# Parametric Bootstrap Procedures for Small Area Prediction Variance

Andreea L. Erciulescu*  Wayne A. Fuller†

**Abstract**
A parametric bootstrap procedure is proposed for the mean squared error of the predictor based on a unit level model. It is demonstrated that the proposed procedure has smaller bootstrap error than a classical double bootstrap procedure with the same number of samples. Applications to a logit model under different types of auxiliary information are discussed.

**Key Words:** Unit Level Model, Parametric Bootstrap, Double Bootstrap, Mean Squared Error of Prediction, Auxiliary Information

## 1. Introduction

The estimation of prediction mean squared error (MSE) for small area models is complicated, particularly in a nonlinear model setting. Double bootstrap is proposed as a method to estimate the prediction mean squared error by Hall and Maiti (2006). They construct nonnegative, bias-corrected MSE estimates using a double bootstrap procedure. They consider area level models and a unit level binomial model with fixed known covariates. Pfeffermann and Correa (2012) suggest a double bootstrap procedure in which the bias in the estimator is estimated as a function of parameters and of a bootstrap estimator of bias. Davidson and MacKinnon (2007) introduce a fast double bootstrap procedure for bootstrap testing.

A number of papers consider sampling variability in the auxiliary variables. Ghosh, Sinha and Kim (2006) consider an area level linear model with random auxiliary variable mean, estimated jointly with the small area mean. Ybarra and Lohr (2008) consider an area level linear model with auxiliary mean estimated with error. Datta, Rao and Torabi (2010), following Ghosh and Sinha (2007), studied a nested error linear regression model with area level covariate subject to measurement error.

We study unit level generalized linear mixed models under situations where the mean of an auxiliary variable is subject to estimation error. We propose a parametric bootstrap procedure for prediction mean squared error estimation and compare the proposed procedure with a classical double bootstrap procedure using a simulation study. Estimation with different types of auxiliary information is illustrated.

*Iowa State University, 1215 Department of Statistics, Ames, IA 50010
†Iowa State University, 1214 Department of Statistics, Ames, IA 50010

## 2. Models

### 2.1 Unit level nonlinear model

Consider the unit level generalized linear mixed model

$$y_{ij} = g(\mathbf{x}_{ij}, \boldsymbol{\beta}, b_i) + e_{ij}, \tag{1}$$

$$\mathbf{x}_{ij} = \boldsymbol{\mu}_x + \boldsymbol{\delta}_i + \boldsymbol{\epsilon}_{ij} =: \boldsymbol{\mu}_{xi} + \boldsymbol{\epsilon}_{ij}, \tag{2}$$

$$\tilde{\mathbf{x}}_{ij'} = \boldsymbol{\mu}_{xi} + \boldsymbol{\epsilon}_{ij'}, \tag{3}$$

$i = 1, ..., m$, where $m$ is the number of areas and $j = 1, ..., n_i$, where $n_i$ is the number of units within area $i$. The vector $(y_{ij}, \mathbf{x}_{ij})$ is observed. In addition to $\mathbf{x}_{ij}$, a vector of auxiliary information, $\tilde{\mathbf{x}}_{ij'}$, is also available where $j' = 1, ..., n_i'$, $n_i'$ is the number of additional observations in area $i$. The vector of random variables $(b_i, \boldsymbol{\delta}_i, e_{ij}, \boldsymbol{\epsilon}_{ij})$ is unobserved, and $\boldsymbol{\beta}$ is the vector of coefficients. Of interest is the small area mean of $\mathbf{y}$

$$\theta_i = \int g(\mathbf{x}_{ij}, \boldsymbol{\beta}, b_i) dF_{\mathbf{x_i}}(\mathbf{x}), \tag{4}$$

where $F_{\mathbf{x_i}}(\mathbf{x})$ is the distribution of $\mathbf{x}$ 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. The nature of the estimation-prediction problem is determined by the distributional properties of the vector $(b_i, \boldsymbol{\delta}_i, e_{ij}, \boldsymbol{\epsilon}_{ij})$.

As an example of model (1), consider a Bernoulli response variable $\mathbf{y}$, 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 $F_{\mathbf{x_i}}$, and let $b_i$ be independent and identically distributed, with a density $f_b$ with mean 0 and variance $\sigma_b^2$. The mean of $y$ given $(\mathbf{x}_{ij}, b_i)$ is

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

where $\mathbf{x}_{ij} = (1, x_{ij}), \tilde{\mathbf{x}}_{ij'} = (1, \tilde{x}_{ij'})$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. We assume that $b_i \sim NI(0, \sigma_b^2),\ \delta_i \sim NI(0, \sigma_\delta^2),\ \epsilon_{ij} \sim NI(0, \sigma_\epsilon^2)$ and that the elements of $(b_i, \delta_i, \epsilon_{ij})$ are mutually independent.

### 2.2 Predictors of $\theta_i$

We present predictions for $\theta_i$, for different cases of auxiliary information, given known parameters.

#### *2.2.1 Known Covariate Mean*

Let $\mu_{xi}$ be known, and let the form of the distribution of $x$ be specified. Then, given known parameters, the minimum mean square error (MMSE) predictor of the small area mean of $y$ is

$$\hat{\theta}_i = E\left[\hat{\theta}(b) | (\mathbf{x}_i, \mathbf{y}_i)\right],$$

where

$$\hat{\theta}(b) = \int_{\mathbf{x}} g(x, \boldsymbol{\beta}, b) dF_{\mathbf{x}}(x)$$

and

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}(b) \prod_{t=1}^{n_i} f(y_{it}|b_i) f(x_{it}|\mu_{xi}) f_b(b_i) db_i}{\int_b \prod_{t=1}^{n_i} f(y_{it}|b_i) f(x_{it}|\mu_{xi}) f_b(b_i) db_i}. \tag{7}$$

In some finite population situations, the entire finite population of $x$ values may be known and the integral expression for $\hat{\theta}(b)$ in (7) is the sum over the population.

#### 2.2.2 *Unknown Random Covariate Mean*

Given known parameters, the MMSE predictor of the small area mean of $y$ is

$$\hat{\theta}_i = E\left[\hat{\theta}(b, \delta)|(\mathbf{x}_i, \mathbf{y}_i)\right],$$

where $\mathbf{x}_i = (x_{i,1}, x_{i,2}, ..., x_{i,n_i}), \mathbf{y}_i = (y_{i,1}, y_{i,2}, ..., y_{i,n_i})$, and

$$\hat{\theta}(b, \delta) = \int g(\mu_x + \delta + \epsilon, \boldsymbol{\beta}, b) dF_{\boldsymbol{\epsilon}}(\epsilon)$$

and

$$\hat{\theta}_i = \frac{\int_b \int_\delta \hat{\theta}(b, \delta) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_i) f(x_{it}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}{\int_b \int_\delta \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_i) f(x_{it}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}. \tag{8}$$

#### 2.2.3 *Unknown Random Covariate Mean,* $\tilde{\mathbf{x}}_i = (\tilde{x}_{i,1}, \tilde{x}_{i,2}, ..., \tilde{x}_{i,n'_i})$ *observed*

Given known parameters, the MMSE predictor of the small area mean of $y$ is

$$\hat{\theta}_i = E\left[\hat{\theta}(b, \delta)|(\mathbf{x}_i, \mathbf{y}_i, \tilde{\mu}_{xi})\right],$$

where

$$\hat{\theta}(b, \delta) = \int g(\mu_x + \delta + \epsilon, \boldsymbol{\beta}, b) dF_{\boldsymbol{\epsilon}}(\epsilon),$$

$$\hat{\theta}_i = \frac{\int_b \int_\delta \hat{\theta}(b, \delta) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_i) f(x_{it}|\delta_i) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}{\int_b \int_\delta \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_i) f(x_{it}|\delta_i) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta) d_{b_i}(b)}. \tag{9}$$

and $\tilde{\mu}_{xi} = (n'_i)^{-1} \sum_{j'=1}^{n'_i} \tilde{x}_{ij'}$.

## 3. Bootstrap estimation

Let $\boldsymbol{\psi}$ be the parameter that defines the distribution of the sample observations. Let $\hat{\boldsymbol{\psi}}$ 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 above, let $\alpha_i$ be the MSE of the prediction error for area $i$, as defined in (5). For the nonlinear small area model with known $\mu_{xi}$, the vector of parameters is $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta}, \sigma_\epsilon^2)$. For the nonlinear small area models with unknown $\mu_{xi}$, the vector of parameters is $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta}, \sigma_\epsilon^2, \mu_x, \sigma_\delta^2)$.

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 $r_{1,1}, r_{1,2}, ..., r_{1,B_1}$. Let $\boldsymbol{\psi}_k^*$ be the estimator of $\boldsymbol{\psi}$ from the $k$th bootstrap sample generated using $DG(\hat{\boldsymbol{\psi}}, r_{1,k})$. The bootstrap estimator of prediction MSE for area $i$ is

$$\hat{\alpha}_i^* = B_1^{-1} \sum_{k=1}^{B_1} (\hat{\theta}_{i,k}^* - \theta_{i,k}^*)^2 =: B_1^{-1} \sum_{k=1}^{B_1} \alpha_{i,k}^* = \bar{\alpha}_i^*, \tag{10}$$

where $\theta^*_{i,k}$ is the true small area mean generated for the $k$th 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 $k$th bootstrap sample. The estimator (10) is called the level-one bootstrap estimator.

In the double bootstrap, a sample of $\alpha^{**}_i$ is generated using $\boldsymbol{\psi}^*$ from the level-one generated sample. The bias in $\alpha^*_i$ is estimated as the difference between $\alpha^{**}_i$ and $\alpha^*_i$. Typically a large number of $\alpha^{**}_i$ is generated for each $\alpha^*_i$ and the average used to estimate the bias as

$$\hat{\Delta}_{\alpha^*_i} = B_1^{-1} \sum_{k=1}^{B_1} (B_2^{-1} \sum_{t=1}^{B_2} \alpha^{**}_{i,k,t} - \alpha^*_{i,k}), \tag{11}$$

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 $r_{2,k,t}, k = 1, 2, ..., B_1, t = 1, 2, ..., B_2$, are independent random numbers, independent of $r_{1,k}$. One form of the bias adjusted estimator is

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

A simpler double bootstrap procedure suggested by Davidson and MacKinnon (2007) generates a single $\alpha^{**}_i$ for each $\alpha^*_i$ . Let $r_{2,1}, r_{2,2}, ..., r_{2,B_1}$ 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 $DG(\boldsymbol{\psi}^*_k, r_{2,k})$. Then a (classic) double bootstrap estimator 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 alternative bootstrap estimator, define $\alpha^*_{i,k,2}$ to be calculated from data generated with $DG(\hat{\boldsymbol{\psi}}, r_{2,k})$. Then a bias adjusted (double bootstrap) estimator is

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

The quantity $\alpha^{**}_{i,k} - \alpha^*_{i,k}$ is a one-degree-of-freedom estimator of the bias. One might use $r_{2,1}$ as $r_{1,2}$, $r_{2,2}$ as $r_{1,3}$, etc. Then, a form of (14) becomes

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

where $\alpha^*_{i,k+1}$ is generated with $DG(\hat{\boldsymbol{\psi}}, r_{1,k+1})$ 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 $\tilde{\alpha}^{**}_{i,T}$ can have smaller variance than $\tilde{\alpha}^{**}_{i,C}$ of (13).

## 4. 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, $(\mathbf{y}, \mathbf{x}, \tilde{\mathbf{x}})$, is generated using model

(1 - 3) with $\sigma_b^2 = 0.25, \mu_x = 0, \sigma_\delta^2 = 0.16$, and $\sigma_\epsilon^2 = 0.36$. The vector of coefficients for the fixed effects is $(\beta_0, \beta_1) = (-0.8, 1)$ and, for each unit, the probability that $y_{ij} = 1$ is

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

The population mean of $g(x_{ij}, \boldsymbol{\beta}, b_i)$ 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: Model (1-2), known auxiliary mean $\mu_{xi}$
- Model 2: Model (1-2), unknown random auxiliary mean $\mu_{xi}$
- Model 3: Model (1-3), unknown random auxiliary mean $\mu_{xi}$, observed $\tilde{\mathbf{x}}$

The models are fitted as generalized linear mixed models (GLMMs), using the *glmer* function in the *lme4* package in R. The true small area mean of $\mathbf{y}$ is given by (4) and the predicted area means of $\mathbf{y}$ are given in (7 - 9), with estimated $(\mu_x, \beta_0, \beta_1, \sigma_b^2, \sigma_\delta^2, \sigma_\epsilon^2)$. The true variance of $\tilde{\mu}_{xi}$ is used in $f(\tilde{\mu}_{xi}|\delta_i)$ of (9). The integrals in (4, 7 - 9) were approximated using a K-point approximation to the normal distribution, with $K = 20$ as in Erciulescu and Fuller (2013).

### 4.1 Refinement of Prediction MSE Estimators

We consider the bootstrap estimators of the prediction MSE of $\theta_i$ given in (10),(13),(15), with $B_1 - 1$ terms in the summation, because of the lagged values in the telescoping bootstrap procedure.

The fact that $\sigma_b^2 \geq 0, \sigma_\delta^2 \geq 0$ and that some unrestricted estimators of $\sigma_b^2, \sigma_\delta^2$ can equal zero must be recognized in constructing estimators. We bound the estimator of $\sigma_b^2$ with 0.003 and bound the estimator of $\sigma_\delta^2$ with 0.002. If $\hat{\sigma}_{b,k}^2 = 0.003$ or $\hat{\sigma}_{b,k}^2 > 0.003$ and $\hat{\sigma}_{b,k}^{2*} \leq 0.003$, we set $\alpha_{i,k}^{**}$ equal to $\alpha_{i,k}^*$. That is, the estimated bias is zero for such samples. The proportion of sample estimators of $\hat{\sigma}_b^2$ that hit the bound is 0.015, the proportion of level one estimators of $\hat{\sigma}_b^{2*}$ that hit the bound is 0.104.

The coefficient of variation for $\hat{\sigma}_b^2$ 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 $\hat{\sigma}_b^2$ based on 400 samples was about $-0.12$, which is approximately equal to eighteen Monte Carlo standard errors.

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

$$\begin{aligned} \hat{\alpha}_{i,C}^{**} &= 0.77\bar{\alpha}_i^*, \text{ if } \bar{\alpha}_i^{*-1}\bar{\alpha}_i^{**} < 0.77 \\ &= \tilde{\alpha}_{i,C}^{**}, \text{ otherwise }, \end{aligned} \tag{17}$$

where $0.77$ is the $0.025$ point of the chi-square distribution with $99$ $(B_1 - 1)$ 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 bound defined in (17) are 0.0375, 0.0325 and 0.0175, 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 bound defined in (17) are 0.0400, 0.0375 and 0.1000, for the areas of sizes 2, 10 and 40, respectively.

### 4.2 MSE for Different Types of Auxiliary Information

Table 1 contains estimates of the $\alpha = MSE$ for three models exploiting 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 auxiliary mean is known. The small area mean predictor for Model 2 is the conditional expected value formula given in (8). Notice that in the construction of the small area predictor for Model 3, given in (9), the conditioning is also on the additional source of information, $\tilde{\mu}_{xi}$, available for the areas. By including the ten additional unit level observations, the estimated MSE is closer to the MSE of the known mean case than to the MSE for the case with no additional information.

**Table 1**: MSE for Different Types Auxiliary Information

| Size | $\bar{y}$ | Model 1 | Model 2 | Model 3 |
|---|---|---|---|---|
| 2 | 101.91 | 9.18 | 13.23 | 10.74 |
| | (1.09) | (0.18) | (0.26) | (0.21) |
| 10 | 20.66 | 7.28 | 8.28 | 7.78 |
| | (0.27) | (0.16) | (0.18) | (0.17) |
| 40 | 5.17 | 3.69 | 3.82 | 3.74 |
| | (0.07) | (0.07) | (0.08) | (0.08) |

Model 1: known $\mu_{xi}$ , Model 2: random $\mu_{xi}$, no $\tilde{\mathbf{x}}$ , Model 3: random $\mu_{xi}$, observed $\tilde{\mathbf{x}}$

### 4.3 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 3, where the area mean $\mu_{xi}$ is random and auxiliary information $\tilde{\mathbf{x}}$ 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. The first line is the Monte Carlo estimate of the prediction MSE, $\hat{\alpha}$ , where the MSE for a Monte Carlo sample is the average of squared prediction errors for the 12 areas with the same sample size, and the samples are generated with the true values. The next four lines are 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

$$
\begin{aligned}
RelBias &= (\hat{\alpha}^{EST}_{\cdot,s} - \hat{\alpha}_{\cdot,s})/\hat{\alpha}_{\cdot,s}, \\
CV &= \sqrt{(400-1)^{-1}\sum_{\zeta=1}^{400}(\hat{\alpha}^{EST}_{\zeta,s} - \hat{\alpha}^{EST}_{\cdot,s})^2}/\hat{\alpha}_{\cdot,s}, \\
Bias/sd &= (\hat{\alpha}^{EST}_{\cdot,s} - \hat{\alpha}_{\cdot,s})/\sqrt{(400-1)^{-1}\sum_{\zeta=1}^{400}(\hat{\alpha}^{EST}_{\zeta,s} - \hat{\alpha}^{EST}_{\cdot,s})^2}, \\
Bias/se &= (\hat{\alpha}^{EST}_{\cdot,s} - \hat{\alpha}_{\cdot,s})/\sqrt{(400(400-1))^{-1}\sum_{\zeta=1}^{400}(\hat{\alpha}^{EST}_{\zeta,s} - \hat{\alpha}^{EST}_{\cdot,s})^2},
\end{aligned}
$$

where $\zeta$ indexes the Monte Carlo samples, $s$ denotes the sample size for a group of areas, $\hat{\alpha}_{\cdot,s} = (400)^{-1}\sum_{\zeta=1}^{400}\hat{\alpha}_{\zeta,s}$ is the average of the Monte Carlo prediction error estimators, $\hat{\alpha}^{EST}_{\cdot,s} = (400)^{-1}\sum_{\zeta=1}^{400}\hat{\alpha}^{EST}_{\zeta,s}$ 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 averaged over the areas with same sample size $s$. For example, $\hat{\alpha}^*_{\zeta,s} = 12^{-1}\sum_{i=1}^{12}\hat{\alpha}^*_{\zeta,s,i}$, where $\hat{\alpha}^*_{\zeta,s,i}$ is the $i$th area in the set with sample size $s$. The estimated prediction MSEs have CV's of about $40\%, 31\%$ and $20\%$ for 100 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 nearly $9\%$ of the true value for sample size 2.

**Table 2**: Monte Carlo properties of prediction MSE estimators
($B_1 = 100, B_2 = 1$ and $400$ MC samples, Variances multiplied by $10^3$)

| size | | $\hat{\alpha}^*$ | $\hat{\alpha}^{**}_T$ | $\hat{\alpha}^{**}_C$ |
|---|---|---|---|---|
| 2 | V($\hat{\theta}-\theta$) | 10.7382 | 10.7382 | 10.7382 |
| | RelBias | -0.1412 | -0.0877 | -0.0866 |
| | CV($\hat{\alpha}$) | 0.3939 | 0.4567 | 0.4592 |
| | Bias/sd | -0.3584 | -0.1920 | -0.1886 |
| | Bias/se | -7.1686 | -3.8407 | -3.7716 |
| 10 | V($\hat{\theta}-\theta$) | 7.7769 | 7.7769 | 7.7769 |
| | RelBias | -0.1300 | -0.0661 | -0.0633 |
| | CV($\hat{\alpha}$) | 0.3069 | 0.3619 | 0.3646 |
| | Bias/sd | -0.4235 | -0.1827 | -0.1736 |
| | Bias/se | -8.4703 | -3.6534 | -3.4724 |
| 40 | V($\hat{\theta}-\theta$) | 3.7431 | 3.7431 | 3.7431 |
| | RelBias | -0.0767 | -0.0225 | -0.0245 |
| | CV($\hat{\alpha}$) | 0.2047 | 0.2398 | 0.2442 |
| | Bias/sd | -0.3748 | -0.0938 | -0.1003 |
| | Bias/se | -7.4950 | -1.8768 | -2.0067 |

The variance of an estimator of the prediction MSE has two components. The first, that we call the 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 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 $r'_{1,k}, r'_{2,k}, k = 1, ..., B_1$ for the second set are independent of the sequences of random seeds $r_{1,k}, r_{2,k}, k = 1, ..., B_1$ 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 difference between the two prediction MSE estimates,

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

The estimated between variance component is the difference between the estimated total variance and the estimated within variance component,

$$Var_{between}^{EST} = Var_{total}^{EST} - Var_{within}^{EST},$$

where $Var_{total}^{EST}$ is the variance estimate for $B_1 = 100$.

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 entries in the table are averages over the Monte Carlo samples and are multiplied by $10^6$. The components may be functions of the parameters.

**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 | $\alpha^*$ | $\alpha_T^{**}$ | $\alpha_C^{**}$ |
|---|---|---|---|---|
| Between | 2 | 17.6962 | 23.5806 | 23.5806 |
| Within | | 0.1982 | 0.4685 | 0.7362 |
| Total | | 17.8944 | 24.0491 | 24.3168 |
| Between | 10 | 5.6088 | 7.6819 | 7.6819 |
| Within | | 0.0897 | 0.2374 | 0.3576 |
| Total | | 5.6985 | 7.9193 | 8.0395 |
| Between | 40 | 0.5617 | 0.7456 | 0.7456 |
| Within | | 0.0252 | 0.0599 | 0.0898 |
| Total | | 0.5869 | 0.8055 | 0.8354 |

The within variance for the double bootstrap is a function of the variance of $\alpha^*$, the variance of $\alpha^{**}$ and the covariance between $\alpha^*$ and $\alpha^{**}$. The within variance component for the prediction MSE estimators $(\hat{\alpha}^*, \hat{\alpha}^{**})$ and the corresponding within covariance component are given in Table 4 for $(B_1 = 1, B_2 = 1)$. Using the entries in this table, one can calculate the variance of estimated prediction MSE for $B_1$ level-one samples combined with $B_2$ level-two samples for each level-one sample,

$$Var_{within}(\hat{\alpha}_C^{**}) = 4B_1^{-1}Var_{within}(\hat{\alpha}^*) - 4B_1^{-1}Cov_{within}(\hat{\alpha}^*, \hat{\alpha}^{**}) + B_1^{-1}B_2^{-1}Var_{within}(\hat{\alpha}^{**}).$$

Table 5 contain estimates of the within variance components for $\alpha_C^{**}$ for different double bootstrap designs, that is for different combinations of $B_1$ level-one samples combined with $B_2$ level-two samples for each level-one sample. The choice of $(B_1, B_2)$ pair has an effect on the estimated within variance of the bootstrap prediction MSE estimator. In Appendix A, we derive the optimal bootstrap design and prove that $B_2 = 1$ is the optimal choice for the number of level-two bootstrap samples.

**Table 4**: Estimated Within Bootstrap Variance and Covariance of the MSE $(B_1 = 1, B_2 = 1)$ $(\times 10^6)$

| Size | $V(\alpha^*)$ | $V(\alpha^{**})$ | $C(\alpha^*, \alpha^{**})_T$ | $C(\alpha^*, \alpha^{**})_C$ |
|---|---|---|---|---|
| 2 | 19.8237 | 36.2553 | 17.1754 | 10.4831 |
| 10 | 8.9691 | 14.9861 | 6.7807 | 3.7763 |
| 40 | 2.5174 | 2.8295 | 1.7268 | 0.9799 |

**Table 5**: Within Bootstrap Variance, Classic $(\times 10^6)$

| Size | $V_{100,1}$ | $V_{1000,1}$ | $V_{5000,1}$ | $V_{100,50}$ | $V_{20,10}$ | $V_{50,4}$ |
|---|---|---|---|---|---|---|
| 2 | 0.7362 | 0.0736 | 0.0147 | 0.3809 | 2.0494 | 0.9285 |
| 10 | 0.3576 | 0.0358 | 0.0072 | 0.2107 | 1.1135 | 0.4904 |
| 40 | 0.0898 | 0.0090 | 0.0018 | 0.0621 | 0.3216 | 0.1371 |

**Table 6**: Within Bootstrap Variance, Telescoping $(\times 10^6)$

| Size | $V_{100,1}$ | $V_{1000,1}$ | $V_{5000,1}$ |
|---|---|---|---|
| 2 | 0.4685 | 0.0469 | 0.0094 |
| 10 | 0.2374 | 0.0237 | 0.0047 |
| 40 | 0.0599 | 0.0060 | 0.0012 |

Table 6 contains estimates of the within variance components for $\alpha_T^{**}$ for different double bootstrap designs, $(B_1 = 100, 1000, 5000, B_2 = 1)$.

Consider the predictor MSE estimators for the areas of size $n_i = 2$. Using the results in tables 5 and 6, we conclude that increasing the number of bootstrap samples to $B_1 = 5000$ reduces the within variance component to about $0.06\%$ of the total variance for the classic method and to about $0.04\%$ of the total variance for the telescoping method. For the classic bootstrap method based on a total of 200 samples, the estimated within variance component is about $3\%$ of the total variance for the design $(B_1 = 100, B_2 = 1)$, about $4\%$ of the total variance for the design $(B_1 = 50, B_2 = 4)$ and about $8\%$ of the total variance for the design $(B_1 = 20, B_2 = 10)$. Also, for the classic bootstrap method based on a total of 10000 samples, the estimated within variance component is about $0.3\%$ of the total variance for the design $(B_1 = 5000, B_2 = 1)$ and about $2\%$ of the total variance for the design $(B_1 = 100, B_2 = 50)$.

### 4.4 Equal Efficiency Bootstrap Designs

We give bootstrap sample sizes such that the bootstrap variance of the estimated prediction MSE is the same under different bootstrap sampling procedures.

Table 7 contains the number of level-one bootstrap samples needed in the classic bootstrap method in order to produce prediction MSE estimates as efficient as the prediction MSE estimates produced using the telescoping bootstrap method with $(B_1 = 100, B_2 = 1)$. The last column in Table 7 contains the total number of bootstrap samples for each procedure, for each design.

**Table 7**: Equal efficiency bootstrap procedures, Model 3

| Bootstrap Method/Design | Size | $B_1$ | Total |
|---|---|---|---|
| Telescoping $(100, 1)$ | 2 | 100 | 200 |
| Classic $(B_1, 1)$ | | 159 | 318 |
| Classic $(B_1, 50)$ | | 84 | 4284 |
| Telescoping $(100, 1)$ | 10 | 100 | 200 |
| Classic $(B_1, 1)$ | | 157 | 314 |
| Classic $(B_1, 50)$ | | 90 | 4590 |
| Telescoping $(100, 1)$ | 40 | 100 | 200 |
| Classic $(B_1, 1)$ | | 150 | 300 |
| Classic $(B_1, 50)$ | | 103 | 5253 |

## 5. Conclusions

We present a parametric double bootstrap procedure for estimation of the mean squared error of the predictor for a unit level nonlinear model. We show that the fast double bootstrap procedure, where the number of level-two bootstrap samples is $B_2 = 1$, has superior bootstrap efficiency relative to classic double bootstrap procedure with $B_2 > 1$. The double bootstrap reduces the prediction MSE estimation bias to about $50\%$ of that of the level one bootstrap. The double bootstrap increases the standard error of the prediction MSE estimator by about $15$ to $20\%$ relative to that of the level one bootstrap.

We used a simulation study of a unit level binomial model to compare the impact of different levels of auxiliary information. The estimated minimum mean square error estimates for the small area means were obtained by conditioning on the observations, including the area means of the auxiliary information. The results indicate that the random model for the covariates has potential to reduce the prediction MSE relative to that of the fixed model when additional auxiliary information is available and included in the estimation.

## REFERENCES


Battese, G.E., Harter, R.M., Fuller, W.A. (1988), "An error component model for prediction of county crop areas using survey and satellite data," *Journal of the American Statistical Association*, 83, 28-36.

Berrige D. and Crouchley R. (2011), *Multivariate Generalized Linear Mixed Models Using R*, CRC Press.

Beale, E.M.L. (1962). "Some uses of computers in operations research." *Industrielle Organisation*, 31, 51-52.

Datta, G.S. and Lahiri, P. (2000), "A unified measure of uncertainty of best linear unbiased predictors in small area estimation problems," *Statistica Sinica*, 10, 613-627.

Datta, G.S., Rao, J.N.K. and 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. and 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. and MacKinnon, J.G. (2007), "Improving the reliability of bootstrap tests with the fast double bootstrap," *Computational Statistics and Data Analysis*, 51, 3259-3281.

Erciulescu, A.L. and Fuller, W.A. (2013), "Small Area Prediction of the Mean of a Binomial Random Variable," *Survey Research Methods Section, JSM Proceedings*, 855-863.

Ghosh, M. and 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. and Maiti T. (2006), "On parametric bootstrap methods for small area prediction,"*J.R. Statist. Soc. B*, 68, 2, 221-238.

Harville D.A., (1985), "Decomposition of Prediction Error," *Journal of the American Statistical Association*, 80, 389, 132-138.

Jeong K.M., Son J. (2009), "Estimation of Small Area Proportions Based on Logistic Mixed Model," *The Korean Journal of Applied Statistics*, 22(1), 153-161.

Kackar, R. and Harville, D.A. (1984) "Approximations for standard errors of estimators of fixed and random effects in mixed linear models, " *Journal of the American Statistical Association*, 79, 853-862.

Pfeffermann, D. and Correa, S. (2012)."Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation." *Biometrika*, 99, 457-472.

Pfeffermann D, Glickman H. (2004), "Mean Square Error Approximation in Small Area Estimation by Use of Parametric and Nonparametric Bootstrap," *ASA Section on Survey Research Methods*, 4167-4178.

Prasad, N.G.N. and Rao, J.N.K. (1990), "The estimation of mean squared error of small area estimators.," *Journal of the American Statistical Association*, 85, 163-171.

Rao J.N.K. (2003), "Small Area Estimation," *Wiley Series in Survey Methodology*.

Stroup, W.W. (2012), "Generalized Linear Mixed Models: Modern Concepts, Methods and Applications," *Chapman and Hall*, CRC Texts in Statistical Science, Taylor and Francis.

Wang, J. and 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. and Lohr, S.L. (2008), " Small area estimation when auxiliary information is measured with error," *Biometrika*, 95, 919-931.

## 6. Appendix A, Optimal bootstrap design

The number and type of bootstrap samples can change the within sample variance, but will have little impact on the between sample variance. The variance of the double bootstrap procedure (14) that uses $B_1$ level-one samples and $B_2$ level-two samples is

$$V(\hat{\alpha}_C) = 4B_1^{-1}V(\hat{\alpha}^*) - 4B_1^{-1}C(\hat{\alpha}^*, \hat{\alpha}^{**}) + B_1^{-1}B_2^{-1}V(\hat{\alpha}^{**}).$$

We would like to minimize $V(\hat{\alpha}_C)$ with respect to the restriction $B_1B_2 + B_1 = k$, where k is a constant representing the total number of bootstrap samples.

To consider this problem as a Lagrangian multiplier problem, let

$$L(B_1, B_2, \lambda) = 4B_1^{-1}V(\hat{\alpha}^*) - 4B_1^{-1}C(\hat{\alpha}^*, \hat{\alpha}^{**}) + B_1^{-1}B_2^{-1}V(\hat{\alpha}^{**}) + \lambda(B_1B_2 + B_1 - k)$$

where $\lambda$ the Lagrangian multiplier. The resulting system of three equations is:

$$\begin{array}{rcl} 0 & = & -4B_1^{-2}V(\hat{\alpha}^*) + 4B_1^{-2}C(\hat{\alpha}^*, \hat{\alpha}^{**}) - B_1^{-2}B_2^{-1}V(\hat{\alpha}^{**}) + \lambda(B_2 + 1) \\ 0 & = & -B_1^{-1}B_2^{-2}V(\hat{\alpha}^{**}) + \lambda B_1 \\ 0 & = & B_1B_2 + B_1 - k. \end{array}$$

The solution is

$$B_2 = \sqrt{\frac{V(\hat{\alpha}^{**})}{4(V(\hat{\alpha}^*) - C(\hat{\alpha}^*, \hat{\alpha}^{**}))}},$$

and

$$B_1 = k\left(\sqrt{\frac{V(\hat{\alpha}^{**})}{4(V(\hat{\alpha}^*) - C(\hat{\alpha}^*, \hat{\alpha}^{**}))}} + 1\right)^{-1}$$

For the parameters in the simulation study, $B_2 = 1$ is the optimal choice.