STATISTICS IN TRANSITION new series and *SURVEY METHODOLOGY*
Joint Issue: Small Area Estimation 2014
Vol. 17, No. 1, pp. 9–24

SMALL AREA PREDICTION UNDER ALTERNATIVE MODEL SPECIFICATIONS

Andreea L. Erciulescu[1], **Wayne A. Fuller**[2]

ABSTRACT

Construction of small area predictors and estimation of the prediction mean squared error, given different types of auxiliary information are illustrated for a unit level model. Of interest are situations where the mean and variance of an auxiliary variable are subject to estimation error. Fixed and random specifications for the auxiliary variables are considered. The efficiency gains associated with the random specification for the auxiliary variable measured with error are demonstrated. A parametric bootstrap procedure is proposed for the mean squared error of the predictor based on a logit model. The proposed bootstrap procedure has smaller bootstrap error than a classical double bootstrap procedure with the same number of samples.

Key words: unit level model, parametric bootstrap, double bootstrap, measurement error, auxiliary information.

1. Introduction

Small area estimation procedures use models and auxiliary data to construct estimates for subpopulations that are more efficient than the direct estimators for those subpopulations. Modeling provides potential for gains by postulating a distribution for the unknown parameters. The presence of variables that are correlated with the variable of interest provides potential for efficiency gains when there is knowledge about the distribution of those variables. In most of the small area literature the small area population means of the auxiliary variables are assumed to be known. We are interested in the situation where only estimates of the parameters of the distribution of the auxiliary variables are available. Our study was motivated by a situation where the sample used for small area estimation was a subsample of a larger survey. The larger survey furnished estimates of the distribution of the auxiliary variables.

[1] National Institute of Statistical Sciences and USDA NASS, 1400 Independence Ave. SW, Room 6040 F, Washington, DC 20250.
[2] Iowa State University, 1214 Department of Statistics, Ames, IA 50010.

A number of papers consider measurement error in the auxiliary variables used in the linear regression model. See Fuller and Harter (1987), Ghosh, Sinha and Kim (2006), Ghosh and Sinha (2008), Torabi, Datta and Rao (2009), Ybarra and Lohr (2008) and Datta, Rao and Torabi (2010). In contrast, we study unit level mixed models where the observed explanatory variables are measured without error, but the parameters of the distribution of the auxiliary variables are known subject to estimation error. We consider auxiliary information obtained from a sample, including the limit case of a complete sample.

Because there are no closed-form estimators for the prediction mean squared error (MSE) for most nonlinear models, bootstrap methods have been suggested. See Hall and Maiti (2006) and Pfeffermann and Correa (2012). We propose parametric bootstrap procedures based on the work of Davidson and MacKinnon (2007).

This paper is organized in sections. In Section 2.2 we present predictors of small area means assuming a unit level generalized linear mixed model, with alternative specifications for the auxiliary information. In Section 2.4 we describe parametric double bootstrap procedures for MSE estimation. Section 3.2 contains simulation results comparing the prediction MSEs for the logit model under alternative model specifications and alternative types of data for the auxiliary variables. Simulation comparisons of alternative bootstrap prediction MSE estimators are given in Section 3.3.

2. Unit Level Nonlinear Models

2.1. Introduction

The unit level generalized linear mixed model considered in this study is

$$E\left[y_{ij}|\boldsymbol{x}_{ij}, b_i\right] = g(\boldsymbol{x}_{ij}\boldsymbol{\beta}, b_i), \tag{1}$$

$$\boldsymbol{x}_{ij} = \boldsymbol{\mu}_{xi} + \boldsymbol{\varepsilon}_{ij}, \tag{2}$$

$i = 1, \ldots, m$, where m is the number of areas, j is the index for units in the area, $\boldsymbol{\beta}$ is a vector of coefficients, $\boldsymbol{\mu}_{xi}$ is the area mean of the auxiliary variable, and b_i is the area random effect. It is assumed that the b_i are independent and identically distributed, with a density f_b with mean 0 and variance σ_b^2, mutually independent of ε_{ij}, where the ε_{ij} are independent and identically distributed random variables with a density f_ε with mean 0 and variance σ_ε^2. The vector $(y_{ij}, \boldsymbol{x}_{ij}), i = 1, \ldots, m, j = 1, \ldots, n_i$ is observed.

Additional information on the distribution of $\boldsymbol{x}_{ij}$ may be available. Possibilities include a second sample of $\boldsymbol{x}_{ij}$ observations, or an estimator of $\boldsymbol{\mu}_{xi}$,

or complete knowledge of the distribution function. The area means of $\boldsymbol{x}$ can be treated as fixed or as random variables. If random, we assume

$$\boldsymbol{\mu}_{xi} = \boldsymbol{\mu}_{x} + \boldsymbol{\delta}_{i}, \tag{3}$$

where δ_i are independent and identically distributed, with a density f_δ with mean 0 and variance σ_δ^2. Assume δ_i are independent of $b_k, e_{ij}, \varepsilon_{rt}$, for all i, k, r and t, where $e_{ij} = y_{ij} - g(\boldsymbol{x}_{ij}\boldsymbol{\beta}, b_i)$.

Of interest is the i^{th} small area mean of $\boldsymbol{y}$

$$\theta_i = \int g(\boldsymbol{x}\boldsymbol{\beta}, b_i) dF_{\boldsymbol{x}_i}(\boldsymbol{x}), \tag{4}$$

where $F_{\boldsymbol{x}_i}(\boldsymbol{x})$ is the distribution of $\boldsymbol{x}$ in area i. Also of interest is the prediction mean squared error

$$\alpha_i = E(\hat{\theta}_i - \theta_i)^2, \tag{5}$$

where $\hat{\theta}_i$ is the predictor. We assume throughout that the area population is large so that we need not consider finite population corrections.

The nature of the estimation-prediction problem is determined by the distributional properties of the vector $(b_i, \boldsymbol{\delta}_i, \boldsymbol{\varepsilon}_{ij})$. The nonlinear model is more complicated than the linear model for several reasons. First, parameter estimation is more difficult because no closed form estimator exists. Likewise, closed form estimators of the mean squared error do not exist. Lastly, the small area mean of the auxiliary variable is not sufficient for the estimation of θ_i.

As an example of model (1), consider a Bernoulli response variable $\boldsymbol{y}$, with realizations y_{ij} for m different areas and n_i different units within each area. To simplify the presentation, we consider scalar x_{ij} for the remainder of our discussion. Let x_{ij} be independent and identically distributed, following a distribution $F_{\boldsymbol{x}_i}$. Let the expected value of y_{ij} given $(\boldsymbol{x}_{ij}, b_i)$ be

$$g(\boldsymbol{x}_{ij}\boldsymbol{\beta}, b_i) = \frac{exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + b_i)}{1 + exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + b_i)}, \tag{6}$$

where $\boldsymbol{x}_{ij} = (1, x_{ij})$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. The model is the generalized linear mixed model with logit link.

2.2. Predictors of $\boldsymbol{\theta_i}$

We present predictors of θ_i for model (6), under alternative specifications for $\boldsymbol{x_{ij}}$ and for different levels of auxiliary information, given known parameters $(\sigma_b^2, \sigma_\varepsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x)$.

2.2.1. Known Covariate Distribution

Let the distribution of x_{ij} be known and let $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ be a random sample of (x_{ij}, y_{ij}), where $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}), \boldsymbol{y}_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,n_i})$. Then, given known parameters, the minimum mean squared error (MMSE) predictor of the i^{th} small area mean of $\boldsymbol{y}$ is

$$\begin{aligned} \hat{\theta}_i &= E[\theta_i(b)|(\boldsymbol{x}_i, \boldsymbol{y}_i)] \\ &= \frac{\int_b \theta_i(b) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_b(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_b(b)}, \end{aligned} \tag{7}$$

where

$$\theta_i(b) = \int g(\boldsymbol{x}\boldsymbol{\beta}, b) dF_{\boldsymbol{x}_i}(\boldsymbol{x}).$$

In some finite population situations, the entire finite population of $\boldsymbol{x}$ values may be known and the integral expression for $\theta_i(b)$ in (7) is the sum over the population. In the simulations for this model we assume $x_{ij} \sim NI(\mu_{xi}, \sigma_\varepsilon^2)$ with μ_{xi} known and σ_ε^2 known.

2.2.2. Sample Estimated Covariate Distribution

Let an estimator of the distribution of x_{ij} be given by a sample $(x_{ij}, w_{ij}), j = 1, \ldots, r_i$, where w_{ij} are weights such that the sample cumulative distribution function (CDF) is unbiased for the population CDF. Then, given known $(\sigma_b^2, \boldsymbol{\beta})$, the predictor of the i^{th} small area mean of $\boldsymbol{y}$ is

$$\begin{aligned} \hat{\theta}_i &= E[\theta_i(b)|(\boldsymbol{x}_i, \boldsymbol{y}_i)] \\ &= \frac{\int_b \theta_i(b) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_b(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_b(b)}, \end{aligned} \tag{8}$$

where

$$\theta_i(b) = \sum_{j=1}^{r_i} w_{ij} g(\boldsymbol{x_{ij}}\boldsymbol{\beta}, b).$$

The sample used to estimate the CDF could be the original sample with $r_i = n_i$ or the estimation sample could be the original sample augmented by an

additional probability sample of size n_i' selected from the area population. See Ghosh et al. (2009) for an example using the sample CDF.

2.2.3. Unknown Random Covariate Mean

Assume the form of the distribution of $\boldsymbol{x}$ for area i is known, with unknown parameters $(\mu_{xi}, \sigma_\varepsilon^2)$. Assume μ_{xi} satisfies (3). Then, given known $(\sigma_b^2, \sigma_\varepsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x)$, the MMSE predictor of the i^{th} small area mean of $\boldsymbol{y}$ is

$$\begin{aligned} \hat{\theta}_i &= E[\theta_i(b,\delta)|(\boldsymbol{x}_i, \boldsymbol{y}_i)] \\ &= \frac{\int_b \int_\delta \theta_i(b,\delta) \prod_{t=1}^{n_i} f(y_{it}|x_{it},b) f(x_{it}|\delta) dF_\delta(\delta) dF_b(b)}{\int_b \int_\delta \prod_{t=1}^{n_i} f(y_{it}|x_{it},b) f(x_{it}|\delta) dF_\delta(\delta) dF_b(b)}, \end{aligned} \tag{9}$$

where

$$\theta_i(b,\delta) = \int g[(\mu_x + \delta + \boldsymbol{\varepsilon})\,\boldsymbol{\beta}, b] dF_{\boldsymbol{\varepsilon}_i}(\boldsymbol{\varepsilon}).$$

In the simulations we assume $x_{ij} \sim NI(\mu_{xi}, \sigma_\varepsilon^2)$ and $\delta_i \sim NI(0, \sigma_\delta^2)$.

2.2.4. Unknown Random Covariate Mean, Additional Information $\tilde{\boldsymbol{x}}_i$

Let the random model assumptions of Section 2.2.3 hold. Let a vector of n_i' observations on x_{ij}, denoted by $\tilde{\boldsymbol{x}}_i$, be available. Then, given known $(\sigma_b^2, \sigma_\varepsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x)$, the MMSE predictor of the i^{th} small area mean of $\boldsymbol{y}$ is

$$\begin{aligned} \hat{\theta}_i &= E[\theta_i(b,\delta))|(\boldsymbol{x}_i, \boldsymbol{y}_i, \tilde{\boldsymbol{x}}_i)], \\ &= \frac{\int_b \int_\delta \theta_i(b,\delta) \prod_{t=1}^{n_i} f(y_{it}|x_{it},b) f(x_{it}|\delta) \prod_{t'=1}^{n_i'} f(\tilde{x}_{it'}|\delta) dF_\delta(\delta) dF_b(b)}{\int_b \int_\delta \prod_{t=1}^{n_i} f(y_{it}|x_{it},b) f(x_{it}|\delta) \prod_{t'=1}^{n_i'} f(\tilde{x}_{it'}|\delta) dF_\delta(\delta) dF_b(b)}, \end{aligned}$$

where

$$\theta_i(b,\delta) = \int g[(\mu_x + \delta + \boldsymbol{\varepsilon})\boldsymbol{\beta}, b]\, dF_{\boldsymbol{\varepsilon}_i}(\boldsymbol{\varepsilon}).$$

In the simulations we assume $\tilde{x}_{ij'} \sim NI(\mu_{xi}, \sigma_\varepsilon^2)$, so $\tilde{\mu}_{xi} = (n_i')^{-1} \sum_{j'=1}^{n_i'} \tilde{x}_{ij'}$ is a sufficient statistic for μ_{xi} and the predictor simplifies to

$$\hat{\theta}_i = \frac{\int_b \int_\delta \theta_i(b,\delta) \prod_{t=1}^{n_i} f(y_{it}|x_{it},b) f(x_{it}|\delta) f(\tilde{\mu}_{xi}|\delta) dF_\delta(\delta) dF_b(b)}{\int_b \int_\delta \prod_{t=1}^{n_i} f(y_{it}|x_{it},b) f(x_{it}|\delta) f(\tilde{\mu}_{xi}|\delta) dF_\delta(\delta) dF_b(b)}. \tag{10}$$

2.3. Estimation

In practice, the vector of parameters $\boldsymbol{\psi} = (\sigma_b^2, \sigma_\varepsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x)$ is not known and needs to be estimated. Consider the model specified by (1), (2), (3), (6), with additional information $\tilde{\boldsymbol{x}}_i$ available, as described in Section 2.2.4. The likelihood is

$$L(\sigma_b^2, \sigma_\varepsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x | \boldsymbol{x}, \boldsymbol{y}, \tilde{\boldsymbol{x}}) = \prod_{i=1}^{m} L_i,$$

where

$$\begin{aligned} L_i &= \int_b \int_\delta \prod_{j=1,j'=1}^{n_i,n_i'} f(y_{ij}, x_{ij}, \tilde{x}_{ij'} | b, \delta, \psi) f(b|\psi) f(\delta|\psi) d\delta db \\ &= \int_b \prod_{j=1}^{n_i} f(y_{ij}|b, x_{ij}, \beta) f(b|\sigma_b^2) db \int_\delta \prod_{j=1}^{n_i+n_i'} f(x_{ij}^*|\delta, \mu_x, \sigma_\varepsilon^2) f(\delta|\sigma_\delta^2) d\delta, \end{aligned}$$

and $\boldsymbol{x}^* = (\boldsymbol{x}, \tilde{\boldsymbol{x}})$ is the vector of all available auxiliary information.

Notice that the likelihood $L(\sigma_b^2, \sigma_\varepsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x | \boldsymbol{x}, \boldsymbol{y}, \tilde{\boldsymbol{x}})$ factors into $L(\sigma_b^2, \boldsymbol{\beta} | \boldsymbol{y}, \boldsymbol{x})$ and $L(\sigma_\varepsilon^2, \sigma_\delta^2, \mu_x | \boldsymbol{x}, \tilde{\boldsymbol{x}})$. Hence, the parameters $(\sigma_\varepsilon^2, \sigma_\delta^2, \mu_x)$ can be estimated separately from the estimation of the parameters $(\sigma_b^2, \boldsymbol{\beta})$. Estimation of $(\sigma_\varepsilon^2, \sigma_\delta^2, \mu_x)$ can be based on maximizing the likelihood for the linear mixed model specified in (2) and (3), with additional information $\tilde{\boldsymbol{x}}_i$ available.

Numerical integration methods are required for construction of estimates and predictions.

2.4. Bootstrap MSE Estimation

In this section we consider estimation of the MSE of $\hat{\theta}_i$ as a predictor of θ_i. Let $\boldsymbol{\psi}$ be the parameter that defines the distribution of the sample observations, and let $\widehat{\boldsymbol{\psi}}$ be an estimator of $\boldsymbol{\psi}$. Let $\boldsymbol{\alpha}$ be a vector of parameters of interest and let $\boldsymbol{\alpha}^*$ be a parametric bootstrap (simulation) estimator of $\boldsymbol{\alpha}$. For the models considered in Section 2.2, let α_i be the MSE of the prediction error for area i, as defined in (5). For the nonlinear small area model with known distribution for x_{ij}, the vector of parameters is $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta})$. For the nonlinear small area models with unknown random μ_{xi}, the vector of parameters is $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta}, \sigma_\varepsilon^2, \mu_x, \sigma_\delta^2)$. Because there is no closed form expression for the prediction MSE given in (5), we consider bootstrap MSE estimation.

A sample generated with $\boldsymbol{\psi}$ and random number seed r is said to be created with data generator $(\boldsymbol{\psi}, r)$, denoted $DG(\boldsymbol{\psi}, r)$. Let B_1 bootstrap samples be generated using random number seeds $r_{1,1}, r_{1,2}, \ldots, r_{1,B_1}$. Let $\boldsymbol{\psi}_k^*$ be the estimator of $\boldsymbol{\psi}$ from the kth bootstrap sample generated using $DG(\widehat{\boldsymbol{\psi}}, r_{1,k})$. The bootstrap estimator of prediction MSE for area i is

$$\hat{\alpha}_i^* = B_1^{-1} \sum_{k=1}^{B_1} (\hat{\theta}_{i,k}^* - \theta_{i,k}^*)^2 =: B_1^{-1} \sum_{k=1}^{B_1} \alpha_{i,k}^* = \bar{\alpha}_i^*, \qquad (11)$$

where $\theta_{i,k}^*$ is the true small area mean generated for the kth bootstrap sample, $\hat{\theta}_{i,k}^*$ is the sample predictor of $\theta_{i,k}^*$ and $\alpha_{i,k}^*$ is the prediction squared error for the kth bootstrap sample. The estimator (11) is called the level-one bootstrap estimator.

In the double bootstrap, a sample estimator, denoted by α_i^{**}, is generated using $\boldsymbol{\psi}^*$ from the level-one generated sample. Typically a large number of α_i^{**} is generated for each α_i^* and the bias adjusted estimator is

$$\tilde{\alpha}_i^{**} = B_1^{-1} \sum_{k=1}^{B_1} 2\alpha_{i,k}^* - B_1^{-1} B_2^{-1} \sum_{k=1}^{B_1} \sum_{t=1}^{B_2} \alpha_{i,k,t}^{**}. \tag{12}$$

where $\alpha_{i,k,t}^{**}$ is generated using $DG(\boldsymbol{\psi}_k^*, r_{2,k,t})$, B_1 is the number of level-one bootstrap samples, B_2 is the number of level-two bootstrap samples per level-one sample, and the $r_{2,k,t}, k = 1,2,\ldots,B_1, t = 1,2,\ldots,B_2$, are independent random numbers, independent of $r_{1,k}$.

We use a double bootstrap estimator based on the work of Davidson and MacKinnon (2007) who give a fast double bootstrap procedure for bootstrap testing. See also Giacomini, Politis and White (2013). In the fast double bootstrap, a single α_i^{**} is generated for each α_i^*. Let $r_{2,1}, r_{2,2}, \ldots, r_{2,B_1}$ be a second independent sequence of random numbers. Given the sequence of random numbers, define $\alpha_{i,k}^{**}$ to be calculated from data generated with $DG(\boldsymbol{\psi}_k^*, r_{2,k})$. The (classic) double bootstrap estimator used in this study is

$$\tilde{\alpha}_{i,C}^{**} = B_1^{-1} \sum_{k=1}^{B_1} (2\alpha_{i,k}^* - \alpha_{i,k}^{**}) = 2\bar{\alpha}_i^* - \bar{\alpha}_i^{**}. \tag{13}$$

To construct an even more efficient bootstrap estimator, define $\alpha_{i,k,2}^*$ to be calculated from data generated with $DG(\widehat{\boldsymbol{\psi}}, r_{2,k})$. Then a bias adjusted (double bootstrap) estimator is

$$\hat{\alpha}_i^{**} = B_1^{-1} \sum_{k=1}^{B_1} (\alpha_{i,k}^* + \alpha_{i,k,2}^* - \alpha_{i,k}^{**}), \tag{14}$$

where the quantity $\alpha_{i,k}^{**} - \alpha_{i,k}^*$ is a one-degree-of-freedom estimator of the bias. If one uses $r_{2,1}$ as $r_{1,2}$, $r_{2,2}$ as $r_{1,3}$, etc., a form of (14) becomes

$$\tilde{\alpha}_{i,T}^{**} = B_1^{-1} \sum_{k=1}^{B_1} (\alpha_{i,k}^* + \alpha_{i,k+1}^* - \alpha_{i,k}^{**}), \tag{15}$$

where $\alpha_{i,k+1}^*$ is generated with $DG(\widehat{\boldsymbol{\psi}}, r_{1,k+1})$ and $\alpha_{i,k}^{**}$ is generated with $DG(\boldsymbol{\psi}_k^*, r_{1,k+1})$. We call the estimator (15) a telescoping bootstrap because it is of the form (14) using lagged values of $\alpha_{i,k}^*$. If the use of $r_{2,k}$ in place of an

independent random number results in positive correlation between $\alpha_{i,k}^{*}$ and $\alpha_{i,k-1}^{**}$, then $\tilde{\alpha}_{i,T}^{**}$ will have smaller simulation variance than $\tilde{\alpha}_{i,C}^{**}$ of (13).

3. Simulations

In the simulation study we consider $m = 36$ areas with unit level observations x_{ij} in three groups of 12 areas, with sizes $n_i \in \{2,10,40\}$. The number of additional unit level observations is $n_{i,} = 10$, for each area i. Each sample, $(\boldsymbol{y}, \boldsymbol{x}, \tilde{\boldsymbol{x}})$, is generated using model (1 - 3) with $\sigma_b^2 = 0.25, \mu_x = 0, \sigma_\delta^2 = 0.16$, and $\sigma_\varepsilon^2 = 0.36$. The vector of coefficients for the fixed effects is $(\beta_0, \beta_1) = (-0.8,1)$ and $\boldsymbol{x}_{ij} = (1, x_{ij})$. For each unit, the probability that $y_{ij} = 1$ is

$$g(\boldsymbol{x_{ij}\beta}, b_i) = \frac{\exp(-0.8+x_{ij}+b_i)}{1+\exp(-0.8+x_{ij}+b_i)}. \quad (16)$$

The population mean of $g(\boldsymbol{x_{ij}\beta}, b_i)$ is 0.334 with variance 0.029. An area with $\mu_{xi} = 0.4$ has mean 0.412 with variance 0.028. Four hundred Monte Carlo samples were generated satisfying the model.

The estimation models are:

- Model 1: Specified by (1) and (6) and described in Section 2.2.1. Known normal distribution for x_{ij}. The distribution of y_{ij} is

$$f(y_{ij}|x_{ij}, b_i) = I(y_{ij}, 1)g(\boldsymbol{x_{ij}\beta}, b_i) + I(y_{ij}, 0)(1 - g(\boldsymbol{x_{ij}\beta}, b_i)),$$

  where $I(y_{ij}, .)$ is the indicator function, and $g(\boldsymbol{x_{ij}\beta}, b_i)$ is defined in (16). The distribution of b_i is $N(0,0.25)$.

- Model 2: Specified by (1) and (6) and described in Section 2.2.2. Sample estimated distribution of $\boldsymbol{x}$ based on the original sample $\boldsymbol{x}$.
- Model 2*: Specified by (1) and (6) and described in Section 2.2.2. Sample estimated distribution of $\boldsymbol{x}$ based on the original sample $\boldsymbol{x}$ augmented by a sample $\tilde{\boldsymbol{x}} = (\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \ldots, \tilde{\boldsymbol{x}}_m)$.
- Model 3: Specified by (1), (2), (6) and described in Section 2.2.3. Unknown random auxiliary mean μ_{xi}. Distributions of y_{ij} and b_i are the same as those for Model 1. The distribution of x_{ij} is defined by the random model given in Section 2.2.3.
- Model 4: Specified by (1), (2), (3), (6) and described in Section 2.2.4. Unknown random auxiliary mean μ_{xi} and observed $\tilde{\boldsymbol{x}} = (\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \ldots, \tilde{\boldsymbol{x}}_m)$.

The models are fitted as generalized linear mixed models, using the *lmer* and *glmer* functions in the *lme4* package in R. The true i^{th} small area mean of $\boldsymbol{y}$ is given by (4) and the predicted i^{th} area means of $\boldsymbol{y}$ are given in (7 - 10), with estimated $(\mu_x, \beta_0, \beta_1, \sigma_b^2, \sigma_\delta^2, \sigma_\varepsilon^2)$. The integrals in (4, 7 - 10) were approximated using a 26-point approximation to the normal distribution.

3.1. Refinement of Prediction MSE Estimators

Wang and Fuller (2003) suggested the estimator of σ_δ^2 be bounded by

$$K_{\delta,s} = 0.5\left[\hat{V}(\hat{\sigma}_\delta^2 | \sigma_\delta^2 = 0)\right]^{0.5},$$

where $\hat{V}(\hat{\sigma}_\delta^2 | \sigma_\delta^2 = 0)$ is the estimated variance of $\hat{\sigma}_\delta^2$, given $\sigma_\delta^2 = 0$. Because of the large degrees of freedom for $\hat{\sigma}_\varepsilon^2$, we set $K_{\delta,s}$ equal to the true value of 0.008 in the simulations,

$$K_{\delta,s} = 0.5[2m(m-1)^{-1}(\textstyle\sum_{i=1}^{m}((n_i + n_{i\prime})^{-1}\sigma_\varepsilon^2)^{-2})^{-1}]^{0.5} = 0.008.$$

Similarly, we bound the estimator of σ_b^2 by

$$K_{b,s} = 0.5\left[V(\hat{\sigma}_b^2 | \sigma_b^2 = 0)\right]^{0.5} = 0.006.$$

The proportion of sample estimators $\hat{\sigma}_b^2$ that hit the bound is 0.025, the proportion of level one estimators of $\hat{\sigma}_b^{2*}$ that hit the bound is 0.111. If $\hat{\sigma}_{b,k}^2 = 0.006$ we set $\alpha_{i,k}^{**}$ equal to $\alpha_{i,k}^{*}$. That is, the estimated bias is zero for such samples.

Using (13), one can obtain an unacceptable double bootstrap prediction MSE estimator, where the estimated bias for a sample is greater than the estimate. In practice, one would increase the number of bootstrap samples. Rather than build such a procedure into our Monte Carlo algorithm, we defined bounds for the estimator. Thus, the final estimator is

$$\hat{\alpha}_{i,C}^{**} = \begin{cases} 1.60\bar{\alpha}_i^{*}, & if \quad \bar{\alpha}_i^{*-1}\bar{\alpha}_i^{**} > 1.60 \\ 0.83\bar{\alpha}_i^{*}, & if \quad \bar{\alpha}_i^{*-1}\bar{\alpha}_i^{**} < 0.83 \\ \tilde{\alpha}_{i,C}^{**}, & otherwise \ , \end{cases} \tag{17}$$

where 0.83 and 1.60 are the 0.025 and 0.975 points of the chi-square distribution with 199 $(B_1 - 1)$ degrees of freedom, and $\tilde{\alpha}_{i,C}^{**}$ is defined in (13). The analogous definition holds for the telescoping estimator of (14). See Hall and Maiti (2006) for an alternative definition of the direct double bootstrap estimates.

The proportions of sample estimators of $\hat{\alpha}_{i,T}^{**}$ that hit the lower bound defined in (17) are 0.016, 0.016 and 0.013, for the areas of sizes 2, 10 and 40, respectively. The proportions of sample estimators of $\hat{\alpha}_{i,T}^{**}$ that hit the upper bound defined in (17) are 0.026, 0.069 and 0.084, for the areas of sizes 2, 10 and 40, respectively. Due to larger variability in the classic double bootstrap estimators, the proportions of sample estimators of $\hat{\alpha}_{i,C}^{**}$ that hit the lower bound defined in (17) are 0.058, 0.048 and 0.041, for the areas of sizes 2, 10 and 40, respectively, and the proportions of sample estimators of $\hat{\alpha}_{i,C}^{**}$ that hit the upper bound defined in (17) are 0.155, 0.201 and 0.183, for the areas of sizes 2, 10 and 40, respectively.

3.2. MSE for Different Types of Auxiliary Information

The coefficient of variation for $\hat{\sigma}_b^2$ calculated for the 400 Monte Carlo samples is about 0.64, approximately the CV of a Chi-square with five degrees of freedom. The Monte Carlo relative bias of the estimator of $\hat{\sigma}_b^2$ is about -0.12, which is approximately equal to eighteen Monte Carlo standard errors.

Table 1 contains estimates of the prediction MSE, denoted by α, for fixed and random models with different amounts of auxiliary information. The simulation MSE standard errors are presented in parantheses below the MSE values. The smallest MSE is for Model 1, where the covariate distribution is known. The next smallest MSE is for Model 4, where the form of the covariate distribution is known, the covariate mean is random and the auxiliary information is available. The largest MSE is for Model 2, where the covariate distribution is not specified. The small area mean predictor for Model 3 is the conditional expected value formula given in (9). Notice that in the construction of the small area predictor for Model 4, given in (10), the conditioning is also on the additional source of information, $\boldsymbol{x}$, available for the areas.

The extra observations on x_{ij} represent additional information available about the distribution of $\boldsymbol{x}$ for the area. Hence, the large gain in efficiency associated with $\boldsymbol{x}$ for sample size two (compare 10.94 for Model 2^* to 17.29 for Model 2). Model 3 differs from Model 2 in that the distribution of x_{ij} is assumed to be normal and the area mean is also assumed to be normally distributed. Adding these distributional assumptions changes the MSE from 17.29 to 13.22 for sample size two. The effect of added information is smaller for the random μ_{xi} models (models 2^* and 4) than for the fixed μ_{xi} models (models 2 and 3).

The contribution of the variance of the estimation error in the mean of $\boldsymbol{x}$ to the MSE depends on the importance of $\boldsymbol{x}$ in the model and on the size of the samples. With $n_i = 2$, the MSE with known area mean of $\boldsymbol{x}$ is 57% of the MSE with no additional information on the distribution of $\boldsymbol{x}$. The reduction in MSE from adding independent observations on $\boldsymbol{x}$ is related to the sizes of the two samples and to the model. If the small area mean of $\boldsymbol{x}$ is fixed, the original sample is ten observations and the added sample is ten observations, the MSE falls midway

between that with no additional information and that with complete information. With fixed small area mean of $\boldsymbol{x}$, an original sample of size 2 and an added sample of size 10, the expected added variance is one sixth of that of the original sample. In this simulation the effect of treating the mean as random is equivalent to adding 2.25 observations on $\boldsymbol{x}$.

Table 1. MSE for different types auxiliary information (entries multiplied by 10^3)

Size	$\bar{y}$	Model 1	Model 2	Model 2*	Model 3	Model 4
2	102.14 (6.13)	9.88 (0.71)	17.29 (1.24)	10.94 (0.79)	13.22 (0.92)	10.72 (0.76)
10	20.15 (1.40)	7.15 (0.52)	8.56 (0.63)	7.87 (0.57)	8.26 (0.60)	7.76 (0.56)
40	5.14 (0.37)	3.46 (0.25)	3.81 (0.27)	3.74 (0.27)	3.78 (0.27)	3.72 (0.27)

Model 1: known distribution for x_{ij},
Model 2: unknown distribution for x_{ij}, with no $\tilde{\boldsymbol{x}}$,
Model 2*: unknown distribution for x_{ij}, with observed $\tilde{\boldsymbol{x}}$,
Model 3: random μ_{xi}, with no $\tilde{\boldsymbol{x}}$,
Model 4: random μ_{xi}, with observed $\tilde{\boldsymbol{x}}$

3.3. Monte Carlo Properties of Prediction MSE Estimators

The relative performances of bootstrap prediction MSE estimators under the different types of auxiliary information are similar. Therefore, we only present properties of prediction MSE estimators for Model 4, where the area mean μ_{xi} is random and auxiliary information $\tilde{\boldsymbol{x}}$ is available.

Table 2 contains results for $(\hat{\alpha}^{*}, \hat{\alpha}_T^{**}, \hat{\alpha}_C^{**})$ for the three area sample sizes, in groups of five lines. Each line is the average of the results for the 12 areas with the same sample size. The first line is the Monte Carlo estimates of the prediction MSE, $\hat{\alpha}$. The next four lines are of the bias relative to the mean, the coefficient of variation, the bias relative to the standard deviation and the bias relative to the standard error. The definitions are

$$RelBias = \sum_{is=1}^{12} \left(\hat{\alpha}_{.,is}^{EST} - \hat{\alpha}_{.,is}\right) \Big/ \sum_{is=1}^{12} \hat{\alpha}_{.,is} \,,$$

$$CV = \sum_{is=1}^{12} \sqrt{(400-1)^{-1} \sum_{\zeta=1}^{400} (\hat{\alpha}_{\zeta,is}^{EST} - \hat{\alpha}_{.,is}^{EST})^2} \Big/ \sum_{is=1}^{12} \hat{\alpha}_{.,is} \,,$$

$$\frac{Bias}{sd} = \sum_{is=1}^{12} \left(\hat{\alpha}_{.,is}^{EST} - \hat{\alpha}_{.,is}\right) \Bigg/ \sum_{is=1}^{12} \sqrt{(400-1)^{-1} \sum_{\zeta=1}^{400} (\hat{\alpha}_{\zeta,is}^{EST} - \hat{\alpha}_{.,is}^{EST})^2} \ ,$$

$$\frac{Bias}{se} = Bias/(20sd) \ ,$$

where ζ indexes the Monte Carlo samples, i denotes an area from a group of areas of sample size s, $\hat{\alpha}_{.,is} = (400)^{-1} \sum_{\zeta=1}^{400} \hat{\alpha}_{\zeta,is}$ is the average of the Monte Carlo prediction error estimators, $\hat{\alpha}_{.,is}^{EST} = (400)^{-1} \sum_{\zeta=1}^{400} \hat{\alpha}_{\zeta,is}^{EST}$ is the average of the bootstrap prediction MSE estimators, and $\hat{\alpha}^{EST} \in \{\hat{\alpha}^{*}, \hat{\alpha}_T^{**}, \hat{\alpha}_C^{**}\}$ is the bootstrap estimator for an area. The estimated prediction MSEs have CVs of about $40\%, 32\%$ and 22% for 200 bootstrap samples for sample sizes 2, 10, and 40, respectively.

In all cases the telescoping double bootstrap, denoted with a subscript T, has lower MSE than the classic double bootstrap, denoted with a subscript C. The estimators $\hat{\alpha}_T^{**}$ and $\hat{\alpha}_C^{**}$ have the same bias if the bound (17) is not used. The double bootstrap reduces the absolute value of the bias for all the sample sizes. However, the absolute bias of the double bootstrap is about 6% of the true value for sample size 2.

Table 2. Monte Carlo properties of prediction MSE estimators
($B_1 = 200, B_2 = 1$ and 400 MC samples, variances multiplied by 10^3)

Size	*Measure*	$\hat{\alpha}^{*}$	$\hat{\alpha}_T^{**}$	$\hat{\alpha}_C^{**}$
2	$V(\hat{\theta} - \theta)$	10.723	10.723	10.723
	RelBias	-0.143	-0.058	-0.062
	$CV(\hat{\alpha})$	0.403	0.456	0.477
	Bias/sd	-0.355	-0.127	-0.130
	Bias/se	-7.097	-2.537	-2.609
10	$V(\hat{\theta} - \theta)$	7.758	7.758	7.758
	RelBias	-0.133	-0.032	-0.039
	$CV(\hat{\alpha})$	0.318	0.365	0.385
	Bias/sd	-0.417	-0.087	-0.102
	Bias/se	-8.336	-1.738	-2.034
40	$V(\hat{\theta} - \theta)$	3.721	3.721	3.721
	RelBias	-0.082	0.016	0.009
	$CV(\hat{\alpha})$	0.222	0.260	0.286
	Bias/sd	-0.372	0.062	0.032
	Bias/se	-7.430	1.249	0.636

The variance of an estimator of the prediction MSE has two components. The first, that we call *between*, is the variance one would obtain if one used an infinite number of bootstrap samples. The second, that we call *within*, is the variability due to the fact that our set of bootstrap samples is a sample of samples.

We estimate these two components using two independent sets of bootstrap samples. That is, for each Monte Carlo sample, we generate two sets of $(B_1 = 100, B_2 = 1)$ samples. The sequences of random seeds $r_{1,k}', r_{2,k}', k = 1, \ldots, B_1$ for the second set are independent of the sequences of random seeds $r_{1,k}, r_{2,k}, k = 1, \ldots, B_1$ for the first set. Let $(\hat{\alpha}^*, \hat{\alpha}^{**}, \hat{\alpha}_T^{**}, \hat{\alpha}_C^{**})$ be the prediction MSE estimates for the first group of bootstrap samples and let $(\hat{\alpha}_2^*, \hat{\alpha}_2^{**}, \hat{\alpha}_{T2}^{**}, \hat{\alpha}_{C2}^{**})$ be the prediction MSE estimates for the second group of bootstrap samples. The within variance component for $B_1 = 100$ is estimated by half of the mean of squared differences between the two prediction MSE estimates,

$$Var_{within}^{EST} = (12)^{-1} \sum_{is=1}^{12} \left((400)^{-1} \sum_{\zeta=1}^{400} (\hat{\alpha}_{\zeta,is}^{EST} - \hat{\alpha}_{2,\zeta,is}^{EST})^2 \right)/2.$$

The variance components for the prediction MSE estimators $(\hat{\alpha}^*, \hat{\alpha}_T^{**}, \hat{\alpha}_C^{**})$ are given in Table 3 for $(B_1 = 100, B_2 = 1)$. The estimated between variance component is the difference between the estimated total variance and the estimated within variance component. The entries in the table are averages over the areas of the same sample size and over the Monte Carlo samples.

Table 3. Estimated variance components for variance of estimated prediction MSE
(Within is for 100 bootstrap samples. All variances have been multiplied by 10^6)

Source of Variation	Size	α^*	α_T^{**}	α_C^{**}
Between	2	17.886	23.040	23.040
Within		2.099	3.903	10.599
Total		19.985	26.943	33.639
Between	10	5.562	7.267	7.267
Within		1.099	2.324	5.376
Total		6.661	9.591	12.643
Between	40	0.544	0.725	0.725
Within		0.264	0.613	1.300
Total		0.808	1.338	2.025

The between component for the level one bootstrap is about 75% of the between component for the double bootstrap procedures. This is not surprising as bias reduction procedures often increase the variance. The bootstrap sampling variance, the within component, for the classic double bootstrap is about five times that of the level one bootstrap. The telescoping bootstrap is 2.1 to 2.7 times as efficient as the classic double bootstrap.

4. Summary

We used a simulation study of a unit level logistic model to compare the impact of different levels of auxiliary information. The minimum mean squared error predictors for the small area means were obtained by conditioning on the information available for an area. That information is the unit level response realizations, the unit level covariate observations, and the sometimes available additional unit level auxiliary information. We considered fixed and random mean models for the covariates, as well as known and unknown distribution for the covariates. The percentage effect on the prediction MSE of including auxiliary information in the estimation is smaller for the random mean model than for the fixed mean model for the covariates because using a random model is equivalent to adding observations.

We presented a parametric double bootstrap procedure for the prediction MSE for the unit level logistic model. The fast double bootstrap procedure, where the number of level-two bootstrap samples is $B_2 = 1$, has superior bootstrap efficiency relative to the classic double bootstrap procedure with $B_2 > 1$. The double bootstrap reduces the prediction MSE estimation bias to less than 50% of the bias of the level-one bootstrap. The double bootstrap increases the standard error of the prediction MSE estimator by 13 to 17% relative to that of the level-one bootstrap.

Acknowledgement

The work of Erciulescu was conducted while she was a student in the Departments of Statistics at Iowa State University. This research was partially supported by USDA NRCS CESU agreement 68-7482-11-534.

REFERENCES

DATTA, G. S., RAO, J. N. K., SMITH, D., (2005). On measuring the variability of small area estimators under a basic area level model, Biometrika, 92, 183−196.

DATTA, G. S., RAO, J. N. K., SMITH, D., (2012). Amendments and Corrections: On measuring the variability of small area estimators under a basic area level model, Biometrika, 99, 2, 509.

DATTA, G. S., RAO, J. N. K., TORABI, M., (2010). Pseudo-empirical Bayes estimation of small area means under a nested error linear regression model with functional measurements errors, Journal of Statistical Planning and Inference, 140, 2952−2962.

DAVIDSON, R., MACKINNON, J. G., (2007). Improving the reliability of bootstrap tests with the fast double bootstrap, Computational Statistics and Data Analysis, 51, 3259−3281.

FULLER, W. A., HARTER, R. M., (1987). The multivariate components of variance model for small area estimation, Small Area Statistics: An International Symposium, Platek, R., Rao, J. N. K., Sarndal, C.E. and Singh, M.P. (Eds.), John Wiley, New York, 103−123.

GHOSH, M., KIM, D., SINHA, K., MAITI, T., KATZOFF, M., PARSONS, V. L., (2009). Hierarchical and Empirical Bayes small domain estimation and proportion of persons without health insurance for minority subpopulations, Survey Methodology, 35, 53−66.

GHOSH, M., SINHA, K., (2007). Empirical Bayes estimation in finite population sampling under functional measurement error models, Journal of Statistical Planning and Inference, 137, 2759−2773.

HALL, P., MAITI, T., (2006). On parametric bootstrap methods for small area prediction, J.R. Statist. Soc. B, 68, 2, 221−238.

PFEFFERMANN, D., CORREA, S., (2012). Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation, Biometrika, 99, 457−472.

TORABI, M., DATTA, G., RAO, J. N. K., (2009). Empirical Bayes Estimation of Small Area Means under a Nested Error Linear Regression Model with Measurement Errors in the Covariates, Scandinavian Journal of Statistics, 36, 355−368.

WANG, J., FULLER, W. A., (2003). The mean squared error of small area spedictors constructed with estimated area variances, Journal of the American Statistical Association, 98, 716−723.

YBARRA, L. M. R., LOHR, S. L., (2008). Small area estimation when auxiliary information is measured with error, Biometrika, 95, 919−931.