Proceedings of Statistics Canada Symposium 2001
Achieving Data Quality in a Statistical Agency: A Methodological Perspective

ON MEASURING THE QUALITY OF INDIRECT SMALL AREA ESTIMATES

J.N.K Rao[1]

ABSTRACT

Samples sizes in small areas are typically very small. As a result, customary area-specific direct estimators of small area means do not provide acceptable quality in terms of MSE. Indirect estimators that borrow strength from related areas through linking models based on auxiliary data are therefore now widely used for small area estimation. Such linking models are either implicit (as in the case of synthetic estimators) or explicit (as in the case of model based estimators). In the frequentist approach the quality of an indirect estimator is measured by its estimated MSE while the posterior variance of the small area mean is used in the Bayesian approach. In this paper, I will review some recent work on estimating MSE and the evaluation of posterior variance.

KEY WORDS: Composite Estimators; Linking Models; MSE; Posterior Variance.

1. INTRODUCTION

Sample surveys are typically designed to provide direct estimators with small sampling coefficient of variation (CV) for large areas (or domains). In fact, survey practitioners often stress that nonsampling errors, including measurement and coverage errors and nonresponse, contribute much more than sampling errors to total mean squared error (MSE) which is often used as a measure of quality of estimators. However, sampling errors play a dominant role in small area estimation because sample sizes in small areas are rarely large enough to provide area-specific direct estimators with acceptable quality in terms of sampling MSE (or CV). In fact, sample sizes are often zero in many small areas of interest. For example, in the estimation of county and school district counts of poor school age children in USA, using Current Population Survey (CPS) data and census and administrative data, CPS sample size is zero in many of the counties (National Research Council, 2000).

It is necessary to employ indirect estimators for small areas that borrow information from related areas through linking models, using census and administrative data. Such indirect estimators are based on either implicit or explicit linking models. Traditional indirect estimators based on implicit models include synthetic and composite estimators. But the estimation of sampling MSE of such estimators presents difficulties (Section 2). On the other hand, indirect estimators based on explicit models have received a lot of attention in recent years because of the following advantages over the traditional indirect estimators: (i) Explicit model-based methods make specific allowance for local variation through complex error structures in the model that link the small areas. (ii) Models can be validated from the sample data. (iii) Methods can handle complex cases such as cross-sectional and time series data and multivariate data. (iv) Stable area-specific measures of variability associated with the estimates may be obtained, unlike overall measures commonly used for the traditional indirect estimators.

In this paper, we provide an overview of developments on sampling MSE estimation of indirect estimators with emphasis on explicit model-based methods. Evaluation of posterior variance, using hierarchical Bayes

[1] J.N.K. Rao, Carleton University, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada, K1S 5B6

(HB) methods, is also briefly discussed; posterior variance is used as a measure of quality of the estimators in the Bayesian approach.

2. TRADITIONAL INDIRECT ESTIMATORS

Synthetic estimators, $\hat{Y}_i(S)$, of small area totals, Y_i, are widely used because of their simplicity and efficiency. Suppose that reliable direct estimators, $\hat{\bar{Y}}.g$, of large poststrata means, $\bar{Y}.g$, and the population cell counts $\{N_{ig}\}$ are both available. A simple synthetic estimator of Y_i is then given by

$$\hat{Y}_i(S) = \sum_g N_{ig} \hat{\bar{Y}}.g \ . \tag{2.1}$$

This estimator has a small variance and performs well in terms of MSE if the implicit model $\bar{Y}_{ig} \approx \bar{Y}.g$ holds for all *i*, where $\bar{Y}_{ig}$ is (*i*,*g*)-th the cell mean.

An unbiased estimator of MSE of $\hat{Y}_i(s)$ is given by

$$\begin{aligned} mse\left[\hat{Y}_i(S)\right] &= \left(\hat{Y}_i(S) - \hat{Y}_i\right)^2 - v\left(\hat{Y}_i(S) - \hat{Y}\right) + v\left(\hat{Y}_i(S)\right), \\ &\approx \left(\hat{Y}_i(S) - \hat{Y}\right)^2 - v\left(\hat{Y}_i\right), \end{aligned} \tag{2.2}$$

where $\hat{Y}_i$ is an unbiased direct estimator of Y_i and $v(.)$ denotes a variance estimator. The variance estimators of $\hat{Y}_i(S) - \hat{Y}_i$ and $\hat{Y}_i(S)$ are easily obtained using the jackknife method, especially for stratified multistage designs such as the CPS. Although (2.2) is unbiased, it is very unstable and may take negative values. A popular method to avoid instability is to take the average of $mse\left[\hat{\bar{Y}}_i(S)\right] = mse\left[\hat{Y}_i(S)\right] / N_i^2$ over the areas *i* and then use

$$mse_a\left(\hat{Y}_i(S)\right) = mse_a\left(\hat{\bar{Y}}_i(S)\right) N_i^2 \tag{2.3}$$

as the estimator of MSE, where $\hat{\bar{Y}}_i(S) = \hat{Y}_i(S) / N_i$ is the synthetic estimator of the mean $\bar{Y}_i$, N_i is the known area size and

$$mse_a\left[\hat{\bar{Y}}_i(S)\right] = \tfrac{1}{m} \sum_i mse\left[\hat{\bar{Y}}_i(S)\right]. \tag{2.4}$$

The MSE estimator (2.3) is stable but not area-specific except for the multiplier N_i^2.

Marker (1995) proposed an alternative MSE estimator that is more area-specific than (2.3) and also stable. It assumes that the squared bias, $b^2\left[\hat{\bar{Y}}_i(S)\right]$, of $\hat{\bar{Y}}_i(S)$ is approximated equal to the average squared bias:

$$b^2\left[\hat{\bar{Y}}_i(S)\right] \approx b_a^2\left[\hat{\bar{Y}}_i(S)\right] = mse_a\left[\hat{\bar{Y}}_i(S)\right] - \tfrac{1}{m} \sum_i \tfrac{1}{N_i^2} v\left[\hat{Y}_i(S)\right]. \tag{2.5}$$

Under this assumption, Marker's MSE estimator is given by

$$mse_M\left[\hat{Y}_i(S)\right] = v\left[\hat{Y}_i(S)\right] + N_i^2 b_a^2\left[\hat{\bar{Y}}_i(S)\right]. \tag{2.6}$$

For the simple synthetic estimator (2.1), $mse_M\left[\hat{Y}_i(S)\right]$ depends on area specific cell counts $\{N_{ig}\}$ through $v\left[\hat{Y}_i(S)\right]$ whereas $mse_a\left[\hat{Y}_i(S)\right]$ depends only on the overall size N_i of the *i*-th area. In this sense, Marker's MSE estimator is more area- specific.

Composite estimators of the form

$$\hat{Y}_i(C) = \hat{\phi}_i \hat{Y}_i + (1-\hat{\phi}_i)\hat{Y}_i(S) \tag{2.7}$$

are also often used, where $\hat{\phi}_i$ is an estimator of the optimal weight $\phi_i(opt)$ that minimizes $MSE(\hat{Y}_i(C))$. Such estimated weights are of the form

$$\hat{\phi}_i = mse[\hat{Y}_i(S)] / (\hat{Y}_i(S) - \hat{Y}_i)^2 \tag{2.8}$$

and highly unstable. To avoid the instability of $\hat{\phi}_i$, an average weight $\hat{\phi}_\bullet$ is often used in (2.7). An estimator of $MSE[\hat{Y}_i(C)]$ may be obtained along the above lines for $\hat{Y}_i(S)$ by changing $\hat{Y}_i(S)$ to $\hat{Y}_i(C)$ in (2.2) – (2.6). As a result, $\hat{Y}_i(C)$ has the same limitations as $\hat{Y}_i(S)$, in addition to difficulties of getting a stable weight $\hat{\phi}_i$.

External evaluations of indirect estimators are often conducted by comparing the estimators to true values. Gonzalez et al. (1996) studied count-synthetic estimators

$$\hat{P}_i(S) = \left(\sum_g N_{ig}\hat{P}_{ig}\right) \Big/ \left(\sum_g N_{ig}\right) \tag{2.9}$$

of proportions P_i, where $\hat{P}_{.g}$ is the direct estimator of post-stratum proportion $P_{.g}$. They studied the performance of $\hat{P}_i(S)$ relative to the direct estimator $\hat{P}_i$, using data from the 1980 U.S. National Natality Survey. For external evaluation, they obtained true P_i for three characteristics (low birth weight, prenatal care and Apgar score) and used the unbiased MSE estimator.

$$mse[\hat{P}_i(S)] = (\hat{P}_i(S) - P_i)^2 . \tag{2.10}$$

Performance of $\hat{P}_i(S)$ was judged on the basis of estimated CVs computed from (2.10). On the other hand, balanced repeated replication (BRR) was used to estimate the variance of $\hat{P}_i$. However, caution should be exercised in drawing conclusions from such comparisons because (2.10) is highly unstable. In this case, only overall comparisons may be reliable. For example, one could compare the average or the median absolute relative error (*ARE*), where $ARE_i = |est_i - true_i| / true_i$ for the *i*-th area. A more refined comparison may be implemented by comparing average or median ARE within specified categories of the small areas. National Research Council (2000) reported such comparisons in the context of county estimates of poor school age children.

3. SMALL AREA MODELS

Two types of basic small area models have been studied in the literature. In the first type, called basic area-level model, only area-specific auxiliary data $z_i = (z_{1i},..., z_{pi})^T$, related to small area means $\overline{Y}_i (i = 1,..., m)$ or to suitable functions $\theta_i = g(\overline{Y}_i)$, are used to develop linking models of the form $\theta_i \overset{ind}{\sim} N(z_i^T\beta, \sigma_v^2)$. The linking model is combined with the sampling model $\hat{\theta}_i|\theta_i \overset{ind}{\sim} N(\theta_i, \psi_i)$ where $\hat{\theta}_i$ is a direct estimator of θ_i with known sampling variance ψ_i. Model-based estimators of θ_i and $\overline{Y}_i = g^{-1}(\theta_i)$ are obtained from the combined model using empirical Bayes (EB) or hierarchical Bayes (HB) methods. Measures of variability associated with the estimators are also obtained as indicators of quality of the estimators. The basic area-level model with $\theta_i = \log Y_i$ has been recently used to produce EB county estimates of poor school-age children in U.S.A. (National Research Council, 2000). Using these estimates, the US Department of Education allocates over 7 billion dollars of federal funds annually to counties. Various extensions of the basic area level model have been proposed to handle correlated sampling errors, spatial dependence of θ_i's, time series and cross-sectional data and others (see Rao, 1999 for a recent overview).

In the second type, called basic unit-level model, unit-level auxiliary variables $\mathbf{x_{ij}} = (x_{1ij},...,x_{pij})^T$ are related to the unit y-values y_{ij} through a one-way nested error regression model $y_{ij} = x_{ij}^T\beta + v_i + e_{ij}$, where $v_i \overset{iid}{\sim} N(0,\sigma_v^2)$ are independent of $e_{ij} \overset{iid}{\sim} N(0,\sigma^2)$. The parameters of interest are the small area means $\overline{Y}_i = \sum_j y_{ij} / N_i \left(j = 1, N_i, i = 1,...,m \right)$. Various extensions of the basic unit-level model have been proposed to handle binary responses, multivariate responses, two-stage sampling within small area, and others (see Rao, 1999). In this paper, we focus on the basic area level model for simplicity, and review some recent work on estimating MSE and evaluating posterior variance.

4. EMPIRICAL BAYES (EB) METHOD

4.1 Estimator of θ_i

Under the basic area-level model, the best estimator of θ_i, in the sense of minimum MSE, is given by the conditional expectation $E(\theta_i|\hat{\theta}_i)$:

$$E(\theta_i|\hat{\theta}_i) = \gamma_i\hat{\theta}_i + (1-\gamma_i)\mathbf{z_i^T\beta}, \tag{4.1}$$

where $\gamma_i = \sigma_v^2/(\sigma_v^2 + \psi_i)$. This estimator is also called the Bayes estimator and denoted $\hat{\theta}_i^B$. It depends on the unknown model parameters $(\boldsymbol{\beta},\sigma_v^2)$. Replacing $(\boldsymbol{\beta},\sigma_v^2)$ by suitable estimators $(\hat{\boldsymbol{\beta}},\hat{\sigma}_v^2)$, obtained from the marginal distribution of $\hat{\theta}_i$'s, namely $\hat{\theta}_i \overset{ind}{\sim} N(\mathbf{z_i^T\beta}, \sigma_v^2 + \psi_i)$, we obtain the empirical Bayes or empirical best (EB) estimator $\hat{\theta}_i^{EB}$:

$$\hat{\theta}_i^{EB} = \hat{\gamma}_i\hat{\theta}_i + (1-\hat{\gamma}_i)\mathbf{z_i^T}\hat{\boldsymbol{\beta}}, \tag{4.2}$$

where $\hat{\gamma}_i = \hat{\sigma}_v^2/(\hat{\sigma}_v^2 + \psi_i)$. The form (4.2) shows that $\hat{\theta}_i^{EB}$ is a weighted average of the direct estimator, $\hat{\theta}_i$, and the regression synthetic estimator, $\mathbf{z_i^T}\hat{\boldsymbol{\beta}}$ with weights $\hat{\gamma}_i$ and $1-\hat{\gamma}_i$ respectively; $\hat{\gamma}_i$ is a measure of between area variability relative to total variability associated with area *i*. The estimator $\hat{\theta}_i^{EB}$ is model unbiased in the sense of $E(\hat{\theta}_i^{EB} - \theta_i) = 0$.

For a given σ_v^2, the estimator of $\boldsymbol{\beta}$ is the weighted least squares estimator $\tilde{\boldsymbol{\beta}}(\sigma_v^2)$ and it does not require normality assumption. Similarly, without normality assumption σ_v^2 may be estimated by a simple method of moments (Prasad and Rao, 1990) or by solving the following moment equation iteratively for σ_v^2 (Fay and Herriot, 1979):

$$h(\sigma_v^2) = \sum_i \left(\hat{\theta}_i - z_i^T\tilde{\beta}(\sigma_v^2)\right)^2 / (\sigma_v^2 + \psi_i) = m - p. \tag{4.3}$$

The resulting estimators $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\sigma}_v^2)$ and $\hat{\sigma}_v^2$ substituted in (4.1) lead to empirical best linear unbiased prediction (EBLUP) estimator, not depending on normality. It is identical to $\hat{\theta}_i^{EB}$ given by (4.2). Under normality, one could use maximum likelihood (ML) or restricted maximum likelihood (REML) estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_v^2$ which remain consistent for $\boldsymbol{\beta}$ and σ_v^2 even under nonnormality (Jiang, 1996). The estimator of $\overline{Y}_i$ is taken as $g^{-1}(\hat{\theta}_i^{EB})$. More complex estimators of $\overline{Y}_i$ with reduced bias have also been proposed.

For example, with $g(\bar{Y}_i) = \log \bar{Y}_i$ one could use $\exp\left[\hat{\theta}_i^{EB} + \frac{1}{2} mse\left(\hat{\theta}_i^{EB}\right)\right]$ assuming that $\hat{\theta}_i^{EB}$ is normal (or $\exp\left(\hat{\theta}_i^{EB}\right)$ is lognormal), where $mse\left(\hat{\theta}_i^{EB}\right)$ is an estimator of $MSE\left(\hat{\theta}_i^{EB}\right) = E\left(\hat{\theta}_i^{EB} - \theta_i\right)^2$; see National Research Council (2000).

4.2 MSE estimators

Under normality, the conditional (or posterior) distribution of θ_i given $\hat{\theta}_i$ is $N\left(\hat{\theta}_i^{B}, g_{1i}(\sigma_v^2)\right)$, where

$$g_{1i}(\sigma_v^2) = \gamma_i \psi_i . \tag{4.4}$$

A naive EB approach uses the estimated posterior distribution $N\left(\hat{\theta}_i^{EB}, g_{1i}(\hat{\sigma}_v^2)\right)$ for making inference on θ_i. In particular, the mean of the estimated posterior distribution, $\hat{\theta}_i^{EB}$, is used as the estimator of θ_i, and the variance of the estimated posterior distribution, $g_{1i}(\hat{\sigma}_v^2)$, as the measure of variability. This approach can lead to severe underestimation of MSE of $\hat{\theta}_i^{EB}$ because $g_{1i}(\hat{\sigma}_v^2)$ ignores the variability associated with $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_v^2$. In the EBLUP approach, a naive estimator of $MSE\left(\hat{\theta}_i^{EB}\right)$, without normality assumption, is given by

$$mse_N\left(\hat{\theta}_i^{EB}\right) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2), \tag{4.5}$$

where

$$g_{2i}(\sigma_v^2) = (1-\gamma_i)^2 \mathbf{z}_i^{\mathbf{T}} \left[\sum_i \mathbf{z}_i \mathbf{z}_i^{\mathbf{T}} \big/ (\sigma_v^2 + \psi_i)\right]^{-1} \mathbf{z}_i . \tag{4.6}$$

The last term $g_{2i}(\hat{\sigma}_v^2)$ in (4.5) accounts for the variability of $\hat{\boldsymbol{\beta}}$ but not of $\hat{\sigma}_v^2$. Note that $mse_N\left(\hat{\theta}_i^{EB}\right)$ is an improvement over the naive EB measure, $g_{1i}(\hat{\sigma}_v^2)$.

Methods that account for the variability of both $\hat{\beta}$ and $\hat{\sigma}_v^2$ have been studied extensively in recent years. We summarize some of the developments here. An accurate approximation of $MSE\left(\hat{\theta}_i^{EB}\right)$ under normality is given by

$$MSE\left(\hat{\theta}_i^{EB}\right) \approx g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2), \tag{4.7}$$

where

$$g_{3i}(\sigma_v^2) = \left[\psi_i^2 \big/ (\sigma_v^2 + \psi_i)^4\right] E\left(\hat{\theta}_i - \mathbf{z}_i^{\mathbf{T}} \boldsymbol{\beta}\right)^2 h(\sigma_v^2) \tag{4.8}$$

$$= \left[\psi_i^2 \big/ (\sigma_v^2 + \psi_i)^3\right] h(\sigma_v^2) \tag{4.9}$$

and $h(\sigma_v^2)$ is the asymptotic variance of $\hat{\sigma}_v^2$ for large m. Neglected terms in the approximation (4.7) are of lower order than m^{-1}. The approximation (4.7) is valid for all the methods of estimating σ_v^2 considered in section 4.1.

For the Prasad-Rao (PR) moment estimator of σ_v^2, we have

$$h_{PR}(\sigma_v^2) = 2m^{-2} \sum_i (\sigma_v^2 + \psi_i)^2 . \tag{4.10}$$

The asymptotic variance for the Fay-Herriot (FH) estimator of σ_v^2 is

$$h_{FH}(\sigma_v^2) = 2m \left[\sum_i (\sigma_v^2 + \psi_i)^{-1}\right]^{-2} \tag{4.11}$$

(Datta, Rao and Smith, 2001). It follows from (4.10) and (4.11) that

$$h_{FH}(\sigma_v^2) \le h_{PR}(\sigma_v^2) \tag{4.12}$$

with equality holding when $\psi_i = \psi$ for all *i*. For the ML or REML estimators of σ_v^2, we have

$$h_{ML}\left(\sigma_v^2\right) = h_{REML}\left(\sigma_v^2\right) = 2\left[\sum_i \left(\sigma_v^2 + \psi_i\right)^{-2}\right]^{-1} \leq h_{FH}\left(\sigma_v^2\right). \tag{4.13}$$

It follows from (4.12) and (4.13) that

$$h_{ML}\left(\sigma_v^2\right) = h_{REML}\left(\sigma_v^2\right) \leq h_{FH}\left(\sigma_v^2\right) \tag{4.14}$$

with equality holding when $\psi_i = \psi$ for all *i*. Thus, ML or REML lead to the smallest MSE followed by FH.

Turning to MSE estimation, an estimator correct to the same order of approximation as (4.7) is given by

$$mse\left(\hat{\theta}_i^{EB}\right) \approx g_{1i}\left(\hat{\sigma}_v^2\right) + g_{2i}\left(\hat{\sigma}_v^2\right) + 2g_{3i}\left(\hat{\sigma}_v^2\right). \tag{4.15}$$

This estimator is nearly unbiased for $MSE\left(\hat{\theta}_i^{EB}\right)$ in the sense that its bias is of lower order than m^{-1}. The approximation (4.15) is valid for the REML and PR estimators of σ_v^2 but not for ML and FH estimators of σ_v^2.

Lahiri and Rao (1995) showed that (4.15), using PR estimator of σ_v^2, is robust to nonnormality of the θ_i's in the sense that near unbiasedness remains valid. Note that normality of the direct estimators $\hat{\theta}_i$'s is still assumed but it is less restrictive than the normality of θ_i's because of the central limit theorem's effect on $\hat{\theta}_i$'s. It is not known if the robustness remains valid under the REML estimator of σ_v^2.

For ML and FH estimators, an extra term $g_{10}\left(\hat{\sigma}_v^2\right)$ is added (4.15). This extra term for ML is positive (Datta and Lahiri, 2000). Therefore, ignoring this term and using (4.15) with ML estimator $\hat{\sigma}_v^2$ would lead to underestimation of MSE. On the other hand, the extra term for FH is negative (Datta, Rao and Smith, 2001). Therfore, ignoring this term and using (4.15) with FH estimator $\hat{\sigma}_v^2$ would lead to overestimation of MSE.

A criticism of the MSE estimator (4.15) and its modification for ML or FH is that it is not area-specific in the sense that it does not depend on the direct estimator $\hat{\theta}_i$ although $\mathbf{z_i}$ is involved in the g_{2i}-term. But it is easy to find other choices, using the form (4.8) for $g_{3i}\left(\sigma_v^2\right)$. For example, we can use

$$mse_1\left(\hat{\theta}_i^{EB}\right) = g_{1i}\left(\hat{\sigma}_v^2\right) + g_{2i}\left(\hat{\sigma}_v^2\right) + g_{3i}\left(\hat{\sigma}_v^2\right) + \left[\psi_i^2 / \left(\sigma_v^2 + \psi_i\right)^4\right]\left(\hat{\theta}_i - \mathbf{z_i^T}\hat{\boldsymbol{\beta}}\right)^2 h\left(\hat{\sigma}_v^2\right) \tag{4.16}$$

(Rao, 2000). The last term of (4.16) is less stable then $g_{3i}\left(\hat{\sigma}_v^2\right)$ but it is of lower order than the leading term $g_{1i}\left(\hat{\sigma}_v^2\right)$. As a result, the variablility of $mse_1\left(\hat{\theta}_i^{EB}\right)$ should be comparable to the variability of $mse\left(\hat{\theta}_i^{EB}\right)$, at least for moderate to large *m*.

In the above discussion, we have used MSE to measure the variability of $\hat{\theta}_i^{EB}$. Another approach tries to correct the underestimation induced by the estimated posterior distribution by mimicking the hierarchical Bayes (HB) approach (Section 5). Under this approach, it is necessary to entertain a prior distribution on the model parameters $\boldsymbol{\beta}$ and σ_v^2 to arrive at the posterior variance $V\left(\theta_i \middle| \hat{\boldsymbol{\theta}}\right)$ which is used as a measure of variability under the HB framework, where $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_m\right)^T$. Kass and Steffey (1989) used an asymptotic first order approximation to the posterior variance which does not depend on the form of the prior distribution on $\boldsymbol{\beta}$ and σ_v^2. Laird and Louis (1987) first expressed $V\left(\theta_i \middle| \hat{\boldsymbol{\theta}}\right)$ as

$$V\left(\theta_i \middle| \hat{\boldsymbol{\theta}}\right) = E_{\boldsymbol{\beta},\sigma_v^2}\left[V\left(\theta_i \middle| \hat{\theta}_i, \boldsymbol{\beta}, \sigma_v^2\right)\right] + V_{\boldsymbol{\beta},\sigma_v^2}\left[E\left(\theta_i \middle| \hat{\theta}_i, \boldsymbol{\beta}, \sigma_v^2\right)\right], \tag{4.17}$$

where $E_{\boldsymbol{\beta},\sigma_v^2}$ and $V_{\boldsymbol{\beta},\sigma_v^2}$ respectively denote the expectation and variance with respect to the posterior distribution $f\left(\boldsymbol{\beta},\sigma_v^2 \middle| \hat{\boldsymbol{\theta}}\right)$. The last two terms of (4.17) were then estimated separately using the parametric bootstrap method (see Ghosh and Rao, 1994). Note that the last term of (4.17) accounts for the underestimation while the second term is roughly equal to $g_{1i}\left(\hat{\sigma}_v^2\right)$, the variance of the estimated posterior distribution of θ_i. Butar and Latiri (1997) showed that the bias of Laird-Louis estimator (as an estimator of MSE) is of order m^{-1}, unlike the bias of (4.15) or (4.16). By correcting this bias, they obtained an MSE estimator which is identical to $mse_1\left(\hat{\theta}_i^{EB}\right)$ given by (4.16). Kass and Steffey's (1989) first order estimator is also biased. A second order estimator of Kass and Steffey depends on the prior distribution of $\boldsymbol{\beta}$ and σ_v^2.

4.3 Jackknife estimator of MSE

Jiang, Lahiri and Wang (1999) proposed a jackknife method of estimating MSE that is applicable to general models with block diagonal covariance structures, where the blocks correspond to the small areas. We illustrate the jackknife method for the basic area-level model. It uses the following orthogonal decomposition of MSE:

$$\begin{aligned} MSE\left(\hat{\theta}_i^{EB}\right) &= E\left(\hat{\theta}_i^{B}-\theta_i\right)^2 + E\left(\hat{\theta}_i^{EB}-\hat{\theta}_i^{B}\right)^2 \\ &= g_{1i}\left(\sigma_v^2\right) + E\left(\hat{\theta}_i^{EB}-\hat{\theta}_i^{B}\right)^2 . \end{aligned} \tag{4.18}$$

Write $\hat{\theta}_i^{EB}$ as $\hat{\theta}_i = k\left(\hat{\theta}_i, \hat{\boldsymbol{\delta}}\right)$, where $\boldsymbol{\delta} = \left(\boldsymbol{\beta}^T, \sigma_v^2\right)^T$. The EB estimator $\hat{\theta}_i^{EB}$ may be expressed as $k\left(\hat{\theta}_i, \hat{\boldsymbol{\delta}}\right)$. The jackknife steps are as follows:

(i) Calculate $\hat{\boldsymbol{\delta}}(l) = \left(\hat{\boldsymbol{\beta}}(l), \hat{\sigma}_v^2(l)\right)$, the estimator of $\boldsymbol{\delta}$ when the *l*-th area data $\left(\hat{\theta}_l, \mathbf{z}_l\right)$ is deleted. Let $\hat{\theta}_i^{EB}(l) = k\left(\hat{\theta}_i, \hat{\boldsymbol{\delta}}(l)\right)$. Note that $\hat{\theta}_i$ remains unchanged in $\hat{\theta}_i^{EB}(l)$.

(ii) Calculate

$$M_{2i} = \tfrac{m-1}{m} \sum_{i=1}^{m} \left(\hat{\theta}_i^{EB}(l) - \hat{\theta}_i^{EB}\right)^2 . \tag{4.19}$$

M_{2i} is a jackknife estimator of the last term in (4.18).

(iii) Adjust the bias of $g_{1i}\left(\hat{\sigma}_v^2\right)$ (as an estimator of $g_{1i}\left(\sigma_v^2\right)$) using the jackknife bias reduction method. The adjusted estimator of $g_{1i}\left(\sigma_v^2\right)$ is

$$\hat{M}_{1i} = g_{1i}\left(\hat{\sigma}_v^2\right) - \tfrac{m-1}{m} \sum_{i=1}^{m} \left[g_{1i}\left(\hat{\sigma}_v^2(l)\right) - g_{1i}\left(\hat{\sigma}_v^2\right)\right]. \tag{4.20}$$

(iv) Calculate the jackknife estimator of MSE as

$$mse_J\left(\hat{\theta}_i^{EB}\right) = \hat{M}_{1i} + \hat{M}_{2i} . \tag{4.21}$$

The jackknife MSE estimator (4.21) is approximately unbiased in the sense of bias lower order than m^{-1}. It is also area-specific, similar to (4.16). But the computations can be cumbersome if iterative calculations are involved, as in the case of ML or REML because the estimates of model parameters $\boldsymbol{\delta}$, are recomputed *m* times by deleting each area in turn. The computations can be significantly reduced for ML or REML by using a single step of the Newton-Raphson algorithm with the estimate $\hat{\boldsymbol{\delta}}$ from the full sample as starting value.

Pfeffermann and Tiller (2001) used a bootstrap version of the Jiang-Lahiri method in the context of Kalman filter time series models. This approach should be applicable to small area models as well.

4.4 Alternative MSE estimators

Booth and Hubert (1998) argued that the conditional MSE of EB estimator given the *i*-th area data is more relevant as a measure of variability than the unconditional MSE because the former is area-specific. Fuller (1989) earlier proposed a similar measure in the context of linear mixed models. But the MSE estimator (4.16) and the jackknife estimator (4.21) show that it is possible to get area-specific estimator of the unconditional MSE. In fact, for the basic area-level model (4.16) is closely related to Fuller's estimator of the conditional MSE.

It is more appealing to survey practitioners to consider the estimation of sampling MSE of $\hat{\theta}_i^{EB}$, i.e., $MSE_p(\hat{\theta}_i^{EB}) = E_p(\hat{\theta}_i^{EB} - \theta_i)^2$, where the expectation, E_p, is with respect to the sampling distribution $f(\hat{\theta}_i|\theta_i)$. Rivest and Belmonte (2000) derived a design-unbiased estimator, $mse_p(\hat{\theta}_i^{EB})$, using the PR estimator of σ_v^2. In the simple case of known $\boldsymbol{\beta}$ and σ_v^2, we have $\hat{\theta}_i^{EB} = \hat{\theta}_i^B$ and

$$mse_p(\hat{\theta}_i^B) = \psi_i\gamma_i + \left(\frac{\psi_i}{\psi_i + \sigma_v^2}\right)^2 \left[(\hat{\theta}_i - z_i^T\beta)^2 - \psi_i - \sigma_v^2\right]. \tag{4.22}$$

On the other hand, $mse(\hat{\theta}_i^B) = g_{1i}(\sigma_v^2) = \psi_i\gamma_i$. Rivest and Bellmonte (2000) studied the relative performance of $mse_p(\hat{\theta}_i^B)$ and $mse(\hat{\theta}_i^B)$ in estimating $MSE_p(\hat{\theta}_i^B)$ for the special case of $\psi_i = \psi$. They calculated the ratio of average MSE of $mse_p(\hat{\theta}_i^B)$ to the average MSE of $mse_p(\hat{\theta}_i^B)$, where the average is over the small areas *i*. This ratio reduces to $R = (\psi^2 + 2\psi\sigma_v^2)/\sigma_v^4$ which is greater than 1 if $\sigma_v^2/\psi < 2.4$. When shrinking is not light, i.e., when γ_i is small, $mse(\hat{\theta}_i^B)$ performs much better than $mse_p(\hat{\theta}_i^B)$. For example, if $\gamma_i = \frac{1}{2}$ or $\sigma_v^2/\psi = 1$, we have $R = 3$.

4.5 Comparison of MSE estimators

Datta, Rao and Smith (2001) conducted a simulation study of the relative bias of MSE estimators studied in Section 4.2. They considered a simple model: $\hat{\theta}_i \overset{ind}{\sim} N(\theta_i, \psi_i)$ and $\theta_i \overset{ind}{\sim} N(0,1)$, $i = 1,\ldots,m$ and $m = 15, 30$, and two patterns for ψ_i: (a) moderate variation over $i: 0.7$ to 0.3; (b) large variation over $i: 4.0$ to 0.1. They generated 10,000 samples for each pattern of ψ_i and m. For the pattern (a), the MSE estimators based on FH, ML, REML and PR are comparable in terms of relative bias, but FH performed well over all combinations of m and ψ_i-patterns. These results strongly suggest that the FH-based MSE estimator is robust over ψ_i-patterns while the FH-based estimator, $\hat{\theta}_i^{EB}$, maintains good efficiency.

5. HIERARCHRICAL BAYES (HB) METHOD

A disadvantage of $\hat{\theta}_i^{EB}$ is that the weight $\hat{\gamma}_i$ can take zero value in which case $\hat{\theta}_i^{EB}$ reduces to the regression synthetic estimator $\mathbf{z}_i^T\hat{\boldsymbol{\beta}}$. Thus, the direct estimators, $\hat{\theta}_i$, receive zero weight even when the sample size in some areas is not small. This difficulty was encountered in using a state model to produce EB state estimates of poor school-age children in U.S.A. (National Research Council, 2000). The HB method avoids this difficulty by producing positive weights in all cases. Moreover, the HB approach is straight forward, inferences are ″exact″ and complex problems can be handled using recently developed Monte Carlo Markov Chain (MCMC) methods. A prior distribution on the model parameters $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \sigma_v^2)^T$

is specified and inferences are based on the posterior distribution $f\left(\theta_i|\hat{\boldsymbol{\theta}}\right)$; in particular θ_i is estimated by the posterior mean $E\left(\theta_i|\hat{\boldsymbol{\theta}}\right)$ and its precision is measured by the posterior variance $V\left(\theta_i|\hat{\boldsymbol{\theta}}\right)$ given by (4.17).

MCMC methods are used to generate simulated samples $\left\{\theta_1^{(j)},...,\theta_m^{(j)}; j=1,...,J\right\}$ from the joint posterior distribution $f\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}\right)$. For the basic area level model, simulated samples can be readily generated from the conditional distributions $\boldsymbol{\beta}|\boldsymbol{\theta},\sigma_v^2,\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}|\boldsymbol{\beta},\sigma_v^2,\hat{\boldsymbol{\theta}}$ and $\sigma_v^{-2}|\boldsymbol{\beta},\boldsymbol{\theta},\hat{\boldsymbol{\theta}}$ using Gibbs sampling. We have

$$\hat{\theta}_i^{HB} = E\left(\theta_i|\hat{\boldsymbol{\theta}}\right) \approx \tfrac{1}{J}\sum_{j=1}^{J}\theta_i^{(j)} = \theta_i^{(.)} \tag{5.1}$$

and

$$V\left(\theta_i|\hat{\boldsymbol{\theta}}\right) \approx \tfrac{1}{J}\sum_{j=1}^{J}\left(\theta_i^{(j)}-\theta_i^{(.)}\right)^2 . \tag{5.2}$$

More efficient estimators (in terms of reduced simulation error) can also be obtained. HB estimate of $\overline{Y}_i$ and its posterior variance can also be readily obtained from (5.1) and (5.2) by changing $\theta_i^{(j)}$ to $g^{-1}\left(\theta_i^{(j)}\right)=\overline{Y}_i^{(j)}$.

Bell (1999) applied the HB approach to the state model mentioned above, using improper (diffuse) priors on $\boldsymbol{\beta}$ and σ_v^2 : $f(\boldsymbol{\beta})$ = constant and $f\left(\sigma_v^2\right)$ = constant $\left(0<\sigma_v^2<\infty\right)$. He obtained HB estimates with positive weights in all cases.

Recently, Yong and Rao (2001) applied the HB approach to handle unmatched sampling and linking models and applied it to Canadian census undercount estimation. In this application, c_i = census count, u_i = number missing and y_i is a post-enumeration survey estimator of u_i with known variance ξ_i. The sampling model is given by $y_i \mid u_i \overset{ind}{\sim} N\left(u_i,\xi_i\right)$ and the linking model is given by $\theta_i = \log\left\{u_i/\left(u_i+c_i\right)\right\} \overset{ind}{\sim} N\left(\mathbf{z}_i^T\boldsymbol{\beta},\sigma_v^2\right)$. Note that the basic area-level model uses matched sampling and linking models, but the assumption $E_p\left[\hat{\theta}_i\right]=\theta_i$ may not be valid if small area sample size is small, where $\hat{\theta}_i = \log\left\{y_i/\left(y_i+c_i\right)\right\}$.

The HB approach is attractive but caution should be exercised when using MCMC methods (see e.g., Rao (1999)). For example, the Gibbs sampler could lead to seemingly reasonable inferences about a nonexistent posterior distribution. This happens when the posterior distribution is improper and yet all the Gibbs conditional distributions are proper (Hobert and Casella, 1996). Another difficulty with MCMC is that the convergence diagnostic tools can fail to detect the sort of convergence failure they were designed to identify (Cowles and Carlin, 1996).

REFERENCES

Bell, W.R. (1999), "Accounting for Uncertainty About Variances in Small Area Estimation", *Bulletin of the International Statistical Institute*, pp.

Booth, J.G., and Hobert, J.P. (1998), "Standard Errors of Predictors in Generalized Linear Mixed Models", *Journal of the American Statistical Association*, 6, pp. 363-372.

Butar, P.B., and Lahiri, P. (1997), "On the Measure of Uncertainty of Empirical Bayes Small Area Estimators", Unpublished Report, Lincoln, Nebraska: University of Nebraska-Lincoln.

Cowles, M.K., and Carlin, B.P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review", *Journal of the American Statistical Association*, 91, pp. 883-904.

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, pp. 613-627.

Datta, G.S, Rao, J.N.K., and Smith, D.D. (2001), "On Measures of Uncertainty of Small Area Estimates in the Fay-Herriot Model", Unpublished Report, Athens, Georgia: University of Georgia.

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

Fuller, W.A. (1989), "Prediction of True Values for the Measurement Error Model", paper presented at the Conference on Statistical Analysis of Measurement Error Models, Humboldt State University.

Ghosh, M., and Rao, J.N.K. (1994), "Small Area Estimation: an Appraisal", (with discussion), *Statistical Science*, 9, pp. 55-93.

Gonzalez, Jr., J.F., Placek, P.J. and Scott, C. (1996), "Synthetic Estimation in Follow back Surveys at the National Center for Health Statistics", in W.L. Schaible (ed.) *Indirect Estimators in U.S. Federal Programs*, New York: Springer, pp. 16-27.

Hobert, J.P., and Casella, G. (1996), "The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models", *Journal of the American Statistical Association*, 91, pp. 1461-1479.

Jiang, J. (1996), "REML Estimation: Asymptotic Behavior and Related Topics", *Annals of Statistics*, 24, pp. 255-286.

Jiang, J., Lahiri, P., and Wan, S. (1999), "Jackknifing the Mean Squared Error of Empirical Best Predictor", Unpublished Report, Lincoln, Nebraska: University of Nebraska-Lincoln.

Kass, R.E, and Steffey, D. (1989), "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)", *Journal of the American Statistical Association*, 84, pp. 717-726.

Lahiri, P, and Rao, J.N.K. (1995), "Robust Estimation of Mean Squared Error of Small Area Estimators", *Journal of the American Statistical Association*, 82, pp. 758-766.

Laird, N.M., and Louis, T.A. (1987), "Empirical Bayes Confidence Intervals Based on Bootstrap Samples", *Journal of the American Statistical Association*, 82, pp. 739-750.

Marker, D.A. (1995), *Small Area Estimation: A Bayesian Perspective*, unpublished Ph.D. thesis, Ann Arbor, Michigan: University of Michagan.

National Research Council (2000), *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology*, C.F. Citro and G. Kalton (eds.), Committee on National Statistics, Washington, D.C.: National Academy Press.

Pfefferman, D, and Tiller, R. (2001), "Bootstrap Approximation to Prediction MSE for state-space Models with Estimated Parameters", Unpublished Report, Jerusalem, Israel: Hebrew University.

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, pp. 163-171.

Rao, J.N.K. (1999), "Some Recent Advances in Model-Based Small Area Estimation", *Survey Methodology*, 25, pp. 175-186.

Rao, J.N.K. (2000), "EB and EBLUP in Small Area Estimation" in S.E. Ahmed and N. Reid (eds.) *Empirical Bayes Likelihood Inference*, New York: Springer, pp. 33-43.

Rivest, L-P., and Belmonte, E. (2000), "A Conditional Mean Squared Error of Small Area Estimators", *Survey Methodology*, 26, pp. 67-78.

You, Y., and Rao, J.N.K. (2002), "Small Area Estimation Using Unmatched Sampling and Linking Models", *Canadian Journal of Statistics*, 29, in press.