The Canadian Journal of Statistics
Vol. 46, No. 3, 2018, Pages 482–500
La revue canadienne de statistique

Benchmarked small area prediction

Emily BERG and Wayne A. FULLER

Department of Statistics, Iowa State University, Ames, IA 50010, U.S.A.

Key words and phrases: Bootstrap; empirical Bayes; Generalized linear mixed model; natural exponential family.

MSC 2010: Primary 62D05

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 predictions is equal to the same weighted sum of the original observations. Two benchmarking procedures for nonlinear models are proposed: a linear additive adjustment and a method based on an augmented model for the expectation function. Variance estimators for benchmarked predictors are presented and vetted through simulation studies. The benchmarking procedures are applied to county estimates of the proportion of area in cropland using data from the National Resources Inventory. *The Canadian Journal of Statistics* 46: 482–500; 2018 © 2018 Statistical Society of Canada

Résumé: L'estimation pour des petits domaines comporte souvent des prévisions formulées à partir d'un modèle estimé, suivi d'une étape d'ajustement. Dans cette dernière, les prévisions sont modifiées afin que leur somme pondérée réponde à des contraintes. La plus commune consiste à exiger que le résultat d'une somme pondérée des prévisions soit identique au meme calcul effectué sur les données originales. Les auteurs proposent deux méthodes d'ajustement pour les modèles non linéaires : un ajustement linéaire additif et une méthode basée sur un modèle augmenté pour la fonction d'espérance. Les auteurs présentent les estimateurs de la variance des prévisions ajustées et les évaluent avec des études de simulation. Ils utilisent des données du National Resources Inventory pour illustrer les procédures d'évaluation en estimant la proportion des superficies en culture au niveau des comtés. *La revue canadienne de statistique* 46: 482–500; 2018 © 2018 Société statistique du Canada

1. INTRODUCTION

Small area predictors based on random models are used to improve the mean squared error (MSE) when direct estimators are judged unreliable for domains of interest. 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. In the context of small area estimation, the operation of imposing the constraint is called benchmarking. Bell, Datta, & Ghosh (2012) and Pfeffermann, Sikov, & Tiller (2014) review constrained estimation procedures for small area estimation.

Imposing a benchmarking constraint on small area predictors typically reduces the model efficiency of the predictors relative to predictors constructed under a model-based optimality criterion. An important reason to impose the constraint is that forcing the small area estimates to sum to previously released estimates prevents users from obtaining inconsistent results. Statistical reasons to impose the constraint relate to robustness and design-consistency. Imposing the constraint may reduce the bias associated with an imperfect model (Pfeffermann & Barnard,

* *Author to whom correspondence may be addressed.*
E-mail: emilyb@iastate.edu

1991). Benchmarking can also lead to design-consistent estimators for aggregations of small areas. Our procedures are motivated primarily by situations in which benchmarking is an operational requirement to ensure consistency with published estimates.

Wang, Fuller, & Qu (2008) define a class of benchmarking procedures that minimize a criterion. The class encompasses the benchmarking used in Pfeffermann & Barnard (1991), Battese, Harter, & Fuller (1988), and Isaki, Tsay, & Fuller (2000). You, Rao, & Hidiroglou (2013) derive a second-order unbiased estimator of the MSE of the You & Rao (2002) self-benchmarked predictor. Bell, Datta, & Ghosh (2012) discuss theoretical properties of a class of benchmarked predictors for the linear model with known variances. They extend the Wang, Fuller, & Qu (2008) estimators to enforce a vector of restrictions (instead of a single restriction) and derive optimal benchmarked estimators using a quadratic loss function. Steorts & Ghosh (2013) derive a second-order approximation to the MSE of the benchmarked empirical Bayes (EB) predictor in the Fay & Herriot (1979) model and show that the increase in MSE due to benchmarking is the same order as the estimation error. Ugarte, Militino, & Goicoa (2009) benchmark small area estimates to a synthetic estimator. Datta et al. (2011) and Ghosh & Steorts (2013) develop a benchmarking procedure for a general hierarchical Bayesian model. Pfeffermann, Sikov, & Tiller (2014) develop a hierarchical benchmarking procedure for linear, cross-sectional time series models.

In linear benchmarking procedures, predictors may fall outside of the bounds of a restricted parameter space (Ghosh, Kubokawa, & Kawakubo, 2015). A common approach to ensure that the benchmarked predictors remain in the parameter space is to use raking (Pfeffermann, Sikov, & Tiller, 2014). In the context of a lognormal area-level model, Ghosh, Kubokawa, & Kawakubo (2015) specify a Kulback-Leibler loss function that leads to a raking adjustment. Berg (2010) and Berg & Fuller (2009) consider multiple benchmarking restrictions in the context of a nonlinear model. Nandram & Sayit (2011) incorporate the constraint in a Bayesian analysis of a beta-binomial model. Montanari, Ranalli, & Vicarelli (2010) impose the benchmarking restriction by adding the constraint to a penalized quasi-likelihood for a logistic mixed model.

We develop benchmarking procedures for a general class of nonlinear models. We consider benchmarking procedures that are simple to implement, allow MSE estimation, and provide predictions in the parameter space. In Section 2, we introduce the context, defining two classes of nonlinear models. In Section 3, we present two benchmarking procedures for nonlinear models: an additive adjustment and a procedure based on an augmented model. Section 4 contains simulations for models in which the response variable is a proportion. In Section 5, we illustrate the procedures using the proportion of cropland in Minnesota counties. Section 6 concludes with issues to consider in application.

2. NONLINEAR AREA-LEVEL MODELS

We introduce two classes of nonlinear models that encompass many models used for small area estimation. Examples of additive error models, the type of model defined in Section 2.1, include the natural exponential quadratic variance function models of Ghosh & Maiti (2004) and models satisfying the posterior linearity condition of Lahiri (1990). Generalized linear mixed models (GLMMs) are in the class of non-additive error models, defined in Section 2.2. For the additive error models, estimators are defined by first and second moments of the observations, while the estimation procedures for the non-additive error models use full distributional assumptions.

2.1. Additive Error Model

We define an additive-error area-level model as one in which the observed direct estimator, y_i, satisfies

$$E[y_i|\theta_i] = \theta_i, \quad Var\{y_i|\theta_i\} = v_1(\theta_i, \alpha_i),$$

$$\theta_i = g(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + u_i, \tag{1}$$

for $i = 1, \ldots, m$, where m is the number of small areas, θ_i is the small area mean, $\boldsymbol{x}_i$ is a known vector, u_i is a random effect, $v_1(\cdot)$ and $g(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ are known functions, $\boldsymbol{\beta}$ is an unknown parameter to estimate, and α_i is known. Additive error models can be expressed as

$$y_i = g(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + u_i + e_i,$$

where $e_i = y_i - \theta_i$. Assume $E[(e_i, u_i)^\top] = \mathbf{0}$, and

$$\text{Cov}\{(u_i, e_i)\} = \text{diag}\{\sigma_{ui}^2, \sigma_{ei}^2\},$$

where $\sigma_{ui}^2 = v_2(\boldsymbol{x}_i, \boldsymbol{\beta}, \sigma_u^2)$, v_2 is a known function, and σ_u^2 is an unknown parameter to estimate. To define the estimators of $\boldsymbol{\beta}$ and σ_u^2, we treat σ_{ei}^2 as known. In practice, σ_{ei}^2 may be obtained by smoothing a direct estimator of the design-variance of $y_i - \theta_i$, such as a jackknife estimator, using a generalized variance function (Rao & Molina 2015, p. 77; Wolter, 2007).

Define an estimator of $\boldsymbol{\beta}$ for an additive error model by $\widehat{\boldsymbol{\beta}}$, the solution of

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

where $\partial g(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}})/\partial \boldsymbol{\beta}$ is the partial derivative of $g(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ evaluated at $\widehat{\boldsymbol{\beta}}$,

$$\widehat{\sigma}_{ui}^2 = v_2(\boldsymbol{x}_i, \widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2),$$

and $\widehat{\sigma}_u^2$ is a method of moments estimator of σ_u^2 defined in the Supporting Information. (Also see Berg & Fuller, 2012, 2014 for estimation of σ_u^2.) Under standard regularity conditions,

$$\begin{aligned} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= (\boldsymbol{H}^\top \Sigma_{aa}^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^\top \Sigma_{aa}^{-1} \boldsymbol{a} + o_p(m^{-\frac{1}{2}}) \\ &=: \boldsymbol{M}\boldsymbol{a} + o_p(m^{-\frac{1}{2}}), \end{aligned}$$

where $\boldsymbol{H}$ is the matrix with $\boldsymbol{h}_i = \partial g(\boldsymbol{x}_i^\top \boldsymbol{\beta})/\partial \boldsymbol{\beta}^\top$ as the i^{th} row, Σ_{aa} is a diagonal matrix with $\sigma_{ui}^2 + \sigma_{ei}^2$ as the i^{th} diagonal element, and the i^{th} element of $\boldsymbol{a}$ is $a_i = y_i - g(\boldsymbol{x}_i^\top \boldsymbol{\beta})$. The variance of the limiting distribution of $\widehat{\boldsymbol{\beta}}$ is

$$Var\{\widehat{\boldsymbol{\beta}}\} = (\boldsymbol{H}^\top \Sigma_{aa}^{-1} \boldsymbol{H})^{-1}.$$

The minimum MSE convex combination of y_i and $g(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ is $\theta_i(\gamma_i) = y_i - (1 - \gamma_i)(y_i - g(\boldsymbol{x}_i^\top \boldsymbol{\beta}))$, where $\gamma_i = \sigma_{ui}^2/(\sigma_{ui}^2 + \sigma_{ei}^2)^{-1}$. An estimator of $\theta_i(\gamma_i)$ is

$$\widehat{\theta}_i = y_i - (1 - \widehat{\gamma}_i)(y_i - g(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}})), \tag{3}$$

where $\widehat{\gamma}_i = \widehat{\sigma}_{ui}^2 (\widehat{\sigma}_{ui}^2 + \sigma_{ei}^2)^{-1}$. Under regularity conditions and using approximation methods such as Rao (2003, p. 103), the variance of the approximate distribution of the prediction error is

$$Var\{\widehat{\theta}_i - \theta_i\} = \gamma_i \sigma_{ei}^2 + (1 - \gamma_i)^2 \boldsymbol{h}_i^\top Var\{\widehat{\boldsymbol{\beta}}\} \boldsymbol{h}_i + [\dot{v}_{2i}(\sigma_u^2)]^2 (1 - \gamma_i)^2 Var\{\widehat{\sigma}_u^2\}/(\sigma_{ui}^2 + \sigma_{ei}^2), \tag{4}$$

where $\dot{v}_{2i}(\sigma_u^2)$ is the derivative of $v_2(\boldsymbol{x}_i, \boldsymbol{\beta}, \sigma_u^2)$ with respect to σ_u^2.

2.2. Non-Additive Error Models

We use the term non-additive error model for a model in which

$$\theta_i = g_2(\boldsymbol{x}_i^\top \boldsymbol{\beta} + u_i)$$

and $g_2(\cdot)$ is nonlinear. We assume the distribution functions for u_i and $y_i|(\theta_i, \psi_i)$ are $F(u_i|\sigma_u^2)$ and $F(y_i|\theta_i, \psi_i)$, respectively. A classical non-additive error model is the area-level GLMM in which the conditional distribution for $y_i|(\theta_i, \psi_i)$ is in the exponential family, and $u_i \sim \mathrm{N}(0, \sigma_u^2)$.

An EB predictor of θ_i is

$$\widehat{\theta}_i = \widehat{\theta}_i(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2), \tag{5}$$

where

$$\widehat{\theta}_i(\boldsymbol{\beta}, \sigma_u^2) = E[\theta_i|y_i; (\boldsymbol{\beta}, \sigma_u^2)],$$

which denotes the conditional expectation of θ_i given y_i under the distribution defined by the true parameter values, and $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)$ is the maximum likelihood estimator of $(\boldsymbol{\beta}, \sigma_u^2)$. The MSE of $\widehat{\theta}_i$ of (5) is

$$\mathrm{MSE}(\widehat{\theta}_i) = E[Var\{\theta_i|y_i; (\boldsymbol{\beta}, \sigma_u^2)\}] + E[(\widehat{\theta}_i - \widehat{\theta}_i(\boldsymbol{\beta}, \sigma_u^2))^2],$$

where $Var\{\theta_i|y_i; (\boldsymbol{\beta}, \sigma_u^2)\} = O(1)$, and $E[(\widehat{\theta}_i - \widehat{\theta}_i(\boldsymbol{\beta}, \sigma_u^2))^2] = O(m^{-1})$ (Lahiri et al., 2007).

An estimator of the leading term in the MSE of $\widehat{\theta}_i$ is

$$Var\{\theta_i|y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\} = E[\theta_i^2|y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)] - \widehat{\theta}_i^2, \tag{6}$$

where numerical approximations may be needed to compute the integrals defining (6). In the approximations below, we evaluate the conditional expectations at the parameter estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}_u^2$. For replicates $r = 1, \ldots, R$, generate $u_i^{(r)} \sim F(u_i|\widehat{\sigma}_u^2)$ and let $\theta_i^{(r)} = g_2(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}} + u_i^{(r)})$. An approximation for $E[\theta_i|y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)]$ is

$$\widehat{E}[\theta_i|y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)] = \frac{R^{-1} \sum_{r=1}^R \theta_i^{(r)} f(y_i|\theta_i^{(r)}, \psi_i)}{R^{-1} \sum_{r=1}^R f(y_i|\theta_i^{(r)}, \psi_i)}, \tag{7}$$

where $f(y_i|\theta_i, \psi_i)$ is the probability density or mass function corresponding to $F(y_i|\theta_i, \psi_i)$. Likewise, an approximation for the conditional variance of θ_i, given y_i, is

$$\widehat{Var}\{\theta_i|y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\} = \sum_{r=1}^R w_i^{(r)} \{(\theta_i^{(r)}) - \widehat{E}[\theta_i|y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)]\}^2,$$

where $w_i^{(r)} = f(y_i|\theta_i^{(r)}, \psi_i)(\sum_{r=1}^R f(y_i|\theta_i^{(r)}, \psi_i))^{-1}$.

We use the parametric bootstrap to estimate $E[(\widehat{\theta}_i - \widehat{\theta}_i(\boldsymbol{\beta}, \sigma_u^2))^2]$ and to estimate the bias of $Var\{\theta_i \mid y_i, (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\}$. For the bootstrap, the $(\theta_1^{(r)}, \ldots, \theta_m^{(r)})$ and $(y_1^{(r)}, \ldots, y_m^{(r)})$ are generated for $r = 1, \ldots, R$ using $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)$ as the generating vector. Let $\widehat{\boldsymbol{\beta}}^{(r)}$ and $\widehat{\sigma}_u^{2(r)}$ denote the parameter estimates based on the r^{th} bootstrap sample. An MSE estimator that uses a multiplicative bias correction is

$$\widehat{\mathrm{MSE}}^{EB}_{i,mult} = [Var\{\theta_i \mid y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\}]^2 \left[R^{-1} \sum_{r=1}^{R} Var\{\theta_i | y_i^{(r)}; (\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)})\} \right]^{-1} + R^{-1} \sum_{r=1}^{R} (\widehat{\theta}_i^{(r)}(\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)}) - \widehat{\theta}_i^{(r)}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2))^2, \tag{8}$$

where $\widehat{\theta}_i^{(r)}(\boldsymbol{\beta}, \sigma_u^2) = E[\theta_i | y_i^{(r)}; (\boldsymbol{\beta}, \sigma_u^2)]$.

A symmetric bootstrap confidence interval is

$$\{\theta : \widehat{\theta}_i - [Var\{\theta_i \mid y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\}]^{0.5} \widehat{q}_{D(i)} \le \theta \le \widehat{\theta}_i + [Var\{\theta_i \mid y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\}]^{0.5} \widehat{q}_{D(i)}\},$$

where $\widehat{q}_{D(i)}$ satisfies

$$\frac{1}{|D(i)|} \frac{1}{R} \sum_{i \in D(i)} \sum_{r=1}^{R} I\left[\frac{|\widehat{\theta}_i^{(r)}(\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)}) - \theta_i^{(r)}|}{\sqrt{Var\{\theta_i | y_i^{(r)}; (\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)})\}}} \le \widehat{q}_{D(i)} \right] = 1 - \alpha, \tag{9}$$

the nominal confidence level is $(1 - \alpha)100\%$, and $D(i)$ is a group of areas containing area i. For instance, $D(i)$ may be a single area or a collection of areas with the same sample size as area i. We invert the confidence interval (9) and define the MSE estimator,

$$\widehat{MSE}^{EB}_{i,c} = Var\{\theta_i \mid y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\} \widehat{q}^2_{D(i)} [\Phi^{-1}(1 - \alpha/2)]^{-2}, \tag{10}$$

where $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal distribution. We call the MSE estimator (10) the inversion MSE estimator.

3. BENCHMARKING PROCEDURES FOR NONLINEAR MODELS

We propose two benchmarking procedures for the models defined in Section 2: a linear additive adjustment and a method based on an augmented model. The benchmarking procedures are applicable to both the additive error and non-additive error models of Section 2. In defining the benchmarking procedures, we use $\widehat{\theta}_i$ to denote a predictor of θ_i such as (3) or (5), based on an optimality criterion. When it is necessary to emphasize the dependence of $\widehat{\theta}_i$ on the parameter estimators, we will write

$$\widehat{\theta}_i = \widehat{\theta}_i(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2), \tag{11}$$

where $\widehat{\theta}_i(\boldsymbol{\beta}, \sigma_u^2)$ is the optimal estimator constructed with the true parameters. Let the benchmarking restriction be

$$\boldsymbol{W}^\top \widehat{\boldsymbol{\theta}}^w = \boldsymbol{W}^\top \boldsymbol{y}, \tag{12}$$

where $\boldsymbol{y} = (y_1, \ldots, y_m)'$, $\boldsymbol{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)^\top$ has full column rank, $\boldsymbol{w}_i = (w_{1i}, \ldots, w_{ri})^\top$, and the j^{th} row of $\boldsymbol{W}^\top$ defines the j^{th} restriction of a set of r restrictions.

3.1. Linear Additive Benchmarking

Wang, Fuller, & Qu (2008) show that the linear predictor for the linear model that satisfies (12) and minimizes

$$Q(\boldsymbol{t}) = \sum_{i=1}^{m} \phi_i E[(t_i - \theta_i)^2] \quad \text{with} \quad \boldsymbol{t} = (t_1, \ldots, t_m)^\top$$

can be expressed as

$$\tilde{\theta}_i^w = \tilde{\theta}_i + \phi_i^{-1} \boldsymbol{w}_i^\top \widetilde{\boldsymbol{\beta}}_{\ell a}, \tag{13}$$

where

$$\widetilde{\boldsymbol{\beta}}_{\ell a} = \left(\sum_{j=1}^m \phi_j^{-1} \boldsymbol{w}_j (\phi_j^{-1} \boldsymbol{w}_j)^\top \phi_j \right)^{-1} \sum_{j=1}^m \boldsymbol{w}_j \phi_j^{-1} \phi_j (y_j - \tilde{\theta}_j), \tag{14}$$

$\{\phi_i : i = 1, \ldots, m\}$ is a set of constants specified by the investigator, and $\tilde{\theta}_i$ is the best linear unbiased predictor (BLUP) of θ_i for the linear model. Although $\phi_j^{-1}\phi_j$ in (14) may seem awkward because $\phi_j^{-1}\phi_j$ is clearly 1, the expression (14) illustrates that $\widetilde{\boldsymbol{\beta}}_{\ell a}$ can be viewed as a generalized least squares coefficient, where $\phi_j^{-1}\boldsymbol{w}_j$ is the explanatory variable, ϕ_j is the weight, and $y_j - \tilde{\theta}_j$ is the dependent variable. An equivalent expression for $\widetilde{\boldsymbol{\beta}}_{\ell a}$ is

$$\widetilde{\boldsymbol{\beta}}_{\ell a} = \left(\sum_{i=1}^m \boldsymbol{z}_i \psi_i^{-1} \boldsymbol{z}_i^\top \right)^{-1} \sum_{i=1}^m \boldsymbol{z}_i \psi_i^{-1} (y_i - \boldsymbol{x}_i^\top \widetilde{\boldsymbol{\beta}}), \tag{15}$$

where $\psi_i = \phi_i^{-1}(1-\gamma_i)^{-2}$, $\boldsymbol{z}_i = \psi_i(1-\gamma_i)\boldsymbol{w}_i$, $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \Sigma_{aa}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \Sigma_{aa}^{-1} \boldsymbol{y}$, and $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)^\top$. The predictor (13) applies immediately to the nonlinear models of Section 2 with $\tilde{\theta}_i = \widehat{\theta}_i(\widehat{\boldsymbol{\beta}}, \sigma_u^2)$, where $\widehat{\theta}_i(\boldsymbol{\beta}, \sigma_u^2)$ is defined following (11).

In practice, an estimated predictor can be benchmarked using the procedures defined previously with σ_{ui}^2 evaluated at an estimate and permitting ϕ_i to depend on an estimate $\widehat{\sigma}_{ui}^2$. We let $\widehat{\phi}_i$, $\widehat{\theta}_{\ell a,i}$, and $\widehat{\boldsymbol{\beta}}_{\ell a}$ denote ϕ_i, $\tilde{\theta}_i^w$, and $\widetilde{\boldsymbol{\beta}}_{\ell a}$, respectively, evaluated at $\widehat{\sigma}_{ui}^2$.

The predictor (13) is not constrained to fall in the parameter space for models with restricted parameter space, but in situations we have studied, benchmarked predictors outside the parameter space are rare. Should such a situation occur, one can choose ϕ_i to place predictions in the parameter space. Assume all of the original predictions are in the interior of the parameter space. One can fix the benchmarked prediction for area i equal to the initial prediction for area i by setting $\widehat{\phi}_i^{-1} = 0$ and $\widehat{\phi}_i^{-1}\widehat{\phi}_i = 1$ in the adjustment expressions. A possible new $\widehat{\phi}_i^{-1}$ for elements with initial predictions close to the boundary is

$$\widehat{\phi}_{*i}^{-1} = 0.5(\boldsymbol{w}_i^\top \widehat{\boldsymbol{\beta}}_{\ell a})^{-1} |K_i - \widehat{\theta}_i|$$

where K_i is the boundary closest to $\widehat{\theta}_i$. The new benchmarked predictor will fall between the prediction and the boundary unless observation i is extremely important in $\widehat{\boldsymbol{\beta}}_{\ell a}$.

To estimate the variance of $\widehat{\theta}_{\ell a,i}$, we use a linearization approximation for the additive error models and the bootstrap for non-additive error models. The MSE estimator for the additive error model is defined in (A1) and (A2) of the appendix. For non-additive error models, we apply (8) or (10) with the benchmarked predictor in place of the EB predictor. See (A4) and (A5) of the Appendix for details of how we apply the bootstrap procedure to the benchmarked predictors.

Remark 1. A ratio adjustment is a common way to benchmark predictors with a restricted parameter space. Simple ratio adjustment (raking) is not a linear predictor but can be written as predictor (13), where $\phi_i^{-1} = \widehat{\theta}_i w_i^{-1}$ and

$$\widehat{\beta}_{\ell a} = \left(\sum_{i=1}^m \widehat{\theta}_i w_i \right)^{-1} \sum_{i=1}^m w_i (y_i - \widehat{\theta}_i). \tag{16}$$

An estimator of the variance of the predictor obtained from the ratio benchmarking adjustment is

$$\widehat{Var}\{\widehat{\theta}_{\ell a,i} - \theta_i\} = \widehat{Var}\{\widehat{\theta}_i - \theta_i\} + (\widehat{\theta}_i \widehat{\beta}_{\ell a})^2, \tag{17}$$

where $\widehat{\beta}_{\ell a}$ is defined in (16). Variance estimation for the ratio adjustment is discussed in more detail in the Supporting Information.

3.2. Augmented Model Benchmarking Procedure

If the mean function and $\widehat{\theta}_i$ satisfy parameter restrictions for all i, an alternative benchmarking procedure begins with the augmented model, $\boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{z}_i^\top \boldsymbol{\beta}_{aug}$, where $\boldsymbol{z}_i$ is a specified vector with the same dimension as the vector $\boldsymbol{w}_i$ defining the restrictions (12) such that $\sum_{i=1}^m \boldsymbol{z}_i(1-\gamma_i)\boldsymbol{w}_i^\top$ is invertible. We define the benchmarked predictor $\widehat{\theta}_i^{aug}$ by

$$\widehat{\theta}_i^{aug} = \widehat{\theta}_i((\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\beta}}_{aug}^\top)^\top, \widehat{\sigma}_u^2) \tag{18}$$

which is (11) evaluated at $\boldsymbol{\beta} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\beta}}_{aug}^\top)^\top$. The benchmarking restriction is achieved provided $\widehat{\boldsymbol{\beta}}_{aug}$ satisfies $S_w(\widehat{\boldsymbol{\beta}}_{aug}) = \boldsymbol{0}$, where

$$S_w(\boldsymbol{\beta}_{aug}) = \sum_{i=1}^m \boldsymbol{w}_i[y_i - \widehat{\theta}_i((\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}_{aug})^\top, \widehat{\sigma}_u^2)]. \tag{19}$$

We may take $\boldsymbol{z}_i = \phi_i^{-1}\boldsymbol{w}_i$. The root of the estimating equation 19 can be obtained using Newton-type methods. For the additive error model predictors defined in (3), the estimating equation 19 simplifies to

$$S_w(\boldsymbol{\beta}_{aug}) = \sum_{i=1}^m \boldsymbol{w}_i(1-\widehat{\gamma}_i)(y_i - g(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}} + \boldsymbol{z}_i^\top \boldsymbol{\beta}_{aug})). \tag{20}$$

For non-additive error models in which $\widehat{\theta}_i$ is an estimator of the conditional mean (5), the estimating equation 19 takes the form

$$S_w(\boldsymbol{\beta}_{aug}) = \sum_{i=1}^m \boldsymbol{w}_i(y_i - E[\theta_i|y_i; (\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}_{aug})^\top, \widehat{\sigma}_u^2]). \tag{21}$$

The form of $\boldsymbol{z}_i$ in (15) motivates alternative choices of $\boldsymbol{z}_i$ for additive error models. By analogy with (15), one choice of $\boldsymbol{z}_i$ is

$$\boldsymbol{z}_i = d_{g,i}(0)^{-1}\psi_i(1-\gamma_i)\boldsymbol{w}_i^\top, \tag{22}$$

where $d_{g,i}(0)$ is the derivative of $g(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ with respect to $\boldsymbol{x}_i^\top \boldsymbol{\beta}$ evaluated at $\widehat{\boldsymbol{\beta}}$, $\psi_i = \phi_i^{-1}(1-\gamma_i)^{-2}$, and ϕ_i is specified. For the choice of $\boldsymbol{z}_i$ in (22), the root of the estimating equation can be obtained by the following iterative sequence. Let $\widehat{\boldsymbol{\beta}}_{aug}^{(0)}$ be a specified initial value. For $j = 1, 2, \ldots$, define

$$\widehat{\boldsymbol{\beta}}_{aug}^{(j)} = \widehat{\boldsymbol{\beta}}_{aug}^{(j-1)} +$$

$$(\boldsymbol{Z}^\top \boldsymbol{D}_g(0)\Psi^{-1}\boldsymbol{D}_g(0)\boldsymbol{Z})^{-1}\sum_{i=1}^{m} z_i d_{g,i}(0)\psi_i^{-1}(y_i - g(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}} + z_i^\top\widehat{\boldsymbol{\beta}}_{aug}^{(j-1)})), \tag{23}$$

where $\boldsymbol{Z} = (z_1, \ldots, z_m)^\top$, $\boldsymbol{D}_g(0)$ is the diagonal matrix with the derivative of $g(\boldsymbol{x}_i^\top\boldsymbol{\beta})$ with respect to $\boldsymbol{x}_i^\top\boldsymbol{\beta}$ evaluated at $\widehat{\boldsymbol{\beta}}$ on the diagonal, and Ψ is a diagonal matrix with $\psi_i = \phi_i^{-1}(1-\gamma_i)^{-2}$ on the diagonal. The iteration is terminated when $||\widehat{\boldsymbol{\beta}}_{aug}^{(j)} - \widehat{\boldsymbol{\beta}}_{aug}^{(j-1)}||$ is less than a specified tolerance. In our experience, $\widehat{\boldsymbol{\beta}}_{aug}^{(0)} = \boldsymbol{0}$ can be used as a starting value. By the definition of z_i in (22), an equivalent form for the iterative sequence (23) is

$$\widehat{\boldsymbol{\beta}}_{aug}^{(j)} = \widehat{\boldsymbol{\beta}}_{aug}^{(j-1)} + (\widetilde{\boldsymbol{Z}}^\top\Psi\widetilde{\boldsymbol{Z}})^{-1}\sum_{i=1}^{m}\widetilde{z}_i(y_i - g(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}} + z_i^\top\widehat{\boldsymbol{\beta}}_{aug}^{(j-1)})), \tag{24}$$

where $\widetilde{\boldsymbol{Z}} = (\boldsymbol{I}-\Gamma)\boldsymbol{W}$, $\widetilde{z}_i = (1-\gamma_i)\boldsymbol{w}_i$, $\boldsymbol{W}$ is defined in (12), and $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_m)$. A related approach is to define z_i in two steps. First, define $\widehat{\boldsymbol{\beta}}_{aug}^{1}$ to minimize $Q_{aug}(\boldsymbol{\beta}_{aug})$, where

$$Q_{aug}(\boldsymbol{\beta}_{aug}) = \sum_{i=1}^{m}\psi_i^{-1}[y_i - g(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}} + (z_i^{(0)})^\top\boldsymbol{\beta}_{aug})]^2,$$

and $z_i^{(0)} = d_{g,i}(0)^{-1}\psi_i(1-\gamma_i)\boldsymbol{w}_i^\top$. Then, define $\widehat{\boldsymbol{\beta}}_{aug}$ through the iteration (23) with $\widehat{\boldsymbol{\beta}}_{aug}^{1}$ as a starting value.

We use Taylor approximations to obtain an estimator of the MSE of the augmented model predictors for the additive error models and use the bootstrap for the non-additive error models. The MSE estimator for the additive error model is defined in (A3) of the Appendix. For the non-additive error model, we apply (8) or (10) with the benchmarked predictor in place of the EB predictor, as described in (A4) and (A5) of the Appendix.

4. SIMULATIONS

We evaluate the properties of the estimators using two simulation models. The simulation study of Section 4.1 uses an additive error model and is designed to demonstrate the impacts of different choices of z_i in the augmented model approach. In Section 4.2, we compare the augmented benchmarking approach to the linear additive approach using the binomial-logistic model.

4.1. Additive Error Simulation Study

Our example is a simplification of the small area model that motivated the research on benchmarking. 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 & Patak (2009), Berg & Fuller (2012, 2014) for details about the data and the model.

Letting $\widehat{p}_i$, $p_{c,i}$, and n_i be the direct estimator, Census proportion, and number of primary sampling units, respectively, for area i, we consider the model

$$\begin{aligned}
\widehat{p}_i &= \theta_i + e_i,\\
\theta_i &= g(\boldsymbol{x}_i^\top\boldsymbol{\beta}) + u_i,\\
g(\boldsymbol{x}_i^\top\boldsymbol{\beta}) &= [1+\exp(\beta_0+\beta_1 x_i)]^{-1}\exp(\beta_0+\beta_1 x_i),\\
x_i &= \log[(p_{c,i})(1-p_{c,i})^{-1}] + \log[(p_{c,1})(1-p_{c,1})^{-1}],
\end{aligned} \tag{25}$$

TABLE 1: Parameters for simulation areas.

Area	1–4	5–8	9–12	13–16	17–20	21–24	25–28	29–32	33–36	37–40
x_i	0	−2.48	−1.39	0	−2.48	−2.48	−1.39	0	−2.48	−1.39
g_i	0.2	0.75	0.5	0.2	0.75	0.75	0.5	0.2	0.75	0.5
Sample size	16	16	16	30	30	60	60	204	204	204
$10\sigma_{ei}^2$	0.1	0.15	0.26	0.07	0.11	0.04	0.07	0.01	0.01	0.02

$\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$, $\boldsymbol{x}_i = (1, x_i)^\top$, and $(u_i, e_i) \sim (\mathbf{0}, \text{diag}(\sigma_{ui}^2, n_i^{-1}\sigma_{ei}^2))$. The model for the variance of the random effects is

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

where the definition of x_i is the interaction in a saturated loglinear model fit to the Census counts, as in Berg & Fuller (2012), and σ_u^2 is a parameter to be estimated. The simulation is built to reflect 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 $\widehat{p}_i$ from a mixture of beta-binomial distributions with parameters chosen to give the specified first and second moments. Berg & Fuller (2012) describes the simulation procedure in detail.

We simulate proportions for 40 areas, with sample sizes and parameters as in LFS. The areas are divided into 10 groups, each of size four, where each area in a group has the same value for x_i, n_i, and σ_{ei}^2. The proportions and sample sizes for the simulation are given in Table 1, where g_i denotes $g(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ of (25). The MSE of the EBLUP depends on both the sample size and the proportion. Comparing results from areas with the 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 the same proportion provides information about the effect of sample sizes. The value of σ_u^2 for the simulation is 0.005, which is similar to the estimate obtained from the LFS data.

4.1.1. Model estimation and benchmarking prediction

We use 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 $\sqrt{n_i}e_i$ is

$$\widehat{\sigma}_{ei,w}^2 = \widehat{c}_{h(i)} g(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{(0)})(1 - g(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{(0)})),$$

where $\widehat{\boldsymbol{\beta}}^{(0)}$ is the MLE under an assumption that $n_i\widehat{p}_i$ has a binomial distribution with $\sigma_u^2 = 0$, $h(i)$ denotes the index set of the group of four containing area i,

$$\widehat{c}_{h(i)} = \tilde{\delta}_i \frac{\sum_{k\in h(i)} \widehat{\sigma}_{ek}^2}{\sum_{k\in h(i)} g(\boldsymbol{x}_k^\top \widehat{\boldsymbol{\beta}}^{(0)})(1 - g(\boldsymbol{x}_k^\top \widehat{\boldsymbol{\beta}}^{(0)}))} + (1 - \tilde{\delta}_i)1,$$

$$\tilde{\delta}_i = \frac{10000}{10000 + \widehat{Var}\{\widehat{c}_{h(i)}\}},$$

$$\widehat{Var}\{\widehat{c}_{h(i)}\} = \frac{\sum_{k\in h(i)} g(\boldsymbol{x}_k^\top \widehat{\boldsymbol{\beta}}^{(0)})(1-g(\boldsymbol{x}_k^\top \widehat{\boldsymbol{\beta}}^{(0)}))\Big(\frac{\widehat{\sigma}^2_{ek}}{g(\boldsymbol{x}_k^\top \widehat{\boldsymbol{\beta}}^{(0)})(1-g(\boldsymbol{x}_k^\top \widehat{\boldsymbol{\beta}}^{(0)})} - \widehat{c}_{h(i)}\Big)^2}{(|h(i)|-1)(\sum_{k\in h(i)} g(\boldsymbol{x}_k^\top \widehat{\boldsymbol{\beta}}^{(0)})(1-g(\boldsymbol{x}_k^\top \widehat{\boldsymbol{\beta}}^{(0)})))},$$

and $n_i^{-1}\widehat{\sigma}^2_{ei}$ is the direct estimator of the variance of e_i. Berg & Fuller (2012) provides further detail on the sampling variances for the simulation. The choice of $\tilde{\delta}_i$ is motivated by a Bayes estimator, where the mean of $\widehat{c}_{h(i)}$ has a prior distribution that is normal with a prior mean of 1 and a prior variance of 10,000. If $\widehat{\sigma}^2_{ek} = 0$ for all $k \in h(i)$, then $\widehat{c}_{h(i)} = 1$.

Let the predictor be of the form (3), replace y_i with $\widehat{p}_i$, and use the working model estimator of γ_i to obtain

$$\widehat{\theta}_i = \widehat{p}_i - (1-\widehat{\gamma}_i)[\widehat{p}_i - g(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}})],$$

where $\widehat{\gamma}_i = (n_i^{-1}\widehat{\sigma}^2_{ei,w} + \widehat{\sigma}^2_{ui})^{-1}\widehat{\sigma}^2_{ui}$, and $\widehat{\boldsymbol{\beta}}$ is defined by (2) with $\widehat{\sigma}^2_{ui}$ obtained by application of the procedure described in Section 1 of the Supporting Information.

The proportions have a structure of four 2×10 tables with row 1 of table k containing $\{\widehat{p}_i : i \bmod 4 = k\}$ and the column for element i containing $(\widehat{p}_i, 1-\widehat{p}_i)^\top$. We consider a multivariate benchmarking restriction that preserves row and column margins of each two-way table. A predictor $\widehat{\theta}_i^w$ is benchmarked if

$$\sum_{i=1}^{40} w_i \widehat{\theta}_i^w \delta_{ki} = \sum_{i=1}^{40} w_i \widehat{p}_i \delta_{ki},$$

where $\delta_{ki} = I(i \bmod 4 = k)$ for $k = 0, 1, 2, 3$. In this notation, w_i and $\sum_{i=1}^{40} w_i \widehat{p}_i \delta_{ki}$ are, respectively, column and row margins of the two-way table. We consider three benchmarked estimators that guarantee $0 \le \widehat{\theta}_i^w \le 1$. The first two benchmarking procedures use the augmented model benchmarking method defined in (22)–(23) of Section 3.2 with

$$\boldsymbol{z}_i = \frac{\psi_i(1-\gamma_i)w_i(\delta_{1i}, \ldots, \delta_{4i})^\top}{g(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}})(1-g(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}))},$$

and two different values for ψ_i defined by

$$\psi_i = (1-\widehat{\gamma}_i)^{-2}(n_i^{-1}\widehat{\sigma}^2_{ei} + \widehat{\sigma}^2_{ui})^{-1} n_i^{-1}\widehat{\sigma}^2_{ei}\widehat{\sigma}^2_{ui} \tag{26}$$

and

$$\psi_i = (1-\widehat{\gamma}_i)^{-2} w_i^{-1}, \tag{27}$$

where $n_i^{-1}\widehat{\sigma}^2_{ei}$ is a design-consistent estimator of the variance of e_i. The ψ_i in (26) is that proposed by Battese, Harter, & Fuller (1988), and we refer to the ψ_i in (26) as ''Aug BHF.'' We refer to ψ_i in (27) as ''Aug w^{-1}.'' We compare the augmented model procedures of Section 3.2 to benchmarking based on iterative proportional fitting (raking). As discussed in Berg (2010, p. 58), the raked predictor $\widehat{\theta}_i^w w_i$ for a single table of proportions with indexes $i, i+4, \ldots, i+9$ is approximately a linear additive predictor with

$$\widehat{\boldsymbol{\beta}}^{rak}_{\ell a} = (\boldsymbol{X}^\top \boldsymbol{D}_\theta \boldsymbol{X})^- \boldsymbol{X}^\top (\widetilde{\boldsymbol{y}} - \widehat{\boldsymbol{\theta}}), \tag{28}$$

TABLE 2: Simulation MSE for predictions.

						Increase in MSE		
Set	n_i	g_i	w_i	$Var\{e_i\}$	MSE($\widehat{\theta}_i$)	Raking	Aug BHF	Aug w^{-1}
1	16	0.01	0.20	100	18	−0.1	−0.0	−0.1
1	16	0.01	0.50	265	21	−0.1	−0.0	−0.2
1	204	0.25	0.20	8	6	0.1	0.1	0.2
1	204	0.25	0.50	21	12	0.3	0.5	0.2
2	16	0.25	0.20	100	18	19	27	28
2	16	0.25	0.50	265	21	46	68	29
2	204	0.01	0.20	8	6	18	0.0	28
2	204	0.01	0.50	21	12	42	0.0	24

For set 1, w_i increases as n_i increases. For set 2, w_i decreases as n_i increases. All MSEs and variances have been multiplied by 10,000.

where $\boldsymbol{X} = (\boldsymbol{J}_{10} \otimes \boldsymbol{I}_2, \boldsymbol{I}_{10} \otimes \boldsymbol{J}_2)$, $\boldsymbol{D}_\theta = \text{diag}(w_i\widehat{\theta}_i, w_i(1-\widehat{\theta}_i), \ldots, w_{i+9}\widehat{\theta}_{i+9}, w_{i+9}(1-\widehat{\theta}_{i+9}))$, $\boldsymbol{D}_w = \text{diag}(w_i\boldsymbol{I}_2, \ldots, w_{i+9}\boldsymbol{I}_2)$, $\widetilde{\boldsymbol{y}} = (\widehat{p}_i w_i, (1-\widehat{p}_i)w_i, \ldots, \widehat{p}_{i+9}w_{i+9}, (1-\widehat{p}_{i+9})w_{i+9})^\top$, $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_i w_i, (1-\widehat{\theta}_i)w_i, \ldots, \widehat{\theta}_{i+9}w_{i+9}, (1-\widehat{\theta}_{i+9})w_{i+9})^\top$, $\boldsymbol{I}_r$ is the identity matrix of dimension r, and $\boldsymbol{J}_r$ is the $r-$dimensional vector of ones.

4.1.2. Comparison of efficiencies of benchmarked predictors

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

The results for the set of parameters where the weights w_i increase as the sample size increases are given in the rows of Table 2 corresponding to set 1. The w_i are 0.01 for areas with sample size 16 and w_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), $\widehat{\gamma}_i$ is close to one, so $\widehat{\theta}_i$ 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 $\{\widehat{\theta}_i : i = 1, \ldots, m\}$ is relatively small. 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 5% of the original MSE.

The results for the simulation where the weights decrease as the sample size increases are given in the rows of Table 2 corresponding to set 2. In this set, the amount added to the MSE due to benchmarking is large. For raking, the increase in MSE is about equal to the original MSE for $n_i = 16$ and more than twice the original MSE for $n_i = 204$. For set 2, the increase in MSE due to raking is approximately proportional to g_i ($20/0.2 \approx 45/0.5 \approx 100$). The approximation for the raked predictor is in the scale of totals. For proportions, the increase in MSE due to raking is approximately $\boldsymbol{L}_p\boldsymbol{D}_\theta\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{rak}_{\ell a}(\widehat{\boldsymbol{\beta}}^{rak}_{\ell a})^\top\boldsymbol{X}^\top\boldsymbol{D}_\theta\boldsymbol{L}_p^\top$, where $\boldsymbol{L}_p$ is the block-diagonal matrix defining the linear approximation that converts totals to proportions and has $w_k^{-1}((1-\widehat{p}_i, -(1-\widehat{p}_i))^\top, (-\widehat{p}_i, \widehat{p}_i)^\top)^\top$ as block i. For the augmented model with the weights proposed by BHF, the increase in MSE is associated with the weight, and 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.

TABLE 3: Empirical coverages of normal theory 95% confidence intervals.

Method	Set	$n_i = 16$	$n_i = 30$	$n_i = 60$	$n_i = 204$
Aug BHF	1	0.95	0.94	0.95	0.96
Aug BHF	2	0.95	0.95	0.95	0.96
Aug w^{-1}	1	0.95	0.95	0.95	0.96
Aug w^{-1}	2	0.96	0.95	0.96	0.95
Raking	1	0.95	0.95	0.95	0.96
Raking	2	0.97	0.96	0.97	0.98

Simulation set 1: w_i increases as n_i increases. Simulation set 2: w_i decreases as n_i increases.

The results in Table 2 illustrate that the nature of the added variance can be determined by the choice of ψ_i.

To examine the effect of the size of σ_u^2 on MSE, we conducted simulations with $\sigma_u^2 = 0.06$ as well as $\sigma_u^2 = 0.005$. For both values of σ_u^2, the impact of benchmarking on the MSE is greater for the w_i inversely related to n_i than for w_i directly related to n_i. For large n_i and the w_i used for set 2, the raking and Aug w^{-1} benchmarking procedures substantially increase the MSE relative to the MSE of $\widehat{\theta}_i$, while the BHF benchmarking procedure has negligible impact on the MSE, regardless of the size of σ_u^2. The change in MSE due to benchmarking is typically smaller for $\sigma_u^2 = 0.06$ than for $\sigma_u^2 = 0.005$, which is expected because the values of $\widehat{\theta}_i$ are closer to the direct estimators for $\sigma_u^2 = 0.06$ than for $\sigma_u^2 = 0.005$. The full simulation output is provided in the Supporting Information.

4.1.3. Evaluation of MSE estimators for benchmarked predictors

Table 3 contains the empirical coverages of nominal 95% prediction intervals for predictions for the three benchmarking procedures. The results are averaged across areas with the same sample size, for each set of parameters. The prediction intervals are constructed as $\widehat{\theta}_i^w \pm t_{n_i}(0.975)\widehat{MSE}_{i,w}^{0.5}$, where $\widehat{\theta}_i^w$ denotes a benchmarked predictor and $\widehat{MSE}_{i,w}$ the corresponding MSE estimator. The MSE estimator $\widehat{MSE}_{i,w}$ is computed using (A3) for the augmented model procedures, with a modification to account for the use of the working model for σ_{ei}^2. (Section 3 of the Supporting Information outlines the modification to the variance estimator for the use of the working model. Also see Berg & Fuller, 2012, 2014.) The MSE for raking is based on the approximation (28) for the raked predictor as a linear additive predictor. The empirical coverages for set 1 are between 94% and 96%, slightly smaller than the empirical coverages for set 2, which range between 95% and 98%. We consider the empirical coverages to be satisfactory.

4.2. Non-Additive Error Simulation Study

Assume a binomial random variable for area i,

$$y_i \sim \text{Binomial}(n_i, \theta_i),$$

where

$$\theta_i = \frac{\exp(\beta + u_i)}{1 + \exp(\beta + u_i)}, \quad u_i \sim \text{N}(0, \sigma_u^2),$$

TABLE 4: MC MSEs of alternative predictors.

n_i	Bayes	EB	Lin. Add.	Aug
2	1.490	1.587	1.587	1.585
6	1.115	1.187	1.187	1.184
18	0.644	0.679	0.679	0.681

Bayes = Bayes predictor constructed with true parameters; EB = empirical Bayes predictor constructed with estimated parameters; Lin. Add. = benchmarked predictor based on the linear additive approach; Aug = benchmarked predictor based on the augmented model approach.

and $i = 1, \ldots, m$. The simulation parameters are $\beta = -1.2$, and $\sigma_u^2 = 0.8$. The sample sizes n_i are 2 for 20 areas, 6 for 20 areas, and 18 for the remaining 20 areas. An estimator of the minimum MSE predictor is

$$\widehat{\theta}_i(\widehat{\beta}, \widehat{\sigma}_u^2) = E[\theta_i \mid y_i; (\widehat{\beta}, \widehat{\sigma}_u^2)],$$

where $\widehat{\beta}$ and $\widehat{\sigma}_u^2$ are approximate MLEs obtained from Gauss-Hermite quadrature with 10 points (nAGQ = 10 in the R function glmer), and $E[\theta_i \mid y_i; (\widehat{\beta}, \widehat{\sigma}_u^2)]$ is evaluated using the numerical approximation defined in (7) with $R = 500$. We use a lower bound of 0.006 for the estimator of σ_u^2, as proposed in Erciulescu (2015).

The benchmarking restriction is

$$\sum_{i=1}^{m} n_i \widehat{\theta}_i^w = \sum_{i=1}^{m} n_i y_i.$$

We benchmark the $\widehat{\theta}_i$ using the linear additive approach of (13) and the augmented model approach of (18). For the linear additive approach, we use

$$\widehat{\phi}_{i,1}^{-1} = R^{-1} \sum_{r=1}^{R} \theta_i^{(r)}(1 - \theta_i^{(r)}) n_i^{-1} + (R-1)^{-1} \sum_{r=1}^{R} (\theta_i^{(r)} - \overline{\theta}^{(\cdot)})^2,$$

where $\theta_i^{(r)} = (1 + \exp(\widehat{\beta} + u_i^{(r)})^{-1} \exp(\widehat{\beta} + u_i^{(r)})$, $u_i^{(r)} \sim \mathrm{N}(0, \widehat{\sigma}_u^2)$, $\overline{\theta}^{(\cdot)} = R^{-1} \sum_{r=1}^{R} \theta_i^{(r)}$, and $R = 500$. For the augmented model approach, we use $z_i = \phi_{i,1}^{-1}$ in (21). To ensure that the benchmarking restriction holds, the same random numbers used to evaluate the conditional means in the estimating equation 19 are used to compute the augmented model predictor (18).

4.2.1. Comparison of MC MSEs of predictors for logistic-normal simulation

Table 4 contains the MC MSEs of the alternative predictors. The Bayes predictor, which is the conditional expectation constructed with the true parameter values, is included to demonstrate the impact of estimating the parameters on the MSE. Benchmarking has little effect on the MSE, and because the parameters β and σ_u^2 are estimated, benchmarking can decrease the MSE relative to the MSE of the EB predictor.

4.2.2. Evaluation of MSE estimators for logistic-normal simulation

We construct MSE estimators and confidence intervals for each predictor. The MSE estimators for the EB predictor are defined in (8) and (10). The MSE estimators for the benchmarked

TABLE 5: MC means of estimated MSEs and empirical coverages of symmetric 95% confidence intervals.

	Area sample size	MC Means of MSE Estimators		Empirical Coverages of 95% CI's	
Predictor	n_i	Mult.	Inversion	Mult.	Inversion
EB	2	1.610	2.633	92.5	95.3
EB	6	1.198	1.768	93.5	95.2
EB	18	0.675	1.054	94.6	96.1
Lin. Add.	2	1.700	2.093	93.0	96.0
Lin. Add.	6	1.323	1.484	94.0	96.6
Lin. Add.	18	0.747	0.850	94.6	97.3
Aug.	2	1.699	2.091	93.0	96.2
Aug.	6	1.321	1.484	93.9	96.5
Aug.	18	0.746	0.850	94.6	97.3

Definitions of EB, Lin. Add., and Aug. are as in Table 4. Mult. and Inversion, respectively, are MSE estimators defined in (8) and (10) for the EB predictor and in (A4) and (A5) for the benchmarked predictors.

predictors are defined in (A4) and (A5). In defining the inversion MSE estimator (10 or A5), we let $D(i)$ be the collection of areas with the same n_i. In the tables below, the label ''Mult.'' refers to the MSE estimators (8) and (A4). The label ''CI'' refers to the MSE estimators (10) and (A5).

Table 5 contains MC means of estimated MSEs and empirical coverages of 95% confidence intervals. The confidence intervals based on the inversion MSE estimators, (10) and (A5), have empirical coverages that are greater than or equal to the nominal level. The coverages of the confidence intervals based on the multiplicative MSE estimators, (8) and (A4), are between 92.5% and 94.6%. The positive bias of the inversion MSE estimator is due to a few extreme MSE estimates.

5. COUNTY ESTIMATION IN THE NATIONAL RESOURCES INVENTORY

We apply the benchmarking methods to data from the National Resources Inventory, a longitudinal survey that monitors status and trends in land cover and land use in the United States (Nusser & Goebel, 1997). One of the parameters of interest in the NRI is the proportion of area devoted to cropland. Interest in sub-state estimates motivated this work. Because the state-level estimates have been published, benchmarking to the state-level estimates is important.

We consider estimating the proportion of cropped acres in each county in Minnesota in 2010. We choose Minnesota because it has a high proportion of cultivated land and also a wide range in the proportions. The estimated proportion of cropland in Minnesota based on the 2010 NRI is 0.389 with an estimated CV of 2.3%.

As a first step, we compute NRI estimators of the proportion of each county in cropland and corresponding estimators of variances. To define the direct estimators at the county level, we use i to index counties ($i = 1, \ldots, 87$) and j to index sampled locations within counties ($j = 1, \ldots, n_i$). The NRI estimator of the proportion is

$$\widehat{p}_i = \frac{\sum_{j=1}^{n_i} w_{ij}\delta_{ji}}{\sum_{j=1}^{n_i} w_{ij}},$$

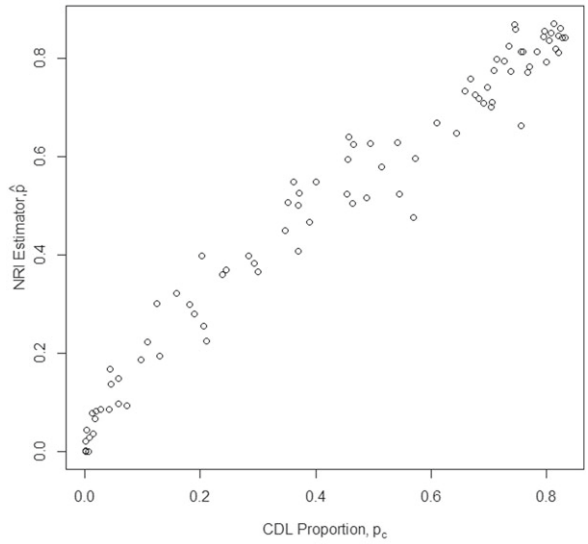

FIGURE 1: 2010 NRI estimates of the proportion of Minnesota counties in cropland in 2010 plotted against the corresponding CDL proportions.

where $\delta_{ij} = I(\text{ location } j \text{ is classified as cropland in 2010})$, and w_{ij} is the weight associated with location j in county i. The estimated coefficients of variation for the estimated proportion of the county classified as cropland range from 3% to 50%, with a mean of 14%.

We obtain auxiliary information from the United States Department of Agriculture National Agricultural Statistics Service geospatial Cropland Data Layer (CDL). The CDL is a classification of 30×30 meter pixels into land cover categories based on satellite data. The covariate is the proportion of CDL pixels in a county classified as cropland. Because many of the CDL proportions are near zero, we define the covariate by

$$p_{c,i} = \frac{\tilde{p}_{c,i} n_i + 0.5}{n_i + 1},$$

where n_i is the number of NRI primary sampling units in county i, and $\tilde{p}_{c,i}$ is the proportion of CDL pixels in a county classified in categories that are cropland according to the NRI definition of cropland. In Figure 1, the NRI estimates of the county-level proportions in cropland are plotted against the corresponding CDL proportions. The correlation between the NRI and CDL proportions is 0.98.

We use the model and estimation procedure from Section 2.1. The model for the direct estimator is $\hat{p}_i = p_{F,i} + u_i + e_i$, where $p_{F,i} = \exp(\beta_0 + \beta_1 x_i)(1 + \exp(\beta_0 + \beta_1 x_i))^{-1}$, $x_i = \text{logit}(p_{c,i}) - \text{logit}(p_{c,1})$, and $(u_i, e_i) \sim \text{diag}(\sigma_u^2 p_{F,i}(1 - p_{F,i}), \sigma_{ei}^2)$. This definition of x_i is a special case of the definition from Berg & Fuller (2012) for a situation in which the response is binary.

We use a generalized variance function for σ_{ei}^2 (Wolter, 2007). An analysis of the relationship between the direct estimates of the variances and preliminary estimates of $p_{F,i}$ suggests a variance model of the form $\sigma_{ei}^2 = 0.0001 + c n_i^{-1} p_{F,i}(1 - p_{F,i})$, where n_i is the number of primary sampling units in county i. We treat the estimate of σ_{ei}^2 as the true variance for this analysis. We estimate c by the ordinary least squares regression of the variance estimate $\hat{\sigma}_{ei}^2$ on $n_i^{-1} \hat{p}_{F,i}^{(0)}(1 - \hat{p}_{F,i}^{(0)})$, holding the intercept fixed at 0.0001. Zero variance estimates are replaced with $0.5 n_i^{-2}$, which is the method used by Berg & Fuller (2012) for the binary case. The initial estimate $\hat{p}_{F,i}^{(0)}$ is an estimator for a model with $\beta_1 = 1$ and $Var\{e_i\} = n_i^{-1} p_{F,i}(1 - p_{F,i})$. The estimate of c

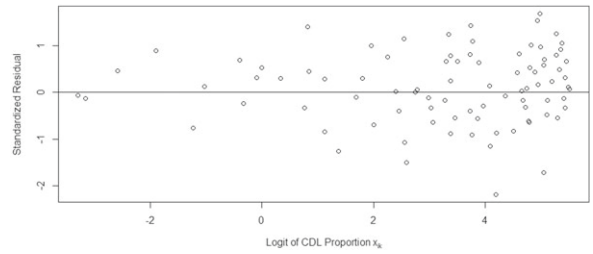

FIGURE 2: Standardized residuals $(\widehat{p}_i - \widehat{p}_{F,i})/\sqrt{\widehat{\sigma}^2_{ui} + \widehat{\sigma}^2_{ei,w}}$ plotted against x_{ik}.

TABLE 6: Ratios of estimated MSEs of benchmarked predictors to estimated MSEs of initial predictors.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Linear	1.00	1.01	1.02	1.03	1.03	1.25
Raking	1.00	1.04	1.06	1.05	1.07	1.09
Augmented	1.00	1.01	1.02	1.03	1.03	1.24

for the Minnesota data is 1.7. The estimates (and standard errors) of these model parameters are $\widehat{\beta}_0 = -2.790(0.086)$, $\widehat{\beta}_1 = 0.806$ (0.022), and $\widehat{\sigma}^2_u = 0.00084$ (0.00180). The plot of the standardized residuals against x_i is given in Figure 2.

We define initial predictors of the form $\widehat{p}_i^{pred} = \widehat{\gamma}_i \widehat{p}_i + (1 - \widehat{\gamma}_i)\widehat{p}_{F,i}$, where $\widehat{p}_{F,i} = [1 + \exp(\widehat{\beta}_0 + \widehat{\beta}_1 x_i)]^{-1} \exp(\widehat{\beta}_0 + \widehat{\beta}_1 x_i)$, and $\widehat{\gamma}_i = [0.001 + \widehat{c} n_i^{-1} \widehat{p}_{F,i}(1 - \widehat{p}_{F,i}) + \widehat{\sigma}^2_u \widehat{p}_{F,i}(1 - \widehat{p}_{F,i})]^{-1} \widehat{\sigma}^2_u \widehat{p}_{F,i}(1 - \widehat{p}_{F,i})$. The estimate of the state-level proportion in cropland based on the initial predictors is 0.386. Although the difference between the state-level proportion based on the predictors (0.386) and the state-level proportion based on the NRI estimates (0.389) is small, the difference is noticeable for an analyst comparing aggregated county-level predictors to the published estimate.

We compare three benchmarking procedures: the augmented model with the BHF value for ψ_i, raking, and the linear additive adjustment with $\phi_i^{-1} = \gamma_i \sigma^2_{ei}$, which corresponds to the BHF ψ_i. Table 6 summarizes the distributions of the ratios of the estimated MSEs of the benchmarked predictors to the estimated MSEs of the initial predictors. The estimated MSEs for raking, the linear additive adjustment, and the augmented model are defined in (17), (A1), and (A3), respectively. Although benchmarking increases the estimated MSE of the predictors, the estimated MSEs for the benchmarked predictors are typically smaller than the estimated MSEs for the direct estimators. For the augmented model and raking procedures, the estimated MSEs of the benchmarked predictors are uniformly smaller than the estimated variances of the direct estimators. For the linear additive procedure, seven estimated MSEs exceed the estimated variances of the corresponding direct estimators.

6. SUMMARY AND RECOMMENDATIONS

We propose two classes of benchmarking procedures for nonlinear small area models: a linear additive adjustment and an augmented model predictor. In application, the analyst can choose a benchmarking procedure within a class on the basis of a weight function. We recommend the linear additive approach. The linear additive approach is computationally simple and widely applicable. Raking can be viewed as a member of the linear additive class and, in our experience,

performs well for situations in which the sampling variance of the proportion is inversely related to the weight defining the restriction.

APPENDIX

SOFTWARE

https://github.com/emilyjb/benchmarked-sae-cjs-2018

MSE ESTIMATION FOR THE ADDITIVE ERROR MODEL

Linear Additive Benchmarking

For additive error models, an estimator of the MSE of $\widehat{\theta}_{\ell a,i}$ is

$$\widehat{Var}\{\widehat{\theta}_{\ell a,i} - \theta_i\} = \widehat{Var}\{\widehat{\theta}_i - \theta_i\} + \widehat{\phi}_i^{-2}\boldsymbol{w}_i^\top \widehat{Var}\{\widehat{\boldsymbol{\beta}}_{\ell a}\}\boldsymbol{w}_i, \tag{A1}$$

where $\widehat{Var}\{\widehat{\theta}_i - \theta_i\}$ is an estimator of (4),

$$\begin{aligned}\widehat{Var}\{\widehat{\boldsymbol{\beta}}_{\ell a}\} &= \widehat{\boldsymbol{M}}_w(\widehat{\Sigma}_{aa} - \widehat{\boldsymbol{H}}(\widehat{\boldsymbol{H}}^\top \widehat{\Sigma}_{aa}^{-1}\widehat{\boldsymbol{H}})^{-1}\widehat{\boldsymbol{H}}^\top)\widehat{\boldsymbol{M}}_w^\top,\\ \widehat{\boldsymbol{M}}_w &= (\boldsymbol{W}^\top\widehat{\Phi}^{-1}\boldsymbol{W})^{-1}\boldsymbol{W}^\top(\boldsymbol{I} - \widehat{\Gamma}),\end{aligned} \tag{A2}$$

$\widehat{\Phi} = \text{diag}(\widehat{\phi}_1, \ldots, \widehat{\phi}_m)$, $\widehat{\Gamma} = \text{diag}(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_m)$, $\widehat{\boldsymbol{H}} = (\widehat{\boldsymbol{h}}_1, \ldots, \widehat{\boldsymbol{h}}_m)^\top$, $\widehat{\boldsymbol{h}}_i = \partial g(\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}})/\partial\boldsymbol{\beta}$ is the vector of partial derivatives of $g(\boldsymbol{x}_i^\top\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ evaluated at $\widehat{\boldsymbol{\beta}}$, and $\widehat{\Sigma}_{aa} = \text{diag}(\widehat{\sigma}_{ui}^2 + \sigma_{ei}^2, \ldots, \widehat{\sigma}_{ui}^2 + \sigma_{ei}^2)$. We justify the variance estimator (A1) for the additive error model in the Supporting Information.

Augmented Model Benchmarking

By the justification in (S19 - S24) of the Supporting Information, an estimator of the MSE of the augmented model predictor for the additive error model is given by

$$\widehat{Var}\{\widehat{\theta}_i^{aug} - \theta_i\} = \widehat{Var}\{\widehat{\theta}_i - \theta_i\} + (1 - \widehat{\gamma}_i)^2\widehat{\boldsymbol{h}}_{i,aug}^\top \widehat{Var}\{\widehat{\boldsymbol{\beta}}_{aug}\}\widehat{\boldsymbol{h}}_{i,aug}, \tag{A3}$$

where $\widehat{\boldsymbol{h}}_{i,aug}$ is the derivative of $g(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \boldsymbol{z}_i^\top\boldsymbol{\beta}_{aug})$ with respect to $\boldsymbol{\beta}_{aug}$ evaluated at $(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\beta}}_{aug}^\top)^\top$.

MSE ESTIMATION FOR THE NON-ADDITIVE ERROR MODEL

Parametric Bootstrap MSE Estimators for Benchmarked Predictors

We provide details of the parametric bootstrap procedure used to estimate the MSE of the benchmarked predictors under the non-additive error model. Let $\widehat{\theta}_i^w(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)$ denote a benchmarked predictor constructed using either the linear additive or augmented model approach. We define MSE estimators for the benchmarked predictors analogous to (8) and (10). An MSE estimator that uses a multiplicative bias correction, as in (8), is

$$\begin{aligned}\widehat{\text{MSE}}_{i,mult}^{Bench} =& [\text{Var}\{\theta_i \mid y_i, (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\}]^2\left[R^{-1}\sum_{r=1}^{R}\text{Var}\{\theta_i|y_i^{(r)}; (\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)})\}\right]^{-1}\\ &+ R^{-1}\sum_{r=1}^{R}(\widehat{\theta}_i^w(\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)}) - \widehat{\theta}_i^{(r)}(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2))^2,\end{aligned} \tag{A4}$$

where $\widehat{\theta}_i^w(\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)})$ is the benchmarked EB predictor for the r^{th} bootstrap sample. The inversion MSE estimator analogous to (10) for the benchmarked predictor is

$$\widehat{MSE}_{i,c}^{Bench} = \text{Var}\{\theta_i \mid y_i; (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_u^2)\}\widehat{q}_{D(i)}^2[\Phi^{-1}(1-\alpha/2)]^{-2}, \tag{A5}$$

where $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal distribution, $\widehat{q}_{D(i)}$ satisfies

$$\frac{1}{|D(i)|}\frac{1}{R}\sum_{i\in D(i)}\sum_{r=1}^{R} I\left[\frac{|\widehat{\theta}_i^w(\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)}) - \theta_k^{(r)}|}{\sqrt{\text{Var}\{\theta_k | y_k^{(r)}; (\widehat{\boldsymbol{\beta}}^{(r)}, \widehat{\sigma}_u^{2(r)})\}}} \le \widehat{q}_{D(i)}\right] = 1-\alpha,$$

and $D(i)$ is a group of areas containing area i.

ACKNOWLEDGEMENTS

This work was partially supported by the United States Department of Agriculture (USDA NRCS CESU Agreement 68-7482-13-529) and by the the National Science Foundation (MMS-1733572).

BIBLIOGRAPHY

Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An error components model for prediction of S crop areas using survey and satellite data. *Journal of the American Statistical Association*, 401, 28–36.

Bell, W. R., Datta, G. S., & Ghosh, M. (2012). Benchmarking small area estimators. *Biometrika*, 100, 1–35.

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

Berg, E. & 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. & Fuller, W. A. (2012). Estimators of error covariance matrices for small area prediction. *Computational Statistics and Data Analysis*, 56, 2949–2962.

Berg, E. & Fuller, W. A. (2014). Small area prediction of proportions with applications to the Canadian Labour Force Survey. *Survey Methodology*, 2, 227–256.

Datta, G. S., Ghosh, M., Steorts, R., & Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, 20, 574–588.

Erciulescu, A. L. (2015). Small area prediction based on unit level models when the covariate mean is measured with error. Graduate Theses and Dissertations. Paper 14674, http://lib.dr.iastate.edu/etd/14674.

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

Ghosh, M. & Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, 91, 95–112.

Ghosh, M., Kubokawa, T., & Kawakubo, Y. (2015). Benchmarked empirical Bayes methods in multiplicative area-level models with risk evaluation. *Biometrika*, 102, 647–659.

Ghosh, M. & Steorts, R. C. (2013). Two-stage benchmarking as applied to small area estimation. *Test*, 22, 670–687.

Hidiroglou, M. A. & 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., & Fuller, W. A. (2000). Estimation of census adjustment factors. *Survey Methodology*, 26, 31–42.

Lahiri, P. (1990). Adjusted Bayes and empirical Bayes estimation in finite population sampling. *Sankhya B*, 52, 50–56.

Lahiri, P., Maiti, T., Katzoff, M., & Parsons, V. (2007). Resampling-based empirical prediction: An application to small area estimation, *Biometrika*, 94, 469–485.

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

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

Nusser, S. M. & Goebel, J. J. (1997). The national resources inventory: A long-term multi-resource monitoring programme. *Environmental and Ecological Statistics*, 4, 181–204.

Pfeffermann, D. & Barnard, C. H. (1991). Some new estimators for small area means with application to assessment of farmland values. *Journal of Business and Economic Statistics*, 9, 73–84.

Pfeffermann, D., Sikov, A., & Tiller, R. (2014). Single-and two-stage cross-sectional and time series benchmarking procedures for small area estimation. *Test*, 23, 631–666.

Rao, J. N. K. (2003). *Small Area Estimation*, Wiley, New Jersey.

Rao, J. N. K. & Molina, I. (2015). *Small Area Estimation*, Wiley, New Jersey.

Steorts, R. C. & Ghosh, M. (2013). On estimation of mean squared errors of benchmarked empirical Bayes estimators. *Statistica Sinica*, 749–767.

Ugarte, M. D., Militino, A. F., & B Goicoa, T. (2009). Benchmarked estimates in small areas using linear mixed models with restrictions. *Test*, 18, 342–364.

Wang, J., Fuller, W. A., & Qu, Y. (2008). Small area estimation under a restriction. *Survey Methodology*, 34, 29–36.

Wolter, K. (2007). *Introduction to variance estimation*, 2nd ed. Springer-Verlag, New York.

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

You, Y., Rao, J. N. K., & Hidiroglou, M. (2013). On the performance of self benchmarked small area estimators under the Fay-Herriot area level model. *Survey Methodology*, 39, 217–230.

Received 17 October 2016
Accepted 4 March 2018