Benchmarked Small Area Prediction

Emily J.Berg[1], Wayne A.Fuller[1], Andreea L.Erciulescu[1]
[1]Center for Survey Statistics and Methodology, Iowa State University, Department of Statistics, Ames, IA 50010

Abstract
Small area estimation often involves constructing predictions with an estimated model followed by a benchmarking step. In the benchmarking operation the predictions are modified so that weighted sums satisfy constraints. The most common constraint is the constraint that a weighted sum of the predictions is equal to the same weighted sum of the original observations. Augmented models as a method of imposing the constraint are investigated for both linear and nonlinear models. Variance estimators for benchmarked predictors are presented.

Key Words: Restricted estimation, Calibration, Mixed linear models, Best linear unbiased prediction, Self-calibrated.

1. Introduction

Small area predictors based on random models are used to improve the mean squared error (MSE) for areas whose direct estimates have large MSEs. In situations where the direct estimates are from survey samples, the random model predictors are often constrained so that the weighted sum of the predictions is equal to the same weighted sum of the direct estimates. See You and Rao (2002) and You and Rao (2003) for examples.

Two reasons are commonly given for imposing the constraint. If there are previously released estimated totals, it is desirable for the small area estimates to sum to those totals. Second, it is felt that imposing the constraint will reduce the bias associated with an imperfect model (see Pfeffermann and Barnard (1991)).

Wang, Fuller and Qu (2008) contains a review of some benchmarking methods and defines a class of benchmarking procedures that minimize a criterion. Bell, Datta, and Ghosh (2012) discuss theoretical properties of a class of benchmarked predictors for the linear model with known variances. They extend the Wang and Fuller (2008) estimators to enforce a vector of restrictions (instead of a single restriction) and derive optimal benchmarked estimators using a quadratic loss function where the weight matrix is not necessarily a diagonal matrix. Berg (2010) and Berg and Fuller (2009) consider multiple benchmarking restrictions in the context of a nonlinear model. In the application of Berg (2010) and Berg and Fuller (2009), the benchmarking restrictions are defined by the row and column sums of a two-way table. Berg (2010) applies a vector modification of the augmented model predictor of Wang, Fuller and Qu (2008) to enforce the restrictions, while Berg and Fuller (2009) use raking. Nandram and Sayit (2011) develop a Bayesian procedure to incorporate a constraint in the context of a beta-binomial model. Montanari, Ranalli, and Vcarelli (2009) impose the benchmarking restriction by adding the constraint to a penalized likelihood for a logistic mixed model.

2. Linear Model

We consider the linear small area model

$$y_i = \theta_i + e_i$$
$$\theta_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i \tag{1}$$

for $i = 1, 2, \ldots, m,$ where m is the number of small areas, θ_i is the small area mean, $\mathbf{x}_i$ is a known fixed vector, u_i is the random small area effect, and e_i is the sampling error. We assume e_i is independent of u_j for all i and j and that

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{ei}^2 & 0 \\ 0 & \sigma_{ui}^2 \end{pmatrix} \right). \tag{2}$$

To introduce prediction subject to a constraint, assume σ_{ui}^2 and σ_{ei}^2, $i = 1, 2, \ldots, m,$ are known and that $(y_i, \mathbf{x}_i)$, $i = 1, 2, \ldots, m,$ are observed. Then the best linear unbiased predictor (BLUP) of θ_i of (1) is

$$\hat{\theta}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}} + \gamma_i (y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}), \tag{3}$$

where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{y}, \tag{4}$$

$$\gamma_i = (\sigma_{ui}^2 + \sigma_{ei}^2)^{-1}\sigma_{ui}^2, \tag{5}$$

$\boldsymbol{\Sigma}_{aa}$ is a diagonal matrix with $\sigma_{ui}^2 + \sigma_{ei}^2$ as the ith diagonal element, $\mathbf{y} = (y_1, y_2, ..., y_m)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)'$. $\mathbf{X}$ is the $m \times k$ matrix with $\mathbf{x}_i'$ as ith row. The variance of the prediction error is

$$V\{\hat{\theta}_i - \theta_i\} = \gamma_i \sigma_{ei}^2 + (1 - \gamma_i)^2 \mathbf{x}_i' V\{\hat{\boldsymbol{\beta}}\}\mathbf{x}_i. \tag{6}$$

The restriction that a linear function of the predictors be equal to the same linear function of the original observations can be written

$$\sum_{i=1}^{m} \omega_i \hat{\theta}_i = \sum_{i=1}^{m} \omega_i y_i \tag{7}$$

or as

$$\sum_{i=1}^{m} \omega_i (1 - \gamma_i)(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}) = 0,$$

where ω_i are fixed coefficients. Wang, Fuller and Qu (2008) showed that the linear predictor that satisfies (7) and minimizes

$$Q(\hat{\boldsymbol{\theta}}^{\omega}) = \sum_{i=1}^{m} \phi_i E(\hat{\theta}_i^{\omega} - \theta_i)^2 \tag{8}$$

can be expressed as

$$\hat{\theta}_i^{\omega} = \hat{\theta}_i + \hat{\alpha}_i \left(\sum_{j=1}^{m} \omega_j y_j - \sum_{j=1}^{m} \omega_j \hat{\theta}_j \right), \tag{9}$$

where

$$\hat{\alpha}_i = \left(\sum_{j=1}^{m} \phi_j^{-1} \omega_j^2 \right)^{-1} \phi_i^{-1} \omega_i .$$

The predictor (9) can be written

$$\hat{\theta}_i^{\omega} = \hat{\theta}_i + \phi_i^{-1} \omega_i \hat{\beta}_{aug}, \tag{10}$$

where

$$\hat{\beta}_{aug} = \left(\sum_{j=1}^{m} \phi_j^{-1} \omega_j^2 \right)^{-1} \sum_{j=1}^{m} \omega_j \phi_j^{-1} \phi_j (y_j - \hat{\theta}_j) \tag{11}$$

and $\hat{\theta}_i$ is defined in (3). We have written (11) to illustrate that $\hat{\beta}_{aug}$ can be viewed as a generalized least squares coefficient, where $\phi_j^{-1}\omega_j$ is the explanatory variable, ϕ_j is the weight and $y_j - \hat{\theta}_j$ is the dependent variable. Battese, Harter and Fuller (1988), Pfeffermann and Barnard (1991), and Isaki, Tsay and Fuller (2000) suggested predictors that are equivalent to (10) with $\phi_i^{-1} = (\sigma_{ei}^2 + \sigma_{ui}^2)^{-1} \sigma_{ei}^2 \sigma_{ui}^2$, $\phi_i^{-1} = \omega_i^{-1} \operatorname{cov}(\hat{\theta}_i, \Sigma_{j=1}^m \omega_j \hat{\theta}_j)$, and $\phi_i^{-1} = (\sigma_{ei}^2 + \sigma_{ui}^2)$, respectively.

The predictor can also be written

$$\hat{\theta}_i^{\omega} = \mathbf{x}_i' \hat{\boldsymbol{\beta}} + z_i \hat{\beta}_{aug} + \gamma_i [y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}} - z_i \hat{\beta}_{aug}] \tag{12}$$

where

$$\hat{\beta}_{aug} = \left(\sum_{i=1}^{m} z_i \psi_i^{-1} z_i \right)^{-1} \sum_{i=1}^{m} z_i \psi_i^{-1} (y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}), \tag{13}$$

$\psi_i = \phi_i^{-1}(1-\gamma_i)^{-2}$ and $z_i = \psi_i (1-\gamma_i)\omega_i$. The alternative expression for (10) given in (12) uses a generalized least squares coefficient with weight ψ_i^{-1}, explanatory variable z_i, and dependent variable $(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}})$. The form (12) can be viewed as a predictor for an augmented model where the original vector of explanatory variables is augmented by $z_i = \psi_i (1-\gamma_i)\omega_i$. The estimated coefficient for z_i is not the usual regression coefficient because z_i arises from a constraint that is not part of the original model. The

predictor (12) is unbiased under model (1), but would be biased under a model that contained z_i as an explanatory vector.

Remark Simple ratio adjustment is not a linear predictor, but can be written as predictor (10),

$$\hat{\theta}^{\omega}_{ratio,i} = \hat{\theta}_i + \hat{\theta}_i \hat{\beta}_{aug}$$

where $\phi_i^{-1} = \hat{\theta}_i \omega_i^{-1}$ and

$$\hat{\beta}_{aug} = \left(\sum_{i=1}^{m} \hat{\theta}_i \omega_i \right)^{-1} \sum_{i=1}^{m} \omega_i (y_i - \hat{\theta}_i). \tag{14}$$

□□

If there are *r* constraints, we write the vector constraint as

$$\sum_{i=1}^{m} \boldsymbol{\omega}_i (y_i - \hat{\theta}_i) = \mathbf{0}, \tag{15}$$

where $\boldsymbol{\omega}_i' = (\omega_{1i}, \omega_{2i}, \ldots, \omega_{ri})$. The predictor that minimizes (8) subject to the vector constraint (15) is

$$\hat{\theta}_i^{\omega} = \hat{\theta}_i + \phi_i^{-1} \boldsymbol{\omega}_i' \hat{\boldsymbol{\beta}}_{aug}, \tag{16}$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{aug} &= (\mathbf{W}'\boldsymbol{\Phi}^{-1}\mathbf{W})^{-1}\mathbf{W}'(\mathbf{y} - \hat{\boldsymbol{\theta}}) \\ &= (\mathbf{W}'\boldsymbol{\Phi}^{-1}\mathbf{W})^{-1}\mathbf{W}'(\mathbf{I} - \boldsymbol{\Gamma})\hat{\mathbf{a}}, \end{aligned} \tag{17}$$

$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)'$, the *i*th element of $\hat{\mathbf{a}}$ is $\hat{a}_i = y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}$, $\mathbf{W}$ is the matrix with $\boldsymbol{\omega}_i'$ as *i*th row, $\boldsymbol{\Phi}$ is a diagonal matrix with ϕ_i as *i*th diagonal element, and $\boldsymbol{\Gamma}$ is a diagonal matrix with γ_i as the *i*th diagonal element.

The prediction error for predictor (16) under model (1) is

$$\hat{\theta}_i^{\omega} - \mathbf{x}_i'\boldsymbol{\beta} - u_i = (1 - \gamma_i)\mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \phi_i^{-1}\boldsymbol{\omega}_i'\hat{\boldsymbol{\beta}}_{aug} + \gamma_i(u_i + e_i) - u_i, \tag{18}$$

where the error in $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{a}, \tag{19}$$

and $a_i = u_i + e_i$ is *i*th element of $\mathbf{a}$. Under model (1), $E(\hat{\boldsymbol{\beta}}_{aug}) = \mathbf{0}$, and

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{aug} &= (\mathbf{W}'\boldsymbol{\Phi}^{-1}\mathbf{W})^{-1}\mathbf{W}'(\mathbf{I} - \boldsymbol{\Gamma})\hat{\mathbf{a}} \\ &= (\mathbf{W}'\boldsymbol{\Phi}^{-1}\mathbf{W})^{-1}\mathbf{W}'(\mathbf{I} - \boldsymbol{\Gamma})(\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{aa}^{-1})\mathbf{a}. \end{aligned} \tag{20}$$

It follows that $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{aug}$ are uncorrelated. Because $u_i - \gamma_i(u_i + e_i)$ is uncorrelated with $u_j + e_j$ for all *i* and *j*, the variance of $\hat{\theta}_i^{\omega} - \theta_i$ is

$$V\{\hat{\theta}_i^{\omega} - \theta_i\} = V\{\hat{\theta}_i - \theta_i\} + \phi_i^{-2}\boldsymbol{\omega}_i' V\{\hat{\boldsymbol{\beta}}_{aug}\}\boldsymbol{\omega}_i, \tag{21}$$

where

$$V\{\hat{\boldsymbol{\beta}}_{aug}\} = (\mathbf{W}'\boldsymbol{\Phi}^{-1}\mathbf{W})^{-1}\mathbf{W}'(\mathbf{I}-\boldsymbol{\Gamma})[\boldsymbol{\Sigma}_{aa} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{X})^{-1}\mathbf{X}'](\mathbf{I}-\boldsymbol{\Gamma})\mathbf{W}(\mathbf{W}'\boldsymbol{\Phi}^{-1}\mathbf{W})^{-1}. \tag{22}$$

Note that only $\mathbf{W}'\boldsymbol{\Phi}^{-1}\mathbf{W}$ of $V\{\hat{\boldsymbol{\beta}}_{aug}\}$ is a function of ϕ_i.

In the typical case σ_{ui}^2 are unknown. Retain model (1), assume $\sigma_{ui}^2 \equiv \sigma_u^2$ is unknown, assume the σ_{ei}^2 are known, and assume (e_i, u_i) is normally distributed. Let $\tilde{\sigma}_u^2$ be an estimator of σ_u^2. Then the estimated best linear unbiased predictor (EBLUP) of θ_i is

$$\tilde{\theta}_i = \mathbf{x}_i\tilde{\boldsymbol{\beta}}_i + \tilde{\gamma}_i(y_i - \mathbf{x}_i\tilde{\boldsymbol{\beta}}_i), \tag{23}$$

where

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\tilde{\boldsymbol{\Sigma}}_{aa}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\boldsymbol{\Sigma}}_{aa}^{-1}\mathbf{y},$$

$$\tilde{\gamma}_i = (\tilde{\sigma}_u^2 + \sigma_{ei}^2)^{-1}\tilde{\sigma}_u^2,$$

and the *i*th element of the diagonal matrix $\tilde{\boldsymbol{\Sigma}}_{aa}$ is $\tilde{\sigma}_u^2 + \sigma_{ei}^2$. Defining $\tilde{\boldsymbol{\beta}}_{aug}$ by (7) with $\tilde{\boldsymbol{\beta}}$ replacing $\hat{\boldsymbol{\beta}}$ and $\tilde{\phi}_i$ replacing ϕ_i, where $\tilde{\phi}_i$ may be a function of $\tilde{\sigma}_u^2$, let

$$\tilde{\theta}_i^{\omega} = \tilde{\theta}_i + \tilde{\phi}_i^{-1}\boldsymbol{\omega}_i\tilde{\boldsymbol{\beta}}_{aug}. \tag{24}$$

Assume that the covariance between $\tilde{\sigma}_u^2$ and $\tilde{\boldsymbol{\beta}}$ is $o(m^{-0.5})$. Then the covariance between $\tilde{\theta}_i$ and $\tilde{\boldsymbol{\beta}}_{aug}$ is $o(m^{-0.5})$ and an estimator of the variance of $\tilde{\theta}_i^{\omega} - \theta_i$ of (24) is

$$\hat{V}\{\tilde{\theta}_i^{\omega} - \theta_i\} = \hat{V}\{\tilde{\theta}_i - \theta_i\} + \tilde{\phi}_i^{-2}\boldsymbol{\omega}_i'\hat{V}\{\tilde{\boldsymbol{\beta}}_{aug}\}\boldsymbol{\omega}_i, \tag{25}$$

where $\hat{V}\{\tilde{\theta}_i - \theta_i\}$ is an estimator of the variance of $\tilde{\theta}_i - \theta_i$,

$$\hat{V}\{\hat{\boldsymbol{\beta}}_{aug}\} = \mathbf{M}_W(\tilde{\boldsymbol{\Sigma}}_{aa} - \mathbf{X}(\mathbf{X}'\tilde{\boldsymbol{\Sigma}}_{aa}^{-1}\mathbf{X})^{-1}\mathbf{X}')\mathbf{M}_W',$$

$$\mathbf{M}_W = (\mathbf{W}'\tilde{\boldsymbol{\Phi}}^{-1}\mathbf{W})^{-1}\mathbf{W}'(\mathbf{I}-\hat{\boldsymbol{\Gamma}}),$$

$\tilde{\boldsymbol{\Phi}} = diag(\tilde{\phi}_i)$ and $\hat{\boldsymbol{\Gamma}} = diag(\hat{\gamma}_i)$.

3. Nonlinear Models

Consider the nonlinear small area model

$$\begin{aligned} y_i &= \theta_i + e_i \\ \theta_i &= g(\mathbf{x}_i, \boldsymbol{\beta}) + u_i, \end{aligned} \tag{26}$$

where the first derivatives of the function $g(\mathbf{x}_i, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ are continuous, $E\{(e_i, u_i)\} = (0, 0)$, $E\{e_i \mid u_j\} = 0$ for all i and j, and $E\{(e_i^2, u_i^2)\} = diag\{\sigma_{ei}^2, \sigma_{ui}^2\}$. Assume that the estimator of $\boldsymbol{\beta}$ satisfies an equation of the form

$$\sum_{i=1}^{m} \frac{\partial g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} (\sigma_{ui}^2 + \sigma_{ei}^2)^{-1} [y_i - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})] = \mathbf{0},$$

where $\partial g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) / \partial \boldsymbol{\beta}$ is the partial derivative of $g(\mathbf{x}_i, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ evaluated at $\hat{\boldsymbol{\beta}}$. Assume $(\sigma_{ui}^2, \sigma_{ei}^2)$, $i = 1, 2, \ldots, m$, are known or are functions of a finite vector of parameters. By standard approximation methods, and given regularity conditions,

$$\begin{aligned} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= (\mathbf{H}'\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{H})^{-1}\mathbf{H}'\boldsymbol{\Sigma}_{aa}^{-1}\mathbf{a} + o_p(m^{-0.5}) \\ &=: \mathbf{Ma} + o_p(m^{-0.5}), \end{aligned} \tag{27}$$

where $\mathbf{H}$ is the matrix with $\mathbf{h}_i = \partial g(\mathbf{x}_i, \boldsymbol{\beta}) / \partial \boldsymbol{\beta}'$ as ith row, $\boldsymbol{\Sigma}_{aa}$ is a diagonal matrix with $\sigma_{ui}^2 + \sigma_{ei}^2$ as ith row and the ith element of $\mathbf{a}$ is $a_i = y_i - g(\mathbf{x}_i, \boldsymbol{\beta})$. Let the predictor be of the form (3) with $g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$ replacing $\mathbf{x}_i'\hat{\boldsymbol{\beta}}$,

$$\hat{\theta}_i = g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) + \gamma_i [y_i - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})]. \tag{28}$$

The approximate variance of the prediction error for known γ_i is

$$V\{\hat{\theta}_i - \theta_i\} \doteq \gamma_i \sigma_{ei}^2 + (1 - \gamma_i)^2 \mathbf{h}_i' V\{\hat{\boldsymbol{\beta}}\} \mathbf{h}_i. \tag{29}$$

One can impose the constraint (15) on the prediction by adding the term $\phi_i^{-1}\boldsymbol{\omega}_i'\hat{\boldsymbol{\beta}}_{la}$ to the predictor to obtain

$$\hat{\theta}_{la,i} = \hat{\theta}_i + \phi_i^{-1}\boldsymbol{\omega}_i'\hat{\boldsymbol{\beta}}_{la} \tag{30}$$

where $\hat{\theta}_i$ is defined in (28),

$$\hat{\boldsymbol{\beta}}_{la} = \left(\sum_{i=1}^{m} \boldsymbol{\omega}_i \phi_i^{-1} \boldsymbol{\omega}_i' \right)^{-1} \sum_{i=1}^{m} \boldsymbol{\omega}_i (y_i - \hat{\theta}_i),$$

$y_i - \hat{\theta}_i = (1-\gamma_i)\hat{a}_i$, and $\hat{a}_i = y_i - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$. The predictor (30) is easy to construct, but the predictor is not constrained to fall in the parameter space for models with restricted parameter space.

If $g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$ and $\hat{\theta}_i$ satisfy the parameter restrictions for all *i*, the original model can be replaced with the augmented model

$$y_i = g\{(\mathbf{x}_i', \mathbf{z}_i'), (\boldsymbol{\beta}, \boldsymbol{\beta}_{aug})\} + u_i + e_i. \tag{31}$$

By analogy to estimator (10), the model is estimated in two steps. The first step is the estimation of $\boldsymbol{\beta}$ by the basic estimation procedure. In the second step

$$\sum_{i=1}^{m} \psi_i^{-1}[y_i - g\{(\mathbf{x}_i', \mathbf{z}_i'), (\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_{aug})\}]^2 \tag{32}$$

is minimized with respect to $\boldsymbol{\beta}_{aug}$, where $\mathbf{z}_i$ is chosen such that

$$\boldsymbol{\omega}_i(1-\gamma_i)\psi_i = \frac{\partial g\{(\mathbf{x}_i', \mathbf{z}_i'), (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{aug})\}}{\partial \boldsymbol{\beta}_{aug}},$$

$$\sum_{i=1}^{m} \boldsymbol{\omega}_i(1-\gamma_i)[y_i - g\{(\mathbf{x}_i', \mathbf{z}_i'), (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{aug})\}] = \mathbf{0}, \tag{33}$$

and the predictor

$$\hat{\theta}_i^{\omega} = g\{(\mathbf{x}_i', \mathbf{z}_i'), (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{aug})\} + \gamma_i(y_i - g\{(\mathbf{x}_i', \mathbf{z}_i'), (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{aug})\}) \tag{34}$$

is benchmarked. By a first order Taylor expansion of $g\{(\mathbf{x}_i', \mathbf{z}_i'), (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{aug})\}$ and using the fact that the true value of $\boldsymbol{\beta}_{aug}$ is the zero vector,

$$\sum_{i=1}^{m} \boldsymbol{\omega}_i(1-\gamma_i)[y_i - g\{(\mathbf{x}_i', \mathbf{z}_i'), (\hat{\boldsymbol{\beta}}, \mathbf{0})\}] - \mathbf{h}_{aug}(\mathbf{z}_i, \mathbf{0})\hat{\boldsymbol{\beta}}_{aug}] = o_p(m^{-0.5})$$

where $\mathbf{h}_{aug,i} = \mathbf{h}_{aug}(\mathbf{z}_i, \mathbf{0}) = \partial g\{(\mathbf{x}_i', \mathbf{z}_i'), (\boldsymbol{\beta}, \mathbf{0})\}/\partial \boldsymbol{\beta}_{aug}$. Thus, an approximation for the variance of $\hat{\boldsymbol{\beta}}_{aug}$ satisfying (33) is

$$V\{\hat{\boldsymbol{\beta}}_{aug}\} \doteq (\mathbf{Z}'\boldsymbol{\Psi}\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{I} - \mathbf{HM})\boldsymbol{\Sigma}_{aa}(\mathbf{I} - \mathbf{HM})'\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Psi}\mathbf{Z})^{-1}, \tag{35}$$

where $\boldsymbol{\Psi}$ is a diagonal matrix with ψ_i as the *i*th diagonal element, $\mathbf{Z}$ is the $m \times r$ matrix with $\mathbf{z}_i'$ as the *i*th row and $\mathbf{H}$, $\boldsymbol{\Sigma}_{aa}$ and $\mathbf{M}$ are defined in (27).

Under model (26), the true value of $\boldsymbol{\beta}_{aug}$ is $\mathbf{0}$ and

$$\hat{\boldsymbol{\beta}}_{aug} = (\mathbf{H}'_{aug}\boldsymbol{\Psi}^{-1}\mathbf{H}_{aug})^{-1}\mathbf{H}'_{aug}\boldsymbol{\Psi}^{-1}(\mathbf{I}-\mathbf{HM})\mathbf{a} + o_p(m^{-0.5}),$$

where the ith row of $\mathbf{H}_{aug}$ is $\mathbf{h}'_{aug,i} = \partial g\{(\mathbf{x}'_i, \mathbf{z}'_i), (\boldsymbol{\beta}, \boldsymbol{\beta}_{aug})\}/\partial\boldsymbol{\beta}'_{aug}$ and $\hat{\boldsymbol{\beta}}_{aug}$ is the value that minimizes (32). To the level of approximation of (35), $\hat{\boldsymbol{\beta}}_{aug}$ is uncorrelated with $\hat{\boldsymbol{\beta}}$. Therefore

$$V\{\hat{\theta}_i^{\omega} - \theta_i\} = V\{\hat{\theta}_i - \theta_i\} + (1-\gamma_i)^2\mathbf{h}'_{aug,i}V\{\hat{\boldsymbol{\beta}}_{aug}\}\mathbf{h}_{aug,i} + o(m^{-1}) \qquad (36)$$

where $\hat{\theta}_i$ is defined by (28). If one has an estimator of $V\{\hat{\theta}_i - \theta_i\}$, an estimator of $V\{\hat{\theta}_i^{\omega} - \theta_i\}$ is constructed by adding an estimator of $(1-\gamma_i)^2\,\mathbf{h}'_{aug,i}V\{\hat{\boldsymbol{\beta}}_{aug}\}\mathbf{h}_{aug,i}$ to the estimator of $V\{\hat{\theta}_i - \theta_i\}$.

4. Example

4.1 Simulation Model

We present an example of estimation for a nonlinear small area model. The model was initially developed to obtain predictors of proportions using direct estimators from the Canadian Labour Force Survey (LFS) and covariates from the previous Canadian Census of Population. See Hidiroglou and Patak (2009) and Berg and Fuller (2012) for details about the data and the model.

Letting $\hat{p}_i$, $p_{C,i}$, and n_i be the direct estimator, Census proportion, and sample size, respectively, for area i, we consider the model,

$$\hat{p}_i = \theta_i + e_i, \qquad (37)$$

where

$$\theta_i = g(\mathbf{x}_i, \boldsymbol{\beta}) + u_i,$$
$$g(\mathbf{x}_i, \boldsymbol{\beta}) = [1+\exp(\beta_0 + \beta_1 x_i)]^{-1}\exp(\beta_0 + \beta_1 x_i),$$
$$x_i = \log\left[p_{C,i}(1-p_{C,i})^{-1}\right],$$

$\mathbf{x}_i = (1, x_i)'$ and

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \sim IND\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} n_i^{-1}\sigma_{ei}^2 & 0 \\ 0 & \sigma_{ui}^2 \end{pmatrix}\right).$$

The model for the variance of the random area effects is

$$\sigma_{ui}^2 = \alpha g(\mathbf{x}_i, \boldsymbol{\beta})(1 - g(\mathbf{x}_i, \boldsymbol{\beta})),$$

where α is a parameter to be estimated. The simulation is built to reflect the situation for the LFS cluster sampling design, so $n_i^{-1}\sigma_{ei}^2$ is the variance of a 2-stage cluster sample of

size n_i. To generate variables that remain in the natural parameter space for proportions, we generate θ_i from a beta distribution and generate $\hat{p}_i$ from a mixture of beta-binomial distributions with parameters chosen to give the specified first and second moments. Berg and Fuller (2012) describe the simulation procedure in detail.

We simulate proportions for ten areas, with sample sizes and parameters as in LFS. The proportions and sample sizes for the simulation are given in Table 1, where g_i denotes $g(\mathbf{x}_i, \boldsymbol{\beta})$ of (37). The MSE of the EBLUP depends on both the sample size and the proportion. Comparing results from areas with same sample size but different proportions provides information about the effect of the mean parameters. On the other hand, comparing results from areas with different sample sizes but same proportion provides information about the effect of sample size. The value of α for the simulation is 0.005, which is similar to the estimate obtained from the LFS data. The information from four sets of generated samples is used to estimate α so that the degrees of freedom estimating variance parameters is similar to that of the LFS study.

Table 1: Parameters for simulation areas

Area	1	2	3	4	5	6	7	8	9	10
g_i	0.20	0.75	0.50	0.20	0.75	0.75	0.50	0.20	0.75	0.50
Sample size	16	16	16	30	30	60	60	204	204	204

4.2 Model Estimation and Benchmarking Prediction

In practice, estimation requires an estimator of the variance of e_i. Because a direct estimator of σ_{ei}^2 may have a large variance or may be correlated with the error in the direct estimator, we consider a working model for σ_{ei}^2. Under a working assumption that the sampling variance is proportional to a multinomial variance, an estimator of the variance of $n_i^{0.5} e_i$ is,

$$\hat{\sigma}_{ei,w}^2 = \hat{c}_i \, g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}^0) \left(1 - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}^0)\right),$$

where $\hat{\boldsymbol{\beta}}^0$ is an initial estimator of $\boldsymbol{\beta}$ that does not depend on an estimate of σ_{ei}^2. Berg and Fuller (2012) gives details on the estimation of the working model.

The model (37) is nonlinear in the parameters and is estimated iteratively. Each iteration involves two steps. The first step is the estimation of $\boldsymbol{\beta}$ by minimizing the quadratic form

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{m} [\hat{p}_i - g(\mathbf{x}_i, \boldsymbol{\beta})]^2 (n_i^{-1} \hat{\sigma}_{ei,w}^2 + \hat{\sigma}_{ui}^2)^{-1}. \tag{39}$$

The second step estimates σ_{ui}^2 with estimated GLS and the residuals from (39).

Let the predictor be of the form (28) replacing y_i with $\hat{p}_i$ and using the working model estimator of γ_i, to obtain

$$\begin{aligned} \hat{\theta}_i &= g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) + \hat{\gamma}_i[\hat{p}_i - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})], \\ \hat{\gamma}_i &= (n_i^{-1}\hat{\sigma}_{ei,w}^2 + \hat{\sigma}_{ui}^2)^{-1}\hat{\sigma}_{ui}^2, \end{aligned} \tag{40}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_{ui}^2$ are the estimators of $\boldsymbol{\beta}$ the variance of u_i, respectively, obtained from the iterative estimation procedure.

A predictor $\hat{\theta}_i^{\omega}$ is benchmarked if

$$\sum_{i=1}^{10} \omega_i \hat{\theta}_i^{\omega} = \sum_{i=1}^{10} \omega_i \hat{p}_i. \tag{41}$$

We consider three different benchmarking methods:

- Ratio adjustment (raking), where $\psi_i = (1-\hat{\gamma}_i)^{-2}\omega_i^{-1}\hat{\theta}_i$;

- Augmented model (31), using the weight $\psi_i = (1-\hat{\gamma}_i)^{-2}(n_i^{-1}\hat{\sigma}_{ei}^2 + \hat{\sigma}_{ui}^2)^{-1} n_i^{-1}\hat{\sigma}_{ei}^2\hat{\sigma}_{ui}^2$, proposed by Battese, Harter and Fuller (1988) (Aug BHF), where $n_i^{-1}\hat{\sigma}_{ei}^2$ is the direct estimator of the variance of e_i.

- Augmented model (31) , where $\psi_i = (1-\hat{\gamma}_i)^{-2}\omega_i^{-1}$ (Aug ω^{-1}).

Appendix 1 describes the implementation of the augmented benchmarking procedure for the simulation.

4.3 Results

We evaluate the performance of estimators constructed under the three benchmarking methods for two simulation sets. In one set ω_i increases as n_i increases and in the other ω_i decreases as n_i increases.

The results for the set of parameters where the weights ω_i increase as the sample size increases, are given in Table 2. The ω_i are 0.01 for areas with sample size 16 and ω_i are 0.25 for areas of sample size 204. The direct estimators with large sample size have more weight in the restriction. Also, for areas with large sample sizes (and large weights), the $\hat{\gamma}_i$ of (40) is close to one, so the EBLUP is close to the direct estimator. As a consequence, the variance of the difference between the weighted sum of the direct estimators (the restriction) and the weighted sum of the EBLUPs is relatively small. Equation (36) gives the increase in MSE caused by augmentation, and the results in Table 1 and document this increase. For the set where the weight increases with the sample

size, benchmarking has a small effect on MSE; the amount added to the MSE of the EBLUP is less than 6% of the original MSE.

Table 2: Simulation MSE for predictions
Simulation set 1: ω_i increases as n_i increases

n_i	$g(x_i, \boldsymbol{\beta})$	ω_i	MSE($\hat{\theta}_i$)	Increase in MSE		
				Raking	Aug BHF	Aug ω^{-1}
16	0.20	0.01	11	0.0	0.0	0.0
16	0.50	0.01	21	0.2	0.0	0.1
204	0.20	0.25	5	0.2	0.2	0.3
204	0.50	0.25	12	0.2	0.5	0.1

□□

The results for the second set of parameters in the simulation, where the weights ω_i decrease as the sample size increases, are given in Table 3. In this set the amount added to the MSE due to benchmarking is large. For raking, the increase in MSE is roughly proportional to the true proportions, an increase about equal to twice the original MSE for n_i=16 and nearly four times the original MSE for n_i=204. For the augmented model with the weights proposed by BHF, the increase in MSE is associated with the weight, but the method gives the smallest average increase in MSE. The amount added to the MSE using the third augmented model is roughly constant, and approximately equal to the variance of the estimated coefficient $\hat{\beta}_{aug}$.The results in Table 3 illustrate that one can determine the nature of the added variance by the choice of ψ_i .

Table 3: Simulation MSE for predictions
Simulation set 2: ω_i decreases as n_i increases

n_i	$g(x_i, \boldsymbol{\beta})$	ω_i	MSE($\hat{\theta}_i$)	Increase in MSE		
				Raking	Aug BHF	Aug ω^{-1}
16	0.20	0.25	11	20	26	28
16	0.50	0.25	21	46	68	30
204	0.20	0.01	5	19	0	28
204	0.50	0.01	12	45	0	28

Table 4 contains the empirical coverages of nominal 95% prediction intervals for the predictions constructed using the weights proposed by BHF. The results are averaged across areas with the same sample size, for each set of parameters. The prediction

intervals are constructed as $\hat{\theta}_i^{\omega} \pm t_{n_i} \hat{MSE}_{i,aug}^{0.5}$, where the MSE is computed using Taylor approximations similar to (36). The MSE estimator is approximately unbiased even if the working model used for the variance of e_i differs from the true variance. (See Appendix 2 for details of the MSE estimator.) The empirical coverages for set 1 are between 93% and 95%, slightly smaller than the empirical coverages for set 2, that range between 93% and 96%. We consider the empirical coverages to be satisfactory.

Table 4: Empirical Coverages (BHF)
Simulation set 1: ω_i increases as n_i increases
Simulation set 2: ω_i decreases as n_i increases

	$n_i = 16$	$n_i = 30$	$n_i = 60$	$n_i = 204$
Set 1	0.94	0.93	0.94	0.95
Set 2	0.96	0.93	0.93	0.95

5. Conclusions

We presented an augmented approach for benchmarking nonlinear small area models, considering multiple benchmarking restrictions. In the simulation example we used a nonlinear small area model for proportions and analyzed the benchmarking effect on MSE. When the weights were inversely related to the variance, there was a small increase in MSE. When the weights were positively related to the variance, there was a large increase in MSE. We considered different weights in the objective function and the weights proposed by Battese, Harter and Fuller (1988) gave the smallest average amount added to MSE. The augmented model with $\phi_i^{-1} = \omega_i^{-1}$ gave a nearly constant increase in MSE.

Appendix 1: Augmented Model Benchmarking for the Simulation

The benchmarked predictor for area *i* based on the augmented model with weight ψ_i is of the form,

$$\hat{\theta}_i^{\omega} = \hat{\gamma}_i \hat{p}_i + (1 - \hat{\gamma}_i)\, g\{(\mathbf{x}_i', z_i), (\hat{\boldsymbol{\beta}}, \hat{\beta}_{aug})\},$$

where $z_i = \left\{ g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) \left(1 - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})\right)\right\}^{-1} \omega_i (1 - \hat{\gamma}_i)\psi_i$ and $\mathbf{x}_i = (1, x_i)'$. The $\hat{\beta}_{aug}$ is obtained through iteration as follows. Starting with $\hat{\beta}_{aug}^{(0)} = 0$, let

$$\hat{\beta}_{aug}^{(j+1)} = \hat{\beta}_{aug}^{(j)} + \left(\sum_{i=1}^{m} z_i^2 \psi_i^{-1} [g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})(1 - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}))]^2 \right)^{-1} \sum_{i=1}^{m} z_i \psi_i^{-1} [g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})(1 - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}))]\; a_i^{(j)},$$

where $a_i^{(j)} = \hat{p}_i - g\{(\mathbf{x}_i', z_i), (\hat{\boldsymbol{\beta}}, \hat{\beta}_{aug}^{(j)})\}$, and terminate the iteration when $|\hat{\beta}_{aug}^{(j)} - \hat{\beta}_{aug}^{(j+1)}|$ is less than 10^{-5}. The benchmarking restriction is satisfied because

$$\sum_{i=1}^{m} z_i [g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})(1-g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}))]\psi_i^{-1} a_i^{(j)} = \sum_{i=1}^{m} \omega_i (1-\hat{\gamma}_i) a_i^{(j)} \approx 0$$

in the last iteration.

Appendix 2: MSE Estimator for Simulation

In the approximation for the MSE in (36), $\hat{\boldsymbol{\beta}}$ and $\hat{\beta}_{aug}$ are uncorrelated with $\gamma_i(u_i + e_i) - u_i$. We define an MSE estimator to account for a correlation between ($\hat{\boldsymbol{\beta}}$, $\hat{\beta}_{aug}$) and $\gamma_i(u_i + e_i) - u_i$ that may result from misspecification of the working model for the variance of e_i. A Taylor approximation for the error in the vector of benchmarked predictors is

$$\hat{\boldsymbol{\theta}}^{\omega} - \boldsymbol{\theta} \approx \mathbf{D}_{\gamma}(\mathbf{u} + \mathbf{e}) - \mathbf{u} + \left(\mathbf{I} - \mathbf{D}_{\gamma}\right)\mathbf{L}_{aug}(\mathbf{u} + \mathbf{e}),$$

where $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}^{\omega}$ are the vectors of benchmarked predictors and true values, respectively, $\mathbf{u}$ and $\mathbf{e}$ are the vectors of area effects and sampling errors, $\mathbf{D}_{\gamma}$ is a diagonal matrix with γ_i on the diagonal, $\mathbf{I}$ is an identity matrix of dimension m, and $\mathbf{L}_{aug}$, defined below, is based on a linear approximation for $\hat{\boldsymbol{\beta}}$ and $\hat{\beta}_{aug}$. To define $\mathbf{L}_{aug}$, let $\mathbf{D}_g$ and $\boldsymbol{\Psi}$ be diagonal matrices with $g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})(1-g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}))$ and ψ_i, respectively, on the diagonal, let $\mathbf{X}'$ be the matrix with i^{th} column $\mathbf{x}_i$, and let $\mathbf{Z}$ be the column vector with i^{th} element z_i. Then,

$$\mathbf{L}_{aug} = \mathbf{D}_g(\mathbf{X}, \mathbf{Z})(\mathbf{L}'_{\beta}, \mathbf{L}'_{\delta})',$$

where

$$\mathbf{L}_{\beta} = \left(\mathbf{X}'\mathbf{D}_g\mathbf{V}_m^{-1}\mathbf{D}_g\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}_g\mathbf{V}_m^{-1},$$
$$\mathbf{L}_{\delta} = \mathbf{M}\left(\mathbf{I} - \mathbf{D}_g\mathbf{X}\mathbf{L}_{\beta}\right),$$
$$\mathbf{M} = (\mathbf{Z}'\mathbf{D}_g\,\boldsymbol{\Psi}^{-1}\mathbf{D}_g\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}_g\,\boldsymbol{\Psi}^{-1},$$

and $\mathbf{V}_m$ is a diagonal matrix with $\hat{\alpha}g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})(1-g(\mathbf{x}_i, \hat{\boldsymbol{\beta}})) + n_i^{-1}\hat{\sigma}_{ei,w}^2$ on the diagonal. An approximation for the MSE of $\hat{\boldsymbol{\theta}}^{\omega}$, the vector of benchmarked predictors constructed with the true γ_i, is

$$\mathbf{MSE}_{aug} = \alpha\mathbf{M}_{aug}\mathbf{D}_g\mathbf{M}'_{aug} + \left(\mathbf{M}_{aug} + \mathbf{I}\right)\mathbf{V}\{\boldsymbol{e}\}\left(\mathbf{M}_{aug} + \mathbf{I}\right)',$$

where $\mathbf{M}_{aug} = \mathbf{D}_{\gamma} - \mathbf{I} + \left(\mathbf{I} - \mathbf{D}_{\gamma}\right)\mathbf{L}_{aug}$. The MSE estimator is the sum of an estimator of $\mathbf{MSE}_{aug}$ and a second term to account for the variance of the estimator of $\mathbf{D}_{\gamma}$. See Berg and Fuller (2012) for the estimator of the variance of the estimator of $\mathbf{D}_{\gamma}$.

Acknowledgements

This research was funded in part by the USDA Natural Resources Conservation Service via cooperative agreement 68748211534.

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 the American Statistical Association*, 28-36.

Bell, W.R., Datta, G.S., Ghosh ,M. (2012). Benchmarking Small Area Estimators. *Biometrika*,1-35.

Berg, E. and Fuller, W.A. (2009). A small area procedure for estimating population counts. *Proceedings of the Section on Survey Research Methods. American Statistical Association*. 3811-3825.

Berg, E. (2010). A small area procedure for estimating population proportions. Unpublished Dissertation, Iowa State University.

Berg, E. and Fuller, W.A. (2012). Estimators of Error Covariance Matrices for Small Area Prediction. *Computational Statistics and Data Analysis.* 56, 2949 - 2962.

Hidiroglou, M.A. and Patak, Z. (2009). An Application of Small Area Estimation Techniques to the Canadian Labour Force Survey. *Proceedings of the Survey Methods Section*. Statistical Society of Canada.

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

Montanari, G.E., Ranalli, G.M., and Vicarelli, C. (2010). A comparison of small area estimators of counts aligned with direct higher level estimates. Scientific Meeting of the Itialian Statistical Society. http://homes.stat.unipd.it/mgri/SIS2010/Program/contributedpaper/678-1393-1-DR.pdf.

Nandram, B., and Sayit, S. (2011). A Bayesian analysis of small area probabilities under a constraint. *Survey Methodology*. 37, 137--152.

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 and Economic Statistics*, 73-84.

Wang, J., Fuller, W.A. and Qu, Y. (2008). Small area estimation under a restriction. *Survey Methodology*, 29-36.

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.