Transformed Fay-Herriot model with measurement error in covariates

Sepideh Mosaferi, Malay Ghosh, and Rebecca C. Steorts

Department of Statistics, Iowa State University, Ames, Iowa, USA; Department of Statistics, University of Florida, Gainesville, Florida, USA; Department of Statistical Science and Computer Science, Duke University, Durham, North Carolina, USA

ABSTRACT

Statistical agencies are often asked to produce small area estimates (SAEs) for positively skewed variables. When domain sample sizes are too small to support direct estimators, effects of skewness of the response variable can be large. As such, it is important to appropriately account for the distribution of the response variable given available auxiliary information. Motivated by this issue and in order to stabilize the skewness and achieve normality in the response variable, we propose an area-level log-measurement error model on the response variable. Then, under our proposed modeling framework, we derive an empirical Bayes (EB) predictor of positive small area quantities subject to the covariates containing measurement error. We propose a corresponding mean squared prediction error (MSPE) of EB predictor using both a jackknife and a bootstrap method. We show that the order of the bias is $O(m^{-1})$, where $m$ is the number of small areas. Finally, we investigate the performance of our methodology using both design-based and model-based simulation studies.

ARTICLE HISTORY

Received 13 March 2020
Accepted 6 March 2021

KEYWORDS

Small area estimation; official statistics; Bayesian methods; jackknife; parametric bootstrap; applied statistics; simulation studies

1. Introduction

Typically, in small area measurement error models, both the response variable and covariate can be any real number (see Ybarra and Lohr (2008), Arima et al. (2017)). However, statistical agencies are often asked to produce small area estimates (SAEs) for skewed variables, which are also positive in $\mathbb{R}^+$. For instance, the Census of the Governments (CoG) provides information on roads, tolls, airports, and other similar information at the local-government level as defined by the United States Census Bureau (USCB). Another example includes the United States National Agricultural Statistics Service (NASS), which provides estimates regarding crop harvests (see Bellow and Lahiri (2011)). The United States Natural Resources Conservation Service (NRCS) provides estimates regarding roads at the county-level (e.g., Wang and Fuller (2003)), and the Australian Agricultural and Grazing Industries Survey provides estimates of the total expenditures of Australian farms (e.g., Chandra and Chambers (2011)).

When domain sample sizes are too small to support direct estimators, the effect of skewness can be quite large, and it is critical to account for the distribution of the response variable given auxiliary information at hand. For a review of the SAE literature, we refer to recent work by Rao and Molina (2015) and Pfefferman (2013). The case of positively skewed response variables is one such that the governing parameter in the Box-Cox transformation is zero. Due to the fact that...
the covariate in the model may be positively skewed and contains measurement error, this has received less attention in the literature. Throughout this paper, we explain the problem which is beyond a simple substitution and address some of its difficulties.

1.1. Census of the governments

As mentioned in Sec. 1, our proposed framework is motivated by data that are positively skewed. One such data set is the Census of Governments (CoG), which is a survey data collected by the United States Census Bureau (USCB) periodically that provides comprehensive statistics about governments and governmental activities. Data are reported on government organizations, finances, and employment. For example, data from organizations refer to location, type, and characteristics of local governments and officials. Data from finances/employment refer to revenue, expenditure, debt, assets, employees, payroll, and benefits.

We utilize data from the CoG from 2007 and 2012 (https://www.census.gov/econ/overview/go0100.html). In the CoG, the small areas consist of the 48 states of the contiguous United States. These 48 areas contain 86,152 local governments defined by the USCB, such as airports, toll roads, bridges, and other federal government corporations. The parameter of interest is the average number of full-time employees per government at the state level from the 2012 data set, which can be defined as the total number of full-time employees from all local governments divided by the total number of local governments per state. The covariate of interest is the average number of full-time employees per government at the state level from the 2007 data set. After studying residual plots and histograms, we observe skewed patterns in the average number of full-time employees in both the 2007 and 2012 data sets, which partially motivate our proposed framework.

1.2. Our contribution

Motivated by issues that statistical agencies face with skewed response variables, we make several contributions to the literature. In order to stabilize the skewness and achieve normality in the response variable, we propose an area-level log-measurement error model on the response variable (Eq. (1)). In addition, we propose a log-measurement error model on the covariates (Eq. (2)). Next, under our proposed modeling framework, we derive an EB predictor of positive small area quantities subject to the covariates containing measurement error. In addition, we propose a corresponding estimate of the MSPE using a jackknife and a parametric bootstrap, where we illustrate that the order of the bias is $O(m^{-1})$ under standard regularity conditions. We illustrate the performance of our methodology in both model-based simulation and design-based simulation studies. We summarize our conclusions and provide directions for future work.

The article is organized as follows. Sec. 1.3 details the prior work related to our proposed methodology. In Sec. 2, we propose a log-measurement error model for the response variable. In addition, we consider a measurement error model of the covariates with a log transformation. Further, we derive the EB predictor under our framework. Sec. 2.2 provides the MSPE for our EB predictor. We provide a decomposition of the MSPE to include the uncertainty of the EB predictor through unknown parameters. Sec. 3 provides two estimators of the MSPE, namely a jackknife and a parametric bootstrap, where we prove that the order of the bias is $O(m^{-1})$ under standard regularity conditions. Sec. 4 provides both design-based and model-based simulation studies. Sec. 5 provides a discussion and directions for future work.

1.3. Prior work

In this section, we review the prior literature most relevant to our proposed work. There is a rich literature on the area-level Fay-Herriot model, where various additive measurement error models
have been proposed on the covariates. Ybarra and Lohr (2008) proposed the first additive measurement error model on the covariates. More specifically, the authors considered covariate information from another survey that was independent of the response variable. More recently, Berg and Chandra (2014) have proposed an EB predictor and an approximately unbiased MSE estimator under a unit-level log-normal model, where no measurement error is assumed present in the covariates. Turning to the Bayesian literature, Arima et al. (2017), Arima, Datta, and Liseo (2015a) and (2015b) have provided fully Bayesian solutions to the measurement error problem for both unit-level and area-level small area estimation problems.

Next, we discuss related literature regarding the proposed jackknife and parametric bootstrap estimator of the MSPE of the Bayes estimators, where the order of the bias is \(O(m^{-1})\), under standard regularity conditions. Our proposed jackknife estimator of the MSPE contrasts that of Jiang, Lahiri, and Wan (2002), who proposed an MSE using an orthogonal decomposition, where the leading term in the MSE does not depend on the area-specific response and is nearly unbiased. Given that the authors can make an orthogonal decomposition, they can show that the order of the bias of the MSE is \(o(m^{-1})\), which contrasts our proposed approach. Under our approach, the leading term depends on the area-specific response, and thus, the bias is of order \(O(m^{-1})\). Turning to the bootstrap, we utilize methods similar to Butar and Lahiri (2003). Using this approach, we propose a parametric bootstrap estimator of the MSPE of our estimator. In a similar manner to the jackknife, the order of the bias for the parametric bootstrap estimator of the MSPE is \(O(m^{-1})\).

2. Area-level logarithmic model with measurement error

Consider \(m\) small areas and let \(Y_i (i = 1, ..., m)\) denote the population characteristic of interest in area \(i\), where often the information of interest is a population mean or proportion. A primary survey provides a direct estimator \(y_i\) of \(Y_i\) for some or all of the \(m\) small areas. In this section, we propose a measurement error model suitable for the inference of positively skewed response variable \(y_i\). To achieve normality in the response variable, we therefore propose an area-level log-measurement error model on \(Y_i\):

Consider the following model:

\[
Z_i = \theta_i + e_i, \tag{1}
\]

where \(Z_i := \log y_i\), \(\theta_i := \log Y_i\), and \(e_i\) is the sampling error distributed as \(e_i \sim \mathcal{N}(0, \psi_i)\). Assume

\[
\theta_i = \sum_{k=1}^{p} \beta_k \log X_{ik} + \nu_i,
\]

where \(X_{ik}\) is the \(k\)-th covariate of the \(i\)-th small area, which is unknown but is observed by \(x_{ik}\). The regression coefficient \(\beta_k\) is unknown and must be estimated, and \(\nu_i\) is the random effect distributed as \(\nu_i \sim \mathcal{N}(0, \sigma^2_i)\), where \(\sigma^2_i\) is unknown.

Our measurement error model for the case of positively skewed \(X_{ik}\)’s is proposed as

\[
w_{ik} := \log x_{ik} = \log X_{ik} + \eta_{ik}, \quad k = 1, ..., p,
\]

or in a vector form

\[
w_i = W_i + \eta_i, \quad \eta_i \sim \mathcal{N}_p(0, \Sigma_i), \tag{2}
\]

where \(w_i = (w_{i1}, ..., w_{ip})^T\) and \(W_i = (W_{i1}, ..., W_{ip})^T\) for \(W_{ik} = \log X_{ik}\). Note that in Eq. (2), \(W_i\) is non-stochastic within the class of functional measurement error models (c.f. Fuller (2006)). We assume \(\Sigma_i\) is known, and if it is unknown, it can be estimated using microdata or from another independent survey. We refer to Arima et al. (2017) for further details of estimating \(\Sigma_i\).
Now, one can write

\[
\begin{align*}
\begin{cases}
z_i &= W_i^\top \beta + \nu_i + \beta^\top \eta_i + \varepsilon_i \\
\theta_i &= W_i^\top \beta + \nu_i + \beta^\top \eta_i
\end{cases}
\end{align*}
\]

where \( \beta = (\beta_1, \ldots, \beta_p)^\top \). Thus, for the pair \((z_i, \theta_i)\), we have the following joint normal distribution

\[
\begin{align*}
\begin{pmatrix}
z_i \\
\theta_i
\end{pmatrix}
\sim N_2\left[
\begin{pmatrix}
W_i^\top \beta \\
W_i^\top \beta
\end{pmatrix},
\begin{pmatrix}
\beta^\top \Sigma_i \beta + \sigma^2_\nu + \psi_i & \beta^\top \Sigma_i \beta + \sigma^2_\nu \\
\beta^\top \Sigma_i \beta + \sigma^2_\nu & \beta^\top \Sigma_i \beta + \sigma^2_\nu + \psi_i
\end{pmatrix}
\right].
\end{align*}
\]

We assume all the sources of errors \((e_i, \nu_i, \eta_i)\) for \(i = 1, \ldots, m\) are mutually independent throughout the rest of the paper.

**Remark 1.** Eq. (1) is a Fay-Herriot model for \(z_i\), however, the parameter of interest is \(Y_i := \exp(\theta_i)\) rather than \(\theta_i\). Slud and Maiti (2006) and Ghosh, Kubokawa, and Kawakubo (2015) used a similar model in the absence of measurement errors in the covariates.

Next, we give the following conditional distribution \(\theta_i|z_i\) to later justify our Bayesian interpretation of the unknown interested parameter \(Y_i\):

\[
\theta_i|z_i \sim N\left(W_i^\top \beta + \frac{\beta^\top \Sigma_i \beta + \sigma^2_\nu + \psi_i}{\beta^\top \Sigma_i \beta + \sigma^2_\nu + \psi_i} (z_i - W_i^\top \beta), \beta^\top \Sigma_i \beta + \sigma^2_\nu - \frac{(\beta^\top \Sigma_i \beta + \sigma^2_\nu)^2}{\beta^\top \Sigma_i \beta + \sigma^2_\nu + \psi_i}\right),
\]

i.e.,

\[
\theta_i|z_i \sim N(\gamma_i z_i + (1 - \gamma_i)W_i^\top \beta, \gamma_i \psi_i),
\]

where \(\gamma_i = (\beta^\top \Sigma_i \beta + \sigma^2_\nu)/ (\beta^\top \Sigma_i \beta + \sigma^2_\nu + \psi_i)\).

Recall that the parameter of interest is \(Y_i := \exp(\theta_i)\) after transforming from the logarithmic scale back to the original scale. Therefore, the corresponding Bayes predictor is given by \(\hat{Y}_i := E(Y_i|z_i)\). By using the moment generating function of the normal distribution of \(\theta_i|z_i\), the Bayes predictor has the form of \(\hat{Y}_i = \exp\{\gamma_i z_i + (1 - \gamma_i)W_i^\top \beta + \gamma_i \psi_i/2\}\). In practice, \(W_i\) is unobserved, and since \(E(w_i) = W_i\), we can replace it with the observed \(w_i\). Also, \(\beta\) and \(\sigma^2_\nu\) are unknowns, and we need to replace them with their consistent estimators. Therefore, the EB predictor of \(Y_i\) is

\[
\hat{Y}_i^{EB} = \exp\left\{\gamma_i z_i + (1 - \gamma_i)W_i^\top \beta + \frac{\gamma_i \psi_i}{2}\right\}.
\]

**2.1. Estimation of unknown parameters**

In this section, we discuss estimation of the unknown parameters \(\beta\) and \(\sigma^2_\nu\). First, an estimator of \(\beta\) is obtained by solving the equation

\[
\sum_{i=1}^{m} [D_i(w_iw_i^\top - \Sigma_i)] \beta = \sum_{i=1}^{m} D_iw_i z_i.
\]

The justification for Eq. (4) is as follows. Let \(z = (z_1, \ldots, z_m)^\top\) and \(W^\top = (W_1, \ldots, W_m)\). Then, \(z \sim N_m(W\beta, D^{-1})\) where \(D^{-1} = \text{diag}(D_1^{-1}, \ldots, D_m^{-1})\) and \(D_i^{-1} = \beta^\top \Sigma_i \beta + \sigma^2_\nu + \psi_i\). Hence, an estimator of \(\beta\) is obtained by solving

\[
\beta = (W^\top D W)^{-1} W^\top D z = \left(\sum_{i=1}^{m} D_i W_i W_i^\top\right)^{-1} \sum_{i=1}^{m} D_i W_i z_i.
\]
Now, notice that \( E(w_i w_i^\top) = W_i W_i^\top + \Sigma_i \) and \( E(w_i) = W_i \). Hence, we estimate \( \beta \) from

\[
\sum_{i=1}^{m} [D_i(w_i w_i^\top - \Sigma_i)] \beta = \sum_{i=1}^{m} D_i w_i z_i.
\]

However, \( D_i \) is not known as both \( \beta \) and \( \sigma^2_\nu \) are unknowns. Take \( E(z_i - w_i^\top \beta)^2 = \sigma^2_\nu + \psi_i \). Then, \( \sigma^2_\nu \) can be estimated from

\[
m^{-1} \sum_{i=1}^{m} (z_i - w_i^\top \beta)^2 - m^{-1} \sum_{i=1}^{m} \psi_i.
\]

If the above is less than zero, estimate \( \sigma^2_\nu \) as zero. One can estimate \( \beta \) and \( \sigma^2_\nu \) by iteratively solving the Eqs. (4) and (5).

### 2.2. Mean squared prediction error of the EB predictor

In this section, we first define the MSPE of the EB predictor \( \hat{Y}^{\text{EB}}_i \). Second, we show that the cross-product term of the MSPE of the EB predictor \( \hat{Y}^{\text{EB}}_i \) is exactly zero. Now, we introduce notation that will be used throughout the rest of the paper. Let

\[
M_{1i} := E[(\hat{Y}_i - Y_i)^2|z_i] = \exp{\{\psi_i|\gamma_i\}}\exp{\{\psi_i|\gamma_i\}} - 1 \exp{\{2[\gamma_i z_i + (1 - \gamma_i) w_i^\top \beta] \}}
\]

\[
M_{2i} := E[(\hat{Y}^{\text{EB}}_i - \hat{Y}_i)^2|z_i], \quad M_{3i} := E[(\hat{Y}^{\text{EB}}_i - \hat{Y}_i)(\hat{Y}_i - Y_i)|z_i].
\]

Note that we estimate \( W_i \) with \( w_i \), and the term \( M_{1i} \) depends on the area-specific response variable \( z_i \) unlike Jiang, Lahiri, and Wan (2002), and its estimator has bias of order \( O(m^{-1}) \). Since we wish to include the uncertainty of the EB predictor \( \hat{Y}^{\text{EB}}_i \) with respect to the unknown parameters \( \beta \) and \( \sigma^2_\nu \), we decompose the MSPE into three terms using Definition 1.

**Definition 1.** The MSPE of the EB predictor \( \hat{Y}^{\text{EB}}_i \) is

\[
\text{MSPE}(\hat{Y}^{\text{EB}}_i) = E[(\hat{Y}^{\text{EB}}_i - Y_i)^2|z_i] = E[(\hat{Y}_i - Y_i)^2|z_i] + E[(\hat{Y}^{\text{EB}}_i - \hat{Y}_i)^2|z_i] + 2E[(\hat{Y}^{\text{EB}}_i - \hat{Y}_i)(\hat{Y}_i - Y_i)|z_i] = M_{1i} + M_{2i} + 2M_{3i},
\]

where we show below that \( M_{3i} = 0 \).

To show that the cross product term, \( M_{3i} \) goes to 0, recall the Bayes estimator is

\[
E[\hat{Y}_i] = E[Y_i|z_i] \Rightarrow E[\hat{Y}_i - Y_i|z_i] = 0.
\]

Consider

\[
M_{3i} = E[(\hat{Y}^{\text{EB}}_i - \hat{Y}_i)(\hat{Y}_i - Y_i)|z_i] = E[(\hat{Y}^{\text{EB}}_i - \hat{Y}_i)E((\hat{Y}_i - Y_i)|z_i)] = 0.
\]

### 3. Jackknife and parametric bootstrap estimators of the MSPE

In this section, we propose two estimators for the MSPE of the EB predictor \( \hat{Y}^{\text{EB}}_i \). First, we propose a jackknife estimator of the MSPE. Second, we propose a parametric bootstrap estimator of
the MSPE. The expectation of the proposed measure of uncertainty based on both methods is correct up to the order $O(m^{-1})$ for the EB predictor.

3.1. Jackknife estimator of the MSPE

In this section, we propose a jackknife estimator of the MSPE of the EB predictor $\hat{Y}_{i}^{EB}$, denoted by $\text{mspe}_j(\hat{Y}_{i}^{EB})$. We prove the order of the bias of $\text{mspe}_j(\hat{Y}_{i}^{EB})$ is correct up to the order $O(m^{-1})$ under six regularity conditions. We propose the following jackknife estimator:

$$\text{mspe}_j(\hat{Y}_{i}^{EB}) = \hat{M}_{1i,j} + \hat{M}_{2i,j}$$

where the data are based on all the areas other than $j$-th area. Therefore, let

$$\hat{M}_{1i} := M_{1i}(\hat{\sigma}_v^2, \hat{\beta})$$

$$= \exp \{ \psi(\hat{\gamma}) \} \left[ \exp \{ \psi(\hat{\gamma}) \} - 1 \right] \exp \left\{ 2 \left[ \hat{\gamma} \cdot z_i + (1 - \hat{\gamma})w_i^T \hat{\beta} \right] \right\},$$

$$\hat{M}_{1(-j)} = \exp \{ \psi(\hat{\gamma}_{(-j)}) \} \left[ \exp \{ \psi(\hat{\gamma}_{(-j)}) \} - 1 \right] \exp \left\{ 2 \left[ \hat{\gamma}_{(-j)} \cdot z_i + (1 - \hat{\gamma}_{(-j)})w_i^T \hat{\beta}_{(-j)} \right] \right\},$$

and

$$\hat{Y}_{i}^{EB} = \exp \left\{ \hat{\gamma}_{(-j)} \cdot z_i + (1 - \hat{\gamma}_{(-j)})w_i^T \hat{\beta}_{(-j)} + \frac{\psi(\hat{\gamma}_{(-j)})}{2} \right\}.$$ 

Note that for all $[\cdot]_{(-j)}$ cases, the estimators of $\phi = (\beta, \sigma_v^2) \in \Theta$ where $\Theta$ is a compact set such that $\Theta \subseteq (\mathbb{R}^p, \mathbb{R}^+)$. Let $\ell(\cdot|z_i)$ denote the conditional likelihood function. We define the corresponding first, second, and third derivatives of the conditional likelihood function by $\ell_i(\phi|z_i), \ell''_i(\phi|z_i)$, and $\ell'''_i(\phi|z_i)$, respectively. Now, assume the following six regularity conditions:

Condition 1. Define $\phi^\dagger = (\beta, \sigma_v^2) \in \Theta$ where $\Theta$ is a compact set such that $\Theta \subseteq (\mathbb{R}^p, \mathbb{R}^+)$. Assume $\phi$ is a consistent estimator for $\phi$, i.e., $\hat{\phi} \rightarrow \phi$. Assume $\ell_i(\phi|z_i)$ and $\ell'''_i(\phi|z_i)$ both exist for $i = 1, \ldots, m$, almost surely in probability.

Condition 4. Assume $E\{\ell_i(\phi|z_i)|\phi\} = 0$ for $i = 1, \ldots, m$.

Condition 5. Assume $\ell''_i(\phi|z_i)$ is a continuous function of $\phi$ for $i = 1, \ldots, m$, almost surely in probability, where $E\{\ell''_i(\phi|z_i)\}$ is positive definite, uniformly bounded away from $0$, and is a measurable function of $z_i$.

Condition 6. Assume $E\{|\ell_i'(\phi|z_i)|^{4+\delta}\}, E\{|\ell''_i(\phi|z_i)|^{4+\delta}\}$, and $E\{\sup_{c \in (-\epsilon, \epsilon)} |\ell'''_i(\phi + c|z_i)|^{4+\delta}\}$ are uniformly bounded for $i = 1, \ldots, m$ under some $\epsilon > 0$ and $\delta > 0$.

Theorem 1. Assume Conditions 1–6 hold. Then

$$E\left[ \text{mspe}_j(\hat{Y}_{i}^{EB}) \right] = \text{MSPE}(\hat{Y}_{i}^{EB}) + O(m^{-1}).$$
Proof. Define

\[
E(\text{mspe}_f(\hat{Y}_i^{EB})) = E(\hat{M}_{1i,j} + \hat{M}_{2i,j})
= E\left(\hat{M}_{1i} - \frac{m-1}{m} \sum_{j=1}^{m} \left[ (\hat{M}_{1i} - \hat{M}_{1i(-j)} | z_i) \right] \right)
+ \frac{m-1}{m} E\left(\sum_{j=1}^{m} \left[ (\hat{Y}_i^{EB} - \hat{Y}_{i(-j)}^{EB})^2 | z_i \right] \right).
\]

Also, define a remainder term \( r_i \) that is bounded in absolute value by \( R_i \) such that,
\[
|r_i| \leq \max\{1, |\ell'(\phi|z_i)|^3, |\ell''(\phi|z_i)|^3, |\ell'''(\phi|z_i)|^3 \} \equiv R_i.
\]

First, we prove \( \hat{M}_{1i,j} \) has a bias of order \( O(m^{-1}) \). Using a Taylor series expansion, we find that

\[
\hat{M}_{1i} = M_{1i} + M_{1i}^T(\phi)(\hat{\phi} - \phi) + \frac{1}{2} M_{1i}^{\sigma T}(\phi)(\hat{\phi} - \phi)^2 + \frac{1}{6} M_{1i}^{\sigma \sigma T}(\phi^*)(\hat{\phi} - \phi)^3,
\]

for \( \phi^* \) between \( \phi \) and \( \hat{\phi} \). Also, \( M_{1i}^T(\phi), M_{1i}^{\sigma T}(\phi), \) and \( M_{1i}^{\sigma \sigma T}(\phi^*) \) stand for the first, second, and third derivatives of \( M_{1i} \) with respect to \( \phi \). Let \( \hat{\phi}^T = (\hat{\beta}, \hat{\sigma}_i^2) \), and it follows that

\[
\hat{M}_{1i} - \hat{M}_{1i(-j)} = \hat{M}_{1i}^T(\hat{\phi})(\hat{\phi} - \hat{\phi}_{(-j)}) + \frac{1}{2} \hat{M}_{1i}^{\sigma T}(\hat{\phi})(\hat{\phi} - \hat{\phi}_{(-j)})^2 + \frac{1}{6} \hat{M}_{1i}^{\sigma \sigma T}(\hat{\phi}^*)(\hat{\phi} - \hat{\phi}_{(-j)})^3,
\]

for some \( \hat{\phi}^* \) between \( \hat{\phi}_{(-j)} \) and \( \hat{\phi} \). In order to approximate the solution \( \hat{\phi} \) of the equation \( f(\tau) = \sum_{i=1}^{m} \ell'(\tau|z_i) = 0 \) in iteration \( (\xi + 1) \), we use Householder’s method (Householder (1970), Theorem 4.4.1). See also Theorem 1 of Lohr and Rao (2009):

\[
\tau_{\xi+1} = \tau_{\xi} - \frac{f(\tau_{\xi})}{f'(\tau_{\xi})} \left[ 1 + \frac{\tau_{\xi} f''(\tau_{\xi})}{2(f'(\tau_{\xi}))^2} \right].
\]

By taking the initial value \( \tau_0 = \phi \), we have

\[
\hat{\phi} - \phi = -\sum_{i=1}^{m} \ell'_i(\phi|z_i) \left\{ 1 + \sum_{k=1}^{m} \ell''_k(\phi|z_k) \frac{\sum_{r=1}^{m} \ell'''_r(\phi|z_r)}{2(\sum_{k=1}^{m} \ell''_k(\phi|z_k))^2} \right\} + O_p(|\hat{\phi} - \phi|^{3}),
\]

and

\[
\hat{\phi}_{(-j)} - \hat{\phi}_{(-j)} = \frac{\ell'_j(\phi|z_j)}{\sum_{k \neq j} \ell'_k(\phi|z_k)} \left[ 1 - \frac{\sum_{k \neq j} \ell''_k(\phi|z_k)}{2(\sum_{k \neq j} \ell''_k(\phi|z_k))^2} \right] + O_p(|\hat{\phi} - \hat{\phi}_{(-j)}|^{3}).
\]

By taking conditional expectation and using Theorem 2.1 of Jiang, Lahiri, and Wan (2002), we find that

\[
E(\hat{\phi} - \phi|z_i) = -\frac{\ell'_i(\phi|z_i) + \varphi}{\sum_{i=1}^{m} E(\ell'_i(\phi|z_i))} + r_i o(m^{-1}),
\]

where
By combining the above results, we find that
\[ E\{\hat{M}_{1i,i} - M_{1i,i}\} = -M_{1i,i}(\phi|z_i)\ell_i'(\phi|z_i)/\varphi + r_i o(m^{-1}). \]
Hence, \(E(\hat{M}_{1i,i}) = M_{1i,i} + O(m^{-1}).\)

Second, we prove \(\hat{M}_{2i} \) has a bias of order \(o(m^{-1}).\) Let
\[ \hat{Y}_i - \hat{Y}_{i(-j)} := h(\hat{\phi}|z_i) - h(\hat{\phi}_{(-j)}|z_i), \]
and \(h(\phi|z_i) = E(Y_{iEB}|z_i, \phi).\) Using a Taylor series expansion, we find that
\[ \hat{Y}_i - \hat{Y}_{i(-j)} = h^T(\hat{\phi}|z_i)(\hat{\phi} - \hat{\phi}_{(-j)}) + \frac{1}{2} h''^T(\hat{\phi}_{(-j)}|z_i)(\hat{\phi} - \hat{\phi}_{(-j)})^2, \]
where
\[ h^T(\hat{\phi}|z_i) = \left( \frac{\partial h(\hat{\phi}|z_i)}{\partial \beta}, \frac{\partial h(\hat{\phi}|z_i)}{\partial \sigma^2_{z_i}} \right), \quad h''^T(\hat{\phi}|z_i) = \left( \frac{\partial^2 h(\hat{\phi}|z_i)}{\partial^2 \beta}, \frac{\partial^2 h(\hat{\phi}|z_i)}{\partial^2 \sigma^2_{z_i}} \right), \]
and \(\hat{\phi}_{(-j)}\) is between \(\hat{\phi}_{(-j)}\) and \(\hat{\phi}.\) Using an additional Taylor series expansion, we find that
\[ \sum_{j=1}^{m} E((\hat{Y}_{i(-j)} - \hat{Y}_i|z_i)^2) = \{h^T(\hat{\phi}|z_i)\}^2 \times \frac{\sum_{j=1}^{m} E((\ell_j'(\phi|z_i))^2)}{\varphi^2} + r_i o(m^{-1}). \]
Similarly,
\[ E((\hat{Y}_i - \hat{Y}_i|z_i)^2) = \{h^T(\phi|z_i)\}^2 \times \frac{\sum_{j=1}^{m} E((\ell_j'(\phi|z_i))^2)}{\varphi^2} + r_i o(m^{-1}). \]
By combining the above results, we find that
\[ E(\hat{M}_{2i,i}) = M_{2i,i} + o(m^{-1}). \]
Finally,
\[ E(\text{mspe}_f(\hat{Y}_i)) = E(\hat{M}_{1i,i}) + E(\hat{M}_{2i,i}) \]
\[ = \{M_{1i,i} + O(m^{-1})\} + \{M_{2i,i} + o(m^{-1})\} \]
\[ = M_{1i,i} + M_{2i,i} + O(m^{-1}). \]
Hence, \(E[\text{mspe}_f(\hat{Y}_i)] = \text{MSPE}(\hat{Y}_i) + O(m^{-1}).\) \(\square\)
3.2. Parametric bootstrap estimator of the MSPE

In this section, we propose a parametric bootstrap estimator of the MSPE of the EB predictor \( \hat{Y}_i^{EB} \), which we denote it by \( \text{mspe}_B(\hat{Y}_i^{EB}) \). We prove that the order of the bias is correct up to order \( O(m^{-1}) \). Specifically, we extend Butar and Lahiri (2003) to find a parametric bootstrap of our proposed EB predictor. To introduce the parametric bootstrap method, consider the following bootstrap model:

\[
\begin{align*}
\text{mspe}_B(\hat{Y}_i^{EB}) &= 2M_{1i}(\hat{\sigma}_\nu^2, \hat{\beta}) - E_*\left[M_{1i}(\hat{\sigma}_\nu^2, \hat{\beta}^*)|z_i^*)\right] + E_*\left[(\hat{Y}_i^{EB})^2 - \hat{Y}_i^{EB} \right]^2 |z_i^*],
\end{align*}
\]

which has bias of order \( O(m^{-1}) \) as shown in the Theorem 2.

**Theorem 2.** Assume \( E_*(\hat{\sigma}_\nu^2 - \hat{\sigma}_\nu^2) = O_p(m^{-1}) \) and \( E_*({\hat{\beta}}^* - \hat{\beta}) = O_p(m^{-1}) \). The bootstrap estimator of the MSPE has bias of order \( O(m^{-1}) \), i.e.,

\[ E\left[\text{mspe}_B(\hat{Y}_i^{EB})\right] = \text{MSPE}(\hat{Y}_i^{EB}) + O(m^{-1}). \]

**Proof.** Let

\[ E_*\left[M_{1i}(\hat{\sigma}_\nu^2, \hat{\beta})|z_i^*)\right] = M_{1i}(\hat{\sigma}_\nu^2, \hat{\beta}) + O_p(m^{-1}). \]

Assume that \( R_m^* = O_p(m^{-1}) \) such that \( mR_m^* \) is bounded in probability under the parametric bootstrap model in Eq. (8). Consider the following Taylor series expansion:

\[ \hat{Y}_i^{EB^*} - \hat{Y}_i^{EB} = (\hat{\phi}^* - \hat{\phi})^T h'(\hat{\phi}|z_i) + R_m^*, \]

such that \( \hat{\phi}^* = (\hat{\beta}^*, \hat{\sigma}_\nu^2). \)

Using an argument similar to the proof of Theorem 1,

\[ \text{mspe}_B(\hat{Y}_i^{EB}) = 2M_{1i} - \left[\hat{M}_{1i} + O_p(m^{-1})\right] + \hat{M}_{2i} + o_p(m^{-1}) \]

Substituting Eq. (10) into Eq. (9), we find that

\[ \text{mspe}_B(\hat{Y}_i^{EB}) = \hat{M}_{1i} + \hat{M}_{2i} + O_p(m^{-1}). \]
This suggests that

\[ E \left[ \text{mspe}_B(\hat{Y}_i^{\text{EB}}) \right] = \text{MSPE}(\hat{Y}_i^{\text{EB}}) + O(m^{-1}). \]

\[ \square \]

### 4. Experiments

In this section, we investigate the performance of the EB predictors in comparison to the direct estimators through design-based and model-based simulation studies. In addition, we evaluate the MSPE estimators using both a jackknife and a parametric bootstrap.

#### 4.1. Design-based simulation study

In this section, we consider a design-based simulation study using the CoG data set as described in Sec. 1.1.

##### 4.1.1. Design-based simulation setup

We describe the design-based simulation setup. The parameter of interest is average number of full-time employees per government at the state level from 2012 data set. The covariate is the average number of full-time employees per government at the state level from the 2007 data set. There are observed skewed patterns in the average number of full-time employees in both 2007 and 2012, which motivates our proposed framework.

For the response variable, we select a total sample of 7,000 governmental units proportionally allocated to the states and for the covariates, we select a total sample of 70,000 units and the survey-weighted averages were then calculated. The measurement error variance \( \Sigma_i \) was obtained from a Taylor series approximation, where \( \text{Var}(x_i) \) was estimated from the formula of variance in simple random sampling without replacement at each state. The \( \psi_i \)'s were estimated by a Generalized Variance Function (GVF) method (see Fay and Herriot (1979)). We assume the sampling variances to be known throughout the estimation procedure.

For the design-based simulation, we draw 1,000 samples and estimate the parameters from each sample. We evaluate our proposed predictors by empirical MSE per each state \( i \):

\[ \text{EMSE}(\hat{Y}_i) = \frac{1}{R} \sum_{r=1}^{R} \left[ \hat{Y}^{(r)}_i - Y_i^{(r)} \right]^2, \]

where \( R = 1,000 \) is the total number of replications, and \( \hat{Y}_i \) is the estimator of \( Y_i \). In addition, when the parametric bootstrap it used, we take \( B = 1,000 \) bootstrap samples. We use the same number of replications and bootstrap samples in the design and model-based simulation studies.

##### 4.1.2. Design-based simulation results

In this section, we provide the results of the design-based simulation study.

##### 4.1.2.1. Investigating the performance of the proposed estimators

Recall that the covariate of interest is the average number of full-time employees per government at the state level from 2007 data set, and we wish to predict the average number of full-time employees per government at the state level in 2012. To do so, we give the predictors for each state as well as their corresponding EMSE’s in Tables 1 and 2. More specifically, we compare the following three estimators:

1) \( \hat{y}_i \): the direct estimator,
$\hat{Y}_i$: the EB predictor, assuming the true covariate $w_i$ and ignoring $\Sigma_i$ in our model,

$\hat{Y}_{iEB}$: the EB predictor, assuming the true covariate $w_i$ has measurement error, where $\Sigma_i$ is included in our model.
We observe that in most cases the $\text{EMSE}(\hat{Y}_{i}^{EB})$ is smaller than the $\text{EMSE}(\tilde{Y}_{i})$.

However, we observe that our proposed EB predictor does not always outperform the direct estimator, which we further explore in our model-based simulation studies in Sec. 4.2.

4.1.2.2. Jackknife versus parametric bootstrap estimators. Next, we consider the performance of MSPE estimators, i.e., the jackknife and bootstrap, with respect to the true MSE, i.e., $\text{EMSE}(\hat{Y}_{i}^{EB})$ in Figure 1. The results are given on the logarithmic scale, and we observe that the distribution of jackknife is closer to the distribution of true MSE when compared to the bootstrap. Therefore, we recommend the jackknife given that it slightly overestimates the true MSE. As already mentioned, given that our proposed estimator does not uniformly beat the direct estimator in terms of the EMSE, we conduct a model-based simulation study in Sec. 4.2 to investigate this and provide further insight.

4.2. Model-based simulation study

In this section, we describe our model-based simulation study to further investigate the performance of the proposed EB predictor $\hat{Y}_{i}^{EB}$. Second, we compare the proposed jackknife and parametric bootstrap estimators, $\text{mspe}_{J}(\hat{Y}_{i}^{EB})$ and $\text{mspe}_{B}(\hat{Y}_{i}^{EB})$. Third, we investigate how often the variance estimates $\hat{\sigma}_{r}^{2}$ are zero. Finally, we investigate how the regression parameter changes when $\Sigma_{i}$ is misspecified. Our goal through this model-based simulation study is to understand how one could improve the EB predictor through future research and to further understand its underlying behavior.
Table 3. Model-based simulation setup with definition of parameters and distributions.

| Simulation Setup: |
|-------------------|
| Generate $W_i$ from a Normal(5,9) and $\psi_i$ from a Gamma(4.5,2) |
| Take $\theta_i = 3W_i + \nu_i$, $z_i = \theta_i + e_i$, and $w_i = W_i + \eta_i$ |
| $\nu_i \sim$ Normal$(0, \sigma^2)$, $e_i \sim$ Normal$(0, \psi_i)$, and $\eta_i \sim$ Normal$(0, \Sigma_i)$ |
| Take $y_i = \exp(x_i)$ and $y_i = \exp(\theta_i)$ |
| Parameter Definition: |
| Let $m = 20, 50, 100$, and 500 (number of small areas) |
| Let $\sigma^2 = 2$ (for all cases) |
| Let $k \in \{0, 20, 50, 80, 100\}$ |
| $\Sigma_i \in \{0, d\}$, where $d = 2$ or $4$ |
| Allow $k\%$ of the $\Sigma_i$’s randomly receive $d$ and the rest 0. |

4.2.1. Model-based simulation setup

In this section, we provide the setup of our model-based simulation study in Table 3. This setup follows Eqs. (1) and (2). We are interested in comparing the following four estimators:

1) $y_i$: the direct estimator,
2) $\hat{Y}_i$: the EB predictor, assuming the true covariate $W_i$,
3) $\hat{Y}_{iEB}$: the EB predictor, assuming the true covariate $w_i$ and ignoring $\Sigma_i$ in our model,
4) $\tilde{Y}_{iEB}$: the EB predictor, assuming the true covariate $w_i$ has measurement error, where $\Sigma_i$ is included in our model.

We compare these four estimators (for each area $i$) using the empirical MSE:

$$EMSE(\hat{Y}_i) = \frac{1}{R} \sum_{r=1}^{R} \left( \hat{Y}_i^{(r)} - Y_i^{(r)} \right)^2,$$

where $\hat{Y}_i$ is the estimator of $Y_i$.

In order to evaluate the jackknife and parametric bootstrap estimators of $\hat{Y}_{iEB}$, we consider the relative bias, denoted by $RB_j(\hat{Y}_{iEB})$ and $RB_B(\hat{Y}_{iEB})$, respectively. More specifically, the relative biases are defined as follows for each area $i$:

$$RB_j(\hat{Y}_{iEB}) = \left\{ \frac{1}{R} \sum_{r=1}^{R} mspe_j^{(r)}(\hat{Y}_{iEB}^{(r)}) - EMSE(\hat{Y}_{iEB}) \right\} / EMSE(\hat{Y}_{iEB}),$$

$$RB_B(\hat{Y}_{iEB}) = \left\{ \frac{1}{R} \sum_{r=1}^{R} mspe_B^{(r)}(\hat{Y}_{iEB}^{(r)}) - EMSE(\hat{Y}_{iEB}) \right\} / EMSE(\hat{Y}_{iEB}).$$

4.2.2. Model-based simulation results

In this section, we summarize our results of the model-based simulation study.

4.2.2.1. Investigating the performance of the proposed estimators. In this section, we investigate the performance of the proposed estimators. Table 4 provides the four estimators given in Sec. 4.2.1 with their empirical MSEs, where we average the results over all the small areas and re-scale them using the logarithmic scale. When $k = 0$, the MSE’s for all EB predictors are the same since the term $\Sigma_i$ vanishes and $w_i$ is the same as $W_i$. Overall, as the value of $k$ increases, the empirical MSE increases for almost all predictors. We observe there are cases in which the EB predictors cannot outperform the direct estimators based on the simulation results.

Table 4 illustrates that there are cases in which the EMSE($\tilde{Y}_{iEB}$) is larger than the EMSE($y_i$). In fact, the EB predictors cannot outperform the direct estimators due to propagated errors in the term $\beta^T \Sigma_i \beta$, which is present in the term $\gamma_i$ in the EB predictors through the simulations; see Eq. (3). Therefore, as the measurement error variance $\Sigma_i$ increases, we have shown that the MSE
of our proposed EB predictors can also increase. This is the main point that one should notice when using a log-model with measurement error.

4.2.2.2. Jackknife versus parametric bootstrap estimators. We compare the jackknife MSPE estimator of the EB predictor $\hat{Y}_{\text{EB}}^i$ to that of the bootstrap using the relative bias (see Table 5). In addition, we consider box plots for the jackknife and bootstrap MSPE estimators of the EB predictor $\hat{Y}_{\text{EB}}^i$, where we compare these to box plots of the true values (see Figure 2). Both Table 5 and Figure 2 illustrate that the bootstrap receives a large number of negative values, which is due to the construction of $\hat{M}_{ij}$. Here, we find that the bootstrap grossly underestimates the true values, whereas the jackknife slightly overestimates the true values. This could be due to generating data from the normal distribution and the non-linear transformation in the model. Thus, we would recommend the jackknife in practice.

4.2.2.3. Amount of zeros for the estimates of $\hat{\sigma}_i^2$. Here, we investigate the proportion of zero estimates for $\hat{\sigma}_i^2$, based on iteratively solving the Eqs. (4) and (5). Figure 3 illustrates that as the number of small areas increases, the magnitude of receiving zeros decreases. More specifically, we observe when $m = 20$ and as $k$ increases, $\hat{Y}_{\text{EB}}^i$ and $\hat{Y}_i$ tend to have a proportion of zero estimates of $\hat{\sigma}_i^2$.
Table 5. Comparison of the proposed jackknife and bootstrap estimators from model-based simulations. The results are averaged over all small areas. The results are rescaled by the logarithm of absolute value. Note, * denote that the original value is negative.

| m  | k   | $EMSE(Y_{EB}^kB)$ | mspe$_j(Y_{EB}^kB)$ | mspe$_b(Y_{EB}^kB)$ | RB$_j(Y_{EB}^kB)$ | RB$_b(Y_{EB}^kB)$ |
|----|-----|-------------------|---------------------|---------------------|------------------|------------------|
| 20 | 0   | 111.09            | 120.731             | 118.478*            | 9.641            | 7.389*           |
|    | 20  | 108.425           | 115.222             | 115.884*            | 6.796            | 7.459*           |
|    | 50  | 94.223            | 107.245             | 111.255*            | 13.021           | 17.032*          |
|    | 80  | 97.495            | 113.06              | 129.772*            | 15.563           | 32.277*          |
|    | 100 | 102.48            | 115.336*            | 119.84*             | 13.056*          | 17.36*           |
| 50 | 0   | 106.167           | 112.318*            | 116.383*            | 6.154*           | 10.216*          |
|    | 20  | 100.255           | 110.449             | 106.743*            | 10.194           | 6.489*           |
|    | 50  | 105.547           | 115.365*            | 118.838*            | 9.818*           | 13.291*          |
|    | 80  | 103.615           | 114.127             | 112.717*            | 10.512           | 9.102*           |
|    | 100 | 100.319           | 110.454*            | 112.552*            | 10.134*          | 12.233*          |
| 100| 0   | 92.802            | 98.67               | 102.932*            | 5.866            | 10.13*           |
|    | 20  | 106.172           | 106.788*            | 108.623*            | 1.048*           | 2.534*           |
|    | 50  | 108.214           | 119.666             | 120.032*            | 11.452           | 11.819*          |
|    | 80  | 108.454           | 117.223*            | 119.418*            | 8.769*           | 10.964*          |
|    | 100 | 109.515           | 119.028             | 117.071*            | 9.513            | 7.557*           |
| 500| 0   | 103.465           | 108.38              | 110.455*            | 4.908            | 6.991*           |
|    | 20  | 108.169           | 119.672             | 115.892             | 11.502           | 7.722            |
|    | 50  | 110.006           | 121.191             | 125.754*            | 11.184           | 15.747*          |
|    | 80  | 112.805           | 125.672*            | 129.761*            | 12.867*          | 16.955*          |
|    | 100 | 113.217           | 123.037*            | 124.597*            | 9.819*           | 11.379*          |

Figure 3. The proportion of zero estimates of $\sigma^2_i$ from model-based simulation when we perform 1,000 replications of the simulation study for $k = 0, \ldots, 100, m = 20, 50, 100, 500,$ and $d = 2.$
Denote the second set of experiments by $A$. In addition, we compute the magnitude of bias related to variance $R$ zero estimates of $\sigma_{\nu}^2$ are summarized in Table 6. Recall that $4.2.2.4$. The effect of misspecification of $R$ cautious of this in practical applications, and adjusting for this is of the interest of future work.

Then, we compute the average absolute difference between the respective $\beta$'s by considering:

$$100 \times \frac{1}{R} \sum_{r=1}^{R} \left| \hat{\beta}^{(r)} - \hat{\beta}_{\text{mis}}^{(r)} \right|.$$ 

In addition, we compute the magnitude of bias related to $\hat{\beta}$ and $\hat{\beta}_{\text{mis}}$ with respect to the true value of $\hat{\beta} = 3$ as follows

$$100 \times \frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}^{(r)} - 3) \text{ and } 100 \times \frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}_{\text{mis}}^{(r)} - 3).$$

between 0.3 and 0.5. When $m = 50$ and as $k$ increases, $\hat{Y}_{\text{EB}}^i$ and $\hat{Y}_i$ tend to have a proportion of zero estimates of $\sigma_{\nu}^2$ between 0.15 and 0.4. When $m = 100$ and as $k$ increases, $\hat{Y}_{\text{EB}}^i$ and $\hat{Y}_i$ tend to have a proportion of zero estimates of $\sigma_{\nu}^2$ between 0.05 and 0.3. When $m = 500$ and as $k$ increases, $\hat{Y}_{\text{EB}}^i$ and $\hat{Y}_i$ tend to have a proportion of zero estimates of $\sigma_{\nu}^2$ between 0 and 0.05. One should be cautious of this in practical applications, and adjusting for this is of the interest of future work.

### 4.2.2.4. The effect of misspecification of $\Sigma_i$ on $\beta$

We investigate the effect of mis-specifying the variance $\Sigma_i$ on the estimation of the regression parameter $\beta$. To accomplish this, we conduct an empirical study based on the proposed model-based simulation study in Table 3 for the EB predictor $\hat{Y}_{\text{EB}}^i$. Assume $\beta = 3$, and we consider two sets of experiments for each value of $k$, which are summarized in Table 6. Recall that $\Sigma_i \in \{0, d\}$. Denote the first set of experiments by $A1, B1, C1, D1,$ and $E1$, where we assume $d = 2$. Denote the misspecified value of $d$ by $d_{\text{mis}} = 4$. Denote the second set of experiments by $A2, B2, C2, D2,$ and $E2$, where we assume $d = 4$ and $d_{\text{mis}} = 2$. We conduct both sets of experiments for $m = 20$ and 500. For each experiment, we estimate the unknown parameter $\beta$ under the followings: (1) the true value of $d$ denoted by $\beta$ and (2) the misspecified value of $d_{\text{mis}}$ denoted by $\hat{\beta}_{\text{mis}}$.

Then, we compute the average absolute difference between the respective $\beta$'s by considering:

$$100 \times \frac{1}{R} \sum_{r=1}^{R} \left| \hat{\beta}^{(r)} - \hat{\beta}_{\text{mis}}^{(r)} \right|.$$ 

In addition, we compute the magnitude of bias related to $\hat{\beta}$ and $\hat{\beta}_{\text{mis}}$ with respect to the true value of $\hat{\beta} = 3$ as follows

$$100 \times \frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}^{(r)} - 3) \text{ and } 100 \times \frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}_{\text{mis}}^{(r)} - 3).$$

| $m$ | $k$ | Experiment | $\frac{1}{R} \sum_{r=1}^{R} |\hat{\beta}^{(r)} - \hat{\beta}_{\text{mis}}^{(r)}|$ | $\frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}^{(r)} - 3)$ | $\frac{1}{R} \sum_{r=1}^{R} (\hat{\beta}_{\text{mis}}^{(r)} - 3)$ |
|-----|-----|----------------|---------------------------------|---------------------------------|---------------------------------|
| 20  | 0   | $A1$ ($d = 2, d_{\text{mis}} = 4$) | 0 | -0.026 | -0.026 |
|     |     | $A2$ ($d = 4, d_{\text{mis}} = 2$) | 0 | -0.237 | -0.237 |
| 20  | $B1$ ($d = 2, d_{\text{mis}} = 4$) | 6.034 | 0.314 | -0.046 |
|     | $B2$ ($d = 4, d_{\text{mis}} = 2$) | 5.814 | -0.634 | -0.591 |
| 50  | $C1$ ($d = 2, d_{\text{mis}} = 4$) | 11.287 | -0.184 | -0.864 |
|     | $C2$ ($d = 4, d_{\text{mis}} = 2$) | 10.977 | -0.157 | -0.394 |
| 80  | $D1$ ($d = 2, d_{\text{mis}} = 4$) | 19.641 | 0.842 | 2.177 |
|     | $D2$ ($d = 4, d_{\text{mis}} = 2$) | 18.722 | 0.712 | 0.327 |
| 100 | $E1$ ($d = 2, d_{\text{mis}} = 4$) | 25.774 | 2.650 | 4.285 |
|     | $E2$ ($d = 4, d_{\text{mis}} = 2$) | 25.967 | 4.020 | 2.937 |
| 500 | 0   | $A1$ ($d = 2, d_{\text{mis}} = 4$) | 0 | 0.185 | 0.185 |
|     |     | $A2$ ($d = 4, d_{\text{mis}} = 2$) | 0 | -0.082 | -0.082 |
| 20  | $B1$ ($d = 2, d_{\text{mis}} = 4$) | 1.136 | 0.174 | 0.118 |
|     | $B2$ ($d = 4, d_{\text{mis}} = 2$) | 1.107 | 0.010 | 0.009 |
| 50  | $C1$ ($d = 2, d_{\text{mis}} = 4$) | 2.200 | -0.138 | 0.026 |
|     | $C2$ ($d = 4, d_{\text{mis}} = 2$) | 2.182 | -0.044 | 0.006 |
| 80  | $D1$ ($d = 2, d_{\text{mis}} = 4$) | 3.365 | 0.204 | -0.067 |
|     | $D2$ ($d = 4, d_{\text{mis}} = 2$) | 3.386 | -0.139 | -0.090 |
| 100 | $E1$ ($d = 2, d_{\text{mis}} = 4$) | 5.086 | 0.174 | 0.152 |
|     | $E2$ ($d = 4, d_{\text{mis}} = 2$) | 4.941 | -0.022 | 0.117 |
Table 6 illustrates that the overall misspecification of $\Sigma_i$ leads to bias in $\beta$. When the magnitude of measurement error is zero (i.e., $k=0$), there is no difference between the estimated $\beta$ using $d$ or $d_{mis}$. On the other hand, when the magnitude of $k$ increases and we have more uncertainty in the error variance $\Sigma_o$, values of $\beta$ and $\beta_{mis}$ diverge more from one another, and the magnitude of the bias increases. Also, we observe as the number of small areas increases, the value of bias decreases. One can resolve this bias issue by constructing an adaptive estimator for $\beta_{mis}$ in which its bias is corrected through some techniques such as bootstrap and develop a test of parameter specification.

5. Discussion

In this paper, in order to stabilize the skewness and achieve normality in the response variable, we have proposed an area-level log-measurement error model on the response variable. In addition, we have proposed a measurement error model on the covariates. Second, under our proposed modeling framework, we derived the EB predictor of positive small area quantities subject to the covariates containing measurement error. Third, we proposed a corresponding estimate of MSPE using a jackknife and a bootstrap method, where we illustrated that the order of the bias is $O(m^{-1})$, where $m$ is the number of small areas. Fourth, we have illustrated the performance of our methodology in both design-based simulation and model-based simulation studies, where the EMSE of the proposed EB predictor is not always uniformly better than that of the direct estimator. Our model-based simulation studies have provided further investigation and guidance on the behavior. For example, one fruitful area of future research would be providing a correction to the EB predictor to avoid for such behavior. One way to address this issue is by estimating $\phi = (\beta, \sigma_o^2)^T$ in such a way that its order of bias is smaller than $O(m^{-1})$. This could help to reduce the amount of propagated errors in the EB predictor $\hat{Y}_{EB}$. Another way, which is less theoretical burdensome is by estimating the covariate as we have assumed it follows a functional measurement error model rather than a structural one. We have studied the MSPE of the EB predictor using both the jackknife and bootstrap in simulation studies, where we have shown the jackknife estimator performs better than the bootstrap one under our log model.

References

Arima, S., W. Bell, C. Datta, G. S. Franco, and B. Liseo. 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 (4):1191–209. doi:10.1111/rssa.12321.

Arima, S., G. S. Datta, and B. Liseo. 2015a. Accounting for measurement error in covariates in SAE: An overview in analysis of poverty data by small area methods. M. Pratesi Editor, New York: Wiley.

Arima, S., G. S. Datta, and B. Liseo. 2015b. Bayesian estimators for small area models when auxiliary information is measured with error. *Scandinavian Journal of Statistics* 42 (2):518–29. doi:10.1111/sjos.12120.

Bellow, M. E., and P. S. Lahiri. 2011. An empirical best linear unbiased prediction approach to small area estimation of crop parameters. National Agricultural Statistics Service, University of Maryland.

Berg, E., and H. Chandra. 2011. Small area estimation for skewed data in presence of zeros. *Calcutta Statistical Association Bulletin* 63 (1-4):241–58. doi:10.1177/0008068320110113.

Butar, F. B., and P. Lahiri. 2003. On measures of uncertainty of empirical Bayes small-area estimators. *Journal of Statistical Planning and Inference* 112 (1-2):63–76. doi:10.1016/S0378-3758(02)00323-3.

Chandra, H., and R. Chambers. 2011. Small area estimation for skewed data in presence of zeros. *Calcutta Statistical Association Bulletin* 63 (1-4):241–58. doi:10.1177/0008068320110113.

Fay, R. E., and R. A. Herriot. 1979. Estimates of income for small places: An application of James–Stein procedures to census data. *Journal of the American Statistical Association* 74 (366a):269–77. doi:10.1080/01621459.1979.10482505.

Fuller, W. A. 2006. *Measurement error models*. New York: John Wiley & Sons.

Ghosh, M., T. Kubokawa, and Y. Kawakubo. 2015. Benchmarked empirical Bayes methods in multiplicative area-level models with risk evaluation. *Biometrika* 102 (3):647–59. doi:10.1093/biomet/asv010.

Householder, A. S. 1970. *The numerical treatment of a single nonlinear equation*. McGraw-Hill.
Jiang, J., P. Lahiri, and S. M. Wan. 2002. A unified jackknife theory for empirical best prediction with m-estimation. *The Annals of Statistics* 30 (6):1782–810. doi:10.1214/aos/1043351257.

Lohr, S. L., and J. Rao. 2009. Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models. *Biometrika* 96 (2):457–68. doi:10.1093/biomet/asp003.

Pfefferman, D. 2013. New important developments in small area estimation. *Statistical Science* 28 (1):40–68.

Rao, J. N., and I. Molina. 2015. *Small Area Estimation*. London: Wiley Series in Survey Sampling.

Slud, E. V., and T. Maiti. 2006. Mean-squared error estimation in transformed Fay–Herriot models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (2):239–57. doi:10.1111/j.1467-9868.2006.00542.x.

Wang, J., and W. A. Fuller. 2003. The mean squared error of small area predictors constructed with estimated area variances. *Journal of the American Statistical Association* 98 (463):716–23. doi:10.1198/0162145030000000620.

Ybarra, L. M., and S. L. Lohr. 2008. Small area estimation when auxiliary information is measured with error. *Biometrika* 95 (4):919–31. doi:10.1093/biomet/asn048.