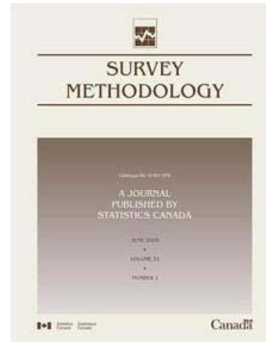
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

On robust small area estimation using a simple random effects model



by N.G.N. Prasad and J.N.K. Rao

June 1999

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 1999
Vol. 25, No. 1, pp. 67-72
Statistics Canada, Catalogue No. 12-001-X

On robust small area estimation using a simple random effects model

N.G.N. Prasad and J.N.K. Rao [1]

Abstract

Robust small area estimation is studied under a simple random effects model consisting of a basic (or fixed effects) model and a linking model that treats the fixed effects as realizations of a random variable. Under this model a model-assisted estimator of a small area mean is obtained. This estimator depends on the survey weights and remains design-consistent. A model-based estimator of its mean squared error (MSE) is also obtained. Simulation results suggest that the proposed estimator and Kott's (1989) model-assisted estimator are equally efficient, and that the proposed MSE estimator is often much more stable than Kott's MSE estimator, even under moderate deviations of the linking model. The method is also extended to nested error regression models.

Key Words: Design consistent; Linking model; Mean squared error; Survey weights.

1. Introduction

Unite-level random effects models are often used in small area estimation to obtain efficient model-based estimators of small area means. Such estimators typically do not make use of the survey weights (*e.g.*, Ghosh and Meeden 1986; Battese, Harter and Fuller 1988; Prasad and Rao 1990). As a result, the estimators are not design consistent unless the sampling design is self-weighting within areas. We refer the reader to Ghosh and Rao (1994) for an appraisal of small area estimation methods.

Kott (1989) advocated the use of design-consistent model-based estimators (*i.e.*, model assisted estimators) because such estimators provide protection against model failure as the small area sample size increases. He derived a design-consistent estimator of a small area mean under a simple random effects model. This model has two components: the basic (or fixed effects) model and the linking model. The basic model is given by

$$y_{ij} = \theta_i + e_{ij},\ j = 1, 2, ..., N_i;\ i = 1, 2, ..., m \qquad (1)$$

where the y_{ij} are the population values and the e_{ij} are uncorrelated random errors with mean zero and variance σ_i^2 for each small area $i(= 1, 2, ..., m)$. For simplicity, we take θ_i as the small area mean $\overline{Y}_i = \sum_j y_{ij}/N_i$, where N_i is the number of population units in the i^{th} area. Note that $\overline{Y}_i = \theta_i + \overline{E}_i$ and $\overline{E}_i = \sum_j e_{ij} / N_i \approx 0$ if N_i is large.

The linking model assumes that θ_i is a realization of a random variable satisfying the model

$$\theta_i = \mu + v_i \qquad (2)$$

where the v_i are uncorrelated random variables with mean zero and variance σ_v^2. Further, $\{v_i\}$ and $\{e_{ij}\}$ are assumed to be uncorrelated.

Assuming that the model (1) also holds for the sample $\{y_{ij},\ j = 1, 2, ..., n_i;\ i = 1, 2, ..., m\}$ and combining the sample model with the linking model, Kott (1989) obtained the familiar unit-level random effects model

$$y_{ij} = \mu + v_i + e_{ij},\ j = 1, 2, ..., n_i;\ i = 1, 2, ..., m, \qquad (3)$$

also called the components-of-variance model. It is customary to assume equal variances $\sigma_i^2 = \sigma^2$, although the case of random error variances has also been studied (Kleffe and Rao 1992; Arora and Lahiri 1997).

Assuming $\sigma_i^2 = \sigma^2$, Kott (1989) derived an efficient estimator $\hat{\theta}_{iK}$ of θ_i which is both model-unbiased under (3) and design-consistent. He also proposed an estimator of its mean squared error (MSE) which is model unbiased under the basic model (1) as well as design-consistent. But this MSE estimator can be quite unstable and can even take negative values, as noted by Kott (1989) in his empirical example. Kott (1989) used his MSE estimators mainly to compare the overall reduction in MSE from using $\hat{\theta}_{iK}$ in place of a direct design-based estimator $\bar{y}_{iw}$ given by (4) below. He remarked that more stable MSE estimators are needed.

The main purpose of this paper is to obtain a pseudo empirical best linear unbiased prediction (EBLUP) estimator of θ_i which depends on the survey weights and is design-consistent (section 2). A stable model-based MSE estimator is also obtained (section 3). Results of a simulation study in section 4 show that the proposed MSE estimator is often much more stable than the MSE estimator of Kott, as measured by their coefficient of variation, even under moderate deviations of the linking model (2). Results under the simple model (3) are also extended to a nested error regression model (section 5).

1. N.G.N. Prasad, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, T6G 2G1; J.N.K. Rao, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, K1S 5B6.

2. Pseudo EBLUP estimator

Suppose $\tilde{w}_{ij}$ denotes the basic design weight attached to the j^{th} sample unit $(j = 1, 2, ..., n_i)$ in the i^{th} area $(i = 1, 2, ..., m)$. A direct design-based estimator of θ_i is then given by the ratio estimator

$$\bar{y}_{iw} = \sum_j \tilde{w}_{ij} y_{ij} \Big/ \sum_j \tilde{w}_{ij} = \sum_j w_{ij} y_{ij} \qquad (4)$$

where $w_{ij} = \tilde{w}_{ij} / \sum_j \tilde{w}_{ij}$. The direct estimator $\bar{y}_{iw}$ is design-consistent but fails to borrow strength from the other areas.

To get a more efficient estimator, we consider the following reduced model obtained from the combined model (3) with $\sigma_i^2 = \sigma^2$:

$$\begin{aligned} \bar{y}_{iw} &= \sum_j w_{ij} (\mu + v_i + e_{ij}) \\ &= \mu + v_i + \bar{e}_{iw}, \end{aligned} \qquad (5)$$

where the $\bar{e}_{iw}$ are uncorrelated random variables with mean zero and variance $\delta_i = \sigma^2 \sum_j w_{ij}^2$. The reduced model (5) is an area-level model similar to the well-known Fay-Herriot model (Fay and Herriot 1979). It now follows from the standard best linear unbiased prediction (BLUP) theory (*e.g.*, Prasad and Rao 1990) that the BLUP estimator of $\theta_i = \mu + v_i$ for the reduced model (5) is given by

$$\tilde{\theta}_i = \tilde{\mu}_w + \tilde{v}_i, \qquad (6)$$

where

$$\tilde{v}_i = \gamma_{iw} (\bar{y}_{iw} - \tilde{\mu}_w)$$

with $\tilde{\mu}_w = \sum_i \gamma_{iw} \bar{y}_{iw} / \sum_i \gamma_{iw}$ and $\gamma_{iw} = \sigma_v^2 / (\sigma_v^2 + \delta_i)$. Note that $\tilde{\theta}_i$ is different from the BLUP estimator under the full model (3). We therefore denote $\tilde{\theta}_i$ as a pseudo-BLUP estimator. The estimator (6) may also be written as a convex combination of the direct estimator $\bar{y}_{iw}$ and $\tilde{\mu}_w$:

$$\tilde{\theta}_i = \gamma_{iw} \bar{y}_{iw} + (1 - \gamma_{iw})\, \tilde{\mu}_w. \qquad (7)$$

The estimator $\tilde{\theta}_i$ depends on the parameters σ_v^2 and σ^2 which are generally unknown in practice. We therefore replace σ_v^2 and σ^2 in (7) by model-consistent estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$ under the original unit-level model (3) to obtain the estimator

$$\hat{\theta}_i = \hat{\gamma}_{iw}\, \bar{y}_{iw} + (1 - \hat{\gamma}_{iw})\, \hat{\mu}_w, \qquad (8)$$

where

$$\hat{\gamma}_{iw} = \hat{\sigma}_v^2 \Big/ \left(\hat{\sigma}_v^2 + \hat{\sigma}^2 \sum_j w_{ij}^2 \right)$$

and

$$\hat{\mu}_w = \sum_i \hat{\gamma}_{iw} \bar{y}_{iw} \Big/ \sum_i \hat{\gamma}_{iw}.$$

The estimator $\hat{\theta}_i$ will be referred to as pseudo-EBLUP estimator. We use standard estimators of σ_v^2 and σ^2, based on the within-area sums of squares

$$Q_w = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$$

and the between-area sums of squares

$$Q_b = \sum_i n_i (\bar{y}_i - \bar{y})^2,$$

where $\bar{y} = \sum_i n_i \bar{y}_i \big/ \sum_i n_i$ is the overall sample mean. We have

$$\hat{\sigma}^2 = Q_w \Big/ \left(\sum_i n_i - m \right)$$

and $\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0)$ where

$$\tilde{\sigma}_v^2 = [Q_b - (m-1)\hat{\sigma}^2] / n^*$$

with

$$n^* = \sum_i n_i - \sum_i n_i^2 \Big/ \sum_i n_i.$$

It may be noted that σ_v^2 and σ^2 are either not estimable or poorly estimated from the reduced model (5) due to identifiability problems. Following Kackar and Harville (1984), it can be shown that the pseudo-EBLUP estimator θ_i is model-unbiased for θ_i under the original model (3) for symmetrically distributed errors $\{v_i\}$ and $\{e_{ij}\}$, not necessarily normal. It is also design consistent, assuming that $n_i \sum_j w_{ij}^2$ is bounded as n_i increases, because $\hat{\gamma}_{iw}$ converges in probability to 1 as $n_i \to \infty$ regardless of the validity of the model (3), assuming $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$ converge in probability to some values, say, $\hat{\sigma}_v^{*2}$ and $\hat{\sigma}^{*2}$.

Kott's (1989) model-based estimator of θ_i is obtained by taking a weighted combination of $\bar{y}_{iw}$ and $\sum_{l \neq i} c_l^{(i)} \bar{y}_l$, that is,

$$f_i(\alpha_i, c^{(i)}) = (1 - \alpha_i)\, \bar{y}_{iw} + \alpha_i \sum_{l \neq i} c_l^{(i)}\, \bar{y}_l,$$

and then minimizing the model mean squared error (MSE) of $f_i(\alpha_i, c^{(i)})$ with respect to α_i and $c_l^{(i)}$ subject to model-unbiasedness condition: $\sum_{l \neq i} c_l^{(i)} = 1$. This leads to

$$\hat{\theta}_{iK} = f_i(\hat{\alpha}_i, \hat{c}^{(i)}) \qquad (9)$$

with

$$\hat{\alpha}_i = \sum_j \left[w_{ij}^2 \Big/ \left\{ \sum_j w_{ij}^2 + \sum_{l \neq i} \hat{c}_l^{(i)^2} \Big/ n_i + \left(1 + \sum_{l \neq i} \hat{c}_l^{(i)^2} \right) (\hat{\sigma}_v^2 / \hat{\sigma}^2) \right\} \right]$$

and

$$\hat{c}_l^{(i)} = \left[(\hat{\sigma}_v^2 / \hat{\sigma}^2) + n_l^{-1} \right] \Big/ \sum_{h \neq i} \left[(\hat{\sigma}_v^2 / \hat{\sigma}^2) + n_h^{-1} \right].$$

The estimator $\hat{\theta}_{iK}$ is also model-unbiased and design-consistent. In a previous version of this paper, we proposed an estimator similar to (9). It uses the best estimators of μ under the unit-level model, based on the unweighted means $\bar{y}_i$, rather than $\hat{\mu}_w$, the best estimator of μ under the reduced model (4), based on the survey-weighted means $\bar{y}_{iw}$.

3. Estimators of MSE

It is straightforward to derive the MSE of the pseudo-BLUP estimator $\tilde{\theta}_i$ under the unit level model (3). We have

$$\mathrm{MSE}(\tilde{\theta}_i) = E(\tilde{\theta}_i - \theta_i)^2 = g_{1i}(\sigma_v^2, \sigma^2) + g_{2i}(\sigma_v^2, \sigma^2) \quad (10)$$

with

$$g_{1i}(\sigma_v^2, \sigma^2) = (1-\gamma_{iw})\, \sigma_v^2$$

and

$$g_{2i}(\sigma_v^2, \sigma^2) = \sigma_v^2 (1-\gamma_{iw})^2 \Big/ \textstyle\sum_i \gamma_{iw}.$$

The leading term, $g_{1i}(\sigma_v^2, \sigma^2)$ is of order $O(1)$, while the second term, $g_{2i}(\sigma_v^2, \sigma^2)$, due to estimation of μ is of order $O(m^{-1})$ for large m.

A naïve MSE estimator of the pseudo-EBLUP estimator $\hat{\theta}_i$ is obtained by estimating $\mathrm{MSE}(\hat{\theta}_i)$ given by (10):

$$\mathrm{mse_N}(\hat{\theta}_i) = g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2) + g_{2i}(\hat{\sigma}_v^2, \hat{\sigma}^2). \quad (11)$$

But (11) could lead to significant underestimation of $\mathrm{MSE}(\hat{\theta}_i)$ because it ignores the uncertainty associated with $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$. Note that

$$\mathrm{MSE}(\hat{\theta}_i) = \mathrm{MSE}(\tilde{\theta}_i) + E(\hat{\theta}_i - \tilde{\theta}_i)^2 \quad (12)$$

under normality of the errors $\{v_j\}$ and $\{e_{ij}\}$ so that $\mathrm{MSE}(\tilde{\theta}_i)$ is always smaller than $\mathrm{MSE}(\hat{\theta}_i)$; see Kackar and Harville (1984).

To get a "correct" estimator of $\mathrm{MSE}(\hat{\theta}_i)$, we first approximate the second order term $E(\hat{\theta}_i - \tilde{\theta}_i)^2$ in (12) for large m, assuming that $\{v_i\}$ and $\{e_{ij}\}$ are normally distributed. Following Prasad and Rao (1990), we have

$$E(\hat{\theta}_i - \tilde{\theta}_i)^2 = g_{3i}(\sigma_v^2, \sigma^2) \quad (13)$$

where the neglected terms are of lower orders than m^{-1}, and

$$g_{3i}(\sigma_v^2, \sigma^2) = \gamma_{iw}(1-\gamma_{iw})^2 \sigma_v^{-2} \{V(\tilde{\sigma}_v^2) - 2(\sigma_v^2/\sigma^2)\mathrm{Cov}(\tilde{\sigma}_v^2, \hat{\sigma}^2) + (\sigma_v^2/\sigma^2)^2 \mathrm{Var}(\hat{\sigma}^2)\}; \quad (14)$$

see Appendix 1. The variances and covariances of $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$ are also given in the Appendix 1. It can be shown that $g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2) + g_{3i}(\hat{\sigma}_v^2, \hat{\sigma}^2)$ is approximately unbiased for $g_{1i}(\sigma_v^2, \sigma^2)$ in the sense that its bias is of lower order than m^{-1} (see Appendix 2). Similarly, $g_{2i}(\hat{\sigma}_v^2, \hat{\sigma}^2)$ and $g_{3i}(\hat{\sigma}_v^2, \hat{\sigma}^2)$ are approximately unbiased for $g_{2i}(\sigma_v^2, \sigma^2)$ and $g_{3i}(\sigma_v^2, \sigma^2)$, respectively. It now follows that an approximately model-unbiased estimator of $\mathrm{MSE}(\hat{\theta}_i)$ is given by

$$\mathrm{mse}(\hat{\theta}_i) = g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2) + g_{2i}(\hat{\sigma}_v^2, \hat{\sigma}^2) + 2g_{3i}(\hat{\sigma}_v^2, \hat{\sigma}^2). \quad (15)$$

For the estimator $\hat{\theta}_{iK}$ given by (9), Kott (1989) proposed an estimator of MSE as

$$\mathrm{mse}(\hat{\theta}_{iK}) = (1-2\hat{\alpha}_i)\, v^*(\bar{y}_{iw}) + \hat{\alpha}_i^2 \left(\bar{y}_{iw} - \sum_{l \neq i} c_l^{(i)} \bar{y}_l \right)^2, \quad (16)$$

where $v^*(\bar{y}_{iw})$ is both a design-consistent estimator of the design-MSE of $\bar{y}_{iw}$ and a model-unbiased estimator of the model-variance of $\bar{y}_{iw}$ under the basic model (1). Since $\hat{\alpha}_i$ converges in probability to zero as $n_i \to \infty$, it follows from (16) that $\mathrm{mse}(\hat{\theta}_{iK})$ is also both design-consistent and model unbiased assuming only the basic model (1). However, $\mathrm{mse}(\hat{\theta}_{iK})$ is unstable and can even take negative values when $\hat{\alpha}_i$ exceeds 0.5, as noted by Kott (1989).

Note that our MSE estimator, $\mathrm{mse}(\hat{\theta}_i)$ is based on the full model (3) obtained by combining the basic model (1) with the linking model (2). However, our simulation results in section 4 show that it may perform well even under moderate deviations from the linking model.

4. Simualtion study

We conducted a limited simulation study to evaluate the performances of the proposed estimator $\hat{\theta}_i$, given by (8), and its estimator of MSE, given by (15), relative to Kott's estimator $\hat{\theta}_{iK}$, given by (9), and its estimator of MSE, given by (16). We studied the performances under two different approaches: (i) For each simulation run, a finite population of $m = 30$ small areas with $N_i = 200$ population units in each area is generated from the assumed unit-level model and then a PPS (probability proportional to size) sample within each small area is drawn independently, using $n_i = 20$. (ii) A fixed finite population is first generated from the assumed unit-level model and then for each simulation run a PPS sample within each small area is drawn independently, employing the fixed finite population. Approach (i) refers to both the design and the linking model whereas approach (ii) is design-based in the sense that it refers only to the design. The errors $\{v_i\}$ and $\{e_{ij}\}$ are assumed to be normally distributed in generating the finite populations $\{y_{ij}, i = 1, 2, ..., 30;\ j = 1, 2, ..., 200\}$. We considered two cases: (1) The linking model (2) is true with $\mu = 50$. (2) The linking model is violated by letting μ vary across areas: $\mu_i = 50, i = 1, 2, ..., 10$; $\mu_i = 55, i = 11, 12, ..., 20$; $\mu_i = 60$,

$i = 21, 22, ..., 30.$ To implement PPS sampling within each area, size measures z_{ij} $(i = 1, 2, ..., 30;\ j = 1, 2, ..., 200)$ were generated from an exponential distribution with mean 200. Using these z-values, we computed selection probabilities $p_{ij} = z_{ij} / \sum_j z_{ij}$ for each area i and then used them to select PPS with replacement samples of sizes $n_i = n$, by taking $n = 20$, and the associated sample values $\{y_{ij}\}$ were observed.

The basic design weights are given by $\tilde{w}_{ij} = n^{-1} p_{ij}^{-1}$ so that $w_{ij} = p_{ij}^{-1} / \sum_j p_{ij}^{-1}$. Using these weights and the associated sample values y_{ij} we compute estimates $\hat{\theta}_i$ and $\hat{\theta}_{iK}$ and associated estimates of MSE, and also the ratio estimate $\bar{y}_{iw}$ for each simulation run; the formula for $v^*(\bar{y}_{iw})$ under PPS sampling is given in Appendix 3. This process was repeated $R = 10{,}000$ times to get from each run $r (= 1, 2, ..., R)$ $\hat{\theta}_i(r)$ and $\hat{\theta}_{iK}(r)$ and associated MSE estimates $\text{mse}_i(\hat{\theta}_i(r))$ and $\text{mse}_i(\hat{\theta}_{iK}(r))$ and also the direct estimate $\bar{y}_{iw}(r)$. Using these values, empirical relative efficiencies (RE) of $\hat{\theta}_i$ and $\hat{\theta}_{iK}$ over $\bar{y}_{iw}$ were computed as

$$\text{RE}(\hat{\theta}_i) = \text{MSE}_*(\bar{y}_{iw}) / \text{MSE}_*(\hat{\theta}_i)$$

and

$$\text{RE}(\hat{\theta}_{iK}) = \text{MSE}_*(\bar{y}_{iw}) / \text{MSE}_*(\hat{\theta}_{iK}),$$

where MSE_* denotes the MSE over $R = 10{,}000$ runs. For example, $\text{MSE}_*(\hat{\theta}_i) = \sum_r [\hat{\theta}_i(r) - \bar{Y}_i(r)]^2 / R$, where $\bar{Y}_i(r)$ is the i^{th} area population mean for the r^{th} run. Note that $\bar{Y}_i(r)$ remains the same over the runs r under the design-based approach because the finite population is fixed over the simulation runs.

Similarly, the relative biases of the MSE estimators were computed as

$$\text{RB}[\text{mse}(\hat{\theta}_i)] = [\text{MSE}_*(\hat{\theta}_i) - E_* \text{mse}(\hat{\theta}_i)] / \text{MSE}_*(\hat{\theta}_i)$$

and

$$\text{RB}[\text{mse}(\hat{\theta}_{iK})] = [\text{MSE}_*(\hat{\theta}_{iK}) - E_* \text{mse}(\hat{\theta}_{iK})] / \text{MSE}_*(\hat{\theta}_{iK}),$$

where E_* denotes the expectation over $R = 10{,}000$ runs. For example, $E_* \text{mse}(\hat{\theta}_i) = \sum_r \text{mse}(\hat{\theta}_i(r)) / R$. Finally, the empirical coefficient of variation (CV) of the MSE estimators were computed as

$$\text{CV}[\text{mse}(\hat{\theta}_i)] = [\text{MSE}_* \{\text{mse}(\hat{\theta}_i)\}]^{1/2} / \text{MSE}_*(\hat{\theta}_i)$$

and

$$\text{CV}[\text{mse}(\hat{\theta}_{iK})] = [\text{MSE}_* \{\text{mse}(\hat{\theta}_{iK})\}]^{1/2} / \text{MSE}_*(\hat{\theta}_{iK}).$$

Note that $\text{MSE}_*[\text{mse}(\hat{\theta}_i)] = \sum_r [\text{mse}(\hat{\theta}_i(r)) - \text{MSE}_*(\hat{\theta}_i)]^2 / R$ and a similar expression for $\text{MSE}_*[\text{mse}(\hat{\theta}_{iK})]$.

Table 1 reports summary measures of the values of percent RE, $|\text{RB}|$ and CV for cases (1) and (2) under approach (i). Summary measures under approach (ii) are reported in Table 2. Summary measures considered are the mean and the median (med) over the small areas $i = 1, 2, ..., 30$.

Table 1
Relative efficiency (RE) of estimators, absolute relative bias (|RB|) and coefficient of variation (CV) of MSE estimators ($\sigma = 5.0$, $n = 20$): Approach (i)

σ_v		RE%		|RB|%		CV%	
		$\hat{\theta}_{iK}$	$\hat{\theta}_i$	mse($\hat{\theta}_{iK}$)	mse($\hat{\theta}_i$)	mse($\hat{\theta}_{iK}$)	mse($\hat{\theta}_i$)
				Case 1			
1	Mean	190	177	15.3	3.5	148	25
	Med	190	182	14.8	2.6	148	25
2	Mean	126	123	5.1	3.2	48	8
	Med	127	124	5.6	2.9	48	8
3	Mean	113	111	3.5	2.7	35	6
	Med	112	111	3.2	3.0	35	6
				Case 2			
1	Mean	108	103	10.4	7.9	39	6
	Med	108	104	11.1	7.7	38	5
2	Mean	108	104	13.3	8.9	39	6
	Med	108	104	13.6	7.9	37	6
3	Mean	104	103	11.5	7.2	37	5
	Med	105	105	13.1	8.0	36	6

Case 1: $\mu_i = 50$, $i = 1, 2, \ldots, 30$; Case 2: $\mu_i = 50$, $i = 1, 2, \ldots, 10$; $\mu_i = 55$, $i = 11, 12, \ldots, 20$; $\mu_i = 60$, $i = 21, 22, \ldots, 30$.

It is clear from Tables 1 and 2 that $\hat{\theta}_{iK}$ and $\hat{\theta}_i$ perform similarly with respect to RE which decreases as σ_v/σ increases. Under approach (ii), RE is large for both cases 1 and 2 when $\sigma_v/\sigma \le 0.4$, whereas it decreases significantly under approach (i) if the linking model is violated (case 2); the direct estimator $\bar{y}_{iw}$ is quite unstable under approach (ii).

Turning to the performance of MSE estimators under approach (i), Table 1 shows that $|\text{RB}|$ of $\text{mse}(\hat{\theta}_i)$ is negligible $(< 4\%)$ when the linking model holds (Case 1) and that it is small $(< 10\%)$ even when the linking model is violated, although it increases. The estimator $\text{mse}(\hat{\theta}_{iK})$ has a larger $|\text{RB}|$ but it is less than 15%. The CV of $\text{mse}(\hat{\theta}_i)$ is much smaller than the CV of $\text{mse}(\hat{\theta}_{iK})$ for both Cases 1 and 2. For example, when the model holds (Case 1) the median CV is 25% for $\text{mse}(\hat{\theta}_i)$ compared to 148% for $\text{mse}\,\hat{\theta}_{iK}$ when $\sigma_v = 1$; the median CV decreases to 8% for $\text{mse}(\hat{\theta}_i)$ compared to 48% for $\text{mse}(\hat{\theta}_{iK})$ when $\sigma_v = 2$. This pattern is retained when the model is violated (Case 2). It may be noted that the probability of $\text{mse}(\hat{\theta}_{iK})$ taking a negative value is quite large (>0.3) when $\sigma_v / \sigma \le 0.4$.

Under approach (ii), Table 2 shows that $|\text{RB}|$ of $\text{mse}(\hat{\theta}_i)$ is larger than the value under approach (i) and ranges from 15% to 25%. On the other hand, $|\text{RB}|$ of $\text{mse}(\hat{\theta}_{iK})$ is smaller and ranges from 4% to 15%. The CV of $\text{mse}(\hat{\theta}_{iK})$, however, is much larger than under approach (i). For example, the median CV for Case 1 is 295% compared to 38% for $\text{mse}(\hat{\theta}_i)$ when $\sigma_v = 1$ which decreases to 122% compared to 23% when $\sigma_v = 2$. A similar pattern holds for case 2 where the fixed finite population is generated from the model with varying means.

Table 2
Relative efficiency (RE) of estimators, absolute relative bias (|RB|) and coefficient of variation (CV) of MSE estimators ($\sigma = 5.0$, $n = 20$): Approach (ii)

σ_v		RE%		|RB|%		CV%	
		$\hat{\theta}_{iK}$	$\hat{\theta}_i$	$\text{mse}(\hat{\theta}_{iK})$	$\text{mse}(\hat{\theta}_i)$	$\text{mse}(\hat{\theta}_{iK})$	$\text{mse}(\hat{\theta}_i)$
				Case 1			
1	Mean	283	281	14.2	25.4	289	39
	Med	275	279	15.0	24.7	295	38
2	Mean	180	182	7.3	19.2	115	24
	Med	177	181	6.9	18.7	122	23
3	Mean	129	129	4.8	14.8	68	24
	Med	129	128	4.2	13.9	65	24
				Case 2			
1	Mean	278	276	15.7	26.8	291	41
	Med	271	275	16.6	26.2	297	40
2	Mean	175	177	8.8	20.7	117	26
	Med	173	177	8.5	20.3	124	25
3	Mean	124	124	6.3	16.2	70	25
	Med	125	124	6.8	15.5	67	26

Case 1: $\mu_i = 50$, $i = 1, 2, \ldots, 30$; Case 2: $\mu_i = 50$, $i = 1, 2, \ldots, 10$; $\mu_i = 55$, $i = 11, 12, \ldots, 20$; $\mu_i = 60$, $i = 21, 22, \ldots, 30$.

To reduce $|\text{RB}|$ of $\text{mse}(\hat{\theta}_i)$ under approach (ii), one could combine it with $\text{mse}(\hat{\theta}_{iK})$ by taking a weighted average, but it appears difficult to chose the appropriate weights. The weighted average will be more stable than $\text{mse}(\hat{\theta}_{iK})$.

5. Nested error regression model

The results in sections 2 and 3 can be extended to nested error regression models

$$y_{ij} = x'_{ij}\beta + v_i + e_{ij},\ \ j = 1, 2, \ldots, n_i;\ \ i = 1, 2, \ldots, m \qquad (17)$$

using the results of Prasad and Rao (1990), where x_{ij} is a p-vector of auxiliary variables with known population mean $\bar{X}_i$ and related to y_{ij}, and β is the p-vector of regression coefficients. The reduced model is given by

$$\bar{y}_{iw} = \bar{x}'_{iw}\beta + v_i + \bar{e}_{iw} \qquad (18)$$

with $\bar{x}'_{iw} = \sum_j w_{ij} x_{ij}$. Model-consistent estimates $\hat{\sigma}_v^2$ and $\hat{\sigma}^2$ are obtained from the unit-level model (17), employing either the method of fitting constants (Prasad and Rao 1990) or REML (restricted maximum likelihood) estimation (Datta and Lahiri 1997).

The pseudo-EBLUP of $\theta_i = \bar{X}'_i\beta + v_i$ is given by

$$\hat{\theta}_i = \hat{\gamma}_{iw}\bar{y}_{iw} + (1 + \hat{\gamma}_{iw})\bar{X}'_i\hat{\beta}_w, \qquad (19)$$

where

$$\hat{\beta}_w = \left(\sum_i \hat{\gamma}_{iw}\bar{x}_{iw}\bar{x}'_{iw}\right)^{-1}\left(\sum_i \hat{\gamma}_{iw}\bar{x}_{iw}\bar{x}_{iw}\right).$$

An approximate model-unbiased estimator of $\text{MSE}(\hat{\theta}_i)$ is given by (15) with

$$g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2) = (1 - \hat{\gamma}_{iw})\hat{\sigma}_v^2$$

as before,

$$g_{2i}(\hat{\sigma}_v^2, \hat{\sigma}2) =$$

$$\hat{\sigma}_v^2(\bar{X}_i - \hat{\gamma}_{iw}\bar{x}_{iw})'\left(\sum_i \hat{\gamma}_{iw}\bar{x}_{iw}\bar{x}'_{iw}\right)^{-1}(\bar{X}_i - \hat{\gamma}_{iw}\bar{x}_{iw}\bar{x}_{iw})$$

and $g_{3i}(\hat{\sigma}_v^2, \hat{\sigma}^2)$, obtained from (14), involves the estimated variances and covariances of $\tilde{\sigma}_v^2$ and $\hat{\sigma}^2$. The latter can be obtained from Prasad and Rao (1990) for the method of fitting constants and from Datta and Lahiri (1997) for REML.

6. Conclusion

We have proposed a model-assisted estimator of a small area mean under a simple unit-level random effects model. This estimator depends on the survey weights and is design-consistent. We have also obtained a model-based MSE estimator. Results of our simulation study have shown that the proposed MSE estimator performs well, even under moderate deviations of the linking model. The proposed approach is also extended to a nested error regression model.

Acknowledgements

This work was supported by research grants from the Natural Sciences and Engineering Research Council of Canada. We are thankful to the Associate Editor and the referee for constructive comments and suggestions.

Appendix 1

Proof of (13):

From general results (Prasad and Rao 1990) we have

$$E(\hat{\theta}_i - \tilde{\theta}_i)^2 \approx tr[A_i(\sigma_v^2, \sigma^2)B_i(\sigma_v^2, \sigma^2)],$$

where $B_i(\sigma_v^2, \sigma^2)$ is the 2×2 covariance matrix of $\tilde{\sigma}_v^2$ and $\hat{\sigma}^2$, and $A_i(\sigma_v^2, \sigma^2)$ is the 2×2 covariance matrix of

$$\left(\frac{\partial\theta_i^*}{\partial\sigma_v^2}, \frac{\partial\theta_i^*}{\partial\sigma^2}\right).$$

Now, noting that

$$\frac{\partial\theta_i^*}{\partial\sigma_v^2} = \frac{\partial\gamma_{iw}}{\partial\sigma_v^2}\bar{y}_{iw} = \left[\frac{\gamma_{iw}(1-\gamma_{iw})}{\sigma_v^2}\right]\bar{y}_{iw},$$

$$\frac{\partial\theta_i^*}{\partial\sigma^2} = \frac{\partial\gamma_{iw}}{\partial\sigma^2}\bar{y}_{iw} = -\left[\frac{\gamma_{iw}(1-\gamma_{iw})}{\sigma^2}\right]\bar{y}_{iw},$$

and $V(\bar{y}_{iw}) = \sigma_v^2 + \delta_i = \sigma_v^2/\gamma_{iw}$, we get

$$A_i(\sigma_v^2, \sigma^2) = [\gamma_{iw}(1-\gamma_{iw})^2\sigma_v^{-2}]\begin{bmatrix} 1 & -\sigma_v^2/\sigma^2 \\ -\sigma_v^2/\sigma^2 & (\sigma_v^2/\sigma)^2 \end{bmatrix},$$

and hence the result (14).

Covariance matrix of $\tilde{\sigma}_v^2$ and $\hat{\sigma}^2$:

Under normality, we have

$$V(\hat{\sigma}^2) = 2\sigma^4 \Big/ \left(\sum\nolimits_i n_i - m\right),$$

$$V(\tilde{\sigma}_v^2) = 2n_*^{-2}$$

$$\left[\sigma^4(m-1)\left(\sum n_i - 1\right)\left(\sum n_i - m\right)^{-1} + 2n_*\sigma^2\sigma_v^2 + n_{**}\sigma_v^4\right]$$

and

$$\mathrm{Cov}(\hat{\sigma}^2, \tilde{\sigma}_v^2) = -(m-1)\,n_*^{-1}V(\hat{\sigma}^2),$$

where

$$n_{**} = \sum n_i^2 - 2\sum n_i^3 \Big/ \sum n_i + \left(\sum n_i^2\right)^2 \Big/ \left(\sum n_i\right)^2;$$

see Searle, Casella and McCulloch (1992, page 428).

Appendix 2

Proof of $E[g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2) + g_{zi}(\hat{\sigma}_v^2, \hat{\sigma}^2)] \approx g_{1i}(\sigma_v^2, \sigma^2)$:

By a Taylor expansion of $g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2)$ around (σ_v^2, σ^2) to second order and noting that $E(\hat{\sigma}^2 - \sigma^2) = 0$ and $E(\hat{\sigma}_v^2 - \sigma_v^2) \approx 0$, we get

$$E[g_{1i}(\hat{\sigma}_v^2, \hat{\sigma}^2) - g_{1i}(\sigma_v^2, \sigma^2)]$$

$$\approx \frac{1}{2}tr[D_i(\sigma_v^2, \sigma^2)B_i(\sigma_v^2, \sigma^2)],$$

where $D_i(\sigma_v^2, \sigma^2)$ is the 2 x 2 matrix of second order derivatives of $g_{1i}(\sigma_v^2, \sigma^2)$ with respect to σ_v^2 and σ^2. It is easy to verify that

$$\frac{1}{2}tr[D_i(\sigma_v^2, \sigma^2)\,B_i(\sigma_v^2, \sigma^2)] = g_{3i}(\sigma_v^2, \sigma^2).$$

Now, noting that $E[g_{3i}(\hat{\sigma}_v^2, \hat{\sigma}^2)] \approx g_{3i}(\sigma_v^2, \sigma^2)$ we get the desired result.

Appendix 3

The design-based estimator of variance of $\bar{y}_{iw}$ under PPS sampling is given by

$$v(\bar{y}_{iw}) = \frac{m}{m-1}\sum_j w_{ij}^2(y_{ij} - \bar{y}_{iw})^2.$$

Kott (1989) model-assisted variance estimator is

$$v^*(\bar{y}_{iw}) = \{V(\bar{y}_{iw})/Ev(\bar{y}_{iw})\}\,v(\bar{y}_{iw})$$

$$= \frac{\left(\sum_j w_{ij}^2\right)\sum_j w_{ij}^2(y_{ij} - \bar{y}_{iw})^2}{\sum_j w_{ij}^2\left(1 - 2w_{ij} + \sum_j w_{ij}^2\right)},$$

where E and V denote expectation and variance with respect to the basic model (1).

References

Arora, V., and Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statistica Sinica*, 7, 1053-1063.

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

Datta, G.S., and Lahiri, P. (1997). A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictor in Small-Area Estimation Problems. Technical Report, University of Nebraska-Lincoln.

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, 269-277.

Ghosh, M., and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. *Journal of the American Statistical Association*, 81, 1058-1069.

Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55-93.

Kackar, R.N., and Harville, D.A. (1984). Approximations for standard errors of estimators for fixed and random effects models. *Journal of the American Statistical Association*, 79, 853-862.

Kleffe, J., and Rao, J.N.K. (1992). Estimation of mean square error of empirical best linear unbiased predictors under a random error variance linear model. *Journal of Multivariate Analysis*, 43, 1-15.

Kott, P. (1989). Robust small domain estimation using random effects modelling. *Survey Methodology*, 15, 3-12.

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

Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*. New York: John Wiley & Sons, Inc.