Transformed Fay-Herriot Model with Measurement Error in Covariates

Sepideh Mosaferi[a] and Rebecca C. Steorts[b]

[a]Department of Statistics, Iowa State University
[b]Department of Statistical Science and Computer Science, Duke University

Abstract

Statistical agencies are often asked to produce small area estimates for skewed variables. When domain sample sizes are too small to support direct estimators, the effect 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 with measurement errors. Motivated by this issue, we first stabilize the skewness to achieve normality in the response variable, and we propose an area-level log-measurement error model on the response variable. Second, under our proposed modeling framework, we derive an adjusted empirical Bayes (EB) predictor of positive small area quantities subject to the covariates containing measurement error. Third, we propose a corresponding mean squared prediction error (MSPE) using a jackknife method and a bootstrap method, where we illustrate that the order of the bias is $O(1/m)$ with m as the total number of small areas. Fourth, we illustrate our methodology using both design-based and model-based simulation studies.

KEYWORDS: small area estimation, official statistics, Bayesian methods, jackknife, parametric bootstrap, simulation studies.

1 Introduction

Statistical agencies are often asked to produce small area estimates (SAEs) for skewed variables or datasets containing outliers. 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). There are many other statistical agencies that provide such information; for example, the United States National Agricultural Statistics Service (NASS) provides estimates regarding crop harvests (Bellow and Lahiri 2011), the United States Natural Resources Conservation Service (NRCS) provides estimates regarding roads at the county-level (Wang and Fuller 2003), and the Australian Agricultural and Grazing Industries Survey provides estimates of the total expenditures of Australian farms (Chandra and Chambers 2011). When domain sample sizes are too small to support direct estimators, the effects of skewness can be quite large, and it is critical to account for the distribution of the response variable given auxiliary (covariate) information at hand. One approach in small area estimation (SAE) is to utilize model-based estimators instead of design-based estimators. For a review of SAE, see Rao and Molina (2015) and Pfefferman (2013).

1.1 Census of the Governments (CoG)

As mentioned in Section 1, our proposed framework is motivated by data that is skewed or contains outliers. One such data set is the Census of Governments (CoG), which is a survey data collected by the USCB periodically to provide comprehensive statistics about governments and governmental activities. Data is reported on government organizations, finances, and employment. For example, data from organizations refers to location, type, and characteristics of local governments and officials. Data from finances/employment refers 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 or outlying response variables, we make four contributions to the literature. First, in order to stabilize the skewness and achieve normality in the response variable, we propose an area-level multiplicative log-measurement error model on the response variable (equations 2.3 and 2.4), contrasting the proposed additive measurement error model of Ybarra and Lohr (2008). In addition, we propose a multiplicative measurement error model on the covariates (equation 2.5). Second, under our proposed modeling framework, we derive Bayes predictors. Specifically, we derive an adjusted empirical Bayes (EB) predictor of positive small area quantities subject to the covariates containing measurement error. Third, 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. Fourth, we illustrate the performance of our methodology in both model-based simulation and design-based simulation studies. Finally, we discuss limitations of our proposed framework and intuition regarding such limitations to provide guidance for future research.

The article is organized as follows. Section 1.3 details the prior work related to our proposed methodology. Section 2 gives our proposed methodology, where we propose a multiplicative measurement error model for the response variable rather than an additive one. Then, in order to stabilize the skewness and achieve normality in the response variable, we consider a logarithmic transformation. In addition, we consider a measurement error model of the covariates with a log transformation. Section 2.1 derives the Bayes and adjusted empirical Bayes predictors under our framework. Section 2.3 provides the MSPE for our adjusted EB predictor. We provide a decomposition of the MSPE to include the uncertainty of the adjusted EB predictor. Section 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. Section 4 provides both design-based and model-based simulation studies. Section 5 provides a discussion and directions for future work.

1.3 Prior Work

As already mentioned, one of our contributions is proposing an area-level *multiplicative measurement error* model for the *response variable* rather than an additive one. Then, in order to stabilize the skewness and achieve normality in the *response variable*, we consider a logarithmic transformation of the *response variable*. In addition, we consider a *multiplicative measurement error* model of the *covariates*. There is a rich literature on the area-level Fay-Herriot model, where authors have considered an *additive measurement error* on the covariates. In fact, the first *additive measurement error* model for the *covariates* was proposed by Ybarra and Lohr (2008). The authors primarily considered covariate information from another survey, which is similar to the application of the CoG. More recently, Berg and Chandra (2014) 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. Therefore, our proposed EB predictor and MSE differ from their framework. Turning to Bayesian approaches, Arima et al. (2017), Arima et al. (2015a), and Arima et al. (2015b) have provided fully Bayesian solutions to the measurement error problem for both unit-level and area-level small area estimation problems.

Given that we make a logarithmic transformation of both the *response* and the *auxiliary* (covariates) variables, there is also a large literature on such transformations. In this paper, we use a Box-Cox transformation due to its simplicity, and discuss more complex transformations in our discussion. For example, Sugasawa and Kubokawa (2015) have considered a modified Box-Cox class of transformation, which is known as the dual-power transformation. They assume there are no measurement errors in the covariates. Other related literature assumes a log-transformation (Bell et al. 2007), Ghosh et al. (2015), and Bell et al. (2016). Other transformations of the response variable could be considered and this is discussed in Section 5.

Finally, we discuss the related literature regarding the proposed jackknife and parametric bootstrap estimator of the MSPE of the adjusted Bayes estimators, where the order of the bias is $O(m^{-1})$, under standard regularity conditions. Our proposed jacknife estimator of the MSPE contrasts that of Jiang et al. (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 the authors can make an orthogonal decomposition, they can easily 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), and propose a parametric bootstrap estimator of the MSPE of our adjusted estimator. In a similar manner to the jacknife, the order of the bias for the parametric bootstrap estimator of the MSPE is $O(m^{-1})$.

2 Area-Level Log-Multiplicative Model with Measurement Error

Consider m small areas and let Y_i $(i = 1, \ldots, m)$ 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 consider a multiplicative model for Y_i. Second, we consider a measurement error model suitable for inference. Given our goal is to stabilize the skewness and achieve normality in the response variable Y_i, we combine these two approaches. Specifically, we propose an area-level multiplicative log-measurement error model on Y_i. In the rest of this section, we give our proposed model. We derive Bayes and adjusted EB predictors under our proposed model. Finally, we derive a corresponding MSPE of the adjusted EB predictor.

Consider the following multiplicative model similar to that of Ghosh et al. (2015):

$$y_i = Y_i \xi_i, \tag{2.1}$$

for a positive response y_i. As the primary goal is to estimate Y_i, consider

$$Y_i = \lambda_i \prod_{k=1}^{p} X_{ik}^{\beta_k}, \tag{2.2}$$

where X_{ik} is the k-th covariate of the i-th small area, which is unknown, but is observed by x_{ik} (see equation 2.5). The regression coefficient β_k is unknown and must be estimated (see Section 2.2).

To achieve normality for the positively skewed response y_i, we apply the following log transformation:

$$z_i := \log y_i = \log(Y_i \xi_i) = \theta_i + e_i, \tag{2.3}$$

where the sampling errors $e_i := \log \xi_i$ and random effects $\nu_i := \log \lambda_i$ are mutually independent and distributed as $e_i \sim N(0, \psi_i)$ and $\nu_i \sim N(0, \sigma_\nu^2)$ for $i = 1, \ldots, m$. Also, we assume that the pairs (z_i, θ_i) are independent (for all $i = 1, ..., m$), leading to the following two-stage hierarchical model:

$$\begin{cases} z_i|\theta_i \sim N(\theta_i, \psi_i), \\ \theta_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma_\nu^2 \sim N(\sum\limits_{k=1}^{p} \beta_k \log X_{ik}, \sigma_\nu^2), \end{cases} \tag{2.4}$$

where $\boldsymbol{X}_i = (X_{i1}, ..., X_{ip})^\top$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$. The variance of the sampling errors ψ_i is assumed to be known (or can be estimated from the survey), and the variance of random effects σ_ν^2 and the regression coefficient $\boldsymbol{\beta}$ is unknown and must be estimated (see Section 2.2).

Remark 2.1. *Equation 2.3 is a Fay-Herriot model for z_i, however, the parameter of interest is $Y_i := \exp(\theta_i)$ rather than θ_i. Slud and Maiti (2006) and Ghosh et al. (2015) used a similar model in the absence of measurement errors in the covariates.*

Let $\boldsymbol{x_i} = (x_{i1}, ..., x_{ip})^\top$ and $\boldsymbol{\eta}_i = (\eta_{i1}, ..., \eta_{ip})^\top$. Let $\odot$ denote the Hadamard (coordinatewise) product, and let Σ_i denote a known positive definite matrix. Here, we propose a multiplicative measurement error model for the covariates $\boldsymbol{x}_i$. This crucially builds off our proposed log-transformed multiplicative model on Y_i while incorporating the following measurement error model:

$$\boldsymbol{x}_i = \boldsymbol{X}_i \odot \boldsymbol{\eta}_i, \qquad \boldsymbol{\eta}_i \sim \text{log-normal}_p(\boldsymbol{0}, \Sigma_i). \tag{2.5}$$

Therefore, $\boldsymbol{X}_i \odot \boldsymbol{\eta}_i$ represents the differences of $\boldsymbol{X}_i$'s from the positively observed values of $\boldsymbol{x}_i$'s, where $\boldsymbol{x}_i$ is independent of $\boldsymbol{\eta}_i$ in different small areas. We assume that all the sources of errors $(e_i, \nu_i, \boldsymbol{\eta}_i)$ are mutually independent. In the situation when Σ_i 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 Σ_i.

As working directly with equation 2.5 is difficult, we consider the following logarithmic transformation:

$$\boldsymbol{w}_i := \log \boldsymbol{x}_i = \log \boldsymbol{X}_i + \log \boldsymbol{\eta}_i = \boldsymbol{W}_i + \boldsymbol{\eta}_i^*, \tag{2.6}$$

where $\boldsymbol{W}_i := \log \boldsymbol{X}_i$, and $\boldsymbol{\eta}_i^* := \log \boldsymbol{\eta}_i$. All logarithmic transformations are applied coordinatewise. Equations 2.3 and 2.6 are now on the same logarithmic scale.

2.1 Bayes and Empirical Bayes Predictors

In this section, we provide the Bayes predictor for θ_i under the assumption that $\boldsymbol{\beta}$ and σ_ν^2 are known. Then we provide the corresponding adjusted Bayes predictor of θ_i. Finally, we provide the back transformed Bayes and adjusted Bayes predictors for $Y_i := \exp(\theta_i)$. Under the two-stage hierarchical model in equation 2.4 and the squared error loss function, the *Bayes predictor* of θ_i is

$$\hat{\theta}_i := E(\theta_i|z_i) = \gamma_i z_i + (1-\gamma_i)\sum_{k=1}^{p}\beta_k W_{ik}, \tag{2.7}$$

where $\gamma_i = \sigma_\nu^2/(\sigma_\nu^2+\psi_i)$. The posterior variance of θ_i is $var(\theta_i|z_i) := E\{(\hat{\theta}_i-\theta_i)^2|z_i\} = \gamma_i\psi_i$, where we assume σ_ν^2 is known and the variance does not depend on the area-specific response z_i. It then follows that the mean squared error of $\hat{\theta}_i$ is

$$\text{MSE}(\hat{\theta}_i) := E\{var(\theta_i|z_i)\} = \gamma_i\psi_i,$$

where the expectation is with respect to the marginal distribution of z_i.

In practice, W_{ik} is unobserved; we replace it with the observed w_{ik} in equation 2.7, which simplifies to

$$\tilde{\theta}_i = \gamma_i z_i + (1-\gamma_i)\sum_{k=1}^{p}\beta_k w_{ik}. \tag{2.8}$$

In addition, w_{ik} contains error. Thus, the shrinkage parameter γ_i must be adjusted; otherwise, the MSE will be larger compared to the direct estimator z_i. The *adjusted Bayes predictor* is

$$\tilde{\theta}_i^A = \gamma_i^* z_i + (1-\gamma_i^*)\sum_{k=1}^{p}\beta_k w_{ik},$$

where $\gamma_i^* = (\sigma_\nu^2 + \boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta})/(\sigma_\nu^2 + \boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta}+\psi_i)$. The MSE of $\tilde{\theta}_i^A$ is smaller than the MSE of z_i (see Ybarra and Lohr (2008) for a similar argument).

Recall that the parameter of interest is $Y_i := \exp(\theta_i)$ after transforming from the logarithmic scale back to the original scale. Our goal is to find the corresponding adjusted EB predictor $\hat{Y}_i^A$. The corresponding *Bayes predictor* is

$$\hat{Y}_i := E(Y_i|z_i) = \exp(\sum_{k=1}^{p}\beta_k W_{ik}) \times E(\exp(\nu_i|z_i)). \tag{2.9}$$

Based upon the normality of ν_i, we have

$$\nu_i^* := \{\nu_i|z_i\} \sim N(\gamma_i^*[z_i - \sum_{k=1}^{p}\beta_k w_{ik}], \psi_i\gamma_i^*). \tag{2.10}$$

Then by using the moment generating function of the normal distribution in equation 2.10, we obtain the *adjusted Bayes predictor* $\tilde{Y}_i^A$ of Y_i from equation 2.9 as

$$\tilde{Y}_i^A = \exp\left\{\gamma_i^* z_i + (1-\gamma_i^*)\sum_{k=1}^{p}\beta_k w_{ik} + \frac{\psi_i\gamma_i^*}{2}\right\}. \tag{2.11}$$

When $\boldsymbol{\beta}$ and σ_ν^2 are unknown, we substitute the consistent estimators $\hat{\beta}$ and $\hat{\sigma}_\nu^2$ into equation 2.11 to obtain the following *adjusted EB predictor* of Y_i:

$$\hat{Y}_i^A = \exp\left\{\hat{\gamma}_i^* z_i + (1-\hat{\gamma}_i^*)\sum_{k=1}^{p}\hat{\beta}_k w_{ik} + \frac{\psi_i\hat{\gamma}_i^*}{2}\right\}.$$

See Section 2.2 for guidance on estimating the unknown parameters $\boldsymbol{\beta}$ and σ_ν^2.

2.2 Estimation of β and σ_ν^2

In this section, we discuss estimation of the unknown parameters $\boldsymbol{\beta}$ and σ_ν^2. We estimate the unknown parameter $\boldsymbol{\beta}$ by the following modified least squares estimator:

$$\hat{\boldsymbol{\beta}}_w = \left\{\sum_{i=1}^{m}\frac{\boldsymbol{w}_i\boldsymbol{w}_i^\top - \Sigma_i}{\boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta} + \sigma_\nu^2 + \psi_i}\right\}^{-1}\left\{\sum_{i=1}^{m}\frac{\boldsymbol{w}_i z_i}{\boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta} + \sigma_\nu^2 + \psi_i}\right\},$$

which has the advantage of taking into account the measurement error by deducting the Σ_i term from the inverse term (see Cheng and Van Ness (1999), pp. 85 and 146). In addition, the estimator is consistent, i.e., $E(\hat{\boldsymbol{\beta}}_w) - \boldsymbol{\beta} = O(m^{-1})$. Section 4.2.2 discusses the consequences of mis-specifying the error variance Σ_i on the coefficient of $\boldsymbol{\beta}$. In fact, we illustrate as the magnitude of misspecification in the measurement error term Σ_i increases, the estimated value of $\boldsymbol{\beta}$ becomes more biased.

Next, we estimate the unknown parameter σ_ν^2 by maximizing the marginal distribution z_i conditional on $\boldsymbol{w}_i$ where $z_i|\boldsymbol{w}_i \sim N(\sum_{k=1}^{p}\beta_k w_{ik}, \boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta} + \sigma_\nu^2 + \psi_i)$. Let us assume the following: assume the likelihood is three-times continuously differentiable and the derivatives are uniformly bounded. In addition, assume $\sigma_\nu^2 \in \mathbb{R}^+$ and $E(\partial^2 \log f(\boldsymbol{z}|\boldsymbol{w})/\partial(\sigma_\nu^2)^2)$ is a finite and positive quantity. Then under these assumptions, the estimator of σ_ν^2 is consistent, i.e., $E(\hat{\sigma}_\nu^2) - \sigma_\nu^2 = O(m^{-1})$. Finally, to estimate these parameters in practice, we use the Fisher Scoring Algorithm, which we review in Appendix A for completeness.

2.3 Mean Squared Prediction Error of the Adjusted EB Predictor

In this section, we first define the MSPE of the adjusted EB predictor $\hat{Y}_i^A$. Second, we show that the cross-product term of the MSPE of the adjusted EB predictor $\hat{Y}_i^A$ tends to zero as $m \to \infty$ (Lemma 2.1). Now, we give some notation that will be used throughout the rest of the paper. Let

$$\begin{aligned} M_{1i} &:= E(\tilde{Y}_i^A - Y_i)^2 \\ &= \exp\left\{\psi_i\gamma_i^*\right\}\Big[\exp\left\{\psi_i\gamma_i^*\right\} - 1\Big]\exp\left\{2\Big[\gamma_i^* z_i + (1-\gamma_i^*)\sum_{k=1}^{p}\beta_k w_{ik}\Big]\right\}, \\ M_{2i} &:= E(\hat{Y}_i^A - \tilde{Y}_i^A)^2, \quad M_{3i} := E(\hat{Y}_i^A - \tilde{Y}_i^A)(\tilde{Y}_i^A - Y_i). \end{aligned}$$

Note the term M_{1i} depends on the area-specific response variable z_i unlike Jiang et al. (2002), and its estimator has bias of order $O(m^{-1})$. Since we wish to include the uncertainty of the adjusted EB predictor $\hat{Y}_i^A$ with respect to the unknown parameters $\boldsymbol{\beta}$ and σ_ν^2, we decompose the MSPE (into three terms) using Definition 2.1.

Definition 2.1. *The MSPE of the adjusted EB predictor $\hat{Y}_i^A$ is*

$$\begin{aligned} MSPE(\hat{Y}_i^A) &= E(\hat{Y}_i^A - Y_i)^2 \\ &\equiv E(\tilde{Y}_i^A - Y_i)^2 + E(\hat{Y}_i^A - \tilde{Y}_i^A)^2 + 2E(\hat{Y}_i^A - \tilde{Y}_i^A)(\tilde{Y}_i^A - Y_i) \\ &= M_{1i} + M_{2i} + 2M_{3i} \end{aligned}$$

Let n_i denote the sample size in the i-th area. Let us assume the following:

$$\min_{1\le i\le m} n_i \ge 1, \quad \max_{1\le i\le m} n_i = K < \infty, \quad n_\top - m \to \infty \tag{2.12}$$

where $n_\top := \sum_{i=1}^{m} n_i$.

Lemma 2.1. *Under the assumptions in equation 2.12, $M_{3i} \to 0$ as $m \to \infty$.*

Proof. Consider the following expression based on the Cauchy-Schwarz inequality:

$$\begin{aligned}
&E\Big(\frac{1}{m}\sum_{i=1}^{m}(\tilde{Y}_i^A - Y_i)(\hat{Y}_i^A - \tilde{Y}_i^A)\Big) \le E\Big(\frac{1}{m}\sum_{i=1}^{m}\Big|(\tilde{Y}_i^A - Y_i)(\hat{Y}_i^A - \tilde{Y}_i^A)\Big|\Big)\\
&\le \frac{1}{m}\sum_{i=1}^{m}\Big[E^{1/2}(\tilde{Y}_i^A - Y_i)^2 E^{1/2}(\hat{Y}_i^A - \tilde{Y}_i^A)^2\Big]\\
&\le \max_{1\le i\le m} E^{1/2}(\tilde{Y}_i^A - Y_i)^2\Big[\frac{1}{m}\sum_{i=1}^{m}E^{1/2}(\hat{Y}_i^A - \tilde{Y}_i^A)^2\Big] \to 0, \quad \text{as} \quad m\to\infty,
\end{aligned}$$

where the last line follows from the asymptotic optimality of $\hat{Y}_i^A$, i.e.,

$$\frac{1}{m}\sum_{i=1}^{m}E(\hat{Y}_i^A - \tilde{Y}_i^A)^2 \to 0, \quad \text{as} \quad m\to\infty.$$

See Ghosh and Sinha (2007) for the definition of asymptotic optimality. □

3 Jackknife and Parametric Bootstrap Estimators of the MSPE

In this section, we propose two estimators for the MSPE of the adjusted EB predictor $\hat{Y}_i^A$, where we ignore the cross-product term M_{3i} following Lemma 2.1. 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 adjusted EB predictor.

3.1 Jackknife Estimator of the MSPE

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

$$\text{mspe}_J(\hat{Y}_i^A) = \hat{M}_{1i,J} + \hat{M}_{2i,J}, \quad \text{where} \tag{3.1}$$

$$\hat{M}_{1i,J} = \hat{M}_{1i} - \frac{m-1}{m}\sum_{j=1}^{m}(\hat{M}_{1i} - \hat{M}_{1i(-j)}) \quad \text{and} \quad \hat{M}_{2i,J} = \frac{m-1}{m}\sum_{j=1}^{m}(\hat{Y}_i^A - \hat{Y}_{i(-j)}^A)^2,$$

such that $(-j)$ denote all areas except the j-th area. Also, let

$$\begin{aligned}\hat{M}_{1i} &:= M_{1i}(\hat{\sigma}_\nu^2, \hat{\boldsymbol{\beta}}) \\ &= \exp\Big\{\psi_i\hat{\gamma}_i^*\Big\}\Big[\exp\Big\{\psi_i\hat{\gamma}_i^*\Big\} - 1\Big]\exp\Big\{2\Big[\hat{\gamma}_i^* z_i + (1-\hat{\gamma}_i^*)\sum_{k=1}^{p}\hat{\beta}_k w_{ik}\Big]\Big\}. \end{aligned} \tag{3.2}$$

Now, we define some notation and establish six regularity conditions used in Theorem 3.1. Let $\ell(\cdot)$ denote the likelihood function. Define the first, second, and third derivatives of the likelihood function with respect to the marginal distribution z_i by $\ell_i^{'}(\phi; z_i)$, $\ell_i^{''}(\phi; z_i)$, and $\ell_i^{'''}(\phi; z_i)$. Establishing Theorem 3.1 depends on the following six regularity conditions: first, define $\phi^\top = (\boldsymbol{\beta}, \sigma_\nu^2) \in \Theta$ where Θ is a compact set and $\Theta \subseteq (\mathbb{R}^p, \mathbb{R}^+)$. Second, assume that $\hat{\phi}$ is a consistent estimator for ϕ. Third, assume that $\ell_i^{'}(\phi; z_i)$ and $\ell_i^{''}(\phi; z_i)$ both exist for $i = 1, \dots m$, almost surely (a.s.) in probability. Fourth, assume that $E\{\ell_i^{'}(\phi; z_i)|\phi\} = 0$ for $i = 1, \dots, m$. Fifth, assume that $\ell_i^{''}(\phi; z_i)$ is a continuous function of ϕ for $i = 1, ..., m$, a.s. 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. Sixth, assume that $E\{|\ell_i^{'}(\phi; z_i)|^{4+\delta}\}$, $E\{|\ell_i^{''}(\phi; z_i)|^{4+\delta}\}$, and $E\{\sup_{c\in(-\epsilon,\epsilon)}|\ell_i^{'''}(\phi + c; z_i)|^{4+\delta}\}$ are uniformly bounded for $i = 1, \dots m$ under some $\epsilon > 0$ and $\delta > 0$.

Theorem 3.1. *If the aforementioned six regularity conditions hold, then*

$$E[mspe_J(\hat{Y}_i^A)] = MSPE(\hat{Y}_i^A) + O(m^{-1}).$$

Proof. Define

$$\begin{aligned}E(\text{mspe}_J(\hat{Y}_i^A)) &\equiv E(\hat{M}_{1i,J} + \hat{M}_{2i,J}) \\ &= E\Big(\hat{M}_{1i} - \frac{m-1}{m}\sum_{j=1}^{m}(\hat{M}_{1i} - \hat{M}_{1i(-j)})\Big) + \frac{m-1}{m}E\Big(\sum_{j=1}^{m}(\hat{Y}_i^A - \hat{Y}_{i(-j)}^A)^2\Big).\end{aligned}$$

Also, define a remainder term r_i that is bounded in absolute value by R_i such that,

$$|r_i| \le \max\{1, |\ell^{'}(\phi; z_i)|^3, |\ell^{''}(\phi; z_i)|^3, |\ell^{'''}(\phi; z_i)|^3\} \equiv R_i.$$

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

$$\hat{M}_{1i} = M_{1i} + M_{1i}^{'\top}(\hat{\phi} - \phi) + \frac{1}{2}M_{1i}^{''\top}(\hat{\phi} - \phi)^2 + \frac{1}{6}M_{1i}^{'''\top}(\hat{\phi} - \phi)^3,$$

where $M_{1i}^{'\top} = \big(\frac{\partial M_{1i}}{\partial \boldsymbol{\beta}}, \frac{\partial M_{1i}}{\partial \sigma_\nu^2}\big)$, $M_{1i}^{''\top} = \big(\frac{\partial(\partial M_{1i})}{\partial^2 \boldsymbol{\beta}}, \frac{\partial(\partial M_{1i})}{\partial^2 \sigma_\nu^2}\big)$, and $M_{1i}^{'''\top} = \big(\frac{\partial(\partial^2 M_{1i})}{\partial^3 \boldsymbol{\beta}}, \frac{\partial(\partial^2 M_{1i})}{\partial^3 \sigma_\nu^2}\big)$. Let $\hat{\phi}^\top = (\hat{\boldsymbol{\beta}}, \hat{\sigma}_\nu^2)$. It follows that

$$\hat{M}_{1i} - \hat{M}_{1i(-j)} = \hat{M}_{1i}^{'\top}(\hat{\phi} - \hat{\phi}_{(-j)}) + \frac{1}{2}\hat{M}_{1i}^{''\top}(\hat{\phi} - \hat{\phi}_{(-j)})^2 + \frac{1}{6}\hat{M}_{1i(-j)}^{'''\top}(\hat{\phi} - \hat{\phi}_{(-j)})^3.$$

Now, observe that

$$\hat{\phi} - \phi = -\frac{\sum_{i=1}^{m}\ell_i^{'}(\phi; z_i)}{\sum_{i=1}^{m}\ell_i^{''}(\phi; z_i)}\left\{1 + \frac{\sum_{k=1}^{m}\ell_k^{'}(\phi; z_k)\sum_{r=1}^{m}\ell_r^{'''}(\phi; z_r)}{2(\sum_{k=1}^{m}\ell_k^{''}(\phi; z_k))^2}\right\} + O_p(|\hat{\phi} - \phi|^3), \text{ and}$$

$$\hat{\phi} - \hat{\phi}_{(-j)} = \frac{\ell_j^{'}(\hat{\phi}; z_j)}{\sum_{k\neq j}^{m}\ell_k^{''}(\hat{\phi}; z_k)}\left[1 - \frac{\ell_j^{'}(\hat{\phi}; z_j)\sum_{k\neq j}^{m}\ell_k^{'''}(\hat{\phi}; z_k)}{2(\sum_{k\neq j}^{m}\ell_k^{''}(\hat{\phi}; z_k))^2}\right] + O_p(|\hat{\phi} - \hat{\phi}_{(-j)}|^3).$$

By taking expectation, 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

$$\varphi = \frac{\sum_{j=1}^{m} E(\ell_j^{'} \ell_j^{''})}{\sum_{j=1}^{m} E\{\ell_j^{''}(\phi; z_j)\}} - \frac{\sum_{j=1}^{m} \sum_{k=1}^{m} E(\ell_j^{'})^2 E(\ell_k^{'''})}{2(\sum_{j=1}^{m} E\{\ell_j^{''}(\phi; z_j)\})^2},$$

and

$$\sum_{j \neq i}^{m} E(\hat{\phi} - \hat{\phi}_{(-j)} | z_i) = \frac{-\ell_i^{'}(\phi; z_i) + \varphi}{\sum_{j=1}^{m} E\{\ell_j^{''}(\phi; z_j)\}} + r_i o(m^{-1}).$$

By combining the above results, we find that

$$E\{\hat{M}_{1i,J} - M_{1i} | z_i\} = -M_{1i}^{'}(\phi; z_i) \ell_i^{'}(\phi; z_i) / \varphi + r_i o(m^{-1}).$$
$$\therefore E(\hat{M}_{1i,J}) = M_{1i} + O(m^{-1}).$$

Second, we prove $\hat{M}_{2i}$ has a bias of order $o(m^{-1})$. Let

$$\hat{Y}_i^A - \hat{Y}_{i(-j)}^A := h(\hat{\phi}, z_i) - h(\hat{\phi}_{(-j)}, z_i),$$

and $h(\phi, z_i) = E(Y_i^A | z_i, \phi)$. Using a Taylor series expansion, we find that

$$\hat{Y}_i^A - \hat{Y}_{i(-j)}^A = h^{'\top}(\hat{\phi}, z_i)(\hat{\phi} - \hat{\phi}_{(-j)}) + \frac{1}{2} h^{''\top}(\hat{\phi}_{(-j)}^*, z_i)(\hat{\phi} - \hat{\phi}_{(-j)})^2,$$

where

$$h^{'\top}(\hat{\phi}, z_i) = \Big(\frac{\partial h(\hat{\phi}, z_i)}{\partial \boldsymbol{\beta}}, \frac{\partial h(\hat{\phi}, z_i)}{\partial \sigma_\nu^2} \Big), \quad h^{''\top}(\hat{\phi}, z_i) = \Big(\frac{\partial(\partial h(\hat{\phi}, z_i))}{\partial^2 \boldsymbol{\beta}}, \frac{\partial(\partial h(\hat{\phi}, z_i))}{\partial^2 \sigma_\nu^2} \Big)$$

and $\hat{\phi}_{(-j)}^*$ is between $\hat{\phi}_{(-j)}$ and $\hat{\phi}$. Using an additional Taylor series expansion, we find that

$$\sum_{j=1}^{m} E\big\{ (\hat{Y}_{i(-j)}^A - \hat{Y}_i^A)^2 | z_i \big\} = \big\{ h^{'\top}(\phi, z_i) \big\}^2 \times \frac{\sum_{j=1}^{m} E\{(\ell_j^{'}(\phi, z_j))^2\}}{\varphi^2} + r_i o(m^{-1}).$$
$$\therefore E(\hat{M}_{2i,J}) = M_{2i} + o(m^{-1}).$$

Finally, the proof is complete as

$$\begin{aligned} E(\text{mspe}_J(\hat{Y}_i^A)) &= E(\hat{M}_{1i,J}) + E(\hat{M}_{2i,J}) \\ &= \{M_{1i} + O(m^{-1})\} + \{M_{2i} + o(m^{-1})\} \\ &= M_{1i} + M_{2i} + O(m^{-1}). \\ \therefore E[\text{mspe}_J(\hat{Y}_i^A)] &= \text{MSPE}(\hat{Y}_i^A) + O(m^{-1}). \end{aligned}$$

□

3.2 Parametric Bootstrap Estimator of the MSPE

In this section, we propose a parametric bootstrap estimator of the MSPE of the adjusted EB predictor $\hat{Y}_i^A$, which we denote it by $\text{mspe}_B(\hat{Y}_i^A)$. 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 adjusted EB predictor, where we ignore the cross-product term (see Lemma 2.1).

To introduce the parametric bootstrap method, consider the following bootstrap model:

$$\begin{aligned} z_i^\star|\boldsymbol{w}_i^\star, \nu_i^\star &\overset{ind}{\sim} N(\sum_{k=1}^{p} \hat{\beta}_k w_{ik}^\star + \nu_i^\star, \psi_i), \\ \boldsymbol{w}_i^\star &\overset{ind}{\sim} N_p(\boldsymbol{W}_i, \Sigma_i), \\ \nu_i^\star &\overset{ind}{\sim} N(0, \hat{\sigma}_\nu^2). \end{aligned} \tag{3.3}$$

Recall that from Definition 2.1, $\text{MSPE}(\hat{Y}_i^A) = M_{1i} + E(\hat{Y}_i^A - \tilde{Y}_i^A)$, where we ignore the cross product term M_{3i}. We use the parametric bootstrap twice. First, we use it to estimate M_{1i} in order to correct the bias of $\hat{M}_{1i} := M_{1i}(\hat{\sigma}_\nu^2, \hat{\boldsymbol{\beta}})$ (see equation 3.2). Second, we use it to estimate $E(\hat{Y}_i^A - \tilde{Y}_i^A)^2$. More specifically, we propose to estimate M_{1i} by $2M_{1i}(\hat{\sigma}_\nu^2, \hat{\boldsymbol{\beta}}) - E_\star[M_{1i}(\hat{\sigma}_\nu^{\star 2}, \hat{\boldsymbol{\beta}}^\star)]$, and $E[\hat{Y}_i^A - \tilde{Y}_i^A]^2$ by $E_\star[\hat{Y}_i^{A\star} - \hat{Y}_i^A]^2$, where $E_\star$ denotes that the expectation is computed with respect to model 3.3 and $\hat{Y}_i^{A\star} = \exp\{\hat{\gamma}_i^{\star\star} z_i + (1 - \hat{\gamma}_i^{\star\star}) \sum_{k=1}^{p} \hat{\beta}_k^\star w_{ik} + \psi_i \hat{\gamma}_i^{\star\star}/2\}$. In addition, $\hat{\gamma}_i^{\star\star} = (\hat{\sigma}_\nu^{\star 2} + \hat{\boldsymbol{\beta}}^{\star\top} \Sigma_i \hat{\boldsymbol{\beta}}^\star)/(\hat{\sigma}_\nu^{\star 2} + \hat{\boldsymbol{\beta}}^{\star\top} \Sigma_i \hat{\boldsymbol{\beta}}^\star + \psi_i)$, where $\hat{\boldsymbol{\beta}}^\star$ and $\hat{\sigma}_\nu^{\star 2}$ are estimators of $\boldsymbol{\beta}$ and σ_ν^2 with respect to the parametric bootstrap model given in equation 3.3 and are estimated according to the Fisher Scoring Algorithm in Appendix A.

Our proposed estimator of $\text{MSPE}(\hat{Y}_i^A)$ is given by

$$\text{mspe}_B(\hat{Y}_i^A) = 2M_{1i}(\hat{\sigma}_\nu^2, \hat{\boldsymbol{\beta}}) - E_\star[M_{1i}(\hat{\sigma}_\nu^{\star 2}, \hat{\boldsymbol{\beta}}^\star)] + E_\star[\hat{Y}_i^{A\star} - \hat{Y}_i^A]^2. \tag{3.4}$$

which has bias of order $O(m^{-1})$ as shown in the Theorem 3.2.

Theorem 3.2. *The bootstrap estimator of the MSPE has bias of order $O(m^{-1})$, such that*

$$E[mspe_B(\hat{Y}_i^A)] = MSPE(\hat{Y}_i^A) + O(m^{-1}).$$

Proof. If $E_\star(\hat{\sigma}_\nu^{\star 2} - \hat{\sigma}_\nu^2) = O_p(m^{-1})$ and $E_\star(\hat{\boldsymbol{\beta}}^\star - \hat{\boldsymbol{\beta}}) = O_p(m^{-1})$, then

$$E_\star[M_{1i}(\hat{\sigma}_\nu^{\star 2}, \hat{\boldsymbol{\beta}}^\star)] = M_{1i}(\hat{\sigma}_\nu^2, \hat{\boldsymbol{\beta}}) + O_p(m^{-1}).$$

Assume that $R_m^* = O_{p^\star}(m^{-1})$ such that $mR_m^\star$ is bounded in probability under the parametric bootstrap model given in equation 3.3. Consider the following Taylor series expansion:

$$\hat{Y}_i^{A\star} - \hat{Y}_i^A = (\hat{\phi}^\star - \hat{\phi})^\top h'(\hat{\phi}, z_i) + R_m^*,$$

such that $\hat{\phi}^{\star\top} = (\hat{\boldsymbol{\beta}}^\star, \hat{\sigma}_\nu^{\star 2})$. Using a similar argument to the proof of Theorem 3.1,

$$E_\star[\hat{Y}_i^{A\star} - \hat{Y}_i^A]^2 = \hat{M}_{2i} + o_p(m^{-1}) \quad \text{and} \quad E_\star[\hat{M}_{1i}^\star] = \hat{M}_{1i} + O_p(m^{-1}) \tag{3.5}$$

Substituting equations 3.5 into equation 3.4, we find that

$$\begin{aligned} \text{mspe}_B(\hat{Y}_i^A) &= 2\hat{M}_{1i} - [\hat{M}_{1i} + O_p(m^{-1})] + \hat{M}_{2i} + o_p(m^{-1}) \\ &= \hat{M}_{1i} + \hat{M}_{2i} + O_p(m^{-1}). \end{aligned}$$

This implies that

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

□

4 Experiments

In this section, we describe our design-based and model-based simulation studies and the results for our proposed framework. All of our code that corresponds with our simulation studies will be released as open source code on `github` with vignettes.

4.1 Design-Based Simulation Study

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

4.1.1 Design-Based Simulation Setup

In this section, 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 Σ_i was obtained from a Taylor series approximation, where $\mathrm{Var}(x_i)$ was estimated from the formula of variance in simple random sampling without replacement at each state. The ψ_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:

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

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.

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 estimators/predictors for each state as well as their corresponding empirical MSEs in Tables 1 and 2. We observe that our proposed adjusted EB predictor does not always outperform the direct estimator, which we further explore in our model-based simulation studies in Section 4.2.

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., EMSE($\hat{Y}_i^A$) 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 adjusted estimator does not uniformly beat the direct estimator in terms of the EMSE, we conduct a model-based simulation study in Section 4.2 to investigate this and provide further insight.

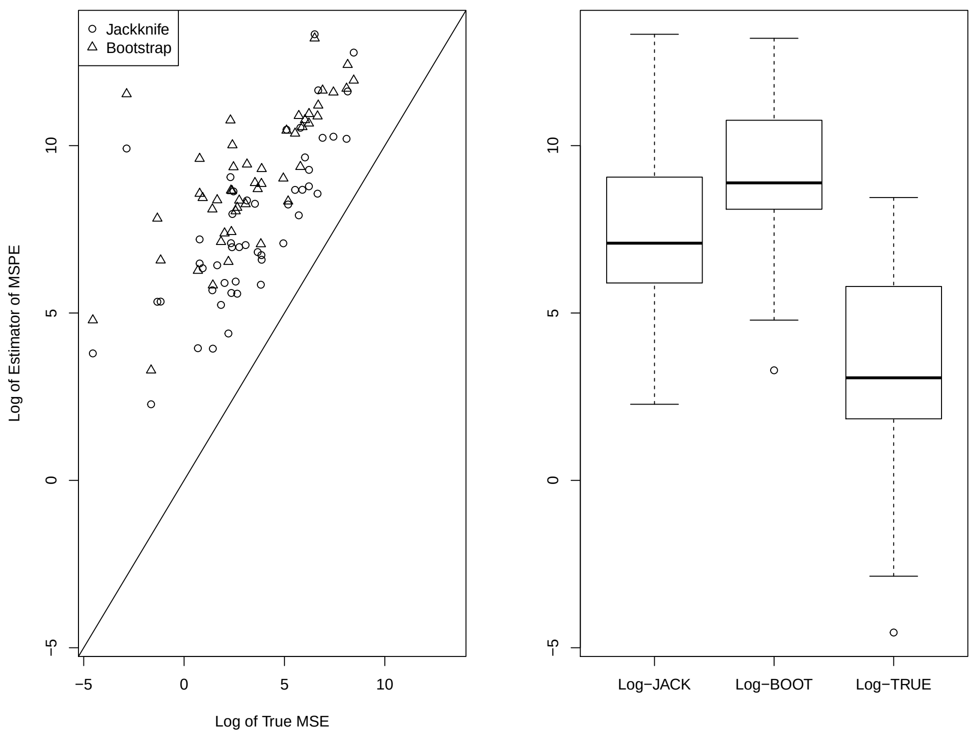

Figure 1: Left: The jackknife and the bootstrap estimators versus the true MSE (EMSE($\hat{Y}_i^A$)), where the results are rescaled logarithmically. Right: Box plots of the jackknife and the bootstrap estimators and the true MSE (EMSE($\hat{Y}_i^A$)), where the results are rescaled logarithmically. In general, the log of the jackknife is closer to the log of true MSE.

4.2 Model-Based Simulation Study

In this section, we describe our model-based simulation study to further investigate the performance of the proposed adjusted EB predictor $\hat{Y}_i^A$. Second, we compare the proposed jackknife and parametric bootstrap estimators, $\text{mspe}_J(\hat{Y}_i^A)$ and $\text{mspe}_B(\hat{Y}_i^A)$. Third, we investigate how often the variance estimates $\hat{\sigma}_u^2$ are zero. Finally, we investigate how the regression parameter changes when Σ_i is misspecified. Our goal through this model-based simulation study is to understand how one could improve the adjusted EB predictor through future research, and to further understand it's underlying behavior.

4.2.1 Model-Based Simulation Setup

In this section, we provide the setup of our model-based simulation study in Table 3.

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) $\tilde{Y}_i$: the EB predictor, assuming the true covariate w_i and ignoring Σ_i in our model,

4) $\hat{Y}_i^A$: the adjusted EB predictor, assuming the true covariate w_i has measurement error, where Σ_i is included in our model

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

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

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

In order to study the magnitude of each term in the MSPE of $\hat{Y}_i^A$ given in Definition 2.1, we use the empirical MSPE (EMSPE) of $\hat{Y}_i^A$ as follows:

$$\text{EMSPE}(\hat{Y}_i^A) \equiv \tilde{M}_i = \tilde{M}_{1i} + \tilde{M}_{2i} + 2\tilde{M}_{3i}, \quad \text{where}$$

$$\tilde{M}_{1i} = \tfrac{1}{R}\sum_{r=1}^{R}\{\tilde{Y}_i^{A(r)} - Y_i^{(r)}\}^2,$$

$$\tilde{M}_{2i} = \tfrac{1}{R}\sum_{r=1}^{R}\{\hat{Y}_i^{A(r)} - \tilde{Y}_i^{A(r)}\}^2, \quad \text{and}$$

$$\tilde{M}_{3i} = \tfrac{1}{R}\sum_{r=1}^{R}\{\hat{Y}_i^{A(r)} - \tilde{Y}_i^{A(r)}\}\{\tilde{Y}_i^{A(r)} - \tilde{Y}_i^{(r)}\}.$$

In order to evaluate the jackknife and parametric bootstrap estimators of $\hat{Y}_i^A$, we consider the relative bias, denoted by $\text{RB}_J(\hat{Y}_i^A)$ and $\text{RB}_B(\hat{Y}_i^A)$, respectively. More specifically, the relative biases are defined as follows for each area i:

$$\text{RB}_J(\hat{Y}_i^A) = \left\{\frac{1}{R}\sum_{r=1}^{R}\text{mspe}_J^{(r)}(\hat{Y}_i^{A(r)}) - \text{EMSE}(\hat{Y}_i^A)\right\}\Big/\text{EMSE}(\hat{Y}_i^A),$$

$$\text{RB}_B(\hat{Y}_i^A) = \left\{\frac{1}{R}\sum_{r=1}^{R}\text{mspe}_B^{(r)}(\hat{Y}_i^{A(r)}) - \text{EMSE}(\hat{Y}_i^A)\right\}\Big/\text{EMSE}(\hat{Y}_i^A).$$

4.2.2 Model-Based Simulation Results

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

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 Section 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 Σ_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. This is not the typical performance that one would expect.

In fact, the adjusted EB predictors cannot outperform the direct estimators due to propagated errors in the term $\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta}$, which is present in the term γ_i^* in the EB predictors. Table 4 illustrates that there are cases in which the EMSE($\hat{Y}_i^A$) is larger than the EMSE(y_i). In order to understand this further, let us consider the following expression:

$$
\begin{aligned}
E[\hat{Y}_i^A - Y_i]^2 &= E\Big[\exp\Big\{\hat{\gamma}_i^* z_i + (1-\hat{\gamma}_i^*)\sum_{k=1}^{p}\hat{\beta}_k w_{ik} + \frac{\psi_i \hat{\gamma}_i^*}{2}\Big\} - Y_i\Big]^2 \\
&= \exp\Big(2\sum_{k=1}^{p}\beta_k W_{ik}\Big)\Big\{\exp(\gamma_i^*\psi_i)(\exp(\gamma_i^*\psi_i)-1) \\
&\quad \times \exp\Big(2(1-\gamma_i^*)^2\boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta} + 2\gamma_i^{*2}(\sigma_\nu^2+\psi_i)\Big)\Big\}.
\end{aligned}
$$

Observe the $\lim_{\boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta}\to\infty}\exp\Big(2(1-\gamma_i^*)^2\boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta} + 2\gamma_i^{*2}(\sigma_\nu^2+\psi_i)\Big) = \infty$ under the assumption that $\boldsymbol{\beta} \neq \mathbf{0}$. Therefore, as the measurement error variance Σ_i increases, we have shown that the MSE of our proposed EB predictors can also increase. This is the behavior that we observe in Table 4. In order to prevent such behavior, a further adjustment should be made to the EB predictors, which we discuss in Section 5.

Decomposition of the MSPE We first explore the decomposition of the MSPE in order to understand how large M_{1i} might be in practice. Table 5 presents the average results for the decomposition of MSPE over all the small areas when $m = 20, 50$, 100, and 500. Often the amount of M_{1i} is larger than the other terms. Next, we explore if the cross-product term M_{3i} in the MSPE (definition 2.1) can be ignored. The term $|\tilde{M}_{3i}|$ cannot generally be ignored in comparison with the term $\tilde{M}_{1i} + \tilde{M}_{2i}$. We also note that the term $\tilde{M}_{3i}$ can be also be negative. Thus, when the cross-product term is not negligible, the proposed $\text{mspe}_J(\hat{Y}_i^A)$ or $\text{mspe}_B(\hat{Y}_i^A)$ may lead to over- or underestimation of the MSPE.

Jackknife versus Parametric Bootstrap Estimators We compare the jackknife MSPE estimator of the adjusted EB predictor $\hat{Y}_i^A$ to that of the bootstrap using the relative bias (see Table 6). In addition, we consider box plots for the jackknife and bootstrap MSPE estimators of the adjusted EB predictor $\hat{Y}_i^A$, where we compare these to box plots of the true values (see Figure 3). Both Table 6 and Figure 3 illustrate that the bootstrap receives a large number of negative values, which is due to the construction of $\hat{M}_{1i}$. Here, we find that the bootstrap grossly underestimates the true values, whereas the jackknife slightly overestimates the true values. Due to this behavior, we would recommend the jackknife in practice.

Amount of Zeros for the Estimates of $\hat{\sigma}_u^2$ Here, we investigate the *proportion* of zero estimates for σ_ν^2 based on the Fisher Scoring Algorithm given in Appendix A. Figure 2 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, $\hat{Y}_i^A$ and $\hat{Y}_i$ tend to have a proportion of zero estimates of

σ_ν^2 between 0.3 and 0.5. When $m = 50$ and as k increases, $\hat{Y}_i^A$ and $\hat{Y}_i$ tend to have a proportion of zero estimates of σ_ν^2 between 0.15 and 0.4. When $m = 100$ and as k increases, $\hat{Y}_i^A$ and $\hat{Y}_i$ tend to have a proportion of zero estimates of σ_ν^2 between 0.05 and 0.3. When $m = 500$ and as k increases, $\hat{Y}_i^A$ and $\hat{Y}_i$ tend to have a proportion of zero estimates of σ_ν^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.

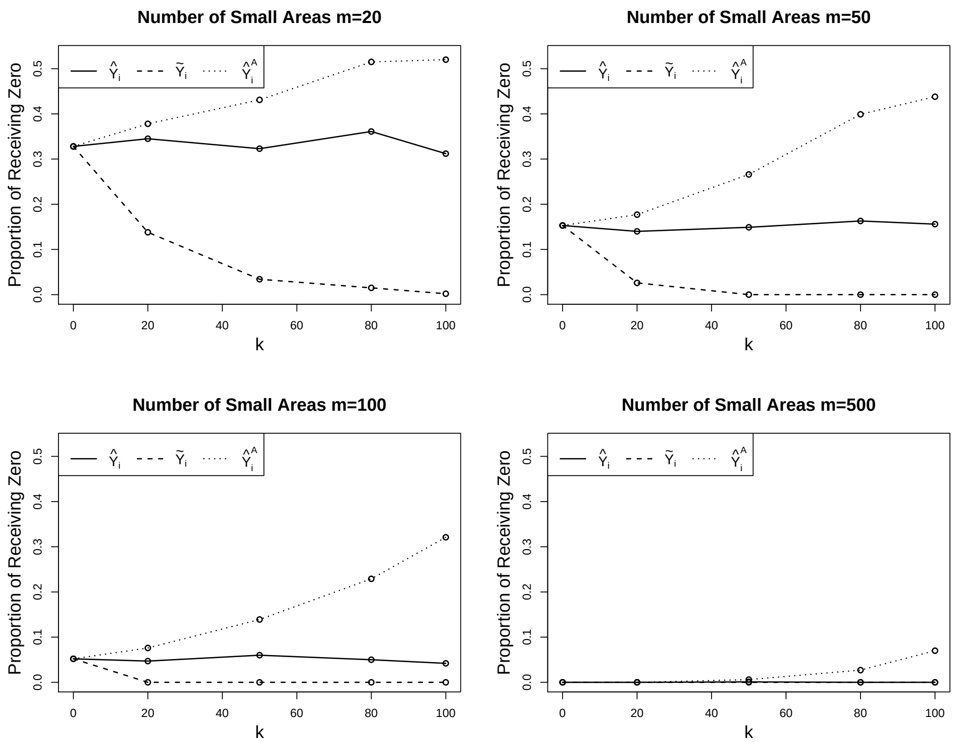

Figure 2: The proportion of zero estimates of σ_ν^2 from model-based simulation when we perform 1,000 replications of the simulation study for $k = 0, \ldots 100$, $m = 20, 50, 100, 500$, and $d = 2$.

The Effect of Misspecification of Σ_i on β We investigate the effect of mis-specifying the variance Σ_i on the estimation of the regression parameter β. To accomplish this, we conduct an empirical study based on the proposed model-based simulation study in Table 3 for the adjusted EB predictor $\hat{Y}_i^A$. Assume $\beta = 3$, and we consider two sets of experiments for each value of k, which are summarized in Table 7. 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 β under the followings: (1) the true value of d denoted by $\hat{\beta}_w$ and (2) the misspecified value of d_{mis} denoted by $\hat{\beta}_{w,\text{mis}}$. Then we compute the average absolute difference between the respective β's by considering the following:

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

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

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

Table 7 illustrates that the overall misspecification of Σ_i leads to bias in β. When the magnitude of measurement error is zero (i.e. $k = 0$), there is no difference between the estimated β using d or d_{mis}. On the other hand, when the magnitude of k increases and we have more uncertainty in the error variance Σ_i, values of $\hat{\beta}_w$ and $\hat{\beta}_{w,\text{mis}}$ 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 $\hat{\beta}_{w,\text{mis}}$ in which its bias is corrected through some techniques such as bootstrap and develop a test of parameter specification.

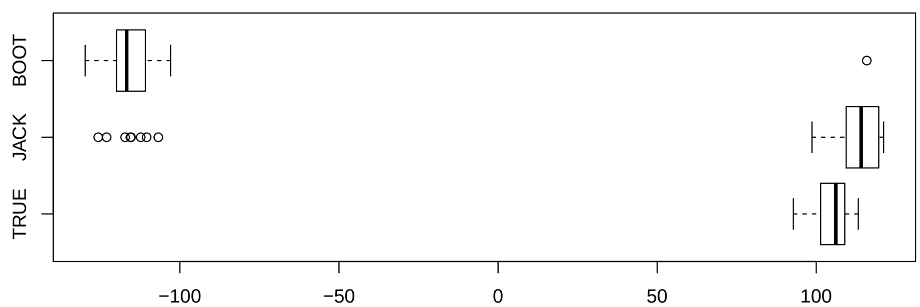

Figure 3: Comparing the distribution of the jackknife and bootstrap estimators with respect to the true MSE (EMSE($\hat{Y}_i^A$)) from the model-based simulations. The results are logarithmically rescaled.

Table 1: Estimators and their empirical MSEs from CoG. The MSEs are rescaled logarithmically.

i	State	n_i	y_i	$\tilde{Y}_i$	$\hat{Y}_i^A$	EMSE(y_i)	EMSE($\tilde{Y}_i$)	EMSE($\hat{Y}_i^A$)
1	RI	10	191.641	202.928	204.907	6.523	5.390	5.103
2	AK	14	132.301	120.112	123.684	4.501	6.153	5.793
3	NV	15	420.299	422.992	431.912	8.077	8.170	8.449
4	MD	19	824.939	784.779	794.784	8.050	5.524	6.503
5	DE	27	64.273	72.674	73.130	3.003	5.113	5.182
6	LA	40	363.838	342.056	343.344	6.017	0.845	-2.863
7	VA	40	560.881	526.612	530.160	6.669	8.265	8.148
8	NH	44	80.695	83.645	83.857	-3.994	2.254	2.387
9	UT	47	126.045	117.729	118.512	2.827	2.873	2.461
10	AZ	49	332.610	349.704	350.915	6.270	7.380	7.439
11	CT	49	187.500	191.494	192.054	4.851	5.457	5.528
12	SC	53	231.790	221.881	222.574	5.158	6.279	6.218
13	WV	53	83.568	84.071	84.315	2.754	2.482	2.336
14	WY	55	43.084	39.629	39.795	2.493	3.873	3.824
15	VT	59	27.717	27.529	27.574	0.474	0.751	0.688
16	ME	65	38.892	40.713	40.789	2.966	1.899	1.840
17	NM	67	93.817	93.360	93.913	3.496	3.331	3.529
18	MA	70	237.066	231.213	231.999	4.218	1.741	2.310
19	TN	72	245.317	232.133	233.405	4.257	6.144	6.023
20	NC	76	372.264	357.791	359.375	5.846	6.997	6.700
21	MS	78	137.977	132.143	132.454	3.891	0.299	0.774
22	ID	92	46.450	45.642	45.784	2.451	1.910	2.016
23	AL	93	162.841	160.389	160.818	3.356	2.131	2.406
24	MT	99	23.296	22.457	22.508	0.461	1.481	1.433
25	KY	104	119.415	121.105	121.576	1.935	2.928	3.134
26	NJ	108	235.215	245.163	245.595	3.898	5.663	5.713
27	GA	109	255.141	243.862	244.539	5.686	6.696	6.648

Table 2: Estimators and their empirical MSEs from CoG. The MSEs are rescaled logarithmically (continued).

i	State	n_i	y_i	$\tilde{Y}_i$	$\hat{Y}_i^A$	EMSE(y_i)	EMSE($\tilde{Y}_i$)	EMSE($\hat{Y}_i^A$)
28	AR	118	68.911	65.223	65.362	-3.160	2.719	2.646
29	OR	120	72.363	72.483	72.653	1.376	1.493	1.648
30	FL	124	377.822	365.035	366.658	7.658	8.148	8.092
31	OK	144	76.686	74.610	74.803	-2.447	1.155	0.926
32	WA	145	93.681	91.238	91.476	3.077	3.920	3.852
33	SD	153	14.757	14.078	14.142	-1.338	-3.583	-4.546
34	IA	155	56.219	55.686	55.790	-4.924	-0.962	-1.332
35	CO	184	74.373	71.579	71.908	0.789	2.907	2.746
36	NE	197	33.713	31.017	31.215	1.326	-0.561	-1.167
37	IN	217	75.994	78.712	78.830	0.619	0.609	0.775
38	ND	217	8.336	7.421	7.469	-1.704	-1.436	-1.644
39	MI	236	85.314	97.710	97.705	-1.220	4.945	4.944
40	WI	246	61.304	61.320	61.451	2.486	2.495	2.569
41	NY	270	280.982	278.784	283.973	6.457	6.275	6.681
42	MO	278	61.972	62.846	62.946	0.093	1.307	1.409
43	MN	289	42.025	45.747	45.727	-1.572	2.367	2.355
44	OH	302	108.261	111.756	111.905	2.364	3.820	3.864
45	KS	309	30.804	30.253	30.321	1.859	2.253	2.208
46	CA	338	238.284	248.558	248.800	6.991	6.244	6.223
47	TX	374	221.105	204.983	205.689	2.570	5.965	5.892
48	PA	386	81.653	83.551	83.668	2.888	3.628	3.666
49	IL	567	64.650	67.278	67.172	1.490	3.110	3.064

Table 3: Model-based simulation setup with definition of parameters and distributions

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

Table 4: Estimators and their empirical MSEs from model-based simulations. The results are averaged over all the small areas and re-scaled logarithmically.

m	k	y_i	$\hat{Y}_i$	$\tilde{Y}_i$	$\hat{Y}_i^A$	EMSE(y_i)	EMSE($\hat{Y}_i$)	EMSE($\tilde{Y}_i$)	EMSE($\hat{Y}_i^A$)
20	0	46.766	50.626	50.626	50.626	102.088	111.09	111.09	111.09
	20	53.054	50.139	49.229	49.864	115.974	109.338	104.398	108.425
	50	42.232	41.174	42.464	43.110	93.496	91.163	92.948	94.223
	80	42.519	44.285	43.469	44.794	93.881	97.557	95.152	97.495
	100	44.682	41.81	45.073	46.331	99.13	92.161	99.938	102.48
50	0	49.624	47.732	47.732	47.732	110.061	106.167	106.167	106.167
	20	44.292	42.851	44.702	45.098	97.644	94.541	99.321	100.255
	50	45.512	44.677	46.071	47.59	99.615	98.56	101.351	105.547
	80	42.703	41.773	45.289	46.469	94.339	93.268	101.068	103.615
	100	43.83	43.201	44.779	45.68	97.144	95.961	98.347	100.319
100	0	42.635	42.241	42.241	42.241	94.625	92.802	92.802	92.802
	20	46.216	45.601	46.179	47.427	103.264	101.412	103.035	106.172
	50	50.93	45.982	49.347	48.519	113.343	103.08	109.718	108.214
	80	50.132	46.586	48.137	48.966	111.618	103.055	106.712	108.454
	100	44.925	44.711	48.009	49.031	100.996	100.764	107.472	109.515
500	0	47.338	45.253	45.253	45.253	107.275	103.465	103.465	103.465
	20	46.382	45.369	47.575	47.652	104.607	102.867	107.896	108.169
	50	53.045	46.854	50.662	49.126	119.208	106.396	114.378	110.006
	80	47.766	44.868	47.706	49.950	108.289	104.921	107.454	112.805
	100	48.586	45.313	49.372	50.449	109.795	103.378	111.069	113.218

Table 5: The average results for the decomposition of empirical MSPE over all the small areas from model-based simulations. The results for $\tilde{M}_1$, $\tilde{M}_2$, $\tilde{M}_3$, and $\tilde{M}$ are rescaled by the logarithm of absolute value and "*" means the original value is negative.

m	k	$\tilde{M}_1$	$\tilde{M}_2$	$\tilde{M}_3$	$\tilde{M}$	$\lvert\tilde{M}_3\rvert/(\tilde{M}_1+\tilde{M}_2)$
20	0	104.005	111.031	107.518*	110.971	0.030
	20	106.392	109.043	107.718	109.514	0.248
	50	93.072	93.684	91.236*	93.998	0.056
	80	97.897	97.584	97.265	98.924	0.307
	100	97.117	102.346	99.675*	102.203	0.069
50	0	101.713	105.946	103.791*	105.702	0.114
	20	101.099	99.266	100.091	101.736	0.315
	50	105.423	102.610	102.104*	105.410	0.034
	80	102.507	101.907	102.206*	99.821	0.478
	100	100.129	99.569	98.418	100.788	0.115
100	0	93.484	92.546	92.670	94.307	0.318
	20	106.091	99.996	102.888*	106.009	0.041
	50	107.648	105.416	106.529*	106.857	0.295
	80	109.659	108.277	108.916	110.449	0.380
	100	107.403	108.663	108.030*	107.154	0.414
500	0	103.680	100.461	101.532	103.921	0.112
	20	108.237	101.446	104.833	108.302	0.033
	50	110.242	107.738	108.316	110.559	0.135
	80	112.656	107.559	110.101*	112.494	0.077
	100	113.091	107.899	110.358*	112.958	0.065

Table 6: Comparison of the proposed jackknife and bootstrap estimators from model-based simulations. The results are averaged over all small ares. The results are rescaled by the logarithm of absolute value. Note, * denote that the original value is negative.

m	k	EMSE($\hat{Y}_i^A$)	mspe$_J$($\hat{Y}_i^A$)	mspe$_B$($\hat{Y}_i^A$)	RB$_J$($\hat{Y}_i^A$)	RB$_B$($\hat{Y}_i^A$)
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*	15.565	32.277*
	100	102.48	115.536*	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*

5 Discussion

We have made four contributions to the literature, which we summarize. First, in order to stabilize the skewness and achieve normality in the response variable, we have proposed an area-level multiplicative log-measurement error model on the response variable. In addition, we have proposed a multiplicative measurement error model on the covariates. Second, under our proposed modeling framework, we derived the adjusted 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 $O(m^{-1})$ under standard regularity conditions. Fourth, we have illustrated the performance of our methodology in both design-based simulation and model-based simulation studies, where we have proposed an adjusted EB predictor, where the corresponding EMSE 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 adjusted EB predictor to avoid for such behavior. One way to address this issue is by estimating $\hat{\phi} = (\boldsymbol{\beta}, \sigma_\nu^2)^\top$ 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 $\hat{Y}_i^A$ and is under our further research.

As mentioned, we have made contributions in estimating the MSPE of the adjusted EB predictor, using both the jackknife and bootstrap. We have illustrated through simulation studies that the cross-product term may not be negligible, which implies that the proposed $\text{mspe}_J(\hat{Y}_i^A)$ or $\text{mspe}_B(\hat{Y}_i^A)$ may lead to over- or underestimation of the MSPE. To address this problem, one could perform a bias correction of our proposed estimator, and potentially show that the order of the bias is $o(m^{-1})$. This also has led to further thoughts regarding our proposed work. For example, it would be of interest to propose an estimator that would correct the bias in M_{1i} term. One could then explore improvements associated with the proposed jackknife and bootstrap estimators of the MSPE.

Finally, while a logarithmic transformation of the response variable y is a standard choice when it is positively skewed, this might not be always appropriate to stabilize the variation and achieve the normality in the response variable y. One alternative transformation is that of Sugasawa and Kubokawa (2015) which is a modified version of the Box-Cox transformation. The Box-Cox transformation itself might not be necessarily compatible with the normality assumption but by $\log y_i$. Sugasawa and Kubokawa (2015) have introduced the dual-power transformation, which is a monotone function of y. It would be of interest to see how this dual-power transformation would integrate with our proposed framework.

Table 7: Percentage of bias related to the consequences of misspecifying the error variance Σ_i on β in the EB predictor $\hat{Y}_i^A$ from model-based simulations. For all cases, we assume the true value for β is 3. Also, $\sigma_\nu^2 = 2$ and $m = 20$ (the smallest one) and 500 (the largest one).

m	k	Experiment	$\frac{1}{R}\sum_{r=1}^{R}\left\lvert\hat{\beta}_w^{(r)} - \hat{\beta}_{w,\text{mis}}^{(r)}\right\rvert$	$\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}_w^{(r)} - 3\right)$	$\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}_{w,\text{mis}}^{(r)} - 3\right)$
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

References

Arima, S., Bell, W., Datta, C., Gauri S. Franco 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**. DOI: https://doi.org/10.1111/rssa.12321.

Arima, S., Datta, G. S. and Liseo, B. (2015a). *Accounting for Measurement Error in Covariates in SAE: An Overview in Analysis of Poverty Data by Small Area Methods*. M. Pratesi Editor, Wiley New York.

Arima, S., Datta, G. S. and Liseo, B. (2015b). “Bayesian Estimators for Small Area Models When Auxiliary Information Is Measured with Error”. *Scandinavian Journal of Statistics*, **42: 518–529**. DOI: http://dx.doi.org/10.1111/sjos.12120.

Bell, W., Basel, W., Cruse, C., Dalzell, L., Maples, J., O'Hara, B. and Powers, D. (2007). *Use of ACS Data to Produce SAIPE Model-Based Estimates of Poverty for Counties*. U.S. Census Bureau Report. Available at: https://www.census.gov/did/www/saipe/publications/files/report.pdf.

Bell, W., Basel, W. and Maples, J. (2016). “An Overview of the U.S. Census Bureau's Small Area Income and Poverty Estimates Program”. *Analysis of Poverty Data by Small Area Estimation (ed. M. Pratesi), ch. 19. Chichester: Wiley.*

Bellow, M. E. and Lahiri, P. S. (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 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**. DOI: https://doi.org/10.1016/S0378-3758(02)00323-3.

Chandra, H. and Chambers, R. (2011). Small area estimation for skewed data in presence of zeros. *Calcutta Statistical Association Bulletin*, **63** 241–258.

Cheng, C. and Van Ness, J. (1999). *Statistical Regression with Measurement Error*. London: Arnold and New York: Oxford University Press.

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**. DOI: http://dx.doi.org/10.1080/01621459.1979.10482505.

Ghosh, M., Kubokawa, T. and Kawakubo, Y. (2015). “Benchmarked Empirical Bayes Methods in Multiplicative Area-Level Models with Risk Evaluation”. *Biometrika*, **102: 647–659**. DOI: https://doi.org/10.1093/biomet/asv010.

Ghosh, M. and Sinha, K. (2007). “Empirical Bayes Estimation in Finite Population Sampling under Functional Measurement Error Models”. *Journal of Statistical Planning and Inference*, **137: 2759–2773**. DOI: http://dx.doi.org/10.1016/j.jspi.2006.08.008.

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**. DOI: http://dx.doi.org/10.1214/aos/1043351257.

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

RAO, J. N. and MOLINA, I. (2015). *Small Area Estimation.* Wiley Series in Survey Sampling, London.

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**. DOI: http://dx.doi.org/10.1111/j.1467-9868.2006.00542.x.

SUGASAWA, S. and KUBOKAWA, T. (2015). "Parametric Transformed Fay–Herriot Model for Small Area Estimation". *Journal of Multivariate Analysis*, **139: 295–311**. DOI: https://doi.org/10.1016/j.jmva.2015.04.001.

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

YBARRA, L. M. and LOHR, S. L. (2008). "Small Area Estimation When Auxiliary Information Is Measured with Error". *Biometrika*, **95: 919–931**. DOI: http://dx.doi.org/10.1093/biomet/asn048.

A Fisher Scoring Algorithm

We briefly give the Fisher scoring algorithm that is used to estimate both $\boldsymbol{\beta}$ and σ_ν^2 as described in Section 2.2. We first define

$$\bigtriangledown\mathcal{L}_{\sigma_\nu^2} := \frac{\partial\mathcal{L}_{\sigma_\nu^2}}{\partial\sigma_\nu^2} = -\frac{1}{2}\sum_{i=1}^m \frac{1}{\boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta}+\sigma_\nu^2+\psi_i} + \frac{1}{2}\sum_{i=1}^m \frac{(z_i - \sum_{k=1}^p \beta_k w_{ik})^2}{(\boldsymbol{\beta}^\top\Sigma_i\boldsymbol{\beta}+\sigma_\nu^2+\psi_i)^2}.$$

Then for $\ell = 0, 1, 2, ...$, the iterative algorithm can be written as

$$\begin{cases} \sigma_\nu^{2(\ell+1)} = \max\left[0, \sigma_\nu^{2(\ell)} + \bigtriangledown\mathcal{L}_{\sigma_\nu^{2(\ell)}}\left\{\frac{1}{2}\sum_{i=1}^m(\boldsymbol{\beta}^{\top(\ell)}\Sigma_i\boldsymbol{\beta}^{(\ell)}+\sigma_\nu^{2(\ell)}+\psi_i)^{-2}\right\}^{-1}\right], \\ \boldsymbol{\beta}_w^{(\ell+1)} = \left\{\sum_{i=1}^m \frac{\boldsymbol{w}_i\boldsymbol{w}_i^\top - \Sigma_i}{\boldsymbol{\beta}^{\top(\ell)}\Sigma_i\boldsymbol{\beta}^{(\ell)}+\sigma_\nu^{2(\ell+1)}+\psi_i}\right\}^{-1}\left\{\sum_{i=1}^m \frac{\boldsymbol{w}_i z_i}{\boldsymbol{\beta}^{\top(\ell)}\Sigma_i\boldsymbol{\beta}^{(\ell)}+\sigma_\nu^{2(\ell+1)}+\psi_i}\right\}. \end{cases}$$

The starting values $\boldsymbol{\beta}^{(0)}$ and $\sigma_\nu^{2(0)}$ can be defined by an ordinary least squares fit. Finally, an adjusted EB predictor $\hat{Y}_i^A$ can be obtained by replacing $\boldsymbol{\beta}$ and σ_ν^2 with their consistent estimators $\hat{\boldsymbol{\beta}}_w$ and $\hat{\sigma}_\nu^2$ into $\tilde{Y}_i^A$.