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

Article

Small area estimation under a restriction

by Junyuan Wang, Wayne A. Fuller and Yongming Qu

June 2008

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2008
Vol. 34, No. 1, pp. 29-36
Statistics Canada, Catalogue No. 12-001-X

Small area estimation under a restriction

Junyuan Wang, Wayne A. Fuller and Yongming Qu [1]

Abstract

Small area prediction based on random effects, called EBLUP, is a procedure for constructing estimates for small geographical areas or small subpopulations using existing survey data. The total of the small area predictors is often forced to equal the direct survey estimate and such predictors are said to be calibrated. Several calibrated predictors are reviewed and a criterion that unifies the derivation of these calibrated predictors is presented. The predictor that is the unique best linear unbiased predictor under the criterion is derived and the mean square error of the calibrated predictors is discussed. Implicit in the imposition of the restriction is the possibility that the small area model is misspecified and the predictors are biased. Augmented models with one additional explanatory variable for which the usual small area predictors achieve the self-calibrated property are considered. Simulations demonstrate that calibrated predictors have slightly smaller bias compared to those of the usual EBLUP predictor. However, if the bias is a concern, a better approach is to use an augmented model with an added auxiliary variable that is a function of area size. In the simulation, the predictors based on the augmented model had smaller MSE than EBLUP when the incorrect model was used for prediction. Furthermore, there was a very small increase in MSE relative to EBLUP if the auxiliary variable was added to the correct model.

Key Words: Components-of-variance model; Best linear unbiased prediction; Calibration; Design consistent; Mixed linear models.

1. Introduction

There are situations in which it is desirable to derive reliable estimators for small geographical areas or small subpopulations from existing survey data. However, sample sizes for the areas may be such that the usual survey estimators yield unacceptably large standard errors. This makes it reasonable to use a model-based estimator. See Rao (2003) for a complete discussion of small area estimation.

A model for small area estimation is

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + b_i, \tag{1}$$

$$Y_i = y_i + e_i, \qquad i = 1, ..., n, \tag{2}$$

where y_i are unobservable small area means, Y_i are observable survey estimators, $\boldsymbol{x}_i'$ are known vectors, $\boldsymbol{\beta}$ is the vector of regression parameters, b_i are independent and identically distributed random variables with $E(b_i) = 0$ and $V(b_i) = \sigma_b^2$, and e_i are sampling errors with $E(e_i \mid y_i) = 0$ and $V(e_i \mid y_i) = \sigma_{ei}^2$. Combining (1) and (2), we obtain

$$Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + b_i + e_i, \qquad i = 1, ..., n, \tag{3}$$

which is a special case of the mixed linear model.

Assuming the variance components σ_b^2 and σ_{ei}^2 to be known, the best linear unbiased estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \left[\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} = \left[\sum_{i=1}^{n}(\sigma_b^2 + \sigma_{ei}^2)^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\right]^{-1}\left[\sum_{i=1}^{n}(\sigma_b^2 + \sigma_{ei}^2)^{-1}\boldsymbol{x}_i Y_i\right], \tag{4}$$

where $\boldsymbol{X}' = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$, $\boldsymbol{Y}' = (Y_1, ..., Y_n)$, and $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{Y}) = \mathrm{diag}(\sigma_b^2 + \sigma_{e1}^2, ..., \sigma_b^2 + \sigma_{en}^2)$. Furthermore, the best linear unbiased predictor (BLUP) of y_i is

$$\tilde{y}_i^H = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + \gamma_i(Y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}), \tag{5}$$

where

$$\gamma_i = (\sigma_b^2 + \sigma_{ei}^2)^{-1}\sigma_b^2. \tag{6}$$

See Henderson (1963) and Rao (2003). When the variance components are unknown, we replace the variance components in (4) and (6) with estimators to obtain $\hat{y}_i^H$, the empirical BLUP or EBLUP.

The survey estimator of the total of all survey areas is often judged to be of adequate precision. In such cases, the practitioner may prefer to use the design consistent estimator of the total and to require that the weighted sum of the small area predictors equal the design consistent estimator. Thus, it is desirable to have small area predictors $\hat{y}_i$ that satisfy

$$\sum_{i=1}^{n}\omega_i\hat{y}_i = \sum_{i=1}^{n}\omega_i Y_i, \tag{7}$$

where ω_i are sampling weights such that $\sum_{i=1}^{n}\omega_i Y_i$ is a design consistent estimator of the total (or mean). A number of procedures have been suggested for constructing predictors to satisfy (7). Such procedures are often called "benchmarking" or "calibration", *e.g.*, Mantel, Singh and Barcau (1993) and You and Rao (2003).

To review such procedures, let $\tilde{\boldsymbol{y}}^H = (\tilde{y}_1^H, ..., \tilde{y}_n^H)'$ denote the BLUP predictor of $\boldsymbol{y} = (y_1, ..., y_n)'$ defined in (5), where

1. Junyuan Wang, Director of Biostatistics, Global Biostatistics and Programming, Wyeth Research, 35 Cambridge Park Drive, Cambridge, MA 02140; Wayne A. Fuller, Professor, Department of Statistics, Iowa State University, Ames, IA 50011; Yongming Qu, Sr. Research Scientist, Eli Lilly and Company, Lilly Corporation Center, Indianapolis, IN 26285.

$$\tilde{\boldsymbol{y}}^H = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{b}}, \tag{8}$$

$\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b}}$ are any solutions to

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1} \\ \boldsymbol{\Sigma}_e^{-1}\boldsymbol{X} & \boldsymbol{\Sigma}_e^{-1} + \boldsymbol{\Sigma}_b^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{\Sigma}_e^{-1}\mathbf{Y} \\ \boldsymbol{\Sigma}_e^{-1}\mathbf{Y} \end{bmatrix}, \tag{9}$$

$\boldsymbol{\Sigma}_b = \sigma_b^2 \boldsymbol{I}_n$, and $\boldsymbol{\Sigma}_e = \text{diag}(\sigma_{e1}^2, ..., \sigma_{en}^2)$. Equation (9) is called the mixed model equation. Finding a solution to a mixed model equation (9) is equivalent to finding a solution to the minimization problem

$$\min_{\boldsymbol{\beta},\boldsymbol{b}} \{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{b})'\boldsymbol{\Sigma}_e^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{b}) + \boldsymbol{b}'\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}\}. \tag{10}$$

Pfeffermann and Barnard (1991) proposed the modified predictor

$$\hat{\boldsymbol{y}}^{\text{PB}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{PB}} + \hat{\boldsymbol{b}}^{\text{PB}}, \tag{11}$$

where $\hat{\boldsymbol{\beta}}^{\text{PB}}$ and $\hat{\boldsymbol{b}}^{\text{PB}}$ are any solutions to the minimization problem (10) with $\boldsymbol{\beta}$ and $\boldsymbol{b}$ subject to the constraint

$$\sum_{i=1}^n \omega_i(\boldsymbol{x}_i'\boldsymbol{\beta} + b_i) = \sum_{i=1}^n \omega_i Y_i. \tag{12}$$

This leads to the predictor

$$\hat{y}_i^{\text{PB}} = \tilde{y}_i^H + [\text{Var}(\tilde{y})]^{-1}\text{cov}(\tilde{y}_i^H, \tilde{y})\left[\sum_{j=1}^n \omega_j Y_j - \tilde{y}\right], \tag{13}$$

where $\tilde{y} = \sum_{i=1}^n \omega_i \tilde{y}_i^H$, $\text{cov}(\tilde{y}_i^H, \tilde{y}) = \omega_i \gamma_i \sigma_{ei}^2 + \sum_{j=1}^n \omega_j (1-\gamma_i)(1-\gamma_j)\boldsymbol{x}_i' V(\hat{\boldsymbol{\beta}})\boldsymbol{x}_j$, and $\text{Var}(\tilde{y}) = \sum_{i=1}^n \omega_i \text{cov}(\tilde{y}_i^H, \tilde{y})$.

Isaki, Tsay and Fuller (2000) imposed the restriction by a procedure that, approximately, constructs the best predictors of $n-1$ quantities that are uncorrelated with $\sum_{i=1}^n \omega_i Y_i$. After some matrix operations, the Isaki-Tsay-Fuller (ITF) predictor can be rewritten as

$$\hat{y}_i^{\text{ITF}} = \hat{y}_i^H + \left[\sum_{j=1}^n \omega_j^2 \widehat{\text{Var}}(Y_j)\right]^{-1} \omega_i \widehat{\text{Var}}(Y_i)\left(\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \hat{y}_j^H\right), \tag{14}$$

where $\widehat{\text{Var}}(Y_i)$ is an estimator of $\sigma_b^2 + \sigma_{ei}^2$.

Note that the Pfeffermann-Barnard (PB) predictor (13), and the ITF predictor (14) have the form

$$\hat{y}_i^{\boldsymbol{a}} = \hat{y}_i + a_i\left(\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \hat{y}_j\right), \tag{15}$$

where $\sum_{i=1}^n \omega_i a_i = 1$. In other words, we may consider imposing restriction (7) to be an adjustment problem. To make an adjusted predictor $\hat{y}_i^{\boldsymbol{a}}$ satisfy (7), the difference $\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \hat{y}_j$ is allocated to small area predictor $\hat{y}_i$ using a_i.

Using the unit level model, You and Rao (2002) proposed an estimator of $\boldsymbol{\beta}$ such that the resulting predictors satisfy (7). They called such predictors self-calibrated. Applying their procedure to the area model (3), we have

$$\hat{y}_i^{\text{YR}} = \hat{\gamma}_i Y_i + (1 - \hat{\gamma}_i)\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\text{YR}}, \tag{16}$$

where

$$\hat{\boldsymbol{\beta}}_{\text{YR}} = \left[\sum_{i=1}^n \omega_i(1-\hat{\gamma}_i)\boldsymbol{x}_i\boldsymbol{x}_i'\right]^{-1}\sum_{i=1}^n \omega_i(1-\hat{\gamma}_i)\boldsymbol{x}_i Y_i. \tag{17}$$

Any predictor that has the self-calibrated property, such as the You and Rao (YR) predictor (16), is a predictor of the form (15) since the difference $\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \hat{y}_j$ is equal to zero.

We will derive the "best" predictor of the form (15) in Section 2. The results will lead to a unifying view of several BLUP based predictors. In Section 3, we propose an alternative approach that has the self-calibrated property. We will briefly discuss the mean square error (MSE) in Section 4 and use simulation studies to compare the predictors in Section 5. Conclusions and discussion will be given in Section 6.

2. "Best" linear unbiased predictor under a restriction

To find the "best" linear unbiased predictor for $\boldsymbol{y}$ that satisfies restriction (7), we first assume the parameters for the variance components are known. According to Lemma 1 of Pfeffermann and Barnard (1991), it is impossible to compare predictors that satisfy restriction (7) component-by-component to find the best one. Therefore, some kind of overall criterion is required. A natural criterion is

$$Q(\hat{\boldsymbol{y}}^{\boldsymbol{a}}) = \sum_{i=1}^n \varphi_i E(\hat{y}_i^{\boldsymbol{a}} - y_i)^2, \tag{18}$$

where the φ_i, $i = 1, ..., n$ are a chosen set of positive weights.

Theorem 1. *Assume the random effects model*

$$Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + b_i + e_i, \qquad i = 1, ..., n,$$

where the b_i have independent identical distributions with mean zero and variance σ_b^2, the e_i have independent distributions with mean zero and variance σ_{ei}^2, and $\boldsymbol{b} = (b_1, ..., b_n)'$ is independent of $\boldsymbol{e} = (e_1, ..., e_n)'$. Assume σ_b^2 and σ_{ei}^2 are known, and $\boldsymbol{\beta}$ is unknown. Let $\tilde{y}_i^H$ be the BLUP of y_i defined in (5)*. Let*

$$\hat{y}_i^{\breve{a}} = \tilde{y}_i^H + \breve{a}_i \left(\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \tilde{y}_j^H \right), \qquad (19)$$

where $\breve{a}_i = (\sum_{i=1}^n \varphi_i^{-1} \omega_i^2)^{-1} \varphi_i^{-1} \omega_i$ *and* ω_i *are the fixed weights of (7). Then* $\hat{\boldsymbol{y}}^{\breve{a}} = (\hat{y}_1^{\breve{a}}, ..., \hat{y}_n^{\breve{a}})'$ *is the unique predictor among all linear unbiased predictors that satisfy* (7) *and minimize criterion* (18).

Proof: See Appendix A.

Remark 1. When the variance components are unknown, we replace the variance components in (6) with suitable estimators to obtain the empirical BLUP or EBLUP, denoted by $\hat{y}_i^H$. Thus, we have the modified predictor

$$\hat{y}_i^{\breve{a}} = \hat{y}_i^H + \breve{a}_i \left(\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \hat{y}_j^H \right). \qquad (20)$$

Remark 2. The criterion (18) defines a "loss function", where the choice of the weights φ_i depends on the problem under consideration. For example, a statistician can decide to assign higher weights to "more important" areas and lower weights to "less important" areas. Often, φ_i is function of the variance components. In some cases, one can choose φ_i so that the derived predictors have certain desirable properties. For example, $\varphi_i = [\widehat{\text{Var}}(Y_i)]^{-1}$ gives the ITF predictors, which are BLUP (in the traditional sense) in the subspace that is orthogonal to $(\omega_1, ..., \omega_n)'$ in the space spanned by $\boldsymbol{Y}$.

Remark 3. If $\varphi_i = \omega_i [\text{cov}(\tilde{y}_i^H, \tilde{y})]^{-1}$, where $\tilde{y} = \sum_{j=1}^n \omega_j \tilde{y}_j^H$, we have the predictor (13) derived by Pfeffermann and Barnard (1991). When $\varphi_i = [\widehat{\text{Var}}(\hat{y}_j^H)]^{-1}$, we have the predictor used by Battese, Harter, and Fuller (1988).

3. An alternative way to impose the restriction

We have discussed a family of predictors in which the total for the small area predictors is equal to the total of the direct survey estimates. Implicit in the imposition of restriction (7) is the possibility that the small area predictor of the total is biased due to a misspecified model (3). In practical applications, model misspecification is a valid concern since the true mechanism that generates $\boldsymbol{Y}$ is unknown.

A common misspecification occurs when the explanatory variables used in the model are not the same as the ones that generated $\boldsymbol{Y}$. Thus, the direction of the overall bias may not be the same as the direction of the bias for a particular small area. In this case, the predictors of form (15) may increase the bias for some small areas compared to the bias before adjustment. Mantel *et al.* (1993) concluded that "Generally the effect of benchmarking here is a slight improvement in the overall bias at the cost of some deterioration with respect to the other evaluation measures".

Since the bias is nonzero if there is nonzero correlation between ω_i and $(Y_i - \hat{y}_i)$, the bias can be reduced by including ω_i in the model. That is, for a given model, one approach is to use the augmented model

$$\boldsymbol{Y} = \boldsymbol{X}_1 \boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{e}, \qquad (21)$$

where $\boldsymbol{X}_1 = (\boldsymbol{X}, \boldsymbol{\omega})$ and $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)'$, to obtain the BLUP or EBLUP. With $\boldsymbol{\omega}$ in the model, the adjustment needed to meet restriction (7) will often be much smaller than the adjustment for the model without $\boldsymbol{\omega}$.

Using an augmented model approach, we can go one step further and construct predictors that satisfy restriction (7). First, assume the variances σ_{ei}^2 are known. Note that

$$\sum_{j=1}^n \omega_j Y_j - \sum_{j=1}^n \omega_j \hat{y}_j^H = \sum_{i=1}^n \omega_i (1 - \gamma_i)(Y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}) \qquad (22)$$

and $\omega_i (1 - \gamma_i) \text{Var}(Y_i) = \omega_i \sigma_{ei}^2$. Using the theory of the linear model, we can show that the predictor constructed with the augmented model

$$\boldsymbol{Y} = \boldsymbol{X}_2 \boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{e}, \qquad (23)$$

where $\boldsymbol{X}_2 = (\boldsymbol{X}, \boldsymbol{\omega}_e)$ and $\boldsymbol{\omega}_e = (\omega_1 \sigma_{e1}^2, ..., \omega_n \sigma_{en}^2)'$, has the self-calibrated property when the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$ is used. Note that this approach gives a predictor that is different than the You-Rao predictor (16).

If the σ_{ei}^2 are unknown, we replace σ_{ei}^2 in $\boldsymbol{\omega}_e$ with its estimator $\hat{\sigma}_{ei}^2$. As long as the $\hat{\sigma}_{ei}^2$ in $\boldsymbol{\omega}_e$ is the same as the $\hat{\sigma}_{ei}^2$ used in constructing γ_i, the predictors have the self-calibrated property. If the σ_{ei}^2 has the form $\sigma_e^2 f(\boldsymbol{u}_i)$, where σ_e^2 is unknown, but $\boldsymbol{u}_i$ and $f(.)$ are known, one can construct the variables for model (23) using $\boldsymbol{\omega}_e = (\omega_1 f(\boldsymbol{u}_1), ..., \omega_n f(\boldsymbol{u}_n))'$ without estimating σ_e^2. For example, if $\sigma_{ei}^2 = m_i^{-1} \sigma_e^2$, the $\boldsymbol{\omega}_e$ for model (23) is $(m_i^{-1} \omega_1, ..., m_n^{-1} \omega_n)'$.

4. The MSE of the modified predictors

One can show that any predictor of the form (15) can be written as

$$\tilde{\boldsymbol{y}}^{\boldsymbol{a}} = \boldsymbol{Y} - \boldsymbol{C}_{\boldsymbol{a}}^{-1} \boldsymbol{B} \boldsymbol{C}_{\boldsymbol{a}} (\boldsymbol{I}_n - \boldsymbol{\Gamma})(\boldsymbol{Y} - \boldsymbol{X} \hat{\boldsymbol{\beta}}) \qquad (24)$$

by letting $\boldsymbol{C}_{\boldsymbol{a}} = \boldsymbol{A}_{\boldsymbol{a}} \boldsymbol{T}$, where

$$\boldsymbol{A}_{\boldsymbol{a}} = \begin{pmatrix} 1 & \boldsymbol{0}_{n-1}' \\ -\boldsymbol{a}_{n-1} & \boldsymbol{I}_{n-1} \end{pmatrix},$$

$$\boldsymbol{a}_{n-1} = (a_2, ..., a_n)'.$$

Therefore, the estimator for the variance of $\tilde{\boldsymbol{y}}^{a}$ defined in (14) proposed in Isaki *et al.* (2000) can be used to estimate the MSE of any predictor of the form (15). Often, the MSE of an adjusted predictor is close to the MSE of the predictor before the adjustment.

The augmented model (23) has the self-calibrated property, thus the MSE can be estimated using the formula for usual EBLUP predictors.

5. Simulation study

5.1 Simulation setup

To study the empirical properties of small area predictors described in Section 2 and 3, we use data designed to simulate a large national survey in which state estimates are of interest. Table 1 contains the approximate populations of the 50 states of the United States in the year 2000. The sample sizes m_i given in the table are approximately proportional to the square roots of the state populations.

Table 1 Population and sample sizes for the simulation

State	Population (in 1,000)	Sample size (m_i)	State	Population (in 1,000)	Sample size (m_i)
1	33,640	58	26	4,000	20
2	21,160	46	27	3,610	19
3	19,360	44	28	3,240	18
4	16,000	40	29	3,240	18
5	12,250	35	30	2,890	17
6	12,250	35	31	2,890	17
7	11,560	34	32	2,560	16
8	10,240	32	33	2,560	16
9	8,410	29	34	2,250	15
10	8,410	29	35	1,960	14
11	7,840	28	36	1,690	13
12	7,290	27	37	1,690	13
13	6,250	25	38	1,690	13
14	6,250	25	39	1,210	11
15	5,760	24	40	1,210	11
16	5,760	24	41	1,210	11
17	5,760	24	42	1,210	11
18	5,290	23	43	1,000	10
19	5,290	23	44	810	9
20	5,290	23	45	810	9
21	4,840	22	46	810	9
22	4,410	21	47	640	8
23	4,410	21	48	640	8
24	4,410	21	49	640	8
25	4,000	20	50	490	7

A total of 10,000 samples were generated. Each sample in the simulation study was composed of observations generated from model

$$Y_{ij} = \boldsymbol{x}_i'\boldsymbol{\beta} + b_i + \varepsilon_{ij}, \tag{25}$$

where $\boldsymbol{x}_i' = (1,\ z_i)$, $\boldsymbol{\beta}' = (6.0,\ 3.0)$, $z_i = \mathrm{Pop}_i^{0.2} - \overline{\mathrm{Pop}^{0.2}}$, Pop_i is the population of state i in millions, $\overline{\mathrm{Pop}^{0.2}}$ is the mean of $\mathrm{Pop}_i^{0.2}$, $b_i \sim NI(0,\ 1)$, and $\varepsilon_{ij} \sim NI(0,\ 16)$. The b_i's and e_{ij}'s are independent. The model for the state observations becomes

$$Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + b_i + e_i, \tag{26}$$

where $Y_i = m_i^{-1}\sum_{j=1}^{m_i} Y_{ij}$ and $e_i = m_i^{-1}\sum_{j=1}^{m_i}\varepsilon_{ij}$. With the sample sizes given in Table 1, the γ_i defined in (6) is 0.784 for the largest state (California) and 0.304 for the smallest state (Wyoming).

To investigate the performance of five predictors, EBLUP, Pfeffermann-Bernard (PB), Isaki-Tsay-Fuller (ITF), You-Rao (YR), and augmented model (23) (AUG2), two estimation models were used. The first model is a misspecified model with $\boldsymbol{x}_i' = 1$ in the notation of (26). This model is called model (A). Correspondingly, the data generating model (26) with $\boldsymbol{x}_i' = (1,\ z_i)$ is called model (B).

Following the method outlined in Wang and Fuller (2003), the estimator of σ_b^2 is

$$\hat{\sigma}_b^2 = \max\{0.5[\hat{V}(\tilde{\sigma}_b^2)]^{0.5},\ \tilde{\sigma}_b^2\}, \tag{27}$$

where

$$\tilde{\sigma}_b^2 = \sum_{i=1}^{50} c_i\left[\frac{50}{50-k}(Y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})^2 - \hat{\sigma}_{ei}^2\right], \tag{28}$$

$$\hat{V}(\tilde{\sigma}_b^2) = \sum_{i=1}^{50} c_i^2\left\{\left[\frac{50}{50-k}(Y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})^2 - \hat{\sigma}_{ei}^2\right] - \tilde{\sigma}_b^2\right\}^2, \tag{29}$$

$c_i = (\sum_{i=1}^{50} m_i^{0.5})^{-1} m_i^{0.5}$, $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ is the ordinary least squares estimator of the regression coefficient of Y_i on $\boldsymbol{x}_i$, k is the dimension of the vector $\boldsymbol{x}_i$, $\hat{\sigma}_{ei}^2 = m_i^{-1}s_i^2$, and $s_i^2 = (m_i - 1)^{-1}\sum_{j=1}^{m_i}(Y_{ij} - Y_i)^2$ is the sample variance of area i.

The EBLUP predictor is

$$\hat{y}_i = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} + \hat{\gamma}_i(Y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}), \tag{30}$$

where

$$\hat{\gamma}_i = (\hat{\sigma}_b^2 + \hat{\sigma}_{ei}^2)^{-1}\hat{\sigma}_b^2, \tag{31}$$

and

$$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \left[\sum_{i=1}^{n}(\sigma_b^2 + \hat{\sigma}_{ei}^2)^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\right]^{-1}\left[\sum_{i=1}^{n}(\hat{\sigma}_b^2 + \hat{\sigma}_{ei}^2)^{-1}\boldsymbol{x}_i Y_i\right] \tag{32}$$

is the generalized least squares estimator of $\boldsymbol{\beta}$. The restriction considered is $\sum_{i=1}^{50}\omega_i\hat{y}_i^{a} = \sum_{i=1}^{50}\omega_i Y_i$, where

$$\omega_i = \left(\sum_{i=1}^{50}\mathrm{Pop}_i\right)^{-1}\mathrm{Pop}_i. \tag{33}$$

With EBLUP predictor, PB and ITF are derived using (13) and (14). The YR predictor is derived use (16) with $\hat{\gamma}_i$ defined in (31) and $\hat{\boldsymbol{\beta}}_{\mathrm{YR}}$ defined in (17). The AUG2

predictor is of the form (30) with $\boldsymbol{x}_i' = (1, \omega_i \hat{\sigma}_{ei}^2)$ under the augmented model (A) and $\boldsymbol{x}_i' = (1, z_i, \omega_i \hat{\sigma}_{ei}^2)$ under the augmented model (B).

For each of the 10 predictors, the criterion

$$Q(\hat{\boldsymbol{y}}) = 0.02 \sum_{i=1}^{50} \varphi_i (\hat{y}_i - y_i)^2 \qquad (34)$$

was calculated, where $\varphi_i = (\gamma_i m_i^{-1} \sigma_e^2)^{-1}$. Note that φ_i^{-1} is the variance of the predictor of y_i constructed with known parameters.

5.2 Simulation results

The estimator of σ_b^2 constructed under model (B) has a Monte Carlo mean of 1.001 with a standard deviation of 0.386. When model (A) is used for estimation, the Monte Carlo mean of the estimator of σ_b^2 is 1.720 with a standard deviation of 0.521. The mean square of $3z_i$ is 0.636. Thus, the estimation procedure using model (A) incorporates much of the fixed area effect of model (B) into the random effect. With a bigger $\hat{\sigma}_b^2$, the $\hat{\gamma}_i$ is bigger and a higher proportion of Y_i was used to construct the predictors, which partially offsets the impact of model misspecification. Under augmented model (A), the estimator of σ_b^2 is 1.197, *i.e.*, a much smaller portion of the fixed area effect of model (B) is incorporated into the random effect.

Table 2 contains some summary statistics for predictors based on 10,000 simulated samples. The empirical bias in $\sum \omega_i (Y_i - \hat{y}_i)$ is zero for all predictors when model (B) or its augmented model is used because the prediction model matches the data-generating model. The difference $\sum \omega_i (Y_i - \hat{y}_i)$ has a simulated standard deviation of 0.022 for the usual EBLUP predictor. The standard deviations of $\sum \omega_i (Y_i - \hat{y}_i)$ are 0 for the other four predictors because the predictors satisfy the restriction $\sum_{i=1}^{50} \omega_i \hat{y}_i = \sum_{i=1}^{50} \omega_i Y_i$.

Table 2 Monte Carlo properties of small area predictors (average of 10,000 samples generated by model B)

Quantity		EBLUP	PB	ITF	YR	AUG2
		Predictor constructed under model (A)				
$\sum \omega_i (Y_i - \hat{y}_i)$	Mean	-0.100	0.000	0.000	0.000	0.000
	(SD)	(0.027)	(0.000)	(0.000)	(0.000)	(0.000)
$Q(\hat{\boldsymbol{y}})$	Mean	1.438	1.446	1.419	1.558	1.298
		Predictor constructed under model (B)				
$\sum \omega_i (Y_i - \hat{y}_i)$	Mean	-0.000	0.000	0.000	0.000	0.000
	(SD)	(0.022)	(0.000)	(0.000)	(0.000)	(0.000)
$Q(\hat{\boldsymbol{y}})$	Mean	1.203	1.202	1.202	1.208	1.219

Prediction based on model (A), or its augmented model, is biased because the data generation model (B) contains a function of the population size. The simulated mean of the weighted difference $\sum \omega_i (Y_i - \hat{y}_{iA})$ is -0.100, where $\hat{y}_{iA}$ is the EBLUP predictor. The *t*-statistic for the weighted bias is -3.70. The simulated variance of the weighted mean of the predictions is

$$V\left\{\sum_{i=1}^{50} \omega_i \hat{y}_{iA}\right\} = 0.060.$$

The estimated mean square error of the model (A) prediction of $\sum \omega_i y_i$ for data generated by model (B) is

$$\text{MSE}\left\{\sum_{i=1}^{50} \omega_i (\hat{y}_{iA} - y_i)\right\} = V\left\{\sum_{i=1}^{50} \omega_i \hat{y}_{iA}\right\} + \text{Bias}^2$$

$$= 0.060 + (-0.100)^2 = 0.070. \qquad (35)$$

The variance of $\sum \omega_i Y_i$ is

$$V\left\{\sum_{i=1}^{50} \omega_i Y_i\right\} = \sum_{i=1}^{50} \omega_i^2 (\sigma_b^2 + m_i^{-1} \sigma_e^2) = 0.0622. \qquad (36)$$

Thus, the use of $\sum \omega_i Y_i$ as the estimator of $\sum \omega_i y_i$ is about 12.5% more efficient than the predictor $\sum \omega_i \hat{y}_{iA}$ based on the model (A). Due to calibration, the MSE of the four predictors (PB, ITF, YR, and AUG2) of $\sum \omega_i y_i$ have the same MSE as the MSE of the directly estimated mean $\sum \omega_i Y_i$. The squared bias would be a much larger proportion of the mean square error if there were more small areas.

The value of criterion Q for EBLUP is 1.20 under model (B). Thus, estimation of the parameters increased the average variance of the predictors about 20% relative to the use of known parameters. If the predictions are made using the known σ_{ei}^2, the value of criterion Q for EBLUP is 1.06. Therefore, estimation of $\hat{\sigma}_{ei}^2$ contributes the most to the increase in the variability. Because the bias is zero when model (B) is used for estimation, the adjustments that the restricted predictors make are small. Consequently, the adjustment predictors give criteria values very similar to those of the unadjusted predictors. The YR predictors have slightly larger criterion values than the corresponding PB and ITF predictors because the YR predictor uses an inefficient estimator of $\boldsymbol{\beta}$. Predictors based on the augmented model have slightly larger criterion value Q because the model has a redundant variable. The less than 2% increase in Q is on the order of n^{-1}, which is the expected loss from adding an unnecessary parameter in a least square prediction.

The value of criterion Q for the EBLUP under model (A) is 1.438 compared to 1.203 for the EBLUP under model (B). This is the penalty due to the model misspecification. Among adjustment procedures based on model (A), the ITF procedure has the smallest value for Q. There is little difference among the PB, ITF, and the EBLUP predictors. The Q for the You-Rao procedure is about 8% larger than that for the EBLUP. The predictor based on the augmented

model has a Q about 11% smaller than that of the EBLUP predictor. In a sense, the augmented model is less misspecified.

The augmentation approach not only calibrates the small area predictors, it also reduces the bias at the area level when the model is misspecified. Tables 3 and 4 contain Monte Carlo properties of predictors for some selected areas. In the tables, the estimated biases are normalized by $(\gamma_i m_i^{-1} \sigma_e^2)^{0.5}$, the square root of the MSE of the BLUP predictor with known parameters, and the estimated MSE's are normalized by $\gamma_i m_i^{-1} \sigma_e^2$. When the correct model (B) is used, the Monte Carlo bias at the individual area is close to zero. See Table 3. Also there is little difference in the MSE's of the different procedures, with the augmented model having slightly (less than 2%) larger MSE. Again, Table 3 shows that the calibration process has little effect on the MSE of the area predictors in compared to the EBLUP predictor.

Table 3
Monte Carlo Properties of individual area predictors using versions of model (B) (10,000 samples generated by model B)

State	Quantity	EBLUP	PB	ITF	YR	AUG2
1	Bias	0.011	0.012	0.012	0.013	0.011
	MSE	1.100	1.101	1.105	1.104	1.120
2	Bias	0.000	0.001	0.001	0.002	0.001
	MSE	1.072	1.072	1.072	1.076	1.092
14	Bias	-0.001	-0.001	-0.001	-0.001	-0.001
	MSE	1.058	1.058	1.058	1.058	1.074
26	Bias	0.015	0.015	0.015	0.015	0.018
	MSE	1.078	1.077	1.078	1.079	1.092
38	Bias	-0.005	-0.005	-0.005	-0.006	-0.003
	MSE	1.123	1.122	1.122	1.132	1.135
50	Bias	0.012	0.012	0.012	0.009	0.014
	MSE	1.222	1.222	1.222	1.247	1.246

If a misspecified model, such as model (A) is used, the bias in the sum of the EBLUP predictors as an estimator of the total is negative for the example because the states with a negative bias have large ω_i. See Table 4. The adjustment procedures such as PB or ITF allocate the bias to all the small areas. Thus, the adjustment reduces the negative bias of area predictors with large negative bias and increases the positive bias of predictors with large positive bias. This results in a smaller MSE for larger states and a slightly larger MSE for smaller states. The YR predictor has larger bias than the ITF predictor. On the other hand, predictors constructed with $\boldsymbol{x}_i' = (1, \omega_i \hat{\sigma}_{ei}^2)$, *i.e.*, the augmented model (A), are much superior to those constructed under model (A). The bias is reduced for areas, large or small.

Table 4
Monte Carlo Properties of individual area predictors using versions of model (A) (10,000 samples generated by model B)

State	Quantity	EBLUP	PB	ITF	YR	AUG2
1	Bias	-0.597	-0.225	-0.052	-0.473	-0.030
	MSE	1.471	1.165	1.130	1.331	1.139
2	Bias	-0.496	-0.227	-0.170	-0.358	-0.070
	MSE	1.330	1.134	1.115	1.207	1.124
14	Bias	-0.121	0.004	-0.031	0.055	-0.025
	MSE	1.100	1.085	1.086	1.089	1.105
26	Bias	0.057	0.157	0.115	0.249	0.053
	MSE	1.126	1.148	1.136	1.188	1.132
38	Bias	0.380	0.453	0.406	0.601	0.202
	MSE	1.340	1.405	1.361	1.571	1.233
50	Bias	0.922	0.980	0.931	1.178	0.537
	MSE	2.196	2.316	2.215	2.767	1.577

6. Conclusions

In this paper, several calibrated predictors are reviewed. We offer a fresh look at the benchmarking restriction (7). Imposing the restriction is viewed as an adjustment problem and a criterion that unifies the derivation of calibrated predictors is presented. The criterion approach to the problem opens the door for consideration of other predictors.

Implicit in the imposition of the restriction is the possibility that the small area model is misspecified and the predictors are biased. When the model is misspecified, the calibration adjustment only adjusts for the overall bias, not for the bias at the small area level. The augmented model approach leads to a self-calibrated predictor and reduces the bias at the small area level. Also, variance estimation for the externally calibrated predictors is relatively complex, while variance estimation for self-calibrated predictors is straightforward. In summary, if the bias is a concern, use of a self-calibrated augmented model is preferred to external calibration.

7. Appendix

Proof of Theorem 1: Let $\tilde{y}_i^H$ be the BLUP of y_i and let $\hat{y}_i$ be any linear unbiased predictor of y_i. By standard results for BLUP (see, for example, Robinson (1991) and Harville (1976)), we have

$$\text{cov}(\tilde{y}_i^H - y_i,\ \hat{y}_j - \tilde{y}_j^H) = 0, \quad \text{if} \quad i \neq j, \qquad \text{(A.1)}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, n$. Let $R(\hat{\boldsymbol{y}}^a)$ denote the collection of all linear unbiased predictors that satisfy (7). For any $\hat{\boldsymbol{y}} \in R(\hat{\boldsymbol{y}}^a)$, by (A.1), we have

$$E\{(\hat{y}_i - y_i)^2\} = E\{(\tilde{y}_i^H - y_i)^2\} + E\{(\hat{y}_i - \tilde{y}_i^H)^2\}. \quad \text{(A.2)}$$

Therefore,

$$Q(\hat{\boldsymbol{y}}) = \sum_{i=1}^{n} \varphi_i E\{(\hat{y}_i - y_i)^2\}$$

$$= \sum_{i=1}^{n} \varphi_i E\{(\tilde{y}_i^H - y_i)^2\} + \sum_{i=1}^{n} \varphi_i E\{(\hat{y}_i - \tilde{y}_i^H)^2\}. \quad \text{(A.3)}$$

Since $\hat{\boldsymbol{y}}$ satisfies (7), we have $\sum_{i=1}^{n} \omega_i \hat{y}_i = \sum_{i=1}^{n} \omega_i Y_i$ and, for the $\hat{y}_i^{\breve{a}}$ defined in (19),

$$\hat{y}_i^{\breve{a}} = \tilde{y}_i^H + \breve{a}_i \left[\sum_{j=1}^{n} \omega_j (Y_j - \tilde{y}_j^H) \right]$$

$$= \tilde{y}_i^H + \breve{a}_i \left[\sum_{j=1}^{n} \omega_j (\hat{y}_j - \tilde{y}_j^H) \right].$$

By (A.1),

$$Q(\hat{y}_i^{\breve{a}}) = \sum_{i=1}^{n} \varphi_i E(\hat{y}_i^{\breve{a}} - y_i)^2$$

$$= \sum_{i=1}^{n} \varphi_i E\left\{ \left[(\tilde{y}_i^H - y_i) + \breve{a}_i \left(\sum_{j=1}^{n} \omega_j \hat{y}_j - \sum_{j=1}^{n} \omega_j \tilde{y}_j^H \right) \right]^2 \right\}$$

$$= \sum_{i=1}^{n} \varphi_i E\{(\tilde{y}_i^H - y_i)^2\} + E\left\{ \left[\sum_{j=1}^{n} \omega_j (Y_j - \tilde{y}_j^H) \right]^2 \right\} \sum_{i=1}^{n} \varphi_i \breve{a}_i^2. \quad \text{(A.4)}$$

Since $\breve{a}_i = (\sum_{i=1}^{n} \varphi_i^{-1} \omega_i^2)^{-1} \varphi_i^{-1} \omega_i$ in (A.4), we have

$$Q(\hat{y}_i^{\breve{a}}) = \sum_{i=1}^{n} \varphi_i E\{(\tilde{y}_i^H - y_i)^2\}$$

$$+ E\left\{ \left[\sum_{j=1}^{n} \omega_j (\hat{y}_j - \tilde{y}_j^H) \right]^2 \right\} \left(\sum_{i=1}^{n} \varphi_i^{-1} \omega_i^2 \right)^{-1}. \quad \text{(A.5)}$$

Note that

$$E\left\{ \left[\sum_{j=1}^{n} \omega_j (\hat{y}_j - \tilde{y}_j^H) \right]^2 \right\} \le \sum_{j=1}^{n} \sum_{k=1}^{n} \omega_j \omega_k g_j g_k$$

$$= \left(\sum_{i=1}^{n} \omega_i g_i \right)^2, \quad \text{(A.6)}$$

where $g_j = \{E[(\hat{y}_j - \tilde{y}_j^H)^2]\}^{0.5}$. By Cauchy's inequality,

$$\left(\sum_{j=1}^{n} \omega_j g_j \right)^2 \le \left(\sum_{i=1}^{n} \varphi_i^{-1} \omega_i^2 \right) \left(\sum_{i=1}^{n} \varphi_i g_i^2 \right)$$

$$= \left(\sum_{i=1}^{n} \varphi_i^{-1} \omega_i^2 \right) \left[\sum_{i=1}^{n} \varphi_i E\{(\hat{y}_j - \tilde{y}_j^H)^2\} \right]. \quad \text{(A.7)}$$

Combining (A.3), (A.5), (A.6), and (A.7), we have $Q(\hat{y}_i^{\breve{a}}) \le Q(\hat{\boldsymbol{y}})$.

To show the uniqueness of $\hat{y}_i^{\breve{a}}$, we need to check when the inequalities (A.6) and (A.7) become equalities. Inequality (A.6) becomes an equality if and only if

$$\hat{y}_j - \tilde{y}_j^H = c_j^0 + c_j^1 (\hat{y}_1 - \tilde{y}_1^H) \quad \text{(A.8)}$$

for some constants c_j^0 and c_j^1, $j = 2, ..., n$. Inequality (A.7) becomes an equality if and only if

$$\sqrt{\varphi_i^{-1} \omega_i^2} \sqrt{v_j g_j^2} - \sqrt{v_j^{-1} \omega_j^2} \sqrt{\varphi_i g_i^2} = 0, \quad \text{(A.9)}$$

or, equivalently,

$$\varphi_i^2 \omega_i^{-2} E\{(\hat{y}_i - \tilde{y}_i^H)^2\} = v_j^2 \omega_j^{-2} E\{(\hat{y}_j - \tilde{y}_j^H)^2\}. \quad \text{(A.10)}$$

Also,

$$\sum_{i=1}^{n} \omega_i \hat{y}_i = \sum_{i=1}^{n} \omega_i Y_i. \quad \text{(A.11)}$$

Combining (A.8), (A.10), and (A.11), we have that the equality holds if and only if $\hat{y}_j = \hat{y}_i^{\breve{a}}$. Thus, we have shown that $\hat{y}_i^{\breve{a}}$ is the unique linear unbiased predictor that satisfies (7) and minimizes criterion (18).

Acknowledgements

This research was funded in part by cooperative agreement 68-3A75-14 between the USDA Natural Resources Conservation Service and Iowa State University.

References

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

Harville, D.A. (1976). Extension of the Gauss-Markov Theorem to include the estimation of random effects. *Annals of Statistics*, 384-395.

Henderson, C.R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*, (Eds., W.D. Hanson and H.F. Robinson), National Academy of Sciences–National Research Council, Washington, Publication 982, 141-163.

Isaki, C.T., Tsay, J.H. and Fuller, W.A. (2000). Estimation of census adjustment factors. *Survey Methodology*, 31-42.

Mantel, H.J., Singh, A.C. and Barcau, M. (1993). Benchmarking of small area estimators. In *Proceedings of International Conference on Establishment Surveys*, American Statistical Association, Washington, DC, 920-925.

Pfeffermann, D., and Barnard, C.H. (1991). Some new estimators for small-area means with application to the assessment of Farmland values. *Journal of Business & Economic Statistics*, 73-84.

Rao, J.N.K. (2003). *Small area estimation*. New York: John Wiley & Sons, Inc.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science*, 15-51.

Wang, J., and Fuller, W.A. (2003). The mean square error of small area estimators constructed with estimated area variances. *Journal of American Statistical Association*, 716-723.

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

You, Y., and Rao, J.N.K. (2003). Pseudo hierarchical Bayes small area estimation combining unit level models and survey weights. *Journal of Statistical Planning and Inference*, 197-208.