ON THE ESTIMATION OF MEAN SQUARE ERROR OF SMALL AREA PREDICTORS

N.G.N. Prasad and J.N.K. Rao

University of Alberta and Carleton University

1. Introduction

Small area estimation has received considerable attention in recent years due to growing demand for reliable small area statistics. The usual survey estimates, based only on the data from a given small area (domain), are likely to be unreliable due to smallness of sample size in the domain. Therefore, alternative estimators which "borrow strength" from other areas have been proposed in the literature to improve the efficiency. These estimators use models, either explicitly or implicitly, that "connect" the small areas through supplementary data (e.g., census and administrative data). Simple synthetic estimators, for example, are based on implicit modelling.

In this paper, three small area models, due to Battese and Fuller (1982), Dempster et al. (1981) and Fay and Herriot (1979) respectively, are investigated. The best linear unbiased predictor (BLUP) under each model is obtained, using the general theory of Henderson (1975) for a mixed linear model. A weighted jackknife estimator of BLUP is also derived. Second order approximations to the mean square error (MSE) of estimated BLUP and the estimate of MSE are obtained, under normality. Robust estimates of the MSE approximation are also derived, using the weighted jackknife method. Finally, the results of a Monte Carlo study, on the efficiency of estimated BLUPs and the accuracy of the proposed approximations to MSE and its estimates, are reported.

2. Three Models

2.1. Nested error regression model

Battese and Fuller (1982) proposed a nested error regression model in the context of estimating (or predicting) mean hectares under corn for 12 counties (small areas) in north-central Iowa, using Landsat Satellite data in conjunction with survey data. The scatter plot of y (hectare of corn in a segment) against x (number of sampled units observed in the i-th small area, is large so that Ei = 0 noting that E(eij) = 0. Thus, the problem is to estimate the small area mean

\[
\mu_i = \bar{x}_i \beta + v_i, \quad i = 1, ..., t. \tag{2.2}
\]

Note that the \(\mu_i\)'s are random variables, as in the prediction approach to survey sampling (Royall, 1970).

2.2. Random regression coefficients model

A more general model with random slopes was proposed by Dempster et al. (1981). Their random regression coefficients model, in the context of small area estimation, may be written as

\[
Y_{ij} = x_{ij}' \beta_i + e_{ij}, \tag{2.3}
\]

where \(Y_{ij}, x_{ij}, \beta_i\) and \(e_{ij}\) are as defined in the model (2.1), and \(v_i = (v_{i1}, ..., v_{ik})'\) are independent of the \(e_{ij}\)'s and independently distributed with mean vector 0 and covariance matrix \(\psi\), say. The i-th small area population mean is given by

\[
\mu_i = \bar{x}_i \beta_i + \bar{v}_i. \tag{2.4}
\]

In this paper, we confine ourselves to the special case \(k=1\), i.e., one concomitant variable and regression through origin. Hence, this special case does not cover the Fuller-Battese model.

2.3. Fay-Herriot model

In the context of estimating per capita income for small areas (population less than 1000), Fay and Herriot (1979) assumed that a k-vector of benchmark variables \(x_i = (x_{i1}, ..., x_{ik})\), related to the small area mean \(\mu_i\), is available for each small area i, and that the \(\mu_i\) are independent \(N(x_i \beta, A)\), where \(\beta\) is a k-vector of parameters. They further assume that the sample mean vector \(\bar{y} = (\bar{y}_1, ..., \bar{y}_k)'\) given \(\mu = (\mu_1, ..., \mu_k)'\) is normally distributed with mean \(1_{1\times k} \mu\) and known covariance matrix \(\text{diag}(D_1, ..., D_k)\).

The Fay-Herriot model can be restated as a linear model:

\[
\bar{y}_i = u_i + e_i, \quad u_i = x_i \beta + v_i \tag{2.5}
\]

where the \(e_i\)'s and the \(v_i\)'s are independent and \(N(0,D_i)\) and \(N(0,A)\) respectively. The normality assumption is not needed in deriving the BLUP. Note also that the auxiliary information at the random intercepts.

The i-th small area population mean may be written as \(\mu_i = x_i \beta + v_i\), where \(x_i\) and \(v_i\) are the population means of \(x_{ij}\) and \(e_{ij}\) for the i-th small area. We assume that \(x_i\) is known, e.g., the mean number of pixels from Satellite data. Also, we assume that \(N_i\) the number of population units in the i-th small area, is large so that \(E_i = 0\) noting that \(E(e_{ij}) = 0\). Thus, the problem is to estimate the small area means.
3. Best Linear Unbiased Predictors

3.1. General mixed model

The models in Section 2 are special cases of a general mixed linear model

\[ y = X\beta + Zv + e \]  

where X and Z are known matrices and v and e are mutually independent random vectors with zero means and covariance matrices G and R respectively, depending on some parameters \( \theta \) called variance components. Henderson (1975) has shown that the BLUP of \( y = k'\beta + m'v \) is given by

\[ t(\theta, y) = k'(\beta) + m'GZ'(V^{-1} - X') \]  

where \( V = R + GZ' \) and \( \beta = (X'V^{-1}X)^{-1}(X'V^{-1}y) \). He has also given a method for evaluating \( \beta \) without actually inverting \( V \), but \( V^{-1} \) can be explicitly obtained for all the three models using the following matrix lemma:

**Lemma 3.1.** (Graybill (1969); Theorem 8.3.3).

Let \( C = D + ab' \), where \( D = \text{diag}(d_{ij}) \), and \( a \) and \( b \) are the \( i \)-th elements of \( a \) and \( b \) respectively. Then \( C^{-1} = D^{-1} + ab' \), where \( \gamma = (a + ab'd_{ij})^{-1} \), \( a' = a'd_{ii} \) and \( b' = b'd_{ii} \).

3.2. Special cases

Writing the nested error regression model (2.1) in terms of the general mixed model (3.1), we have

\[ y = \text{col} \text{col} (y_{ij}), \quad x = \text{col} \text{col} (x_{ij}) \]  

\[ z = \text{diag} \text{col} (1), \quad v = \text{col} (v_{i}), \]  

\[ e = \text{col} \text{col} (e_{ij}), \]  

and \( G = \sigma_v^2 I, \quad R = \sigma_e^2 I \). If \( \theta = (\sigma_v^2, \sigma_e^2) \) is the BLUP \( t(\theta, y) = t(\theta) \) depends on the variance components \( \theta \), but in practice the components of \( \theta \) will be unknown. Actually, the BLUP depends only on the ratios \( \theta_i/\theta_m \), where \( \theta = (\theta_1, \ldots, \theta_m) \), e.g., the BLUP (3.4) depends only on \( \sigma_v^2/\sigma_e^2 \). It is customary to estimate the BLUP by replacing \( \theta \) by an asymptotically consistent estimator \( \hat{\theta} \). The resulting estimator, \( t(\hat{\theta}, y) = t(\hat{\theta}) \), of \( y \) will remain unbiased provided \( \hat{\theta} \) is even and translation invariant, i.e., \( t(\theta+y) = t(\theta) \), and the distributions of \( v \) and \( e \) are both symmetric (not necessarily normal); see Kackar and Harville (1984). However, the MSE of \( t(\theta) \) will be increased relative to \( MSE(t(\theta)) \), as shown in Section 4.

Various methods of estimating \( \theta \) for a general mixed model are available (see Harville's (1977) review paper), but in this paper we confine ourselves to the well-known method of fitting of constants, also called Henderson's method 3. We now spell out the expressions for \( \sigma_v^2 \) and \( \sigma_e^2 \) in the case of Battese-Fuller model. Let \( \hat{\sigma}_{ij}^2 \) be the estimated residuals obtained from the ordinary least squares regression of \( y_{ij} - \bar{y_i} \) on \( \{x_{ij1} - \bar{x_i1}, \ldots, x_{ijk} - \bar{x_ik}\} \). Similarly, a second set of residuals \( \hat{u}_{ij} \) are obtained by performing ordinary least squares regression of

\[ t_1(\sigma_v^2, y) = \bar{y}_i - \bar{y}_{\hat{u}_{ij}} \]  

where \( \bar{y}_{\hat{u}_{ij}} \) is the sample mean of \( x_{ij} \) for the \( i \)-th small area. Battese and Fuller (1982) also obtained the predictor (3.3) using heuristic arguments, but they have not shown that it is in fact a BLUP.

Similar calculations for the random regression coefficients model (2.3) with \( k=1 \) lead to the BLUP of \( \mu_i \) given by

\[ t_1(\sigma_v^2, y) = \bar{x}_i - \bar{y}_{\hat{u}_{ij}} \]  

where \( \sigma_v^2 = \sigma_v^2/\sigma_e^2 \),

\[ \sigma_i^2 = \sigma_v^2/\sigma_e^2 (\sum x_{ij}^2)^{-1} \]  

\[ \sigma_v^2 = \text{var}(v_i) \]  

and \( \hat{\beta} = (\sum y_i(\sum x_{ij}^2)^{-1})(\sum y_i)^{-1} \).

For the Fay-Herriot model (2.5), the BLUP of \( \mu_i \) is obtained as

\[ t_1(A_i, y) = x_i - A_i D_i^{-1}(\bar{y}_i - \bar{x}_i) \]  

where \( \bar{x}_i \) is the sample mean of \( x_i \) for the \( i \)-th small area. Battese and Fuller (1982) also obtained the predictor (3.3) using heuristic arguments, but they have not shown that it is in fact a BLUP.

Similar calculations for the random regression coefficients model (2.3) with \( k=1 \) lead to the BLUP of \( \mu_i \) given by

\[ t_1(\sigma_v^2, y) = \bar{y}_i - \bar{y}_{\hat{u}_{ij}} \]  

where \( \sigma_v^2 = \sigma_v^2/\sigma_e^2 \),

\[ \sigma_i^2 = \sigma_v^2/\sigma_e^2 (\sum x_{ij}^2)^{-1} \]  

\[ \sigma_v^2 = \text{var}(v_i) \]  

and \( \hat{\beta} = (\sum y_i(\sum x_{ij}^2)^{-1})(\sum y_i)^{-1} \).

For the Fay-Herriot model (2.5), the BLUP of \( \mu_i \) is obtained as

\[ t_1(A_i, y) = x_i - A_i D_i^{-1}(\bar{y}_i - \bar{x}_i) \]  

where \( \bar{x}_i \) is the sample mean of \( x_i \) for the \( i \)-th small area. Battese and Fuller (1982) also obtained the predictor (3.3) using heuristic arguments, but they have not shown that it is in fact a BLUP.

Similar calculations for the random regression coefficients model (2.3) with \( k=1 \) lead to the BLUP of \( \mu_i \) given by

\[ t_1(\sigma_v^2, y) = \bar{y}_i - \bar{y}_{\hat{u}_{ij}} \]  

where \( \sigma_v^2 = \sigma_v^2/\sigma_e^2 \),

\[ \sigma_i^2 = \sigma_v^2/\sigma_e^2 (\sum x_{ij}^2)^{-1} \]  

\[ \sigma_v^2 = \text{var}(v_i) \]  

and \( \hat{\beta} = (\sum y_i(\sum x_{ij}^2)^{-1})(\sum y_i)^{-1} \).

For the Fay-Herriot model (2.5), the BLUP of \( \mu_i \) is obtained as

\[ t_1(A_i, y) = x_i - A_i D_i^{-1}(\bar{y}_i - \bar{x}_i) \]  

where \( \bar{x}_i \) is the sample mean of \( x_i \) for the \( i \)-th small area. Battese and Fuller (1982) also obtained the predictor (3.3) using heuristic arguments, but they have not shown that it is in fact a BLUP.
\( y_{ij} \) on \( \{x_{ij1}, \ldots, x_{ijk}\} \). Then, unbiased quadratic estimators of \( \sigma_y^2 \) and \( \sigma^2 \) are obtained as
\[
\sigma_y^2 = (n-t-k+\lambda)^{-1} \sum_i \hat{e}_{ij}^2 \quad (3.6)
\]
and
\[
\sigma^2 = n^{-1} \sum_i \hat{e}_{ij}^2 - (n-k) \sigma_y^2 \quad (3.7)
\]
where
\[
(t-\lambda)n_* = n - t \text{tr}[(X'X)^{-1} t \sum_{i=1}^t n_{ij} x_{ij1} x_{ij1}']
\]
and \( \lambda = 0 \) if the model (2.1) has no intercept term and \( \lambda = 1 \) otherwise.

For the random regression coefficients model (2.3) with one auxiliary variable \( (k = i) \) we obtain
\[
\sigma_y^2 = (n_t)^{-1} \sum_i \hat{e}_{ij}^2 \quad (3.8)
\]
\[
\sigma^2 = n^{-1} \sum_i \hat{e}_{ij}^2 - (n-1) \sigma_y^2 \quad (3.9)
\]
where
\[
\hat{e}_{ij} = y_{ij} - x_{ij} (\bar{x}_i x_{ij1}) (\Sigma x_{ij1}^2)^{-1}
\]
and
\[
\hat{n}_* = \Sigma x_{ij1}^2 - \{ \Sigma (\Sigma x_{ij1}^2) \} (\Sigma x_{ij1}^2)^{-1}.
\]

An unbiased quadratic estimator of \( \lambda \) in the Fay-Herriot model can be obtained as
\[
\hat{\lambda} = (t-k)^{-1} \sum_i \hat{\lambda}_i, \quad \hat{\lambda}_i = \hat{n}_i (n_i - 1) - B_2 (\Sigma x_{ij1} - \bar{x}_i)^2 - B_1 (\Sigma x_{ij1} - \bar{x}_i)^2
\]
where
\[
\bar{x}_i = \Sigma x_{ij1}/n.
\]

It is possible for \( \sigma_y^2 \), defined by (3.7) or (3.9), or \( \hat{\lambda} \), given by (3.10), to be negative. In practice if \( \hat{\lambda} \) or \( \sigma_y^2 \) is negative, we set it equal to zero. Fay and Herriot (1979) obtained an estimator \( \lambda \) as a solution of the nonlinear equation
\[
\sum_i \hat{\lambda}_i = 0 \quad (3.11)
\]
where
\[
\hat{\lambda}_i = \hat{y}_i - \bar{y}_i, \quad \hat{\beta} = (X'X)^{-1} (X'y - \bar{y}).
\]

The estimator \( \hat{\lambda} \) remains unchanged, but for notational consistency we denote it as \( \sigma^2 \).

3.4. Jackknife estimator of BLUP

For the three small area models, alternative estimators of BLUP can be obtained by the jackknife method. For simplicity, we confine ourselves to the nested error regression model with one auxiliary variable \( \{i.e., k = 2, x_{ij1} = 1\} \), \( x_{ij2} = x_{ij1} \), \( \beta_1 = \alpha \), \( \beta_2 = \beta \). The estimator is obtained by replacing \( \hat{y}_i \) in \( t_i(\sigma^2, \gamma) \) by a jackknife estimator \( \hat{y}_i(J) \) which is approximately unbiased, i.e., \( E(\hat{y}_i(J) - \gamma_i) = o(t^{-1}) \) while \( E(\hat{y}_i(y_i - \gamma_i) = o(t^{-1}) \) for large \( t \) and bounded \( n_i \).

Fuller and Harter (1966) have also obtained an alternative estimator of \( \hat{y}_i \), but its approximate unbiasedness depends on normality of the errors \( \{\varepsilon_{ij}\} \) and \( \{\epsilon_{ij}\} \) unlike \( \hat{y}_i(J) \). For the Fay-Herriot model, Morris (1983) used \( 1 - [t/(t-k-1)]D_1/A + D_1 \) as an estimator of \( \hat{y}_i = \alpha + \beta_1 + \beta_2 \), using that in the equal variance case, \( D_1 = D \), it is exactly unbiased under normality.

We modify the estimator \( \sigma^2 \) slightly in order to construct weighted pseudo-values, similar to Hinkley’s (1977) for the standard regression model. The modified estimator \( \sigma^2 \) is obtained along the lines of Arvesen (1969) for the ANOVA model
\[
\hat{y}_i = \mu + v_i + e_i
\]
where
\[
\hat{y}_i = y_{ij} - x_{ij} \hat{\beta} = (X'X)^{-1} x_{ij} (X'y - \bar{y}).
\]

The estimator \( \sigma^2 \) remains unchanged, but for notational consistency we denote it as \( \sigma^2 \).

Our jackknife procedure is essentially based on deleting the residuals \( \{e_{ij}, \epsilon_{ij}, j = 1, \ldots, n_i\} \) in turn for \( i = 1, \ldots, t \) and then computing the estimates of \( \sigma^2 \) and \( \sigma^2 \) as above:
\[
\sigma^2(-i) = (n-t-\lambda_i)^{-1} \sum_i \hat{e}_{ij}^2 - \hat{\lambda}_i \quad (3.12)
\]
\[
\sigma^2(-i) = n_1^{-1} \sum_i \hat{e}_{ij}^2 - \hat{\lambda}_i \quad (3.13)
\]
The estimators $\hat{\sigma}_e^{-2}$ and $\hat{\sigma}_v^{-2}$ are both unbiased, and $\hat{\sigma}_e^{-2}$ and $\hat{\sigma}_v^{-2}$ both can be expressed as weighted sums of $\hat{\sigma}_e^{-2}$ and $\hat{\sigma}_v^{-2}$ respectively, similar to Wu's (1986) representation for the standard regression model:

$$\hat{\sigma}_e^{-2} = (t-1)^{-1} \sum_{i=1}^t \hat{\sigma}_e^{-2} (t-i)$$

and

$$\hat{\sigma}_v^{-2} = (t-1)^{-1} \sum_{i=1}^t \hat{\sigma}_v^{-2} (t-i),$$

where

$$a_i = 1/(n-1)^{-1}, b_i = 1-n^{-1}.$$

For a general function, $g(\hat{\sigma}_e^{-2}, \hat{\sigma}_v^{-2})$, of $\hat{\sigma}_e^{-2}$ and $\hat{\sigma}_v^{-2}$, our jackknife estimator is given by

$$g_{jk}(\hat{\sigma}_e^{-2}, \hat{\sigma}_v^{-2}) = t^{-1} \sum_{j=1}^t g_j,$$

where

$$g_j(\hat{\sigma}_e^{-2}, \hat{\sigma}_v^{-2}) = g(\hat{\sigma}_e^{-2} + a_i \hat{\sigma}_e^{-2} (t-i), \hat{\sigma}_v^{-2} - b_i \hat{\sigma}_v^{-2} (t-i)), $$

are the weighted pseudo-values. In the linear case, $g(\hat{\sigma}_e^{-2}, \hat{\sigma}_v^{-2}) = \hat{\sigma}_e^{-2} + \hat{\sigma}_v^{-2}$ for some known constants $\hat{\sigma}_e$ and $\hat{\sigma}_v$, the jackknife estimator reduces to the unbiased estimator $\hat{\sigma}_e^{-2} + \hat{\sigma}_v^{-2}$, while it is approximately unbiased in the nonlinear case (see Prasad, 1985). The jackknife estimator $\hat{\gamma}_{1j}$ is a special case of (3.16) with $g(\hat{\sigma}_e^{-2}, \hat{\sigma}_v^{-2}) = \hat{\gamma}_j = \hat{\sigma}_e^{-2} + \hat{\sigma}_v^{-2} (1-n^{-1})^{-1}$.

The estimated BLUP, $t_{1j}(\hat{\sigma}_e^{-2}, \vec{y})$, is given by (3.3) if $\hat{\gamma}_{1j}$ and $\hat{\beta} = \hat{\beta}(\hat{\sigma}_e^{-2})$ are replaced by $\hat{\gamma}_{1j}$ and $\hat{\beta}(\hat{\sigma}_e^{-2})$ respectively.

4. Second Order Approximation to MSE

Kackar and Harville (1984) have shown that

$$\text{MSE}[t(\hat{\theta})] = \text{MSE}[t(\theta)] + E[t(\hat{\theta})-t(\theta)]^2$$

(4.1)

under normality, provided $\hat{\theta}$ is translation invariant, i.e., $\hat{\theta}(y+bx) = \hat{\theta}(y)$ for all $y$ and $b$. That is, the use of $\text{MSE}[t(\theta)]$ leads to underestimation of actual MSE by an amount $E[t(\hat{\theta})-t(\theta)]^2$. Henderson (1975) has given an exact expression for $\text{MSE}[t(\theta)]$, but the second term of (4.1) is in general not tractable, except in special cases, e.g., Peixoto (1982) obtained $\text{MSE}[t(\hat{\theta})]$ for the one-way, balanced ANOVA model, $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$ with $n_i = r$. Kackar and Harville (1984) obtained a Taylor approximation

$$E[t(\hat{\theta})-t(\theta)]^2 = E[d(\theta)'](\hat{\theta}-\theta)^2$$

(4.2)

where $d(\theta) = \partial t(\theta)/\partial \theta = d(y, \theta)$. Using (4.2), they proposed a further approximation

$$E[d(\theta)'](\hat{\theta}-\theta)^2 \approx \text{tr}[A(\theta)E(\hat{\theta}-\theta)(\hat{\theta}-\theta)'].$$

(4.3)

where $A(\theta)$ is the covariance matrix of $d(\theta)$. General conditions are given in Prasad (1985) under which the precise order of neglected terms in the approximation (4.2) and (4.3) is $O(t^{-1})$ for large $t$. The three small area models satisfy these conditions.

For the Battese-Fuller model, $\text{MSE}[t_{1j}(\hat{\sigma}_e^{-2}, \vec{y})]$ under arbitrary distributions of $(\epsilon_{ij})$ and $(\epsilon_{ij})$ is obtained as

$$\text{MSE}[t_{1j}(\hat{\sigma}_e^{-2}, \vec{y})] = (1-\hat{\gamma}_1)\hat{\sigma}_e^{-2}$$

(4.4)

and

$$\text{MSE}[t_{1j}(\hat{\sigma}_v^{-2}, \vec{y})] = (1-\hat{\gamma}_1)\hat{\sigma}_v^{-2}$$

(4.5)

The second terms on r.h.s. of (4.5) and (4.6) are both of order $O(t^{-1})$ for large $t$. Under normality, the approximation to $E[t(\hat{\beta})-t(\theta)]^2$ for the Battese-Fuller model reduces to

$$E[t_{1j}(\hat{\sigma}_e^{-2}, \vec{y})-t_{1j}(\theta, \vec{y})]^2$$

(4.7)

Similarly, for the random regression coefficient model

$$E[t_{1j}(\hat{\sigma}_v^{-2}, \vec{y})-t_{1j}(\theta, \vec{y})]^2$$

(4.8)

In the case of Fay-Herriot model, we get

$$E[t_{1j}(\hat{\sigma}_v^{-2}, \vec{y})-t_{1j}(\theta, \vec{y})]^2$$

(4.9)

Ignoring the uncertainty in $\hat{\sigma}_e^{-2}$ and $\hat{\beta}$ and using $\text{MSE}[t_{1j}(\sigma_e^{-2}, \vec{y})]$ and $\text{MSE}[t_{1j}(\hat{\sigma}_e, \vec{y})]$ as approximations to $\text{MSE}[t_{1j}(\hat{\sigma}_e^{-2}, \vec{y})]$ and $\text{MSE}[t_{1j}(\hat{\sigma}_v^{-2}, \vec{y})]$ respectively could lead to serious understatement since the neglected terms are of the same order, $O(t^{-1})$, as the term due to estimating $\hat{\beta}$ in the MSE of BLUP.

5. Estimators of MSE

We now obtain estimators of $\text{MSE}[t_{1j}(\hat{\sigma}_e^{-2}, \vec{y})]$ for the three models. The bias of these estimators is of order lower than $O(t^{-1})$, under normality,
i.e.,  \( \text{E} \) \( \text{mse}[t_1(\hat{\theta},y)] - \text{MSE}[t_1(\hat{\theta},y)] = o(t^{-1}) \),
where \( \text{mse}[t_1(\hat{\theta},y)] \) denotes an estimator of
\( \text{MSE}[t_1(\hat{\theta},y)] \). Details of the derivation are
given only for the nested error regression model.

5.1 Normality-based estimators

The MSE approximation may be written as

\[
\text{MSE}[t_1(\sigma^2, y)] = g_1(\sigma^2) + g_2(\sigma^2) + g_3(\sigma^2),
\]

(5.1)

where

\[
g_1(\sigma^2) = (1-\gamma_1)\sigma^2
\]

(5.2)

is of order \( O(1) \), and

\[
g_2(\sigma^2) = (x_1'\gamma_1x_1)^{-1}(x_1'\gamma_1x_1)'
\]

(5.3)

\[
g_3(\sigma^2) = \left[ n^{-2} \sigma^4 (\sum_{i,j=1}^{n} \sigma^2)^{-1} \right]^{3} \left[ \sum_{i,j=1}^{n} \sigma^4 \text{var}(\sigma^2) + \right]
\]

\[
\sum_{i,j=1}^{n} \text{cov}(\sigma^2, \sigma^2) \]

(5.4)

are both of order \( O(t^{-1}) \). The formulae for
\( \text{var}(\sigma^2) \), \( \text{var}(\sigma^2) \) and \( \text{cov}(\sigma^2, \sigma^2) \) under normality
are given in Appendix I. The estimators of
\( g_2(\sigma^2) \) and \( g_3(\sigma^2) \) are simply given by \( g_2(\hat{\sigma}^2) \) and
\( g_3(\hat{\sigma}^2) \), respectively, correct to \( O(t^{-1}) \). However,
\( g_1(\sigma^2) \) is not correct to the desired order of
approximation since its bias is of order \( O(t^{-1}) \).

The bias of \( g_1(\sigma^2) \) to \( O(t^{-1}) \) is obtained
by making a Taylor expansion of \( g_1(\sigma^2) \) about the
point \( \sigma^2 \) and then taking its expectation:

\[
E_g(\sigma^2) - g_1(\sigma^2) = -g_3(\sigma^2) + o(t^{-1}).
\]

Therefore, \( g_1(\sigma^2) + g_3(\sigma^2) \) is correct to \( O(t^{-1}) \),
in estimating \( g(\sigma^2) \). It now follows from (5.1) that an estimator of MSE correct to \( O(t^{-1}) \) is given by

\[
\text{mse}[t_1(\sigma^2, y)] = g_1(\hat{\sigma}^2) + g_2(\hat{\sigma}^2) + 2g_3(\hat{\sigma}^2).
\]

(5.5)

Similarly, for the random regression coefficient model, \( \text{mse}[t_1(\sigma^2, y)] \) is given by (5.5) with

\[
g_1(\sigma^2) = (1-\gamma_1)\sigma^2
\]

(5.6)

and

\[
g_2(\sigma^2) = (x_1'\gamma_1x_1)^{-1}(x_1'\gamma_1x_1)'
\]

(5.7)

\[
g_3(\sigma^2) = \left[ n^{-2} \sigma^4 (\sum_{i,j=1}^{n} \sigma^2)^{-1} \right]^{3} \left[ \sum_{i,j=1}^{n} \sigma^4 \text{var}(\sigma^2) + \right]
\]

\[
\sum_{i,j=1}^{n} \text{cov}(\sigma^2, \sigma^2) \]

(5.8)

The formulae for \( \text{var}(\sigma^2) \), \( \text{var}(\sigma^2) \) and \( \text{cov}(\sigma^2, \sigma^2) \)
under normality are given in Appendix I.

Turning to the Fay-Herriot model an estimator
of MSE \( t_1(\sigma, y) \), correct to \( O(t^{-1}) \), is given by

\[
\text{mse}[t_1(\hat{\sigma}, y)] = D_1(\hat{\sigma} + D_1)^{-1}x_1'x_1^{-1}x_1'
\]

\[
+ 2D_1^{-1}(\hat{\sigma} + D_1)^{-1} \text{estvar}(\hat{\sigma}),
\]

(5.9)

where \( \text{estvar}(\hat{\sigma}) \) is the value of \( \text{var}(\hat{\sigma}) \) evaluated
at \( \sigma = \hat{\sigma} \), and the formula for \( \text{var}(\hat{\sigma}) \) under normality
is given in Appendix I.

5.2 Weighted jackknife estimators

A robust estimator of the MSE approximation,
correct to \( O(t^{-1}) \), can be obtained by the weighted
jackknife method of Section 3.4. The robust
estimators of \( g_1(\sigma^2) \) and \( g_2(\sigma^2) \) are given by \( g_1(\hat{\sigma}^2) \) (see (3.16)) and \( g_2(\hat{\sigma}^2) \), respectively,
for the nested error regression model. Letting
\( Q_1(e) = e^2 + e_1(\hat{\sigma}^2 - e_1) \) and \( Q_1(v) = e^2 + t v_1(\hat{\sigma}^2 - e_1) \), robust jackknife estimators of
\( \text{var}(\sigma^2) \), \( \text{var}(\sigma^2) \) and \( \text{cov}(\sigma^2, \sigma^2) \) are given by

\[
v_j(\hat{\sigma}^2) = [t(t-1)]^{-1} [Q_1(e) - \bar{Q}(e)]^2
\]

(5.10)

\[
v_j(\hat{\sigma}^2) = [t(t-1)]^{-1} [Q_1(v) - \bar{Q}(v)]^2
\]

(5.11)

and

\[
\text{cov}_j(\hat{\sigma}^2, \hat{\sigma}^2) = [t(t-1)]^{-1} [Q_1(e) - \bar{Q}(e)] [Q_1(v) - \bar{Q}(v)]
\]

(5.12)

Note that \( \bar{Q}(e) = \sigma^2 \) and \( \bar{Q}(v) = \sigma^2 \). The jack-
knife estimators (5.10)–(5.12) can be shown to be consistent
at \( t + \infty \) (Prasad, 1985). It now
follows that a robust estimator of the approxima-
tion to MSE \( t_1(\sigma^2, y) \) is given by

\[
\text{mse}_j[t_1(\sigma^2, y)]] = g_1(\hat{\sigma}^2) + g_2(\hat{\sigma}^2) + 2g_3(\hat{\sigma}^2),
\]

(5.13)

where

\[
\text{mse}_j[t_1(\sigma^2, y)] = g_1(\hat{\sigma}^2) + g_2(\hat{\sigma}^2) + n^{-1}(\hat{\sigma}^2 + \sigma^2)^{-3} \left[ \sum_{i,j=1}^{n} \sigma^4 \text{var}(\sigma^2) + \sum_{i,j=1}^{n} \text{cov}(\sigma^2, \sigma^2) \right].
\]

(5.14)

It should be noted that \( \text{mse}_j[t_1(\sigma^2, y)] \) is not a robust estimator of the true MSE since the MSE
approximation itself depends on normality of \( \{v_i\} \)
and \( \{e_{ij}\} \).

Turning to the random regression model, we
again modify \( \hat{\sigma}^2 \) to

\[
\hat{\sigma}^2 = n^{-1} [\sum_{i,j=1}^{n} \sigma^2] - \sum_{i,j=1}^{n} \bar{x}_{ij}(n-1)^{-1} \sum_{i,j=1}^{n} \bar{x}_{ij}^{-1}
\]

(5.15)

where

\[
\bar{x}_{ij} = n^{-1} \sum_{i,j=1}^{n} \bar{x}_{ij} \sum_{i,j=1}^{n} \bar{x}_{ij}^{-1}.
\]
The estimator $\widehat{\sigma_e^2}$, given by (3.8), remains unchanged but we denote it as $\overline{\sigma_e^2}$ for notational consistency. The robust estimator of MSE[$t_i(\overline{\sigma_e^2},y)$] is again given by (5.13), provided $a_i$ and $b_i$ are changed to

$$\tilde{a}_i = 1-(n_i-1)(n-t)^{-1}$$

and

$$\tilde{b}_i = 1-\tilde{a}_i^{-1}$$

respectively, where

$$\overline{n}_i = \sum_{j=1}^{n_i} x_{ij}^2 - (n_i-1)x_i^2, \quad \overline{x}_{ij}^2 = \sum_{j=1}^{n_i} x_{ij}$$

Finally, for the Fay-Herriot model with a single concomitant variable, a jackknife estimator of $g(A)$ is given by

$$g_j(A) = \sum_i (g(A) + t_i c_i[g(A) - g(A(-i))])$$

where

$$Q_i(g) = g(A) + t_i c_i[g(A) - g(A(-i))]$$

A(-i) = (n-1)c_i[(n-1)A - z_i]_i$$

and

$$z_i = (n-1antiago-1)c_i[(n-1)\hat{A} - z_i]$$

It can be shown that the bias of $g_j(A)$ is of lower order than $O(t^{-1})$. A jackknife estimator of var($\hat{A}$) is given by

$$v_j(A) = t(n-1)c_i[(n-1)\hat{A} - z_i]^2$$

which is consistent as $t \to \infty$. Putting these results together, we get the following robust estimator of MSE approximation:

$$\text{mse}_j[t_i(A,y)] = \sum_i (g(A) + t_i c_i[g(A) - g(A(-i))])^2 + 2D_i^2(A + D_i)^{-3} v_j(A)$$

where

$$g_i(A) = AD_i/(A + D_i).$$

6. Monte Carlo Study

6.1. Objectives.

A Monte Carlo study under the nested error regression model was conducted to study the finite sample properties of estimated BLUP's. In particular, we have studied the efficiency of estimated BLUP, $t_i(\overline{\sigma_e^2},y)$, under normality of errors $(v_i)$ and $(e_{ij})$, ranged from 123% to 184% with respect to $\overline{y}_i$(syn), and from 142% to 274% with respect to $\overline{y}_i$(reg). The relative efficiency with respect to $\overline{y}_i$(syn) increases as $n_i$ increases from 2 to 6, while the relative efficiency with respect to $\overline{y}_i$(reg) exhibited an opposite trend, i.e., it decreased as $n_i$ increases.

Table 1 reports the percent gain in efficiency of the jackknife estimator $t_{ij}(\overline{\sigma_e^2},y)$ over $t_i(\overline{\sigma_e^2},y)$, given by $100 \times \text{MSE}[t_{ij}(\overline{\sigma_e^2},y)] - \text{MSE}[t_i(\overline{\sigma_e^2},y)] / \text{MSE}[t_i(\overline{\sigma_e^2},y)]$. The values reported in Tables 1-5 are averages over small areas having the same $n_i$-value. It is evident from Table 1 that the gain in efficiency of $t_{ij}(\overline{\sigma_e^2},y)$ is small (≤5%) and that it decreases as $n_i$ increases, under both normality and deviations from normality.
6.4. Accuracy of second order approximation to MSE

Table 2 gives the percent relative error (R.E.) of the second order approximation, MSE*, i.e., 
\[ \frac{\text{MSE}^* - \text{MSE}}{\text{MSE}} \times 100 \]

The percent relative error under normality of both \( v_i \) and \( e_{ij} \) is small (R.E. \( \leq 2\% \)). The approximation is quite satisfactory under deviations from normality for \( e_{ij} \) but assuming that \( v_i \) are normal (R.E. \( < 5\% \)); R.E. is negligible (< 1%) under uniform distribution of \( e_{ij} \). When both errors are exponential, the approximation leads to considerable overstatement of MSE, R.E. ranging from 9.5% to 22%. The approximation is also not quite satisfactory when both errors are double exponential, R.E. ranging from 7% to 12%. Under uniform distributions for both \( v_i \) and \( e_{ij} \), the approximation in fact leads to a slight understatement, R.E. ranging from -0.5% for \( n_i = 2 \) to -4% for \( n_i = 6 \).

The accuracy of MSE* depends on the negligibility of the cross-product term 
\[ 2E[t_i(\delta^2,y)-t_i(\delta^2,y)](t_i(\delta^2,y)-\mu_i)/\text{MSE}, \]
which is exactly zero under normality, and the accuracy of the approximation (4.7) to 
\[ E[t_i(\delta^2,y)-t_i(\delta^2,y)]^2. \]

The value of cross-product term ranged from \(-5\%\) to \(-15\%\) under exponential distributions and \(-4\%\) to \(-9\%\) under double exponential distributions, as \( n_i \) increased from 2 to 6, compared to \(-1.4\%\) to \(-3.7\%\) under exponential for \( e_{ij} \) only and \(-1\%\) to \(-2.4\%\) under double exponential for \( e_{ij} \) only. These results imply that the formula (4.1) for MSE under normality leads to considerable overstatement when both \( v_i \) and \( e_{ij} \) are exponential or double exponential. In practice, however, it may be more realistic to assume that the random effects \( v_i \) are approximately normal.

Turning to the accuracy of the approximation (4.7) to 
\[ E[t_i(\delta^2,y)-t_i(\delta^2,y)]^2, \]
the difference relative to MSE ranged from \(4\%\) to \(9\%\) under exponential distributions and \(3\%\) to \(6\%\) under double exponential distributions, compared to \(0.5\%\) to \(2.1\%\) under exponential for \( e_{ij} \) only and \(0.2\%\) to \(2.0\%\) under double exponential for \( e_{ij} \) only. The overstatement of the approximation, therefore, is substantial when both \( v_i \) and \( e_{ij} \) are exponential or double exponential.

6.5. Relative bias of estimators of MSE

Tables 3 and 4 report the percent relative biases of normality-based and weighted jackknife estimators of MSE, denoted by 
\[ \text{Bias}_N = \frac{E[\text{MSE}(t_i(\delta^2,y))] - \text{MSE}(t_i(\delta^2,y))}{\text{MSE}(t_i(\delta^2,y))} \times 100 \]
and 
\[ \text{Bias}_j = \frac{E[\text{MSE}(t_i(\delta^2,y)) - \text{MSE}(t_i(\delta^2,y))]}{\text{MSE}(t_i(\delta^2,y))} \times 100. \]

It is seen from Table 3 that \( \text{Bias}_N \) is small \((< 5\%)\) when both \( v_i \) and \( e_{ij} \) are normal or uniform. It is also small when \( v_i \) is normal and \( e_{ij} \) uniform. It ranges from \(2\%\) to \(17\%\) under double exponential distributions and from \(-1.5\%\) to \(-17\%\) under exponential distributions, as \( n_i \) increases from 2 to 6, compared to \(0.5\%\) to \(10.0\%\) under double exponential for \( e_{ij} \) only and \(1\%\) to \(13\%\) under exponential for \( e_{ij} \) only. The normality-based estimator of MSE, therefore, leads to considerable overestimation as \( n_i \) increases, when both \( v_i \) and \( e_{ij} \) are exponential or double exponential. This is also true to a lesser extent under exponential or double exponential for \( e_{ij} \) only.

Turning to the jackknife estimator, Table 4 shows that \( \text{Bias}_j \) is somewhat larger than \( \text{Bias}_N \) when both \( v_i \) and \( e_{ij} \) are normal, ranging from \(2\%\) to \(8\%\). It is, however, small \((< 2\%)\) under uniform for both errors or for \( e_{ij} \) only. On the other hand, \( \text{Bias}_j \) is considerable under exponential and double exponential distributions, ranging from \(8\%\) to \(10\%\) and \(5\%\) to \(11\%\) respectively, as \( n_i \) increases from 2 to 6. It is smaller than \( \text{Bias}_N \) under double exponential and exponential for \( e_{ij} \) only, ranging from \(0.5\%\) to \(9\%\) and \(1\%\) to \(9\%\) respectively; for \( n_i = 5 \) and 6, the values of \( \text{Bias}_j \) are somewhat larger than the corresponding values of R.E., i.e., the jackknife estimator did not track the approximation to MSE very well. Overall, however, the jackknife estimator tracked the approximation to MSE better than the normality-based estimator, but the approximation itself is not very accurate under exponential or double exponential distributions for both \( v_i \) and \( e_{ij} \), as noted before.

We have also evaluated the percent underestimation of MSE when the robust estimator, \( g_{ij}(\delta^2) + \delta_{ij}(\delta^2) \), of MSE \( t_i(\delta^2,y) \) is used as an estimator of MSE \( t_i(\delta^2,y) \). The robust estimator leads to about \(8\%\) to \(15\%\) underestimation under normality of \( v_i \) and \(11\%\) to \(16\%\) underestimation when both \( v_i \) and \( e_{ij} \) are uniform. The underestimation is slightly less when both errors are double exponential or exponential, and it decreases as \( n_i \) increases.

Our investigation has shown that the jackknife estimator of MSE may be satisfactory, in the sense of providing not overly conservative standard errors, except when both errors are exponential or double exponential. As noted earlier, it may be realistic in practice to assume that \( v_i \) are approximately normal. It would be desirable, however, to develop more accurate approximations to MSE, by evaluating the cross-product term 
\[ 2E[(t_i(\delta^2,y)-t_i(\delta^2,y))(t_i(\delta^2,y)-\mu_i)], \]
and construct robust estimators of the improved approximations to MSE.

<table>
<thead>
<tr>
<th>( n_i )</th>
<th>(N,N)</th>
<th>(N,DE)</th>
<th>(N,E)</th>
<th>(N,U)</th>
<th>(DE,DE)</th>
<th>(E,E)</th>
<th>(U,U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.5</td>
<td>2.9</td>
<td>3.2</td>
<td>3.9</td>
<td>3.1</td>
<td>3.5</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>5.6</td>
<td>2.2</td>
<td>2.1</td>
<td>3.0</td>
<td>4.1</td>
<td>1.0</td>
<td>4.0</td>
</tr>
<tr>
<td>4</td>
<td>3.2</td>
<td>1.5</td>
<td>1.6</td>
<td>2.1</td>
<td>4.5</td>
<td>0.5</td>
<td>3.1</td>
</tr>
<tr>
<td>5</td>
<td>2.2</td>
<td>1.2</td>
<td>1.4</td>
<td>1.2</td>
<td>0.1</td>
<td>0</td>
<td>3.0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.4</td>
<td>-0.2</td>
<td>-0.5</td>
<td>1.4</td>
</tr>
</tbody>
</table>

N = normal, DE = double exponential, E = exponential U = uniform
Table 2. Percent Relative Error of Second Order Approximation to MSE of Estimated BLUP for (N,N), (N,DE), (N,E), (DE,DE), (E,E), (U,U) Distributions of \( \{v_i\} \) and \( \{e_{ij}\} \) Respectively.

<table>
<thead>
<tr>
<th>( n )</th>
<th>(N,N)</th>
<th>(N,DE)</th>
<th>(N,E)</th>
<th>(N,U)</th>
<th>(DE,DE)</th>
<th>(E,E)</th>
<th>(U,U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.0</td>
<td>1.0</td>
<td>2.3</td>
<td>0.7</td>
<td>7.0</td>
<td>9.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>3</td>
<td>1.5</td>
<td>3.5</td>
<td>4.8</td>
<td>0.4</td>
<td>10.0</td>
<td>18.0</td>
<td>-1.8</td>
</tr>
<tr>
<td>4</td>
<td>2.0</td>
<td>3.5</td>
<td>5.2</td>
<td>0.4</td>
<td>12.0</td>
<td>21.0</td>
<td>-2.5</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>3.5</td>
<td>4.0</td>
<td>-0.3</td>
<td>12.0</td>
<td>22.0</td>
<td>-3.0</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>3.0</td>
<td>4.2</td>
<td>-0.3</td>
<td>12.0</td>
<td>19.0</td>
<td>-4.0</td>
</tr>
</tbody>
</table>

N = normal, DE = double exponential, E = exponential, U = uniform

Table 3. Percent Relative Bias of Normality-Based Estimator of MSE for (N,N), (N,DE), (N,E), (N,U), (DE,DE), (E,E), (U,U) Distributions of \( \{v_i\} \) and \( \{e_{ij}\} \) Respectively.

<table>
<thead>
<tr>
<th>( n )</th>
<th>(N,N)</th>
<th>(N,DE)</th>
<th>(N,E)</th>
<th>(N,U)</th>
<th>(DE,DE)</th>
<th>(E,E)</th>
<th>(U,U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0</td>
<td>2.0</td>
<td>-1.5</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3.5</td>
<td>2.0</td>
<td>0.9</td>
<td>6.0</td>
<td>7.0</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>2.0</td>
<td>7.0</td>
<td>6.0</td>
<td>1.5</td>
<td>7.0</td>
<td>11.0</td>
<td>1.5</td>
</tr>
<tr>
<td>5</td>
<td>3.0</td>
<td>9.0</td>
<td>9.0</td>
<td>2.8</td>
<td>12.0</td>
<td>16.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>6</td>
<td>5.0</td>
<td>10.0</td>
<td>13.0</td>
<td>5.0</td>
<td>17.0</td>
<td>17.0</td>
<td>-1.5</td>
</tr>
</tbody>
</table>

N = normal, DE = double exponential, E = exponential, U = uniform

Table 4. Percent Relative Bias of Weighted Jackknife Estimator of MSE for (N,N), (N,DE), (N,E), (N,U), (DE,DE), (E,E), (U,U) Distributions of \( \{v_i\} \) and \( \{e_{ij}\} \) Respectively.

<table>
<thead>
<tr>
<th>( n )</th>
<th>(N,N)</th>
<th>(N,DE)</th>
<th>(N,E)</th>
<th>(N,U)</th>
<th>(DE,DE)</th>
<th>(E,E)</th>
<th>(U,U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0.3</td>
<td>5.0</td>
<td>8.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>3</td>
<td>2.0</td>
<td>3.0</td>
<td>4.0</td>
<td>0.0</td>
<td>8.0</td>
<td>12.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>4</td>
<td>4.0</td>
<td>4.5</td>
<td>5.0</td>
<td>0.3</td>
<td>11.0</td>
<td>14.0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>6.0</td>
<td>7.0</td>
<td>1.1</td>
<td>10.0</td>
<td>16.0</td>
<td>1.0</td>
</tr>
<tr>
<td>6</td>
<td>8.0</td>
<td>9.0</td>
<td>9.0</td>
<td>1.1</td>
<td>11.0</td>
<td>18.0</td>
<td>1.5</td>
</tr>
</tbody>
</table>

N = normal, DE = double exponential, E = exponential, U = uniform

APPENDIX I

VARIANCE OF ESTIMATED VARIANCE COMPONENTS

1. Nested error regression model

Under normality, Battese and Fuller (1982) have shown that

\[
\text{var}(\hat{\sigma}_e^2) = 2\sigma_e^4(n-t-k+1)^{-1}\text{tr}(X'X)^{-1}X'X
\]

and

\[
\text{cov}(\hat{\sigma}_v^2, \hat{\sigma}_e^2) = -2\sigma_v \sigma_e(n-t)^{-1}\text{tr}(X'X)^{-1}X'X
\]

where

\[\sigma_v^2 = \Sigma_{i=1}^t (x_i'x_i)^{-1}, \quad \sigma_e^2 = \Sigma_{i=1}^t e_i^2 \]

2. Random regression coefficient model

Under normality, it is easily shown that

\[
\text{var}(\hat{\sigma}_v^2) = 2\sigma_v^4(n-t)^{-1}\text{tr}(X'X)^{-1}X'X
\]

and

\[
\text{cov}(\hat{\sigma}_v^2, \hat{\sigma}_e^2) = -2\sigma_v \sigma_e(n-t)^{-1}\text{tr}(X'X)^{-1}X'X
\]

3. Fay-Herriot model

Under normality, it is easily shown that

\[
\text{var}(\hat{\eta}) = 2(n-t)^{-2}\text{tr}(X'X)^{-1}X'X
\]

REFERENCES


