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 hectres 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 (hectres of corn in a segment) against x (number of pixels of corn) indicated that the segments in a particular county fall roughly around a regression line, but they tend to cluster a bit above or below the line. One way of representing this type of phenomenon is to use a nested error regression model

$$y_{ij}=x_{ij}\beta + v_i + e_{ij},\ i=1,\ldots,t;j=1,\ldots,n_i \quad (2.1)$$

where y_{ij} is the character of interest for the j-th sampled unit in the i-th small area, $x_{ij} = (x_{ij1},\ldots,x_{ijk})'$ is a $1 \times k$ vector of corresponding auxiliary values, $\beta = (\beta_1,\ldots,\beta_k)'$, is a k-vector of unknown parametres and n_i is the number of sampled units observed in the i-th small area ($\Sigma n_i=n$). The random errors v_i's are assumed to be independent $N(0,\sigma_v^2)$, independent of e_{ij}'s which are assumed to be independent $N(0,\sigma_e^2)$. The normality assumption, however, is not necessary in deriving the BLUP. The model (2.1) can also be viewed as a random intercept model by taking $x_{ij1}=1$, $\beta_1=\alpha$. The variables $\alpha_i=\alpha+v_i$ are the random intercepts.

The i-th small area population mean may be written as $\mu_i=\bar{X}_i\beta+v_i+\bar{E}_i$, where $\bar{X}_i$ and $\bar{E}_i$ are the population means of x_{ij} and e_{ij} for the i-th small area. We assume that $\bar{X}_i$ 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 $\bar{E}_i \doteq 0$ noting that $E(e_{ij})=0$. Thus, the problem is to estimate the small area means

$$\mu_i = \bar{X}_i\beta + v_i,\quad i = 1,\ldots,t\ . \quad (2.2)$$

Note that the μ_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

$$y_{ij} = x_{ij}\beta_i + e_{ij}$$
$$\beta_i = \beta + v_i,\quad j=1,\ldots,n_i;\ i=1,\ldots,t \quad (2.3)$$

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

$$\mu_i = \bar{X}_i\beta_i = \bar{X}_i\beta + \bar{X}_i v_i\ . \quad (2.4)$$

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 bench mark variables $x_i=(x_{i1},\ldots,x_{ik})$, related to the small area mean μ_i, is available for each small area i, and that the μ_i are independent $N(x_i\beta,A)$, where β is a k-vector of parameters. They further assume that the sample mean vector $\bar{y} = (\bar{y}_1,\ldots,\bar{y}_k)'$ given $\mu = (\mu_1,\ldots,\mu_k)' = \underset{1\le i\le k}{\text{col}}(\mu_i)$, is normally distributed with mean vector μ and known covariance matrix $\text{diag}(D_1,\ldots,D_k)$.

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

$$\bar{y}_i = \mu_i + e_i;\quad \mu_i = x_i\beta + v_i \quad (2.5)$$

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 unit level is not needed, unlike in the Battese-Fuller model.

3. Best Linear Unbiased Predictors

3.1. General mixed model

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

$$y = X\beta + Zv + e \tag{3.1}$$

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 θ called variance components. Henderson (1975) has shown that the BLUP of $\mu = k'\beta + m'v$ is given by

$$t(\theta,y) = k'\tilde{\beta} + m'GZ'V^{-1}(y-X\tilde{\beta}) \tag{3.2}$$

where $V = R+ZGZ'$ and $\tilde{\beta} = (X'V^{-1}X)^{-1}(X'V^{-1}y)$. He has also given a method of evaluating $\tilde{\beta}$ 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 $D = \underset{1\le i\le t}{\text{diag}}\,(d_{ii})$, and α is a scalar such that $\alpha \ne -(\Sigma a_i b_i d_{ii}^{-1})^{-1}$ where a_i and b_i are the i-th elements of a and b respectively. Then $C^{-1} = D^{-1}+\gamma a^* b^{*\prime}$, where $\gamma=-\alpha(1+\alpha\Sigma a_i b_i d_{ii}^{-1})^{-1}$, $a_i^* = a_i d_{ii}^{-1}$ and $b_i^* = b_i d_{ii}^{-1}$.

3.2. Special cases

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

$$y = \underset{1\le i\le t}{\text{col}}\ \underset{1\le j\le n_i}{\text{col}}\ (y_{ij}),\quad X = \underset{1\le i\le t}{\text{col}}\ \underset{1\le j\le n_i}{\text{col}}\ (x_{ij})$$

$$Z = \underset{1\le i\le t}{\text{diag}}\ \underset{1\le j\le n_i}{\text{col}}\ (1),\quad v = \underset{1\le i\le t}{\text{col}}\ (v_i),$$

$$e = \underset{1\le i\le t}{\text{col}}\ \underset{1\le j\le n_i}{\text{col}}\ (e_{ij}),$$

and $G = \sigma_v^2 I_t$, $R = \sigma_e^2 I_n$, $\theta = (\sigma_e^2,\sigma_v^2)' = \sigma^2$.

Now applying Lemma 2.1 with $C = V_i = \sigma_e^2 I_{n_i} + \sigma_v^2 1_{n_i} 1'_{n_i}$, $D = \sigma_e^2 I_{n_i}$, $\alpha = \sigma_v^2$, $a = b = 1_{n_i}$, we obtain $V_i^{-1} = (\sigma_e^2)^{-1}[I_{n_i} - \gamma_i n_i^{-1} 1_{n_i} 1'_{n_i}]$ and $V^{-1} = \underset{1\le i\le t}{\text{diag}}\,(V_i^{-1})$, where $\gamma_i = \sigma_v^2(\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-1}$.

Taking now $k = \bar{X}'_i$ and $m = \underset{1\le \ell\le t}{\text{col}}\ (\delta_{i\ell})$ in (3.2), where $\delta_{i\ell} = 1$ if $\ell = i$ and $\delta_{i\ell} = 0$ if $\ell \ne i$, we get the BLUP of μ_i as

$$t_i(\sigma^2,y) = \bar{X}_i\tilde{\beta} + \gamma_i(\bar{y}_i - \bar{x}_i\tilde{\beta}) \tag{3.3}$$

where $\bar{x}_i$ 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 μ_i given by

$$t_i(\sigma^2,y) = \bar{X}_i\tilde{\beta} + \gamma_i(\bar{y}_i - \bar{x}_i\tilde{\beta}) \tag{3.4}$$

where $\sigma^2 = (\sigma_e^2,\sigma_v^2)$,

$$\gamma_i = \sigma_v^2\{\sigma_v^2 + \sigma_e^2(\sum_j x_{ij}^2)^{-1}\}^{-1} ,$$

$\sigma_v^2 = \text{var}\,(v_i)$ and

$$\tilde{\beta} = \{\sum_i \gamma_i(\sum_j x_{ij}y_{ij})(\sum_j x_{ij}^2)^{-1}\}(\sum_i \gamma_i)^{-1}.$$

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

$$t_i(A,\bar{y}) = x_i\tilde{\beta} + \frac{A}{A+D_i}(\bar{y}_i - x_i\tilde{\beta}) \tag{3.5}$$

where $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}\bar{y}$ and $V = \underset{1\le i\le t}{\text{diag}}\,(A+D_i)$. Under normality, the predictor (3.5) is a Bayes estimator, as shown by Fay and Herriot (1979). Note that $t_i(A,\bar{y})$ tends to the usual survey estimator $\bar{y}_i$ as $D_i/(A+D_i) \to 0$ and to the synthetic estimator $x_i\tilde{\beta}$ as $A/(A+D_i) \to 0$. Thus the BLUP is a weighted average of the two estimators, where the weight $w = A/(A+D_i)$ reflects the uncertainty, A , in the model for the μ_i's relative to the total variance $A+D_i$.

Schaible et al. (1977) also considered similar composite estimators, but the weight is obtained by minimizing the mean square error under repeated sampling from a fixed finite population.

3.3. Estimated BLUP

The BLUP $t(\theta,y) = t(\theta)$ depends on the variance components θ , but in practice the components of θ will be unknown. Actually, the BLUP depends only on the ratios θ_i/θ_m, where $\theta = (\theta_1,\ldots,\theta_m)'$ e.g., the BLUP (3.4) depends only on σ_v^2/σ_e^2. It is customary to estimate the BLUP by replacing θ by an asymptotically consistent estimator $\hat{\theta}$. The resulting estimator, $t(\hat{\theta},y) = t(\hat{\theta})$, of μ will remain unbiased provided $\hat{\theta}$ is even and translation invariant, i.e., $\hat{\theta}(-y) = \hat{\theta}(y)$, and the distributions of v and e are both symmetric (not necessarily normal); see Kackar and Harville (1984). However, the MSE of $t(\hat{\theta})$ will be increased relative to $\text{MSE}[t(\theta)]$, as shown in Section 4.

Various methods of estimating θ 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 $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ in the case of Battese-Fuller model. Let $\{\hat{e}_{ij}\}$ be the estimated residuals obtained from the ordinary least squares regression of $y_{ij} - \bar{y}_i$ on $\{x_{ij1} - \bar{x}_{i.1},\ldots,x_{ijk} - \bar{x}_{i.k}\}$. Similarly, a second set of residuals $\{\hat{u}_{ij}\}$ are obtained by performing ordinary least squares regression of

y_{ij} on $\{x_{ij1},\ldots,x_{ijk}\}$. Then, unbiased quadratic estimators of σ_v^2 and σ_e^2 are obtained as

$$\hat{\sigma}_e^2 = (n-t-k+\lambda)^{-1}\Sigma\Sigma\hat{e}_{ij}^2 \tag{3.6}$$

and

$$\hat{\sigma}_v^2 = n_*^{-1}[\Sigma\Sigma\ \hat{u}_{ij}^2 - (n-k)\hat{\sigma}_e^2], \tag{3.7}$$

where

$$(t-\lambda)n_* = n-tr[(X'X)^{-1}\sum_{i=1}^{t} n_i^2\bar{x}_i\bar{x}_i']$$

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 = 1) we obtain

$$\hat{\sigma}_e^2 = (n-t)^{-1}\ \Sigma\Sigma\hat{e}_{ij}^2 \tag{3.8}$$

$$\hat{\sigma}_v^2 = \tilde{n}_*^{-1}[\Sigma\Sigma\ \hat{u}_{ij}^2 - (n-1)\hat{\sigma}_e^2], \tag{3.9}$$

where

$$\hat{e}_{ij} = y_{ij}-x_{ij}(\sum_j x_{ij}y_{ij})(\sum_j x_{ij}^2)^{-1}$$

and

$$\tilde{n}_* = \Sigma\Sigma\ x_{ij}^2 - \{\sum_i(\sum_j x_{ij}^2)^2\}(\Sigma\Sigma\ x_{ij}^2)^{-1}\ .$$

An unbiased quadratic estimator of A in the Fay-Herriot model can be obtained as

$$\hat{A} = (t-k)^{-1}[\sum_{i=1}^{t}\hat{u}_i^2 - \sum_{i=1}^{t}D_i(1-x_i(X'X)^{-1}x_i')] \tag{3.10}$$

where

$$\hat{u}_i = \bar{y}_i-x_i\hat{\beta},\quad \hat{\beta} = (X'X)^{-1}X'\bar{y}\ .$$

It is possible for $\hat{\sigma}_v^2$, defined by (3.7) or (3.9), or $\hat{A}$, given by (3.10), to be negative. In practice if $\hat{A}$ or $\hat{\sigma}_v^2$ is negative, we set it equal to zero. Fay and Herriot (1979) obtained an estimator A as a solution of the nonlinear equation

$$\Sigma\frac{(\bar{y}_i - \tilde{\nu}_i)^2}{\tilde{A} + D_i} = t - k \tag{3.11}$$

where $\tilde{\nu}_i = x_i\tilde{\beta}$, $\tilde{\beta} = (X'\tilde{V}^{-1}X)^{-1}(X'\tilde{V}^{-1}\bar{y})$ and $\tilde{V} = \underset{1\le i\le t}{\text{diag}}(\tilde{A}+D_i)$, noting that the $\bar{y}_i$'s are marginally independent $N(x_i\beta,\tilde{A}+D_i)$. The resulting estimator of μ_i is an empirical Bayes estimator, under normality.

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., $k=2$, $x_{ij1}=1$, $x_{ij2}=x_{ij}$, $\beta_1 = \alpha$, $\beta_2 = \beta$). The estimator is obtained by replacing $\hat{\gamma}_i$ in $t_i(\hat{\sigma}^2,y)$ by a jackknife estimator $\tilde{\gamma}_i(J)$ which is approximately unbiased, i.e., $E[\tilde{\gamma}_i(J)-\gamma_i] = o(t^{-1})$ while $E[\hat{\gamma}_i-\gamma_i] = O(t^{-1})$ for large t and bounded n_i. Fuller and Harter (1986) have also obtained an alternative estimator of γ_i, but its approximate unbiasedness depends on normality of the errors $\{v_i\}$ and $\{e_{ij}\}$ unlike $\tilde{\gamma}_i(J)$. For the Fay-Herriot model, Morris (1983) used $1-[(t-k-2)/(t-k)]D_i/(\tilde{A}+D_i)$ as an estimator of $w_i = A/(A+D_i)$ in (3.5), noting that in the equal variance case, $D_i = D$, it is exactly unbiased under normality.

We modify the estimator $\hat{\sigma}_v^2$ slightly in order to construct weighted pseudo-values, similar to Hinkley's (1977) for the standard regression model. The modified estimator $\tilde{\sigma}_v^2$ is obtained along the lines of Arvsen (1969) for the ANOVA model $y_{ij} = \mu + v_i + e_{ij}$:

$$\tilde{\sigma}_v^2 = n_*^{-1}[\Sigma\Sigma\ \hat{u}_{ij}^2 - \sum_i\frac{\lambda_{1i}}{\lambda_{2i}}\sum_j\hat{e}_{ij}^2] \tag{3.12}$$

where n_*, $\hat{u}_{ij}$ and $\hat{e}_{ij}$ are as before, and

$$\lambda_{1i} = (n_i-1) - B_1\sum_j(x_{ij} - \bar{x}_i)^2$$

$$\lambda_{2i} = n_in^{-1}(n-1) - B_2\sum_j(x_{ij} - \bar{x})^2$$

with

$$B_1 = [\Sigma\Sigma(x_{ij}-\bar{x}_i)^2]^{-1} \text{ and } B_2 = [\Sigma\Sigma(x_{ij}-\bar{x})^2]^{-1}$$

and $\bar{x} = \Sigma n_i\bar{x}_i/n$. The estimator $\tilde{\sigma}_v^2$ is seen to be unbiased, noting that

$$E(\sum_j\hat{e}_{ij}^2) = \lambda_{1i}\sigma_e^2,\quad E(\sum_j\hat{u}_{ij}^2) = n_i^*\sigma_v^2 + \lambda_{2i}\sigma_e^2\ ,$$

where

$$n_i^* = n_i-2n^{-1}n_i^2+n^{-1}n_i\Sigma n_i^2+B_2^2[\Sigma n_i^2(\bar{x}_i-\bar{x})^2]\sum_j(x_{ij}-\bar{x})^2$$
$$-2B_2n_i^2(\bar{x}_i-\bar{x})^2+2B_2[\Sigma n_i^2(\bar{x}_i-\bar{x})^2]n_i(\bar{x}_i-\bar{x})n^{-1},$$

and

$$\Sigma\lambda_{1i}=n-t-1,\quad \Sigma n_i^*=n-\Sigma n_i^2/n-B_2\Sigma n_i^2(\bar{x}_i-\bar{x})^2 = n_*\ .$$

The estimator $\hat{\sigma}_e^2$ remains unchanged, but for notational consistency we denote it as $\tilde{\sigma}_e^2$.

Our jackknife procedure is essentially based on deleting the residuals $\{\hat{e}_{ij},\hat{u}_{ij},j=1,\ldots,n_i\}$ in turn for $i=1,\ldots,t$ and then computing the estimates of σ_e^2 and σ_v^2 as above:

$$\tilde{\sigma}_e^2(-i) = (n-t-1-\lambda_{1i})^{-1}(\Sigma\Sigma\ \hat{e}_{ij}^2 - \sum_j\hat{e}_{ij}^2) \tag{3.13}$$

$$\tilde{\sigma}_v^2(-i)=(n_*-n_i^*)^{-1}[(\Sigma\Sigma\hat{u}_{ij}^2 - \sum_i\frac{\lambda_{2i}}{\lambda_{1i}}\sum_j\hat{e}_{ij}^2)-$$

$$-(\sum_j u_{ij}^2 - \frac{\lambda_{2i}}{\lambda_{1i}} \sum_j \hat{e}_{ij}^2)]. \quad (3.14)$$

The estimators $\tilde{\sigma}_e^2(-i)$ and $\tilde{\sigma}_v^2(-i)$ are both unbiased, and $\tilde{\sigma}_e^2$, $\tilde{\sigma}_v^2$ both can be expressed as weighted sums of $\tilde{\sigma}_e^2(-i)$ and $\tilde{\sigma}_v^2(-i)$ respectively, similar to Wu's (1986) representation for the standard regression model:

$$\tilde{\sigma}_e^2 = (t-1)^{-1}\Sigma a_i \tilde{\sigma}_e^2(-i) \text{ and } \tilde{\sigma}_v^2 = (t-1)^{-1}\Sigma b_i \tilde{\sigma}_v^2(-i), \quad (3.15)$$

where

$$a_i = 1-(n-t-1)^{-1}\lambda_{1i}, \; b_i = 1-n_*^{-1}n_i^* .$$

For a general function, $g(\sigma_e^2, \sigma_v^2)$, of σ_e^2 and σ_v^2, our jackknife estimator is given by

$$g_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) = t^{-1}\Sigma Q_i(g), \quad (3.16)$$

where

$$Q_i(g) = g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2) - t[g\{\tilde{\sigma}_e^2 + a_i(\tilde{\sigma}_e^2(-i) - \tilde{\sigma}_e^2),$$
$$\tilde{\sigma}_v^2 + b_i(\tilde{\sigma}_v^2(-i) - \tilde{\sigma}_v^2)\} - g(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)] \quad (3.17)$$

are the weighted pseudo-values. In the linear case, $g(\sigma_e^2, \sigma_v^2) = \ell_1\sigma_e^2 + \ell_2\sigma_v^2$ for some known constants ℓ_1 and ℓ_2, the jackknife estimator $g_J(\tilde{\sigma}_e^2, \tilde{\sigma}_v^2)$ reduces to the unbiased estimator $\ell_1\tilde{\sigma}_e^2 + \ell_2\tilde{\sigma}_v^2$, while it is approximately unbiased in the nonlinear case (see Prasad, 1985). The jackknife estimator $\tilde{\gamma}_{iJ}$ is a special case of (3.16) with $g(\sigma_e^2, \sigma_v^2) = \gamma_i = \sigma_v^2(\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-1}$. The estimated BLUP, $t_{iJ}(\tilde{\sigma}^2, y)$ is given by (3.3) if γ_i and $\tilde{\beta} = \tilde{\beta}(\sigma^2)$ are replaced by $\tilde{\gamma}_i(J)$ and $\tilde{\beta}(\tilde{\sigma}^2)$ respectively.

4. Second Order Approximation to MSE

Kackar and Harville (1984) have shown that

$$MSE[t(\hat{\theta})] = MSE[t(\theta)] + E[t(\hat{\theta}) - t(\theta)]^2 \quad (4.1)$$

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

$$E[t(\hat{\theta})-t(\theta)]^2 \doteq E[d(\theta)'(\hat{\theta}-\theta)]^2 \quad (4.2)$$

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

$$E[d(\theta)'(\hat{\theta}-\theta)]^2 \doteq tr[A(\theta)E(\hat{\theta}-\theta)(\hat{\theta}-\theta)'] \quad (4.3)$$

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) to $E[t(\hat{\theta})-t(\theta)]^2$ is $o(t^{-1})$ for large t. The three small area models satisfy these conditions.

For the Battese-Fuller model, $MSE[t_i(\sigma^2,y)]$ under arbitrary distributions of $\{v_i\}$ and $\{e_{ij}\}$ is obtained as

$$MSE[t_i(\sigma^2,y)] = (1-\gamma_i)\sigma_v^2$$
$$+ (\bar{X}_i - \gamma_i\bar{x}_i)(X'V^{-1}X)^{-1}(\bar{X}_i - \gamma_i\bar{x}_i) \quad (4.4)$$

using Henderson's (1975) general result, where $(X'V^{-1}X)^{-1}$ is the covariance matrix of $\tilde{\beta}$, so that the second term on r.h.s. of (4.4) is of order $O(t^{-1})$ for large t. Similarly, for the random regression coefficient model and the Fay-Herriot model we get

$$MSE[t_i(\sigma^2,y)] = (1-\gamma_i)\sigma_v^2 + (\bar{X}_i - \gamma_i\bar{x}_i)^2$$
$$\times[\Sigma\Sigma x_{ij}^2(\sigma_e^2 + \sigma_v^2\sum_j x_{ij}^2)^{-1}]^{-1} \quad (4.5)$$

and

$$MSE[t_i(A,\bar{y})] = AD_i(A+D_i)^{-1} + x_i(X'V^{-1}X)^{-1}x_i'. \quad (4.6)$$

Again, 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 $E[t(\hat{\theta})-t(\theta)]^2$ for the Battese-Fuller model reduces to

$$E[t_i(\hat{\sigma}^2,y) - t_i(\sigma^2,y)]^2$$
$$\doteq n_i^{-2}(\sigma_v^2 + \sigma_e^2 n_i^{-1})^{-3} var(\hat{\sigma}_e^2\sigma_v^2 - \hat{\sigma}_v^2\sigma_e^2) + o(t^{-1}). \quad (4.7)$$

Similarly, for the random regression coefficient model

$$E[t_i(\hat{\sigma}^2,y) - t_i(\sigma^2,y)]^2$$
$$\doteq \bar{X}_i^2(\sum_j x_{ij}^2)(\sigma_v^2\sum_j x_{ij}^2 + \sigma_e^2)^{-3} var(\hat{\sigma}_e^2\sigma_v^2 - \hat{\sigma}_v^2\sigma_e^2) + o(t^{-1}). \quad (4.8)$$

In the case of Fay-Herriot model, we get

$$E[t_i(\hat{A},\bar{y}) - t_i(A,\bar{y})]^2 \doteq D_i^2(A+D_i)^{-3} var(\hat{A}) + o(t^{-1}). \quad (4.9)$$

Ignoring the uncertainty in $\hat{\sigma}^2$ and $\hat{A}$ and using $MSE[t_i(\sigma^2,y)]$ and $MSE[t_i(A,\bar{y})]$ as approximations to $MSE[t_i(\hat{\sigma}^2,y)]$ and $MSE[t_i(\hat{A},\bar{y})]$ respectively could lead to serious understatement since the neglected terms are of the same order, $O(t^{-1})$, as the term due to estimating β in the MSE of BLUP.

5. Estimators of MSE

We now obtain estimators of $MSE[t_i(\hat{\theta},y)]$ for the three models. The bias of these estimators is of order lower than $O(t^{-1})$, under normality, i.e., $E\ \text{mse}[t_i(\hat\theta,y)]-\text{MSE}[t_i(\hat\theta,y)] = o(t^{-1})$, where $\text{mse}[t_i(\hat\theta,y)]$ denotes an estimator of $\text{MSE}[t_i(\hat\theta,y)]$. 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

$$\text{MSE}[t_i(\hat\sigma^2,y)] \doteq g_1(\sigma^2)+g_2(\sigma^2)+g_3(\sigma^2), \quad (5.1)$$

where

$$g_1(\sigma^2) = (1-\gamma_i)\sigma_v^2 \quad (5.2)$$

is of order $O(1)$, and

$$g_2(\sigma^2) = (X_i-\gamma_i\bar{x}_i)'(X'V^{-1}X)^{-1}(\bar{X}_i-\gamma_i\bar{x}_i) \quad (5.3)$$

$$g_3(\sigma^2) = n_i^{-2}(\sigma_v^2+\sigma_e^2 n_i^{-1})^{-3}[\sigma_e^4\text{var}(\hat\sigma_v^2) + \sigma_v^4\text{var}(\hat\sigma_e^2)-2\sigma_e^2\sigma_v^2\text{cov}(\hat\sigma_v^2,\hat\sigma_e^2)] \quad (5.4)$$

are both of order $O(t^{-1})$. The formulae for $\text{var}(\hat\sigma_v^2)$, $\text{var}(\hat\sigma_e^2)$ and $\text{cov}(\hat\sigma_v^2,\hat\sigma_e^2)$ 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, $g_1(\hat\sigma^2)$ is not correct to the desired order of approximation since its bias is or order $O(t^{-1})$. The bias of $g_1(\hat\sigma^2)$ to order $O(t^{-1})$ is obtained by making a Taylor expansion of $g_1(\hat\sigma^2)$ about the point σ^2 and then taking its expectation:

$$Eg_1(\hat\sigma^2) - g_1(\sigma^2) = -g_3(\sigma^2) + o(t^{-1}).$$

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

$$\text{mse}_N[t_i(\hat\sigma^2,y)]=g_1(\hat\sigma^2)+g_2(\hat\sigma^2)+2g_3(\hat\sigma^2). \quad (5.5)$$

Similarly, for the random regression coeficient model, $\text{mse}_N[t_i(\hat\sigma^2,y)]$ is given by (5.5) with

$$g_1(\sigma^2) = (1-\gamma_i)\sigma_v^2 \quad (5.6)$$

$$g_2(\sigma^2) = (\bar{X}_i-\gamma_i\bar{x}_i)^2 \times[\sum_i\sum_j x_{ij}^2(\sigma_e^2+\sigma_v^2\sum_j x_{ij}^2)^{-1}]^{-1} \quad (5.7)$$

and

$$g_3(\sigma^2) = \bar{X}_i^2(\sum_j x_{ij}^2)(\sigma_e^2+\sigma_v^2\sum_j x_{ij}^2)^{-3} \times [\sigma_e^4\text{var}(\hat\sigma_v^2)+\sigma_v^4\text{var}(\hat\sigma_e^2) -2\sigma_e^2\sigma_v^2\text{cov}(\hat\sigma_e^2,\hat\sigma_v^2)]. \quad (5.8)$$

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

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

$$\text{mse}_N[t_i(\hat{A},\bar{y})] = \hat{A}D_i(\hat{A}+D_i)^{-1}+x_i(X'\hat{V}^{-1}X)^{-1}x_i' +2D_i^2(\hat{A}+D_i)^{-3}\text{estvar}(\hat{A}), \quad (5.9)$$

where $\text{estvar}(\hat{A})$ is the value of $\text{var}(\hat{A})$ evaluated at $A = \hat{A}$, and the formula for $\text{var}(\hat{A})$ 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 $g_{1J}(\tilde\sigma^2)$ (see (3.16)) and $g_2(\tilde\sigma^2)$ respectively, for the nested error regression model. Letting $Q_i(e) = \tilde\sigma_e^2+ta_i(\tilde\sigma_e^2-\tilde\sigma_e^2(-i))$ and $Q_i(v) = \tilde\sigma_v^2 + tb_i(\tilde\sigma_v^2-\tilde\sigma_v^2(-i))$, robust jackknife estimators of $\text{var}(\tilde\sigma_e^2)$, $\text{var}(\tilde\sigma_v^2)$ and $\text{cov}(\tilde\sigma_e^2,\tilde\sigma_v^2)$ are given by

$$v_J(\tilde\sigma_e^2) = [t(t-1)]^{-1}\Sigma[Q_i(e)-\bar{Q}(e)]^2 = t(t-1)^{-1}\Sigma a_i^2[\tilde\sigma_e^2(-i)-\tilde\sigma_e^2] \quad (5.10)$$

$$v_J(\tilde\sigma_v^2) = [t(t-1)]^{-1}\Sigma[Q_i(v)-\bar{Q}(v)]^2 = t(t-1)^{-1}\Sigma b_i^2[\tilde\sigma_v^2(-i)-\tilde\sigma_v^2] \quad (5.11)$$

and

$$\text{cov}_J(\tilde\sigma_v^2,\tilde\sigma_e^2) = [t(t-1)]^{-1}\Sigma[Q_i(e)-\bar{Q}(e)]\times[Q_i(v)-\bar{Q}(v)] = t(t-1)^{-1}\Sigma a_ib_i[\tilde\sigma_e^2(-i)-\tilde\sigma_e^2]\times[\tilde\sigma_v^2(-i)-\tilde\sigma_v^2]. \quad (5.12)$$

Note that $\bar{Q}(e) = \tilde\sigma_e^2$ and $\bar{Q}(v) = \tilde\sigma_v^2$. The jackknife estimators (5.10)-(5.12) can be shown to be consistent at $t \to \infty$ (Prasad, 1985). It now follows that a robust estimator of the approximation to $\text{MSE}[t_i(\tilde\sigma^2,y)]$ is given by

$$\text{mse}_J[t_i(\tilde\sigma^2,y)] = g_{1J}(\tilde\sigma^2)+g_2(\tilde\sigma^2)+n_i^{-1}(\tilde\sigma_v^2+\tilde\sigma_e^2 n_i^{-1})^{-3} [\tilde\sigma_e^4 v_J(\tilde\sigma_v^2) +\tilde\sigma_v^4 v_J(\tilde\sigma_e^2)-2\tilde\sigma_e^2\tilde\sigma_v^2\text{cov}_J(\tilde\sigma_e^2,\tilde\sigma_v^2)]. \quad (5.13)$$

It should be noted that $\text{mse}_J[t_i(\tilde\sigma^2,y)]$ 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 $\hat\sigma_v^2$ to

$$\tilde\sigma_v^2 = \tilde{n}_*^{-1}[\sum_i\sum_j\hat{u}_{ij}^2 -\sum_i\tilde\lambda_i(n_i-1)^{-1}\sum_j\hat{e}_{ij}^2] \quad (5.14)$$

where

$$\tilde\lambda_i = n_i - \sum_j x_{ij}^2(\Sigma\Sigma\, x_{ij}^2)^{-1} .$$

The estimator $\hat{\sigma}_e^2$, given by (3.8), remains unchanged but we denote it as $\tilde{\sigma}_e^2$ for notational consistency. The robust estimator of $MSE[t_i(\tilde{\sigma}^2,y)]$ 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

$$\tilde{b}_i = 1-\tilde{n}_i^*\tilde{n}_*^{-1}$$

respectively, where

$$\tilde{n}_i^* = \sum_j x_{ij}^2-(\sum_j x_{ij}^2)^2(\sum\sum_{ij} x_{ij}^2)^{-1},\ \Sigma\tilde{n}_i^*=\tilde{n}_*.$$

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

$$g_J(\hat{A}) = \bar{Q}(g) = \frac{1}{t}\Sigma Q_i(g),$$

where

$$Q_i(g) = g(\hat{A})+tc_i[g(\hat{A})-g(\hat{A}(-i))],$$

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

$$c_i = 1-(t-1)^{-1}[1-x_i^2(\Sigma x_i^2)^{-1}],$$

$$\hat{A} = (t-1)^{-1}\Sigma c_i\hat{A}(i)$$

and

$$z_i = \hat{u}_i^2-D_i-x_i^2(\Sigma x_i^2D_i)(\Sigma x_i^2)^{-1}+2D_ix_i^2(\Sigma x_i^2)^{-1}.$$

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

$$v_J(\hat{A}) = t(t-1)^{-1}\Sigma c_i^2[\hat{A}(-i)-\hat{A}]^2$$

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

$$mse_J[t_i(\hat{A},\bar{y})] = g_{1J}(\hat{A})+x_i(X'\hat{V}^{-1}X)^{-1}x_i' + 2D_i^2(A+D_i)^{-3}v_J(\hat{A}),$$

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, $t_i(\tilde{\sigma}^2,y)$, under normality of errors $\{v_i\}$ and $\{e_{ij}\}$ relative to the regression synthetic estimator $\bar{y}_i(syn) = \hat{\alpha}+\hat{\beta}\bar{x}_i$ and an approximately unbiased regression estimator $\bar{y}_i(reg) = \bar{y}_i+\hat{\beta}(\bar{X}_i-\bar{x}_i)$, where $\hat{\alpha}$ and $\hat{\beta}$ are the ordinary least squares estimators of α and β respectively. The efficiency of jackknife estimator of BLUP, $t_{iJ}(\tilde{\sigma}^2,y)$, relative to $t_i(\tilde{\sigma}^2,y)$ is also evaluated under normality and deviations from normality. The accuracy of second order approximation to $MSE[t_i(\tilde{\sigma}^2,y)]$, say MSE^*, is also investigated, under normality and deviations from normality. Finally, the relative biases of normality based estimator of MSE, (5.5), and jackknife estimator of MSE, (5.13), are studied.

6.2. Description of the experiment

Battese and Fuller (1982) used the nested error regression model, $y_{ij} = \alpha+\beta x_{ij}+v_i+e_{ij}$, to predict mean corn hectares, $\bar{Y}_i$, for 12 counties (small areas) in Iowa, where x_{ij} is the number of pixels of corn for the j-th sample segment of county i and the population means $\bar{X}_i$ are known. They obtained $\hat{\sigma}_e^2 = 292$ and $\hat{\sigma}_v^2 = 64$ and generalized least squares estimates of α and β given by 5.5 and 0.388 respectively. In their data set $n_i = 1$ for three of the counties. We pooled these three small area to satisfy the requirement $n_i \geq 2$ for our jackknife method. We increased the number of small areas, t, to 20 from 10 by duplicating the x_{ij}, n_i and $\bar{X}_i$ reported by Battese and Fuller. We then generated 10,000 independent sets of normal deviates $\{e_{ij},\ j=1,\ldots,n_i;\ i=1,\ldots,20\}$ and $\{v_i,\ i=1,\ldots,20\}$ from $N(0,\sigma_e^2 = 292)$ and $N(0,\sigma_v^2 = 64)$, and then using the x_{ij}-values obtained 10,000 sets of $\{y_{ij},\ j=1,\ldots,n_i;\ i=1,\ldots,20\}$ from the model

$$y_{ij} = 5.5 + 0.388x_{ij} + v_i + e_{ij},$$

using $\alpha = 5.5$ and $\beta = 0.388$. Similarly, independent data sets were generated from the following nonnormal distributions for both v_i and e_{ij}: double exponential (symmetric, long-tailed), uniform (short-tailed) and exponential (positively skewed), all with means zero and variances 292 and 64 for v_i and e_{ij} respectively. Data sets $\{y_{ij}\}$ were also generated from normal distribution for $\{v_i\}$ and uniform, exponential and double exponential distributions for $\{e_{ij}\}$. Monte Carlo values of $MSE[t_i(\tilde{\sigma}^2,y)]$, $MSE[t_{iJ}(\tilde{\sigma}^2,y)]$ etcetra were computed from the 10,000 data sets so generated.

6.3. Efficiency of estimated BLUPs

The relative efficiency of estimated BLUP, $t_i(\tilde{\sigma}^2,y)$, under normality of both $\{v_i\}$ and $\{e_{ij}\}$, ranged from 123% to 184% with respect to $\bar{y}_i(syn)$, and from 142% to 274% with respect to $\bar{y}_i(reg)$. The relative efficiency with respect to $\bar{y}_i(syn)$ increases as n_i increases from 2 to 6, while the relative efficiency with respect to $\bar{y}_i(reg)$ exhibited an opposite trend, i.e., it decreased as n_i increases.

Table 1 reports the percent gain in efficiency of the jackknife estimator $t_{iJ}(\tilde{\sigma}^2,y)$ over $t_i(\tilde{\sigma}^2,y)$, given by $100\times\{MSE[t_i(\tilde{\sigma}^2,y)] - MSE[t_{iJ}(\tilde{\sigma}^2,y)]/MSE[t_i(\tilde{\sigma}^2,y)]\}$. 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 $t_{iJ}(\tilde{\sigma}^2,y)$ 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., $\{(MSE^*-MSE)/MSE\} \times 100$. 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 $2E[t_i(\tilde{\sigma}^2,y)-t_i(\sigma^2,y)][t_i(\sigma^2,y)-\mu_i]/MSE$, which is exactly zero under normality, and the accuracy of the approximation (4.7) to $E[t_i(\tilde{\sigma}^2,y)-t_i(\sigma^2,y)]^2$. The value of cross-product term ranged from -5% to -15% under exponential distributions and -4% to -8% 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 $E[t_i(\tilde{\sigma}^2,y)-t_i(\sigma^2,y)]^2$, 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 $Bias_N = \{E\ mse_N[t_i(\tilde{\sigma}^2,y)]-MSE[t_i(\tilde{\sigma}^2,y)]\}/MSE[t_i(\tilde{\sigma}^2,y)]\times 100$ and $Bias_J = \{E\ mse_J[t_i(\tilde{\sigma}^2mt)]-MSE[t_i(\tilde{\sigma}^2,y)] / MSE[t_i(\tilde{\sigma}^2,y)] \times 100$.

It is seen from Table 3 that $Bias_N$ is small ($\leq$ 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 $Bias_J$ is somewhat larger than $Bias_N$ when both $\{v_{ij}\}$ 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, $Bias_J$ is considerable under exponential and double exponential distributions, ranging from 8% to 18% and 5% to 11% respectively, as n_i increases from 2 to 6. It is smaller than $Bias_N$ 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 $Bias_J$ 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, $g_{1J}(\tilde{\sigma}^2) + g_2(\tilde{\sigma}^2)$, of $MSE[t_i(\tilde{\sigma}^2,y)]$ is used as an estimator of $MSE[t_i(\tilde{\sigma}^2,y)]$. 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 $2E[t_i(\tilde{\sigma}^2,y)-t_i(\sigma^2,y)][t_i(\sigma^2,y)-\mu_i]$, and construct robust estimators of the improved approximations to MSE.

Table 1. Percent Gain in Efficiency of $t_{iJ}(\tilde{\sigma}^2,y)$ over $t_i(\tilde{\sigma}^2,y)$ for (N,N), (N,DE), (N,E), (N,U), (DE,DE), (E,E), (U,U) Distributions of $\{v_i\}$ and $\{e_{ij}\}$ respectively.

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

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.

n_i	(N,N)	(N,DE)	(N,E)	(N,U)	(DE,DE)	(E,E)	(U,U)
2	2.0	1.0	2.3	0.7	7.0	9.5	-0.5
3	1.5	3.5	4.8	0.4	10.0	18.0	-1.8
4	2.0	3.5	5.2	0	12.0	21.0	-2.5
5	1.5	3.5	4.0	-0.3	12.0	22.0	-3.0
6	0.5	3.0	4.2	-0.3	12.0	19.0	-4.0

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.

n_i	(N,N)	(N,DE)	(N,E)	(N,U)	(DE,DE)	(E,E)	(U,U)
2	-1.0	0.5	1.0	0	2.0	-1.5	2.5
3	0	3.5	2.0	0.9	6.0	7.0	3.0
4	2.0	7.0	6.0	1.5	7.0	11.0	1.5
5	3.0	9.0	9.0	2.8	12.0	16.0	-1.0
6	5.0	10.0	13.0	5.0	17.0	17.0	-1.5

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.

n_i	(N,N)	(N,DE)	(N,E)	(N,U)	(DE,DE)	(E,E)	(U,U)
2	2.0	0.5	1.0	0.3	5.0	8.0	-1.0
3	2.0	3.0	4.0	0.0	8.0	12.0	-1.0
4	4.0	4.5	5.0	0.3	11.0	14.0	0
5	5.0	6.0	7.0	1.1	10.0	16.0	1.0
6	8.0	9.0	9.0	1.3	11.0	18.0	1.5

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

$$\text{var}(\hat{\sigma}_e^2) = 2\sigma_e^4(n-t-k+\lambda)^{-1}$$

$$\text{var}(\hat{\sigma}_v^2) = 2n_*^{-2}[(n-t-k+\lambda)^{-1}(t-\lambda)^2\sigma_e^4 + 2n_*\sigma_e^2\sigma_v^2 + n_{**}\sigma_v^4]$$

and

$$\text{cov}(\hat{\sigma}_e^2,\hat{\sigma}_v^2) = -(t-\lambda)n_*^{-1}\ \text{var}(\hat{\sigma}_e^2),$$

where

$$n_{**} = \Sigma n_i^2 - \text{tr}[(X'X)^{-1}\Sigma n_i^3\, \bar{x}_i'\, \bar{x}_i] .$$

2. Random regression coefficient model

Under normality, it is easily shown that

$$\text{var}(\hat{\sigma}_e^2) = 2\sigma_e^4(n-t)^{-1}$$

$$\text{var}(\hat{\sigma}_v^2) = 2\tilde{n}_*^{-2}\left[\frac{(t-1)^2}{n-t}\sigma_e^4 + 2\tilde{n}_*\sigma_e^2\sigma_v^2 + \tilde{n}_{**}\sigma_v^4\right]$$

and

$$\text{cov}(\hat{\sigma}_e^2,\hat{\sigma}_v^2) = 2(t-1)\tilde{n}_*^{-1}\ \text{var}(\hat{\sigma}_e^2)$$

where

$$\tilde{n}_* = \sum_i\sum_j x_{ij}^2 - [\sum_i(\sum_j x_{ij}^2)^2](\sum_i\sum_j x_{ij}^2)^{-1}$$

and

$$\tilde{n}_{**} = \sum_i(\sum_j x_{ij}^2)^2 - [\sum_i(\sum_j x_{ij}^2)^3](\sum_i\sum_j x_{ij}^2)^{-1}.$$

3. Fay-Herriot model

Under normality, it is easily shown that

$$\text{var}(\hat{A}) = 2(t-k)^{-2}\{(t-k)A^2+2A\Sigma D_i(1-x_i(X'X)^{-1}x_i') + \Sigma D_i^2-2\Sigma D_i^2x_i(X'X)^{-1}x_i' + \text{tr}[(X'X)^{-1}X'DX(X'X)^{-1}X'DX]\}$$

where $D = \text{diag}(D_1,\ldots,D_t)$.

REFERENCES

Arvesen, J. and Schmitz, T. (1970). Robust procedures for variance component problems using the jackknife. *Biometrics, 26,* 677-86.

Battese, G.E. and Fuller, W.A. (1982). An error component model for prediction of county crop areas using survey and satellite data. Unpublished Technical Report, Iowa State University, Ames, Iowa.

Dempster, A.P., Rubin, D.B. and Tsutakawa, R.K. (1981). Estimation in covariance components models. *J. Amer. Statist. Assoc., 76,* 341-53.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc., 74,* 269-77.

Graybill, F.A. (1969). *Introduction to Matrices with Applications in Statistics.* Wadsworth, California.

Harville, D.A. (1977). Maximum likelihood approach to variance component estimation and related problems. *J. Amer. Statist. Assoc., 72,* 320-40.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics, 31,* 423-47.

Hinkley, D.V. (1977). Jackknifing in unbalanced situations. *Technometrics, 19,* 285-92.

Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc., 79,* 853-62.

Peixoto, J.L. (1982). Estimation of Random Effects in the Balanced One-way Classification. Ph.D. thesis, Iowa State University, Ames,Iowa.

Prasad, N.G.N. (1985). Small Area Estimation and Measurement of Response Error Variance in Surveys. Unpublished Ph.D. Thesis, Carleton University, Ottawa, Canada.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika, 57,* 377-87.

Schaible, W.L., Brock, D and Schnack, G.A. (1977). An empirical comparison of the simple inflation, synthetic and composite estimators for small area statistics. Proc. Amer. Statist. Assoc., Soc. Statist. Sec., 1017-21.