A weighted jackknife MSPE estimator in small-area estimation

Shijie Chen, Research Triangle Institute
P. Lahiri, University of Maryland at College Park
Shijie Chen, STRD, RTI International, RTP, Durham NC, 27709, USA (schen@rti.org)

Key Words: Method of moments, mixed linear models, prediction, variance components.

Abstract:

The paper presents a new weighted jackknife method to estimate the mean squared prediction error (MSPE) of an empirical best linear unbiased predictor (EBLUP) of a general mixed effect in the context of a mixed linear normal model. This new MSPE estimator has the same second order asymptotic property of the Taylor series method. However, unlike the Taylor series method, our simulation results demonstrate that the proposed weighted jackknife method performs well for small samples and departure from various assumptions required to obtain the second order asymptotic properties.

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 Ghosh and Rao (1994), Lahiri and Meza (2002), among others.

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 correct MSPE estimators. See Prasad and Rao (1990), Datta and Lahiri (2000), Butar and Lahiri (2002), Jiang *et al.* (2002), among others, for various approaches.

This paper is a follow-up of the recent jackknife method of Jiang *et al.* (2002) in the following new directions:

(i) Unlike Jiang *et al.* (2002), our method covers the important method of moments method of variance component estimation.

(ii) In case of normality our proposed method exploits normality to do more exact calculations and thereby increases the efficiency if the data indeed follow normal distribution. When the distribution is not normal, weighted jackknife version of Jiang *et al* (2002) is also suggested in the concluding remarks section.

(iii) Our method uses weights using leverage values allowing smaller weights to small-areas with extreme covariate values.

(iv) Special attention has been given to achieve good properties of our weighted jackknife especially when m is small. This ensures non-negative estimates of MSPE estimators, a problem with the method proposed by Jiang *et al.* (2002). See Bell (2001).

In section 2, we define the BLUP and EBLUP of a general mixed effect. The weighted jackknife method is proposed in section 3. The proposed weighted jackknife estimator is second order accurate. 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

The research of P. Lahiri was supported in part by National Science Foundation Grant SES-9978145 and a grant from the Center of Excellence in Health Statistics of the University of Michigan, Ann Arbor.

study are reported in section 5. Unlike other methods available in the literature, our proposed method turns out to be amazingly robust against departures from a variety of model assumptions needed to prove the second order asymptotic property.

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 $v_i \overset{ind}{\sim} N_{b_i}(0, G_i)$ and $e_i \overset{ind}{\sim} N_{n_i}(0, R_i)$, $i = 1, ..., m$. We assume that $G_i = G_i(\psi)$ $(b_i \times b_i)$ and $R_i = R_i$ (ψ) $(n_i \times n_i)$ possibly depend on $\psi = (\psi_1, ..., \psi_q)'$, a $q \times 1$ vector of fixed variance components. Write $y = col_{1\le i\le m} y_i$, $e = col_{1\le i\le m} e_i$, $X = col_{1\le i\le m}(X_i)$, $Z = diag_{1\le i\le m}(Z_i)$, $G(\psi) = diag_{1\le i\le m} G_i$, $v = col_{1\le i\le m} v_i$ and $R(\psi) = diag_{1\le i\le m} R_i$. We assume that X has full column rank p. Let $\Sigma(\psi) = R(\psi) + ZG(\psi)Z'$, the variance-covariance matrix of y. With these notation 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 $\tilde{\theta}(y;\psi) = h'\tilde{\beta}(\psi) + s'(\psi)[y - X\tilde{\beta}(\psi)]$, where $s(\psi) = \Sigma^{-1}(\psi)ZG(\psi)\lambda$ and $\tilde{\beta}(\psi) = [X'\Sigma^{-1}(\psi)X]^{-1}[X'\ \Sigma^{-1}(\psi)y]$.

In practice ψ is unknown and is estimated from the data. Let $\hat{\psi}$ be a consistent estimator of ψ considered in 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 weighted 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). Using the well-known identity due to Kackar and Harville (1984):

$$\begin{aligned} MSPE[\hat{\theta}(y;\hat{\psi})] &= g_{1i}(\psi) + g_{2i}(\psi) \\ &\quad + E[\hat{\theta}(y;\hat{\psi}) - \hat{\theta}(y;\psi)]^2, \quad (2) \end{aligned}$$

where $g_1(\psi) = \lambda' G(\psi)\lambda - s'(\psi)ZG(\psi)\lambda$, and $g_2(\psi) = [h - X's(\psi)]'(X'\Sigma^{-1}(\psi)X)^{-1}[h - X's(\psi)]$. We propose to use a weighted jackknife method twice - once to estimate the first two terms of (2) by correcting the bias of $g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi})$ and then to estimate the third term involving uncertainty due to the estimation of ψ. The weighted jackknife MSPE estimator is then given by:

$$\begin{aligned} & mse^{WJ} \\ = \; & g_1(\hat{\psi}) + g_2(\hat{\psi}) \\ & - \sum_{u=1}^m w_u \left(g_1(\hat{\psi}_{-u}) + g_2(\hat{\psi}_{-u}) - [g_1(\hat{\psi}) + g_2(\hat{\psi})] \right) \\ & + \sum_{u=1}^m w_u [\hat{\theta}(y;\hat{\psi}_{-u}) - \hat{\theta}(y;\hat{\psi})]^2. \qquad (3) \end{aligned}$$

The weights w_u's satisfy certain regularity conditions (see Chen and Lahiri 2002). A possible choice of w_u is $w_u = 1 - X_u'(X'X)^{-1}X_u$ which gives smaller weights to unusual observations depending on their leverage values.

Under the regularity conditions of Prasad and Rao (1990) and Jiang *et al.* (2002), it can be shown that mse^{WJ} is second order correct as an estimator of $MSPE[\hat{\theta}(y;\hat{\psi})]$. See Chen and Lahiri (2002) for details. Notice that because of the weighted jackknife bias adjustment of $g_1(\hat{\psi}) + g_2(\hat{\psi})$ it possible for mse^{WJ} to yield negative estimates, especially when m is small and the regularity conditions are violated. The same problem arises for the unweighted jackknife method considered in Jiang *et al.* (2002). We now discuss a remedy for this situation. 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]. In case mse^{WJ} yields a negative estimate we approximate the bias correction in the weighted jackknife formula by

$$\begin{aligned} & \sum_{u=1}^m w_u \left(g_1(\hat{\psi}_{-u}) + g_2(\hat{\psi}_{-u}) - [g_1(\hat{\psi}) + g_2(\hat{\psi})] \right) \\ \doteq \; & b_{\hat{\psi}}'(\hat{\psi}) \nabla g_1(\hat{\psi}) - tr[L(\hat{\psi})\Sigma(\hat{\psi})L'(\hat{\psi})v_{WJ}], \qquad (4) \end{aligned}$$

where $L(\psi) = col_{1\le d\le q} L_d'(\psi)$, $L_d(\psi) = \frac{\partial}{\partial\psi_d} s(\psi) (d = 1, \cdots, q)$, $v_{WJ} = \sum_{u=1}^m w_u(\hat{\psi}_{-u} - \hat{\psi})(\hat{\psi}_{-u} - \hat{\psi})'$, 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 (4) 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 accurate. We shall use the following expression: $b_{\hat{\psi}_{ML}}(\psi) \doteq \frac{1}{2}v_{WJ}col_{1\le d\le q}tr[I_\beta^{-1}(\psi)\frac{\partial}{\partial\psi_d}I_\beta(\psi)]$, where $I_\beta(\psi)$ is the part of the information matrix corresponding to β.

For small m, it is possible that some components of $\hat{\psi}$ may go out of the range. In such a case, we truncate those components by the boundary values. In the weighted jackknife method, we shall not change these estimates for sub-samples. This point is illustrated in the next section.

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 $\hat{A}$ is a consistent estimator of A and $\hat{\beta}$ are consistent estimators of A and β respectively.

The weighted jackknife MSPE estimator of $\hat{\theta}_i(y_i; \hat{A})$ described given by (3) reduces to:

$$\begin{aligned} & mse_i^{1WJ} \\ = \; & g_{1i}(\hat{A}) + g_{2i}(\hat{A}) \\ - \; & \sum_{u=1}^m w_u(g_{1i}(\hat{A}_{-u}) + g_{2i}(\hat{A}_{-u}) - (g_{1i}(\hat{A}) + g_{2i}(\hat{A}))) \\ + \; & \sum_{u=1}^m w_u[\hat{\theta}_i(y_i; \hat{A}_{-u}) - \hat{\theta}_i(y_i; \hat{A})]^2, \end{aligned} \quad (5)$$

where $w_u = \frac{1}{D_u}x_u'\left(\sum_{j=1}^m \frac{x_jx_j'}{D_j}\right)x_u$, $g_{1i}(\hat{A}) = \frac{\hat{A}D_i}{\hat{A}+D_i}$, $g_{2i}(\hat{A}) = \frac{D_i^2}{(\hat{A}+D_i)^2}x_i'\left(\sum_{j=1}^m \frac{1}{\hat{A}+D_j}x_jx_j'\right)^{-1}x_i$. If $\hat{A} = 0$, we estimate MSPE by $g_{2i}(\hat{A})$.

It can be checked that for the Fay-Herriot model the bias of $\hat{A}_{ML}$ is $b_{\hat{A}_{ML}}(A) = -tr\{[\sum_{u=1}^m(A + D_u)^{-1}x_ux_u']^{-1}[\sum_{u=1}^m(A + D_u)^{-2}x_ux_u']\}/\sum_{u=1}^m(A + D_u)^{-2}$ and $\nabla g_1(A) = (\frac{D_i}{A+D_i})^2 > 0$. Note that the right hand side of (4) is given by $-g_{3i}^*(\hat{A}) = -D_i^2(\hat{A} + D_i)^{-3}v_{WJ}$ for ANOVA and REML method, where $v_{WJ} = \sum_{u=1}^m w_u(\hat{A}_u - \hat{A})^2$. For the ML method, this is given by: $b_{\hat{A}_{ML}}(\hat{A}_{ML})\nabla g_1(\hat{A}_{ML}) - g_{3i}^*(\hat{A}_{ML})$.

5. Monte Carlo Simulations

In this section, we investigate the performances of different MSPE estimators for small m through Monte Carlo simulations. For this purpose, we consider the Fay-Herriot model and consider the following four MSPE estimators for the EBLUP (with A estimated by the method of moments) given in section 4:

(i) Naive MSPE estimator given by $mse_i^N = g_{1i}(\hat{A}) + g_{2i}(\hat{A})$;

(ii) The Prasad-Rao MSPE estimator given by $mse_i^{PR} = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + 2g_{3i}(\hat{A})$, where the expressions for g_{1i} and g_{2i} are given in section 4 and $g_{3i} = \frac{2D_i^2}{m^2(\hat{A}+D_i)^3}\sum_{j=1}^m(\hat{A} + D_j)^2$. When $\hat{A} = 0$, we use $g_{2i}(\hat{A})$ in order to achieve better simulation results.

(iii) The proposed weighted jackknife mse_i^{WJ} defined in (5).

We consider $m = 12$, $p = 1$, $\beta = 1$, and $A = 10$. For the first eleven areas, we consider the following combinations of the sampling variance and covariate values: $(D, x):$ $(10, 1); (9, 1.5); (14, 2); (14, 2.5); (11, 3);$ $(10, 3.5); (10, 4); (13, 5); (4, , 6); (3, 7); (14, 8)$. To study the effect of the covariate on the accuracies of different MSPE estimators, we change $x = x_{12}$ for the last area, keeping $D = D_{12}$ fixed. Similarly, to study the effect of the sampling variance on the accuracies of different MSPE estimators, we change the sampling variance D for the last area keeping $x = x_{12}$ fixed.

For a specific simulation, $R = 10,000$ independent samples of $(v_i,\ e_i)$, $i = 1, \cdots, 12$, are generated from the Fay-Herriot model. We then calculate the relative bias (RB) of each of the three MSPE estimators for all the 12 areas using the Monte Carlo method as follows:

$$RB_i = 100\frac{E(mspe_i) - MSPE_i}{MSPE_i},\ i = 1, \cdots, 12,$$

where $mspe_i$ denotes an estimator of $MSPE_i$, the MSPE of EBLUP $\hat{\theta}_i(y_i; \hat{A})$ of the true small-area mean θ_i $(i = 1, \cdots, 12)$, and the expectation "E" is approximated by the Monte Carlo method. We report RB's for the last area and summary statistics (mean and the standard deviation) of the RB's for the rest of 11 areas.

We illustrate the effect of x through leverage value defined as $h = h_{12} = \frac{x^2}{\sum_{j=1}^{12} x_j^2}$. Note that h is an increasing function of $x > 0$. We increase h from 0 to 1 by increasing x. Table 1 displays RB's of different MSPE estimators for different x. The naive estimator underestimates MSPE in general, the magnitude of the underestimation being severe when h approaches 1. While the adjusted Prasad-Rao estimators performs well for small to moderate h, it tends to overestimate for outlying small-areas, i.e., when h approaches 1. Our weighted jackknife estimator is very robust for different x's - it does extremely well in protecting against outlying x. Table 2 presents the means and the standard deviations of the RB's for the rest of the 11 components. Here again our weighted jackknife method is usually the winner. It is interesting to note that the adjusted Prasad-Rao method performs slightly better than the weighted jackknife method in this table when h is close to 1.

Table 3 displays RB's of different MSPE estimators for different D/A for the last small-area. Here again, the naive estimator generally underestimates the true MSPE, the underestimation being severe for large D/A. The RB's for the adjusted Prasad-Rao and the weighted jackknife methods exhibit an interesting pattern for varying D/A. The absolute RB's generally increases with increasing $|D/A - 1|$. Both the methods usually overestimate when $D/A < 1$ and underestimate when $D/A > 1$. Small values of D/A cause severe overestimation for the adjusted Prasad-Rao method. In comparison, our weighted jackknife MSPE estimator is quite robust.

In Table 4 we display the means and the standard deviations for the rest of the 11 components for different values of D/A. In this table also, the robustness of our weighted jackknife method in comparison with the naive and the adjusted Prasad-Rao method is clearly demonstrated.

6. Concluding Remarks

The paper brings out usefulness of weighted jackknife method for mixed linear normal models. The extension of the method to cover nonlinear nornormal model as given in Jiang *et al.* (2002) is currently under investigation. Our preliminary results indicate that an adjustment of the method proposed by Jiang *et al.* (2002) with weight of the form $w_u = 1+O(m^{-1})$ yields a weighted jackknife method which enjoys the desirable second order asymptotic property.

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. (2002), "On the measures of uncertainty of empirical Bayes small-area estimators," *Journal of Statistical Planning and Inference,* in press.

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

[4] 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.

[5] 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.

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

[7] Jiang, J., Lahiri, P., and Wan, S.(2002), "A Unified Jackknife Method", *Annals of Statistics* , in press.

[8] 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.

[9] 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.

Table 1: Relative Biases (%) of MSE Estimators for area=12

h	Naive	P-R	W.J.
0	-21.17	3.99	3.46
0.3	-17.18	2.58	0.73
0.5	-13.90	3.49	-0.61
0.8	-7.92	6.76	-2.81
0.98	-2.36	11.28	-3.07

Table 2: Summary statistics of RB's for the first 11 areas

	h	Naive	P-R	W.J.
MEAN	0	-18.7997	6.93240	5.04580
	0.3	-19.2918	6.86435	5.31373
	0.5	-19.8263	6.80661	5.51134
	0.8	-20.6288	6.45830	5.81070
	0.98	-21.3220	6.05728	6.67634
STD	0	2.77507	7.29355	2.91998
	0.3	2.14786	6.72557	3.00014
	0.5	1.62392	6.27256	3.20507
	0.8	1.08782	5.03241	3.75387
	0.98	1.45733	3.95403	4.80431

Table 3: Relative Biases (%) of MSE Estimators for area=12

D	Naive	P-R	W.J.
0.1	-6.0128	64.7182	11.0703
0.5	-11.1043	14.4996	3.8854
1	-13.6912	4.4930	-0.1128
2	-15.1193	-2.0299	-3.3317
10	-16.7781	-7.3950	-7.6992

Table 4: Summary statistics of RB's for the first 11 areas

	D	Naive	P-R	W.J.
MEAN	0.1	-24.3902	2.3620	6.38752
	0.5	-20.1185	5.2909	5.84857
	1	-19.8493	6.5309	5.60050
	2	-19.7484	9.6408	5.23917
	10	-22.6196	89.6253	3.58639
STD	0.1	2.00918	2.5557	3.67880
	0.5	1.15896	4.9565	3.67170
	1	1.53662	6.0253	3.30198
	2	2.03694	7.8697	2.92664
	10	2.97276	47.7501	3.87454