Model †</th>
<th>Coefficient</th>
<th>Estimate (standard error)</th>
<th>Variance Component</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD8 as time-dependent</td>
<td>$\alpha$</td>
<td>1675.50 (138.27)</td>
<td>$\sigma^2_1$</td>
<td>468832</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>$-388.17$ (38.06)</td>
<td>$\sigma^2_2$</td>
<td>47843</td>
</tr>
<tr>
<td></td>
<td>$\theta_1$</td>
<td>$-163.41$ (146.61)</td>
<td>$\sigma_{12}$</td>
<td>$-103226$</td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>0.26 (0.03)</td>
<td>$\sigma^2$</td>
<td>477810</td>
</tr>
</tbody>
</table>

CD8 as time-independent (initial value) | $\alpha$ | 1735.88 (188.60) | $\sigma^2_1$ | 429957 |
|          | $\beta$    | $-417.51$ (40.57) | $\sigma^2_2$ | 55206 |
|          | $\theta_1$ | $-105.28$ (146.61) | $\sigma_{12}$ | $-102537$ |
|          | $\theta_2$ | 0.27 (0.10) | $\sigma^2$ | 529062 |

* Data from P2C2 HIV Infection Study (The P2C2 Study Group, 1996)
† CD4$_{ij}$ = ($\alpha$ + $a_i$) + ($\beta$ + $b_i$)AGE$_{ij}$ + $\theta_1$GENDER$_i$ + $\theta_2$CD8 + $e_{ij}$

Table II. Simulation results for random intercept and slope predictions *,†

<table>
<thead>
<tr>
<th>True $\alpha$’s</th>
<th>$\tilde{\alpha}_{i,\text{BLUP}}$</th>
<th>$\tilde{\alpha}_{i,\text{LX}}$</th>
<th>$\tilde{\alpha}_{i,G}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1675.5</td>
<td>1680.8</td>
<td>1681.2</td>
</tr>
<tr>
<td>Variance</td>
<td>468832</td>
<td>376252</td>
<td>475834</td>
</tr>
<tr>
<td>Prediction MSE</td>
<td>--</td>
<td>98600</td>
<td>105400</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>True $\beta$’s</th>
<th>$\tilde{\beta}_{i,\text{BLUP}}$</th>
<th>$\tilde{\beta}_{i,\text{LX}}$</th>
<th>$\tilde{\beta}_{i,G}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$-388.2$</td>
<td>$-389.4$</td>
<td>$-386.2$</td>
</tr>
<tr>
<td>Variance</td>
<td>47843</td>
<td>16115</td>
<td>48693</td>
</tr>
<tr>
<td>Prediction MSE</td>
<td>--</td>
<td>31593</td>
<td>40375</td>
</tr>
</tbody>
</table>

* Data simulated to mimic model (21) with parameters equal to estimates in Table I (top)
† Predictions computed assuming parameter values that generated the data
Table III. Simulation results for $Y_{it}^*$ predictions *†

<table>
<thead>
<tr>
<th></th>
<th>True $Y_{it}^*$'s</th>
<th>$\tilde{Y}_{it,\text{BLUP}}$</th>
<th>$\tilde{Y}_{it,LX}$</th>
<th>$\tilde{Y}_{it,G}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1156.4</td>
<td>1158.4</td>
<td>1158.3</td>
<td>1158.4</td>
</tr>
<tr>
<td>Variance</td>
<td>289054</td>
<td>177184</td>
<td>289249</td>
<td>288880</td>
</tr>
<tr>
<td>Prediction MSE</td>
<td>--</td>
<td>110636</td>
<td>128884</td>
<td>124112</td>
</tr>
</tbody>
</table>

* Data simulated to mimic model (21) with parameters equal to estimates in Table I (bottom)
† Predictions computed assuming parameter values that generated the data
FIGURE LEGENDS

Figure 1. EBLUP (panel A) and LX ECB (panel B) vs. Ghosh ECB predictions for random intercepts ($\alpha_i$) based on the fit of model (21) with CD8 count as a time-dependent covariate (Table I).

Figure 2. EBLUP (panel A) and LX ECB (panel B) vs. Ghosh ECB predictions for random slopes ($\beta_i$) based on the fit of model (21) with CD8 count as a time-dependent covariate (Table I).

Figure 3. EBLUP (dark circle) vs. Ghosh ECB (open circle) predictions of $Y_{it}^* = \alpha_i + \beta_i t_i + \theta_1 \text{GENDER}_i + \theta_2 \text{CD8}_i$ for females (panel A) and males (panel B), with initial CD8 count as a time-independent covariate. Plotted population-average regression lines are computed at the overall mean initial CD8 cell count.