

Invited Paper

A Review of Empirical Best Linear Unbiased Prediction For the Fay-Herriot Small-Area Model

Partha Lahiri*

ABSTRACT

The Fay-Herriot model, a simple mixed regression model, has played an important role in small-area estimation. In this paper, we firstly motivate the use of empirical best linear unbiased predictors (EBLUP) using several special cases of the Fay-Herriot model. We then critically examine different issues involving estimation and prediction, including the variance component estimation and the measure of uncertainty of an EBLUP.

Key words: Mean squared errors, mixed linear model, Henderson's method III, variance components.

1. INTRODUCTION

Various government agencies (e.g., the U.S. Census Bureau, Statistics Canada, Central Statistical Office of U.K., etc.) are in need of producing reliable small-area statistics. A small-area (or small domain) generally refers to a subgroup of a large target population. The subgroup may refer to a small geographic region (e.g., state, county, municipality, etc.), a particular demographic group (e.g., black female in the age group 18-24) or a demographic group within a small geographic region. Small-area statistics are needed in regional planning and fund allocation in many government programs and thus the importance of producing reliable small-area statistics cannot be over-emphasized.

There is a long history of small-area estimation. Brackstone (1987) cited the use of small-area statistics in 11th century England and 17th century Canada when the small-area statistics were mostly based on administrative or census records. Nowadays most private and government agencies rely on sample surveys for publishing official statistics. With the help of a well-designed sample survey, quality data on a variety of variables can be frequently collected at a lower cost. Because of the heavy reliance on survey data by different statistical agencies, a natural question is: can sample survey data be used in estimating small-area characteristics?

Clearly, a design-based estimator, which uses only the sample survey data for the particular small-area of interest, is unreliable due to the small sample size for the small-area. For example, in a statewide telephone survey of sample size 4,300 in the state of Nebraska, U.S.A., only 14 observations are available to estimate the prevalence of alcohol abuse in the Boone county, a small county in Nebraska. The problem is even more severe for direct survey estimation of the prevalence rate for white female in the age group 25-44 in this county since only one observation is available from the survey. See Meza, *et al.* (2003) for details.

*Milton Mohr Distinguished Professor of Statistics, University of Nebraska-Lincoln, Adjunct Professor, The Joint Program in Survey Methodology, University of Maryland at College Park and Senior Scientist, The Gallup Organization, Princeton; Email: plahiri@survey.umd.edu

Different strategies can be considered at the design stage to improve on small-area statistics. For example, more stratification and less clustering are likely to increase the efficiency of small-area statistics. Also, one can develop a suitable sample allocation formula so as to increase the efficiency of design-based estimators for small-areas. Ideally, these strategies should be considered in such a way that will not result in a considerable increase in the total survey cost and permit estimation of large area characteristics with a desired level of accuracy. Readers interested in learning more about these design strategies are referred to Rao (2003) and the papers cited therein.

Even after adopting good practices at the design stage, standard design-based methods do not usually offer a small-area statistic with an acceptable level of accuracy. In order to improve on design-based estimators, several indirect and model-based methods have been proposed in the literature. These improved estimation procedures use *implicit* or *explicit* models, which *borrow strength* from related sources such as administrative and census records, and previous survey data.

In order to estimate per-capita income for small areas (populations less than 1,000), Fay and Herriot (1979) used an empirical Bayes method, which combines the U.S. Current Population Survey data with various administrative and census records. In order to incorporate both the sampling and model errors, Fay and Herriot (1979) used a two-level model, which can be either viewed as a Bayesian model or a mixed regression model. Their empirical Bayes estimator [also an empirical best linear unbiased predictor (EBLUP)] performed better than the direct survey estimator and a synthetic estimator used earlier by the U.S. Census Bureau. The Fay-Herriot model and the associated EBLUP are now widely used in small-area estimation and related problems.

The Fay-Herriot model has been implemented using a Hierarchical Bayes (HB) approach, which is straightforward, in the sense that the posterior distribution, once obtained, can be used for all inferential purposes. However, it requires a specification of the prior distribution on the regression coefficients and the variance component (known as hyper-parameters in the Bayesian literature). The method requires checking the propriety of the posterior distribution of the parameter(s) of interest and the convergence of the Markov Chain Monte Carlo (MCMC) method often used to approximate the posterior distribution. In the passing, we note that MCMC methods could be very time consuming for more complex multi-level small-area models.

In an EBLUP approach, the best linear unbiased predictor (BLUP) of the small area is first produced using the general theory of Henderson (1975) and then the unknown variance component(s) is (are) estimated by a standard method [e.g., maximum likelihood (ML), residual maximum likelihood (REML), ANOVA, etc.]. The resultant predictor, i.e., the BLUP with estimated variance component(s), is known as an EBLUP of the true small-area mean. Unlike a HB approach, an EBLUP approach does not require any specification of the prior distribution on the hyper-parameters and it generally takes considerably less time in producing the small-area estimates. However, for some data set, the standard methods may produce an unreasonable zero estimate of a variance component, especially when the number of small-areas is small. In the context of the Fay-Herriot model, we critically examine this issue and review a recent paper by Lahiri (2003) who suggested some remedies. Another challenging problem in an EBLUP approach is to obtain a reliable measure of uncertainty of an EBLUP that captures all sources of variabilities. We critically review different methods of producing uncertainty measures that have been proposed in the literature for the last fifteen

years. In this paper, we do not attempt to cite all the papers that use the Fay-Herriot model or its extension. Such references can be found in Rao (2003).

2. EBLUP WITH NO AUXILLARY VARIABLE

The following hypothetical example reminds us of the deficiencies of the usual design-based estimators for small-area estimation and the need for considering estimators that are possibly not design-unbiased.

A large-scale national sample survey produces the following unbiased estimates of per-capita income for two of its states (small-areas): State A: \$15,000 and State B: \$30,000. Based on these estimates, a decision is made to provide federal assistance to State A than State B. Is it a fair allocation of federal money? We cannot, of course, support this fund allocation plan if it turns out that the standard errors for these estimates are: State A: \$15,000, State B: \$2,000. In this case, the huge difference in the per capita income estimates for these two states may not be real and can be explained by the very unreliable (reflected by high standard error) unbiased estimate of State A.

What can we do to improve a situation described above? Should we consider a slightly biased estimator if it reduces the variability of this estimator? To this end, we revisit the well-celebrated James Stein estimator in the small-area estimation context. Let y_i be the direct estimator of the true mean θ_i of the i th small-area, $i = 1, 2, \dots, m$. Usually y_i is an average or a weighted average of observations in the i th small-area and is typically design-unbiased (or nearly so) for θ_i , $i = 1, 2, \dots, m$. Let $y_i \stackrel{ind}{\sim} N(\theta_i, 1)$ $i = 1, 2, \dots, m$ and consider the simultaneous estimation of $\theta = (\theta_1, \dots, \theta_m)'$. Under the sum of squared error loss, the frequentist's risk of $y = (y_1, \dots, y_m)'$, as an estimator of θ , is given by

$$R(\theta, y) = \sum_{i=1}^m E[(y_i - \theta_i)^2 | \theta_i] = m$$

It is well-known that for $m \geq 3$, y is uniformly (i.e., for all θ) inferior to the well-celebrated James-Stein estimator $\hat{\theta}^{JS} = (\hat{\theta}_1^{JS}, \dots, \hat{\theta}_m^{JS})'$, where

$$\hat{\theta}_i^{JS} = \left(1 - \frac{m-2}{s}\right)y_i, \quad s = \sum_{i=1}^m y_i^2,$$

and

$$R(\theta, \hat{\theta}^{JS}) \leq m - \frac{(m-2)^2}{m-2 + \sum_i \theta_i^2}$$

If $\theta_i = 0$, ($i = 1, \dots, m$) then

$$R(\theta, \hat{\theta}^{JS}) \leq [m - (m-2)] = 2$$

Thus, the largest reduction is obtained when $\theta_i = 0$, ($i = 1, \dots, m$) and m large. However, this is hardly realistic since some variation in the θ_i 's is expected.

The above discussions suggest that the justification of the James-Stein estimator lies in the following two-level model:

Model 1: A Two-Level Model

- Level 1: $y_i | \theta_i \stackrel{ind}{\sim} N(\theta_i, 1)$, $i = 1, \dots, m$;
- Level 2: $\theta_i \stackrel{ind}{\sim} N(0, A)$, $i = 1, \dots, m$;

We assume that $A > 0$ since otherwise $\theta_i = 0$, ($i = 1, \dots, m$). Efron and Morris (1973) used the above two-level model (also a Bayesian model) in order to provide the following empirical Bayes interpretation of the James-Stein estimator:

Under Model 1, the posterior distribution of θ_i is given by: $\theta_i | y_i \stackrel{ind}{\sim} N[(1-B)y_i, 1-B]$, where $B = \frac{1}{1+A}$. We note that B is strictly less than 1 since $A > 0$. The Bayes estimator of θ_i , under the squared error loss function, is then given by

$$\hat{\theta}_i^B = (1-B)y_i,$$

When B is unknown, it can be estimated by $\hat{B} = \frac{m-2}{s}$, an unbiased estimator of B under the marginal distribution of y . The following empirical Bayes estimator of θ_i is obtained from the Bayes estimator by plugging in \hat{B} for B :

$$\hat{\theta}_i^{EB} = (1-\hat{B})y_i$$

Note that $\hat{\theta}_i^{EB} = \hat{\theta}_i^{JS}$.

The empirical Bayes estimator $\hat{\theta}_i^{EB}$ can be also motivated by a linear empirical Bayes approach where the normality assumption for both levels of Model 1 is replaced by the assumption of posterior linearity, i.e., $E[\theta_i | y_i] = a + by_i$, where a and b are constants. It is also an EBLUP under the following simple random effects model:

$$y_i = v_i + e_i$$

where the sampling errors $\{e_i\}$ and the random effects $\{v_i\}$ are uncorrelated with $v_i \sim (0, A)$ and $e_i \sim (0, 1)$, $i = 1, \dots, m$. Without the assumption of normality, \hat{B} is no longer unbiased and the uniform dominance of the James-Stein estimator over y , if true, has not been established.

It is interesting to look at the James-Stein estimator from a survey sampler's viewpoint where y_i 's are treated as fixed and the entire randomness is induced by the sampling design. Let y_i be a design-unbiased estimator of the true small-area mean θ_i ($i = 1, \dots, m$). It is easy to see that the usual design-based bias and variance of this estimator $-B\theta_i$ and $(1-B)^2$, respectively. Thus, on the average, the estimator underestimates θ_i . Also, as B approaches 1, the absolute bias increases but the variance decreases. A robust way to compare $(1-B)y_i$ with the design-unbiased estimator y_i is to compare the design-based mean square error, defined as $MSE_d(e_i) = E_d[e_i - \theta_i]^2$, where e_i is an arbitrary estimator of θ_i and E_d is the expectation with respect to the sampling design. It is easy to see that

$$MSE_d[(1-B)y_i] = (1-B)^2 + B^2\theta_i^2.$$

Thus, the best case for this estimator is when $\theta_i \approx 0$ and $B \approx 1$.

We point out that B is an important factor that provides relevant information about the choice of the James-Stein estimator over the direct estimator. As discussed earlier, we should seriously consider the James-Stein estimator when B is very close to 1 and m is large. However, B is generally unknown and thus one may consider \hat{B} to understand the utility of the James-Stein shrinker for a given data set. For some data set, \hat{B} may be more than 1 in which case it is usually truncated to 1 yielding an unreasonable estimate of B . Note that the ML method could have a similar problem. One must obtain a reasonable estimator of B before using this to determine the utility of the James-Stein shrinker. The problem can be solved when some information on B is available through a probability distribution on $(0,1)$. To a Bayesian, this is a prior distribution on B ; to a non-Bayesian, this is a part of a multi-level model (in this case, a three level model).

The real problem is when nothing is known about the distribution on B . In this case, there is a difference between the Bayesian and non-Bayesian estimation. In a non-Bayesian approach, B is treated as a fixed unknown parameter while in a Bayesian approach, a vague prior distribution is assumed for B (i.e., B is assumed to be a random variable under a Bayesian approach). However, this apparent difference between the Bayesian and non-Bayesian methods of estimation can be made smaller. The problem associated with the ML approach can be easily rectified in a non-Bayesian framework by choosing alternate measures of central tendency (e.g., median, mean, etc.) of the probability distribution obtained by standardizing the marginal likelihood function. Note that the ML estimator of B essentially corresponds to the mode of this standardized marginal likelihood. The problem does not seem to be with the marginal likelihood but the choice of the measure of central tendency of the standardized marginal likelihood. Needless to say that the Bayes estimator of B with a uniform vague prior on $(0,1)$ is identical to the average of the standardized likelihood.

In a Bayesian approach, there is a debate as to which vague prior for B should be chosen. Of course, it is difficult to reach a consensus among the Bayesians unless there is an agreement about the evaluation criteria for the vague prior selection. Using a Monte Carlo simulation study, Chen (2001) compared the two-level marginal mean square error of different HB estimators of B under a variety of prior distributions on B for different values

of B . As expected, there is no unique prior distribution on B that emerges as the best for all B but certain priors emerge as viable choices.

As argued earlier, there is no need to introduce a vague prior distribution on B in order to obtain an estimate of B that is strictly less than 1 – one can simply choose a different measure of central tendency of the standardized marginal likelihood. As an alternative, Lahiri (2003) considered certain smooth estimators of B which can be viewed as a weighted average of 1 and the usual unbiased estimator of B , the weight assigned to 1 being a decreasing function of the test statistic for testing $H_0 : A = 0$.

The above two-level model lays the foundation of more complex commonly used multi-level small-area models. We shall now discuss some simple useful extensions of Model 1. Efron and Morris (1975) extended the model by replacing the known zero mean in the second level by an unknown mean μ . They illustrated the superiority of an empirical Bayes estimator over the direct estimator in the context of estimating the 1970 season batting averages of 18 major league baseball players based on batting averages of these players 45 at bats. For this model, the Bayes estimator of θ_i is given by:

$$\hat{\theta}_i(y_i; B) = \mu + (1 - B)(y_i - \mu)$$

Efron and Morris (1975) estimated μ and B by their unbiased estimators $\bar{y} = m^{-1} \sum_{i=1}^m y_i$, and $B = (m - 3) / \sum_{i=1}^m (y_i - \bar{y})^2$, respectively. An empirical Bayes estimator of θ_i is then given by

$$\hat{\theta}_i = \bar{y} + (1 - B)(y_i - \bar{y}).$$

The data analysis given in Efron and Morris (1975) clearly shows that their empirical Bayes estimators perform much better than the direct estimators on the average. An extension of the Efron-Morris model to the unbalanced sampling variances can be found in Carter and Rolph (1974) who estimated the probability that a box-reported alarm signals a structural fire given the alarm box location.

Model-based small area estimation is usually thought to be important only in the presence of strong auxiliary information. However, the above two examples demonstrate the utility of empirical Bayes estimators or EBLUP's even when there is no auxiliary information. The main reason for the success of the empirical Bayes method in these two situations is that the true small-area means appear to be more or less exchangeable yielding high values of the factor B or B_i in the unbalanced case. For example, an unbiased estimate of B is 0.791 for the baseball data. When no auxiliary information is available or when the relationship between the main variable and the auxiliary variable(s) is not clear cut, one may seriously consider a method similar to Efron and Morris (1975) or Carter and Rolph (1974) on the group of small-areas which can be approximately treated as exchangeable.

3. EBLUP WITH AUXILLARY VARIABLES

The discussions in Section 2 do not by any means imply that auxiliary information is not needed in a small-area estimation problem. An intelligent use of strong auxiliary information will certainly improve on small-area estimation. For example, Fay and Herriot (1979) considered the following extension of the Carter-Rolph model where they replaced the common mean μ by a linear regression:

The Fay-Herriot Model:

- Level 1: $y_i | \theta_i \stackrel{ind}{\sim} N(\theta_i, D_i), \quad i = 1, \dots, m;$
- Level 2: $\theta_i \stackrel{ind}{\sim} N(x_i' \beta, A), \quad i = 1, \dots, m.$

In Fay and Herriot (1979), the sampling variances D_i were assumed to be known although they were really estimated using a generalized variance function technique. In the above model, x_i represents a vector of known auxiliary variables such as tax return data, housing data, etc.

Under the Fay-Herriot model, the Bayes estimator of θ_i is given by:

$$\hat{\theta}_i(y_i; \phi) = (1 - B_i)y_i + B_i x_i' \beta,$$

where $B_i = \frac{D_i}{D_i + A}$, $\phi = (\beta, A)'$.

If A is known, β can be estimated by

$$\hat{\beta}(A) = \left(\sum_{i=1}^m \frac{1}{D_i + A} x_i x_i' \right)^{-1} \left(\sum_{i=1}^m \frac{1}{D_i + A} x_i y_i \right),$$

a weighted least square estimator of β . Replacing β by $\hat{\beta}(A)$ we get the following empirical Bayes estimator of θ_i :

$$\hat{\theta}_i^{EB} = \hat{\theta}_i(y_i; A) = (1 - B_i)y_i + B_i x_i' \hat{\beta}(A).$$

We note that $\hat{\theta}_i(y_i; A)$ is also the BLUP under the following mixed regression model:

$$y_i = x_i' \beta + v_i + e_i$$

where the sampling errors $\{e_i\}$ and the random effects $\{v_i\}$ are uncorrelated with $v_i \sim (0, A)$ and $e_i \sim (0, D_i), i = 1, \dots, m$.

When both β and A are unknown we propose the following empirical Bayes (same as EBLUP) estimator of θ_i :

$$\hat{\theta}_i(y_i; \hat{A}) = (1 - \hat{B}_i)y_i + \hat{B}_i x_i' \hat{\beta}(\hat{A}),$$

where $\hat{B}_i = \frac{D_i}{D_i + \hat{A}}$. Different estimators of A such as the standard ANOVA estimator, ML,

REML and the Fay-Herriot estimator have been proposed in the literature. While all of them perform well for large m , they have the problem of producing zero estimates for A , especially for small and moderate m . A Bayesian method can always produce a positive solution for A . But, this, in turn, generates another problem of choice of prior for A . Certain non-Bayesian strictly positive estimators of A have been discussed by Lahiri (2003).

The assumption of known sampling variances D_i are often justified by means of the standard asymptotic theory of transformation of variable and/or empirical evidence of the relationship between the estimated coefficient of variation and relevant census or administrative variable(s). Thus, usually once an EBLUP is found, it needs to be transformed back in the original scale. A naïve method calls for taking a reverse transformation. However, Lahiri (2003) suggested alternative Bayesian and non-Bayesian solutions, which need to be evaluated through simulations.

4. MEAN SQUARE ERROR ESTIMATION

We define the two-level marginal mean square prediction error as $MSPE(\hat{\theta}_i) = E[(\hat{\theta}_i - \theta_i)^2]$, where the expectation is taken over the marginal distribution of y under the Fay-Herriot model. For the last two decades, several attempts have been made to estimate $MSPE(\hat{\theta}_i)$ accurately. We will now explain some of the approaches in estimating $MSPE$.

Note that by the Kackar-Harville identity [see Kackar and Harville, 1984], we have

$$MSPE[\hat{\theta}_i(y_i; \hat{A})] = g_{1i}(A) + g_{2i}(A) + G_{3i}(A), \quad (1)$$

where

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

$$g_{2i}(A) = \frac{D_i^2}{(A + D_i)^2} x_i' \left(\sum_{j=1}^m \frac{1}{A + D_j} x_j x_j' \right)^{-1} x_i,$$

$$G_{3i}(A) = E[\hat{\theta}_i(y_i; \hat{A}) - \hat{\theta}_i(y_i; A)]^2.$$

Note that $g_{1i}(A) + g_{2i}(A)$ is the MSPE of the BLUP, $\hat{\theta}_i(y_i; \hat{A})$, and $G_{3i}(A)$ is the additional uncertainty due to the estimation of the variance component A .

A naïve *MSPE* estimator is obtained by estimating the *MSPE* of the BLUP and is given by:

$$mse_i^N = g_{1i}(\hat{A}) + g_{2i}(\hat{A}).$$

Intuitively, this naïve *MSPE* estimator is likely to underestimate the true *MSPE* since it fails to incorporate the additional uncertainty due to the estimation of A . In fact, Prasad and Rao (1990) showed that the order of this underestimation is $O(m^{-1})$ under the following regularity conditions:

$$(r.1) \quad 0 < D_L \leq D_i \leq D_U < \infty, \quad \forall i = 1, \dots, m;$$

$$(r.2) \quad \sup_{i \geq 1} h_{ii} = O\left(\frac{1}{m}\right),$$

where $h_{ii} = x_i (\sum_{j=1}^m x_j x_j')^{-1} x_i$ is the leverage. Interestingly, the naïve *MSPE* estimator even underestimates the true *MSPE* of the BLUP, the order of underestimation being $O(m^{-1})$.

Prasad and Rao (1990) proposed the following *MSPE* estimator when A is estimated by the usual method of moments:

$$mse_i^{PR} = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + 2g_{3i}(\hat{A}),$$

where $g_{3i} = \frac{2D_i^2}{m^2(\hat{A} + D_i)^3} \sum_{j=1}^m (\hat{A} + D_j)^2$. This estimator was obtained using the argument that under regularity conditions (r.1) and (r.2),

$$(a) \quad G_{3i}(A) = g_{3i}(A) + o(m^{-1});$$

$$(b) \quad E[g_{1i}(\hat{A})] = g_{1i}(A) - g_{3i}(A) + o(m^{-1});$$

(c) The terms $g_{2i}(A)$ and $g_{3i}(A)$ are of the order $O(m^{-1})$ and \hat{A} is a consistent estimator of A .

The Prasad-Rao formula and other *MSPE* estimators based on the Taylor series approximation are often criticized on the ground that the formulae are obtained using regularity conditions (r.1) and (r.2) and large m which may not be satisfied in many applications. While the Bayesian posterior variances do not need these regularity conditions and large m , they have a different problem of selection of a prior distribution for A . Chen and Lahiri (2003b) studied the *exact* relative contributions of the three terms in the right hand side of (1) for a special case of the Fay-Herriot model when $x_i = 1$, $D_i = D$, $i = 1, \dots, m$ and \hat{A} is the usual untruncated unbiased quadratic estimator of A given in Prasad and Rao (1990). We restate their results in our notation.

The relative contributions of the three terms in the right hand side of (1) are given by:

$$(i) h_1(B, m) = \frac{g_1(A)}{g_1(A) + g_2(A) + G_3(A)} = \frac{1 - B}{1 - B + \frac{B}{m} + \frac{m-1}{m-3} \frac{2B}{m}};$$

$$(ii) h_2(B, m) = \frac{g_2(A)}{g_1(A) + g_2(A) + G_3(A)} = \frac{\frac{B}{m}}{1 - B + \frac{B}{m} + \frac{m-1}{m-3} \frac{2B}{m}};$$

$$(iii) h_3(B, m) = \frac{G_3(A)}{g_1(A) + g_2(A) + G_3(A)} = \frac{\frac{m-1}{m-3} \frac{2B}{m}}{1 - B + \frac{B}{m} + \frac{m-1}{m-3} \frac{2B}{m}}.$$

Chen and Lahiri (2003) noted the following:

- (a) The relative contributions do not depend on D .
- (b) For fixed $0 < B < 1$, $h_1(B, m)$ is an increasing function in m with $\lim_{m \uparrow \infty} h_1(B, m) = 1$. On the other hand, $h_2(B, m)$ and $h_3(B, m)$ are decreasing functions of m with $\lim_{m \uparrow \infty} h_2(B, m) = \lim_{m \uparrow \infty} h_3(B, m) = 0$.
- (c) For fixed m , $h_1(B, m)$ is a decreasing function in B with $\lim_{B \downarrow 0} h_1(B, m) = 1$ and $\lim_{B \uparrow 1} h_1(B, m) = 0$. On the other hand, both $h_2(B, m)$ and $h_3(B, m)$ are increasing functions of B with

$$\lim_{B \downarrow 0} h_2(B, m) = \lim_{B \downarrow 0} h_3(B, m) = 0;$$

$$\lim_{B \uparrow 1} h_2(B, m) = \frac{1}{1 + \frac{2(m-1)}{m-3}};$$

$$\lim_{B \uparrow 1} h_3(B, m) = \frac{1}{1 + \frac{m-1}{2(m-3)}}.$$

In the presence of an auxiliary variable, Chen and Lahiri (2003b) showed that the relative contribution of $g_2(A)$ is an increasing function of the leverage and may be even larger than the relative contribution of $g_1(A)$.

The above results point out the difficulties in establishing an appropriate asymptotic framework, which works for all situations. The simulation study of Chen and Lahiri (2003b) shows that the Taylor series method tends to overestimate the true $MSPE$ considerably compared to the jackknife and the parametric bootstrap methods and this overestimation increases with the increase of B . It is interesting to note that both the jackknife and the parametric bootstrap methods, like the Bayesian method, are not derived using any

approximation but both enjoy the same second-order property of the Prasad-Rao Taylor series method under the regularity conditions (r.1) and (r.2).

We now review the jackknife and the parametric bootstrap methods. For the Fay-Herriot model, the jackknife *MSPE* estimator proposed by Jiang, Lahiri and Wan (2002) reduces to:

$$mspe_i^{JLW} = g_{1i}(\hat{A}) - \frac{m-1}{m} \sum_{u=1}^m (g_{1i}(\hat{A}_{-u}) - g_{1i}(\hat{A})) + \frac{m-1}{m} \sum_{u=1}^m (\tilde{\theta}_i(y_i; \hat{A}_{-u}) - \tilde{\theta}_i(y_i; \hat{A}))^2,$$

where

$$\tilde{\theta}_i(y_i; \hat{A}_{-u}) = \frac{D_i}{\hat{A}_{-u} + D_i} x_i \hat{\beta}_{-u} + \frac{\hat{A}_{-u}}{\hat{A}_{-u} + D_i} y_i,$$

$$\hat{\beta}_{-u} = \left(\sum_{j \neq u} \frac{D_j}{\hat{A}_{-u} + D_j} x_j x_j' \right)^{-1} \sum_{j \neq u} \frac{D_j}{\hat{A}_{-u} + D_j} x_j y_j.$$

Jiang, Lahiri and Wan (2002) motivated their jackknife estimator as follows:

(a) Note that

$$MPSE[\hat{\theta}_i(y_i; \hat{A})] = g_{1i}(A) + E[\hat{\theta}_i(y_i; \hat{A}) - \hat{\theta}_i(y_i; \phi)]^2.$$

Jiang, Lahiri and Wan (2002) exploited the above identity in order to cover a general non-normal nonlinear mixed model. In the above, $\hat{\theta}_i(y_i; \phi)$ is the best predictor (BP) of θ_i under the assumption $E[\theta_i | y_i] = a + by_i$, where a and b are constants.

(b) The second term of the jackknife formula corrects the bias of $g_{1i}(\hat{A})$.

(c) The third term of the jackknife formula is carefully devised so as to capture the additional uncertainties due to the estimation of β and A . We stress that both $\hat{\beta}$ and \hat{A} need to be recomputed using the deleted samples.

Chen and Lahiri (2002) proposed a different jackknife *MSPE* estimator especially designed for a mixed linear normal model. For the Fay-Herriot case, this is given by:

$$mspe_i^{CL} = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) - \sum_{u=1}^m w_u (g_{1i}(\hat{A}_{-u}) + g_{2i}(\hat{A}_{-u}) - [g_{1i}(\hat{A}) + g_{2i}(\hat{A})]) + \sum_{u=1}^m w_u [\hat{\theta}_i(y_i; \hat{A}_{-u}) - \hat{\theta}_i(y_i; \hat{A})]^2.$$

Chen and Lahiri (2002) suggested two choices of $w_u = \frac{m-1}{m}$ and $w_u = x_u' (\sum_{j=1}^m x_j x_j') x_u$. Note that $mspe_i^{CL}$ is different from $mspe_i^{JLW}$ in two respects. First,

Chen and Lahiri (2002) used more exact calculations by exploiting the Kackar-Harville identity, which is valid for the normal assumption. Secondly, the method also adjusts the $g_{2i}(\hat{A})$ term for bias. Although in the standard second-order asymptotic sense this adjustment is not needed, we may not ignore this bias correction when the relative contribution from $g_{2i}(A)$ is significant.

Butar and Lahiri (2003) proposed the following parametric bootstrap *MSPE* estimator:

$$mspe_i^{BL} = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) - E_* \left(g_{1i}(\hat{A}^*) + g_{2i}(\hat{A}^*) - [g_{1i}(\hat{A}) + g_{2i}(\hat{A})] \right) \\ + E_* [\hat{\theta}_i(y_i; \hat{A}^*) - \hat{\theta}_i(y_i; \hat{A})]^2$$

In the above, E_* is the bootstrap expectation, i.e., expectation with respect to the Fay-Herriot model with β and A replaced by $\hat{\beta}$ and \hat{A} , respectively. We obtain \hat{A}^* using the formula for \hat{A} with the original sample replaced by the bootstrap sample. In practice, Monte Carlo methods are employed to approximate the bootstrap expectations.

Note that because of the bias correction term, the jackknife and parametric bootstrap methods could produce negative estimates. This was first observed by Bell (2001) in the context of jackknife *MSPE* estimator. But this can be easily corrected as noted by Chen and Lahiri (2003) who recommended the following *MSPE* estimator in case $mspe_i^{CL}$ yields a negative value:

$$mspe_i^{CL} = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + D_i^2 (\hat{A} + D_i)^{-3} v_{WJ} + \sum_{u=1}^m w_u [\hat{\theta}_i(y_i; \hat{A}_u) - \hat{\theta}_i(y_i; \hat{A})]^2.$$

where $v_{WJ} = \sum_{u=1}^m w_u (\hat{A}_u - \hat{A})^2$. Similar corrections can be made for $mspe_i^{JLW}$ and $mspe_i^{BL}$.

For small m , \hat{A} could yield a zero estimate. This is problematic for all the *MSPE* estimators described earlier. In order to achieve good small sample properties, Chen and Lahiri (2003a) suggested using $g_{2i}(\hat{A})$ for a *MSPE* estimate whenever $\hat{A} = 0$.

Chen and Lahiri (2003b) conducted an extensive simulation study to compare the Taylor series, jackknife and parametric bootstrap *MSPE* estimators for a special case of the Fay-Herriot model when $D_i = D, i = 1, \dots, m$. The Chen-Lahiri jackknife estimator and the Butar-Lahiri parametric bootstrap are very robust for different values of B and the leverage, and overall they perform very well compared to the other methods. The performance of the Jiang-Lahiri-Wan jackknife and the Prasad-Rao Taylor series *MSPE* estimators depend very much on the combination of values of B and the leverage.

For a special case of the Fay-Herriot model when $D_i = D$ and $x_i \beta = \mu$ ($i = 1, \dots, m$), Butar and Lahiri (2003) showed that their parametric bootstrap *MSPE* estimator is identical to a measure of uncertainty proposed by Morris (1983) up to the

order $O(m^{-1})$ if an unbiased estimator of $B = D/(A + D)$ is chosen in the EBLUP formula. This is also true for the Chen-Lahiri jackknife $MSPE$ estimator. Thus, the parametric bootstrap and the Chen-Lahiri jackknife $MSPE$ estimators are close to a Bayesian solution since Morris obtained his uncertainty measure by approximating the posterior variance using flat uniform priors on the μ and B .

For the above model and the standard untruncated unbiased quadratic estimator of A , Lahiri (1995) approximated $mspe_i^{JLW}$ up to the order $O_p(m^{-1})$ and obtained the following result:

$$mspe_i^{JLW} \doteq g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + \frac{D^2}{m(\hat{A} + D)}(b_2 - 1) + \frac{D^2}{m(\hat{A} + D)^2}(b_2 - 1)u_i^2 - \frac{2D^2}{m(\hat{A} + D)^{3/2}}\sqrt{b_1}u_i$$

where $b_1 = m_3^2/(\hat{A} + D)^3$, $b_2 = m_4/(\hat{A} + D)^2$ and $u_i = (y_i - \bar{y})$. Here b_1 and b_2 can be viewed as estimated skewness and kurtosis for the marginal distribution of y_i 's. Under normality, $b_1 \approx 0$ and $b_2 \approx 3$ and so in this case $mspe_i^{JLW}$ reduces to

$$mspe_i^{JLW} \doteq g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + \frac{2D^2}{m(\hat{A} + D)} + \frac{2D^2}{m(\hat{A} + D)^2}u_i^2.$$

It is reasonable to expect that in this case $mspe_i^{JLW}$ is identical to $mspe_i^{CL}$ and $mspe_i^{BL}$ correct up to the order $O_p(m^{-1})$.

We can define the first level (or conditional) mean square prediction error as: $CMSPE(\hat{\theta}_i) = E[(\hat{\theta}_i - \theta_i)^2 | \theta]$, where the expectation is taken over the first level of the model, i.e., conditioning on the θ_i 's. Rivest and Belmonte (1999) proposed an unbiased estimator of $CMSPE$. Hwang and Rao (1987) obtained a similar unbiased estimator earlier and, using a Monte Carlo simulation study, showed that the Prasad-Rao $MSPE$ estimator is more stable than such an unbiased estimator of $CMSPE$. Interestingly, their simulation results showed that the Prasad-Rao $MSPE$ estimator tracks the $CMSPE$ very well under a moderate deviation from the level 2 model. However, the Prasad-Rao $MSPE$ estimator could perform poorly compared to the $CMSPE$ unbiased estimator for an outlying small-area. This may be due to the fact that the Prasad-Rao $MSPE$ estimator is not area specific in terms of the main variable. The jackknife and the parametric bootstrap $MSPE$ estimators are likely to perform better than the Prasad-Rao estimator in this situation as both of them are area-specific with respect to the main variable. However, this conjecture remains to be validated through a Monte Carlo simulation study.

5. CONCLUDING REMARKS

The Fay-Herriot model assumes known sampling variances and thus all the $MSPE$ estimators discussed in this paper underestimate the true uncertainty of an EBLUP. Nonetheless, the Fay-Herriot model has been widely used in practice because of its simplicity and its ability to produce EBLUP's that are design-consistent. As an alternative to the Fay-

Herriot aggregate level model, one can explore modeling the individual responses. However, this poses the challenging problem of modeling complex survey data and producing estimates that are design-consistent. We have not reviewed various complex small area modeling and design-consistency issues in this paper but will refer the readers to the relevant papers cited in Rao (2003).

ACKNOWLEDGMENT

The author thanks the Associate Editor for his comments on an earlier version of this paper and Dr. William R. Bell of the U.S. Census Bureau for some helpful discussions on variance component estimation. The research was supported in part by the Gallup Organization and a grant from the Center of Excellence in Health Studies of the University of Michigan, Ann Arbor.

References

- BELL, W. (2001), Discussion with "Jackknife in the Fay-Herriot Model with an Example," *Proceeding of the Seminar on Funding Opportunity in Survey Research*, 98-104.
- BRACKSTONE, G. J. (1987), "Small area data: policy issues and technical challenges," In *Small Area Statistics* (R. Platek, J.N.K. Rao, C.E. Sarndal and M.P. Singh eds.) 3-20, Wiley, New York.
- BUTAR, F., AND LAHIRI, P. (2003), "On measures of uncertainty of empirical Bayes small-area estimators," *Journal of Statistical Planning and Inference*, 112, 63-76.
- CARTER, G. M., AND ROLPH, J.E. (1974), "Empirical Bayes methods applied to estimating fire alarm probabilities," *Journal of the American Statistical Association* 69, 880-885.
- CHEN, S. (2001), "Empirical Best Prediction and Hierarchical Bayes Methods in Small-Area Estimation," Ph.D. Dissertation, Department of Mathematics and Statistics, University of Nebraska-Lincoln.
- CHEN, S. AND LAHIRI, P. (2002), "A weighted jackknife MSPE estimator in small-area estimation," *Proceedings of the Survey Research Methods Section of the ASA*.
- CHEN, S. AND LAHIRI, P. (2003a), "A comparison of different MSPE estimators of EBLUP for the Fay-Herriot model," to appear in the *Proceedings of the Survey Research Methods Section of the ASA*.
- CHEN, S. AND LAHIRI, P. (2003b), "Jackknifing the mean square prediction error of EBLUP for the Fay-Herriot model," unpublished manuscript.
- EFRON, B. AND MORRIS C. (1973), "Stein's estimation rule and its competitors: an empirical Bayes approach," *Journal of the American Statistical Association* 68, 117-130.
- EFRON, B. AND MORRIS C. (1975), "Data analysis using Stein's estimator and its generalizations," *Journal of the American Statistical Association* 70, 311-9.

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 the American Statistical Association* 74, 269-277.

HENDERSON, C. R. (1975), Best linear unbiased estimation and prediction under a selection model, *Biometrics* 31, 423-447.

JIANG, J., LAHIRI, P., AND WAN, S. (2002), "A Unified Jackknife Method", *Annals of Statistics*, 30, 1782-1810.

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.

LAHIRI, P. (1995). "A jackknife measure of uncertainty of linear empirical Bayes estimators," unpublished manuscript.

LAHIRI, P. (2003), "Some thoughts on small-area estimation," unpublished manuscript.

MEZA, J. CHEN, S., AND LAHIRI, P. (2003), "Estimation of lifetime alcohol abuse for Nebraska counties," unpublished manuscript.

MORRIS, C. N. (1983), "Parametric empirical Bayes inference: theory and applications," *Journal of the American Statistical Association* 78, 47-59.

PRASAD, N.G.N. AND RAO, J.N.K. (1990), "The Estimation of Mean Squared Errors of Small Area Estimators," *Journal of the American Statistical Association* 85, 163-171.

RAO, J. N. K. (2003), *Small Area Estimation*, John Wiley & Sons.

RIVEST, L.P. AND BELMONTE, E. (1999), "The conditional mean square errors of small area estimators in survey sampling" Technical Report, Laval University.

