Mean Square Error Approximation in Small Area Estimation By Use of Parametric and Nonparametric Bootstrap

Danny Pfeffermann, Hebrew University, Israel &University of Southampton, UK
Hagit Glickman, Central Bureau of Statistics, Israel

ABSTRACT

The computation of reliable MSE estimators in small area prediction problems is a complicated process. This is so because the models in use and the small sample sizes within the areas require the accounting for the contribution to the error resulting from estimating the model parameters. In this paper we study the use of parametric and nonparametric bootstrap methods for MSE estimation. The parametric method consists of generating parametrically a large number of area bootstrap samples from the model fitted to the original data, re-estimating the model parameters for each bootstrap sample and then estimating the separate components of the MSE. The nonparametric method generates the samples by bootstrapping standardized residuals estimated for the original data. The bootstrap methods are compared to other methods in common use in a simulation study that examines also the robustness of the various methods to non-normality of the model error terms.

1. Model, predictors and prediction MSEs

We study the use of the bootstrap methods by focusing for convenience on the small area model considered by Fay and Herriot (1979), but as becomes apparent below, the use of these methods is very general. The Fay-Herriot (F-H) model is defined as,

$$y_i = \mathrm{x}_i^T \beta + u_i + e_i\,, \quad i = 1,\ldots,m, \qquad (1)$$

where y_i is the direct small area estimator (based only on the sample from area i), x_i is a $p \times 1$ vector of known area level characteristics with fixed regression coefficients $\beta = (\beta_1,\ldots,\beta_p)^T$, $u_i \overset{iid}{\sim} N(0,\sigma_u^2)$ is the area random effect, representing the area characteristics not accounted for by the *x*-values and $e_i \overset{iid}{\sim} N(0,\psi_i)$ is the sampling error. The true (population) small area mean is $\theta_i = \mathrm{x}_i^T \beta + u_i$. It is assumed that the sampling error variances, ψ_i, are known. (Reliable estimates of these variances are ordinarily obtained from statistical bureaus carrying out the surveys.) Let $y = (y_1,\ldots,y_m)^T$, $u = (u_1,\ldots,u_m)^T$ and $e = (e_1,\ldots,e_m)^T$.

The purpose is to predict the true area means, θ_i. The mean square error (MSE) of a predictor $\hat{\theta}_i$ is defined to be, $MSE(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$. For known β and σ_u^2, the *best predictor* (BP) of $\hat{\theta}_i$ under the model (minimum MSE) is,

$$\begin{aligned}\hat{\theta}_i^{BP} &= \hat{\theta}_i[\mathrm{y};\sigma_u^2,\beta] \\ &= \mathrm{x}_i^T\beta + \gamma_i\left(y_i - \mathrm{x}_i^T\beta\right),\end{aligned} \qquad (2)$$

where $\gamma_i = \sigma_u^2/(\psi_i + \sigma_u^2)$ is the shrinkage factor. The MSE of $\hat{\theta}_i^{BP}$ is,

$$\mathrm{MSE}(\hat{\theta}_i^{BP}) = g_{1i}(\sigma_u^2) = \gamma_i \cdot \psi_i. \qquad (3)$$

Notice that equation (3) holds without the normality assumptions for the error terms u_i and e_i. The normality assumptions guarantee that the predictor defined by (2) attains the minimum MSE. When β is unknown (but σ_u^2 is known), the *best linear unbiased predictor* (BLUP) of θ_i is obtained by replacing β by its generalized least square estimator,

$$\tilde{\beta} = \tilde{\beta}(\mathrm{y};\sigma_u^2) = \left[\sum_i \frac{1}{\psi_i + \sigma_u^2}\mathrm{x}_i\mathrm{x}_i^T\right]^{-1}\left[\sum_i \frac{1}{\psi_i + \sigma_u^2}\mathrm{x}_i y_i\right],$$

that is,

$$\begin{aligned}\hat{\theta}_i^{BLUP} &= \hat{\theta}_i[\mathrm{y};\sigma_u^2,\tilde{\beta}(\mathrm{y};\sigma_u^2)] \\ &= \mathrm{x}_i^T\tilde{\beta} + \gamma_i\left(y_i - \mathrm{x}_i^T\tilde{\beta}\right).\end{aligned} \qquad (4)$$

The mean square error of the BLUP is,

$$MSE(\hat{\theta}_i^{BLUP}) = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2), \qquad (5)$$

where

$$g_{2i}(\sigma_u^2) = (1-\gamma_i)^2 \mathrm{x}_i^T \left[\sum_{j=1}^{m} \frac{1}{\psi_j + \sigma_u^2} \mathrm{x}_j \mathrm{x}_j^T \right]^{-1} \mathrm{x}_i$$
$$= (1-\gamma_i)^2 \mathrm{x}_i^T Var(\tilde{\beta}) \mathrm{x}_i .$$

Notice that the BLUP property and the MSE expression in (5) are valid without the normality assumptions.

In practice, both β and σ_u^2 are unknown and need to be estimated form the data. An *empirical BLUP* (EBLUP) is then obtained by replacing σ_u^2 by an estimator $\hat{\sigma}_u^2 = \hat{\sigma}_u^2(y)$ in the expression of the BLUP, yielding

$$\begin{aligned}\hat{\theta}_i^{EBLUP} &= \hat{\theta}_i[\mathrm{y}; \hat{\sigma}_u^2, \hat{\beta}(\hat{\sigma}_u^2)] \\ &= \mathrm{x}_i^T \hat{\beta} + \hat{\gamma}_i \left(y_i - \mathrm{x}_i^T \hat{\beta} \right),\end{aligned} \qquad (6)$$

where $\hat{\gamma}_i$ and $\hat{\beta}$ are obtained from γ_i and $\tilde{\beta}$ by replacing σ_u^2 by $\hat{\sigma}_u^2$.

In what follows we consider two widely used estimators for the variance σ_u^2 of the random effects. The first estimator is the method of moments ("fitting of constants") estimator (Prasad and Rao, 1990),

$$\begin{aligned}\tilde{\sigma}_{PR}^2 &= (m-p)^{-1} \\ &\times \sum_i [(y_i - \mathrm{x}_i^T \hat{\beta}_{OLS})^2 - \psi_i (1-\tilde{h}_i)],\end{aligned} \qquad (7)$$

where $\hat{\beta}_{OLS} = \left[\sum_i \mathrm{x}_i \mathrm{x}_i^T\right]^{-1} \left[\sum_i \mathrm{x}_i y_i\right]$ and $\tilde{h}_i = \mathrm{x}_i^T \left[\sum_i \mathrm{x}_i \mathrm{x}_i^T\right]^{-1} \mathrm{x}_i$. Notice that $\tilde{\sigma}_{PR}^2$ can be negative and hence, $\hat{\sigma}_{PR}^2 = \max(\tilde{\sigma}_{PR}^2, 0)$.

The second estimator is the Fay-Herriot estimator, $\hat{\sigma}_{FH}^2$, obtained by solving iteratively the equation,

$$\sum_i (y_i - \mathrm{x}_i^T \tilde{\beta})^2 / (\sigma_u^2 + \psi_i) = m - p \qquad (8)$$

and setting $\hat{\sigma}_{FH}^2 = 0$ if no positive solution exists. The estimators defined by (7) and (8) are,

(i) even functions of y, i.e., $\hat{\sigma}_u^2(\mathrm{y}) = \hat{\sigma}_u^2(-\mathrm{y})$, and

(ii) translation invariant functions, i.e., $\hat{\sigma}_u^2(\mathrm{y} + \mathrm{Xd}) = \hat{\sigma}_u^2(\mathrm{y})$ for any given vector $\mathrm{d} \in R^p$ and all y, where $X = [\mathrm{x}_1 ... \mathrm{x}_\mathrm{n}]^T$.

2. Bootstrap Methods for Estimating the MSE of the EBLUP

2.1. *MSE Decomposition*

The prediction error of the EBLUP can be decomposed as,

$$\begin{aligned}&\hat{\theta}_i^{EBLUP} - \theta_i = \\ &(\hat{\theta}_i^{BLUP} - \theta_i) + (\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP}),\end{aligned} \qquad (9)$$

where $\hat{\theta}_i^{BLUP}$ is defined by (4). Hence, by (5),

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}_i^{EBLUP}) &= g_1(\sigma_u^2) + g_2(\sigma_u^2) \\ &+ E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2 \\ &+ 2E(\hat{\theta}_i^{BLUP} - \theta_i)(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP}).\end{aligned} \qquad (10)$$

Under the normality assumptions for the random effects and the sampling errors, and for estimators $\hat{\sigma}_u^2$ satisfying (i) and (ii) above, the cross product expectation $CPE = E(\hat{\theta}_i^{BLUP} - \theta_i)(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})$ is zero, see Harville (1985). However, for other distributions of the model error terms the cross product expectation can be of similar magnitude as the second and third terms on the right hand side of (10). Lahiri and Rao (1995) developed a second order approximation for the cross-product expectation for the case where $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$ that only requires that the sampling errors are normally distributed. This approximation involves the forth moment of the distribution of the random effects.

In what follows we propose bootstrap methods for estimating the separate components of the MSE defined by (10), assuming the F-H model defined by (1).

2.2. *Parametric Bootstrap*

The method consists of the following steps:

1. Generate independently a large number B of random effects, $u^b = (u_1^b, \ldots, u_m^b)^T$ and sampling errors, $e^b = (e_1^b, \ldots, e_m^b)^T$ and hence bootstrap direct estimators $y^b = (y_1^b, \ldots, y_m^b)^T$, $b = 1, \ldots, B$ from the model (1) with normal error terms and hyper-parameters equal to ψ_i, $\hat{\sigma}_u^2(\mathrm{y})$ and

$\hat{\beta}[\mathrm{y};\hat{\sigma}_u^2(\mathrm{y})]$ where y defines the original sample.

2. Re-estimate the hyper-parameters σ_u^2 and β for each of the bootstrap samples using the same method as used for the original sample, yielding the estimators $\hat{\sigma}_u^2(\mathrm{y}^b)$, $\hat{\beta}[\mathrm{y}^b;\hat{\sigma}_u^2(\mathrm{y}^b)]$, and also $\hat{\beta}(\mathrm{y}^b;\hat{\sigma}_u^2(\mathrm{y}))$.

3. Estimate the MSE of the EBLUP as,

$$\hat{\mathrm{MSE}}(\hat{\theta}_i^{EBLUP}) = 2\left[g_{1i}(\hat{\sigma}_u^2(\mathrm{y})) + g_{2i}(\hat{\sigma}_u^2(\mathrm{y}))\right] - \bar{g}_{1i}^{PB} - \bar{g}_{2i}^{PB} + puc^{PB}, \qquad (11)$$

where

$$puc^{PB} = B^{-1}\sum_b \Big\{\hat{\theta}_i[y^b;\hat{\sigma}_u^2(y^b),\hat{\beta}(y^b;\hat{\sigma}_u^2(y^b))] - \hat{\theta}_i[y^b;\hat{\sigma}_u^2(y);\hat{\beta}(y^b;\hat{\sigma}_u^2(y))]\Big\}^2$$

estimates the contribution to the MSE resulting from parameter uncertainty as defined by the third term on the right side of (10), and $\bar{g}_{di}^{PB} = B^{-1}\sum_b g_{di}(\hat{\sigma}_u^2(y^b))$, d=1,2.

Using similar arguments to Pfeffermann and Tiller (2003), it follows that under mild regularity conditions the MSE estimator (11) has bias of order $O(1/m^2)$. (Pfeffermann and Tiller consider MSE estimation for state predictors that use estimated model parameters in the context of state-space models that contain the F-H model as a simple special case.)

Notice that although the method is described for the case of normal error terms, the same method can also be applied for other known distributions of the error terms. However, in this case, the cross-product expectation in (10) may no longer vanish and needs to be estimated as well. The MSE estimator is modified in this case by adding the following expression to the estimator (11),

$$cpe^{PB} = B^{-1}\sum_b \Big\{\hat{\theta}_i[y^b;\hat{\sigma}_u^2(y^b);\hat{\beta}(y^b;\hat{\sigma}_u^2(y^b))] - \hat{\theta}_i[y^b;\hat{\sigma}_u^2(y),\hat{\beta}(y^b;\hat{\sigma}_u^2(y))]\Big\} \times \{\hat{\theta}_i[y^b;\hat{\sigma}_u^2(y),\hat{\beta}(y^b;\hat{\sigma}_u^2(y))] - \theta_i^b\}, \qquad (12)$$

where $\theta_i^b = \mathrm{x}_i^T\hat{\beta}(\mathrm{y};\hat{\sigma}_u^2(\mathrm{y})) + u_i^b$ is the 'true' mean generated for area i in bootstrap sample *b*.

An alternative bootstrap estimator again resulting from Pfeffermann and Tiller (2003), is obtained by replacing (11) by

$$\hat{\mathrm{MSE}}(\hat{\theta}_i^{EBLUP}) = \left[g_{1i}(\hat{\sigma}_u^2(\mathrm{y})) + g_{2i}(\hat{\sigma}_u^2(\mathrm{y}))\right] - \bar{g}_{1i}^{PB} - \bar{g}_{2i}^{PB} + mse^{PB} \qquad (13)$$

where

$$mse^{PB} = B^{-1} \times \sum_b \{\hat{\theta}_i[y^b;\hat{\sigma}_u^2(y^b),\hat{\beta}(y^b;\hat{\sigma}_u^2(y^b))] - \theta_i^b\}^2$$

is the MSE of the EBLUP under the bootstrap model. The bootstrap MSE, mse^{PB} can be viewed as a 'naïve' MSE estimator since it ignores the bias resulting from generating the bootstrap samples with an estimator of σ_u^2 instead of the true variance.

The estimator (13) is equivalent asymptotically to the estimator (11) but it has the advantage of potential robustness against sampling from non-normal distributions. To see this, let E_* denotes the expectation with respect to the bootstrap model, i.e., when generating the area direct estimators with hyper-parameters $\hat{\sigma}_u^2(y)$ and $\hat{\beta}(y;\hat{\sigma}_u^2(y))$. Then, analogously to (10),

$$E_*\left[\hat{\theta}_i^{EBLUP} - \theta_i\right]^2 = g_1(\hat{\sigma}_u^2) + g_2(\hat{\sigma}_u^2) + E_*\left[\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP}\right]^2 + 2E_*\left[\hat{\theta}_i^{BLUP} - \theta_i\right]\left[\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP}\right] \qquad (14)$$

Hence,

$$E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})^2 + 2E(\hat{\theta}_i^{BLUP} - \theta_i)(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BLUP})$$

can be estimated by

$$E_*\left[\hat{\theta}_i^{EBLUP} - \theta_i\right]^2 - g_{1i}(\hat{\sigma}_u^2) - g_{2i}(\hat{\sigma}_u^2),$$

or by $mse^{PB} - g_{1i}(\hat{\sigma}_u^2) - g_{2i}(\hat{\sigma}_u^2)$. The estimator (13) is now obtained by adding the bias reduced estimator $2\left[g_{1i}(\hat{\sigma}_u^2(\mathrm{y})) + g_{2i}(\hat{\sigma}_u^2(\mathrm{y}))\right] - \bar{g}_{1i}^{PB} - \bar{g}_{2i}^{PB}$ of $g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2)$ to this expression. Notice that for distributions such that the cross-product expectation in (10) is of order

$O(1/m)$, the MSE estimator (13) has bias of order $O(1/m^2)$.

From simulation studies carried out so far we find that the two bootstrap estimators defined by (11) (with the addition of (12) for the non-normal cases) and (13) perform almost identically even for small samples.

2.3. *Nonparametric Bootstrap*

The nonparametric bootstrap uses the original estimates of β and σ_u^2 in order to generate bootstrap replications of estimated standardized combined error terms $(u_i + e_i)$. The method consists of the following steps:

1. Calculate estimated standardized residuals,

$$\hat{r}_i = (y_i - \mathrm{x}_i^T \hat{\beta}) / c_i^{1/2},\ i = 1...m, \qquad (15)$$

where

$$c_i = (\hat{\sigma}_u^2 + \psi_i) - \{\mathrm{x}_i^T [\sum_i \frac{1}{(\hat{\sigma}_u^2 + \psi_i)} \mathrm{x}_i \mathrm{x}_i^T]^{-1} \mathrm{x}_i\},$$

$\hat{\sigma}_u^2 = \hat{\sigma}_u^2(\mathrm{y})$ and $\hat{\beta} = \hat{\beta}[\mathrm{y}; \hat{\sigma}_u^2(\mathrm{y})]$.

2. Select a large number B of sets of standardized residuals r^b ($b = 1, \ldots, B$), where each set $r^b = (r_1^b, \ldots, r_m^b)$ is a simple random sample with replacement from the empirical standardized residuals $\hat{r}_i$, $i = 1, \ldots, m$ defined by (15).

3. Calculate the bootstrap direct estimators, $y_i^b = r_i^b c_i^{1/2} + \mathrm{x}_i^T \hat{\beta}$; $i = 1, \ldots, m$, $b = 1, \ldots, B$.

4. Re-estimate the hyper-parameters σ_u^2 and β for each of the bootstrap samples obtained in (3) using the same method as used for the original sample, yielding the estimates $\hat{\sigma}_u^2(\mathrm{y}^b)$, $\hat{\beta}[\mathrm{y}^b; \hat{\sigma}_u^2(\mathrm{y}^b)]$ and $\hat{\beta}[\mathrm{y}^b; \hat{\sigma}_u^2(\mathrm{y})]$.

5. Estimate the MSE of the EBLUP as,

$$\mathrm{M\hat{S}E}(\hat{\theta}_i^{EBLUP}) = 2\{g_{1i}[\hat{\sigma}_u^2(\mathrm{y})] + g_{2i}[\hat{\sigma}_u^2(\mathrm{y})]\} - \bar{g}_{1i}^{NPB} - \bar{g}_{2i}^{NPB} + puc^{NPB} \qquad (16)$$

where puc^{NPB} and $\bar{g}_{di}^{NPB}$, d=1,2 are defined similarly to (11).

<u>Comment</u>: The estimator (16) is essentially the same as (11). Notice, however, that by bootstrapping the estimated standardized residuals it is no longer possible to generate 'true' bootstrap area means θ_i^b and hence to compute a non-parametric MSE estimator that is equivalent to (13). Also, like the parametric estimator (11), the estimator (16) assumes implicitly that the cross product expectation in (10) vanishes, or that the two error terms of the F-H model are normally distributed.

3. Other MSE Estimators Proposed in the Literature

3.1. *Methods based on Taylor approximations*

Prasad and Rao (1990) show that the MSE of the EBLUP based on an estimator $\hat{\sigma}_u^2$ can be approximated as,

$$\mathrm{MSE}[\hat{\theta}_\mathrm{i}^{\mathrm{EBLUP}}(\hat{\sigma}_u^2)] = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + g_{3i}(\sigma_u^2) Var(\hat{\sigma}_u^2) \qquad (17)$$

where $g_{3i}(\sigma_u^2) = \psi_i^2(\psi_i + \sigma_u^2)^{-3}$. For the case where σ_u^2 is estimated by the method of moments estimator $\hat{\sigma}_{PR}^2$, the Prasad-Rao MSE estimator with bias of order $o(1/m)$ is,

$$\mathrm{M\hat{S}E}[\hat{\theta}_\mathrm{i}^{\mathrm{EBLUP}}(\hat{\sigma}_{PR}^2)] = g_{1i}(\hat{\sigma}_{PR}^2) + g_{2i}(\hat{\sigma}_{PR}^2) + 2 g_{3i}(\hat{\sigma}_{PR}^2) V_{PR} \qquad (18)$$

where,

$V_{PR} = Var(\hat{\sigma}_{PR}^2) \cong 2m^{-2} \sum_{i=1}^m (\sigma_u^2 + \psi_i)^2$.

Datta, Rao and Smith (2005) consider the case where σ_u^2 is estimated by the F-H estimator. The authors develop a MSE estimator with bias of order $o(1/m)$ that has the form,

$$\mathrm{M\hat{S}E}[\hat{\theta}_\mathrm{i}^{\mathrm{EBLUP}}(\hat{\sigma}_{FH}^2)] = g_{1i}(\hat{\sigma}_{FH}^2) + g_{2i}(\hat{\sigma}_{FH}^2) + 2 g_{3i}(\hat{\sigma}_{FH}^2) V_{FH} - g_{4i}(\hat{\sigma}_{FH}^2) \qquad (19)$$

where

$V_{FH} = Var(\hat{\sigma}_{FH}^2) \cong 2m[\sum_{i=1}^m (\sigma_u^2 + \psi_i)^{-1}]^{-2}$

and

$$g_{4i}(\hat{\sigma}_{FH}^2) = 2\left[1 - \gamma_i(\hat{\sigma}_{FH}^2)\right]^2 \times \qquad (20)$$
$$\left[m \sum_{i=1}^m v_i^{-2} - [\sum_{i=1}^m v_i^{-1}]^2\right] \times \left[\sum_i v_i^{-1}\right]^{-3};$$

$v_i = (\psi_i + \hat{\sigma}^2_{FH})$.

3.2. *Jackknife MSE estimators*
An alternative approach for estimating the MSE of the EBLUP is the use of Jackknife procedures. Jiang, Lahiri and Wan (2002) proposed a jackknife method for empirical best prediction (minimum MSE), which includes the Fay-Herriot model as a special case. Recall that under normality of the error terms $\hat{\theta}_i^{EBLUP}$ is the empirical best predictor of θ_i. Similarly to (9),

$$\hat{\theta}_i^{EBLUP} - \theta_i = (\hat{\theta}_i^{BP} - \theta_i) + (\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BP}) \quad (21)$$

where $\hat{\theta}_i^{BP}$ is defined by (2) (with known β and σ_u^2). Hence, by (3), the MSE of $\hat{\theta}_i^{EBLUP}$ is decomposed as

$$\text{MSE}(\hat{\theta}_i^{EBLUP}) = g_{1i}(\sigma_u^2) + E(\hat{\theta}_i^{EBLUP} - \hat{\theta}_i^{BP})^2. \quad (22)$$

Utilizing this decomposition, the Jiang, Lahiri and Wan jackknife estimator is given by

$$M\hat{S}E(\hat{\theta}_i^{EBLUP}) = g_{1i}(\hat{\sigma}_u^2(y)) - \frac{m-1}{m}\sum_{j=1}^{m}\left[g_{1i}(\hat{\sigma}_u^2(y_{-j})) - g_{1i}(\hat{\sigma}_u^2(y))\right] + \frac{m-1}{m}\sum_{j=1}^{m}\left\{ \hat{\theta}_i[y;\hat{\sigma}_u^2(y_{-j}),\hat{\beta}(y_{-j};\hat{\sigma}_u^2(y_{-j}))] - \hat{\theta}_i[y;\hat{\sigma}_u^2(y),\hat{\beta}(y;\hat{\sigma}_u^2(y))]\right\}^2, \quad (23)$$

where y_{-j} is the vector of observations after deleting the j[th] area direct estimator and

$$\hat{\beta}(y_{-j};\hat{\sigma}_u^2(y_{-j})) = \left[\sum\nolimits_{i\neq j}[\psi_i + \hat{\sigma}_u^2(y_{-j})]^{-1}\mathrm{x}_i\mathrm{x}_i^T\right]^{-1} \times\left[\sum\nolimits_{i\neq j}[\psi_i + \hat{\sigma}_u^2(y_{-j})]^{-1}\mathrm{x}_i y_i\right].$$

Chen and Lahiri (2003) utilize the decomposition (10) for developing the following jackknife estimator:

$$M\hat{S}E(\hat{\theta}_i^{EBLUP}) = g_{1i}(\hat{\sigma}_u^2(y)) + g_{2i}(\hat{\sigma}_u^2(y)) - \frac{m-1}{m}\sum_{j=1}^{m}\Big[g_{1i}(\hat{\sigma}_u^2(y_{-j})) + g_{2i}(\hat{\sigma}_u^2(y_{-j})) - g_{1i}(\hat{\sigma}_u^2) - g_{2i}(\hat{\sigma}_u^2)\Big] \quad (24)$$
$$+ \frac{m-1}{m}\sum_{j=1}^{m}\left\{ \hat{\theta}_i[y;\hat{\sigma}_u^2(y_{-j}),\hat{\beta}(y;\hat{\sigma}_u^2(y_{-j}))] - \hat{\theta}_i[y;\hat{\sigma}_u^2(y),\hat{\beta}(y;\hat{\sigma}_u^2(y))]\right\}^2$$

For the case where the estimator (24) is negative, the authors recommend replacing the third term of the right hand side of (24) by $\psi_i^2[\psi_i + \hat{\sigma}_u^2(y)]^{-3}\hat{V}(\hat{\sigma}_u^2)$ where $\hat{V}(\hat{\sigma}_u^2) = \frac{m-1}{m}\sum_{j=1}^{m}[\hat{\sigma}_u^2(y_{-j}) - \hat{\sigma}_u^2(y)]^2$ is the jackknife variance estimator of $\hat{\sigma}_u^2$. They also approximate the fourth term on the right hand side of 24 by $\psi_i^2[\psi_i + \hat{\sigma}_u^2(y)]^{-4}(y_i - x_i^T\hat{\beta})\hat{V}(\hat{\sigma}_u^2)$ where $\hat{\beta} = \hat{\beta}[y;\hat{\sigma}_u^2(y)]$. Thus, the alternative jackknife MSE estimator of Chen and Lahiri (2003) is given by,

$$M\hat{S}E(\hat{\theta}_i^{EBLUP}) = g_{1i}(\hat{\sigma}_u^2(y)) + g_{2i}(\hat{\sigma}_u^2(y)) + \left[\frac{\psi_i^2}{(\psi_i + \hat{\sigma}_u^2(y))^3} + \frac{\psi_i^2}{(\psi_i + \hat{\sigma}_u^2(y))^4}(y_i - x_i^T\hat{\beta})\right] \times\ \hat{V}(\hat{\sigma}_u^2). \quad (25)$$

<u>Comment</u>: For normal error terms (u_i, e_i), the three jackknife MSE estimators defined by (23)-(25) have bias of order $o(1/m)$. All these estimators assume that the cross product expectation is zero and thus rely on the normality of the random effects and the sampling errors.

4. Simulation study

In order to assess the finite-sample performance of the bootstrap MSE estimators described in Section 2 and compare it to the performance of the estimators described in Section 3, we conducted a Monte Carlo simulation experiment. The experiment consists of generating a large number of samples of direct estimators from the Fay-Herriot model defined by (1) and computing the EBLUP and the corresponding MSE estimators for each sample. Following Datta, Rao and Smith (2005), we used for convenience a model with no auxiliary variables, such that $x_i^T\beta = \mu$, with μ =0 (but assumed unknown and hence estimated for every sample). The sample consists of the direct estimators in 15 small areas, divided into 5 groups of 3 areas, with different sampling error variances in different groups. The five sampling error variances are, $\psi_i = 2.0, 0.6, 0.5, 0.4, 0.2$. The random effect variance is $\sigma_u^2 = 1$, so that the values ψ_i define also the ratios ψ_i / σ_u^2, which determine the extent of shrinkage of the small area predictors. (See the definition of the shrinkage factor γ_i below equation 2. The values of ψ_i and σ_u^2 are the same as in Datta, Rao and smith, 2005). Since small areas within the same group are exchangeable, all the simulation results are reported as averages over the 3 areas in each group.

We consider three different combinations of distributions for the random effects u_i and the sampling errors e_i,

1- The u_i`s and the $e_i$`s are sampled from normal distributions, as in the original F-H model,

2- The u_i`s are sampled from the location exponential distribution and the $e_i$`s are sampled from the normal distribution,

3- The u_i`s and the $e_i$`s are sampled from location exponential distributions.

The location exponential distributions were defined such that the variances are the same as the variances of the corresponding normal distributions; $E(-1,1)$ for the random effects distribution and $E(-1,\sqrt{\psi_i})$ for the distributions of the sampling errors. The second and third combinations of distributions have been considered in order to study the robustness of the various MSE estimators to possible deviations from the normality assumptions.

We first generated 50,000 sets of direct estimators for each case and computed the EBLUP using both $\hat{\sigma}_{PR}^2$ and $\hat{\sigma}_{FH}^2$ for estimating the variance σ_u^2. This enables us to compute the 'true' MSE of the corresponding two EBLUP estimators. Next we generated 10,000 new samples of direct estimators for each case and computed for each sample the Prasad-Rao MSE estimator (Eq.18), the estimator of Datta, Rao and Smith (Eq. 20), the three jackknife estimators (Eqs. 23, 24 and 25), and the parametric and nonparametric bootstrap estimators described in Section 3 (Eqs. 11 and 16 respectively). As it turned out, the Jackknife estimator defined by Eq. 25 (hereafter, JK-ACL) outperforms the other two Jackknife estimators in every case and hence we only report the results obtained for this estimator. The bootstrap estimators are based on $B = 500$ replications. As mentioned in Section 2, the estimators $\hat{\sigma}_{PR}^2$ and $\hat{\sigma}_{FH}^2$ can be zero, in which case we followed Chen and Lahiri (2003) and replaced the three jackknife MSE estimates by $g_2(\hat{\sigma}_u^2)$. (No changes have been made to the other estimators when σ_u^2 was estimated by zero).

Tables 1-3 show the true MSE and the percent relative bias of the MSE estimators based on Taylor approximations (the Prasad Rao estimator when $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$, and the estimator of Datta, Rao and Smith when $\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$), the parametric and nonparametric bootstrap estimators, the 'naïve' parametric bootstrap estimator mse^{BP} defined below (13) and the jackknife estimator JK-ACL (Eq. 25). Tables 4-6 show the corresponding percent relative root MSE of the MSE estimators. The results are shown separately for each of the sampling variances ψ_i, distinguishing between the case where $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$ and the case where $\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$. When the error terms are sampled from the exponential distributions, we also show the results obtained for the parametric bootstrap estimator and the naïve bootstrap estimator that use the correct distributions.

Table 1 shows the results for the case where the distributions of the two error terms are normal. This is the case where all the methods, except for the naïve bootstrap estimator have bias of order $o(1/m)$ and indeed, the biases are generally low, except for the Taylor based estimator with $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$, where the bias increases quite drastically as the sampling variance decreases, similarly to the results in Datta, Rao and Smith (2005). As explained in Section 2, the naïve bootstrap estimator mse^{BP} has a relatively large negative bias in all the cases. All the estimators except mse^{BP} have very small biases for the case where $\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$ and the biases are generally smaller than the biases obtained when $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$, particularly for the smaller variances ψ_i.

Table 2 shows the results obtained when the random effects are sampled from the location exponential distribution but the sampling errors are sampled from the normal distributions. The results of the parametric bootstrap under the correct model, i.e., when generating the bootstrap samples using the correct distributions of the random effects and the sampling errors are denoted PB-NE and mse^{BP}-NE, whereas the results of the parametric bootstrap when the model is misspecified and the bootstrap samples are generated from normal distributions are denoted PB-NN and mse^{BP}-NN.

The results in Table 2 reveal that all the estimators except for the Taylor based estimator with $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$ perform relatively well in this case, despite the nonnormality of the random effects, although the biases are generally higher than in Table 1 where the random effects are generated from the normal distribution. The large biases observed for the Taylor based estimator with $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$ are somewhat surprising in view of the theoretical results developed by Lahiri and Rao (1995), that show that this MSE estimator has bias of order $o(1/m)$ even for the case where the random effects are generated from distributions other than normal, but notice that in this experiment we only use the direct estimators for 15 areas. The nonparametric bootstrap estimator and the jackknife estimator perform better than all the other estimators, but the jackknife estimator has a relatively large bias when $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$ and $\psi_i = 0.2$ (small sampling error variance). The parametric bootstrap estimators PB-NE and PB-NN have similar biases with relative bias of less than 6% in all the cases. The naïve bootstrap estimators mse^{PB}-NN and mse^{PB}-NE have again large negative biases.

Table 3 shows the results obtained for the case where both the random effects and the sampling errors are generated from the location exponential distributions. The results of the parametric bootstrap estimators under the correct model are denoted in this table as PB-EE and mse^{BP}-EE.

The relative biases in this table are much larger than in Tables 1 and 2 except in the case of the parametric bootstrap estimator PB-EE that uses the correct distributions for generating the bootstrap samples. Thus, all the other methods are sensitive to the deviation from normality of the sampling error distribution considered in the present experiment. The nonparametric bootstrap and the jackknife estimators again perform very similarly and in most cases better than the corresponding Taylor based estimators.

Tables 4-6 show the percent relative root MSE of the MSE estimators under the three scenarios of the distributions of the random effects and the sampling errors. For the case where both distributions are normal (Table 4), the Taylor based estimators have the lowest MSEs with the parametric bootstrap estimators coming next. The nonparametric bootstrap and the jackknife estimators perform generally similarly, but the jackknife estimator has a very large MSE when $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$ and $\psi_i = 0.2$, which is explained by its relatively large bias in this case (see Table 1).

The relative performance of the various estimators is similar in the case where the sampling errors have a normal distribution but the random effects are generated from the location exponential distribution (Table 5). The one profound exception is the naïve bootstrap estimator mse^{BP}-NE, which in this case has much lower MSEs than the other estimators in all the cases, despite having large negative biases (Table 2). This result is explained by the fact that the addition of bias correction terms to the naïve MSE estimator, $g_{1i}(\hat{\sigma}_u^2) + g_{2i}(\hat{\sigma}_u^2)$, underlying all the other methods generally increases the variances of the resulting estimators and hence the MSE of the MSE estimators. See, Singh *et. al* (1998) for similar findings and discussion. (The naïve estimator is obtained by substituting $\hat{\sigma}_u^2$ for σ_u^2 in the MSE of the BLUP, see Eq. 5. This estimator ignores the contribution to the MSE resulting from the use of an estimated variance, i.e., the use of the EBLUP instead of the BLUP.)

The naïve bootstrap estimator again performs best when the sampling errors are also generated from the location exponential distribution (Table 6) and $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$, but when $\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$ and the variance of the sampling errors gets small, the Taylor based estimator has the smallest MSEs. The parametric bootstrap estimator mse^{BP}-EE that uses the correct distributions has somewhat larger MSEs (but lower than the MSEs of mse^{BP}-NN), despite having the lowest biases in this case. As in the other tables, the nonparametric bootstrap and the jackknife estimators perform similarly, except for the case $\psi_i = 0.2$ and $\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$, where the jackknife estimator has a much larger MSE.

5. Summary Remarks

The results of this study indicate that the EBLUP that uses the estimator $\hat{\sigma}_{FH}^2$ for estimating the variance of the random effects has uniformly a lower MSE than the EBLUP that uses the estimator $\hat{\sigma}_{PR}^2$ (compare the true MSE in the various tables). On the other hand, no single method of MSE estimation dominates all the other methods in terms of bias and MSE (of the MSE estimators). When the sampling errors have a normal distribution, the nonparametric bootstrap and the Jackknife estimators have in general the smallest biases, with the parametric bootstrap method coming next. The Taylor based method works well in terms of bias when the random effect variance is estimated by $\hat{\sigma}_{FH}^2$, but not in the case where the variance is estimated by $\hat{\sigma}_{PR}^2$. We emphasize again, however, that our results so far are restricted to 15 small areas. When the distribution of the sampling errors is non-normal, all the MSE estimators have appreciable biases, except for the parametric bootstrap method that generates the bootstrap samples from the correct distributions.

Almost all the published studies on the estimation of the MSE of the EBLUP in small area estimation focus solely on the bias of the MSE estimators. We recognize the fact that analytical comparisons of the MSE of MSE estimators to the right order is not simple and may not even be feasible, but there is no reason why this important property of any given MSE estimator is not explored empirically. As our results indicate very clearly, a MSE estimator with negligible bias may actually have a much larger variance and hence a larger MSE than another estimator with an appreciable bias. For example, the nonparametric bootstrap and the jackknife estimators have much smaller biases than the Taylor based estimator of Prasad and Rao (1990) when the random effects variance is estimated by the method of moments and the model error terms are normal, (left hand side of Table 1), and yet, the Taylor based estimator has smaller MSEs except in the case of the smallest sampling error variance ($\psi_i = 0.2$, Table 4).

The present article explores also the effect of deviations from normality of the distributions of the model error terms on the performance of the MSE estimators. Lahiri and Rao (1995) show theoretically that the Prasad-Rao MSE estimator is robust to deviation from normality of the distribution of the random effects in the sense of preserving the order of the bias and the empirical results in the present study suggest that this is true also for the bootstrap methods and the jackknife estimator. However, all the methods except for the parametric bootstrap method that generates the samples from the correct distributions yield estimators with large biases when the distribution of the sampling errors is also non-normal. Clearly, the use of the latter estimator requires knowledge of the correct distribution of the sampling errors, which may not be available, but the results of this study suggest that more attention should be paid for the development of new procedures for the identification of the distribution of the sampling errors in small area estimation problems.

References

Chen, S. and Lahiri, P. (2003). A Comparison of different MSPE estimators of EBLUP for the Fay-Herriot model. *Proceedings of* the *Survey Research methods section of the ASA*, 905-911.

Datta, G.S., Rao, J.N.K., and Smith, D.D. (2005). On measures of uncertainty of small area estimators in the Fay-Herriot model. *Biometrika* (to appear).

Fay, R.E. and Herriot, R.A. (1979). Estimation of income from small places: an application of James-Stein Procedures to census data. *Journal of the American Statistical Association,* **74**, 269-277.

Harville, D. A. (1985). Decomposition of prediction error. *Journal of the American Statistical Association,* **80**, 132-138.

Jiang, J., Lahiri, P. and Wan, S. (2002). A unified jackknife method. *Annals of Statistics,* **30**, 1782-1810.

Lahiri, P. and Rao, J.N.K. (1995). Robust estimation of mean square error of small area estimators, *Journal of the American Statistical Association,* **90**, 758-766.

Prasad, N.G.N. and Rao, J.N.K. (1990). On estimation of the mean squared error of small-are estimators. *Journal of the American Statistical Association,* **85**, 163-171.

Pfeffermann, D. and Tiller, R. B. (2003). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. Methodology working Paper # MO3/05, Southampton Statistical Sciences Research Institute, Southampton, U.K.

Singh, A.C., Stukel, D. and Pfeffermann, D. (1998). Bayesian versus frequentist measures of error in small area estimation. *Journal of the Royal Statistical Society, Series B*, **60**, 377-396.

Table 1: True MSE of EBLUP and Percent Relative Bias of MSE Estimators Based on Taylor approximations (Taylor), Parametric Bootstrap (PB), Nonparametric Bootstrap (NPB) and Jackknife (JK-ACL). 15 Areas, Model Error terms generated from *normal distributions*

	$\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$					$\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$				
ψ_i	2.0	0.6	0.5	0.4	0.2	2.0	0.6	0.5	0.4	0.2
100*MSE	78.3	43.6	38.7	33.7	19.6	77.0	41.9	37.0	31.9	17.9
Taylor	0.2	7.3	9.4	11.2	34.2	-2.0	-0.0	0.5	-0.2	3.7
PB	-2.6	-2.8	-2.4	-3.1	0.5	-1.2	-0.6	-0.2	-1.0	1.8
mse^{BP}	-8.3	-10.2	-9.7	-10.4	-6.2	-6.1	-6.7	-6.3	-6.9	-3.6
NPB	0.0	-1.2	-0.8	-1.7	1.4	1.5	1.0	1.3	0.4	3.1
JK–ACL	-2.3	-1.1	-0.4	-0.8	6.3	-0.6	-0.3	0.0	-0.9	1.4

Table 2: True MSE of EBLUP and Percent Relative Bias of MSE Estimators Based on Taylor approximations (Taylor), Parametric Bootstrap (PB), Nonparametric Bootstrap (NPB) and Jackknife (JK-ACL). 15 Areas, Random Effects Generated from *location exponential distribution*, Sampling Errors generated from *normal distribution*.

	$\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$					$\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$				
ψ_i	2.0	0.6	0.5	0.4	0.2	2.0	0.6	0.5	0.4	0.2
100*MSE	73.2	41.1	36.6	31.8	19.2	73.4	39.7	34.9	30.0	17.2
Taylor	0.3	13.2	18.1	24.3	71.2	-4.8	-0.9	0.7	1.6	8.5
PB–NN	-4.6	-6.1	-5.7	-6.0	-4.3	-4.4	-3.0	-1.9	-1.9	0.0
mse^{BP}–NN	-8.8	-11.9	-11.5	-11.9	-9.9	-8.3	-8.2	-7.3	-7.3	-5.0
NPB	2.0	-0.3	0.2	-0.2	1.1	1.0	1.5	2.5	2.5	3.9
JK–ACL	0.0	0.8	2.1	1.7	9.8	0.4	1.5	2.3	2.0	3.9
PB–NE	-3.9	-5.9	-5.3	-5.8	-4.2	-3.3	-3.2	-2.2	-2.3	-0.5
mse^{BP}–NE	-12.1	-15.3	-14.7	-14.9	-12.4	-11.4	-12.8	-12.0	-11.8	-9.1

Table 3: True MSE of EBLUP and Percent Relative Bias of MSE Estimators Based on Taylor approximations (Taylor), Parametric Bootstrap (PB), Nonparametric Bootstrap (NPB) and Jackknife (JK-ACL). 15 Areas, Random Effects and Sampling Errors generated from *location exponential distributions*.

	$\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$					$\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$				
ψ_i	2.0	0.6	0.5	0.4	0.2	2.0	0.6	0.5	0.4	0.2
100*MSE	90.1	44.8	40.4	34.3	20.8	88.3	42.8	38.1	31.8	18.0
Taylor	-21.3	4.0	8.4	19.6	81.7	-22.2	-10.1	-9.5	-5.8	5.7
PB–NN	-26.1	-18.2	-18.9	-17.1	-15.7	-22.1	-12.6	-12.7	-10.2	-6.1
mse^{BP} –NN	-28.5	-22.2	-22.9	-21.3	-19.7	-25.0	-17.0	-17.1	-14.7	-10.5
NPB	-19.1	-11.6	-12.3	-10.4	-9.0	-16.0	-6.7	-6.9	-4.3	-0.3
JK–ACL	-20.1	-11.2	-11.5	-9.3	-4.7	-16.1	-7.6	-7.8	-5.4	-1.7
PB–EE	-6.6	-7.4	-8.8	-7.6	-7.6	-5.1	-2.1	-3.1	-1.4	0.8
mse^{BP} –EE	-13.3	-16.9	-18.3	-17.3	-16.6	-11.9	-11.5	-12.6	-11.1	-8.7

Table 4: True MSE and Percent Relative Root MSE of MSE Estimators Based on Taylor approximations (Taylor), Parametric Bootstrap (PB), Nonparametric Bootstrap (NPB) and Jackknife (JK-ACL). 15 Areas, Model Error terms generated from *normal distributions*

	$\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$					$\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$				
ψ_i	2.0	0.6	0.5	0.4	0.2	2.0	0.6	0.5	0.4	0.2
100*MSE	78.3	43.6	38.7	33.7	19.6	77.0	41.9	37.0	31.9	17.9
Taylor	39.5	20.6	20.0	22.0	58.6	36.9	20.3	17.8	14.5	9.4
PB	42.5	29.1	27.3	25.2	21.6	37.3	22.8	20.8	18.4	13.0
mse^{BP}	40.2	29.2	27.6	26.0	21.1	35.8	23.7	22.0	21.1	14.8
NPB	47.4	33.2	31.5	29.3	25.8	38.3	24.1	22.2	19.8	15.0
JK–ACL	46.6	29.4	28.6	29.4	59.9	40.5	24.3	22.5	20.1	15.2

Table 5: True MSE and Percent Relative Root MSE of MSE Estimators Based on Taylor approximations (Taylor), Parametric Bootstrap (PB), Nonparametric Bootstrap (NPB) and Jackknife (JK-ACL). 15 Areas, Random Effects Generated from *location exponential distribution*, Sampling Errors generated from *normal distribution*.

	$\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$					$\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$				
ψ_i	2.0	0.6	0.5	0.4	0.2	2.0	0.6	0.5	0.4	0.2
100*MSE	73.2	41.1	36.6	31.8	19.2	73.4	39.7	34.9	30.0	17.2
Taylor	51.7	30.7	30.8	33.7	76.9	49.1	29.7	26.9	23.3	17.8
PB–NN	55.1	39.0	37.0	34.6	29.5	49.8	33.0	30.9	28.3	21.7
mse^{BP} –NN	51.7	38.0	36.1	34.7	28.9	47.6	32.7	30.8	28.5	22.4
NPB	63.4	45.9	44.1	41.6	36.5	54.6	37.1	35.2	32.5	25.9
JK–ACL	63.5	42.0	41.3	40.0	68.8	58.0	37.7	35.1	31.7	24.0
PB–NE	52.3	39.0	37.4	35.3	30.1	51.1	38.3	36.8	35.0	30.9
mse^{BP} –NE	28.4	27.4	27.2	26.7	25.2	19.5	18.4	18.2	18.0	16.9

Table 6: True MSE and Percent Relative Root MSE of MSE Estimators Based on Taylor approximations (Taylor), Parametric Bootstrap (PB), Nonparametric Bootstrap (NPB) and Jackknife (JK-ACL). 15 Areas, Random Effects and Sampling Errors generated from *location exponential distributions*.

	$\hat{\sigma}_u^2 = \hat{\sigma}_{PR}^2$					$\hat{\sigma}_u^2 = \hat{\sigma}_{FH}^2$				
ψ_i	2.0	0.6	0.5	0.4	0.2	2.0	0.6	0.5	0.4	0.2
100*MSE	90.1	44.8	40.4	34.3	20.8	88.3	42.8	38.1	31.8	18.0
Taylor	47.9	9.3	14.1	37.0	197.4	49.3	30.5	26.8	21.2	10.7
PB–NN	55.6	45.0	43.3	41.5	36.6	50.7	37.1	35.1	32.8	26.4
mse^{BP} –NN	51.8	40.7	39.3	37.1	32.0	48.9	35.7	33.9	31.3	24.2
NPB	59.2	48.6	46.6	44.9	39.8	52.9	39.8	37.6	35.7	29.7
JK–ACL	63.5	44.8	43.7	43.5	69.6	59.7	39.8	37.3	34.6	27.5
PB–EE	46.1	42.9	41.2	40.3	34.2	48.2	35.3	33.2	31.5	26.8
mse^{BP} –EE	42.2	36.6	35.7	34.5	31.6	42.5	31.3	30.0	28.1	23.0