JIRSS (2003)
Vol. 2, No. 2, pp 145-169

Some New Developments in Small Area Estimation

J. N. K. Rao

Carleton University, School of Mathematics and Statistics, Ottawa, Canada.
(jrao@math.carleton.ca)

Abstract. Small area estimation has received a lot of attention in recent years due to growing demand for reliable small area statistics. Traditional area-specific estimators may not provide adequate precision because sample sizes in small areas are seldom large enough. This makes it necessary to employ indirect estimators based on linking models. Basic area level and unit level models have been extensively studied in the literature to derive empirical best linear unbiased prediction (EBLUP), empirical Bayes (EB) and hierarchical Bayes (HB) small area estimators and associated measures of variability. In this paper, I will cover several important new developments related to model-based small area estimation.

1 Introduction

Due to cost and other considerations, sample surveys are typically designed to provide area-specific (or direct) estimators with small sampling coefficient of variation (CV) for large areas (or domains).

Received: August 2003
Key words and phrases: Area level,empirical Bayes, hierarchical Bayes, linking models, unit level.

In fact, survey practitioners often stress that nonsampling errors, including measurement and coverage errors, 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 seldom large enough to provide direct estimators with acceptable quality in terms of sampling MSE (or CV). In fact, sample sizes can be zero in many small areas of interest. For example, data from the Current Population Survey (CPS) are used to estimate county (and school district) counts of poor school age children in the United States, but the CPS sample sizes are zero in many of the counties (National Research Council, 2000).

Due to difficulties with direct estimators, it is often necessary to employ indirect estimates that borrow information from related areas through explicit (or implicit) linking models, using census and administrative data associated with the small areas. Indirect estimators based on explicit linking models have received a lot of attention in recent years because of the following advantages over the traditional indirect estimators based on implicit models: (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, binary or count data, spatially-correlated data and multivariate data. (iv) Area-specific measures of variability associated with the estimates may be obtained, unlike overall measures commonly used with the traditional indirect estimators.

Basic area level and unit level models have been extensively studied in the literature to derive empirical best linear unbiased prediction (EBLUP), empirical Bayes (EB) and hierarchical Bayes (HB) small area estimators of totals (or means) and associated measures of variability. The EBLUP method is applicable for linear mixed models that cover the basic area level and unit level models. On the other hand, EB and HB methods are more generally applicable, covering generalized linear mixed models that are used to handle categorical (e.g., binary) and count data. MSE is used as a measure of variability under the EBLUP and EB approaches, while the HB approach uses the posterior variance as a measure of variability, assuming a prior distribution on the model parameters. We refer the reader to Rao (2003) for an extensive account of EBLUP, EB and HB methods for

small area estimation. Recent review papers on small area estimation include Ghosh and Rao (1994), Rao (1999) and Pfeffermann (2002).

In this paper, I will cover several important new developments related to model-based small area estimation, including unmatched sampling and linking area level models, use of sampling weights in unit level models, jackknife methods for MSE estimation, and MSE estimation for area level models when the sampling variances are estimated. Section 2 gives the basic small area models and some extensions. Results for the basic area level model under EBLUP (or EB) estimation are presented in Section 3. Jackknife estimation of MSE of EB estimators that can handle generalized linear mixed models is studied in Section 4. A pseudo-EBLUP method for the basic unit level model that takes account of survey weights is given in Section 5. HB estimation under unmatched sampling and linking models is studied in Section 6, as well as "matching" priors that lead to well-calibrated inferences. Section 7 presents some recent applications of EBLUP, EB and HB methods. Finally, some practical issues are discussed in Section 8.

2 Small area models

Two types of basic small area models have been studied in the literature. In the first type, called the basic area level model, only area-specific auxiliary data $\mathbf{z}_i = (z_{1i}, \ldots, z_{pi})^T$, related to some suitable functions $\theta_i = g(Y_i)$ of the small area total Y_i $(i = 1, \ldots, m)$, are used to develop a linking model of the form $\theta_i = \mathbf{z}_i^T \boldsymbol{\beta} + v_i$ with $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$, where σ_v^2 is the model variance. The linking model is combined with the matching sampling model $\hat{\theta}_i = \theta_i + e_i$, where $\hat{\theta}_i = g(\hat{Y}_i)$ is a direct estimator of θ_i and $e_i|\theta_i \overset{\text{ind}}{\sim} N(0, \psi_i)$ with known sampling variance ψ_i. The combined model, $\hat{\theta}_i = \mathbf{z}_i^T \boldsymbol{\beta} + v_i + e_i$, is a special case of the linear mixed model.

The basic area level model has at least two limitations. First, the assumption of known sampling variances, ψ_i, is restrictive, although methods based on generalized variance functions (GVF) have been proposed to produce smoothed estimates of the ψ_i's. Secondly, the assumption $E(e_i|\theta_i) = 0$ may not be tenable if the small area sample size, n_i, is small and θ_i is a nonlinear function of the total Y_i, even if the director estimator $\hat{Y}_i$ is design-unbiased for Y_i. It is more realistic to use the sampling model $\hat{Y}_i = Y_i + f_i$ with $E(f_i|Y_i) = 0$, which simply says that $\hat{Y}_i$ is design-unbiased for Y_i. Further, we assume that

$V(f_i|Y_i) = \sigma_i^2$, where the sampling variance may depend on Y_i; for example, $\sigma_i^2 = Y_i^2 c_i^2$, where c_i is the known coefficient of variation of $\hat{Y}_i$ ascertained from fitting GVF's. The sampling model is now unmatched with the linking model in the sense that they cannot be combined directly to produce a linear mixed model. Various extensions of the basic area level (also called Fay-Herriot model) have been proposed to handle correlated sampling errors, spatial dependence of the model errors v_i and time-series and cross-sectional data (see Rao, 2003, Chapter 8).

In the second type, called basic unit level model, unit level auxiliary variables $\mathbf{x}_{ij} = (x_{1ij}, \ldots, x_{pij})^T$ are related to the unit y-values y_{ij} through a nested error linear regression model $y_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + v_i + e_{ij}$, where $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ and independent of $e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$. Various extensions of the basic unit-level model have been proposed to handle binary responses, two-stage sampling within areas, multivariate responses and others (see Rao, 2003, Chapters 8, 9 and 10). For example, for binary responses y_{ij}, we may assume that $y_{ij} \overset{\text{ind}}{\sim}$ Bernoulli (p_{ij}) and that the p_{ij}'s are linked by assuming a logistic regression model $\log\{p_{ij}/(1-p_{ij})\} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + v_i$, where $v_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$. This is a special case of generalized linear mixed models. Malec et al. (1999) used a two-level logistic regression model on the p_{ij}'s with random slopes $\boldsymbol{\beta}_i$.

3 Basic area level model: EB

3.1 Estimation of θ_i

Under the basic area level model, the best estimator of θ_i in the sense of minimum MSE is given by $E(\theta_i|\hat{\theta}_i, \boldsymbol{\beta}, \sigma_v^2)$ which depends on the model parameters $\boldsymbol{\beta}$ and σ_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 $\hat{\theta}_i \overset{\text{ind}}{\sim} N(\mathbf{z}_i^T\boldsymbol{\beta}, \sigma_v^2 + \psi_i)$, we obtain the empirical Bayes or empirical best (EB) estimator:

$$\hat{\theta}_i^{\text{EB}} = \hat{\gamma}_i\hat{\theta}_i + (1-\hat{\gamma}_i)\mathbf{z}_i^T\hat{\boldsymbol{\beta}}, \tag{3.1}$$

where $\hat{\gamma}_i = \hat{\sigma}_v^2/(\hat{\sigma}_v^2 + \psi_i)$. The form (3.1) shows that the EB estimator of θ_i 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. The weight $\hat{\gamma}_i$ is a measure of between area variability relative to total variability associated with area i. The estimator $\hat{\theta}_i^{\text{EB}}$ is unbiased for

θ_i under the combined model, i.e., $E(\hat{\theta}_i^{\rm EB} - \theta_i) = 0$, but the resulting estimator $g^{-1}(\hat{\theta}_i^{\rm EB})$ of Y_i is biased. Note that $g^{-1}(\hat{\theta}_i^{\rm EB})$ is not equal to the EB estimator $\hat{Y}_i^{\rm EB}$ obtained by evaluating $E[g^{-1}(\theta_i)|\hat{\theta}_i, \boldsymbol{\beta}, \sigma_v^2]$ at $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_v^2$.

The EB estimator $\hat{\theta}_i^{\rm EB}$ is design-consistent as the sampling variance ψ_i goes to zero, provided the direct estimator is design-consistent. This follows by noting that $\hat{\theta}_i^{\rm EB} \to \hat{\theta}_i$ as $\psi_i \to 0$.

Under normality assumption, maximum likelihood (ML) or residual maximum likelihood (REML) method may be used to estimate $\boldsymbol{\beta}$ and σ_v^2 from the marginal distribution $\hat{\theta}_i \overset{\rm ind}{\sim} N(\mathbf{z}_i^T\boldsymbol{\beta}, \sigma_v^2 + \psi_i)$. Alternatively, 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):

$$a(\sigma_v^2) = \sum_{i=1}^{m} (\hat{\theta}_i - \mathbf{z}_i^T \tilde{\boldsymbol{\beta}}(\sigma_v^2))^2/(\sigma_v^2 + \psi_i) = m - p, \tag{3.2}$$

where $\tilde{\boldsymbol{\beta}}(\sigma_v^2)$ is the weighted least squares estimator of $\boldsymbol{\beta}$ for given σ_v^2. The resulting estimators $\hat{\sigma}_v^2$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\sigma}_v^2)$ lead to the EBLUP estimator of θ_i from (3.1). The EBLUP estimator does not depend on normality.

3.2 MSE estimation

Methods of estimating $\mathrm{MSE}(\hat{\theta}_i^{\rm EB})$ that account for the variability of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_v^2$ have been studied extensively in the literature, where $\mathrm{MSE}(\hat{\theta}_i^{\rm EB}) = E(\hat{\theta}_i^{\rm EB} - \theta_i)^2$ and the expectation is with respect to the combined model (see Rao, 2003, Chapter 7). An accurate approximation to $\mathrm{MSE}(\hat{\theta}_i^{\rm EB})$ under normality is given by

$$\mathrm{MSE}(\hat{\theta}_i^{\rm EB}) \approx g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2) \tag{3.3}$$

where the leading term $g_{1i}(\sigma_v^2) = \gamma_i\psi_i$ with $\gamma_i = \sigma_v^2/(\sigma_v^2 + \psi_i)$ is the contribution to MSE assuming $\boldsymbol{\beta}$ and σ_v^2 are known,

$$g_{2i}(\sigma_v^2) = (1-\gamma_i)^2 \mathbf{z}_i^T \left[\sum_{i=1}^{m} \mathbf{z}_i\mathbf{z}_i^T/(\sigma_v^2 + \psi_i)\right]^{-1} \mathbf{z}_i \tag{3.4}$$

accounts for the variability of $\hat{\boldsymbol{\beta}}$ and the term

$$g_{3i}(\sigma_v^2) = \left[\psi_i^2/(\sigma_v^2 + \psi_i)^4\right] E(\hat{\theta}_i - \mathbf{z}_i^T\boldsymbol{\beta})^2 h(\sigma_v^2) \tag{3.5}$$

$$= \left[\psi_i^2/(\sigma_v^2 + \psi_i)^3\right] h(\sigma_v^2) \tag{3.6}$$

accounts for the variability of $\hat{\sigma}_v^2$, where $h(\sigma_v^2)$ is the asymptotic variance of $\hat{\sigma}_v^2$ for large m. Neglected terms in the approximation (3.3) are of lower order than m^{-1}, and the approximation is valid for the Prasad-Rao (PR), Fay-Herriot (FH), ML and REML methods of estimating σ_v^2. Datta, Rao and Smith (2002) showed that

$$h_{\mathrm{ML}}(\sigma_v^2) = h_{\mathrm{REML}}(\sigma_v^2) \leq h_{\mathrm{FH}}(\sigma_v^2) \leq h_{\mathrm{PR}}(\sigma_v^2). \tag{3.7}$$

It follows from (3.6) and (3.7) that ML (or REML) leads to the smallest MSE approximation followed by FH and PR.

Comparing the leading term $\gamma_i\psi_i$ of (3.3) with ψ_i, the MSE of the direct estimator $\hat{\theta}_i$, it is clear that the EB estimator $\hat{\theta}_i^{\mathrm{EB}}$ leads to large gain in efficiency when γ_i is small, i.e., when σ_v^2, the variability of the model errors v_i, is small relative to the total variability, $\sigma_v^2+\psi_i$. Note that ψ_i is the design variance of $\hat{\theta}_i$.

Turning to MSE estimation, an estimator correct to the same order approximation as (3.3) is given by

$$\mathrm{mse}(\hat{\theta}_i^{\mathrm{EB}}) \approx g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2). \tag{3.8}$$

The estimator (3.8) is approximately unbiased for $\mathrm{MSE}(\hat{\theta}_i^{\mathrm{EB}})$ in the sense that its bias is of lower order than m^{-1}, provided $\hat{\sigma}_v^2$ is based on REML or PR. For ML and FH methods of estimating σ_v^2, an extra term $g_{0i}(\hat{\sigma}_v^2)$ is added to (3.8). This extra term for ML is positive (Datta and Lahiri, 2000). Therefore, ignoring this term and using (3.8) 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, 2002). Therefore, ignoring this term and using (3.8) with FH estimator $\hat{\sigma}_v^2$ would lead to overestimation of MSE.

Lahiri and Rao (1995) showed that the MSE estimator (3.8) using the PR estimator of σ_v^2 is robust to nonnormality of the random effects in the sense that approximate unbiasedness remains valid, provided the normality of the sampling errors, e_i, holds. The latter assumption is less restrictive than the normality of the v_i's because of the central limit theorem effect on the direct estimators $\hat{\theta}_i$. It is not known if the robustness property is also valid under REML, ML and FH methods.

A criticism of the MSE estimator (3.8) and its modification for ML and FH is that it is not area-specific in the sense that it does not explicitly depend on $\hat{\theta}_i$ although the area-specific auxiliary data $\mathbf{z}_i$ is involved in the $g_{2i}(\hat{\sigma}_v^2)$–term. Rao (2000) used the expression (3.5) for $g_{3i}(\sigma_v^2)$ to get an alternative area-specific estimator of $g_{3i}(\sigma_v^2)$:

$$\tilde{g}_{3i}(\hat{\sigma}_v^2, \hat{\theta}_i) = [\psi_i^2/(\sigma_v^2+\psi_i)^4](\hat{\theta}_i - \mathbf{z}_i^T\hat{\boldsymbol{\beta}})^2 h(\hat{\sigma}_v^2). \tag{3.9}$$

Using (3.9), we get two different area-specific MSE estimators for REML or PR:

$$\text{mse}_1(\hat{\theta}_i^{\text{EB}}) \approx g_{1i}(\hat{\theta}_v^2) + g_{2i}(\hat{\sigma}_v^2) + g_{3i}(\hat{\sigma}_v^2) + \tilde{g}_{3i}(\hat{\sigma}_v^2, \hat{\theta}_i) \tag{3.10}$$

and

$$\text{mse}_2(\hat{\theta}_i^{\text{EB}}) \approx g_{1i}(\hat{\theta}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2\tilde{g}_{3i}(\hat{\sigma}_v^2, \hat{\theta}_i). \tag{3.11}$$

The term $\tilde{g}_{3i}(\hat{\sigma}_v^2, \hat{\theta}_i)$ is less stable than $g_{3i}(\hat{\sigma}_v^2)$ but it is of lower order than the leading term $g_{1i}(\hat{\sigma}_v^2)$ in (3.10) and (3.11). As a result, the coefficient of variation (CV) of $\text{mse}_1(\hat{\theta}_i^{\text{EB}})$ should be comparable to the CV of $\text{mse}(\hat{\theta}_i^{\text{EB}})$, at least for moderate to large m. Fuller (1989) estimated the conditional MSE of $\hat{\theta}_i^{\text{EB}}$ given the i-th area direct estimator $\hat{\theta}_i$. His area-specific MSE estimator is closely related to the unconditional MSE estimator (3.10). Butar and Lahiri (1997) obtained an area-specific MSE estimator by correcting the bias of Laird and Louis' (1987) estimator of variability based on the parametric bootstrap method. This bias-corrected MSE estimator is identical to (3.10) which is obtained in a straightforward manner from the formula (3.5) for $g_{3i}(\sigma_v^2)$.

Datta, Rao and Smith (2002) conducted a simulation study of the relative bias (i.e., bias/MSE) and CV of MSE estimators based on PR, ML, REML and FH. They used a simple linking model $\theta_i = \mu + v_i$ with $\mu = 0$ and three different distributions for v_i : $N(0, 1)$, double exponential $(0, 1)$ and location exponential $(-1, 1)$, each distribution with mean zero and variance 1. The sampling errors e_i were generated from $N(0, \psi_i)$ for specified ψ_i–patterns to reflect moderate to large variation in the ψ_i's. They generated 10,000 samples for each pattern of ψ_i's and $m = 15, 30$. For the pattern with moderate ψ–variation, all the MSE estimators are comparable in terms of relative bias, while FH outperformed for the pattern with large ψ_i–variation. In the latter case, the other methods lead to considerable over-estimation for the areas with small ψ_i. The FH method also performed well in terms of CV of the MSE estimator, particularly for the areas with small ψ_i when the ψ_i–variation is large. These empirical results strongly suggest that the FH-based MSE estimator is robust over ψ_i–patterns, while the FH-based estimator, $\hat{\theta}_i^{\text{EB}}$, maintains good efficiency.

It is more appealing to survey practitioners to consider the estimation of sampling MSE of $\hat{\theta}_i^{\text{EB}}$, i.e., $\text{MSE}_p(\hat{\theta}_i^{\text{EB}}) = E_p(\hat{\theta}_i^{\text{EB}} - \theta_i)^2$, where the expectation E_p is with respect to the sampling design $p(\cdot)$, i.e., the distribution of sampling errors given the θ_i's. Rivest and

Belmonte (2000) derived a design-unbiased estimator of $\mathrm{MSE}_p(\hat{\theta}_i^{\mathrm{EB}})$ using the PR estimator of σ_v^2. The leading term of this MSE estimator is area-specific, i.e., depends on $\hat{\theta}_i$, unlike the leading term $g_{1i}(\hat{\sigma}_v^2)$ of the model-based MSE estimator, $\mathrm{mse}(\hat{\theta}_i^{\mathrm{EB}})$. However, it is highly unstable relative to $\mathrm{mse}(\hat{\theta}_i^{\mathrm{EB}})$ unless more weight is attached to the direct estimator $\hat{\theta}_i$, i.e., $1-\hat{\gamma}_i$ is small.

3.3 Unknown sampling variances ψ_i

In sections 4.1 and 4.2 we assumed that the sampling variances, ψ_i, are known, but this is a restrictive assumption. Wang (2000) and Rivest and Vandal (2002) studied the effect of estimating ψ_i on the MSE of the EB estimator (3.1) with $\hat{\gamma}_i$ replaced by $\hat{\sigma}_v^2/(\hat{\sigma}_v^2+\hat{\psi}_i)$, where $\hat{\psi}_i$ is an estimator of ψ_i. For example, suppose that we have a random sample $y_{ij} \overset{\text{iid}}{\sim} N(\theta_i, \sigma^2)$, $j=1,\ldots,n_i(\geq 2)$ from the i-th area and $\hat{\theta}_i = \bar{y}_i$, the sample mean. In this case $\hat{\psi}_i = s_i^2/n_i$ is design-unbiased for ψ_i, where s_i^2 is the sample variance. Further, $\bar{y}_i$ and $\hat{\psi}_i$ are independently distributed with $\hat{\psi}_i \approx N[\psi_i, \delta_i = 2\psi_i^2/(n_i-1)]$. Under this set-up, Rivest and Vandal (2002) obtained an appropriate MSE estimator by adding the term $2\hat{\delta}_i\hat{\sigma}_v^4/(\hat{\psi}_i+\hat{\sigma}_v^2)^3$ to (3.8) to account for the estimation of ψ_i, where $\hat{\delta}_i = 2\hat{\psi}_i^2/(n_i-1)$. If the sample sizes n_i are small, then (3.8) can underestimate the MSE quite severely, unlike the Rivest-Vandal MSE estimator. If $\hat{\psi}_i$ is a smoothed estimator of ψ_i based on GVF model fitting, the contribution from the extra term is of the same order, $O(m^{-1})$, as the g_{3i}–term.

4 Jackkife estimation of MSE

Jiang, Lahiri and Wan (2002) proposed a jackknife method of estimating MSE of EB estimators that is applicable to generalized linear mixed models with block diagonal covariance structures, where the blocks correspond to small areas. This method also leads to approximately unbiased estimators of MSE of EB estimators. For example, consider the case of binary responses $y_{ij} \overset{\text{iid}}{\sim} \text{Bernoulli}(p_i)$, $j=1,\ldots,n_i$ and $\log\{p_i/(1-p_i)\} = \mathbf{z}_i^T\boldsymbol{\beta} + v_i$, $i=1,\ldots,m$, where $\mathbf{z}_i$ is the vector of area-specific covariates, $v_i \overset{\text{iid}}{\sim} N(0,\sigma_v^2)$ and p_i is the i-th area proportion. The minimum MSE (or Bayes) estimator of p_i is given by $\hat{p}_i^B = E(p_i|y_{i\cdot},\boldsymbol{\beta},\sigma_v^2) =: k(y_{i\cdot},\boldsymbol{\beta},\sigma_v^2)$, where $y_{i\cdot} = \Sigma_j y_{ij}$. The EB estimator of p_i is $\hat{p}_i^{\mathrm{EB}} = k(y_{i\cdot},\hat{\boldsymbol{\beta}},\hat{\sigma}_v^2)$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_v^2$ are suitable

estimators of $\boldsymbol{\beta}$ and σ_v^2 obtained from the marginal distribution of y_i's.

The jackknife method makes use of the following orthogonal decomposition of $\text{MSE}(\hat{p}_i^{\text{EB}})$:

$$\text{MSE}(\hat{p}_i^{\text{EB}}) = E(\hat{p}_i^{\text{B}} - p_i)^2 + E(\hat{p}_i^{\text{EB}} - \hat{p}_i^{\text{B}})^2. \tag{4.1}$$

Based on the decomposition (4.1), Jiang et al. (2002) proposed the following jackknife steps to estimate $\text{MSE}(\hat{\theta}_i^{\text{EB}})$:

(1) Calculate $\hat{\boldsymbol{\beta}}(\ell)$ and $\hat{\sigma}^2(\ell)$ deleting the ℓ-th area data $(y_{i.}, \mathbf{x}_i)$. Let $\hat{p}_i^{\text{EB}}(\ell) = k\{y_{i.}, \hat{\boldsymbol{\beta}}(\ell), \hat{\sigma}^2(\ell)\}$ be the EB estimator of p_i based on $\hat{\boldsymbol{\beta}}(\ell)$ and $\sigma^2(\ell)$. Note that $y_{i.}$ remains unchanged.

(2) Calculate the jackknife estimator of the last term in (4.1) as

$$\hat{M}_{2i} = \frac{m-1}{m} \sum_{\ell=1}^{m} \left[\hat{p}_i^{\text{EB}}(\ell) - \hat{p}_i^{\text{EB}}\right]^2. \tag{4.2}$$

(3) The first term $E\left(\hat{p}_i^{B} - p_i\right)^2$ may be written as $E[\tilde{g}_{1i}(y_{i.}, \boldsymbol{\beta}, \sigma_v^2] =: g_{1i}(\boldsymbol{\beta}, \sigma_v^2)$, where $\tilde{g}_{1i}(y_{i.}, \boldsymbol{\beta}, \sigma_v^2) = V(p_i | y_{i.}, \boldsymbol{\beta}, \sigma_v^2)$ is the posterior variance of p_i given $y_{i.}$ and $(\boldsymbol{\beta}, \sigma_v^2)$. Adjust the bias of $g_{1i}(\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$ (as an estimator of $g_{1i}(\boldsymbol{\beta}, \sigma_v^2)$) using the jackknife bias reduction method. The bias-adjusted estimator is given by

$$\begin{aligned} \hat{M}_{1i} &= g_{1i}(\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2) - \frac{m-1}{m} \\ &\quad \sum_{\ell=1}^{m} \left[g_{1i}(\hat{\boldsymbol{\beta}}(\ell), \hat{\sigma}_v^2(\ell)) - g_{1i}(\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)\right]. \end{aligned} \tag{4.3}$$

Note that the leading term $g_{1i}(\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$ is not area-specific in the sense of not depending on $y_{i.}$.

(4) Calculate the jackknife estimator of MSE as

$$\text{mse}_J(\hat{p}_i^{\text{EB}}) = \hat{M}_{1i} + \hat{M}_{2i}. \tag{4.4}$$

Booth and Hobert (1998) argued that for non-normal responses the MSE estimator should be area-specific because the posterior variance of p_i given $(\boldsymbol{\beta}, \sigma_v^2)$ depends on $y_{i.}$, unlike the case of linear mixed models. They proposed the conditional MSE, given the i-th area data $(y_i, \mathbf{z}_i)$, as the relevant measure of variability, and estimated the

conditional MSE. Rao (2003, Chapter 9) addressed this criticism by simply modifying the bias-adjusted estimator $\hat{M}_{1i}$. Instead of evaluating the expectation of $\tilde{g}_{1i}(y_{i.}, \boldsymbol{\beta}, \sigma_v^2)$ with respect to the marginal distribution of $y_{i.}$ (using numerical integration), he proposed to adjust the bias of $\tilde{g}_{1i}(y_{i.}, \hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$ as an estimator of $g_{1i}(y_{i.}, \boldsymbol{\beta}, \sigma_v^2)$. This leads to

$$\begin{aligned} \tilde{M}_{1i}(y_{i.}) &= \tilde{g}_{1i}(y_{i.}, \hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2) - \frac{m-1}{m} \\ &\quad \sum_{\ell=1}^{m} \left[\tilde{g}_{1i}(y_{i.}, \hat{\boldsymbol{\beta}}(\ell), \hat{\sigma}_v^2(\ell)) - \tilde{g}_{1i}(y_{i.}, \hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2) \right] \end{aligned} \tag{4.5}$$

which is area-specific including the leading term $\tilde{g}_{1i}(y_{i.}, \hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$. The modified jackknife estimator of MSE is given by

$$\text{mse}_J^*(\hat{p}_i^{\text{EB}}) = \tilde{M}_{1i}(y_{i.}) + \hat{M}_{2i}. \tag{4.6}$$

Note that (4.6) is not only area-specific but also computationally simpler than (4.4) because it avoids the evaluation of the expectation of $\tilde{g}_{1i}(y_{i.}, \boldsymbol{\beta}, \sigma_v^2)$ with respect to the marginal distribution of $y_{i.}$. Properties of the modified jackknife MSE estimator (4.6) are under investigation (jointly with Sharon Lohr). For the case of a linear mixed model, $\tilde{M}_{1i}(y_{i.}) = \hat{M}_{1i}$ and hence (4.6) is identical to (4.4).

5 Basic unit level model: PSEUDO-EB

We now turn to the basic unit level model, $y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}$, and assume that the model holds for the sample, i.e., no sample selection bias within areas. If the areas are also sampled, then we assume the absence of selection bias for sampled areas as well. The mean for the i-th area, $\bar{Y}_i$, may be approximated as $\mu_i = \bar{\mathbf{X}}_i^T \boldsymbol{\beta} + v_i$, assuming that the number of population units in the i-th area, N_i, is large, where $\bar{\mathbf{X}}_{i.}$ is the population mean of $\mathbf{x}$ for the i-th area. Assuming $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ and independent of $e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and estimating the model parameters $\boldsymbol{\beta}$ and σ_v^2 from the marginal distribution of the sampled y_{ij}'s, we get the EB estimator of μ_i as

$$\hat{\mu}_i^{\text{EB}} = \hat{\gamma}_i \left[\bar{y}_i + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_i)^T \hat{\boldsymbol{\beta}} \right] + (1 - \hat{\gamma}_i) \bar{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}}, \tag{5.1}$$

where $\hat{\gamma}_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \hat{\sigma}_e^2 / n_i)$, $(\bar{y}_i, \bar{x}_i)$ are the i-th area sample means and $(\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2)$ are the estimators of $(\boldsymbol{\beta}, \sigma_v^2)$; see Battese, Harter and Fuller (1988). This estimator is also the EBLUP estimator without

normality assumption, provided $\boldsymbol{\beta}$ and σ_v^2 are estimated by a moments method such as the method of fitting-of-constants. As $n_i \to \infty$, the EB estimator converges to the "survey regression" estimator $\bar{y}_i + (\bar{\mathbf{X}}_i - \bar{\mathbf{x}}_i)^T \hat{\boldsymbol{\beta}}$, while it converges to the "synthetic regression" estimator $\bar{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}}$ as n_i decreases. We refer to Rao (2003), Chapter 7 for a detailed discussion of EBLUP estimation and MSE estimation.

A drawback of (5.1) is that it is purely model-based and does not take account of the survey weights w_{ij}. As a result, it is not design-consistent as n_i increases, unless the sampling design is self-weighting within areas, i.e., $w_{ij} = w_i$. On the other hand, the EBLUP estimator under the area level model is design-consistent. It is desirable to ensure design-consistency for the unit level models as well because n_i could be moderately large for some of the areas under consideration. Also, it is desirable to ensure that the estimators of area totals automatically add up to the direct survey regression estimator of the large area total. You and Rao (2002a) developed a pseudo-EBLUP estimator of μ_i that satisfies both the desirable properties. We give a brief account of their approach.

We first obtain a survey-weighted area level model from the unit level model by taking a weighted average using normalized weights $\tilde{w}_{ij} = w_{ij}/w_{i.}$, where $w_{i.} = \Sigma_j w_{ij}$:

$$\bar{y}_{iw} = \Sigma_j \tilde{w}_{ij} y_{ij} = \bar{\mathbf{x}}_{iw}^T \boldsymbol{\beta} + v_i + \bar{e}_{iw}, \tag{5.2}$$

where $\bar{e}_{iw} = \Sigma_j \tilde{w}_{ij} e_{ij}$ with $E(\bar{e}_{iw}) = 0$ and $V(\bar{e}_{iw}) = \sigma_v^2 \Sigma_j \tilde{w}_{ij}^2 =: \sigma_e^2 \delta_{iw}$, and $\bar{\mathbf{x}}_{iw} = \Sigma_j \tilde{w}_{ij} \mathbf{x}_{ij}$. Then the BLUP estimator of μ_i from the aggregated model (5.2) is obtained as

$$\begin{aligned} \tilde{\mu}_{iw}(\boldsymbol{\beta}, \sigma_v^2, \sigma_v^2) &= \bar{\mathbf{X}}_i^T \boldsymbol{\beta} + \gamma_{iw}(\bar{y}_{iw} - \bar{\mathbf{x}}_{iw}^T \boldsymbol{\beta}), & (5.3) \\ &= \bar{\mathbf{X}}_i^T \boldsymbol{\beta} + \tilde{v}_{iw}(\boldsymbol{\beta}, \sigma_e^2, \sigma_v^2) & (5.4) \end{aligned}$$

where $\gamma_{iw} = \sigma_v^2/(\sigma_v^2 + \sigma_e^2 \delta_{iw})$ and $\tilde{v}_{iw}(\boldsymbol{\beta}, \sigma_e^2, \sigma_v^2) = \gamma_{iw}(\bar{y}_{iw} - \bar{\mathbf{x}}_{iw}^T \boldsymbol{\beta})$ is the BLUP of v_i from the aggregated model (5.2). This estimator depends on the model parameters $\boldsymbol{\beta}$ and σ_v^2.

We estimate $\boldsymbol{\beta}$, given σ_e^2 and σ_v^2, using the following weighted estimating equations based on the unit level deviations $y_{ij} - x_{ij}^T \boldsymbol{\beta} - \tilde{v}_{iw}(\boldsymbol{\beta}, \sigma_e^2, \sigma_v^2)$:

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} w_{ij} \mathbf{x}_{ij} \left[y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta} - \tilde{v}_{iw}(\boldsymbol{\beta}, \sigma_e^2, \sigma_v^2) \right] = \mathbf{0}. \tag{5.5}$$

Denote the solution of (5.5) as $\tilde{\boldsymbol{\beta}}_w(\sigma_e^2, \sigma_v^2)$. Now we replace σ_v^2 and σ_e^2 by suitable estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ from the unit level model to get

$\hat{\boldsymbol{\beta}}_w = \tilde{\boldsymbol{\beta}}_w(\hat{\sigma}_e^2, \sigma_v^2)$. This leads to a pseudo-EBLUP estimator of μ_i as

$$\hat{\mu}_{iw}^P = \tilde{\mu}_{iw}(\hat{\boldsymbol{\beta}}_w, \hat{\sigma}_e^2, \hat{\sigma}_v^2) = \bar{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}}_w + \hat{\gamma}_{iw}(\bar{y}_{iw} - \bar{\mathbf{x}}_{iw}^T \hat{\boldsymbol{\beta}}_w), \tag{5.6}$$

where $\hat{\gamma}_{iw} = \hat{\sigma}_v^2/(\hat{\sigma}_v^2 + \hat{\sigma}_e^2 \delta_{iw})$. It is also possible to use design-weighted estimators of σ_e^2 and σ_v^2 (Korn and Granbard, 2003). You and Rao (2002) showed that the estimated area totals, $N_i \hat{\mu}_{iw}^P$, add up to the direct survey regression estimator of the total $\Sigma_i N_i \mu_i$:

$$\Sigma_i N_i \hat{\mu}_{iw}^P = \hat{Y}_w + (\mathbf{X} - \hat{\mathbf{X}}_w)^T \hat{\boldsymbol{\beta}}_w, \tag{5.7}$$

where $\hat{Y}_w = \Sigma_i \Sigma_j w_{ij} y_{ij}$, $\hat{\mathbf{X}}_w = \Sigma_i \Sigma_j w_{ij} \mathbf{x}_{ij}$. Thus, the pseudo-EBLUP estimators $\hat{\mu}_{iw}^P$ satisfy the benchmarking property automatically without any post-adjustment, unlike in the case of EBLUP estimators $\hat{\mu}_i^{\mathrm{EB}}$. In the latter case, Battese et al. (1988) proposed an efficient post-adjustment of the EBLUP estimators $\hat{\mu}_i^{\mathrm{EB}}$ to ensure agreement with the direct estimator of the total $\Sigma_i N_i \mu_i$.

We have assumed that the random small area effects v_i are normally distributed in the basic unit level model $y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}$. The MSE estimator of the purely model-based estimator $\hat{\mu}_i^{\mathrm{EB}}$ derived under normality of the v_i's is not robust to deviations from normality, unlike the case of basic area level model studied in Section 3.2. It would be worthwhile to study EB inference under semi-nonparametric (SNP) representations of the density of v_i. Zhang and Davidian (2001) approximated the density of v_i by a SNP representation which includes normality as a special case, and it provides flexibility in capturing nonnormality through a user-chosen tuning parameter. Maiti (2001) used a finite mixture of normal distributions for the distribution of v_i, and developed hierarchical Bayes (HB) estimates of small area means, assuming a prior distribution on the model parameters. EB estimation of small area means and associated MSE estimation under broad classes of densities of v_i, such as the above, would be practically useful.

6 Hierarchical Bayes (HB) approach

We illustrate the HB approach using the basic area level model of Section 2. Under this approach, a prior distribution on the model parameters $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \sigma_v^2)^T$ is specified, and inferences are then based on the posterior distribution, $f(\theta_i|\hat{\boldsymbol{\theta}})$, of θ_i given the data $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^T$. In particular, θ_i is estimated by its posterior mean $E(\theta_i|\hat{\boldsymbol{\theta}})$, called the HB estimator $\hat{\theta}_i^{\mathrm{HB}}$. The variability of $\hat{\theta}_i^{\mathrm{HB}}$ is measured by the

posterior variance $V(\theta_i|\hat{\boldsymbol{\theta}})$. The calculation of $\hat{\theta}_i^{\mathrm{HB}}$ and $V(\theta_i|\hat{\boldsymbol{\theta}})$ involves integrations with respect to the posterior distribution of $\boldsymbol{\beta}$, σ_v^2, $f(\boldsymbol{\beta}, \sigma_v^2|\hat{\boldsymbol{\theta}})$. However, Monte Carlo Markov chain (MCMC) methods can be used to generate J simulated samples $\{\theta_1^{(j)}, \ldots, \theta_m^{(j)};\ j = 1, \ldots, J\}$ directly from the joint posterior $f(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, where J is sufficiently large and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$. Using the simulated samples, we approximate $\hat{\theta}_i^{\mathrm{HB}}$ by the mean $\theta_i^{(\cdot)} = J^{-1}\Sigma_j \theta_i^{(j)}$ and $V(\theta_i|\hat{\boldsymbol{\theta}})$ by the variance $J^{-1}\Sigma_j(\theta_i^{(j)} - \theta_i^{(\cdot)})^2$ of the simulated samples $\theta_i^{(j)}$. The HB estimator of the total Y_i is approximated by the mean $Y_i^{(\cdot)} = J^{-1}\Sigma_j Y_i^{(j)}$ and the posterior variance of Y_i, $V(Y_i|\hat{\boldsymbol{\theta}})$ by the variance $J^{-1}\Sigma_j(Y_i^{(j)} - Y_i^{(\cdot)})^2$, where $Y_i^{(j)} = g^{-1}(\theta_i^{(j)})$.

An advantage of the HB approach is that it is straightforward, the inferences are "exact" unlike the EB (or EBLUP) approach, and it can handle complex small area models using MCMC methods, but it requires the specification of a prior $f(\boldsymbol{\beta}, \sigma_v^2)$ on the model parameters. It would be desirable to select a "matching" prior $f(\boldsymbol{\beta}, \sigma_v^2) \propto f(\sigma_v^2)$ that leads to well-calibrated inferences. In particular, the posterior variance should be approximately unbiased for $\mathrm{MSE}(\hat{\theta}_i^{\mathrm{HB}})$, i.e., $E[V(\theta_i|\hat{\boldsymbol{\theta}})] - \mathrm{MSE}(\hat{\theta}_i^{\mathrm{HB}}) = o(m^{-1})$; asymptotically, $\hat{\theta}_i^{\mathrm{HB}} \approx \hat{\theta}_i^{\mathrm{EB}}$. This will provide a frequentist justification for the posterior variance as a measure of variability Data, Rao and Smith (2002) showed that the matching prior is given by

$$f_i(\sigma_v^2) \propto (\sigma_v^2 + \psi_i)^2 \sum_{\ell=1}^{m} (\sigma_v^2 + \psi_\ell)^{-2}. \tag{6.1}$$

This prior depends collectively on the sampling variances, ψ_ℓ, for all the areas ℓ as well as on the area-specific sampling variance ψ_i. For the balances case, $\psi_i = \psi$, the matching prior reduces to the "flat" prior $f(\sigma_v^2) \propto 1$. Note that the prior (6.1) on the common parameter σ_v^2 is designed for inference on the i-th area so that its dependence on ψ_i may not be problematic.

A disadvantage of the EB estimator $\hat{\theta}_i^{\mathrm{EB}}$ is that the weight $\hat{\gamma}_i$ attached to the direct estimator takes zero value when $\hat{\sigma}_v^2 = 0$, in which case it reduces to the regression synthetic estimators $\mathbf{z}_i^T\hat{\boldsymbol{\beta}}$. Thus, all the direct estimator $\hat{\theta}_i$ receive zero weight even when the sample sizes in some areas are not small. This difficulty was encountered in using a state model to produce EB state estimates of poor school-age children in the United States (National Research Council, 2000). The HB approach avoids this difficulty by producing positive weights in all cases. Bell (1999) applied the HB approach to the state model

using the prior $f(\boldsymbol{\beta}, \sigma_v^2) = f(\boldsymbol{\beta}) f(\sigma_v^2)$ with $f(\boldsymbol{\beta}) \propto 1$ and $f(\sigma_v^2) \propto 1$, and obtained positive weights in all cases. But it is not clear if his method leads to well-calibrated inferences since the matching prior (6.1) is different from the flat prior, especially when the ψ_i-values vary significantly, as in the case of Bell (1999) with $\max(\psi_i)/\min(\psi_i)$ as large as 20.

You and Rao (2002b) used the HB approach to handle the case of unmatched sampling and linking area level models (Section 2) and applied it to Canadian census undercount estimation. In this application, $C_i =$ census count, $Y_i =$ number missing and $\hat{Y}_i$ is a post-census survey estimator of Y_i with known sampling variance σ_i^2 for the i-th province in Canada ($i = 1, \ldots, m = 10$). The σ_i^2's were estimated by fitting a GVF model of the form $V(\hat{Y}_i) \propto C_i^{\gamma}$ and then treating as if known in the sampling model $\hat{Y}_i | Y_i \overset{\text{ind}}{\sim} N(Y_i, \sigma_i^2)$. The linking model is given by $\theta_i = \log\{Y_i/(Y_i + C_i)\} = \beta_0 + \beta_1 \log C_i + v_i$ with $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$. HB estimates of undercounts, Y_i, and undercoverage rates, $U_i = Y_i/(Y_i + C_i)$, and associated coefficients of variation (based on the posterior variance) were calculated, using MCMC methods.

Singh, Folsom and Vaish (2003) studied the basic unit level model, $y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}$ with $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ independent of $e_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$, for the population, and allowed sample selection bias within small areas. They used methods based on survey-weighted estimating functions (EF) to account for the sample selection bias. They also extended the method to generalized linear mixed model such as the logistic mixed model $y_{ij} | p_{ij} \overset{\text{ind}}{\sim}$ Bernoulli(p_{ij}) and $\text{logit}(p_{ij}) = \log\{p_{ij}/(1 - p_{ij})\} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i$; see Rao (2003, pp. 253–254) for a brief account of the EF method. In situations where a sample of areas is selected, the random effects v_i are assumed to be free of sample selection bias.

The HB approach is powerful and attractive, but caution should be exercised when using MCMC methods. For example, MCMC algorithms could lead to seemingly reasonable inferences about a non-existent posterior distribution. This happens when the posterior distribution is improper and yet all the Gibbs conditional distributions, used in generating the MCMC simulated samples, are proper (Hobert and Casella, 1996). Another difficulty with MCMC is that the convergence diagnostic tools can fail to detect the sort of convergence failures they were designed to identify (Cowles and Carlin, 1996). We refer the reader to Rao (2003), Section 10.2.4 for a discussion of practical issues associated with MCMC.

MCMC methods are also extensively used for model determination which plays a vital role in developing model-based small area estimates. In particular, methods based on Bayes factors, posterior predictive densities and cross-validation predictive densities are employed for model determination; see Rao (2003), Section 10.2.6. The criterion of posterior predictive probability is often used to check the overall fit of a proposed model. Sinharay and Stern (2003) conducted a simulation study to investigate the effectiveness of this criterion for model checking, using the basic area level model with no covariates. Their study indicates that it is difficult to detect nonnormality of the random effects v_i using this criterion, unless the extent of violation is huge.

7 Some recent applications

We now give a brief account of some major applications of EB(EBLUP) and HB methods for model-based small area estimation. We refer the reader to Rao (2003), Chapters 7–10 for further details.

(i) Basic area level model

The basic area level model, $\hat{\theta}_i = \theta_i + e_i$ and $\theta_i = \mathbf{z}_i^T\boldsymbol{\beta} + v_i$ with $\theta_i = \log Y_i$, has been used recently to produce EBLUP estimates of poor school-age children, Y_i, for each county in the United States (National Research Council, 2000). Using these counts, the U.S. Department of Commerce allocates annually over $7 billion of funds to counties, and then states distribute these funds among school districts. Data from the Current Population Survey (CPS) are used to calculate the direct estimates $\hat{\theta}_i$, and the area level covariates are based on census and administrative data.

(ii) Basic unit level model

Battese, Harter and Fuller (1988) used the basic unit level model, $y_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + v_i + e_{ij}$, to produce EBLUP estimates of county crop areas in Iowa using LANDSAT satellite data as the unit level covariates $\mathbf{x}_{ij}$. They also calculated standard errors of EBLUP estimates and validated the normality assumption on the error terms v_i and e_{ij}.

(iii) Time series and cross-sectional area level models

Datta, Lahiri and Maiti (2002) applied time series and cross-sectional area level models to obtain EBLUP estimates of median income of four-person families for the fifty American states and the District of Columbia. They used the CPS estimates for nine years (1981–89) as direct estimates, and census data as covariates. They also calculated the coefficient of variation of the EBLUP estimates and demonstrated their superiority over the direct CPS estimates. They also conducted an external evaluation by comparing the 1989 EBLUP estimates to 1990 census estimates for 1989. Datta, Lahiri, Maiti and Lu (1999) applied time series and cross-sectional area level models to obtain HB estimates of monthly unemployment rates for forty-nine U.S. states and the district of Columbia ($m = 30$), and associated standard errors based on the posterior variance. They used the CPS estimates of unemployment rates as direct estimated $\hat{\theta}_{it}$, and Unemployment Insurance (UI) claims rate as covariates z_{it} for the period January 1985–December 1988 ($t = 1, \ldots, T = 48$) to calculate HB estimates, $\hat{\theta}_{iT}^{\mathrm{HB}}$, for the current period T. The linking model used by Datta et al. (1999) accounted for seasonal variation in monthly unemployment rates.

You, Rao and Gambino (2003) applied time series and cross-sectional area level models to data from the Canadian Labour Force Survey to obtain HB estimates and associated standard errors of monthly unemployment rates for Census Metropolitan Areas and Census Agglomerations. Unlike Datta et al. (1999), they used short time series data ($T = 6$ months) and employed simpler models without seasonal parameters.

(iv) Disease mapping area level models

Area level models have also been used in mapping of small area mortality (or incidence) rates of diseases such as cancer. A simple model assumes that the observed small area disease counts $y_i|\lambda_i \overset{\text{ind}}{\sim}$ Poisson$(n_i\lambda_i)$ and $\theta_i = \log \lambda_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, where θ_i is the true incidence rate and n_i is the number exposed in area i (see e.g., Maiti, 1998). Spatial dependence models for θ_i's have also been studied, using conditional autoregression (CAR) that relates each θ_i to a set of neighbourhood areas of area i. Modelling of age-group specific counts, y_{ij}, have also been studied. For example, Nandram, Sedransk and Pickle (1999) assumed that $y_{ij} \overset{\text{ind}}{\sim}$ Poisson$(n_{ij}\lambda_{ij})$ and

$\log \lambda_{ij} = \mathbf{x}_j^T \boldsymbol{\beta} + v_i$ with $v_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, where n_{ij} is the number exposed in age group j and area i, θ_{ij} is the area/age-specific mortality and $\mathbf{x}_j$ is a vector of covariates for age group j. Random slopes, $\boldsymbol{\beta}_i$, instead of fixed $\boldsymbol{\beta}$, for the linking model have also been considered.

Maiti (1998) used the simple model above to obtain HB estimates of lip-cancer incidence rates for Scottish counties. He also studied HB estimation for the lip-cancer data under the spatial dependence model for the β_i's. Estimates of θ_i's were very similar for both the models but standard errors for the spatial linking model were smaller than those under the simpler model of spatial independence. Nandram et al. (1999) used the age-group specific models to obtain age-specific HB estimates of mortality rates for Health Service Areas in the United States for the disease category "all cancers for white males".

(v) Logistic linear mixed models

Malec et al. (1997) studied HB estimation of small area proportions associated with binary responses, using logistic linear mixed models with random slopes. They obtained HB estimates of health-related proportions, from the U.S. National Health Interview Survey, for the 50 states and the District of Columbia and for specified subpopulations within the 51 areas. Malec, Davis and Cao (1999) studied similar models to estimate overweight prevalence for subgroups (small areas) using the U.S. National Health and Nutrition Examination Survey data. Again, HB methods were used but survey weights were incorporated using a pseudo-likelihood. Folsom, Shah and Vaish (1999) studied generalized linear mixed models with age-specific correlated random effects. They produced estimates of prevalence rates for U.S. states and age groups for up to 20 drug use related binary outcomes, using data from the pooled U.S. National Household Survey on Drug Abuse. The population model was assumed to hold for the sample (i.e., the absence of sample selection bias), but survey weights were introduced to obtain pseudo-HB estimates and pseudo-HB standard errors.

8 Some Practical issues

In this section, we provide brief remarks on practical issues related to small area estimation.

(i) Design issues

It is important to consider design issues that have an impact on small area estimation. A proper resolution of design issues could lead to enhancement in the reliability of direct as well as indirect estimates for both planned and unplanned domains (areas). The following measures at the design stage might be useful in minimizing the need for indirect estimators, at least for some planned domains: Use of list frames to replace clusters wherever possible, use of many small strata from which samples are drawn, compromise sample allocations to satisfy reliability requirements at a small area level as well as large area level, integration of surveys through harmonizing questions across surveys of the same population, use of multiple frame surveys, use of "rolling samples" as a method of cumulating data over time. We refer the reader to Rao (2003), Chapter 2, Singh et al. (1994) and Marker (2001) for further details.

Despite the above preventive measures at the design stage, indirect estimates will be needed in practice because it is not possible to anticipate and plan for all possible areas (or domains) and uses of survey data: "the client will always require more than is specified at the design stage" (Fuller, 1999, p. 344).

(ii) Model selection and validation

Methodological developments and applications of model-based estimation are impressive, but caution should be exercised because of the model assumptions.

Good auxiliary information related to the variables of interest plays a vital role in determining suitable linking models. Therefore, more attention should be given to the compilation of auxiliary variables that are good predictors of study variables.

Subject matter specialists or end users should have influence on the choice of models, particularly on the choice of auxiliary variables. However, model diagnostics should be used to find suitable model(s) that fit the data well. Such model diagnostics include residual analysis to detect departures from assumed models, selection of auxiliary variables and case-deletion diagnostics to detect influential observations. We refer the reader to Rao (2003), Chapter 6 for some methods of model validation in the frequentist framework.

Hierarchical Bayes (HB) approach has become very popular in recent years due to its ability to handle complex models using MCMC

methods. However, caution should be exercised in the choice of improper priors on model parameters, as noted in Section 6. Limitations of MCMC methods, such as shortcomings of available convergence diagnostics, should also be noted. Carlin and Louis (2000) made an important observation on the dangers of "plug and play" implementation via MCMC: "Worse, the sheer power of MCMC methods has led to the temptation to fit models larger than the data can readily support without a strongly informative prior structure – now something of a rarity in applied Bayesian work".

HB methods for model validation via MCMC have been extensively developed, but the effectiveness of some criteria for model checking is questionable as noted in Section 6. Further work on effective methods for model checking is needed.

(iii) Area level vs. unit level models

Area level models have wider scope than unit level models because area level auxiliary information is more readily available than unit level auxiliary data. Also, design weights are incorporated by modelling design-weighted direct estimators, and the resulting EB or HB estimators are design consistent. But the assumption of known sampling variances, ψ_i, is quite restrictive. Smoothed estimates of ψ_i's based on GVF model fitting can also cause difficulties in MSE estimation, as noted in Section 3.3. We need more work on obtaining good approximations to the sampling variances as well as methods that incorporate the variability associated with estimated sampling variances in MSE estimation. This task becomes more difficult when using multivariate or time series area level models because sampling covariances are also needed.

Recent work on incorporating survey weights into model-based estimation, via pseudo-EB or pseudo-HB, is promising; in particular, the self-benchmarking property noted in Section 5. But the assumption that the sample selection bias is absent may not be true for some applications. The estimating functions (EF) approach of Singh, Folsom and Vaish (2003), mentioned in Section 6, allows sample selection bias within sampled areas but it assumes that the random effects v_i are free from sample selection bias in situations where a sample of areas is selected. Methods for handling the latter case are needed. Moreover, their method assumes known sampling variances, as in the area level model, and this assumption may be restrictive (see Rao, 2003, Section 10.5.4).

(iv) "Triple-goal" estimation

We focussed on model-based estimation of small area totals or means, but such estimates may not be suitable if the main objective is to produce an ensemble of parameter estimates whose distribution is in some sense close enough to the distribution of area-specific parameters. For example, we may be interested in ranking areas or identifying areas that fall below or above some prespecified level. Shen and Louis (1998) proposed "triple-goal" estimators that can produce good ranks, a good histogram and good area-specific estimators, assuming simple linking models. It would be useful to extend their methods to handle more complex models that are suitable for small area estimation.

(v) Nonsampling errors

We have assumed the absence of measurement errors in the responses and/or the covariates as well as nonresponse. But nonsampling errors can have a substantial effect on small area estimation, and it would be useful to develop suitable designs as well as methods of estimation that can account for nonsampling errors. Nandram and Choi (2002) used HB nonresponse models for binary data, and applied the theory to data from the U.S. National Crime Survey to estimate small area proportions. Measurement errors in the responses, even under an additivity assumption, can lead to biased estimators of quantiles and histograms. In the context of direct estimation, Fuller (1995) proposed methods at the design stage that can lead to bias-adjusted estimators of quantiles and histograms.

(vi) How to handle when explicit data pooling is prohibited

Indirect estimators, studied in previous sections, borrow strength by explicitly pooling data across small areas. Reiter (2000) studied cases where external constraints prohibit explicit data pooling. He proposed methods that may be acceptable under such external constraints and yet yield estimators that are more accurate than the area-specific direct estimators. In particular, he proposed to use the pooled data parameter estimates to specify the model in each area, but estimate the model parameters separately for each area. The proposed methods look interesting and further work would be useful.

Acknowledgment

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988), An error–components model for prediction of crop areas using survey and satellite data. Journal of the American Statistical Association, **83**, 28–36.

Bell, W. R. (1999), Accounting for uncertainty about variances in small area estimation. Bulletin of the International Statistical Institute.

Booth, J. G. and Hobert, J. P. (1998), Standard errors of predictors in generalized linear mixed models. Journal of the American Statistical Association, **93**, 262–272.

Butar, F. B. and Lahiri, P. (2001), On measures of uncertainty of empirical Bayes small-area estimators. Tech. Rep., Department of Statistics, University of Nebraska-Lincoln.

Carlin, B. P. and Louis, T. A. (2000), Empirical Bayes: past, present and future. Journal of the American Statistical Association, **95**, 1286–1289.

Cowles, M. R. and Carlin, B. P. (1996), Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association, **91**, 883–904.

Datta, G. S., Lahiri, P., Maiti, T. and Lu, R. L. (1999), Hierarchical Bayes estimation of unemployment rates for the U.S. states. Journal of the American Statistical Association, **94**, 1074–1082.

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

Datta, G. S., Lahiri, P. and Maiti, T. (2002), Empirical Bayes estimation of median income of four-person families by states using time series and cross-sectional data. Journal of Statistical Planning and Inference, **102**, 83–97.

Datta, G. S., Rao, J. N. K. and Smith, D. D. (2002), On measures of uncertainty of small area estimators in the Fay-Herriot model. Tech. Rep., University of Georgia, Athens.

Fay, R. E. and Herriot, R. A. (1979), Estimation of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, **74**, 269–277.

Folsom, Jr., R., Shah, B. V. and Vaish, A. K. (1999), Substance abuse in states: a methodological report on model-based estimates from the 1994–1996 National Household Surveys on Drug Abuse. Proceedings of the Section on Survey Research Methods, American Statistical Association, 371–375.

Fuller, W. A. (1989), Prediction of true values for the measurement error model. In Conference on Statistical Analysis of Measurement Error Models and Applications, Humboldt State University.

Fuller, W. A. (1995), Estimation in the presence of measurement error. International Statistical Review, **63**, 121–147.

Fuller, W. A. (1999), Environmental surveys over time. Journal of Agriculatural, Biological and Environmental Statistics, **4**, 331–345.

Ghosh, M. and Rao, J. N. K. (1994), Small area estimation: an appraisal (with discussion). Statistical Science, **9**, 65–93.

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**, 1461–1473.

Jiang, J., Lahiri, P. and Wan, S. -M. (2002), A unified jackknife theory for empirical best prediction with M-estimation. Annals of Statistics, **30**, 1782–1810.

Korn, E. L. and Graubard, B. I. (2003), Estimating variance components by using survey data. Journal of Royal Statistical Society, Series B, **65**, 175–190.

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**, 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**, 739–750.

Maiti, T. (1998), Hierarchical Bayes estimation of mortality rates for disease mapping. Journal of Statistical Planning and Inference, **69**, 339–348.

Maiti, T. (2001), Robust generalized linear mixed models for small area estimation. Journal of Statistical Planning and Inference, **98**, 225–238.

Malec, D., Davis, W. W. and Cao, X. (1999), Model-based small area estimates of overweight prevalence using sample selection adjustment. Statistics in Medi-cine, **18**, 3189–3200.

Malec, D., Sedransk, J., Moriarity, C. L. and LeClerc, F. B. (1997), Small area inference for binary variables in National Health Interview Survey. Journal of the American Statistical Association, **92**, 815–826.

Marker, D. A. (2001), Producing small area estimates from national surveys: methods for minimizing use of indirect estimators. Survey Methodology, **27**, 183–188.

Nandram, B., Sedransk, J. and Pickle, L. (1999), Bayesian analysis of mortality rates for U.S. Health Service Areas. Sankhyā, Series B, **61**, 145–165.

Nandram, B. and Choi, J. W. (2002), Hierarchical Bayesian nonresponse models for binary data from small areas with uncertainty about ignorability. Journal of the American Statistical Association, **97**, 381–388.

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.

Pfeffermann, D. (2002), Small area estimation – new developments and directions. International Statistical Review, **70**, 125–143.

Prasad, N. G. N. and Rao, J. N. K. (1990), The estimation of the mean squared error of small-area estimators. Journal of the American Statistical Association, **85**, 163–171.

Rao, J. N. K. (1999), Some recent advances in model-based small area estimation. Survey Methodology, **25**, 175–186.

Rao, J. N. K. (2001), EB and EBLUP in small area estimation. in S. E. Ahmed and N. Reid (Eds.); Empirical Bayes and Likelihood Inference, Lecture Notes in Statistics 148, New York: Springer, 33–43.

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

Rivest, L-P. and Belmonte, E. (2000), A conditional mean squared error of small area estimators. Survey Methodology, **26**, 67–78.

Rivest, L-P. and Vandal, N. (2003), Mean squared error estimation for small areas when the small area variances are estimated. In Proceedings of the International Conference on Recent Advances in Survey Sampling, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada (in press).

Reiter, J. P. (2000), Borrowing strength when explicit data pooling is prohibited. Journal of Official Statistics, **16**, 295–319.

Singh, M. P., Gambino, J., Mantel, H. J. (1994), Issues and strategies for small area data. Survey Methodology, **20**, 3-22.

Singh, A. C., Folsom, Jr., R. E. and Vaish, A. K. (2003), Estimating function based approach to hierarchical Bayes small area estimation for survey data. In Proceedings of the International Conference on Recent Advances in Survey Sampling, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada (in press).

Shen, W. and Louis, T. A. (1998), Triple-goal estimates in two-stage hierarchical models. Journal of the Royal Statistical Society, Series B, **60**, 455–471.

Sinharay, S. and Stern, H. S. (2003), Posterior predictive model checking in hierarchical models. Journal of Statistical Planning and Inference, **111**, 209–221.

You, Y. and Rao, J. N. K. (2002a), A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. Canadian Journal of Statistics, **30**, 431–439.

You, Y. and Rao, J. N. K. (2002b), Small area estimation using unmatched sampling and linking models. Canadian Journal of Statistics, **30**, 3–15.

You, Y. and Rao, J. N. K. and Gambino, J. (2003), Model-based unemployment rate estimation for the Canadian Labour Force Survey: a hierarchical Bayes approach. Survey Methodology, **29**, 25–32.

Wang, J. (2000), Topics in Small Area Estimation with Applications to the National Resources Inventory. Unpublished Ph.D. thesis, Iowa State University, Ames, Iowa.

Zhang, D. and Davidian, M. (2001), Linear mixed models with flexible distributions of random effects for longitudinal data. Biometrics, **57**, 795–802.