Statistica Sinica **35** (2025), 1583-1604
doi:https://doi.org/10.5705/ss.202023.0098

AN UNBIASED PREDICTOR FOR SKEWED RESPONSE VARIABLE WITH MEASUREMENT ERROR IN COVARIATE

Sepideh Mosaferi*, Malay Ghosh and Shonosuke Sugasawa

University of Massachusetts Amherst, University of Florida and Keio University

Abstract: We introduce a new small area predictor when the Fay–Herriot normal error model is fitted to a logarithmically transformed response variable, and the covariate is measured with error. This framework has been previously studied by Mosaferi, Ghosh and Steorts (2023). The empirical predictor given in their manuscript cannot perform uniformly better than the direct estimator. Our proposed predictor in this manuscript is unbiased and can perform uniformly better than the one proposed in Mosaferi, Ghosh and Steorts (2023). We derive an approximation of the mean squared error (MSE) for the predictor. The prediction intervals based on the MSE suffer from coverage problems. Thus, we propose a non-parametric bootstrap prediction interval which is more accurate. This problem is of great interest in small area applications since statistical agencies and agricultural surveys are often asked to produce estimates of right skewed variables with covariates measured with errors. With Monte Carlo simulation studies and two Census Bureau's data sets, we demonstrate the superiority of our proposed methodology.

Key words and phrases: Bayes estimator, prediction interval, transformation.

1. Introduction

Small area estimation concerns producing estimates or predictions of means, totals or quantiles for each of a finite collection of geographic regions, where there are a small number of sampled units in each individual region (area). Classical models used in small area estimation take the form of mixed linear models that result from the concatenation of a model for error in direct sample-based estimators for each area and an additional model that connects areas through the use of covariates and area-specific random effects.

These *linking models* take the direct estimators to be linear combinations of covariates and random effects. We focus here on what is called the *area level model* (Ghosh and Rao, 1994; Pfefferman, 2013; Rao and Molina, 2015, Chap. 4) which uses covariates at the level of the areas. Recently, Mosaferi, Ghosh and Steorts (2023) proposed a model of the Fay–Herriot type and developed an empirical predictor for small area quantities that they are right skewed.

*Corresponding author. E-mail: smosaferi@umass.edu

A complication that arises is that the area-level covariates to be used can be the result of survey sampling (Ybarra and Lohr, 2008), thus producing a small area model with measurement error in the covariates. The predictor given in their manuscript cannot perform uniformly better than the direct estimator because of bias issues. In this manuscript, we propose a new unbiased predictor, which can perform uniformly better than that of Mosaferi, Ghosh and Steorts (2023) and the direct estimator.

Much work with small area models has been devoted to the estimation of MSE. Prasad and Rao (1990) derived a closed form approximation for the estimator of the MSE of an empirical Bayes (EB) predictor under the assumption of normality. Jiang, Lahiri and Wan (2002) proposed a jackknife estimator based on a decomposition of MSE where the leading term is approximately unbiased and does not depend on the area-specific random effects. Butar and Lahiri (2003) developed a bootstrap estimator of MSE.

Here, we derive an approximation of the MSE for the predictor as well as the jackknife estimator of MSE. Prediction intervals based on these suffer from inadequate coverage probabilities. In order to address this shortcoming, we develop prediction intervals based on a non-parametric bootstrap method. In the rest of this section, we list some of the previous works in the literature and highlight our contributions.

1.1. Prior work

Molina and Martín (2018) and Berg and Chandra (2014) worked on the log-transformation model and proposed an EB predictor for the value of the variable of interest for out-of-sample individuals (and for small area means) in a nested-error regression model where no measurement error is assumed present in the covariates. The analytical MSE given in Molina and Martín (2018) has a complex form. Thus, the authors proposed a parametric bootstrap procedure for estimation of the uncertainty following Butar and Lahiri (2003). Slud and Maiti (2006) proposed a large sample approximation to the MSE of the predictor for the transformed Fay–Herriot model without measurement error.

1.2. Our contributions

In this paper, we make several contributions to the literature. First, unlike the earlier works given in Section 1.1, we assume the available covariate in the model is measured with error. Second, we propose a new small area predictor for the skewed response variable in the original scale at the area-level instead of unit-level under presence of measurement error in covariate and make comparisons with the earlier predictor proposed by Mosaferi, Ghosh and Steorts (2023). Third, we explain how to estimate the unknown parameters using unbiased score functions from the marginal likelihood. Finally, we derive an approximation for

the MSE of our proposed predictor and develop prediction intervals based on nonparametric bootstrap techniques.

The rest of the paper is organized as follows. In Section 2, we apply the Fay and Herriot (1979) model to the transformed data with measurement error in covariate and formulate the problem. In Section 3, we derive a new predictor for the response variable in the proposed modeling framework. In Section 4, we explain how to estimate the unknown parameters in the model. In Section 5, we derive an estimator of the MSE of the predictor.

In Section 6, we construct non-parametric bootstrap prediction intervals. In Section 7, using a Monte Carlo simulation study, we make comparisons with other predictors given in the literature. In Section 8, we illustrate our methodology using two data sets from the Census Bureau. The related discussions and possible extensions are given in Section 9. Technical details and additional numerical results are in the Supplementary Material. All the R code implementing the proposed methodology is available at Github repository `https://github.com/SepidehMosaferi/UnbiasedPredictor_SkewedData`.

2. Transformed Fay–Herriot Model with Measurement Errors

Assume response variables y_i $(i = 1, \ldots, m)$ are right skewed. Thus, the log transformation of y_i can stabilize the variation. Let $z_i = \log(y_i)$ and $z_i = \phi_i + e_i$, where $\phi_i = \log(\theta_i)$ such that θ_i is unknown and $e_i \overset{\text{ind.}}{\sim} N(0, \psi_i)$ is the sampling error. We further define $\phi_i = \beta_0 + \beta_1 x_i + v_i$, where the covariate x_i is not observed, and what one observes is the W_i. The linking error is $v_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_v^2)$. The error terms (e_i, v_i) are mutually independent per each i-th small area.

Then, the transformed Fay–Herriot model with measurement error in covariate can be presented with the following hierarchical set-up

$$\begin{aligned} z_i|\phi_i &\overset{\text{ind.}}{\sim} N(\phi_i, \psi_i) \\ \phi_i &\overset{\text{ind.}}{\sim} N(\beta_0 + \beta_1 x_i, \sigma_v^2) \\ W_i &\overset{\text{ind.}}{\sim} N(x_i, C_i), \quad i = 1, \ldots, m. \end{aligned} \tag{2.1}$$

The (z_i, ϕ_i) are assumed to be independent of the W_i. This is because the former constitutes the sampling and linking model, while the later brings in the measurement error part. Here, following Ybarra and Lohr (2008), σ_v^2 is unknown but the sampling variances ψ_i and C_i are assumed to be known, which can be obtained from the asymptotic variances of transformed direct estimates (Carter and Rolph, 1974; Efron and Morris, 1975; Fay and Herriot, 1979).

The primary parameter of interest in the original scale is $\theta_i \equiv \exp(\beta_0 + \beta_1 x_i + v_i)$. Prediction of $\phi_i = \log(\theta_i)$, where $\log(\theta_i) \equiv \beta_0 + \beta_1 x_i + v_i$, is identical to the problem of Ybarra and Lohr (2008).

We will adopt an empirical Bayes approach for estimation of the θ_i. We

will assume a flat prior for all the x_i's, but then estimate β_0, β_1, and σ_v^2 from the resulting marginal likelihood. To this end, first observe that writing $\gamma_i^\star = \sigma_v^2/(\sigma_v^2+\psi_i)$, the conditional posterior distributions

$$\phi_i|\beta_0,\beta_1,\sigma_v^2,x_i,z_i,W_i \overset{\text{ind.}}{\sim} N\{\gamma_i^\star z_i + (1-\gamma_i^\star)(\beta_0+\beta_1 x_i), \gamma_i^\star\psi_i\},$$

and $x_i|\beta_0,\beta_1,\sigma_v^2,z_i,W_i$ are mutually independent, where

$$\pi(x_i|\beta_0,\beta_1,\sigma_v^2,z_i,W_i) \propto (\sigma_v^2+\psi_i)^{-1/2}\exp\left[-\frac{1}{2}\left\{\frac{(z_i-\beta_0-\beta_1x_i)^2}{\sigma_v^2+\psi_i}+\frac{(W_i-x_i)^2}{C_i}\right\}\right].$$

Next we use the identity,

$$\frac{(z_i-\beta_0-\beta_1x_i)^2}{\sigma_v^2+\psi_i}+\frac{(W_i-x_i)^2}{C_i} = \{\beta_1^2C_i+\sigma_v^2+\psi_i\}^{-1}\times(z_i-\beta_0-\beta_1W_i)^2$$
$$+\left\{\frac{\beta_1^2}{\sigma_v^2+\psi_i}+C_i^{-1}\right\}\times\left[x_i-\left\{\frac{\beta_1(z_i-\beta_0)}{\sigma_v^2+\psi_i}+\frac{W_i}{C_i}\right\}\Big/\left\{\frac{\beta_1^2}{\sigma_v^2+\psi_i}+\frac{1}{C_i}\right\}\right]^2. \quad (2.2)$$

Now writing $S_i(\beta_1,\sigma_v^2)=\beta_1^2C_i+\sigma_v^2+\psi_i$, one gets

$$x_i|\beta_0,\beta_1,\sigma_v^2,z_i,W_i \overset{\text{ind.}}{\sim} N\left\{\frac{\beta_1C_i(z_i-\beta_0)+(\sigma_v^2+\psi_i)W_i}{S_i(\beta_1,\sigma_v^2)}, \frac{\sigma_v^2+\psi_i}{S_i(\beta_1,\sigma_v^2)}\right\}.$$

Accordingly,

$$E(\phi_i|\beta_0,\beta_1,\sigma_v^2,z_i,W_i) = E\Big\{E(\phi_i|\beta_0,\beta_1,\sigma_v^2,x_i,z_i,W_i)|\beta_0,\beta_1,\sigma_v^2,z_i,W_i\Big\}$$
$$=\frac{\sigma_v^2}{\sigma_v^2+\psi_i}z_i+\frac{\psi_i}{\sigma_v^2+\psi_i}\left[\beta_0+\beta_1\left\{\frac{\beta_1C_i(z_i-\beta_0)+(\sigma_v^2+\psi_i)W_i}{S_i(\beta_1,\sigma_v^2)}\right\}\right]$$
$$=\frac{\sigma_v^2\,S_i(\beta_1,\sigma_v^2)+\psi_i\beta_1^2C_i}{(\sigma_v^2+\psi_i)S_i(\beta_1,\sigma_v^2)}z_i+\frac{\psi_i}{\sigma_v^2+\psi_i}\left\{\frac{\beta_0(\sigma_v^2+\psi_i)}{S_i(\beta_1,\sigma_v^2)}+\frac{\beta_1(\sigma_v^2+\psi_i)W_i}{S_i(\beta_1,\sigma_v^2)}\right\}$$
$$=\frac{\beta_1^2C_i+\sigma_v^2}{S_i(\beta_1,\sigma_v^2)}z_i+\frac{\psi_i(\beta_0+\beta_1W_i)}{S_i(\beta_1,\sigma_v^2)}=\tilde\gamma_iz_i+(1-\tilde\gamma_i)(\beta_0+\beta_1W_i),$$

where $\tilde\gamma_i\equiv(\beta_1^2C_i+\sigma_v^2)/S_i(\beta_1,\sigma_v^2)$. Further,

$$V(\phi_i|\beta_0,\beta_1,\sigma_v^2,z_i,W_i) = E\Big\{V(\phi_i|\beta_0,\beta_1,\sigma_v^2,x_i,z_i,W_i)|\beta_0,\beta_1,\sigma_v^2,z_i,W_i\Big\}$$
$$+V\Big\{E(\phi_i|\beta_0,\beta_1,\sigma_v^2,x_i,z_i,W_i)|\beta_0,\beta_1,\sigma_v^2,z_i,W_i\Big\}$$
$$=\gamma_i^\star\psi_i+(1-\gamma_i^\star)^2\beta_1^2\frac{(\sigma_v^2+\psi_i)C_i}{S_i(\beta_1,\sigma_v^2)}$$
$$=\frac{\sigma_v^2\psi_i}{\sigma_v^2+\psi_i}+\frac{\psi_i^2}{(\sigma_v^2+\psi_i)^2}\beta_1^2C_i\frac{\sigma_v^2+\psi_i}{S_i(\beta_1,\sigma_v^2)}$$
$$=\frac{1}{\sigma_v^2+\psi_i}\left\{\sigma_v^2\psi_i+\frac{\psi_i^2\beta_1^2C_i}{S_i(\beta_1,\sigma_v^2)}\right\}$$

$$
\begin{aligned}
&= \frac{\psi_i}{(\sigma_v^2+\psi_i)S_i(\beta_1,\sigma_v^2)}\Big\{\sigma_v^2(\beta_1^2C_i+\sigma_v^2+\psi_i)+\beta_1^2C_i\psi_i\Big\}\\
&= \frac{\psi_i}{(\sigma_v^2+\psi_i)S_i(\beta_1,\sigma_v^2)}\Big\{\sigma_v^2(\sigma_v^2+\psi_i)+\beta_1^2C_i(\sigma_v^2+\psi_i)\Big\}\\
&= \tilde{\gamma}_i\psi_i.
\end{aligned}
$$

This leads to the Bayes estimator

$$
\begin{aligned}
\tilde{\theta}_{i,A} &= E\{\exp(\phi_i)|\beta_0,\beta_1,\sigma_v^2,z_i,W_i\}\\
&= \exp\left\{\tilde{\gamma}_i z_i+(1-\tilde{\gamma}_i)(\beta_0+\beta_1W_i)+\frac{\tilde{\gamma}_i\psi_i}{2}\right\},
\end{aligned}
$$

proposed by Mosaferi, Ghosh and Steorts (2023). Also,

$$
\begin{aligned}
V\{\exp(\phi_i)|\beta_0,\beta_1,\sigma_v^2,z_i,W_i\} &= \exp(\tilde{\gamma}_i\psi_i)\{\exp(\tilde{\gamma}_i\psi_i)-1\}\\
&\quad\times\exp[2\{\tilde{\gamma}_i z_i+(1-\tilde{\gamma}_i)(\beta_0+\beta_1W_i)\}].
\end{aligned}
$$

Further, from (2.2), the marginal likelihood of $(\beta_0,\beta_1,\sigma_v^2)$ is given by

$$
L_M(\beta_0,\beta_1,\sigma_v^2)=\prod_{i=1}^{m}S_i^{-1/2}(\beta_1,\sigma_v^2)\,\exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\frac{\tau_i^2(\beta_0,\beta_1)}{S_i(\beta_1,\sigma_v^2)}\right\},\tag{2.3}
$$

where we define $\tau_i(\beta_0,\beta_1)=z_i-\beta_0-\beta_1W_i$. In the reminder of this paper, we will work with the score functions based on the marginal likelihood (2.3) to estimate the vector of unknown parameters.

3. Optimal Predictor in the Original Scale

It is clear that $E(\tilde{\theta}_{i,A})=E(\theta_i)=\exp(\beta_0+\beta_1x_i+\sigma_v^2/2)$ when $C_i=0$. The above equality, however, is not true in general. To this end, we prove the following theorem.

Theorem 1. $E(\tilde{\theta}_{i,A})=\exp[\beta_0+\beta_1x_i+\{\tilde{\gamma}_i(\sigma_v^2+\psi_i)+(1-\tilde{\gamma}_i)\beta_1^2C_i\}/2]$.

Proof. See Supplementary Material.

In view of Theorem 1,

$$
\begin{aligned}
\frac{E(\tilde{\theta}_{i,A})}{E(\theta_i)} &= \exp\left[\frac{1}{2}\{\tilde{\gamma}_i(\sigma_v^2+\psi_i)+(1-\tilde{\gamma}_i)\beta_1^2C_i-\sigma_v^2\}\right]\\
&= \exp\left(\frac{1}{2}d_i\right),\quad\text{(say)}
\end{aligned}
$$

that is

$$E\Big\{\tilde{\theta}_{i,A}\exp\Big(-\frac{1}{2}d_i\Big)\Big\} = E(\theta_i).$$

Therefore, our proposed unbiased predictor is

$$\tilde{\theta}_{i,B} = \tilde{\theta}_{i,A}\exp\Big(-\frac{1}{2}\,d_i\Big). \tag{3.1}$$

With some algebra, d_i can be simplified as $d_i = 2\psi_i\beta_1^2 C_i/S_i(\beta_1,\sigma_v^2)$.

Remark 1. When $C_i = 0$, the true value of x_i can be used for the optimal predictor. By substituting $C_i = 0$ into $\tilde{\theta}_{i,B}$, the optimal predictor is $\tilde{\theta}^0_{i,B} = \exp\{\gamma_i^\star z_i + (1-\gamma_i^\star)(\beta_0+\beta_1 x_i) + \gamma_i^\star\psi_i/2\}$, which is same as the predictor given in Slud and Maiti (2006) and is also identical to predictor $\tilde{\theta}_{i,A}$.

4. Estimating the Unknown Parameters

We are interested in obtaining estimates of the vector of unknown parameters $\boldsymbol{\omega} = (\beta_0,\beta_1,\sigma_v^2)'$. Note that we do not directly use the partial derivatives of the marginal likelihood given in expression (2.3) since they are not unbiased. We denote the log-likelihood of $L_M(.)$ in (2.3) by $\ell_M(\boldsymbol{\omega})$. The score functions of $\ell_M(\boldsymbol{\omega})$ can be defined as follows:

$$U(\boldsymbol{\omega}) = (U_1(\boldsymbol{\omega}),U_2(\boldsymbol{\omega}),U_3(\boldsymbol{\omega}))' = \Big(\frac{\partial\ell_M(\boldsymbol{\omega})}{\partial\beta_0},\frac{\partial\ell_M(\boldsymbol{\omega})}{\partial\beta_1},\frac{\partial\ell_M(\boldsymbol{\omega})}{\partial\sigma_v^2}\Big)'.$$

These score functions are biased for estimating the unknown parameters $\boldsymbol{\omega}$. Thus, we define the unbiased score functions as follows

$$\tilde{U}(\boldsymbol{\omega}) = (\tilde{U}_1(\boldsymbol{\omega}),\tilde{U}_2(\boldsymbol{\omega}),\tilde{U}_3(\boldsymbol{\omega}))' = U(\boldsymbol{\omega}) - E\{U(\boldsymbol{\omega})\}, \tag{4.1}$$

such that $E[\tilde{U}(\boldsymbol{\omega})] = 0$. Original score functions for $U(\boldsymbol{\omega})$ are

$$\begin{aligned}
U_1(\boldsymbol{\omega}) &= \sum_{i=1}^m S_i^{-1}(\beta_1,\sigma_v^2)\tau_i(\beta_0,\beta_1),\\
U_2(\boldsymbol{\omega}) &= -\sum_{i=1}^m S_i^{-1}(\beta_1,\sigma_v^2)\beta_1 C_i + \sum_{i=1}^m S_i^{-1}(\beta_1,\sigma_v^2)W_i\tau_i(\beta_0,\beta_1)\\
&\quad + \sum_{i=1}^m S_i^{-2}(\beta_1,\sigma_v^2)\tau_i^2(\beta_0,\beta_1)\beta_1 C_i,\\
U_3(\boldsymbol{\omega}) &= -\frac{1}{2}\sum_{i=1}^m S_i^{-1}(\beta_1,\sigma_v^2) + \frac{1}{2}\sum_{i=1}^m S_i^{-2}(\beta_1,\sigma_v^2)\tau_i^2(\beta_0,\beta_1),
\end{aligned}$$

and their expected values are

$$E\{U_1(\boldsymbol{\omega})\} = 0,\quad E\{U_2(\boldsymbol{\omega})\} = -\sum_{i=1}^m S_i^{-1}(\beta_1,\sigma_v^2)\beta_1 C_i,\quad E\{U_3(\boldsymbol{\omega})\} = 0,$$

where we emphasize that $E\{W_i\tau_i(\beta_0,\beta_1)\} = -\beta_1 C_i$.

Using expression (4.1), the unbiased score functions for estimating $\boldsymbol{\omega}$ are

$$\begin{aligned}
\tilde{U}_1(\boldsymbol{\omega}) &= \sum_{i=1}^{m} S_i^{-1}(\beta_1,\sigma_v^2)\tau_i(\beta_0,\beta_1) = 0,\\
\tilde{U}_2(\boldsymbol{\omega}) &= \sum_{i=1}^{m} S_i^{-1}(\beta_1,\sigma_v^2)W_i\tau_i(\beta_0,\beta_1) + \sum_{i=1}^{m} S_i^{-2}(\beta_1,\sigma_v^2)\tau_i^2(\beta_0,\beta_1)\beta_1 C_i = 0,\\
\tilde{U}_3(\boldsymbol{\omega}) &= -\frac{1}{2}\sum_{i=1}^{m} S_i^{-1}(\beta_1,\sigma_v^2) + \frac{1}{2}\sum_{i=1}^{m} S_i^{-2}(\beta_1,\sigma_v^2)\tau_i^2(\beta_0,\beta_1) = 0. \qquad (4.2)
\end{aligned}$$

One can solve equations given in (4.2) numerically to find the estimates of unknown parameters. Given the current estimate $\boldsymbol{\omega}^{(r)}$ of $\boldsymbol{\omega}$, by replacing $S_i(\beta_1,\sigma_v^2)$ with $S_i(\beta_1^{(r)},\sigma_v^{2(r)})$ and β_1 with $\beta_1^{(r)}$ in $\tilde{U}_1(\boldsymbol{\omega})$, the solution of β_0 is given by

$$\beta_0 = \left\{\sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r)},\sigma_v^{2(r)})\right\}^{-1}\left\{\sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r)},\sigma_v^{2(r)})(z_i - \beta_1^{(r)}W_i)\right\}.$$

Similarly, by replacing $S_i(\beta_1,\sigma_v^2)$ with $S_i(\beta_1^{(r)},\sigma_v^{2(r)})$ and β_0 with $\beta_0^{(r)}$ in $\tilde{U}_2(\boldsymbol{\omega})$, the solution of β_1 is given by

$$\begin{aligned}
\beta_1 = &\left\{\sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r)},\sigma_v^{2(r)})W_i^2 - \sum_{i=1}^{m} S_i^{-2}(\beta_1^{(r)},\sigma_v^{2(r)})\tau_i^2(\beta_0^{(r)},\beta_1^{(r)})C_i\right\}^{-1}\\
&\times\left\{\sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r)},\sigma_v^{2(r)})W_i(z_i - \beta_0^{(r)})\right\}.
\end{aligned}$$

Therefore, we can develop the iterative Algorithm 1 to solve the estimating equations.

Remark 2. Note that the updating process β_1 in Algorithm 1 is similar to the modified least squares estimator used in Ybarra and Lohr (2008) in which a fixed weight is used instead of $S_i^{-1}(\beta_1,\sigma_v^2)$. Thus, the updating process in the iterative algorithm can be regarded as iteratively reweighted least squares.

Theorem 2. *Define* $\tilde{\sigma}_{ci}^2 := C_i(\sigma_v^2+\psi_i)S_i^{-1}(\beta_1,\sigma_v^2)$. *Based on the properties of the unbiased estimating functions, one can obtain* $[\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}] \xrightarrow{D} N_3(\mathbf{0}, I_{\boldsymbol{\omega}}^{-1})$ *as* $m \to \infty$, *where*

$$I_{\boldsymbol{\omega}} = \begin{bmatrix} \sum_{i=1}^{m} S_i^{-1}(\beta_1,\sigma_v^2) & \sum_{i=1}^{m} S_i^{-1}(\beta_1,\sigma_v^2)x_i & 0\\ \sum_{i=1}^{m} S_i^{-1}(\beta_1,\sigma_v^2)x_i & \sum_{i=1}^{m} S_i^{-1}(\beta_1,\sigma_v^2)(x_i^2+\tilde{\sigma}_{ci}^2) & 0\\ 0 & 0 & \sum_{i=1}^{m} S_i^{-2}(\beta_1,\sigma_v^2)/2 \end{bmatrix}.$$

Proof. See Supplementary Material.

Algorithm 1: Iterative algorithm for solving the unbiased estimating equations (4.2).

1. Set the initial value $\boldsymbol{\omega}^{(0)}$ and $r = 0$.

for $r = 1, \ldots, R$ **do**

2. Update β_0 as

$$\beta_0^{(r+1)} = \left\{ \sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r)}, \sigma_v^{2(r)}) \right\}^{-1} \left\{ \sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r)}, \sigma_v^{2(r)})(z_i - \beta_1^{(r)} W_i) \right\}.$$

3. Update β_1 as

$$\beta_1^{(r+1)} = \left\{ \sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r)}, \sigma_v^{2(r)}) W_i^2 - \sum_{i=1}^{m} S_i^{-2}(\beta_1^{(r)}, \sigma_v^{2(r)}) \tau_i^2(\beta_0^{(r+1)}, \beta_1^{(r)}) C_i \right\}^{-1} \times \left\{ \sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r)}, \sigma_v^{2(r)}) W_i (z_i - \beta_0^{(r+1)}) \right\}.$$

4. Update σ_v^2 (obtain $\sigma_v^{2(r+1)}$) by solving the equation

$$-\frac{1}{2} \sum_{i=1}^{m} S_i^{-1}(\beta_1^{(r+1)}, \sigma_v^2) + \frac{1}{2} \sum_{i=1}^{m} S_i^{-2}(\beta_1^{(r+1)}, \sigma_v^2) \tau_i^2(\beta_0^{(r+1)}, \beta_1^{(r+1)}) = 0.$$

5. If $\|\boldsymbol{\omega}^{(r+1)} - \boldsymbol{\omega}^{(r)}\| < \varepsilon$ with tolerance $\varepsilon > 0$, $\boldsymbol{\omega}^{(r+1)}$ is the final estimate; otherwise, set $r = r + 1$ and go back to Step 2.

end

Result: $\hat{\boldsymbol{\omega}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_v^2)'$.

5. Mean Squared Error Formulae

In this Section, we find an expression for the MSE of $\tilde{\theta}_{i,B}$ which is correct up to $O(m^{-1/2})$ as well as the jackknife estimator of MSE for $\tilde{\theta}_{i,B}^E := \tilde{\theta}_{i,B}(\hat{\boldsymbol{\omega}})$. The MSE of the empirical predictor B is

$$\begin{aligned} \mathrm{MSE}(\tilde{\theta}_{i,B}^E) &= E\Big\{ [(\tilde{\theta}_{i,B} - \theta_i)^2 \Big\} + E\Big\{ (\tilde{\theta}_{i,B}^E - \tilde{\theta}_{i,B})^2 \Big\} \\ &:= R_{1i} + R_{2i}, \end{aligned} \tag{5.1}$$

where R_{1i} is equal to

$$R_{1i} = E\Big\{ (\tilde{\theta}_{i,B} - \theta_i)^2 \Big\} = E\Big\{ (\theta_i - \tilde{\theta}_{i,A})^2 \Big\} + E\Big\{ (\tilde{\theta}_{i,A} - \tilde{\theta}_{i,B})^2 \Big\}.$$

Firstly,

$$\begin{aligned} E\Big\{ (\theta_i - \tilde{\theta}_{i,A})^2 \Big\} &= E\Big[E\Big\{ (\theta_i - \tilde{\theta}_{i,A})^2 | \beta_0, \beta_1, \sigma_v^2, z_i, W_i \Big\} \Big] = E\Big\{ V(\theta_i) | \beta_0, \beta_1, \sigma_v^2, z_i, W_i \Big\} \\ &= E\Big\{ E(\theta_i^2 | \beta_0, \beta_1, \sigma_v^2, z_i, W_i) - \tilde{\theta}_{i,A}^2 \Big\} = E(\theta_i^2) - E(\tilde{\theta}_{i,A}^2). \end{aligned}$$

Secondly, $E\{(\tilde{\theta}_{i,A} - \tilde{\theta}_{i,B})^2\} = E[\tilde{\theta}_{i,A}^2\{1 - \exp(-d_i/2)\}^2]$. By combining the expressions, we have

$$E\Big\{(\tilde{\theta}_{i,B} - \theta_i)^2\Big\} = E(\theta_i^2) + E(\tilde{\theta}_{i,A}^2)\Big\{ - 2\exp\Big(- \frac{d_i}{2}\Big) + \exp(-d_i)\Big\}.$$

Additionally,

$$\begin{aligned}
E(\tilde{\theta}_{i,A}^2) &= E\Big[\exp\Big\{2\tilde{\gamma}_i z_i + 2(1-\tilde{\gamma}_i)(\beta_0 + \beta_1 W_i) + \tilde{\gamma}_i\psi_i\Big\}\Big] \\
&= \exp\Big[2\Big\{\tilde{\gamma}_i(\beta_0 + \beta_1 x_i) + (1-\tilde{\gamma}_i)(\beta_0 + \beta_1 x_i)\Big\}\Big] \\
&\quad \times \exp\Big\{2\tilde{\gamma}_i^2(\sigma_v^2 + \psi_i) + 2(1-\tilde{\gamma}_i)^2\beta_1^2 C_i + \tilde{\gamma}_i\psi_i\Big\} \\
&= \exp\Big\{2(\beta_0 + \beta_1 x_i + \sigma_v^2)\Big\}\exp(-2\sigma_v^2)\exp(-\tilde{\gamma}_i\psi_i) \\
&\quad \times \exp\Big[2\Big\{\tilde{\gamma}_i^2(\sigma_v^2 + \psi_i) + (1-\tilde{\gamma}_i)^2\beta_1^2 C_i + \tilde{\gamma}_i\psi_i\Big\}\Big].
\end{aligned}$$

Note $\tilde{\gamma}_i^2(\sigma_v^2 + \psi_i) + (1-\tilde{\gamma}_i)^2\beta_1^2 C_i + \tilde{\gamma}_i\psi_i = d_i + r^2$. Thus,

$$E(\tilde{\theta}_{i,A}^2) = E(\theta_i^2)\exp\Big(2d_i - \tilde{\gamma}_i\psi_i\Big).$$

As a result

$$\begin{aligned}
R_{1i} &= E\Big\{(\tilde{\theta}_{i,B} - \theta_i)^2\Big\} \\
&= E(\theta_i^2) + E(\theta_i^2)\exp\Big(2d_i - \tilde{\gamma}_i\psi_i\Big)\Big\{ - 2\exp\Big(- \frac{1}{2}\,d_i\Big) + \exp(-d_i)\Big\} \\
&= E(\theta_i^2)\Big\{1 - 2\exp\Big(\frac{3}{2}\,d_i - \tilde{\gamma}_i\psi_i\Big) + \exp(d_i - \tilde{\gamma}_i\psi_i)\Big\} \\
&:= M_{1i}(\boldsymbol{\omega})M_{2i}(\boldsymbol{\omega}),
\end{aligned} \tag{5.2}$$

where $E(\theta_i^2) = \exp\{2(\beta_0 + \beta_1 x_i + \sigma_v^2)\}$. We define $M_{1i}(\boldsymbol{\omega})$ as follows

$$M_{1i}(\boldsymbol{\omega}) := E(\theta_i^2) = E\{\exp(2\phi_i)\} = \exp(2\beta_0 + 2\beta_1 x_i + 2\sigma_v^2).$$

Now note that $E\{\exp(2z_i)\} = \exp(2\beta_0 + 2\beta_1 x_i + 2\sigma_v^2 + 2\psi_i)$. In order to find an unbiased estimator for $M_{1i}(\boldsymbol{\omega})$, one can define $\hat{M}_{1i}(\hat{\boldsymbol{\omega}}) := \exp\{2(z_i - \psi_i)\}$, so that $E\{\hat{M}_{1i}(\hat{\boldsymbol{\omega}})\} = M_{1i}(\boldsymbol{\omega})$.

Now, we have $E\{\hat{M}_{1i}(\hat{\boldsymbol{\omega}}) - M_{1i}(\boldsymbol{\omega})\}^2 = E\{\hat{M}_{1i}^2(\hat{\boldsymbol{\omega}})\} - M_{1i}^2(\boldsymbol{\omega})$, so that

$$\begin{aligned}
E[\hat{M}_{1i}^2(\hat{\boldsymbol{\omega}})] &= E[\exp\{4(z_i - \psi_i)\}] = \exp(-4\psi_i)E\{\exp(4z_i)\} \\
&= \exp(-4\psi_i)\exp(4\beta_0 + 4\beta_1 x_i + 8\sigma_v^2 + 8\psi_i).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&E\{\hat{M}_{1i}(\hat{\boldsymbol{\omega}}) - M_{1i}(\boldsymbol{\omega})\}^2 \\
&= \exp(-4\psi_i)\exp(4\beta_0 + 4\beta_1 x_i + 8\sigma_v^2 + 8\psi_i) \\
&\quad - \exp(-4\psi_i)\exp(4\beta_0 + 4\beta_1 x_i + 4\sigma_v^2 + 4\psi_i) \\
&= \exp(-4\psi_i)\exp(4\beta_0 + 4\beta_1 x_i + 8\sigma_v^2 + 8\psi_i)\{1 - \exp(-4\sigma_v^2 - 4\psi_i)\}.
\end{aligned}$$

One can estimate $E[\hat{M}_{1i}(\hat{\boldsymbol{\omega}}) - M_{1i}(\boldsymbol{\omega})]^2$ by $\Lambda_i(\hat{\boldsymbol{\omega}})$ defined as follows:

$$\Lambda_i(\hat{\boldsymbol{\omega}}) := \exp(-4\psi_i)\exp(4z_i)\{1 - \exp(-4\hat{\sigma}_v^2 - 4\psi_i)\}.$$

For $M_{2i}(\boldsymbol{\omega})$ in expression (5.2), we have

$$M_{2i}(\boldsymbol{\omega}) = 1 - 2\exp\left(\frac{3}{2}\, d_i - \tilde{\gamma}_i\psi_i\right) + \exp(d_i - \tilde{\gamma}_i\psi_i).$$

Define

$$E\{M_{2i}^2(\hat{\boldsymbol{\omega}}) - M_{2i}^2(\boldsymbol{\omega})\} := E\{M_{2i}(\hat{\boldsymbol{\omega}}) - M_{2i}(\boldsymbol{\omega})\}^2 + 2M_{2i}(\boldsymbol{\omega})E\{M_{2i}(\hat{\boldsymbol{\omega}}) - M_{2i}(\boldsymbol{\omega})\},$$

where the terms can be expanded as

$$\begin{aligned}
&\text{(i)} \quad E\{M_{2i}(\hat{\boldsymbol{\omega}}) - M_{2i}(\boldsymbol{\omega})\}^2 \\
&\qquad = E\left[\left\{\exp(\hat{d}_i - \hat{\tilde{\gamma}}_i\psi_i) - \exp(d_i - \tilde{\gamma}_i\psi_i)\right\}\right. \\
&\qquad\quad \left. - 2\left\{\exp\left(\frac{3}{2}\hat{d}_i - \hat{\tilde{\gamma}}_i\psi_i\right) - \exp\left(\frac{3}{2}d_i - \gamma_i\psi_i\right)\right\}\right]^2 \\
&\qquad = O(m^{-1}), \quad \text{following Theorem 2, and} \\
&\text{(ii)} \quad E|M_{2i}(\hat{\boldsymbol{\omega}}) - M_{2i}(\boldsymbol{\omega})| \le E^{1/2}\{M_{2i}(\hat{\boldsymbol{\omega}}) - M_{2i}(\boldsymbol{\omega})\}^2 = O(m^{-1/2}).
\end{aligned}$$

The quantity R_{1i} can be estimated with

$$\begin{aligned}
\hat{R}_{1i} = M_{2i}^2(\hat{\boldsymbol{\omega}})\Lambda_i(\hat{\boldsymbol{\omega}}) &= \left\{1 - 2\exp\left(\frac{3}{2}\hat{d}_i - \hat{\tilde{\gamma}}_i\psi_i\right) + \exp(\hat{d}_i - \hat{\tilde{\gamma}}_i\psi_i)\right\}^2 \\
&\quad \times \exp(-4\psi_i)\exp(4z_i)\{1 - \exp(-4\hat{\sigma}_v^2 - 4\psi_i)\},
\end{aligned}$$

where it has the property that its bias and variance vanish with an order $O(m^{-1/2})$. Details of derivations are given in the Supplementary Material.

In general, there is no closed form expression available for the term R_{2i} in (5.1). One can use the jackknife technique to estimate it as well as the bias of $\hat{R}_{1i}$ for R_{1i}. Therefore, the jackknife estimator of $\mathrm{MSE}(\tilde{\theta}_{i,B}^E)$ is

$$mse_J(\tilde{\theta}_{i,B}^E) = \hat{R}_{1i,J} + \hat{R}_{2i,J},$$

where $\hat{R}_{1i,J} = \hat{R}_{1i} - \{(m-1)/m\}\sum_{j=1}^{m}(\hat{R}_{1i(-j)} - \hat{R}_{1i})$ and $\hat{R}_{2i,J} = \{(m-1)/m\}$ $\sum_{j=1}^{m}(\tilde{\theta}^{E}_{i(-j),B} - \tilde{\theta}^{E}_{i,B})^2$.

Under the regularity conditions 1–3 given in the Supplementary Material, one can show that $E(\hat{R}_{1i,J}) = R_{1i} + O(m^{-1})$ and $E(\hat{R}_{2i,J}) = R_{2i} + o(m^{-1})$ as $m \to \infty$. Thus, $E\{mse_J(\tilde{\theta}^{E}_{i,B})\} = \text{MSE}(\tilde{\theta}^{E}_{i,B}) + O(m^{-1})$. The proof follows along the same lines of Mosaferi, Ghosh and Steorts (2023), and hence we omit it.

The analytical approximation for the MSE of predictor $\tilde{\theta}_{i,B}$ has a complex form. As we will find from our simulations, constructing intervals based on that as well as the jackknife estimator of MSE perform poorly in terms of coverage or length. Additionally, jackknife MSE might yield negative values. Thus, one might prefer to use resampling procedures such as bootstrap to construct the intervals as they are easier with better interpretation and coverage property.

6. Non-parametric Bootstrap Prediction Intervals

In this section, we propose a non-parametric bootstrap approach to approximate the entire distribution of predictor B. We use the percentiles of the bootstrap histogram to obtain highly accurate prediction intervals in terms of coverage. For this purpose, we repeatedly draw samples from the original observed sample.

We construct the bootstrap distribution of predictor B ($\tilde{\theta}_{i,B}$) based on the observed data (W_i, y_i) for $i = 1, \ldots, m$ such that $\tilde{\theta}^{*}_{i,B} = \tilde{\theta}_{i,B}(W^{*}_i, y^{*}_i)$, where (W^{*}_i, y^{*}_i) is obtained from the resampling observed pairs (W_i, y_i); i.e., $\{(W_{j_1}, y_{j_1}), (W_{j_2}, y_{j_2}), \ldots, (W_{j_m}, y_{j_m})\}$ where $j_1, j_2, \ldots, j_m$ is a random sample drawn with replacement from $\{1, 2, \ldots, m\}$. Each bootstrap sample gives a non-parametric bootstrap replication of $\tilde{\theta}_{i,B}$ denoted by $\tilde{\theta}^{*}_{i,B}$. After repeating the bootstrap distribution "BT" times, we obtain $\tilde{\theta}^{*(1)}_{i,B}, \tilde{\theta}^{*(2)}_{i,B}, \ldots, \tilde{\theta}^{*(BT)}_{i,B}$.

We use the upper and lower $\alpha/2$ quantiles of these BT numbers as the prediction interval for θ. Specifically, we propose to use the interval

$$\hat{I}_{i,\alpha} = [\ell_{i,\alpha}, u_{i,\alpha}] = \left[\hat{G}_i^{-1}\left(\frac{\alpha}{2}\right), \hat{G}_i^{-1}\left(1 - \frac{\alpha}{2}\right)\right], \tag{6.1}$$

where $\hat{G}_i(t) = \sum_{bt=1}^{BT} I(\tilde{\theta}^{*(bt)}_{i,B} \leq t)/BT$. In the above expression, $\hat{G}_i(t)$ is the cumulative distribution function (CDF) of BT bootstrap replications, and we let $BT = 2000$ in the application and simulation studies.

Using functional delta method (van der Vaart, 2000, Chap. 20) and Theorem 2, $(\tilde{\theta}_{i,B} - \theta_i) \xrightarrow{D} N(0, \sigma^2)$ for a constant $\sigma^2 > 0$. Let's consider the sampling distribution of standardized $\tilde{\theta}_{i,B}$ (i.e., $T_i = \{\tilde{\theta}_{i,B} - \theta_i\}/\sigma_i$) defined as $G_i(t) \equiv P(T_i \leq t)$ for $t \in \mathbb{R}$. We can obtain the bootstrap estimator of $G_i(t)$ from the bootstrap version $T^{*}_i = \{\tilde{\theta}^{*}_{i,B} - \tilde{\theta}_{i,B}\}/\sigma^{*}_i$ of T_i defined as $\hat{G}^{*}_i(t) \equiv P_{*}(T^{*}_i \leq t)$ for $t \in \mathbb{R}$. As $m \to \infty$ and the bootstrap replications BT increase, the probability distribution $\hat{G}^{*}_i(t)$ weakly converges to $G_i(t)$, i.e., $\hat{G}^{*}_i(t) \to G_i(t)$ for all the

continuity points t of G_i (Hall, 2013, Chaps. 3 and 4) and under the assumption of $\sigma_i^*/\sigma_i \xrightarrow{P} 1$.

One can use the subsequence arguments and the correspondence between convergence almost surely along subsequences and convergence in probability to show the distance between distributions $\hat{G}_i^*$ and G_i goes to zero in probability. Thus, based on Polya's theorem, $\sup_t |\hat{G}_i^*(t) - G_i(t)| = \sup_t |\{\hat{G}_i^*(t) - \Phi(t)\} - \{G_i(t) - \Phi(t)\}| \xrightarrow{P} 0$ for $t \in \mathbb{R}$ and as $m \to \infty$ (Lahiri, 2003, Chap. 2), where $\Phi(t)$ is the CDF of standard normal, and $\hat{G}_i^*(.)$ is a consistent estimator of $G_i(.)$ under the regularity conditions 1 and 2 given in the Supplementary Material and assuming $E[y_i^2]$, $E[||x_i||^2] < \infty$. When model (2.1) is correctly specified,

$$\begin{aligned}
&P\{|\sigma_i^{-1}(\theta_i - \tilde{\theta}_{i,B}^*)| \geq z_{\alpha/2}\} \\
&\leq P\{|\sigma_i^{-1}(\theta_i - \tilde{\theta}_{i,B})| \geq z_{\alpha/2}\} + P\{|\sigma_i^{*-1}(\tilde{\theta}_{i,B} - \tilde{\theta}_{i,B}^*)| \geq z_{\alpha/2}\} \\
&= \alpha + o(1).
\end{aligned}$$

This states that the prediction interval $\hat{I}_{i,\alpha}$ given in (6.1) is asymptotically valid.

7. Simulation Studies

We perform a simulation study to compare the performance of several predictors. For this purpose, we generate data from the model in Section 1. For effective comparisons, x_i is drawn from $N(5, 9)$ (Ybarra and Lohr, 2008) and $\psi_i \sim \text{Gamma}(4.5, 2)$. Then, $\log Y_i = 3x_i + v_i$, $\log y_i = \log Y_i + e_i$, and $W_i = x_i + u_i$. Here, $v_i \sim N(0, \sigma_v^2)$, $e_i \sim N(0, \psi_i)$, and $u_i \sim N(0, C_i)$. The sources of errors v_i, e_i, and u_i are mutually independent. We let $\sigma_v^2 = 2$, and $C_i \in \{0, d\}$, where $d = 2$ or 4 such that only $k\%$ of the C_i's randomly receive d and the rest receive 0, where $k \in \{25, 50, 80, 100\}$. The number of small areas are $m \in \{20, 50, 100\}$.

This set-up of simulation has been previously used by Ybarra and Lohr (2008) which makes our results effectively comparable with theirs. We assume the total number of replications to be $R = 2000$. We compare the performance of four predictors as follows, where we assume Y_i is the truth:

(1) y_i: direct estimator,

(2) $\tilde{\theta}_{i,\text{No-ME}}$: predictor without measurement error,

(3) $\tilde{\theta}_{i,A}$: predictor A, and

(4) $\tilde{\theta}_{i,B}$: predictor B.

For all the predictors, we substitute the estimated values of unknown parameters $\boldsymbol{\omega}$. The resulting predictors are listed in Table 1. Overall, the values of proposed predictor B are much closer to the truth Y_i compared to the rest of other predictors.

Table 1. Comparison of predictors among all the small areas assuming $m = 20$ and $k = 50\%$. The numerical values are in the logarithmic scale.

C_i	Y_i	y_i	$\tilde{\theta}_{i,\text{No-ME}}$	$\tilde{\theta}_{i,A}$	$\tilde{\theta}_{i,B}$
2	16.030	18.222	17.218	19.668	16.337
0	34.128	38.389	36.726	36.726	36.726
2	13.808	20.581	15.308	19.826	14.595
2	-2.514	3.224	0.751	5.094	-0.987
0	9.524	14.283	11.011	11.011	11.011
0	6.371	13.469	8.172	8.172	8.172
2	10.209	14.919	11.995	16.771	10.640
2	18.659	21.433	19.917	23.553	18.665
0	14.746	18.158	15.858	15.858	15.858
0	11.059	16.921	12.780	12.780	12.780
2	26.719	32.409	29.597	34.329	27.178
0	21.941	29.260	24.298	24.298	24.298
2	1.420	8.516	4.125	9.875	4.291
0	14.547	18.936	15.751	15.751	15.751
2	13.990	18.599	16.078	20.317	14.602
2	15.504	22.758	17.554	23.219	15.844
2	16.491	20.650	18.181	22.292	16.827
2	12.325	15.297	13.927	16.460	12.934
0	16.549	20.563	17.501	17.501	17.501
2	25.688	28.582	27.296	30.386	26.005
Avg	14.860	19.758	16.702	19.194	15.951

We observe that when $C_i = 0$ (no measurement error), the values of $\tilde{\theta}_{i,A}$, $\tilde{\theta}_{i,B}$, and $\tilde{\theta}_{i,\text{No-ME}}$ are identical. As the measurement error C_i increases, the values of proposed predictor $\tilde{\theta}_{i,B}$ become much closer to the truth Y_i rather than $\tilde{\theta}_{i,A}$. In the last row of Table 1, we report the average over the values of all small areas, which confirm the previous conclusion.

We compare the performance of predictors based on the empirical MSE defined as follows:

$$\text{EMSE}(\tilde{\theta}_i) = \frac{1}{R}\sum_{r=1}^{R}\left(\tilde{\theta}_i^{(r)} - Y_i^{(r)}\right)^2, \tag{7.1}$$

where $\tilde{\theta}_i$ is the predictor for Y_i, and R is the total number of replications. Based on the results given in Table 2, predictor B is superior to the rest of other predictors. Additionally, we make comparisons with the estimated MSE ($\hat{R}_{1i}$) and jackknife estimator (mse_J). When $C_i = 0$, the empirical MSE's for $\tilde{\theta}_{i,A}$, $\tilde{\theta}_{i,B}$, and $\tilde{\theta}_{i,\text{No-ME}}$ are identical, and when $C_i = 2$, the empirical MSE's for $\tilde{\theta}_{i,B}$ are much smaller than the empirical MSE's for $\tilde{\theta}_{i,A}$. Overall, $\hat{R}_{1i}$ and mse_J are much larger than the empirical MSE of predictor B. In the last row of the table, we report the average over the values of all small areas.

Table 2. Comparison of empirical MSE of predictors among all the small areas assuming $m = 20$ and $k = 50\%$. The numerical values are in the logarithmic scale.

C_i	EMSE(y_i)	EMSE($\tilde{\theta}_{i,\text{No-ME}}$)	EMSE($\tilde{\theta}_{i,A}$)	EMSE($\tilde{\theta}_{i,B}$)	$\hat{R}_{1i}(\tilde{\theta}_{i,B})$	$mse_J(\tilde{\theta}_{i,B})$
2	39.620	37.307	42.255	35.659	81.028	83.227
0	82.052	76.862	76.862	76.862	136.723	137.929
2	48.272	35.470	45.134	34.809	84.564	67.638
2	11.322	5.430	14.972	3.467	4.237	8.242
0	34.702	25.503	25.503	25.503	39.084	38.530
0	34.375	19.240	19.240	19.240	27.859	27.626
2	35.133	27.569	39.131	25.354	56.307	52.168
2	46.958	42.890	50.917	40.981	89.584	90.375
0	41.084	34.251	34.251	34.251	57.158	57.389
0	39.715	28.266	28.266	28.266	34.986	34.773
2	70.133	64.302	72.810	59.415	121.389	117.451
0	65.184	53.869	53.869	53.869	82.810	83.095
2	23.345	11.770	26.675	15.912	22.204	23.868
0	43.877	33.907	33.907	33.907	60.551	60.349
2	42.366	36.899	44.901	35.275	71.884	70.593
2	51.120	39.963	50.437	37.296	81.169	83.368
2	45.980	40.585	48.605	37.410	82.652	85.104
2	35.189	34.129	36.818	31.586	70.900	71.963
0	47.121	37.576	37.576	37.576	70.871	71.125
2	61.381	57.999	64.864	55.730	118.087	120.666
Avg	44.946	37.189	42.350	36.119	69.702	69.274

Table 3. Ratio of average MSE of predictor $\tilde{\theta}_i$ (A, B, or No-ME) to average MSE of direct estimator y_i, where $m = 20$.

Error	Ratio of MSEs		
C_i ($k = 50\%$)	$\tilde{\theta}_{i,\text{No-ME}}$	$\tilde{\theta}_{i,A}$	$\tilde{\theta}_{i,B}$
0	0.006	0.006	0.006
2	0.003	14.548	2.270e-05

For further evaluation, we give the ratio of average MSE of predictor $\tilde{\theta}_i$ to average MSE of direct estimator y_i in Table 3. When $C_i = 0$, the ratio of average MSE's for $\tilde{\theta}_{i,A}$, $\tilde{\theta}_{i,B}$, and $\tilde{\theta}_{i,\text{No-ME}}$ are identical. When $C_i = 2$, this ratio is much smaller for $\tilde{\theta}_{i,B}$ compared to $\tilde{\theta}_{i,A}$ and $\tilde{\theta}_{i,\text{No-ME}}$. Note that, since predictor A is substantially biased (see Section 3), its ratio of MSE to the direct one is very large.

We also provide the relative bias (RB) and the relative root mean squared error (RRMSE) of the main competitors; i.e., predictor A and predictor B in Figure 1. These quantities can be defined as follows:

$$\text{RB}(\tilde{\theta}_i) = \frac{E(\tilde{\theta}_i - Y_i)}{Y_i}, \quad \text{and} \quad \text{RRMSE}(\tilde{\theta}_i) = \frac{\{E(\tilde{\theta}_i - Y_i)^2\}^{1/2}}{Y_i}.$$

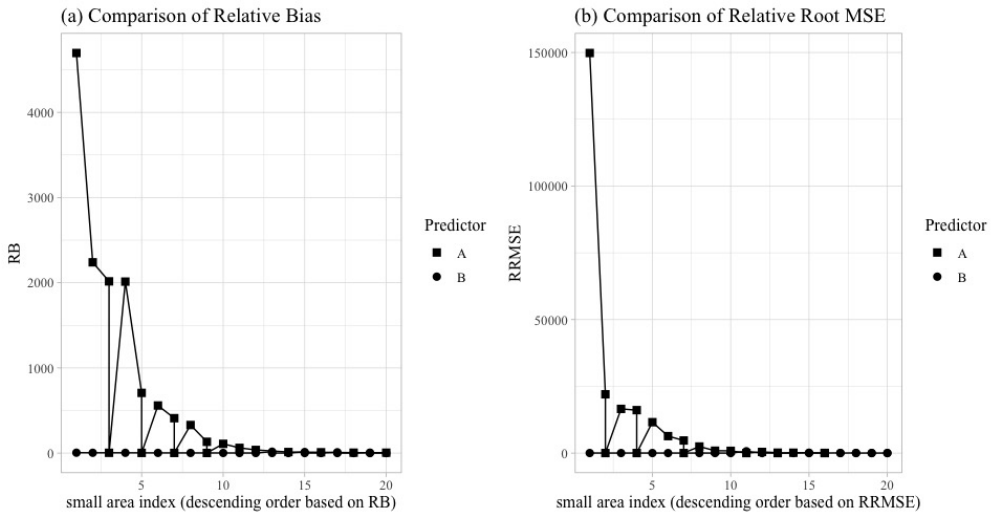

Figure 1. (a) Plot of RB for two predictors A and B. (b) Plot of RRMSE for two predictors A and B. We assume $m = 20$ and $k = 50\%$.

Based on the results in Figure 1, we observe that the RB's and RRMSE's of $\tilde{\theta}_{i,B}$ are very closely centered around 0, but this is not the case for the $\tilde{\theta}_{i,A}$.

For the sake of completeness, we compare the predictors over all possible values of k in Table S2.1 of the Supplementary Material for $m \in \{20, 50, 100\}$. The comparisons are based on the empirical MSE's which are averaged by values of C_i's and confirm the previous arguments.

In order to compare the performance of our proposed prediction intervals for predictor B with the direct estimator, we compare the coverage probabilities and expected lengths of four prediction intervals as follows:

(1) Direct: $[y_i - z_{1-\alpha/2}\sqrt{\psi_i},\ y_i + z_{1-\alpha/2}\sqrt{\psi_i}]$,

(2) Estimated MSE: $[\tilde{\theta}_{i,B} - z_{1-\alpha/2}\sqrt{\hat{R}_{1i}(\tilde{\theta}_{i,B})},\ \tilde{\theta}_{i,B} + z_{1-\alpha/2}\sqrt{\hat{R}_{1i}(\tilde{\theta}_{i,B})}]$,

(3) Jackknife: $[\tilde{\theta}_{i,B} - z_{1-\alpha/2}\sqrt{mse_J(\tilde{\theta}_{i,B})},\ \tilde{\theta}_{i,B} + z_{1-\alpha/2}\sqrt{mse_J(\tilde{\theta}_{i,B})}]$, and

(4) Bootstrap: $[\ell_{i,\alpha},\ u_{i,\alpha}]$ given in (6.1).

We consider nominal coverage $100(1-\alpha)\% = 90\%$, 95%, and 99%. The results are given in Table 4. Overall, the bootstrap method outperforms the other three methods.

The coverage probabilities for bootstrap method are close to the nominal coverage, thus the intervals more accurately include the truth. Prediction intervals based on the direct method, estimated MSE, and jackknife are not reliable as they do not approximately follow the normal theory. Additionally, the latter two usually suffer from coverage problem because of the choice of MSE estimator and small values of m.

Table 4. Comparison of coverage probabilities and mean of the log lengths for the prediction intervals across all the methods. The results for the lengths are averaged over all the small areas. Additionally, we assume $k = 50\%$ for all the cases.

Nominal Coverage	Small Areas m	Direct	Estimated MSE	Jackknife	Bootstrap
90%	20	0.520	0.450	0.420	0.990
		(15.383)	(12.195)	(13.646)	(30.463)
	50	0.688	0.596	0.718	0.928
		(18.483)	(19.754)	(16.708)	(32.969)
95%	20	0.560	0.450	0.420	1.000
		(15.561)	(12.374)	(13.824)	(37.314)
	50	0.696	0.596	0.764	0.948
		(18.662)	(19.932)	(16.887)	(36.755)
99%	20	0.560	0.500	0.500	1.000
		(15.836)	(12.649)	(14.099)	(38.346)
	50	0.736	0.596	0.791	0.992
		(18.936)	(20.207)	(17.161)	(37.692)

8. Applications to Census Bureau's Data Sets

In this Section, we describe the steps used to apply the preceding theory to census data sets. The first application is related to the Census of Governments, where we illustrate our methodology at the state level. The second application is related to the Small Area Income and Poverty Estimates (SAIPE) Program, where we illustrate our methodology at the county level.

8.1. Census of Governments

The purpose of the Census of Governments is providing periodic and comprehensive statistics about governments and governmental activities, and it covers all the states and local governments in the United States. Data are obtained on government organizations, finances, and employment and include location, type, and characteristics of local governments and officials; see `https://www.census.gov/econ/overview/go0100.html` for further information.

Since 1957 the United States Census Bureau collects information from governmental units for years ending in 2 and 7. Here, we utilize data from 2007 and 2012 with 49 states of the Continental United States (excluding Hawaii and District of Columbia) as our small areas of interest.

We define the parameter of interest θ_i to be the mean number of full-time employees per government at state i from the 2012 data set. We define the covariate to be the corresponding mean from the 2007 data set. To define the response y_i, we select sample of sizes 4,000 and 8,000 governmental units from the 2012 data set. Similarly, we construct the covariate W_i from an independent

Table 5. Descriptive statistics for the log lengths of 95% prediction intervals from the Census of Governments. The results are computed using data from all the small areas.

Sample Size	Method	Minimum	25%	Median	Mean	75%	Maximum
4,000	Bootstrap	4.320	4.350	4.360	4.370	4.380	4.410
	Direct	4.330	6.380	6.730	6.970	7.510	11.700
	Jackknife	-7.200	2.950	7.710	5.840	9.660	12.000
8,000	Bootstrap	4.160	4.210	4.230	4.220	4.250	4.270
	Direct	5.370	6.420	7.000	7.150	7.780	10.000
	Jackknife	-15.600	-1.560	6.400	3.760	8.590	11.300

sample of 40,000 and 80,000 governmental units selected from the 2007 data set.

The distributions of response and covariate are displayed in Figures S3.1 and S3.2 of the Supplementary Material for sample sizes 4,000 and 8,000. We observe skewed patterns in both the average number of full-time employees from 2007 and 2012 which motivates our proposed framework. Before a logarithmic transformation, both variables fail the normality assumption, and the normality assumption is more justified after the logarithmic transformation.

The measurement error variances C_i's for the covariates are obtained from a Taylor series approximation because of the logarithmic transformation, and the formula of variance in simple random sampling without replacement is used per each state in the original scale. Additionally, we used Taylor series approximation for the ψ_i. We assume the sampling variances to be known throughout the estimation procedure. Based on our proposed framework, we find the empirical predictor $\tilde{\theta}_{i,B}$.

We construct prediction intervals based on three methods of "Direct", "Jackknife", and "Bootstrap". The box-plots of prediction interval lengths are given in Figure 2. We observe that the distribution of lengths based on non-parametric bootstrap method has less variation in comparison with both direct and jackknife methods. Additionally, the descriptive statistics of lengths for 95% prediction intervals are given in Table 5. We clearly observe more stable results for the descriptive statistics under the bootstrap method.

8.2. Small area income and poverty estimates program

The U.S. Census Bureau's SAIPE program uses Fay and Herriot (1979) model to produce model-based estimates of income and poverty at the state and county levels for various age groups. Since 2005, they use the data from the American Community Survey (ACS) in the modeling. Prior to 2005, data from the Current Population Survey were used. The ACS is the largest U.S. household survey and it almost covers 3.5 million addresses per year.

Despite the large sample size of the ACS, the 1-year direct estimates of the number of related school-aged (5–17 year old) children in poverty are highly

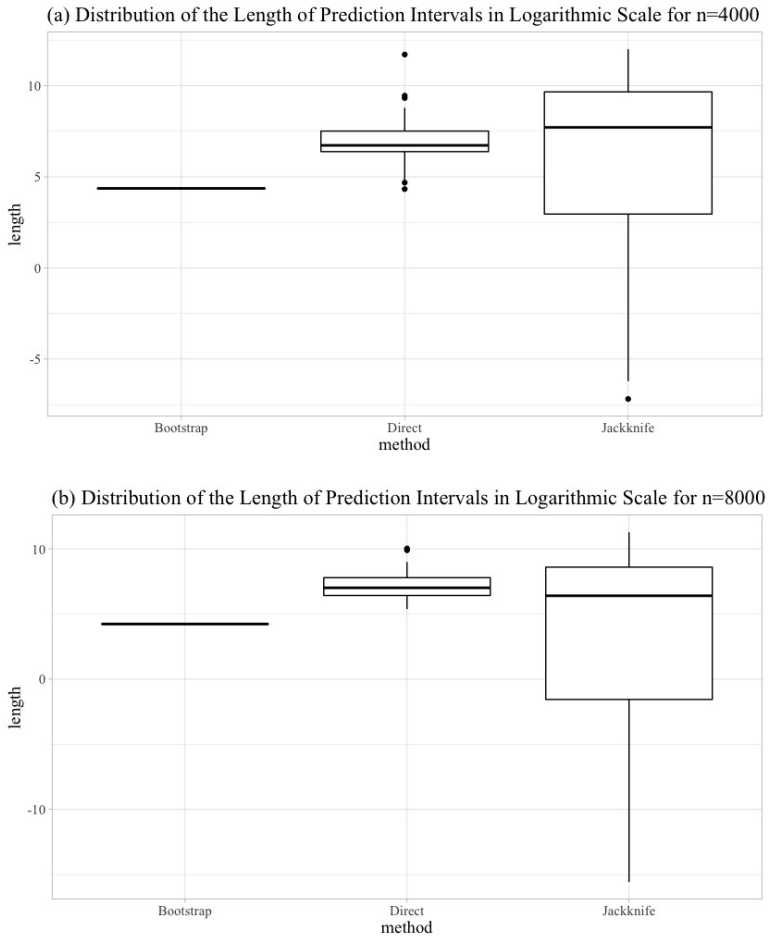

Figure 2. box-plots of prediction interval lengths with $1 - \alpha = 0.95$ based on three methods for the Census of Governments assuming (a) 4,000 and (b) 8,000 sample sizes.

variable for many small counties. Thus, SAIPE program uses the Fay and Herriot (1979) model to borrow strength from covariates such as log number of food-stamp participants, log number of IRS child exemptions in households in poverty, log number of related children aged 5–17 in poverty from previous census, etc. to improve the direct estimate of the logarithm of the single-year ACS poverty counts as the dependent variable. For more information on the SAIPE program, see the SAIPE web page at `https://www.census.gov/programs-surveys/saipe.html`.

Recent research by Huang and Bell (2012) and Franco and Bell (2015) suggests that the most recent previous 5-year ACS estimates instead of outdated census results can be used as a covariate (Arima et al., 2017). However, 5-year ACS estimates are subject to the sampling errors rather than the census results. Thus, it is appropriate to use Fay and Herriot (1979) where covariates are measured with errors and a log transformation is required. As a consequence,

our proposed predictor is well-suited for this scenario.

Here, we assume the parameter of interest is the total population for whom poverty status is determined for the age of 5 to 17 years in 2018 at the county level. The covariate is the corresponding total of 5-year aggregated values; i.e. 2013–2017 at the county level. Additionally, we use the 2018 SNAP data set as a separate covariate which is measured without error.

We summarize these variables as follows:

- y_i: 1-year ACS (2018) estimates for $i = 1, \ldots, 827$.
- W_i: Aggregated 5-year ACS (2013–2017) estimates, measured with errors, and
- x_i: 2018 SNAP data set measured without error as a separate covariate.

Unlike the 5-year aggregated ACS, the information for counties with population less than 65,000 are not publicly released for a single-year ACS. We successfully linked the same m=827 counties across the resources, and we only provide predictions for these publicly available counties. In Figure S3.3 (a) given in the Supplementary Material, we display the scatter plots of these counties.

We observe that both 5-year ACS and SNAP are highly correlated with the response variable y. The skewness in the original scales diminishes after a logarithmic transformation (see Fig. S3.3 (b) in the Supplementary Material). To obtain the variance of estimates, we use the 90% margin of error given in the data set for both the covariate and the response variable. Afterwards, we use a Taylor series approximation to obtain the variance of the logarithm of estimates.

In this section, we intend to study three small area models as follows:

(1) Measurement error model predictor: For this model, we assume the response variable is the 1-year ACS and the covariate is the 5-year ACS (measured with error). Because of skewed patterns in the data set, we need to use a logarithmic transformation. For this purpose, we use predictor B.

(2) Fay–Herriot model predictor: For this model, we assume the response variable is the 1-year ACS and the covariate is the SNAP information (measured without error). Because of skewed patterns in the data set, we again use a logarithmic transformation. For this purpose, we use predictor $\tilde{\theta}^0_{i,B}$, the so-called FHeblup afterwards.

(3) Direct Estimator: The results are solely based on the single-year ACS estimates.

In Figure 3, we compare the distribution of prediction interval lengths for all the counties based on non-parametric bootstrap, jackknife, direct, and FHeblup. We observe that the box-plot for the prediction interval lengths based on the non-parametric bootstrap method has a stable distribution.

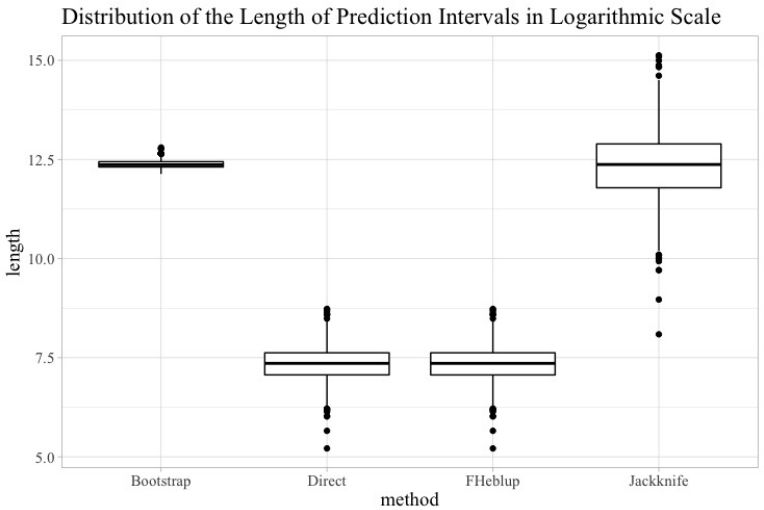

Figure 3. Box-plots of prediction interval lengths with $1 - \alpha = 0.95$ for the SAIPE data set. Note that the results are in the logarithmic scale.

Table 6. Mean, median, and standard deviation for the log lengths of 95% prediction intervals from the SAIPE data set. The results are computed using data from all the small areas.

Method	Median	Mean	Standard Deviation
Bootstrap	12.360	12.400	0.112
Direct	7.358	7.342	0.453
FHeblup	7.357	7.342	0.453
Jackknife	12.370	12.340	0.905

Additionally, box-plots of direct and FHeblup are similar due to the close values of direct estimators and FHeblups (see Fig. S3.4 in the Supplementary Material). This means the values of $\tilde{\theta}^0_{i,B}$ are close to z_i, which requires $\psi_i \approx 0$. This can be confirmed by the range of ψ_i, which is (6.579 e-07, 6.843 e-03). The values of mean, median, and standard deviation of the lengths for all the prediction intervals are given in Table 6.

9. Discussion and Future Work

We propose a new predictor for the skewed response variable under Fay–Herriot model when the covariate is measured with error. Our set-up can be easily extended to the multivariate covariate case, and some of the steps in this regard are given in the Supplementary Material. While the proposed method is for area level model, it would be possible to consider an extension to unit level models, which is left to a future work.

This modeling framework can be of interest for government agencies and survey practitioners dealing with right skewed response variables and covariates that are measured with errors. Our proposed predictor is unbiased and can perform uniformly better than the direct estimator and the alternative predictor gien in the literature by Mosaferi, Ghosh and Steorts (2023). Further, we derive an approximation of the MSE of predictor, which does not perform well for constructing prediction intervals in terms of coverage. Thus, we develop nonparametric bootstrap prediction intervals.

Prediction intervals based on nonparametric bootstrap techniques are easy and allow a good coverage property. In particular, they can be more suitable for real applications as we use the available data sets for generating resamples. It might be of interest to investigate other resampling methods such as parametric bootstrap methods or extend the earlier works of log-MSPE by Jiang, Lahiri and Nguyen (2018) to construct prediction intervals and make comparisons among them for our modeling framework.

Supplementary Material

The online Supplementary Material includes technical details, proofs of the theorems, and additional numerical results.

Acknowledgments

We would like to thank an associate editor and anonymous reviewers who made excellent comments and suggestions that helped us to improve the paper.

References

Arima, S., Bell, W., Datta, G. S., Franco, C. and Liseo, B. (2017). Multivariate Fay–Herriot Bayesian estimation of small area means under functional measurement error. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **180**, 1191–1209.

Berg, E. and Chandra, H. (2014). Small area prediction for a unit-level lognormal model. *Computational Statistics & Data Analysis* **78**, 159–175.

Butar, F. B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *Journal of Statistical Planning and Inference* **112**, 63–76.

Carter, G. M. and Rolph, J. E. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *Journal of the American Statistical Association* **69**, 880–885.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association* **70**, 311–319.

Fay, R. E. and 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.

Franco, C. and Bell, W. R. (2015). Borrowing information over time in binomial/logit normal models for small area estimation. *Statistics in Transition New Series* **16**, 563–584.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science* **9**, 55–76.

Hall, P. (2013). *The Bootstrap and Edgeworth Expansion*. Springer Science & Business Media.

Huang, E. T. and Bell, W. R. (2012). An empirical study on using previous American community survey data versus census 2000 data in SAIPE models for poverty estimates. *Statistics* **4**, 1–35.

Jiang, J., Lahiri, P. and Nguyen, T. (2018). A unified Monte-Carlo jackknife for small area estimation after model selection. *Annals of Mathematical Sciences and Applications* **3**, 405–438.

Jiang, J., Lahiri, P. and Wan, S. M. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *The Annals of Statistics* **30**, 1782–1810.

Lahiri, S. (2003). *Resampling Methods for Dependent Data*. Springer Science & Business Media.

Molina, I. and Martín, N. (2018). Empirical best prediction under a nested error model with log transformation. *The Annals of Statistics* **46**, 1961–1993.

Mosaferi, S., Ghosh, M. and Steorts, R. C. (2023). Transformed Fay–Herriot model with measurement error in covariates. *Communications in Statistics—Simulation and Computation* **52**, 2257–2274.

Pfefferman, D. (2013). New important developments in small area estimation. *Statistical Science* **28**, 40–68.

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

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation*. Wiley Series in Survey Sampling.

Slud, E. V. and Maiti, T. (2006). Mean-squared error estimation in transformed Fay–Herriot models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **68**, 239–257.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Ybarra, L. M. and Lohr, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika* **95**, 919–931.

(Received March 2023; accepted August 2023)