AN EMPIRICAL BAYES DERIVATION OF BEST LINEAR UNBIASED PREDICTORS

ROHANA J. KARUNAMUNI

Received 8 October 2001

Let \((Y_1, \theta_1), \ldots, (Y_n, \theta_n)\) be independent real-valued random vectors with \(Y_i\), given \(\theta_i\), is distributed according to a distribution depending only on \(\theta_i\) for \(i = 1, \ldots, n\). In this paper, best linear unbiased predictors (BLUPs) of the \(\theta_i\)'s are investigated. We show that BLUPs of \(\theta_i\)'s do not exist in certain situations. Furthermore, we present a general empirical Bayes technique for deriving BLUPs.

2000 Mathematics Subject Classification: 62C12.

1. Introduction. Let \((Y_1, \theta_1), \ldots, (Y_n, \theta_n)\) be independent real-valued random vectors satisfying the following:

(i) conditional on \(\theta_i\), \(Y_i\) is distributed according to a distribution depending only on \(\theta_i\), \(E(Y_i | \theta_i) = \theta_i\), and \(\text{Var}(Y_i | \theta_i) = \mu_2(\theta_i)\) with independence over \(\theta_i\), \(i = 1, \ldots, n\);

(ii) \(\theta_i\)'s are independent with \(E(\theta_i) = \mu_i\) and \(\text{Var}(\theta_i) = \tau^2\), \(i = 1, \ldots, n\), where \(\mu_i\)'s and \(\tau\) are fixed numbers;

(iii) \(0 < D_i = E\mu_2(\theta_i) < \infty\), \(i = 1, \ldots, n\), where \(D_i\)'s are fixed numbers.

A special case of the above model is the so-called mixed linear model given by

\[
y_i = x_i' \beta + u_i + e_i, \quad i = 1, \ldots, n,
\]

where \(x_i = (x_{i1}, \ldots, x_{ik})'\), \(\beta\) is a \(k\) vector of unknown parameters, sampling errors \(e_i\) and the random effects \(u_i\) are independently distributed with \(E(e_i) = 0\), \(E(u_i) = 0\), \(\text{Var}(e_i) = D_i\), and \(\text{Var}(u_i) = \tau^2\), \(i = 1, \ldots, n\). The mixed linear model can also be written as

\[
y_i = \theta_i + e_i, \quad \theta_i = x_i' \beta + u_i, \quad i = 1, \ldots, n.
\]

The aim of this paper is about the best linear unbiased predictors (BLUPs) of \(\theta_i\), \(i = 1, \ldots, n\). BLUPs are estimates of the realized value of the random variable \(\theta_i\) and are linear in the sense that they are linear functions of the data, \(y_i\); unbiased in the sense that the average value of the estimate is equal to the average of the quantity being estimated; best in the sense that they have minimum mean squared error within the class of linear unbiased estimators; and predictors to distinguish them from the estimators.
The first derivation of BLUPs seems to have been given by Henderson [8] who studied a more general version of the mixed linear model, namely, \( Y = X\beta + Z\upsilon + e \), where \( Z \) is a known design matrix, while \( e \) is a vector of errors which is uncorrelated with random vector \( \upsilon \). Henderson has described BLUPs as being "joint maximum likelihood estimates," and has assumed that \( \upsilon \) and \( e \) are normally distributed. A number of other derivations have been given over the years. Within the classical school of thought, BLUPs have been shown to have minimum mean squared error within the class of linear unbiased estimators (see Henderson [9] and Harville [5, 6]); in the Bayesian framework, BLUPs have been derived as the posterior mean of the parameter of interest with a noninformative prior for \( \beta \) (see Dempfle [2], and Lindley and Smith [13]); and another derivation of BLUPs has been given by Jiang [10] showing a connection between BLUP and restricted maximum likelihood. In an interesting review article, Robinson [14] has given a wide-ranging account of BLUPs in the mixed model with examples and applications. However, his discussion on empirical Bayes methods and their connection with BLUPs is rather limited—he only states that BLUPs are equivalent to one of the techniques of parametric empirical Bayes methodology, see Robinson [14, Section 5.7]. Commenting on Robinson’s paper, Harville [7] has demonstrated a connection between BLUP and empirical Bayes estimators for a one-way random effects model given by 
\[ y_{ij} = \mu + a_i + e_{ij} \quad (i = 1, \ldots, n; \ j = 1, \ldots, J_i) \]
The purpose of this paper is to investigate the connection between BLUPs and empirical Bayes estimators more closely. In this paper, our discussion is focused on the general model described by (i), (ii), and (iii) at the beginning of the introduction. We first show the nonexistence of BLUPs under certain situations, and then we present a general empirical Bayes technique for deriving BLUPs.

Briefly, our claim is as follows: for \( i = 1, \ldots, n \), suppose \( \delta_i(Y_i, \hat{\mu}_i) \) denotes the linear Bayes estimator of \( \theta_i \) with respect to squared error loss, and suppose that \( \hat{\theta}_i \) denotes the best linear unbiased estimator (BLUE) of \( \theta_i \) based on \( Y_1, \ldots, Y_n \). Then \( \delta_i(Y_i, \hat{\mu}_i) \) is the BLUP of \( \theta_i \), \( i = 1, \ldots, n \), whenever BLUPs exist. This general argument is in line with Robinson’s statement that BLUPs are equivalent to one of the techniques of parametric empirical Bayes methods. Our argument, however, gives a clear process of derivation and would be quite useful when deriving BLUPs in applications.

The main results of the paper are given in Section 2. The proofs are deferred to Section 3.

2. **Main results.** Throughout this section, we assume that \( (Y_1, \theta_1), \ldots, (Y_n, \theta_n) \) are independent random vectors, where the \( Y_i \)'s are observed whereas the \( \theta_i \)'s are not. We first derive expressions for BLUPs of \( \theta_i \), \( i = 1, \ldots, n \), under various conditions on the prior parameters of the \( \theta_i \)'s. Recall that \( \hat{\theta}_i \) is called a BLUP of \( \theta_i \) under squared error loss if \( E(\hat{\theta}_i - \theta_i) = 0 \), \( \hat{\theta}_i \) is a linear combination of the observations \( (Y_1, \ldots, Y_n) \), and \( \hat{\theta}_i \) has the minimum mean squared error \( E(\hat{\theta}_i - \theta_i)^2 \), where \( E \) denotes expectation with respect to all the random variables involved; see, for example, Searle et al. [15, Chapter 7]. In other words, we investigate predictors of the form \( \hat{\theta}_i = \sum_{j=1}^n c_{ij}Y_j \), where \( c_{ij} \geq 0 \) are some constants. In order for \( E(\hat{\theta}_i - \theta_i) = 0 \) to be satisfied, it is required that \( c_{ij} \) satisfy the restriction that \( \sum_{j=1}^n c_{ij}\mu_j = \mu_i \), \( i = 1, \ldots, n \), where \( \mu_k = E(\theta_k) \), \( k = 1, \ldots, n \). When the \( \theta_i \)'s are i.i.d. then the preceding condition reduces to \( \sum_{j=1}^n c_{ij} = 1 \), \( i = 1, \ldots, n \).
**Theorem 2.1.** Let \((Y_1, \theta_1), \ldots, (Y_n, \theta_n)\) be independent real-valued random vectors satisfying the following:

(i) conditional on \(\theta_i\), \(Y_i\) is distributed according to a distribution depending only on \(\theta_i\) and \(E(Y_i \mid \theta_i) = \theta_i\), and \(\text{Var}(Y_i \mid \theta_i) = \mu_2(\theta_i)\) with independence over \(i = 1, \ldots, n\);

(ii) \(\theta_i\)'s are independent with \(E(\theta_i) = \mu_i\) and \(\text{Var}(\theta_i) = \tau^2\) for \(i = 1, \ldots, n\), where \(\mu_i\)'s and \(\tau\) are fixed numbers;

(iii) \(0 < D_i = E\mu_2(\theta_i) < \infty, i = 1, \ldots, n\), where \(D_i\)'s are fixed numbers.

Let \(c_{ij} (\geq 0)\) be constants such that

\[
\sum_{j=1}^{n} c_{ij} \frac{\mu_j}{\sigma_j} = \mu_i, \quad i = 1, \ldots, n.
\]

Then the mean squared error (Bayes risk) of \(\hat{\theta}_i = \sum_{j=1}^{n} c_{ij} Y_j\) as an estimator of \(\theta_i\) is given by

\[
r_i(c) = E \left( \sum_{j=1}^{n} c_{ij} Y_j - \theta_i \right)^2 = \sum_{j=1}^{n} c_{ij}^2 D_j + \tau^2 \left( \sum_{j=1}^{n} c_{ij}^2 + 1 - 2c_{ii} \right).
\]

(2.1)

The values of \(c_{ij}\) that minimize \(r_1(c)\) subject to the restriction \(\sum_{j=1}^{n} c_{ij} \mu_j = \mu_i\) is \(c_{ij}^*\) such that for \(i = 1, \ldots, n\),

\[
c_{ij}^* = \frac{\rho \mu_j}{2(D_j + \tau^2)}, \quad j = 1, \ldots, i-1, i+1, \ldots, n; \quad c_{ii}^* = \frac{2\tau^2 + \rho \mu_i}{2(D_i + \tau^2)},
\]

(2.2)

where \(\rho = \mu_i[D_i(D_i + \tau^2)^{-1}] / \sum_{j=1}^{n} \mu_j^2 [2(D_j + \tau^2)]^{-1}\).

Note that the BLUP \(\sum_{j=1}^{n} c_{ij}^* Y_j\) in **Theorem 2.1** depends on the \(\mu_i\)'s as well. Thus, if \(\mu_i\)'s are completely unknown then there is no BLUP for \(\theta_i\). However, if \(\mu_i\)'s are partially known then a BLUP can be developed as a function of the \(X_j\)'s alone, provided that \(D_i\)'s and \(\tau^2\) are all known. These possibilities are discussed in the following corollaries.

**Corollary 2.2.** Assume that the conditions of **Theorem 2.1** hold. Further, suppose that \(\mu_i = \mu\) for \(i \geq 1\). Then \(\sum_{j=1}^{n} c_{ij}^* Y_j\) reduces to

\[
\left( \frac{D_i}{D_i + \tau^2} \right) \sum_{j=1}^{n} \frac{Y_j(D_j + \tau^2)^{-1}}{D_j + \tau^2} + \left( \frac{\tau^2}{D_i + \tau^2} \right) Y_i.
\]

(2.3)

**Corollary 2.3.** Assume that the conditions of **Theorem 2.1** hold. Further, suppose that \(\mu_i = t_i \beta\), where \(t_i\) are some known constants, \(i = 1, \ldots, n\), and \(\beta\) is an unknown parameter. Then \(\sum_{j=1}^{n} c_{ij}^* Y_j\) reduces to

\[
\left( \frac{D_i}{D_i + \tau^2} \right) t_i \hat{\beta} + \left( \frac{\tau^2}{D_i + \tau^2} \right) Y_i,
\]

(2.4)

where \(\hat{\beta} = [\sum_{j=1}^{n} (D_j + \tau^2)^{-1} t_j^2]^{-1} [\sum_{j=1}^{n} (D_j + \tau^2)^{-1} t_j X_j].\)
**Corollary 2.4.** Assume that the conditions of Theorem 2.1 hold. Further, suppose that \( \mu_i = x_i^\prime \beta \), where \( x_i = (x_{i1}, \ldots, x_{ik})' \) are known vectors and \( \beta \) is a \( k \) vector of unknown parameters, \( i = 1, \ldots, n \). Then \( \sum_{j=1}^{n} c_{ij}^* Y_j \) reduces to

\[
\left( \frac{D_i}{D_i + \tau^2} \right) x_i^* \hat{\beta} + \left( \frac{\tau^2}{D_i + \tau^2} \right) Y_i, \tag{2.5}
\]

where \( \hat{\beta} = [\sum_{j=1}^{n} (D_j + \tau^2)^{-1} x_j^* x_j]^{-1} [\sum_{j=1}^{n} (D_j + \tau^2)^{-1} x_j^* Y_j] \).

We now derive BLUPs using an empirical Bayes technique. Let \( \delta_i(Y_i, \mu_i) = a_i^* Y_i + b_i^* \) denote the linear Bayes estimator of \( \theta_i \) with respect to squared error loss based on the observation \( Y_i, i = 1, \ldots, n \). Then \( a_i^* \) and \( b_i^* \) can be obtained by solving the equations \( \partial \Delta_i / \partial a_i = 0 \) and \( \partial \Delta_i / \partial b_i = 0 \), where \( \Delta_i = E(\theta_i - a_i Y_i - b_i)^2, i = 1, \ldots, n \).

First, assume that the conditions of Theorem 2.1 hold. Then observe that

\[
\frac{\partial \Delta_i}{\partial a_i} = -2E(\theta_i - a_i Y_i - b_i)(Y_i), \quad \frac{\partial \Delta_i}{\partial b_i} = -2E(\theta_i - a_i Y_i - b_i). \tag{2.6}
\]

By setting \( \partial \Delta_i / \partial b_i = 0 \), we obtain \( b_i^* = \mu_i D_i (D_i + \tau^2)^{-1} \). Now substituting the preceding value in \( \partial \Delta_i / \partial a_i = 0 \), we obtain \( a_i^* = \tau^2 (\tau^2 + D_i)^{-1} \). Thus, the Bayes estimator \( \delta_i(Y_i, \mu_i) = a_i^* Y_i + b_i^* \) is given by

\[
\delta_i(Y_i, \mu_i) = \tau^2 (\tau^2 + D_i)^{-1} Y_i + \mu_i D_i (D_i + \tau^2)^{-1}. \tag{2.7}
\]

Suppose now that \( \mu_i = \mu \) for all \( i \geq 1 \). Then, the BLUE of the common \( \mu \) is given by

\[
\hat{\mu} = \frac{\sum_{j=1}^{n} Y_j (D_j + \tau^2)^{-1}}{\sum_{j=1}^{n} (D_j + \tau^2)^{-1}}. \tag{2.8}
\]

Substituting \( \hat{\mu} \) for \( \mu \) in (2.7) yields the empirical Bayes estimator

\[
\hat{\delta}_i = \tau^2 (\tau^2 + D_i)^{-1} Y_i + D_i (\tau^2 + D_i)^{-1} \frac{\sum_{j=1}^{n} Y_j (D_j + \tau^2)^{-1}}{\sum_{j=1}^{n} (D_j + \tau^2)^{-1}}. \tag{2.9}
\]

Note that (2.9) is the same as the BLUP (2.3) derived in Corollary 2.2.

Suppose now that \( \mu_i = t_i \beta \), where \( t_i \) is some known constant, \( i = 1, \ldots, n \), and \( \beta \) is an unknown parameter. Then, the BLUE of \( \mu_i \) is given by

\[
\hat{\mu}_i = t_i \hat{\beta}, \tag{2.10}
\]

where

\[
\hat{\beta} = \frac{\sum_{j=1}^{n} t_j Y_j (D_j + \tau^2)^{-1}}{\sum_{j=1}^{n} t_j^2 (D_j + \tau^2)^{-1}}. \tag{2.11}
\]
is the weighted least squares estimator of $\beta$. Substituting $\hat{\mu}_i$ of (2.10) for $\mu_i$ in (2.7) yields the empirical Bayes estimator

$$\hat{\delta}_i = \tau^2 (\tau^2 + D_i)^{-1} Y_i + D_i (\tau^2 + D_i)^{-1} t_i \hat{\beta},$$

(2.12)

which is the same as the BLUP (2.4) given in Corollary 2.3.

Finally, suppose that $\mu_i = x'_i \beta$, where $x_i$ and $\beta$ are as defined in Corollary 2.4. Then, the BLUE of $\mu_i$ is given by

$$\hat{\mu}_i = x'_i \hat{\beta},$$

(2.13)

where

$$\hat{\beta} = \left[ \sum_{j=1}^n (D_j + \tau^2)^{-1} x'_j x_j \right]^{-1} \left[ \sum_{j=1}^n (D_j + \tau^2)^{-1} x'_j Y_j \right],$$

(2.14)

the weighted least squares estimator of $\beta$. Substituting $\hat{\mu}_i$ of (2.13) for $\mu_i$ in (2.7) yields the empirical Bayes estimator

$$\hat{\delta}_i = \tau^2 (\tau^2 + D_i)^{-1} Y_i + D_i (\tau^2 + D_i)^{-1} x'_i \hat{\beta} + y_i (Y_i - x'_i \hat{\beta}),$$

(2.15)

where $y_i = \tau^2 (D_i + \tau^2)^{-1}$. Again, observe that (2.15) is the same as the BLUP (2.5) derived in Corollary 2.4.

We now present a multivariate extension of Theorem 2.1 to the case that $Y_i$’s and $\theta_i$’s are real-valued random vectors. As a generalization of the univariate case, we assume that $\theta_1, \ldots, \theta_n$ are i.i.d. $p \times 1$ random vectors with a common distribution $G$ having a second moment. For given $\theta_i$, assume that $Y_i$ have a distribution $F_{\theta_i}$, which also has a finite second moment. For a given estimator $\delta(y)$, the mean squared error of $\delta(y)$ is defined as

$$R(\delta) = \text{tr} \left[ E(\delta(Y) - \theta)(\delta(Y) - \theta)' \right],$$

(2.16)

where $\text{tr}(\cdot)$ denotes the trace of the corresponding matrix and expectation $E$ is with respect to all the random variables involved.

**Theorem 2.5.** Let $(Y_1, \theta_1), \ldots, (Y_n, \theta_n)$ be independent random vector pairs satisfying the following:

(i) $\theta_1, \ldots, \theta_n$ are i.i.d. according to $G$ with $D(\theta_i) = \nabla < \infty$ for $i = 1, \ldots, n$;

(ii) $Y_i$, given $\theta_i$, is distributed according to $F_{\theta_i}$ with $D(Y_i | \theta_i) < \infty$ for $i = 1, \ldots, n$;

(iii) $E(Y_i | \theta_i) = \theta_i$ for $i = 1, \ldots, n$,

where $D(\cdot)$ stands for dispersion matrix (i.e., variance-covariance matrix) of the corresponding vector. Then, as an estimator of $\theta_i$, the mean squared error of the estimator $\hat{\delta}_i = \sum_{j=1}^n c_{ij} Y_j$, with $\sum_{j=1}^n c_{ij} = 1$ and $c_{ij} \geq 0$, as defined by (2.16) is given by

$$R(c_i, \hat{\delta}_i) = \sum_{j=1, j \neq i}^n c_{ij}^2 \text{tr} \left[ E(D(Y_j | \theta_i)) + \nabla \right] + c_{ii}^2 \text{tr} \left[ E(D(Y_i | \theta_i)) \right] + (1 - c_{ii})^2 \text{tr}(\nabla),$$

(2.17)

where $i = 1, \ldots, n$. Denote $a_j = \text{tr}[E(D(Y_j | \theta_i)) + \nabla], j = 1, \ldots, i - 1, i + 1, \ldots, n$; $a_i = \text{tr}[E(D(Y_i | \theta_i))]$ and $V = \text{tr}(\nabla)$. Then the value of $c_i = (c_{i1}, \ldots, c_{in})$ which minimizes
\[ R(c_i, \tilde{\delta}_i) \text{ is } c_i^* = (c_{i1}^*, \ldots, c_{in}^*) \text{ such that} \]
\[
c_{ij}^* = \frac{a_i}{a_j [1 + (a_i + V) \sum_{j=1, j \neq i}^n a_i^{-1}]} \quad j = 1, \ldots, i-1, i+1, \ldots, n; \]
\[
c_{ii}^* = \frac{1 + V \left( \sum_{j=1, j \neq i}^n a_j^{-1} \right)}{1 + (a_n + V) \sum_{j=1, j \neq i}^n a_j^{-1}}. \quad (2.18) \]

We now give three examples to illustrate BLUPs obtained using the empirical Bayes (EB) method described above.

**Example 2.6** (normal hierarchy). Consider estimation of \( \theta_i \) in the model \( X_i \mid \theta_i \sim N(\theta_i, \sigma^2) \) for \( i = 1, \ldots, n \), independent, and \( \theta_i \sim N(\mu, \tau^2) \) for \( i = 1, \ldots, n \), independent, where \( \mu \) is unknown. Then the linear Bayes estimator of \( \theta_i \) based on \( X_i \) is
\[
\delta_i(X_i, \mu) = \frac{\sigma^2}{\sigma^2 + \tau^2} \mu + \frac{\tau^2}{\sigma^2 + \tau^2} X_i. \quad (2.19) \]

The BLUE of \( \mu \) is \( \hat{\mu} = \bar{X} \). Now, replacing \( \mu \) by \( \bar{X} \) in the preceding expression yields the EB estimator
\[
\delta_i(X_i, \hat{\mu}) = \frac{\sigma^2}{\sigma^2 + \tau^2} \bar{X} + \frac{\tau^2}{\sigma^2 + \tau^2} X_i. \quad (2.20) \]

It is easy to show that the preceding EB estimator of \( \theta_i \) is also the BLUP (2.3) of \( \theta_i \) given in Corollary 2.2 under the present setup. It is interesting to note that this EB estimator (2.20) can also be obtained as a hierarchical Bayes estimator with an additional (improper) prior, \( \text{Uniform}(-\infty, \infty) \), on \( \mu \). If \( \tau^2 \) is unknown we can estimate \( (\sigma^2 + \tau^2)^{-1} \) by the unbiased estimator \( (n - 3) / \sum (X_i - \bar{X})^2 \) and obtain the EB estimator (see Lindley [12]; Efron and Morris [3, 4])
\[
\bar{X} + \left[ 1 - \frac{(n - 1)\sigma^2}{2 \sum (X_i - \bar{X})^2} \right] (X_i - \bar{X}), \quad (2.21) \]

which is no longer a best linear predictor; indeed, it is not even linear in \( X_i \)’s.

**Example 2.7** (Poisson hierarchy). Suppose that \( X_i \mid \theta_i \sim \text{Poisson}(\theta_i), i = 1, \ldots, n \), independent, and \( \theta_i \sim \text{Gamma}(\alpha, \beta), i = 1, \ldots, n \), independent. Then the linear Bayes estimator of \( \theta_i \) based on \( X_i \) is
\[
\delta_i(X_i, \mu) = \frac{\alpha \beta}{\beta + 1} + \frac{\beta}{\beta + 1} X_i, \quad (2.22) \]

where \( \mu = E(\theta_i) = \alpha \beta \). The BLUE of \( \mu \) is \( \hat{\mu} = \bar{X} \). Thus, replacing \( \alpha \beta \) in the equation for the Bayes estimator, we obtain the EB estimator
\[
\delta_i(X_i, \hat{\mu}) = \frac{1}{\beta + 1} \bar{X} + \frac{\beta}{\beta + 1} X_i, \quad (2.23) \]
which is also the BLUP (2.3) of \( \theta_i \) obtained in Corollary 2.2 under the present setup. Note that if \( \beta \) is also unknown, then there is no BLUP of \( \theta_i \).

**Example 2.8** (a regression model). Consider estimation of \( \theta_i \) in the following regression model: \( X_i \mid \theta_i \sim N(\alpha + \beta t_i, \tau^2) \) for \( i = 1, \ldots, n \), independent, and \( \theta_i \sim N(\alpha, \beta) \) for \( i = 1, \ldots, n \), where \( \alpha \) and \( \beta \) are unknown parameters. Then the linear Bayes estimator of \( \theta_i \) is

\[
\hat{\delta}_i(X_i, \mu_i) = \frac{\sigma^2}{\sigma^2 + \tau^2} \mu_i + \frac{\tau^2}{\sigma^2 + \tau^2} X_i,
\]

where \( \mu_i = \alpha + \beta t_i \).

Recall that the BLUEs of \( \alpha \) and \( \beta \) are the least squares estimators given by

\[
\hat{\alpha} = \bar{X} - \hat{\beta} \bar{T}, \quad \hat{\beta} = \frac{\sum (X_i - \bar{X})(t_i - \bar{T})}{\sum (t_i - \bar{T})^2},
\]

where \( \bar{T} = n^{-1} \sum t_i \). Therefore, the BLUE of \( \mu_i \) is \( \hat{\mu}_i = \hat{\alpha} + \hat{\beta} t_i \). Substituting \( \hat{\mu}_i \) for \( \mu_i \) in the Bayes estimator yields the EB estimator

\[
\hat{\delta}_i(X_i, \hat{\mu}_i) = \frac{\sigma^2}{\sigma^2 + \tau^2} (\hat{\alpha} + \hat{\beta} t_i) + \frac{\tau^2}{\sigma^2 + \tau^2} X_i.
\]

It is easy to see that the EB estimator (2.26) is also the BLUP (2.4) of \( \theta_i \) under the present setup. Again, the EB estimator (2.26) can also be obtained as a hierarchical Bayes predictor, by appending the specification \( (\alpha, \beta) \sim \text{Uniform}(\cdot, \cdot) \times (\cdot, \cdot) \) to the hierarchy in the present example. If \( \tau^2 \) is unknown, we can use the fact that \( E[\sum (X_i - \hat{\alpha} + \hat{\beta} t_i)^2]^{-1} = (n - 4)(\sigma^2 + \tau^2) \) to construct the EB estimator

\[
\hat{\alpha} + \hat{\beta} t_i + \left[ 1 - \frac{(n - 4)\sigma^2}{\sum (X_i - \hat{\alpha} + \hat{\beta} t_i)^2} \right] (X_i - \hat{\alpha} - \hat{\beta} t_i),
\]

which is, again, neither linear nor unbiased for \( \theta_i \).

**Remark 2.9.** Generally speaking, an empirical Bayes estimator can be thought of as a two-stage estimator. Specifically, consider the Bayes model \( Y_i \mid \theta \sim f(y \mid \theta) \), \( i = 1, \ldots, n \) and \( \theta \mid \mu \sim \pi(\theta \mid \mu) \), where \( E(\theta) = \mu \). We, first, obtain an estimate of \( \mu, \hat{\mu}(y) \), based on the marginal distribution of \( Y = (Y_1, \ldots, Y_n) \) with density \( f(y \mid \mu) = \int [\prod_{i=1}^n f(y_i \mid \theta)]\pi(\theta \mid \mu) d\theta \). Then, we substitute \( \hat{\mu}(y) \) for \( \mu \) in \( \pi(\theta \mid \mu) \) and determine the estimator that minimizes the empirical posterior loss \( L(\theta, \hat{\delta}(y))\pi(\theta \mid y, \hat{\mu}(y)) d\theta \). This minimizing estimator is the empirical Bayes estimator. This argument is mathematically equivalent to first obtaining the Bayes estimator \( \delta(Y, \mu) \) and then substituting \( \mu \) by \( \hat{\mu}(y) \) (see, e.g., Lehmann and Casella [11, Chapter 4]).

In the present context, our proposed empirical Bayes estimators can be obtained by minimizing the posterior loss \( L(\theta - a_i - b_i Y_i)^2 \pi(\theta \mid y, \hat{\mu}(y)) d\theta \). Our empirical Bayes argument is somewhat equivalent to the two-stage approach suggested by Bulmer [1, pages 208–209] for the prediction problem of \( v \) in the mixed model of \( Y = X\beta + Z\mu + e \): first form a vector of the data \( Y \) corrected for the fixed effects by \( Y_c = Y - X\hat{\beta} \), where \( \hat{\beta} \) is the BLUE of \( \beta \), and then, under normality assumptions, \( v \) is predicted by \( E(v \mid Y_c) \). Note that our argument, however, make no reference to normality assumptions and can be applied for more general models.
3. Proofs. In this section, we provide proofs of the results presented in Section 2.

**Proof of Theorem 2.1.** Write

\[
 r_i(c) = E \left[ \sum_{j=1}^{n} c_{ij} X_j - \theta_i \right]^2
 = E \left[ \sum_{j=1}^{n} c_{ij} (X_j - \theta_j) \right]^2 + \sum_{j=1}^{n} c_{ij} \theta_j - \theta_i \right]^2 + CPT,
\]

where

\[
 CPT = 2E \left[ \sum_{j=1}^{n} c_{ij} (X_j - \theta_j) \right] \left[ \sum_{j=1}^{n} c_{ij} \theta_j - \theta_i \right]
 = 2E \left[ \sum_{j=1}^{n} c_{ij} (X_j - \theta_j) \right] \left[ \sum_{j=1}^{n} c_{ij} \theta_j - \theta_i \right] = 0.
\]

The first term of the right-hand side of (3.1) can be evaluated as

\[
 E \left[ \sum_{j=1}^{n} c_{ij} (X_j - \theta_j) \right]^2 = \sum_{j=1}^{n} c_{ij}^2 E \left( (X_j - \theta_j)^2 \right) + \sum_{j=k}^{n} c_{ij} c_{ik} E \left( (X_j - \theta_j)(X_k - \theta_k) \right)
 = \sum_{j=1}^{n} c_{ij}^2 E \left( (X_j - \theta_j)^2 \right)
 = \sum_{j=1}^{n} c_{ij}^2 V \left( X_j - \theta_j \right) \quad (3.3)
 = \sum_{j=1}^{n} c_{ij}^2 [E \left( (X_j - \theta_j)/\theta_j \right) + 2E \left( (X_j - \theta_j)/\theta_j \right)]
 = \sum_{j=1}^{n} c_{ij}^2 D_j.
\]

The second term on the right-hand side of (3.1) is equal to

\[
 E \left[ \sum_{j=1}^{n} c_{ij} \theta_j - \theta_i \right]^2 = E \left[ \sum_{j=1}^{n} c_{ij} \theta_j \right]^2 + E \left( \theta_i^2 \right) - 2E \left( \theta_i \sum_{j=1}^{n} c_{ij} \theta_j \right)
 = E \left[ \sum_{j=1}^{n} c_{ij} \theta_j \right]^2 + E \left( \theta_i^2 \right) - 2E \left( \theta_i \sum_{j=1}^{n} c_{ij} \theta_j \right).
\]
Using the independence of $\theta_i$'s and by the restriction $\sum_{j=1}^{n} c_{ij} \mu_j = \mu_i$, we have

\[
E \left( \theta_i \sum_{j=1}^{n} c_{ij} \theta_j \right) = \sum_{j=1}^{n} c_{ij} E(\theta_j \theta_i) = \sum_{j=1}^{n} c_{ij} E(\theta_j \theta_i) + c_{ii} E(\theta_i^2) = \sum_{j=1}^{n} c_{ij} \mu_i + c_{ii} (V(\theta_i) + \mu_i^2) = (\mu_i - c_{ii} \mu_i) \mu_i + c_{ii} (\tau^2 + \mu_i^2) = \mu_i^2 + c_{ii} \tau^2, \tag{3.5}
\]

\[
E \left( \sum_{j=1}^{n} c_{ij} \theta \right)^2 = V \left( \sum_{j=1}^{n} c_{ij} \theta_j \right) + \left( E \left( \sum_{j=1}^{n} c_{ij} \theta_j \right) \right)^2 = \sum_{j=1}^{n} c_{ij}^2 \tau^2 + \left( \sum_{j=1}^{n} c_{ij} \mu_j \right)^2 \tag{3.6}
\]

Now combining (3.4) to (3.6), we obtain

\[
E \left[ \sum_{j=1}^{n} c_{ij} \theta_i \theta_j - \theta_i \theta_j \right] = \tau^2 \sum_{j=1}^{n} c_{ij}^2 + \mu_i^2 + (\tau^2 + \mu_i^2) - 2(\mu_i^2 + c_{ii} \tau^2) = \tau^2 \sum_{j=1}^{n} c_{ij}^2 + \tau^2 - 2c_{ii} \tau^2. \tag{3.7}
\]

The proof of (2.1) is now completed by combining (3.1), (3.3), and (3.7).

In order to find the values of $c_{ij}$ that minimize $r_i(c)$ subject to the restriction $\sum_{j=1}^{n} c_{ij} \mu_j = \mu_i$, we use the Lagrange multiplier method. Write

\[
r_i(c) = \sum_{j=1}^{n} c_{ij}^2 D_j + \tau^2 \left( \sum_{j=1}^{n} c_{ij}^2 + 1 - 2c_{ii} \right) + \rho \left( \mu_i - \sum_{j=1}^{n} c_{ij} \mu_j \right), \tag{3.8}
\]

and we find solutions to the equations $\partial r_i / \partial c_{ij} = 0$ subject to the condition $\sum_{j=1}^{n} c_{ij} \mu_j = \mu_i$. For $i \neq j$ and $j = 1, \ldots, n$,

\[
\frac{\partial r_i}{\partial c_{ij}} = 2D_j c_{ij} + 2\tau^2 c_{ij} - \rho \mu_j, \\
\frac{\partial r_i}{\partial c_{ii}} = 2D_i c_{ii} + 2\tau^2 c_{ii} - 2\tau^2 - \rho \mu_i. \tag{3.9}
\]

Now $\partial r_i / \partial c_{ij} = 0$ implies for $j \neq i$ and $j = 1, \ldots, n$,

\[
c_{ij}^* = \frac{\rho \mu_j}{2D_j + 2\tau^2}, \tag{3.10}
\]
and \( \partial r_i/\partial c_{ii} = 0 \) implies
\[
c_{ii}^* = \frac{2\tau^2 + \rho \mu_i}{2D_i + 2\tau^2}.
\] (3.11)

Using these facts together with \( \sum_{j=1}^n c_{ij}\mu_j = \mu_i \) implies that
\[
\rho = \frac{\mu_i D_i (D_i + \tau^2)^{-1}}{\sum_{j=1}^n \mu_j^2 [2(D_j + \tau^2)]^{-1}}.
\] (3.12)

The proof is now completed by combining (3.10), (3.11), and (3.12).

**Proof of Corollary 2.2.** When \( \mu_j = \mu \) for all \( j \geq 1 \), the restriction \( \sum_{j=1}^n c_{ij}\mu_j = \mu_i \) reduces to \( \sum_{j=1}^n c_{ij} = 1 \) for all \( i \geq 1 \). We now minimize
\[
r_i(c) = \sum_{j=1}^n c_{ij}^2 D_j + \tau^2 \left( \sum_{j=1}^n c_{ij}^2 + 1 - 2c_{ii} \right) + \rho_0 \left( t_i - \sum_{j=1}^n c_{ij} t_j \right),
\] (3.13)
with respect to \( c_{ij} \) subject to \( \sum_{j=1}^n c_{ij} = 1 \). The solutions of \( \partial r_i/\partial c_{ij} = 0 \) are
\[
c_{ij}^* = \frac{\rho_0}{2D_j + 2\tau^2}, \quad \text{for} \ j \neq i; \quad c_{ii}^* = \frac{2\tau^2 + \rho_0}{2D_i + 2\tau^2},
\] (3.14)
where
\[
\rho_0 = \frac{D_i (D_i + \tau^2)^{-1}}{\sum_{j=1}^n [2(D_j + \tau^2)]^{-1}}.
\] (3.15)

Now, by rearranging the terms in \( \sum c_{ij}^* Y_j \) with the above choices of the \( c_{ij}^* \) follows the desired result.

**Proof of Corollary 2.3.** We minimize
\[
r_i(c) = \sum_{j=1}^n c_{ij}^2 D_j + \tau^2 \left( \sum_{j=1}^n c_{ij}^2 + 1 - 2c_{ii} \right) + \rho_1 \left( t_i - \sum_{j=1}^n c_{ij} t_j \right),
\] (3.16)
with respect to \( c_{ij} \) subject to \( \sum_{j=1}^n c_{ij} t_j = t_i \). The rest of the proof is now similar to the proof of Theorem 2.1.

**Proof of Corollary 2.4.** Similar to the proof of Corollary 2.3.
**Proof of Theorem 2.5.** First, we write

\[ R(c_i, \hat{\theta}_i) = \text{tr} \left[ E \left( \left( \sum_{j=1}^{n} c_{ij} (Y_j - \theta_i) \right) \left( \sum_{j=1}^{n} c_{ij} (Y_j - \theta_i) \right)' \right) \right] \]

\[ = \sum_{j=1}^{n} c_{ij}^2 \text{tr} \left[ E( (Y_j - \theta_i) (Y_j - \theta_i)' ) \right] + \sum_{j=k}^{n} c_{ij} c_{ik} \text{tr} \left[ E( (Y_j - \theta_i) (Y_k - \theta_i)' ) \right]. \]

(3.17)

Note that

\[ E((Y_j - \theta_i)(Y_j - \theta_i)') = E((Y_j - \theta_j)(Y_j - \theta_j)') + E((\theta_j - \theta_i)(\theta_j - \theta_i)') \]

\[ + E((\theta_j - \theta_i)(Y_j - \theta_i)') + E((Y_j - \theta_i)(\theta_j - \theta_i)') \]

\[ = E(D(Y_j | \theta_j)) + D(\theta_j - \theta_i) \]

\[ = \begin{cases} E(D(Y_j | \theta_j)) + 2\nabla, & \text{for } j \neq i, \\ E(D(Y_i | \theta_i)), & \text{for } j = i. \end{cases} \]

(3.18)

Also, for \( 1 \leq j \neq k \leq n - 1 \), we have

\[ E((Y_j - \theta_i)(Y_k - \theta_i)') = E[E((Y_j - \theta_i)(Y_k - \theta_i)' | \theta_j, \theta_k)] \]

\[ = E((\theta_j - \theta_i)(\theta_k - \theta_i)') \]

\[ = E[E((\theta_j - \theta_i)(\theta_k - \theta_i)' | \theta_i)] \]

\[ = D(\theta_i) \]

\[ = \nabla. \]

(3.19)

For all other cases, that is, for \( j \neq k \) and either \( j = n \) or \( k = n \), we have

\[ E((Y_j - \theta_j)(\theta_j - \theta_i)') = E[E((Y_j - \theta_j)(\theta_j - \theta_i)' | \theta_j, \theta_i)] \]

(3.20)

\[ = 0. \]

Now, combining (3.18), (3.19), and (3.20) and then substituting in (3.17), we obtain after some simplification \( R(c_i, \hat{\theta}_i) = \sum_{j=1}^{n} c_{ij}^2 \text{tr} [E(D(Y_j | \theta_j)) + \nabla] + c_{ii}^2 \text{tr} [E(D(Y_i | \theta_i))] + (1 - c_{ii})^2 \text{tr}(\nabla). \) This proves (2.17). The derivation of (2.18) easily follows from (2.17) subject to the restriction that \( \sum_{j=1}^{n} c_{ij} = 1 \). This completes the proof of Theorem 2.5.

\[ \square \]

**References**


Rohana J. Karunamuni: Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

E-mail address: rkarunam@ualberta.ca
Submit your manuscripts at http://www.hindawi.com