On mean squared prediction error estimation in small area estimation problems

Shijie Chen, RTI International
P. Lahiri, University of Maryland at College Park
Shijie Chen, STRD, Research Triangle Institute, 1615 M ST N.W. Suite 740, Washington, DC, 20036, USA (schen@rti.org)

Key Words: Mixed Linear Models, Small Area Estimation, Mean Squared Prediction Error, Empirical Best Linear Unbiased Predictor, Jackknife.

Abstract:

In this paper, we consider a Taylor-series approximation to the weighted jackknife mean squared prediction error (MSPE) of an empirical best linear unbiased predictor (EBLUP). Like the Taylor series method, this approximation provides a closed-form expression and saves computation. We compare various MSPE estimators using a Monte Carlo simulation study.

1. Introduction

For effective planning of health, social and other services, and for apportioning government funds, there is a growing demand to produce reliable estimates for smaller geographic areas and sub-populations, called small areas, for which adequate samples are not available. The usual design-based small area estimators are unreliable since they are based on a very few observations that are available from the area. An empirical best linear prediction (EBLUP) approach has been found suitable in many small area estimation problems. The method essentially uses an appropriate mixed linear model which captures various salient features of the sampling design and combines information from censuses or administrative records in conjunction with the survey data. For a review of small area estimation, see Rao (2003).

The estimation of MSPE of EBLUP is a challenging problem. The naive MSPE estimator, i.e., the MSPE of the BLUP with estimated model parameters, usually underestimates the true MSPE. There are two reasons for this underestimation problem. First, it fails to incorporate the extra variabilities incurred due to the estimation of various model parameters and the order of this underestimation is $O(m^{-1})$, where m is the number of the small areas. Secondly, the naive MSPE estimator even underestimates the true MSPE of the BLUP, the order of underestimation being $O(m^{-1})$. Several attempts have been made in the literature to account for these two sources of underestimation and to produce MSPE estimators that are correct up to the order $O(m^{-1})$. These are called second-order unbiased MSPE estimators.

Jiang, Lahiri and Wan (2002) proposed a jackknife method to estimate the MSPE of an empirical best predictor for a general situation. Bell (2001) pointed out that the Jiang-Lahiri-Wan jackknife MSPE estimator could take negative values in certain circumstances. However, Chen and Lahiri (2002, 2003) found that this is not a severe problem in their simulation studies and can be easily rectified by considering an alternative bias correction formula. For the well-known Fay-Herriot model, Chen and Lahiri (2003) provided an approximation to the jackknife MSPE estimator using a Taylor series approximation. Like the Prasad-Rao formula, this provides a closed-form formula. In this paper, we follow up on Chen and Lahiri (2003) and obtain the Taylor series approximation to the jackknife MSPE formula for a general case.

In section 2, we define the BLUP and EBLUP of a general mixed effect. We provide a Taylor series approximation to the jackknife method in section 3. In section 4, the method is illustrated using the simple but important Fay-Herriot model (see Fay and Herriot 1979). To demonstrate the efficiency of our proposed method, results from a Monte carlo simulation study are reported in section 5.

2. The BLUP and EBLUP

Consider the following general normal mixed linear model in small area estimation considered in Prasad and Rao (1990) and Datta and Lahiri (2000) :

$$y_i = X_i\beta + Z_i v_i + e_i,\ i = 1, ..., m, \qquad (1)$$

where X_i $(n_i \times p)$ and Z_i $(n_i \times b_i)$ are known matrices, v_i and e_i are independently distributed with

The research of S. Chen was supported in part by Professional Development Award from the RTI International, RTP NC.

$v_i \stackrel{ind}{\sim} N_{b_i}(0, G_i)$ and $e_i \stackrel{ind}{\sim} N_{n_i}(0, R_i)$, $i = 1, ..., m$. We assume that $G_i = G_i(\psi)$ $(b_i \times b_i)$ and $R_i = R_i\ (\psi)\ (n_i \times n_i)$ possibly depend on $\psi = (\psi_1, ..., \psi_q)'$, a $q \times 1$ vector of fixed variance components. Write $y = col_{1\leq i\leq m} y_i$, $e = col_{1\leq i\leq m} e_i$, $X = col_{1\leq i\leq m}(X_i)$, $Z = diag_{1\leq i\leq m}(Z_i)$, $G(\psi) = diag_{1\leq i\leq m} G_i, v = col_{1\leq i\leq m} v_i$ and $R(\psi) = diag_{1\leq i\leq m} R_i$. We assume that X has full column rank p. Let $\Sigma(\psi) = R(\psi) + ZG(\psi)Z'$, the variance-covariance matrix of y. Note that we can write (1) as

$$y \ = \ X\beta + Zv + e,$$

where v and e are independently distributed with $v \sim N_b(0, G)$, $e \sim N_n(0, R)$, $n = \sum_{i=1}^m n_i$ and $b = \sum_{i=1}^m b_i$.

As in Datta and Lahiri (2000), we are interested in predicting a general mixed effect of $\theta = h'\beta + \lambda' v$, where h and λ are known vectors of order $p \times 1$ and $b \times 1$ respectively. When ψ is known, the BLUP of θ is given by $\hat{\theta}(y; \psi) = h'\hat{\beta}(\psi) + s'(\psi)[y - X\hat{\beta}(\psi)]$, where $s(\psi) = \Sigma^{-1}(\psi)ZG(\psi)\lambda$ and $\hat{\beta}(\psi) = [X'\Sigma^{-1}(\psi)X]^{-1}[X'\ \Sigma^{-1}(\psi)y]$.

In practice ψ is unknown and is estimated from the data. Let $\hat{\psi}$ be a consistent estimator of ψ considered in Fay and Herriot (1979), Prasad and Rao (1990) and Datta and Lahiri (2000). Then an EBLUP of θ is $\hat{\theta}(y; \hat{\psi})$ which is obtained from $\hat{\theta}(y; \psi)$ with ψ replaced by $\hat{\psi}$.

3. A Taylor series approximation to the jackknife MSPE estimator

The MSPE of $\hat{\theta}(y; \hat{\psi})$ is defined as $MSPE[\hat{\theta}(y; \hat{\psi})] = E[\hat{\theta}(y; \hat{\psi}) - \theta]^2$, where E denotes the expectation with respect to model (1).

Chen and Lahiri (2002) proposed the following weighted jackknife MSPE estimator:

$$\begin{aligned} & mse^{WJ} \\ = \ & g_1(\hat{\psi}) + g_2(\hat{\psi}) \\ & - \sum_{u=1}^m w_u \left(g_1(\hat{\psi}_{-u}) + g_2(\hat{\psi}_{-u}) - [g_1(\hat{\psi}) + g_2(\hat{\psi})] \right) \\ & + \sum_{u=1}^m w_u [\hat{\theta}(y; \hat{\psi}_{-u}) - \hat{\theta}(y; \hat{\psi})]^2, \end{aligned} \qquad (2)$$

where $g_1(\psi) = \lambda' G(\psi)\lambda - s'(\psi)ZG(\psi)\lambda$, and $g_2(\psi) = [h - X's(\psi)]'(X'\Sigma^{-1}(\psi)X)^{-1}[h - X's(\psi)]$. The weights satisfy $w_u = 1 + O(m^{-1})$. Chen and Lahiri (2002) provided a specific weight $w_u = 1 - X_u'(X'X)^{-1}X_u$ - but note that it refers to the Fay-Herriot model.

We now consider a Taylor series approximation to mse^{WJ}. To this end, we borrow notation from Datta and Lahiri (2000). Let $b_{\hat{\psi}}(\psi)$ be the bias of $\hat{\psi}$, i.e., $E(\hat{\psi}) - \psi$, correct up to the order $O(m^{-1})$. Let $\nabla g_1(\psi) = (\frac{\partial}{\partial \psi_1} g_1(\psi), ..., \frac{\partial}{\partial \psi_q} g_1(\psi))'$ be the gradient of $g_1(\psi)$ [see Datta and Lahiri (2000) for an expression of the gradient]. We can approximate the bias correction in the weighted jackknife formula by

$$\begin{aligned} & \sum_{u=1}^m w_u \left(g_1(\hat{\psi}_{-u}) + g_2(\hat{\psi}_{-u}) - [g_1(\hat{\psi}) + g_2(\hat{\psi})] \right) \\ \doteq \ & \hat{b}_{WJ}'(\hat{\psi})\nabla g_1(\hat{\psi}) - tr[L(\hat{\psi})\Sigma(\hat{\psi})L'(\hat{\psi})v_{WJ}], \end{aligned} \qquad (3)$$

where $L(\psi) = col_{1\leq d\leq q} L_d'(\psi)$, $L_d(\psi) = \frac{\partial}{\partial \psi_d} s(\psi) (d = 1, \cdots, q)$, $\hat{b}_{WJ} = \sum w_u(\hat{\psi}_{-u} - \hat{\psi})$, the weighted jackknife estimator of the bias of $\hat{\psi}$, $v_{WJ} = \sum_{u=1}^m w_u(\hat{\psi}_{-u} - \hat{\psi})(\hat{\psi}_{-u} - \hat{\psi})'$, a weighted jackknife estimator of the covariance matrix of $\hat{\psi}$ and $\doteq$ means that the neglected terms are of the order $o_p(m^{-1})$. Following the arguments in Datta and Lahiri (2000) it can be shown that for the ANOVA and REML method the first term of the right hand side of (3) is of the order $o_p(m^{-1})$ and hence can be neglected. For the maximum likelihood estimator of ψ, this is, however, of the order $O_p(m^{-1})$ and needs to be kept in order to be second-order unbiased.

We approximate the last term in (2) by

$$\begin{aligned} & \sum_{u=1}^m w_u [\hat{\theta}(y; \hat{\psi}_{-u}) - \hat{\theta}(y; \hat{\psi})]^2 \\ \doteq \ & tr[L(\hat{\psi})[y - X\hat{\beta}(\hat{\psi})][y - X\hat{\beta}(\hat{\psi})]'L'(\hat{\psi})v_{WJ}] \end{aligned}$$

Therefore, we can approximate mse^{WJ} by

$$\begin{aligned} & mse^{AWJ} \\ = \ & g_1(\hat{\psi}) + g_2(\hat{\psi}) \\ & -\hat{b}_{WJ}'(\hat{\psi})\nabla g_1(\hat{\psi}) + tr[L(\hat{\psi})\Sigma(\hat{\psi})L'(\hat{\psi})v_{WJ}] \\ & +tr[L(\hat{\psi})[y - X\hat{\beta}(\hat{\psi})][y - X\hat{\beta}(\hat{\psi})]'L'(\hat{\psi})v_{WJ}]. \end{aligned}$$

There is some possibility that the the third term of mse^{AWJ} could result in a negative value, resulting in a possible negative value for mse^{AWJ}. However, using a Taylor series argument it can be easily seen that the difference between $\hat{b}_{WJ}'(\hat{\psi})$ and the bias expression given in Datta and Lahiri (2000) is of the order $o(m^{-1})$. Thus, if mse^{AWJ} turns out to be negative in a rare situation, we may simply replace $\hat{b}_{WJ}'(\hat{\psi})$ by an alternate bias term for $\hat{\psi}$ (e.g., the formula given in Datta and Lahiri 2000).

4. An Example: The Fay-Herriot Model

In order to estimate per-capita income for small areas (population less than 1,000), Fay and Herriot (1979) considered an aggregate level model and used an empirical Bayes method which combines survey data from the U.S. Current Population Survey with various administrative and census records. Their empirical Bayes estimator worked well when compared to the direct survey estimator and a synthetic estimator used earlier by the Census Bureau. The model can be written as:

$$y_i = x_i'\beta + v_i + e_i,\ i = 1, \cdots, m,$$

where v_i's and e_i's are independent with $v_i \overset{iid}{\sim} N(0, A)$ and $e_i \overset{ind}{\sim} N(0, D_i)$, $D_i\ (i = 1, ..., m)$ being *known*. Here, $n_i = b_i = 1, Z_i = 1, \psi = A,\ R_i(\psi) = D_i$ and $G_i(\psi) = A\ (i = 1, \cdots, m)$.

For the Fay-Herriot model, an EBLUP, say $\hat{\theta}_i(y_i; \hat{A})$, of $\theta_i = x_i'\beta + v_i$ is given by:

$$\hat{\theta}_i(y_i; \hat{A}) \ = \ \frac{D_i}{A + D_i} x_i'\hat{\beta} \ + \ \frac{A}{A + D_i} y_i,$$

where $\hat{\beta}$ is the usual weighted least squares estimator of β and $\hat{A}$ is a consistent estimator of A.

It can be shown that for ANOVA and REML estimators of A

$$\begin{aligned} & mse_i^{AWJ} \\ = \ & g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + \frac{D_i^2}{(\hat{A} + D_i)^3} v_{WJ}(\hat{A}) \\ & + \frac{D_i^2}{(\hat{A} + D_i)^4} (y_i - x_i'\hat{\beta})^2 v_{WJ}(\hat{A}), \qquad (4) \end{aligned}$$

where $g_{1i}(\hat{A}) = \frac{\hat{A}D_i}{\hat{A}+D_i}$, $g_{2i}(\hat{A}) = \frac{D_i^2}{(\hat{A}+D_i)^2} x_i' \left(\sum_{j=1}^m \frac{1}{\hat{A}+D_j} x_j x_j' \right)^{-1} x_i, v_{WJ} = \sum w_u (\hat{A}_{-u} - \hat{A})^2$, the weighted jackknife estimator of $Var(\hat{A})$. The choice for w_u is not unique. For example, we can choose $w_u = \frac{1}{D_u} x_u' \left(\sum_{j=1}^m \frac{x_j x_j'}{D_j} \right) x_u$.

For the maximum likelihood and the Fay-Herriot estimators of A, we have

$$\begin{aligned} & mse_i^{AWJ} \\ = \ & g_{1i}(\hat{A}) + g_{2i}(\hat{A}) - \frac{D_i^2}{(\hat{A} + D_i)^2} \hat{b}_{WJ} \\ & + \frac{D_i^2}{(\hat{A} + D_i)^3} v_{WJ}(\hat{A}) \\ & + \frac{D_i^2}{(\hat{A} + D_i)^4} (y_i - x_i'\hat{\beta})^2 v_{WJ}(\hat{A}), \qquad (5) \end{aligned}$$

where $\hat{b}_{WJ} = \sum w_u(\hat{A}_{-u} - \hat{A})$ is the weighted jackknife estimator of $b_{\hat{A}}(A)$, the bias of $\hat{A}$. For the Fay-Herriot estimator of A, it is interesting to compare (5) with the Datta-Rao-Smith MSPE estimator given by

$$\begin{aligned} & mse_i^{DRS} \\ = \ & g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + 2g_{3i}^* \\ & - \frac{D_i^2}{(\hat{A} + D_i)^2} \frac{m \sum_j \frac{1}{(A+D_j)^2} - (\sum_j \frac{1}{(A+D_j)})^2}{(\sum_j \frac{1}{(A+D_j)})^3} \qquad (6) \end{aligned}$$

where

$$g_{3i}^* = \frac{2mD_i^2}{(A + D_i)^3 (\sum_j \frac{1}{A+D_j})^2}.$$

5. Monte Carlo Simulations

In this section, we investigate the performances of different MSPE estimators for small m through Monte Carlo simulations. Our simulation set-up is similar to the one considered in Datta et al. (2005). We consider the Fay-Herriot model with $x_i'\beta = 0$, $m = 15$, and $A = 1$ and consider five groups of small areas with three areas in each group. Within each group, D_i's remain the same. We consider two different patterns for the D_i's: (a) 0.2, 0.6, 0.5,0.4,0.2, [this is pattern (b) of Datta et al. (2005)] and (b) 20,6,5,4,2. We note that Bell (2001) reported wide variations of D_i's in the context of the U.S. Current Population Survey and the National Health and Interview Survey and so pattern (b) can occur in practice.

As in Datta et al. (2005) we obtained all the results based on 100,000 simulation runs. The ANOVA and the Fay-Herriot methods of estimating the variance component A are considered. Tables 1 and 2 report the percent average relative biases (ARB) for the following MSPE estimators: the Prasad-Rao estimator (Prasad-Rao) for Table 1 (denoted by PR), the Datta-Rao-Smith estimator (Datta et al. 2005) for Table 2 (denoted by DRS), Jiang-Lahiri-Wan jackknife estimator (JLW; see Jiang, Lahiri and Wan 2002), the Chen-Lahiri estimator (denoted by CL; see Chen and Lahiri 2003), and the proposed estimator (denoted by AWJ).

For the D pattern considered by Datta et al. (2005), i.e. pattern (a), we are able to approximately reproduce their results for both the ANOVA and the Fay-Herriot methods of estimating A. For this pattern, the Taylor series (i.e., PR or DRS) is better than the CL in all but one situation. The performance of CL, however, gets better for the Fay-Herriot method of A and the AWJ is comparable to the PR or DRS. In fact, for the method of moments, AWJ is better than PR most of the time.

The performances of different MSPE estimators depend very much on the D_i pattern. For pattern (b), AWJ is a clear winner when A is estimated by the ANOVA method. When the Fay-Herriot method is used to estimate A, the percent average relative bias for the AWJ is the least among all the MSPE estimators. In this situation, the DRS, JLW and CL MSPE estimators tend to overestimate the true MSPE. In contrast, AWJ suffers from a slight underestimation problem.

6. Concluding Remarks

We have considered the case of a mixed linear normal model. Our research indicates that it is difficult to find one MSPE estimator which performs uniformly better that the rest in all situations. In our simulation, the approximated jackknife performed well, although in some cases it tends to underestimate. We have considered the case of a very small number of small areas. The performances of the MSPE estimators are expected to improve with the availability of more small areas.

References

[1] Bell, W. (2001), Discussion with "Jackknife in the Fay-Herriott Model with An Example," *Proceeding of the Seminar on Funding Opportunity in Survey Research* , 98-104.

[2] Butar, F., and Lahiri, P. (2003), "On measures of uncertainty of empirical Bayes small-area estimators," *Journal of Statistical Planning and Inference,* 112, No. 1-2, 63-76.

[3] Chen, S. and Lahiri, P. (2002), "Estimation of the mean square prediction error of EBLUP - a weighted jackknife approach," unpublished manuscript.

[4] Chen, S. and Lahiri, P. (2002), "A weighted jackknife MSPE estimator in small-area estimation," *Proceedings of the Section on Survey Research Methods, American Statistical Association* 473-477.

[5] Chen, S. and Lahiri, P. (2003), "A comparison of different MSPE estimators of EBLUP for the Fay-Herriot model," *Proceedings of the Section on Survey Research Methods, American Statistical Association* 905- 911.

[6] 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.

[7] 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.

[8] 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.

[9] Ghosh, M., and Rao, J.N.K. (1994), "Small Area Estimation: An Appraisal," *Statistical Science* 9, No. 1, 55-93, (with discussion).

[10] Jiang, J., Lahiri, P., and Wan, S.(2002), "A Unified Jackknife Method", *Annals of Statistics* , 30, No. 6, 1782-1810.

[11] Kacker, 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.

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

[13] Rao, J.N.K. (2003), *Small Area Estimation*, Wiley.

Table 1: Percent Relative Biases of Different MSPE Estimators
(A estimated by the ANOVA method)

	Pattern a				Pattern b			
	PR	CL	JLW	AWJ	PR	CL	JLW	AWJ
1	30.8	21.9	28.2	0.2	573.5	34.0	39.0	5.2
2	11.7	17.5	20.4	-1.3	268.1	52.4	53.8	13.4
3	9.0	16.2	18.5	-1.5	213.8	56.8	57.3	15.6
4	7.6	15.9	17.9	-1.3	178.4	59.8	59.7	17.4
5	0.1	12.7	13.0	-2.6	47.7	61.7	59.6	15.0

Table 2: Percent Relative Biases of Different MSPE Estimators
(A estimated by the Fay-Herriot Estimator)

	Pattern a				Pattern b			
	DRS	CL	JLW	AWJ	DRS	CL	JLW	AWJ
1	3.4	11.3	16.5	0.0	111.1	14.8	25.7	-8.5
2	0.3	6.9	9.5	-1.5	50.0	15.5	21.0	-6.8
3	-0.1	5.9	8.0	-1.7	40.4	16.0	20.4	-5.7
4	-0.2	5.4	7.3	-1.6	34.4	16.5	20.4	-4.8
5	-1.7	2.3	3.0	-1.6	18.1	20.4	22.9	1.6