Benchmarked Small Area Prediction

Emily J. Berg¹, Wayne A. Fuller¹, Andreea L. Erciulescu¹
¹Center 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)).

2. Linear Model

We consider the linear small area model

$$y_i = \theta_i + e_i$$

(1)

$$\theta_i = x_i'\beta + u_i$$

for $i = 1, 2, \ldots, m$, where $m$ is the number of small areas, $\theta_i$ is the small area mean, $x_i$ 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

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \sim \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{ei}^2 & 0 \\ 0 & \sigma_{ui}^2 \end{pmatrix}.$$  

(2)

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

$$\hat{\theta}_i = x_i'\hat{\beta} + \gamma_i(y_i - x_i'\hat{\beta}),$$  

(3)

where

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

(4)

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

(5)

$\Sigma_{aa}$ is a diagonal matrix with $\sigma_{ui}^2 + \sigma_{ei}^2$ as the $i$th diagonal element, $y = (y_1, y_2, \ldots, y_m)'$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$. $X$ is the $m \times k$ matrix with $x_i'$ as $i$th row. The variance of the prediction error is

$$V\{\hat{\theta}_i - \theta_i\} = \gamma_i \sigma_{ei}^2 + (1 - \gamma_i)^2 x_i' V\{\hat{\beta}\} x_i.$$  

(6)

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

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

(7)

or as

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

where $\omega_i$ are fixed coefficients. Wang, Fuller and Qu (2008) showed that the linear predictor that satisfies (7) and minimizes
\[ Q(\theta^o) = \sum_{i=1}^{m} \phi_i E(\hat{\theta}_i^o - \theta_i)^2 \]  

(8)

can be expressed as

\[ \hat{\theta}_i^o = \hat{\theta}_i + \hat{\alpha}_i \left( \sum_{j=1}^{m} \omega_j y_j - \sum_{j=1}^{m} \omega_j \hat{\theta}_j \right), \]  

(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

\[ \hat{\theta}_i^o = \hat{\theta}_i + \phi_i^{-1} \omega_i \hat{\beta}_{aug}, \]  

(10)

where

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

(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, \( \phi_j \) 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 \( \phi_i^{-1} = (\sigma^2 + \sigma_u^2)^{-1} \sigma^2 \sigma_u^2, \) \( \phi_i^{-1} = \omega_i^{-1} \text{cov}(\hat{\theta}_i, \sum_{j=1}^{m} \omega_j \hat{\theta}_j), \) and \( \phi_i^{-1} = (\sigma^2 + \sigma_u^2), \) respectively.

The predictor can also be written

\[ \hat{\theta}_i^o = x_i' \hat{\beta} + z_i \hat{\beta}_{aug} + \gamma_i [y_i - x_i' \hat{\beta} - z_i \hat{\beta}_{aug}] \]  

(12)

where

\[ \hat{\beta}_{aug} = \left( \sum_{i=1}^{m} z_i \psi_i^{-1} z_i \right)^{-1} \sum_{i=1}^{m} z_i \psi_i^{-1} (y_i - x_i' \hat{\beta}), \]  

(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 \( \psi_i^{-1}, \) explanatory variable \( z_i, \) and dependent variable \( (y_i - x_i' \hat{\beta}). \) 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),

\[
\hat{\theta}_{\text{ratio},i} = \hat{\theta}_i + \hat{\theta}_i \hat{\beta}_{\text{aug}}
\]

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

\[
\hat{\beta}_{\text{aug}} = \left( \sum_{i=1}^{m} \hat{\theta}_i \omega_i \right)^{-1} \sum_{i=1}^{m} \omega_i (y_i - \hat{\theta}_i).
\]

(14)

If there are \( r \) constraints, we write the vector constraint as

\[
\sum_{i=1}^{m} \omega_i (y_i - \hat{\theta}_i) = 0,
\]

(15)

where \( \omega_i' = (\omega_{i1}, \omega_{i2}, \ldots, \omega_{ir}) \). The predictor that minimizes (8) subject to the vector constraint (15) is

\[
\hat{\theta}_i^o = \hat{\theta}_i + \phi_i^{-1} \omega_i' \hat{\beta}_{\text{aug}},
\]

(16)

where

\[
\hat{\beta}_{\text{aug}} = (W'\Phi^{-1}W)^{-1} W'(y - \hat{\theta})
\]

(17)

\[
= (W'\Phi^{-1}W)^{-1} W'(I - \Gamma) \hat{a},
\]

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

\[
\hat{\theta}_i^o - x_i' \hat{\beta} - u_i = (1 - \gamma_i)x_i' (\hat{\beta} - \beta) + \phi_i^{-1} \omega_i' \hat{\beta}_{\text{aug}} + \gamma_i (u_i + e_i) - u_i,
\]

(18)

where the error in \( \hat{\beta} \) is

\[
\hat{\beta} - \beta = (X' \Sigma_{aa}^{-1} X)^{-1} X' \Sigma_{aa}^{-1} a,
\]

(19)

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

\[
\hat{\beta}_{\text{aug}} = (W'\Phi^{-1}W)^{-1} W'(I - \Gamma) \hat{a}
\]

(20)

\[
= (W'\Phi^{-1}W)^{-1} W'(I - \Gamma)(I - X(X' \Sigma_{aa}^{-1} X)^{-1} X' \Sigma_{aa}^{-1}) a.
\]
It follows that \( \hat{\beta} \) and \( \hat{\beta}_{aug} \) 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 \( \hat{\theta}_i^o - \theta_i \) is

\[
V\{\hat{\theta}_i^o - \theta_i\} = V\{\hat{\theta}_i - \theta_i\} + \phi_i^{-2} \omega_i V\{\hat{\beta}_{aug}\} \omega_i, \tag{21}
\]

where

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

Note that only \( W'\Phi^{-1}W \) of \( V\{\hat{\beta}_{aug}\} \) is a function of \( \phi_i \).

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

\[
\tilde{\theta}_i = x_i \tilde{\beta}_i + \tilde{\gamma}_i (y_i - x_i \tilde{\beta}_i), \tag{23}
\]

where

\[
\tilde{\beta}_i = (X'\tilde{\Sigma}_{aa}^{-1}X)^{-1}X'\tilde{\Sigma}_{aa}^{-1}y, \]

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

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

\[
\tilde{\theta}_i^o = \tilde{\theta}_i + \tilde{\phi}_i^{-1} \omega_i \tilde{\beta}_{aug}. \tag{24}
\]

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

\[
\hat{V}\{\tilde{\theta}_i^o - \theta_i\} = \hat{V}\{\tilde{\theta}_i - \theta_i\} + \tilde{\phi}_i^{-2} \omega_i \hat{V}\{\tilde{\beta}_{aug}\} \omega_i, \tag{25}
\]

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

\[
\hat{V}\{\tilde{\beta}_{aug}\} = M_W (\tilde{\Sigma}_{aa} - X(X'\tilde{\Sigma}_{aa}^{-1}X)^{-1}X')M_W', \]

\[
M_W = (W'\tilde{\Phi}^{-1}W)^{-1}W'(I - \tilde{\Gamma}),
\]

\( \tilde{\Phi} = diag(\tilde{\phi}_i) \) and \( \tilde{\Gamma} = diag(\tilde{\gamma}_i) \).
3. Nonlinear Models

Consider the nonlinear small area model

\[ y_i = \theta_i + e_i \]

\[ \theta_i = g(x_i, \beta) + u_i, \]

where the first derivatives of the function \( g(x_i, \beta) \) with respect to \( \beta \) are continuous, \( E\{e_i, u_i\} = (0, 0), \ E\{e_i | u_j\} = 0 \) for all \( i \) and \( j \), and \( E\{e_i^2, u_i^2\} = \text{diag}\{\sigma_{ui}^2, \sigma_{ei}^2\} \).

Assume that the estimator of \( \beta \) satisfies an equation of the form

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

where \( \frac{\partial g(x_i, \hat{\beta})}{\partial \beta} \) is the partial derivative of \( g(x_i, \beta) \) with respect to \( \beta \) evaluated at \( \hat{\beta} \). 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,

\[ \hat{\beta} - \beta = (H'\Sigma^{-1}_{aa}H)^{-1}H'\Sigma^{-1}_{aa}a + o_p(m^{-0.5}) \]

\[ = Ma + o_p(m^{-0.5}), \quad (27) \]

where \( H \) is the matrix with \( h_i = \frac{\partial g(x_i, \beta)}{\partial \beta} \) as \( i \)th row, \( \Sigma_{aa} \) is a diagonal matrix with \( \sigma_{ui}^2 + \sigma_{ei}^2 \) as \( i \)th row and the \( i \)th element of \( a \) is \( a_i = y_i - g(x_i, \beta) \). Let the predictor be of the form (3) with \( g(x_i, \beta) \) replacing \( x_i \beta \),

\[ \hat{\theta}_i = g(x_i, \hat{\beta}) + \gamma_i[y_i - g(x_i, \hat{\beta})]. \]

The approximate variance of the prediction error for known \( \gamma_i \) is

\[ V\{\hat{\theta}_i - \theta_i\} \geq \gamma_i \sigma_{ei}^2 + (1 - \gamma_i)^2 h_i'V(\hat{\beta})h_i. \]

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

\[ \hat{\theta}_{ia} = \hat{\theta}_i + \phi_i^{-1} \omega_i \hat{\beta}_{la} \]

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

\[ \hat{\beta}_{la} = \left( \sum_{i=1}^{m} \omega_i \phi_i^{-1} \omega_i \right)^{-1} \sum_{i=1}^{m} \omega_i (y_i - \hat{\theta}_i), \]
\[ y_i - \hat{\theta}_i = (1 - \gamma_i)\hat{a}_i, \] and \[ \hat{a}_i = y_i - g(x_i, \hat{\theta}). \] 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 \( g(x_i, \hat{\theta}) \) and \( \hat{\theta}_i \) satisfy the parameter restrictions for all \( i \), the original model can be replaced with the augmented model

\[ y_i = g\{(x'_i, z'_i), (\beta, \beta_{\text{aug}})\} + u_i + e_i. \] (31)

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

\[ \sum_{i=1}^{m} \psi^{-1}_i [(y_i - g(x'_i, z'_i), (\hat{\beta}, \hat{\beta}_{\text{aug}}))]^2 \] (32)

is minimized with respect to \( \beta_{\text{aug}} \), where \( z_i \) is chosen such that

\[ \omega_i (1 - \gamma_i) \psi_i = \frac{\partial g(x'_i, z'_i), (\hat{\beta}, \hat{\beta}_{\text{aug}}))}{\partial \beta_{\text{aug}}}, \]

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

and the predictor

\[ \hat{\theta}_i^{\omega} = g(x'_i, z'_i), (\hat{\beta}, \hat{\beta}_{\text{aug}}) + \gamma_i (y_i - g(x'_i, z'_i), (\hat{\beta}, \hat{\beta}_{\text{aug}})) \] (34)

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

\[ \sum_{i=1}^{m} \omega_i (1 - \gamma_i) [(y_i - g(x'_i, z'_i), (\hat{\beta}, 0))] - h_{\text{aug}, i}(z_i, 0)\hat{\beta}_{\text{aug}}] = o_p(m^{-0.5}) \]

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

\[ V(\hat{\beta}_{\text{aug}}) \approx (\mathbf{Z}' \mathbf{Y} \mathbf{Z})^{-1} \mathbf{Z}' (I - \mathbf{HM}) \Sigma_{\text{aa}} (I - \mathbf{HM})' \mathbf{Z}(\mathbf{Z}' \mathbf{Y} \mathbf{Z})^{-1}, \] (35)

where \( \mathbf{Y} \) is a diagonal matrix with \( \psi_i \) as the \( i \)th diagonal element, \( \mathbf{Z} \) is the \( m \times r \) matrix with \( z'_i \) as the \( i \)th row and \( \mathbf{H}, \Sigma_{\text{aa}} \) and \( \mathbf{M} \) are defined in (27).
Under model (26), the true value of $\hat{\beta}_{aug}$ is 0 and

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

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

Therefore

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

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

$$\hat{p}_i = \theta_i + e_i,  \tag{37}$$

where

$$\theta_i = g(x_i, \beta) + u_i, \quad g(x_i, \beta) = [1 + \exp(\beta_0 + \beta_i x_i)]^{-1} \exp(\beta_0 + \beta_i x_i), \quad x_i = \log[p_{C,i} (1 - p_{C,i})^{-1}]$$

and $x_i = (1, x_i)'$.

The model for the variance of the random area effects is

$$\sigma^2_{ui} = \alpha g(x_i, \beta)(1 - g(x_i, \beta)),$$

where $\alpha$ is a parameter to be estimated. The simulation is built to reflect the situation for the LFS cluster sampling design, so $n_i^{-1} \sigma^2_{ei}$ 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 $\theta_i$ 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 $g(x_i, \hat{\beta})$ 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 $\alpha$ 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 $\alpha$ so that the degrees of freedom estimating variance parameters is similar to that of the LFS study.

<table>
<thead>
<tr>
<th>Table 1: Parameters for simulation areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area 1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>$g_i$ 0.20 0.75 0.50 0.20 0.75 0.75 0.50 0.20 0.75 0.50</td>
</tr>
<tr>
<td>Sample size 16 16 16 30 30 60 60 204 204 204</td>
</tr>
</tbody>
</table>

4.2 Model Estimation and Benchmarking Prediction

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

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

where $\hat{\beta}^0$ is an initial estimator of $\beta$ that does not depend on an estimate of $\sigma_{ei}^2$. 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 $\hat{\beta}$ by minimizing the quadratic form

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

The second step estimates $\sigma_{ui}^2$ 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 $\gamma_i$, to obtain

$$\hat{\theta}_i = g(x_i, \hat{\beta}) + \hat{\gamma}_i [\hat{p}_i - g(x_i, \hat{\beta})],$$

$$\hat{\gamma}_i = (n_i^{-1} \hat{\sigma}_{\epsilon i}^2 + \hat{\sigma}_{ui}^2)^{-1} \hat{\sigma}_{ui}^2,$$  \hspace{1cm} (40)

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

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

$$\sum_{i=1}^{10} \omega_i \hat{\theta}^o_i = \sum_{i=1}^{10} \omega_i \hat{p}_i.$$  \hspace{1cm} (41)

We consider three different benchmarking methods:

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

- Augmented model (31), using the weight $\psi_i = (1 - \hat{\gamma}_i)^{-2} (n_i^{-1} \hat{\sigma}_{\epsilon i}^2 + \hat{\sigma}_{ui}^2)^{-1} n_i^{-1} \hat{\sigma}_{\epsilon i}^2 \hat{\gamma}_i^2 \hat{\sigma}_{ui}^2$, proposed by Battese, Harter and Fuller (1988) (Aug BHF), where $n_i^{-1} \hat{\sigma}_{\epsilon i}^2$ 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 $\omega^{-1}$).

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 $\omega_i$ increases as $n_i$ increases and in the other $\omega_i$ decreases as $n_i$ increases.

The results for the set of parameters where the weights $\omega_i$ increase as the sample size increases, are given in Table 2. The $\omega_i$ are 0.01 for areas with sample size 16 and $\omega_i$ 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: \( \omega_i \) increases as \( n_i \) increases

<table>
<thead>
<tr>
<th>( n_i )</th>
<th>( g(x_i, \beta) )</th>
<th>( \omega_i )</th>
<th>MSE(( \hat{\theta}_i ))</th>
<th>Increase in MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Raking</td>
<td>Aug BHF</td>
</tr>
<tr>
<td>16</td>
<td>0.20</td>
<td>0.01</td>
<td>11</td>
<td>0.0</td>
</tr>
<tr>
<td>16</td>
<td>0.50</td>
<td>0.01</td>
<td>21</td>
<td>0.2</td>
</tr>
<tr>
<td>204</td>
<td>0.20</td>
<td>0.25</td>
<td>5</td>
<td>0.2</td>
</tr>
<tr>
<td>204</td>
<td>0.50</td>
<td>0.25</td>
<td>12</td>
<td>0.2</td>
</tr>
</tbody>
</table>

The results for the second set of parameters in the simulation, where the weights \( \omega_i \) 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 \( \psi_i \).

Table 3: Simulation MSE for predictions
Simulation set 2: \( \omega_i \) decreases as \( n_i \) increases

<table>
<thead>
<tr>
<th>( n_i )</th>
<th>( g(x_i, \beta) )</th>
<th>( \omega_i )</th>
<th>MSE(( \hat{\theta}_i ))</th>
<th>Increase in MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Raking</td>
<td>Aug BHF</td>
</tr>
<tr>
<td>16</td>
<td>0.20</td>
<td>0.25</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>0.50</td>
<td>0.25</td>
<td>21</td>
<td>46</td>
</tr>
<tr>
<td>204</td>
<td>0.20</td>
<td>0.01</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>204</td>
<td>0.50</td>
<td>0.01</td>
<td>12</td>
<td>45</td>
</tr>
</tbody>
</table>

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 $\hat{\theta}_i^{\omega} \pm t_{n_i} \hat{\text{MSE}}_{i,aug}^{0.5}$, 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)

<table>
<thead>
<tr>
<th>Simulation set 1: $\omega_i$ increases as $n_i$ increases</th>
<th>Simulation set 2: $\omega_i$ decreases as $n_i$ increases</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_i = 16$</td>
<td>$n_i = 30$</td>
</tr>
<tr>
<td>Set 1</td>
<td>0.94</td>
</tr>
<tr>
<td>Set 2</td>
<td>0.96</td>
</tr>
</tbody>
</table>

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 $\psi_i$ is of the form,

$$\hat{\theta}_i^{\omega} = \hat{\gamma}_i \hat{\rho}_i + (1 - \hat{\gamma}_i) g\{ (x_i', z_i), (\hat{\beta}, \hat{\beta}_i^{aug})\},$$

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

$$\hat{\beta}_i^{(j+1)} = \hat{\beta}_i^{(j)} + \left( \sum_{i=1}^m z_i \psi_i^{-1} [g(x_i, \hat{\beta})(1 - g(x_i, \hat{\beta}))] \right)^{-1} \sum_{i=1}^m z_i \psi_i^{-1} [g(x_i, \hat{\beta})(1 - g(x_i, \hat{\beta}))] a_i^{(j)},$$

where $a_i^{(j)} = \hat{\rho}_i - g\{ (x_i', z_i), (\hat{\beta}, \hat{\beta}_i^{aug(j)}) \}$, and terminate the iteration when $| \hat{\beta}_i^{(j)} - \hat{\beta}_i^{(j+1)} |$ is less than $10^{-5}$. The benchmarking restriction is satisfied because
\[
\sum_{i=1}^{m} z_i \left[ g(x_i, \hat{\beta})(1 - g(x_i, \hat{\beta})) \right] \psi_i^{-1} a_i^{(j)} = \sum_{i=1}^{m} \omega_i (1 - \hat{y}_i) a_i^{(j)} \approx 0
\]
in the last iteration.

**Appendix 2: MSE Estimator for Simulation**

In the approximation for the MSE in (36), \( \hat{\beta} \) 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 \( (\hat{\beta}, \hat{\beta}_{aug}) \) 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

\[
\hat{\Theta} - \Theta \approx D_{\gamma}(u + e) - u + (I - D_{\gamma})L_{aug}(u + e),
\]

where \( \Theta \) and \( \hat{\Theta} \) are the vectors of benchmarked predictors and true values, respectively, \( u \) and \( e \) are the vectors of area effects and sampling errors, \( D_{\gamma} \) is a diagonal matrix with \( \gamma_i \) on the diagonal, \( I \) is an identity matrix of dimension \( m \), and \( L_{aug} \), defined below, is based on a linear approximation for \( \hat{\beta} \) and \( \hat{\beta}_{aug} \). To define \( L_{aug} \), let \( D_g \) and \( \Psi \) be diagonal matrices with \( g(x_i, \hat{\beta})(1 - g(x_i, \hat{\beta})) \) and \( \psi_i \), respectively, on the diagonal, let \( X' \) be the matrix with \( i^{th} \) column \( x_i \), and let \( Z \) be the column vector with \( i^{th} \) element \( z_i \).

Then,

\[
L_{aug} = D_g(X, Z)(L'_g, L'_g)',
\]

where

\[
L'_g = (X'D_gV_m^{-1}D_gX)^{-1}X'D_gV_m^{-1},
\]

\[
L'_\delta = M(I - D_gXL'_g),
\]

\[
M = (Z'D_g\Psi^{-1}D_gZ)^{-1}Z'D_g\Psi^{-1},
\]

and \( V_m \) is a diagonal matrix with \( \hat{\sigma}^2g(x_i, \hat{\beta})(1 - g(x_i, \hat{\beta})) + n_i^{-1}\hat{\sigma}^2_{ei,w} \) on the diagonal. An approximation for the MSE of \( \hat{\Theta} \), the vector of benchmarked predictors constructed with the true \( \gamma_i \), is

\[
\text{MSE}_{aug} = \alpha M_{aug}D_gM_{aug}' + (M_{aug} + I)V(e)(M_{aug} + I)',
\]

where \( M_{aug} = D_{\gamma} - I + (I - D_{\gamma})L_{aug} \). The MSE estimator is the sum of an estimator of \( \text{MSE}_{aug} \) and a second term to account for the variance of the estimator of \( D_{\gamma} \). See Berg and Fuller (2012) for the estimator of the variance of the estimator of \( D_{\gamma} \).

**Acknowledgements**

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


