Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2006 : Methodological Issues in Measuring Population Health

2006

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2006
Methodological Issues in Measuring Population Health

Robust Mean Squared Prediction Error Estimators of EBLUP of a Small Area Total Under the Fay-Herriot Model

Shijie Chen, P. Lahiri, J.N.K. Rao[1]

Abstract

In this paper we derive a second-order unbiased (or nearly unbiased) mean squared prediction error (MSPE) estimator of empirical best linear unbiased predictor (EBLUP) of a small area total for a non-normal extension to the well-known Fay-Herriot model. Specifically, we derive our MSPE estimator essentially assuming certain moment conditions on both the sampling and random effects distributions. The normality-based Prasad-Rao MSPE estimator has a surprising robustness property in that it remains second-order unbiased under the non-normality of random effects when a simple method-of-moments estimator is used for the variance component and the sampling error distribution is normal. We show that the normality-based MSPE estimator is no longer second-order unbiased when the sampling error distribution is non-normal or when the Fay-Herriot moment method is used to estimate the variance component, even when the sampling error distribution is normal. It is interesting to note that when the simple method-of moments estimator is used for the variance component, our proposed MSPE estimator does not require the estimation of kurtosis of the random effects. Results of a simulation study on the accuracy of the proposed MSPE estimator, under non-normality of both sampling and random effects distributions, are also presented.

KEY WORDS: Mean Squared Prediction Errors; Linear Mixed Model; Variance Components.

1. Introduction

In the context of estimating per-capita income for small places (population less than 1000) from the 1970 Census of Population and Housing, Fay and Herriot (1979) used a small area regression model and demonstrated that the resulting empirical best linear unbiased predictor (EBLUP) have smaller average error than either the traditional survey estimators or an alternative method using county averages. Let Y_i be a direct estimator of the ith small area total θ_i and $x_i = (x_{i1}, \cdots, x_{ip})'$ be a $p \times 1$ vector of associated predictor variables, $i = 1, \cdots, m.$ The Fay-Herriot model may be written as $Y_i = \theta_i + e_i$ and $\theta_i = x_i'\beta + v_i,$ or as a linear mixed model: $Y_i = x_i'\beta + v_i + e_i,$ where the sampling errors $\{e_i\}$ and the random effects $\{v_i\}$ are independent with $e_i \overset{ind}{\sim} N(\theta_i, D_i)$ and $v_i \overset{ind}{\sim} N(0, \psi), i = 1, \cdots, m.$ The sampling variances D_i are assumed to be known, but β and ψ are to be estimated from the data $(Y_i, x_i), i = 1, \cdots, m.$ In practice, D_i's are externally estimated using the generalized variance function (GVF) method; see Wolter (1985), Fay and Herriot (1979), Bell and Otto (1995), among others. The Fay-Herriot (FH) model has been used extensively in small area estimation and related problems for a variety of reasons, including its simplicity, its ability to protect confidentiality of microdata and its ability to produce design-consistent estimators [see Rao's (2003) book, Chapter 7].

While the empirical best linear unbiased predictor (EBLUP) of a small area total or mean under the FH model is easy to obtain, an accurate estimation of its mean square prediction error (MSPE) is a challenging problem. A naive MSPE estimator is given by the MSPE of the best linear unbiased predictor (BLUP) with the model variance ψ replaced by a suitable estimator. But, it usually underestimates the true MSPE of EBLUP mainly for two reasons. First, it fails to incorporate the extra variability incurred due to the estimation of ψ and the order of this underestimation is $O(m^{-1}),$ for large m. Secondly, the naive MSPE estimator underestimates even the true MSPE of the BLUP, the order of the underestimation being $O(m^{-1})$. Prasad and Rao (1990) demonstrated the importance of accounting for these two sources of underestimation, and using a Taylor linearization method produced a second-order unbiased (or nearly unbiased) MSPE estimator of EBLUP when

[1] Shijie Chen, RTI International; P. Lahiri, University of Maryland, College Park; J.N.K. Rao, Carelton University

the variance component is estimated by a simple method-of-moments. The bias of that MSPE estimator is of order $o(m^{-1})$. The derivation of their MSPE estimator involves essentially two main steps. First a second-order correct MSPE expansion is obtained by neglecting all terms of order $o(m^{-1})$. The second step involves the estimation of this second-order correct MSPE approximation such that bias is of lower order, i.e. $o(m^{-1})$. Datta and Lahiri (2000) extended the Prasad-Rao MSPE estimator to cover different methods of estimating ψ. Datta, Rao and Smith (2005) obtained a nearly unbiased MSPE estimator when ψ is estimated by the Fay-Herriot (1979) method-of-moments. Das, Jiang and Rao (2004) generalized the Taylor method to general linear mixed models and obtained nearly unbiased MSPE estimators.

Note that the derivation of EBLUP, using a moment estimator of ψ, does not require normality assumption. However, for the estimation of MSPE, Prasad and Rao (1990), Datta and Lahiri (2000), Das, Jiang and Rao (2004), Datta, Rao and Smith (2005) and others used the normality assumption. One exception is the paper by Lahiri and Rao (1995) who assumed normality of the sampling errors $\{e_i\}$, but replaced the normality of the random effects $\{v_i\}$ by certain moment conditions. They showed that the normality-based Prasad-Rao estimator of the MSPE of EBLUP remains second-order unbiased under this non-normal set-up when the simple method-of-moments estimator of ψ is used. This is indeed a surprising result, demonstrating the robustness of the Prasad-Rao MSPE estimator under unspecified non-normality of the random effects $\{v_i\}$. Does this result hold when the sampling errors $\{e_i\}$ are non-normal or when ψ is estimated by moment estimator of ψ proposed by Fay and Herriot (1979)?.

In Section 2, we briefly review empirical best linear unbiased prediction. In Section 3, we obtain a second-order approximation to the MSPE of EBLUP without normality assumption. In Section 4, we propose a second-order (or nearly) unbiased MSPE estimator of EBLUP, using the approximation to MSPE obtained in Section 3. Finally, some empirical results from a small simulation study are reported in Section 5. Proofs of our results are omitted, for simplicity.

2. EBLUP

The Fay-Herriot model, without the normality assumption, may be written as $Y_i = \theta_i + e_i$ and $\theta_i = x_i'\beta + v_i, i = 1,\cdots,m,$ or as a linear mixed model,

$$Y_i = x_i'\beta + v_i + e_i\ , \quad i = 1,\cdots,m\ , \tag{1}$$

where the sampling errors $\{e_i\}$ and the random effects $\{v_i\}$ are uncorrelated with $e_i \sim [0, D_i, \kappa_{ei}]$ and $v_i \sim [0, \psi, \kappa_v]$, $[\mu, \sigma^2, \kappa]$ representing a probability distribution with mean μ, variance σ^2 and kurtosis κ. We define kurtosis of a distribution as $\kappa = \mu_4/\sigma^4 - 3,$ where σ^2 and μ_4 are the variance and the fourth central moment of the distribution respectively. We assume that $[\beta, \psi, \kappa_v]$ is unknown, but $[D_i, \kappa_{ei}]$ is known.

Define $X' = (x_1,\cdots,x_m)$, and $\Sigma(\psi) = diag\{\psi + D_j; j = 1,...,m\}$ and β can be estimated by $\hat{\beta}(\psi) = [X'\Sigma^{-1}(\psi)X]^{-1}X'\Sigma^{-1}(\psi)Y,$ a weighted least square estimator of β for a given ψ. The BLUP of θ_i under the FH model (1) is given by:

$$\hat{\theta}_i(Y_i;\psi) = B_iY_i + (1-B_i)x_i'\hat{\beta}(\psi),$$

where $B_i = \frac{\psi}{D_i+\psi}, i = 1,...,m.$ An EBLUP of θ_i is then obtained as

$$\hat{\theta}_i(Y_i;\hat{\psi}) = \hat{B}_iY_i + (1-\hat{B}_i)x_i'\hat{\beta}(\hat{\psi}) =: \hat{\theta}_i,$$

where $\hat{B}_i = \frac{\hat{\psi}}{D_i+\hat{\psi}}, i = 1,...,m,$ and $\hat{\psi}$ is a moment estimator of ψ. Note that the BLUP and EBLUP, based on $\hat{\psi}$, do not require normality of $\{e_i\}$ and $\{v_i\}$.

Prasad and Rao (1990) proposed the following simple method-of-moments estimator of ψ:

$$\hat{\psi}_{PR} = max\left\{0, (m-p)^{-1}\sum_{j=1}^{m}\{(Y_u - x_u^{'}\hat{\beta}_{\text{OLS}})^2 - (1-h_{jj})D_j\}\right\},$$

where $\hat{\beta}_{\text{OLS}} = (X^{'}X)^{-1}X^{'}Y,$ the ordinary least square estimator of $\beta,$ and $h_{jj} = x_j^{'}(X^{'}X)^{-1}x_j,$ the leverage for the j th small area $(j = 1,\cdots,m).$ Fay and Herriot (1979) obtained a different moment estimator $\hat{\psi}_{FH},$ by solving the following equation iteratively for ψ :

$$\frac{1}{m-p}Y'Q(\psi)Y - 1 = 0, \quad (2)$$

where $Y'Q(\psi)Y = \sum_{j=1}^{m}(D_j+\psi)^{-1}\{Y_j - x_j^{'}\hat{\beta}(\psi)\}^2$ is the weighted residual sum of squares. Pfeffermann and Nathan (1981) proposed a similar moment method in the context of regression analysis of survey data. For the special case $D_i = D\ (i = 1,\cdots,m)$, $\hat{\psi}_{PR} = \hat{\psi}_{FH}$. In this paper, we focus on EBLUP based on $\hat{\psi}_{PR}$ or $\hat{\psi}_{FH}$. The estimators $\hat{\psi}_{PR}$ and $\hat{\psi}_{FH}$ are typically consistent estimators for large m under the following regularity conditions:

- $(r.1)\ 0 < D_L \le D_j \le D_U < \infty,\ j = 1,\cdots,m,$
- $(r.2)\ \sup_{j\ge1} h_{jj} = O(\frac{1}{m}).$

Under non-normality and regularity conditions, the bias of $\hat{\psi}_{PR}$ is of order $o(m^{-1}).$ However, unless $D_i = D\ (i = 1,\cdots,m)$, the bias of $\hat{\psi}_{FH}$ is of order $O(m^{-1}),$ even under normality, and it is given by

$$E[\hat{\psi}_{FH} - \psi] = b(\hat{\psi}_{FH};\psi,\kappa_v) + o(m^{-1}), \quad (3)$$

where

$$b(\hat{\psi}_{FH};\psi,\kappa_v) = b_N(\hat{\psi}_{FH};\psi) + \alpha(\hat{\psi}_{FH};\psi,\kappa_v)$$

$$b_N(\hat{\psi}_{FH};\psi) = \frac{2[m\text{tr}(\Sigma^{-2}) - \{\text{tr}(\Sigma^{-1})\}^2]}{[\text{tr}(\Sigma^{-1})]^3}$$

$$\alpha(\hat{\psi}_{FH};\psi,\kappa_v) = \frac{[\text{tr}(\Sigma^{-2})]^2 - \text{tr}(\Sigma^{-3})\text{tr}(\Sigma^{-1})}{[\text{tr}(\Sigma^{-1})]^3}\psi^2\kappa_v + \frac{\text{tr}(D^2\Phi\Sigma^{-2})\text{tr}(\Sigma^{-2}) - \text{tr}(\Sigma^{-1})\text{tr}(D^2\Phi\Sigma^{-3})}{[\text{tr}(\Sigma^{-1})]^3},$$

$$D = Diag\{D_j; j = 1,\cdots,m\},$$

$$\Phi = Diag\{\kappa_{ej}; j = 1,\cdots,m\}.$$

In the above, $b_N(\hat{\psi}_{FH};\psi)$ denotes the bias of $\hat{\psi}_{FH}$ up to order $O(m^{-1})$ under normality (see Datta, Rao and Smith, 2005) and $\alpha(\hat{\psi}_{FH};\psi,\kappa_v)$ is the additional non-normality effect on the bias.

The variance of $\hat{\psi}$ may be expressed as

$$var(\hat{\psi}) = var_N(\hat{\psi}) + \eta(\hat{\psi};\psi,\kappa_v), \quad (4)$$

where $var_N(\hat{\psi})$ denotes the variance of $\hat{\psi}$ up to order $O(m^{-1})$ under normality, and $\eta(\hat{\psi},\psi,\kappa_v)$ denotes the additional non-normality effect on the variance. From Datta, Rao and Smith (2005), we have

$$var_N(\hat{\psi}_{PR}) = 2m^{-2}\sum_{j=1}^{m}(\psi + D_j)^2 = 2m^{-2}\text{tr}(\Sigma^{-2}),$$

$$var_N(\hat{\psi}_{FH}) = 2m\left\{\sum_{j=1}^{m}(\psi + D_j)^{-1}\right\}^{-2} = 2m\left\{\text{tr}(\Sigma^{-1})\right\}^{-2}.$$

It can be shown that

$$\eta(\hat{\psi}_{PR};\psi,\kappa_v) = m^{-1}\left\{\kappa_v\psi^2 + \frac{1}{m}\sum_{j=1}^{m}\kappa_{ej}D_j^2\right\} = m^{-1}\left\{\kappa_v\psi^2 + m^{-1}\mathrm{tr}(D^2\Phi)\right\}$$

$$\eta(\hat{\psi}_{FH};\psi,\kappa_v) = \left\{\sum_j(\psi+D_j)^{-1}\right\}^{-2}\sum_{j=1}^{m}\left\{(\psi+D_j)^{-2}[\kappa_v\psi^2+\kappa_{ej}D_j^2]\right\}$$

$$= [\mathrm{tr}(\Sigma^{-1})]^{-2}\left\{\mathrm{tr}(\Sigma^{-2})\kappa_v\psi^2 + \mathrm{tr}(D^2\Phi\Sigma^{-2})\right\}.$$

When both $\{e_i\}$ and $\{v_i\}$ are normal, we have $var(\hat{\psi}_{PR}) = var_N(\hat{\psi}_{PR}), var(\hat{\psi}_{FH}) = var_N(\hat{\psi}_{FH})$ and $var(\hat{\psi}_{FH}) \le var(\hat{\psi}_{PR})$ with equality for the balanced case $D_i = D, (i = 1,\cdots,m),$ see Datta, Rao and Smith (2005). It is interesting to note that the latter result does not extend to the non-normal situation. For the balanced case $D_i = D\ (i=1,\cdots,m),$ we have $b(\hat{\psi}_{PR};\psi,\kappa_v) = b(\hat{\psi}_{FH};\psi,\kappa_v) = 0$ and $var(\hat{\psi}_{PR}) = var(\hat{\psi}_{FH}) = 2m^{-1}(\psi+D)^2 + m^{-1}\{\kappa_v\psi^2 + m^{-1}D^2\mathrm{tr}(\Phi)\}$ simply because in this situation $\hat{\psi}_{PR} = \hat{\psi}_{FH}$.

3. Approximation to MSPE

The MSPE of the EBLUP $\hat{\theta}_i$ is given by $MSPE(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$, where the expectation is taken over the marginal distribution of Y under the non-normal Fay-Herriot model (1). The MSPE of the BLUP of $\hat{\theta}_i(Y_i,\psi)$ is not affected by non-normality and it is given by

$$MSPE[\hat{\theta}_i(Y_i,\psi)] = g_{1i}(\psi) + g_{2i}(\psi),$$

where

$$g_{1i}(\psi) = \frac{\psi D_i}{\psi + D_i},$$

$$g_{2i}(\psi) = \frac{D_i^2}{(\psi+D_i)^2}var[\hat{\beta}(\psi)] = \frac{D_i^2}{(\psi+D_i)^2}x_i'\left[X'\Sigma^{-1}(\psi)X]^{-1}\right]^{-1}x_i.$$

We are interested in approximating the MSPE of EBLUP under non-normality that accounts for the estimation of ψ and is second-order accurate, i.e accurate up to order $O(m^{-1})$.
We decompose the MSPE of EBLUP $\hat{\theta}_i$ as

$$\begin{aligned} MSPE[\hat{\theta}_i(Y_i,\hat{\psi})] &= MSPE[\hat{\theta}_i(Y_i,\psi)] + E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)]^2 \\ &+ 2E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)][\hat{\theta}_i(Y_i,\psi) - \theta_i]. \end{aligned} \quad (5)$$

where $\hat{\theta}_i(Y_i,\hat{\psi}) = \hat{\theta}_i$ and $\hat{\theta}_i(Y_i,\psi)$ is the BLUP. The cross-product term in (5) is zero under normality of $\{v_i\}$ and $\{e_i\}$; see Kacker and Harville (1984), but it is of $O(m^{-1})$ under non-normality and hence not negligible under the non-normal FH model (1). We obtain the following approximations to the last two terms of (5).

Result 1: Under the non-normal FH model (1) and regularity conditions (r.1), (r.2) and (r.3): $\sup_{j\ge1} E|v_i|^{8+\delta}, 0 < \delta < 1$, we have

$$(i) E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)]^2 = g_{3i}(\psi,\kappa_v) + o(m^{-1}),$$

$$(ii) E[\hat{\theta}_i(\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)][\hat{\theta}_i(Y_i,\psi) - \theta_i] = g_{4i}(\psi,\kappa_v) + o(m^{-1}),$$

where

$$g_{3i}(\psi,\kappa_v)=\frac{D_i^2}{(\psi+D_i)^3}var(\hat{\psi}),$$

$$g_{4i}(\psi,\kappa_v)=\frac{\psi D_i^2}{m(\psi+D_i)^3}\left[D_i\kappa_{ei}-\psi\kappa_v\right]c(\hat{\psi};\psi),$$

$c(\hat{\psi}_{PR};\psi)=1$ and $c(\hat{\psi}_{FH};\psi)=m(\psi+D_i)^{-1}\left\{\sum_j(\psi+D_j)^{-1}\right\}^{-1}.$

Thus, a second-order expansion to MSPE of EBLUP $\hat{\theta}_i$ is given by

$$\begin{aligned}&AMSPE_i\\&=g_{1i}(\psi)+g_{2i}(\psi)+g_{3i}(\psi,\kappa_v)+2g_{4i}(\psi,\kappa_v) \qquad (6)\\&=\frac{\psi D_i}{\psi+D_i}+\frac{D_i^2}{(\psi+D_i)^2}var[\hat{\beta}(\psi)]+\frac{D_i^2}{(\psi+D_i)^3}var(\hat{\psi})+\frac{2\psi D_i^2}{m(\psi+D_i)^3}\left[D_i\kappa_{ei}-\psi\kappa_v\right]c(\hat{\psi};\psi)\\&=AMSPE_{i,N}+\frac{D_i^2}{(\psi+D_i)^3}\eta(\hat{\psi};\psi,\kappa_v)+2g_{4i}(\psi,\kappa_v),.\end{aligned}$$

where $AMSPE_{i,N}$ is the normality-based MSPE approximation as given in Prasad and Rao (1990) and Datta, Rao and Smith (2005). The term $g_{3i}(\psi,\kappa_v)$ is the additional uncertainty due to the estimation of the variance component ψ and the term $2g_{4i}(\psi,\kappa_v)$ is needed to adjust for the non-normality. Under the regularity conditions, $g_{1i}(\psi)$ is the leading term [of order $O(1)$] and the remaining terms are all of order $O(m^{-1})$. Note that non-normality affects both $var(\hat{\psi})$ and the cross-product term $2E[\hat{\theta}_i(\hat{\psi},Y)-\hat{\theta}_i(\psi,Y)][\hat{\theta}_i(\psi,Y)-\theta_i]$. When both $\{e_i\}$ and $\{v_i\}$ are normal, the above approximation reduces to the Prasad-Rao (1990) approximation when $\hat{\psi}=\hat{\psi}_{PR}$ and the Datta-Rao-Smith (2005) approximation when $\hat{\psi}=\hat{\psi}_{FH}$. When the $\{e_i\}$ are normal and $\hat{\psi}=\hat{\psi}_{PR}$, the MSPE approximation (6) reduces to the Lahiri-Rao (1995) approximation.

4. Nearly Unbiased Estimator of MSPE

The second-order MSPE approximation $AMSPE_i$, given by (6), involves unknown parameters ψ and κ_v. Let $\hat{\kappa}_v$ be a consistent estimator of κ_v. Then $g_{2i}(\hat{\psi})$, $g_{3i}(\hat{\psi},\hat{\kappa}_v)$, and $g_{4i}(\hat{\psi},\hat{\kappa}_v)$ are second-order unbiased (or nearly unbiased) estimators of $g_{2i}(\psi)$, $g_{3i}(\psi,\kappa_v)$, and $g_{4i}(\psi,\kappa_v)$ respectively since latter functions of ψ and κ_v are already of order $O(m^{-1})$. However, estimation of the the leading term $g_{1i}(\psi)$ in (6) needs special attention since it is of the order $O(1)$.

Under regularity conditions (r.1) and (r.2), it can be shown that

$$E[g_{1i}(\hat{\psi})]=g_{1i}(\psi)-g_{3i}(\psi,\kappa_v)+g_{5i}(\psi,\kappa_v)+o(m^{-1}), \qquad (7)$$

where $g_{5i}(\psi,\kappa_v)=\frac{D_i^2}{(\psi+D_i)^2}b(\hat{\psi};\psi,\kappa_v)$. Using (6) and (7), a second-order unbiased (or nearly unbiased) MSPE estimator is then given by

$$mspe_i=g_{1i}(\hat{\psi})+g_{2i}(\hat{\psi})+2g_{3i}(\hat{\psi},\hat{\kappa}_v)+2g_{4i}(\hat{\psi},\hat{\kappa}_v)-g_{5i}(\hat{\psi},\hat{\kappa}_v). \qquad (8)$$

When $\hat{\psi}=\hat{\psi}_{PR}$, we have

$$mspe_i^{PR}=mspe_{i,N}^{PR}+\frac{2D_i^2}{m(\hat{\psi}+D_i)^3}\left[\hat{\psi}D_i\kappa_{ei}+\frac{1}{m}\sum_{j=1}^{m}\kappa_{ej}D_j^2\right]$$

$$= mspe_{i,N}^{PR} + \frac{2D_i^2}{m(\hat{\psi}+D_i)^3}\left[\hat{\psi}D_i\kappa_{ei} + m^{-1}\mathrm{tr}(D^2\Phi)\right], \quad (9)$$

where $mspe_{i,N}^{PR} = g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}) + \frac{2D_i^2}{m^2(\hat{\psi}+D_i)^3}\sum_{j=1}^{m}(\hat{\psi}+D_j)^2$ is the normality-based MSPE estimator first proposed by Prasad and Rao (1990). It is interesting to note that the MSPE estimator (9) does not require the estimation of κ_v although κ_v is involved in the MSE approximation (6). For normal $\{e_i\}$ but unspecified non-normal $\{v_i\}$, $mspe_i^{PR} = mspe_{i,N}^{PR}$ (Lahiri and Rao, 1995). It is interesting to note that the robustness of the Prasad-Rao MSPE estimator does not extend to the case when $\{e_i\}$ are non-normal. When $\kappa_{ej} > 0,\ j = 1,\cdots,m$, $mspe_i^{PR}$ overestimates whenever $mspe_{i,N}^{PR}$ overestimates.

When $\hat{\psi} = \hat{\psi}_{FH}$, we have

$$mspe_i^{FH} = mspe_i^{DRS} + \frac{2D_i^2}{(\hat{\psi}_{FH}+D_i)^3}\eta(\hat{\psi}_{FH};\hat{\psi}_{FH},\hat{\kappa}_v) + 2g_{4i}(\hat{\psi}_{FH},\hat{\kappa}_v) - \frac{D_i^2}{(\psi+D_i)^2}\alpha(\hat{\psi}_{FH};\hat{\psi}_{FH},\hat{\kappa}_v) \quad (10)$$

where

$$mspe_i^{DRS} = g_{1i}(\hat{\psi}_{FH}) + g_{2i}(\hat{\psi}_{FH}) + \frac{2D_i^2}{(\hat{\psi}_{FH}+D_i)^3}evar_N(\hat{\psi}_{FH}) - \frac{D_i^2}{(\hat{\psi}_{FH}+D_i)^2}b_N(\hat{\psi}_{FH};\hat{\psi}_{FH})$$

is the normality-based estimator proposed by Datta, Rao and Smith (2005). In the above $evar_N(\hat{\psi}_{FH})$ is obtained from $var_N(\hat{\psi}_{FH})$ using $\hat{\psi}_{FH}$ in place of ψ. When both $\{e_i\}$ and $\{v_i\}$ are normal, $mspe_i^{FH}$, given by (10), reduces to $mspe_i^{DRS}$. However, when $\{e_i\}$ are normal but not $\{v_i\}$, $mspe_i^{FH}$ is not identical to $mspe_i^{DRS}$ unless $D_i = D\ (i = 1,\cdots,m)$. Hence $mspe_i^{DRS}$, unlike $mspe_i^{PR}$, is not robust under non-normality of $\{v_i\}$, even when $\{e_i\}$ are normal. It is easy to check that for the balanced case $D_i = D\ (i = 1,\cdots,m),\ mspe_i^{FH} = mspe_i^{PR}$.

We now propose two different estimators of κ_v. To motivate the first estimator, first note that

$$E[\hat{\psi}_{FH}]^2 = Var[\hat{\psi}_{FH}] + [E(\hat{\psi}_{FH})]^2 \quad (11)$$

Now using the second-order approximations to $Var[\hat{\psi}_{FH}]$ and $E(\hat{\psi}_{FH})$, we get

$$E[\hat{\psi}_{FH}]^2 = \psi^2 + k(\psi) + l(\psi)\kappa_v + o(m^{-1}), \quad (12)$$

where

$$k(\psi) = 2m[\mathrm{tr}(\Sigma^{-1})]^{-2} + 2\psi b_N(\hat{\psi}_{FH};\psi) + \mathrm{tr}(D^2\Phi\Sigma^{-2})\left\{[\mathrm{tr}(\Sigma^{-1})]^{-2} + 2\psi\,\mathrm{tr}(\Sigma^{-2})[\mathrm{tr}(\Sigma^{-1})]^{-3}\right\}$$
$$-2\psi\,\mathrm{tr}(D^2\Phi\Sigma^{-3})[\mathrm{tr}(\Sigma^{-1})]^{-2}$$
$$l(\psi) = \psi^2[\mathrm{tr}(\Sigma^{-1})]^{-2}\left[\mathrm{tr}(\Sigma^{-2}) + 2\psi[\mathrm{tr}(\Sigma^{-1})]^{-1}\{[\mathrm{tr}(\Sigma^{-2})]^2 - \mathrm{tr}(\Sigma^{-1})\mathrm{tr}(\Sigma^{-3})\}\right]$$

To get a moment estimator of κ_v we solve

$$\hat{\psi}_{FH}^2 = \hat{\psi}_{FH}^2 + k(\hat{\psi}_{FH}) + l(\hat{\psi}_{FH})\kappa_v \quad (13)$$

for κ_v. The closed-form solution of κ_v is then given by $\hat{\kappa}_v = -k(\hat{\psi}_{FH})/l(\hat{\psi}_{FH})$ if $\hat{\psi}_{FH} > 0$ and 0 otherwise. Since this is a smooth function and $\hat{\psi}_{FH}$ is consistent for ψ, we have consistency of the estimator $\hat{\kappa}_v$.

To obtain the second estimator of κ_v, we replace ψ and $var(\hat{\psi}_{FH})$ in (4) by $\hat{\psi}_{FH}$ and $v_{WJ} = \sum_{u=1} w_u(\hat{\psi}_{FH,(-u)} - \hat{\psi}_{FH})^2$, a weighted jackknife estimator of $var(\hat{\psi}_{FH})$ considered by Chen and Lahiri (2006), where $\hat{\psi}_{FH,(-u)}$ is the Fay-Herriot estimator of ψ, using all but the u th small area data. The resulting equation is given by:

$$k^{å}(\hat{\psi}_{FH}) + l^{å}(\hat{\psi}_{FH})\kappa_v = 0, \tag{14}$$

where

$$k^{å}(\psi) = 2m + \mathrm{tr}(D^2\Phi\Sigma^{-2}) - \{\mathrm{tr}(\Sigma^{-1}\}^2 v_{WJ}$$
$$l^{å}(\psi) = \mathrm{tr}(\Sigma^{-2})\psi^2.$$

Solving (14) for κ_v, we obtain an alternate closed-form estimator of κ_v: $\hat{\kappa}_v^{å} = -k^{å}(\hat{\psi}_{FH})/l^{å}(\hat{\psi}_{FH})$ if $\hat{\psi}_{FH} > 0$ and 0 otherwise. Since this is a smooth function and $\hat{\psi}_{FH}$ and v_{WJ} are consistent for ψ and $var(\hat{\psi}_{FH})$ respectively, we have consistency of the estimator $\hat{\kappa}_v^{å}$.

5. Simulation Study

Finite-sample accuracy of the proposed estimator of MSPE, $mspe_i$, of EBLUP is investigated in this section through Monte Carlo simulation, for the special case $x_i'\beta = \mu$ and $D_j = D, \kappa_{ej} = \kappa_e \ (j = 1, \cdots, m)$. Noting that MSPE is translation invariant (i.e., it remains the same when Y_i is changed to $Y_i - \mu$), we set $\mu = 0$ without loss of generality. Further, we selected the following parameter values: $m = 30, 60$ and nine combinations of $\kappa_v = 0, 3, 6$ and $\kappa_e = 0, 3, 6.$ Note that $\kappa = 0, 3,$ and 6 correspond to normal, double exponential, and *shifted* exponential with mean zero respectively. We set $D = \psi = 1,$ leading to $B = .5.$
We generated 10,000 independent set of variates $\{v_i, e_i, i = 1, \cdots, m\}$ for each case with specified parameters. Simulated values of *MSPE*, *E*[*MSPE* estimator], and *MSE*[*MSPE* estimator] = *E*[*MSPE* estimator – *MSPE*]^2 were then computed from the 10,000 data sets $\{Y_i = v_i + e_i, i = 1, \cdots, m\}$ so generated, and averaged over the small areas. We compared three different MSPE estimators: naive, Prasad-Rao and the proposed based on $\hat{\psi}_{PR}$. Note that in the balanced case $\hat{\psi}_{PR} = \hat{\psi}_{FH}$ and we do not need to estimate κ_v.
Table 1 reports the percent relative biases (RB) for each MSPE estimator. The RB of a MSPE estimator is calculated as

RB = [average *E*(*MSPE* estimator) – average *MSPE*] / (average *MSPE*),

where the average is taken over all the small areas. Note that for the balanced case $D_i = D$, theoretically $E[mspe_1] = \ldots E[mspe_m]$ for any MSPE estimator $mspe_i$, and MSPE of the EBLUP estimator is the same across small areas. Hence, the use of RB given above is appropriate in the balanced case. For each MSPE estimator, Table 1 shows that the absolute RB decreases as the number of small areas, m, increases. For all the cases, the naive estimator leads to underestimation. Results for the Prasad-Rao MSPE estimator and the proposed MSPE estimator are almost identical and RB is negligible whenever the sampling errors $\{e_i\}$ are normally distributed; this is consistent with the previous theory (Lahiri and Rao, 1995). When the $\{e_i\}$ are non-normal, the Prasad-Rao MSPE estimator leads to underestimation, some times as large as 10% for $m = 30$ and interestingly the underestimation decreases slowly as m increases to 60. On the other hand, the proposed MSPE estimator corrects for the underestimation in all cases , but it leads to overestimation which decreases dramatically as m increases to 60.
Table 2 reports the percent relative root mean squared error (RRMSE) of the MSPE estimators. The RRMSE is calculated as

RRMSE = [average square root of *MSE*(*MSPE* estimator)] / (average *MSPE*),

where the average is taken over all the small areas. It shows that the proposed MSPE estimator performs the best, in terms of RRMSE, when the sampling errors are non-normal, while the RRMSE in the normal case are almost identical for the proposed and the Prasad-Rao MSPE estimators. The naïve MSPE estimator leads to relatively large RRMSE due to large squared bias.

6. Acknowledgement

S. Chen was supported in part by the Professional Development Award from the RTI International, RTP NC. J.N.K. Rao was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Finally, the authors wish to thank Ms. Huilin Li for computational support.

References

Bell, W.R. and Otto, M.C. (1995), "Sampling Error Modelling of Poverty and Income Statistics for States," In *Proceedings of the Section on Survey Research Methods,* Washington, D.C. American Statistical Association.

Chen, S. and Lahiri, P. (2006). On mean squared prediction error estimation in small area estimation problems, unpublished manuscript.

Das, K., Jiang, J. and Rao, J. N. K. (2004). Mean squared error of empirical predictor. *Ann. Statist.* 32, 818-840.

Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, 613-627.

Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, 92, 183-196.

Fay, R. E., and Herriot, R. A. (1979). Estimates of Income for Small Places: an Application of James-Stein Procedure to Census Data. *Journal of American Statistical Association* 74, 269-277.

Kackar, R.N., and Harville, D.A. (1984). Approximations for the Standard Errors of Estimation of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association* 79, 853-862.

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

Pfeffermann, D. and Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of American Statistical Association* 76, 681-689.

Prasad, N.G.N., and Rao, J.N.K. (1990). The Estimation of Mean Squared Error of Small Area Estimators. *Journal of American Statistical Association* 85, 163-171.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: Wiley.

Wolter, K. (1985). *Introduction to Variance Estimation* New York: Springer-Verlag.

Table 1. *Simulated values of percent relative bias (RB) of mean squared error estimators for* $\psi = D = 1$. *The variance component is estimated by the Prasad-Rao method of estimating moments.*

Distribution of e		Distribution of v: Normal		Double Exponential		Exponential	
		m=30	m=60	m=30	m=60	m=30	m=60
Normal	Naive	-12.1	-6.64	-6.66	-2.92	-12.29	-6.34
	Prasad-Rao	0.86	-0.11	0.53	0.54	1.61	0.54
	Proposed	0.86	-0.11	0.53	0.54	1.61	0.54
Double Exponential	Naive	-17.9	-10.67	-9.6	-4.96	-17.59	-10.82
	Prasad-Rao	-5.35	-4.22	-2.59	-1.52	-4.21	-4.16
	Proposed	5.76	1.1	1.82	0.42	8.27	1.53
Exponential	Naive	-22.13	-14.17	-11.91	-6.67	-22.65	-14.18
	Prasad-Rao	-9.86	-7.9	-4.85	-3.25	-9.69	-7.6
	Proposed	12.4	2.61	4.33	0.66	15.05	3.86

Table 2. *Simulated values of percent relative root mean square error (RRMSE) of the mean squared error estimators for* $\psi = D = 1$. *The variance component is estimated by the Prasad-Rao method of estimating moments.*

Distribution of e		Distribution of v: Normal		Double Exponential		Exponential	
		m=30	m=60	m=30	m=60	m=30	m=60
Normal	Naive	4.29	2.03	2.48	1.05	5.99	3.2
	Prasad-Rao	2.6	1.57	1.62	0.87	3.87	2.61
	Proposed	2.6	1.57	1.62	0.87	3.87	2.61
Double Exponential	Naive	5.95	2.95	3.04	1.34	7.59	4.2
	Prasad-Rao	3.22	2.13	1.83	1.04	4.47	3.21
	Proposed	1.77	1.57	1.14	0.86	2.61	2.39
Exponential	Naive	7.58	3.88	3.72	1.61	9.3	5.02
	Prasad-Rao	4.04	2.75	2.18	1.21	5.25	3.73
	Proposed	1.61	1.41	0.8	0.8	2.24	1.99