Small area estimation with multiple covariates measured with errors: A nested error linear regression approach of combining multiple surveys

Gauri S. Datta[a,b], Mahmoud Torabi[c], J.N.K. Rao[d], Benmei Liu[e]

[a] *Department of Statistics, University of Georgia, GA 30602, USA*
[b] *US Census Bureasu, DC 20233, USA*
[c] *Departments of Community Health Sciences and Statistics, University of Manitoba, MB, R3E 0W3, Canada*
[d] *School of Mathematics and Statistics, Carleton University, ON, K1S 5B6, Canada*
[e] *National Cancer Institute, DC 20006, USA*

Abstract

Small area estimation has become a very active area of research in statistics. Many models studied in small area estimation focus on one or more variables of interest from a single survey without paying close attention to the nature of the covariates. It is useful to utilize the idea of borrowing strength from covariates to build a model which combines two (or multiple) surveys. In many real applications, there are also covariates measured with errors. In this paper, we study a nested error linear regression model which has multiple unit- or area-level error-free covariates, possibly come from administrative records, and multiple area-level covariates subject to structural measurement error where the data on the latter covariates are obtained from multiple surveys. In particular, we derive empirical best predictors of small area means and estimators of mean squared error of the predictors of small area means. Performance of the proposed approach is studied through a simulation study and also by a real application.

Keywords: Conditional distribution; Jackknife; Linear mixed model; Mean squared prediction error; Measurement error

1. Introduction

Sample surveys are generally conducted to provide reliable estimates of finite population parameters such as totals, means, counts, quantiles, etc. for the nation, census regions or states. In recent years, there has been increasing demand to get such estimates for smaller sub-populations (small areas), such as counties or age-sex-race demographic groups, due to their growing use in formulating policies and programs, allocating government funds, regional planning, marketing decisions at local level, and other uses. However, sample sizes within small areas are often too small to warrant the use of traditional area-specific direct estimates.

Different methods have been proposed in the literature to produce reliable estimates of characteristics of interest for small areas and to obtain measures of error associated with such estimates. These include, among others, the use of synthetic, composite and/or model-based estimators (Jiang and Lahiri [11]; Datta [5]; Pfeffermann [13]; Rao and Molina [15]; Jiang [10] ch. 4). Model-based estimators which borrow strength from related areas have been extensively used in small area estimation (Rao and Molina [15]). In particular, such small area models may be classified into two broad types: (i) Area-level models that relate design-based small area direct estimates to area-specific covariates; such models are used if unit-level data are not available. (ii) Unit-level models that relate the unit values of a study variable to associated unit-level covariates with known covariates area means (obtained, possibly, from administrative records) and area-specific covariates. A comprehensive account of model-based small area estimation under area-level and unit-level models is given by Rao and Molina [15]. In this paper, we focus on empirical best (EB) predictors of small area means under a unit-level nested error linear regression model with measurement errors in some area-level covariate values.

Email address: Mahmoud.Torabi@umanitoba.ca (Mahmoud Torabi)

Battese et al. [3] and Prasad and Rao [14] used a unit-level nested error linear regression model where the covariates are measured without errors. However, there are many circumstances where the covariates are subject to measurement error. In a pioneering paper, Ghosh et al. [8] proposed a nested error linear regression model with a single area-level covariate, which is an area-level covariate subject to structural measurement error (SME). The best predictors (BPs) of small area means obtained by these authors do not account for measurement error variation in the observed covariates, and the EB predictors are then constructed by replacing the unknown model parameters in their derived BPs by estimators obtained using the method of moments. Ghosh and Sinha [9] studied a similar setup, again with an area-level covariate, which is subject to functional measurement error (FME). However, the observed values of the covariate are not properly weighted in the prediction of small area means obtained in these two papers. Torabi et al. [18] and Datta et al. [6] derived more efficient predictors of small area means by appropriate weighting of the observed values of the covariate into the prediction, which was done by conditioning on the observed values of the (random) covariate and the observed data on the response. Arima et al. [1] studied hierarchical Bayes (HB) estimation under the model with the area-level covariate subject to SME, by choosing suitable priors on the model parameters. Torabi [16] and Torabi [17] studied EB estimation using survey weights under the nested error model with the area-level covariate subject to functional and structural measurement errors, respectively. Ybarra and Lohr [19] and Arima et al. [2] studied area-level models with covariates subject to functional measurement error. In the aforementioned papers dealing with unit-level models, both the response and the observed covariate values are assumed to come from the same survey.

In this paper, our goal is to predict the population mean of outcome of interest for each small area with data collected from more than one cross-sectional survey. From the main survey we collect data of the form $\{y_{ij}, \boldsymbol{w}_{ij} : i = 1, \ldots, m; j = 1, \ldots, n_i\}$ on the response variable y_{ij} and error-free covariates $\boldsymbol{w}_{ij} = (w_{ij1}, \ldots, w_{ijp})^\top$, where m is the number of small areas, n_i is the sample size in the ith area, and p is the number of error-free covariates. Note that if one or more error-free covariates are area-level covariates, we can absorb those within the vector $\boldsymbol{w}_{ij}$. Estimation method for the model parameter for $\boldsymbol{w}_{ij}$ that we propose in Section 3 is equally applicable for area-level covariates. From the other external surveys, we collect data on other surrogate covariates $X_{i\ell}, \ell = 1, \ldots, q$, in the form $\{X_{i\ell k} : i = 1, \ldots, m; \ell = 1, \ldots, q; k = 1, \ldots, t_{i\ell}\}$, where from the ℓth survey a sample of size $t_{i\ell}$ from the $T_{i\ell}$ population units in small area i is observed. The corresponding population mean of the lth covariate in area i is denoted by $x_{i\ell}$. Units selected in the external surveys are usually different, and the units on which we make measurement of $\boldsymbol{y}$ and $\boldsymbol{w}$ may not uniquely link to units in an external survey on which we are making measurements on covariates $X_{i\ell}, \ell = 1, \ldots, q$. It is anticipated that the area mean $\gamma_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}, (i = 1, \ldots, m)$, where N_i is the population size of the ith small area, is related to covariates through true area-level covariates means $x_{i\ell}, \ldots, x_{iq}$. However, the q population means, $x_{i\ell}, (\ell = 1, \ldots, q)$, for the m small areas are typically unknown. We need to estimate them from the unit-level samples $\{X_{i\ell k} : i = 1, \ldots, m; \ell = 1, \ldots, q; k = 1, \ldots, t_{i\ell}\}$.

We use the area-level covariate $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iq})^\top$ to specify a unit-level population model for the response values y_{ij} as

$$y_{ij} = \beta_0 + \boldsymbol{\beta}_1^\top \boldsymbol{w}_{ij} + \boldsymbol{\beta}_2^\top \boldsymbol{x}_i + v_i + e_{ij} \qquad (i = 1, \ldots, m; j = 1, \ldots, N_i). \tag{1}$$

Further, we specify a measurement error model on the observed covariate values $X_{i\ell k}$ as

$$X_{i\ell k} = x_{i\ell} + \eta_{i\ell k} \qquad (i = 1, \ldots, m; \ell = 1, \ldots, q; k = 1, \ldots, t_{i\ell}), \tag{2}$$

where $\boldsymbol{x}_i \overset{i.i.d.}{\sim} \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ for the SME and $\boldsymbol{x}_i$ is fixed but unknown for the FME, assuming that the model (2) is appropriate for simple random sampling within each of the q independent survey. In this paper, our focus is on the SME, and the FME will be studied in a separate manuscript. Further, the random errors e_{ij}, measurement errors $\eta_{i\ell k}$, and the area-level random effects v_i are assumed to be mutually independent with $e_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_e^2)$, $\eta_{i\ell k} \overset{ind.}{\sim} \mathcal{N}(0, \sigma_{\eta\ell}^2)$, $(\ell = 1, \ldots, q)$, and $v_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_v^2)$. We expect that the sample sizes, $t_{i\ell}$, from the other surveys are equal or larger than n_i ($t_{i\ell} \geq n_i$). In general, we use the notation $\boldsymbol{\phi}$ which collectively includes the unknown model parameters $\beta_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x, \sigma_v^2, \sigma_e^2$, and $\boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}$, where $\boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}} = (\sigma_{\eta 1}^2, \ldots, \sigma_{\eta q}^2)$.

Based on the available data $d = \{(y_{ij}, \boldsymbol{w}_{ij}) : i = 1, \ldots, m; j = 1, \ldots, n_i; X_{i\ell k} : i = 1, \ldots, m; \ell = 1, \ldots, q; k = 1, \ldots, t_{i\ell}\}$ and the model specified by (1) and (2), we obtain the EB predictor (EBP) of $\gamma_i, (i = 1, \ldots, m)$, assuming that the population model (1) also holds for the sample, that is, no sample selection bias. Under this set-up, the EBP does not depend on survey weight, unlike in the case of informative sampling. We first obtain the BP of γ_i in Section 2. We then obtain the EBP of γ_i by replacing $\boldsymbol{\phi}$ by some suitable estimator $\hat{\boldsymbol{\phi}}$ in the BP (Section 3). In Section 4, we employ the jackknife method to obtain a nearly unbiased estimator of the mean squared prediction error (MSPE) of EBP of γ_i. Section 5 reports the results of a simulation study on the performance of our EBP and associated jackknife MSPE estimator. The proposed method is applied in Section 6 to predict body mass index (BMI) measured in the U.S. National Health and Nutrition Examination Survey (NHANES), with auxiliary information $\boldsymbol{w}_{ij}$ from the NHANES and $X_{i\ell k_\ell}$ from the U.S. National Health Interview Survey (NHIS). Finally, some concluding remarks are given in Section 7. Technical details are deferred to the Appendix.

2. Best predictor

Note that γ_i is a linear function of $\boldsymbol{y}_i = (\boldsymbol{y}_i^{(1)\top}, \boldsymbol{y}_i^{(2)\top})^\top$, where $\boldsymbol{y}_i^{(1)} = (y_{i1}, \ldots, y_{in_i})^\top$ and $\boldsymbol{y}_i^{(2)} = (y_{i(n_i+1)}, \ldots, y_{iN_i})^\top$. We also note the independence of $(\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{w}_i)$, $i = 1, \ldots, m$, where $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iq})^\top$, $\boldsymbol{X}_{i\ell} = (X_{i\ell 1}, \ldots, X_{i\ell t_{i\ell}}, \ldots, X_{i\ell T_{i\ell}})^\top$, $\ell = 1, \ldots, q$, and $\boldsymbol{w}_i = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{iN_i})^\top$. For given $\boldsymbol{\phi}$, we can get the BP (under squared error loss) of the area mean γ_i given the sample data $\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}$ as $\hat{\gamma}_i^B = \mathrm{E}\{\gamma_i | \boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\}$. Let $\boldsymbol{X}_i^{(1)} = (\boldsymbol{X}_{i1}^{(1)}, \ldots, \boldsymbol{X}_{iq}^{(1)})^\top$ where $\boldsymbol{X}_{i\ell}^{(1)} = (X_{i\ell 1}, \ldots, X_{i\ell t_{i\ell}})^\top, (\ell = 1, \ldots, q)$ and $\boldsymbol{w}_i^{(1)} = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{in_i})^\top$. Theorem 1 gives the conditional mean $\mathrm{E}\{\gamma_i | \boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\}$ and the conditional variance $V\{\gamma_i | \boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\}$ of γ_i given $\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}$, and $\boldsymbol{\phi}$.

Theorem 2.1. Under the nested error model given by (1) and (2), the conditional distribution of γ_i given $\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}$ and $\boldsymbol{\phi}$ is normal with mean and variance given by

$$\begin{aligned} \mathrm{E}\{\gamma_i | \boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\} \quad = \quad & (1 - h_i B_{is})\{\bar{y}_i + \boldsymbol{\beta}_1^\top (\bar{\boldsymbol{w}}_{i(P)} - \bar{\boldsymbol{w}}_i)\} \\ & + h_i B_{is}\{\beta_0 + \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_{i(P)} + \boldsymbol{\beta}_2^\top \boldsymbol{\mu}_x\} \\ & + h_i B_{is} \boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_x (\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_{i\eta})^{-1} (\bar{\boldsymbol{X}}_i - \boldsymbol{\mu}_x) \end{aligned} \tag{3}$$

and

$$V\{\gamma_i | \boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\} = h_i \Big\{ h_i B_{is} (\sigma_v^2 + \boldsymbol{\beta}_2^\top \boldsymbol{M}_i^{-1} \boldsymbol{\beta}_2) + \sigma_e^2 / N_i \Big\}, \tag{4}$$

respectively, where $h_i = 1 - n_i/N_i$, $\bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$, $\bar{\boldsymbol{w}}_{i(P)} = N_i^{-1} \sum_{j=1}^{N_i} \boldsymbol{w}_{ij}$, $\bar{\boldsymbol{w}}_i = n_i^{-1} \sum_{j=1}^{n_i} \boldsymbol{w}_{ij}$, $\bar{\boldsymbol{X}}_i = (\bar{X}_{i1}, \ldots, \bar{X}_{iq})^\top$, $\bar{X}_{i\ell} = t_{i\ell}^{-1} \sum_{k=1}^{t_{i\ell}} X_{i\ell k}, (\ell = 1, \ldots, q)$, $\boldsymbol{\Sigma}_{i\eta} = \mathrm{diag}(\sigma_{\eta 1}^2 / t_{i1}, \ldots, \sigma_{\eta q}^2 / t_{iq})$, $\boldsymbol{M}_i = \boldsymbol{\Sigma}_{i\eta}^{-1} + \boldsymbol{\Sigma}_x^{-1}$, and

$$B_{is} = \frac{\sigma_e^2}{\sigma_e^2 + n_i(\sigma_v^2 + \boldsymbol{\beta}_2^\top \boldsymbol{M}_i^{-1} \boldsymbol{\beta}_2)}. \tag{5}$$

Here, the population means $\bar{\boldsymbol{w}}_{i(P)}$'s are assumed to be known. Proof of Theorem 1 is given in the Appendix.

It follows from (4) that $V\{\gamma_i | \boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\}$ does not depend on $\boldsymbol{y}_i^{(1)}$, $\boldsymbol{X}_i^{(1)}$ and $\boldsymbol{w}_i^{(1)}$. Hence, the mean squared prediction error (MSPE) of $\hat{\gamma}_i^B$, $\mathrm{E}(\hat{\gamma}_i^B - \gamma_i)^2$, is equal to the conditional variance of γ_i. Also, note that the right hand side of (4) depends only on $\boldsymbol{\delta}$ which includes $\boldsymbol{\beta}_2, \boldsymbol{\Sigma}_x, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^2, \sigma_v^2, \sigma_e^2$. We denote

$$\mathrm{MSPE}(\hat{\gamma}_i^B) = \mathrm{E}(\hat{\gamma}_i^B - \gamma_i)^2 = g_{1i}(\boldsymbol{\delta}),$$

where $g_{1i}(\boldsymbol{\delta})$ is given by the right hand side of (4). If N_i is large and $n_i/N_i \approx 0$, then

$$g_{1i}(\boldsymbol{\delta}) \approx B_{is}(\sigma_v^2 + \boldsymbol{\beta}_2^T \boldsymbol{M}_i^{-1} \boldsymbol{\beta}_2).$$

3. EB predictor

In practice, the model parameters $\boldsymbol{\phi}$ are unknown and need to be estimated from the data. Let $n = \sum_{i=1}^m n_i$ and $t_\ell = \sum_{i=1}^m t_{i\ell}, (\ell = 1, \ldots, q)$. We estimate $\sigma^2_{\eta\ell}, (\ell = 1, \ldots, q)$, unbiasedly (assuming $t_\ell > m$) by

$$\hat{\sigma}^2_{\eta\ell} = (t_\ell - m)^{-1} \sum_{i=1}^m \sum_{k=1}^{t_{i\ell}} (X_{i\ell k} - \bar{X}_{i\ell})^2.$$

A consistent estimator of $\boldsymbol{\mu}_x$ is given by $\hat{\boldsymbol{\mu}}_x = m^{-1} \sum_{i=1}^m \bar{\boldsymbol{X}}_i$.

To estimate $\boldsymbol{\Sigma}_x$, we note that

$$\mathrm{E}(\bar{\boldsymbol{X}}_i \bar{\boldsymbol{X}}_i^\top) = \boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top + \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_{i\eta},$$

and

$$\mathrm{E}(\sum_{i=1}^m \bar{\boldsymbol{X}}_i \bar{\boldsymbol{X}}_i^\top) = m\boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top + m\boldsymbol{\Sigma}_x + \sum_{i=1}^m \boldsymbol{\Sigma}_{i\eta}.$$

On the other hand,

$$\mathrm{E}(\hat{\boldsymbol{\mu}}_x \hat{\boldsymbol{\mu}}_x^\top) = V(\hat{\boldsymbol{\mu}}_x) + \boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top = \frac{1}{m}\Big(\boldsymbol{\Sigma}_x + \sum_{i=1}^m \boldsymbol{\Sigma}_{i\eta}/m\Big) + \boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top.$$

We can then write, based on the two equations above,

$$\mathrm{E}\Big(\sum_{i=1}^m \bar{\boldsymbol{X}}_i \bar{\boldsymbol{X}}_i^\top - m\hat{\boldsymbol{\mu}}_x \hat{\boldsymbol{\mu}}_x^\top\Big) = (m-1)\boldsymbol{\Sigma}_x + (1 - 1/m) \sum_{i=1}^m \boldsymbol{\Sigma}_{i\eta}.$$

A consistent estimator of $\boldsymbol{\Sigma}_x$ is

$$\hat{\boldsymbol{\Sigma}}_x = \frac{1}{m-1} \sum_{i=1}^m (\bar{\boldsymbol{X}}_i - \hat{\boldsymbol{\mu}}_x)(\bar{\boldsymbol{X}}_i - \hat{\boldsymbol{\mu}}_x)^\top - \frac{1}{m} \sum_{i=1}^m \hat{\boldsymbol{\Sigma}}_{i\eta},$$

where $\hat{\boldsymbol{\Sigma}}_{i\eta} = \mathrm{diag}(\hat{\sigma}^2_{\eta_1}/t_{i1}, \ldots, \hat{\sigma}^2_{\eta_q}/t_{iq})$.

Next, one can estimate $\boldsymbol{\beta}_1$ using

$$\hat{\boldsymbol{\beta}}_1 = \Big\{ \sum_{i=1}^m \sum_{j=1}^{n_i} (\boldsymbol{w}_{ij} - \bar{\boldsymbol{w}}) \boldsymbol{w}_{ij}^\top \Big\}^{-1} \Big\{ \sum_{i=1}^m \sum_{j=1}^{n_i} (\boldsymbol{w}_{ij} - \bar{\boldsymbol{w}}) y_{ij} \Big\},$$

where $\bar{\boldsymbol{w}} = n^{-1} \sum_{i=1}^m n_i \bar{\boldsymbol{w}}_i$. Note that our estimator $\hat{\boldsymbol{\beta}}_1$ remains valid even if some error-free covariates are area-level covariates. To obtain an estimator of $\boldsymbol{\beta}_2$, let

$$\mathrm{E}(S_l) = \boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_x^{(\ell)}, \quad (\ell = 1, \ldots, q),$$

where $\boldsymbol{\Sigma}_x^{(\ell)}$ is the ℓth column of $\boldsymbol{\Sigma}_x$, and

$$S_\ell = \frac{1}{d_\ell - r_\ell} \sum_{i=1}^m t_{i\ell} (\bar{y}_i - \hat{\boldsymbol{\beta}}_1^\top \bar{\boldsymbol{w}}_i)(\bar{X}_{i\ell} - \bar{X}_\ell),$$

where $d_\ell = t_\ell - \sum_{i=1}^m t_{i\ell}^2/t_\ell, r_\ell = \sum_{i=1}^m n_i t_{i\ell}(1 - t_{i\ell}/t_\ell) \bar{\boldsymbol{w}}_i^\top (\mathrm{SST}_W)^{-1} (\bar{\boldsymbol{w}}_i - \bar{\boldsymbol{w}})$, with $\mathrm{SST}_W = \sum_{i=1}^m \sum_{j=1}^{n_i} (\boldsymbol{w}_{ij} - \bar{\boldsymbol{w}}) \boldsymbol{w}_{ij}^\top$ and $\bar{X}_\ell = t_\ell^{-1} \sum_{i=1}^m t_{i\ell} \bar{X}_{i\ell}$. Hence,

$$\mathrm{E}(\mathbf{S}^\top) = \boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_x,$$

and consequently

$$\hat{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{\Sigma}}_x^{-1} \mathbf{S},$$

where $\mathbf{S} = (S_1, \ldots, S_q)^\top$. A consistent estimator of β_0 is then given by

$$\hat{\beta}_0 = \bar{y} - \hat{\boldsymbol{\beta}}_1^\top \bar{\boldsymbol{w}} - \hat{\boldsymbol{\beta}}_2^\top \hat{\boldsymbol{\mu}}_x,$$

where $\bar{y} = n^{-1} \sum_{i=1}^m n_i \bar{y}_i$.

The remaining parameters σ_v^2 and σ_e^2 are consistently estimated by

$$\hat{\sigma}_v^2 = \max\Big[0, \frac{m-1}{n - \sum_{i=1}^m n_i^2/n}\Big\{(\mathrm{MSB}_y - \mathrm{MSW}_y) - \hat{\boldsymbol{\beta}}_1^\top (\mathrm{MSB}_W - \mathrm{MSW}_W)\hat{\boldsymbol{\beta}}_1\Big\} - \hat{\boldsymbol{\beta}}_2^\top \hat{\boldsymbol{\Sigma}}_x \hat{\boldsymbol{\beta}}_2\Big],$$

and

$$\hat{\sigma}_e^2 = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} \Big\{(y_{ij} - \bar{y}_i) - \hat{\boldsymbol{\beta}}_1^\top (\boldsymbol{w}_{ij} - \bar{\boldsymbol{w}}_i)\Big\}^2}{n - m - p},$$

where

$$\mathrm{MSB}_y = (m-1)^{-1} \sum_{i=1}^m n_i (\bar{y}_i - \bar{y})^2,$$

$$\mathrm{MSB}_W = (m-1)^{-1} \sum_{i=1}^m n_i (\bar{\boldsymbol{w}}_i - \bar{\boldsymbol{w}})(\bar{\boldsymbol{w}}_i - \bar{\boldsymbol{w}})^\top,$$

$$\mathrm{MSW}_y = (n-m)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2,$$

$$\mathrm{MSW}_W = (n-m)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (\boldsymbol{w}_{ij} - \bar{\boldsymbol{w}}_i)(\boldsymbol{w}_{ij} - \bar{\boldsymbol{w}}_i)^\top.$$

A consistent estimator, $\hat{B}_{is}$, of B_{is} is obtained from (5) by replacing $\boldsymbol{\delta}$ by $\hat{\boldsymbol{\delta}}$. The EB predictor of γ_i is then given by

$$\begin{aligned} \hat{\gamma}_i^{EB} &= (1 - h_i \hat{B}_{is})\Big\{\bar{y}_i + \hat{\boldsymbol{\beta}}_1^\top (\bar{\boldsymbol{w}}_{i(P)} - \bar{\boldsymbol{w}}_i)\Big\} + h_i \hat{B}_{is}\Big\{\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1^\top \bar{\boldsymbol{w}}_{i(P)} + \hat{\boldsymbol{\beta}}_2^\top \hat{\boldsymbol{\mu}}_x\Big\} \\ &\quad + h_i \hat{B}_{is} \hat{\boldsymbol{\beta}}_2^\top \hat{\boldsymbol{\Sigma}}_x (\hat{\boldsymbol{\Sigma}}_x + \hat{\boldsymbol{\Sigma}}_{i\eta})^{-1} (\bar{\boldsymbol{X}}_i - \hat{\boldsymbol{\mu}}_x). \end{aligned} \tag{6}$$

4. Jackknife estimation of MSPE of EBP of small area means

In this section, we obtain a nearly unbiased estimator of the MSPE of the EB predictor $\hat{\gamma}_i^{EB}$. We estimate $\mathrm{MSPE}(\hat{\gamma}_i^{EB}) = \mathrm{E}(\hat{\gamma}_i^{EB} - \gamma_i)^2$ using the jackknife methods proposed by Jiang et al. [12] and Chen and Lahiri [4]. We have

$$\begin{aligned} \mathrm{MSPE}(\hat{\gamma}_i^{EB}) &= \mathrm{E}(\hat{\gamma}_i^B - \gamma_i)^2 + \mathrm{E}(\hat{\gamma}_i^{EB} - \hat{\gamma}_i^B)^2 + 2\mathrm{E}(\hat{\gamma}_i^B - \gamma_i)(\hat{\gamma}_i^{EB} - \hat{\gamma}_i^B) \\ &=: \mathrm{M}_{1i} + \mathrm{M}_{2i} + 2\mathrm{M}_{3i}, \end{aligned} \tag{7}$$

where $\mathrm{M}_{1i} = g_{1i}(\boldsymbol{\delta})$. In the case of SME, we have $\mathrm{M}_{3i} = 0$, noting that the conditional expectation of γ_i is equal to $\hat{\gamma}_i^B$.

A plug-in estimator of $g_{1i}(\boldsymbol{\delta})$ is $g_{1i}(\hat{\boldsymbol{\delta}})$. We apply the jackknife method of bias reduction to $g_{1i}(\hat{\boldsymbol{\delta}})$ to get a nearly unbiased estimator of $\mathrm{M}_{1i} = g_{1i}(\boldsymbol{\delta})$. Let $\hat{\boldsymbol{\phi}}_{-t}$ be the estimator of $\boldsymbol{\phi}$ obtained by deleting the tth area data set $(\boldsymbol{y}_t^{(1)}, \boldsymbol{X}_t^{(1)}, \boldsymbol{w}_t^{(1)})$ from the full data set $\{(\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}) : i = 1, \ldots, m\}$ and then applying the method-of-moments. This calculation is done for each area t in turn to get m estimators of $\boldsymbol{\phi}$ as $\{\hat{\boldsymbol{\phi}}_{-t} : t = 1, \ldots, m\}$. A weighted jackknife estimator of M_{1i} is given by

$$\hat{\mathrm{M}}_{1iJv} = g_{1i}(\hat{\boldsymbol{\delta}}) - \sum_{t=1}^m v_t \Big\{g_{1i}(\hat{\boldsymbol{\delta}}_{-t}) - g_{1i}(\hat{\boldsymbol{\delta}})\Big\}, \tag{8}$$

where $\nu_t = 1 + O(m^{-1})$ is a suitable weight (Chen and Lahiri [4]). In particular, we use

$$\nu_t = 1 - \boldsymbol{a}_{1t}^\top \Big(\sum_{r=1}^{m} \boldsymbol{a}_{1r}\boldsymbol{a}_{1r}^\top \Big)^{-1} \boldsymbol{a}_{1t}$$

where $\boldsymbol{a}_{1t} = (1, \bar{\boldsymbol{w}}_t^\top, \bar{\boldsymbol{X}}_t^\top)^\top$. An unweighted jackknife estimator of M_{1i}, denoted by $\hat{\mathrm{M}}_{1iJ}$, is obtained by letting $\nu_t = (m-1)/m$, $(t = 1, \ldots, m)$, in (8); (Jiang et al. [12]).

Turning to jackknife estimation of the second term, M_{2i}, in (7), let

$$\hat{\gamma}_i^B = k_i(\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi})$$

be the BP of γ_i expressed as a function of $\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}$, and $\boldsymbol{\phi}$. Then the EB predictor of γ_i may be expressed as

$$\hat{\gamma}_i^{EB} = k_i(\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \hat{\boldsymbol{\phi}}).$$

Now replace $\hat{\boldsymbol{\phi}}$ by $\hat{\boldsymbol{\phi}}_{-t}$ to get

$$\hat{\gamma}_{i,-t}^{EB} = k_i(\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \hat{\boldsymbol{\phi}}_{-t}), \qquad (t = 1, \ldots, m).$$

A weighted jackknife estimator of M_{2i} is then given by

$$\hat{\mathrm{M}}_{2iJ\nu} = \sum_{t=1}^{m} \nu_t (\hat{\gamma}_{i,-t}^{EB} - \hat{\gamma}_i^{EB})^2. \tag{9}$$

An unweighted jackknife estimator of M_{2i}, denoted by $\hat{\mathrm{M}}_{2iJ}$, is obtained by letting $\nu_t = (m-1)/m$, $(t = 1, \ldots, m)$, in (9). By taking the sum of (8) and (9), a weighted jackknife estimator of $\mathrm{MSPE}(\hat{\gamma}_i^{EB})$ is obtained as

$$\mathrm{mspe}_{J\nu}(\hat{\gamma}_i^{EB}) = \hat{\mathrm{M}}_{1iJ\nu} + \hat{\mathrm{M}}_{2iJ\nu}. \tag{10}$$

An unweighted jackknife estimator of $\mathrm{MSPE}(\hat{\gamma}_i^{EB})$ is given by

$$\mathrm{mspe}_{J}(\hat{\gamma}_i^{EB}) = \hat{\mathrm{M}}_{1iJ} + \hat{\mathrm{M}}_{2iJ}. \tag{11}$$

The jackknife estimator (10) is nearly unbiased in the sense that its bias is $o(m^{-1})$ for large m (Jiang et al. [12]).

5. Simulation study

We conduct a simulation study to evaluate the efficiency of the proposed EB predictor $\hat{\gamma}_i^{EB}$. To this end, we assume that the responses y_{ij} for the population units are generated from the model given by (1) and (2) with $\beta_0 = 100, p = q = 2, \boldsymbol{\beta}_1 = (0.1, 0.1)^\top, \boldsymbol{\beta}_2 = (2, 2)^T, \boldsymbol{\mu}_x = (194, 194)^\top, \boldsymbol{\Sigma}_x = \mathrm{diag}(2737, 2737), \sigma_v^2 = 16, \sigma_e^2 = 100, \boldsymbol{\Sigma}_\eta = \mathrm{diag}(25, 25)$, noting that this set up is similar to Ghosh et al. [8] in the case of single area-level covariate from one survey. The population consists of $N = 1400$ units spread across $m = 12$ areas of sizes (N_i) : $50, 250, 50, 100, 200, 150, 50, 150, 100, 150, 100$, and 50. Sample sizes (n_i) within areas for the response variable are taken as $1, 5, 1, 2, 4, 3, 1, 3, 2, 3, 2$ and 1. For the covariates, we use two scenarios ($t_{i\ell} = n_i$ and $t_{i\ell} = 3n_i$) to evaluate the impact of larger sample sizes, $t_{i\ell}$, for the covariates. We first calculate the MSPE of the best estimator $\hat{\gamma}_i^B$, given by (3), assuming the model parameters are known. Table 1 reports $\mathrm{MSPE}(\hat{\gamma}_i^B)$ for the scenarios $t_{i\ell} = n_i$ and $t_{i\ell} = 3n_i$. As observed in Table 1, the $\mathrm{MSPE}(\hat{\gamma}_i^B)$ for the larger sample sizes of covariates ($t_{i\ell} = 3n_i$) are consistently smaller than the corresponding values for the equal sample sizes ($t_{i\ell} = n_i$) with the range of MSPE reduction from 19 % to 34 % across areas.

For the simulation study on MSPE $(\hat{\gamma}_i^{EB})$, we first generate each element of $\boldsymbol{w}_{ij}$ for the population from normal distribution with mean one and variance one and treat them as fixed in the simulation study. We then generate $R = 5,000$ independent sets of normal variates $\{v_i^{(r)} : i = 1, \ldots, m\}, \{e_{ij}^{(r)} : i = 1, \ldots, m; j = 1, \ldots, N_i\}$ with means zero and specified variances σ_v^2 and σ_e^2. We also generate $\{\boldsymbol{x}_i^{(r)} : i = 1, \ldots, m\}$ with mean $\boldsymbol{\mu}_x$ and variance $\boldsymbol{\Sigma}_x$. Using $\{v_i^{(r)}, e_{ij}^{(r)}$

Table 1 MSPE of $\hat{\gamma}_i^B$ for the two scenarios $t_{i\ell} = n_i$ and $t_{i\ell} = 3n_i$.

Area	n_i	$t_{i\ell} = n_i$	$t_{i\ell} = 3n_i$
1	1	68.17	45.19
2	5	14.73	11.89
3	1	68.17	45.19
4	2	34.90	24.82
5	4	18.12	14.16
6	3	23.74	17.80
7	1	68.17	45.19
8	3	23.74	17.80
9	2	34.90	24.82
10	3	23.74	17.80
11	2	34.90	24.82
12	1	68.17	45.19

and $\boldsymbol{x}_i^{(r)}\}$, the values $y_{ij}^{(r)}$ are generated as $y_{ij}^{(r)} = 100 + 0.1w_{ij1} + 0.1w_{ij2} + 2x_{i1}^{(r)} + 2x_{i2}^{(r)} + v_i^{(r)} + e_{ij}^{(r)}; i = 1, \dots, m; j = 1, \dots, N_i; r = 1, \dots, 5000$. The rth simulated population mean for the ith area is given by

$$\gamma_i^{(r)} = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}^{(r)}.$$

Further, $\eta_{i\ell k}^{(r)}$ is generated independently from a normal distribution with mean zero and variance $\sigma_{\eta\ell}^2$ for $i = 1, \dots, m; \ell = 1, 2; k = 1, \dots, t_{i\ell}$ and the observed covariates are taken as $X_{i\ell k}^{(r)} = x_{i\ell}^{(r)} + \eta_{i\ell k}^{(r)}$. From each simulated population, we draw simple random sample $\{y_{ij}^{(r)} : i = 1, \dots, m; j = 1, \dots, n_i\}$ and then, using the sample values $\{y_{ij}^{(r)}\}$ and the covariates values $\{X_{i\ell k}^{(r)}\}$, $\hat{\gamma}_i^{EB(r)}$ is computed for each $r = 1, \dots, 5000$ for the two scenarios $t_{i\ell} = n_i$ and $t_{i\ell} = 3n_i$. The empirical MSPE (EMSPE) of $\hat{\gamma}_i^{EB}$ is then calculated as

$$\text{EMSPE}(\hat{\gamma}_i^{EB}) = \frac{1}{R} \sum_{r=1}^{R} \left\{\hat{\gamma}_i^{EB(r)} - \gamma_i^{(r)}\right\}^2.$$

Table 2 shows that $\hat{\gamma}_i^{EB}$ is substantially more efficient in terms of EMSPE when we use larger sample sizes for the covariates ($t_{i\ell} = 3n_i$) compared to the same sample sizes ($t_{i\ell} = n_i$) with the range of EMSPE reduction from 3 % to 41 % across areas. Note that the result of Table 2 is consistent with the result of Table 1 in a way that the EMSPE of $\hat{\gamma}_i^{EB}$ is slightly larger than the corresponding MSPE of $\hat{\gamma}_i^B$ due to the estimation of model parameters.

We also compare the performance of our method with the naive approach that treats the second measurement error covariate x_{i2} as an error-free covariate. We obtain the naive EB predictor of γ_i, denoted by $\tilde{\gamma}_i^{EB}$, following (6) by setting measurement error variance for the covariate x_{i2} as zero. In other words, the data for the naive method are generated from (1)-(2) where $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{22})^\top, \boldsymbol{w}_{ij} = (w_{ij1}, w_{ij2}, \bar{X}_{i2})^\top, \boldsymbol{\beta}_2$ as β_{21}, and $\boldsymbol{x}_i$ as x_{i1}. We observe that the relative efficiency of the proposed EB predictor to the corresponding naive predictor, $\left\{\text{EMSPE}(\tilde{\gamma}_i^{EB})/\text{EMSPE}(\hat{\gamma}_i^{EB})\right\}$, ranges from 1.20 to 1.50 in the case of $t_i = n_i$ and decreases from 0.99 to 1.37 as t_i increases to $3n_i$.

To evaluate the magnitude of each term in the EMSPE of $\hat{\gamma}_i^{EB}$, we decompose the EMSPE of $\hat{\gamma}_i^{EB}$ as

$$\text{EMSPE}(\hat{\gamma}_i^{EB}) \equiv \text{M}_i = \text{M}_{1i} + \text{M}_{2i} + 2\text{M}_{3i}, \tag{12}$$

where

$$\text{M}_{1i} = R^{-1} \sum_{r=1}^{R} \left\{\hat{\gamma}_i^{B(r)} - \gamma_i^{(r)}\right\}^2,$$

Table 2 Empirical MSPE of $\hat{\gamma}_i^{EB}$ for the two scenarios $t_{i\ell} = n_i$ and $t_{i\ell} = 3n_i$.

Area	n_i	$t_{i\ell} = n_i$	$t_{i\ell} = 3n_i$
1	1	74.80	52.98
2	5	16.99	16.54
3	1	76.22	54.64
4	2	38.55	31.35
5	4	21.64	18.98
6	3	26.98	22.94
7	1	75.87	54.76
8	3	27.11	24.21
9	2	40.01	31.94
10	3	27.07	24.12
11	2	40.37	33.55
12	1	74.38	57.59

$$M_{2i} = R^{-1} \sum_{r=1}^{R} \left\{\hat{\gamma}_i^{EB(r)} - \hat{\gamma}_i^{B(r)}\right\}^2,$$

and

$$M_{3i} = R^{-1} \sum_{r=1}^{R} \left\{\hat{\gamma}_i^{B(r)} - \gamma_i^{(r)}\right\}\left\{\hat{\gamma}_i^{EB(r)} - \hat{\gamma}_i^{B(r)}\right\}.$$

Table 3 presents the result on the decomposition of EMSPE. It is clear from Table 3 that the leading term M_{1i} in (12) makes major contribution to EMSPE for both scenarios $t_{i\ell} = n_i$ and $t_{i\ell} = 3n_i$. On the other hand, the cross-product term M_{3i} is negligible relative to M_{2i} which is substantially smaller than M_{1i}.

Table 3 Components of empirical MSPE of $\hat{\gamma}_i^{EB}$ for the two scenarios $t_{i\ell} = n_i$ and $t_{i\ell} = 3n_i$.

		$t_{i\ell} = n_i$				$t_{i\ell} = 3n_i$			
Area	n_i	M_i	M_{1i}	M_{2i}	M_{3i}	M_i	M_{1i}	M_{2i}	M_{3i}
1	1	74.80	65.94	7.94	0.46	52.98	44.00	10.07	-0.54
2	5	16.99	14.30	2.79	-0.06	16.54	12.12	4.16	0.13
3	1	76.22	68.81	7.66	-0.12	54.64	44.78	10.01	-0.07
4	2	38.55	33.88	5.01	-0.17	31.35	23.78	7.65	-0.04
5	4	21.64	18.21	3.26	0.08	18.98	13.94	4.97	0.03
6	3	26.98	22.88	4.01	0.05	22.94	17.52	6.09	-0.33
7	1	75.87	68.66	7.75	-0.27	54.76	45.22	10.09	-0.28
8	3	27.11	23.08	3.92	0.05	24.21	18.12	6.01	0.04
9	2	40.01	34.73	5.02	0.13	31.94	24.55	7.66	-0.14
10	3	27.07	23.59	3.93	-0.22	24.12	18.08	5.97	0.04
11	2	40.37	35.33	5.09	-0.02	33.55	25.36	7.42	0.39
12	1	74.38	66.56	7.60	0.10	57.59	45.93	10.64	0.51

We also study the performance of the weighted and the unweighted jackknife MSPE estimators, $\text{mspe}_{J\nu}(\hat{\gamma}_i^{EB})$ and $\text{mspe}_J(\hat{\gamma}_i^{EB})$, of the EB estimator $\hat{\gamma}_i^{EB}$. The relative bias (RB) of a MSPE estimator, mspe, is given by

$$\text{RB} = \frac{\text{E(mspe)}}{\text{EMSPE}} - 1,$$

where E(mspe) is estimated by the average of simulated $\text{mspe}^{(r)}, (r = 1, \dots, R)$. Table 4 shows $\text{mspe}_{J\nu}$ given by (10) performs much better than mspe_J given by (11) in terms of RB. Behaviour of RB for areas 2 and 5 shows large RB for mspe_J (39.29% for area 2; 34.71% for area 5) for larger covariate sample sizes $t_{i\ell} = 3n_i$. But for the weighted version, these relative biases are small: $|RB| < 7\%$ for 10 of 12 areas and $< 12.5\%$ for all 12 areas in the case of $t_{i\ell} = 3n_i$.

Table 4 Percent relative bias of jackknife estimators of MSPE of $\hat{\gamma}_i^{EB}$ for the two scenarios $t_{i\ell} = n_i$ and $t_{i\ell} = 3n_i$.

		$t_{i\ell} = n_i$		$t_{i\ell} = 3n_i$	
Area	n_i	$\text{RB}[\text{mspe}_J(\hat{\gamma}_i^{EB})]$	$\text{RB}[\text{mspe}_{J\nu}(\hat{\gamma}_i^{EB})]$	$\text{RB}[\text{mspe}_J(\hat{\gamma}_i^{EB})]$	$\text{RB}[\text{mspe}_{J\nu}(\hat{\gamma}_i^{EB})]$
1	1	3.12	0.24	10.37	1.94
2	5	33.52	16.44	39.29	12.49
3	1	1.50	-1.47	6.89	-1.26
4	2	10.49	4.17	15.86	1.93
5	4	19.73	6.39	34.71	10.79
6	3	13.87	5.36	24.71	6.40
7	1	2.23	-0.86	7.18	-1.42
8	3	14.16	5.10	17.72	0.16
9	2	5.87	0.18	16.32	0.00
10	3	14.51	5.57	20.00	1.27
11	2	5.57	-0.41	7.22	-5.42
12	1	3.80	0.62	1.37	-6.36

6. Application

In this section, we apply our method to data from the 2013–2014 U.S. NHANES, using the 2014 U.S. NHIS as auxiliary information. Following Ybarra and Lohr [19], we consider 50 small domains (demographic subgroups) classified by race and ethnicity (Mexican American, Other Hispanic, White non-Hispanic, Black non-Hispanic and Other), by age group (20–29, 30–39, 40–49, 50–59 and 60–84), and by sex.

Height and weight for each respondent are measured in the NHANES medical examination, and the BMI is calculated as weight/height^2, in units kg/m^2. In the NHIS, by contrast, BMI is calculated using self-reported responses to the height and weight questions in the interview. Since a respondent may not report height and/or weight accurately, the NHIS variable for BMI may not be the same as the NHANES BMI. Instead, the NHIS estimates a characteristic related to BMI and we call this the *reported BMI*. We expect that the estimates of the reported BMI from the NHIS to be highly correlated across areas with the estimates of BMI from the NHANES. The 2013–2014 NHANES had BMI values for 5,588 persons in the demographic subgroups of interest, with domain sample sizes ranging from 31 to 479. The 2014 NHIS had reported BMI for 35,928 persons, with domain sample sizes between 77 and 4,866.

In addition to the reported BMI as a covariate measured with error from the NHIS, we also consider cholesterol level (called CHOLES) from the NHIS as another covariate measured with error. To improve direct estimates of the BMI from NHANES, we also use an error-free covariate current smoker (yes or no) called SMOKE from the same survey. Hence, our observed data for the analysis are $\{(y_{ij}, w_{ij}, X_{i1k}, X_{i2k}) : i = 1, \dots, m = 50; j = 1, \dots, n_i; k = 1, \dots, t_i\}$, where y_{ij} is BMI and w_{ij} is SMOKE from the NHANES, and (X_{i1k}, X_{i2k}) are reported BMI and CHOLES from the NHIS, respectively.

We estimate the model parameters as $\hat{\beta}_0 = 8.79, \hat{\beta}_1 = -1.01, \hat{\beta}_{21} = 0.68, \hat{\beta}_{22} = -0.90, \hat{\boldsymbol{\mu}}_x = (30.62, 0.30), \hat{\boldsymbol{\Sigma}}_x = diag(3.79, 0.03), \hat{\sigma}_v^2 = 1.67, \hat{\sigma}_e^2 = 46.94$ and $\hat{\boldsymbol{\sigma}}_\eta^2 = (217.31, 0.18)$, noting that β_1, β_{21}, and β_{22} refer to SMOKE, reported BMI, and CHOLES, respectively. To calculate the EB predictors, we need the population mean of error-free covariate SMOKE for each small area. However, those population means are not known. To address this problem, we use an EB estimator as proxy to the area mean $\bar{w}_{i(P)} = N_i^{-1} \sum_{j=1}^{N_i} w_{ij}$. We assume that $W_{ij}|\theta_i \sim Bern(\theta_i), j = 1, \dots, N_i$, and $\theta_i|\alpha, a \sim Beta[a\alpha, a(1-\alpha)]$, independently, $i = 1, \ldots, m$. We then obtain the best estimator

$$\bar{w}^B_{i(P)} = \frac{1}{N_i}\Big\{\sum_{j=1}^{n_i} w_{ij} + (N_i - n_i)\mathrm{E}(\theta_i|w_{i1}, \ldots, w_{in_i})\Big\},$$

of $\bar{w}_{i(P)}$, where

$$\mathrm{E}(\theta_i|w_{i1}, \ldots, w_{in_i}; a, \alpha) = \frac{\sum_{j=1}^{n_i} w_{ij} + a\alpha}{n_i + a}.$$

Noting that $\mathrm{E}(\sum_{j=1}^{n_i} w_{ij}) = n_i\alpha$, we get a moment estimator of α as $\hat{\alpha} = n^{-1}\sum_{i=1}^{m} n_i\bar{w}_i$. Further, we estimate a from the estimating equation

$$\sum_{i=1}^{m}\Big(\sum_{j=1}^{n_i} w_{ij}\Big)^2 = \frac{\Big\{\sum_{i=1}^{m} n_i(n_i-1)\Big\}\hat{\alpha}(1-\hat{\alpha})}{a+1} + \Big\{\sum_{i=1}^{m} n_i(n_i-1)\Big\}\hat{\alpha}^2 + n_i\hat{\alpha}.$$

We do not know the population size N_i in our application, but the N_i's are large compared to the corresponding n_i. Hence, we can approximate $\bar{w}^B_{i(P)}$ by $\bar{w}^{*B}_{i(P)} = \mathrm{E}(\theta_i|w_{i1}, \ldots, w_{in_i})$. Substituting $\hat{\alpha}$ and $\hat{a}$ for α and a in $\bar{w}^{*B}_{i(P)}$ leads to the EB estimator $\bar{w}^{EB}_{i(P)} \approx \frac{n_i\bar{w}_i + \hat{a}\hat{\alpha}}{n_i + \hat{a}}$.

Figures 1 and 2 show the boxplots of EB predictors of small area means and the estimated MSPE of EB predictors of small area means, respectively. Figure 2 shows that the MSPE estimators for the unweighted and weighted jackknife approaches are similar. Note that the weight incorporated in the MSE estimator is of order $1/m$. This is the reason for similarity of the two MSE estimators (weighted and unweighted) in the case of relatively large $m(= 50)$; however, in the simulation study we showed that for small $m(= 12)$, the weighted MSE estimator performed better than the unweighted MSE estimator.

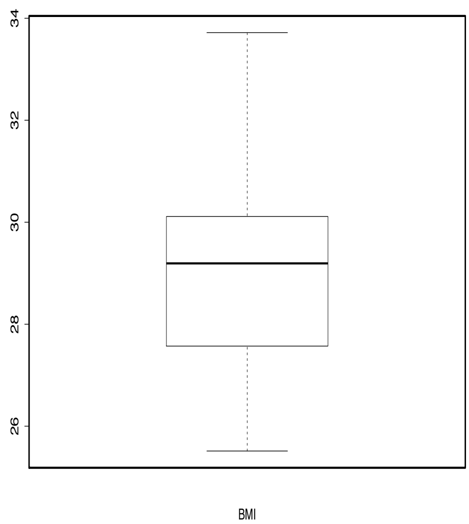

Fig. 1 Boxplot of empirical best predictors of small area means using the structural measurement error model.

7. Concluding remarks

We have derived fully efficient empirical best (EB) predictors of small area means under a nested error linear regression model with multiple error-free unit-level covariates and multiple area-level covariates subject to structural

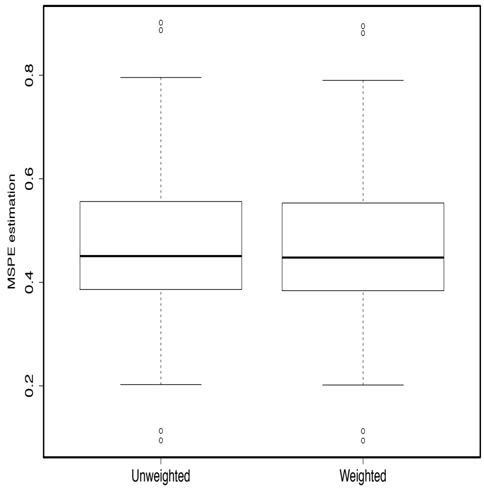

Fig. 2 Boxplots of jackknife estimate of MSPE of small area mean predictors using the structural measurement error model for the unweighted and weighted jackknife approaches.

measurement errors, assuming that the observed covariates can be from other (and possibly bigger) surveys in addition to the response variable survey. We have also proposed jackknife estimators of the mean squared prediction error (MSPE) of the EB predictors. We have shown through a simulation that using covariates with possibly larger sample sizes is more efficient than using the same sample sizes associated with the response variable in terms of empirical MSPE of EB predictors. We have also observed that our proposed approach is more efficient than the naive approach, which ignores the measurement errors in a covariate, in terms of empirical MSPE of the EB predictors.

Supplementary Materials

The supplementary materials provide R codes and corresponding readme files as well as related datasets for the simulation and application conducted in this paper.

Acknowledgements

We would like to thank two referees for constructive comments and suggestions. The research of Mahmoud Torabi and J.N.K. Rao was supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix

Proof of Theorem 2.1. For $j = n_i + 1, \dots, N_i$, we have

$$\mathrm{E}\Big\{y_{ij}|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \beta_0 + \boldsymbol{\beta}_1^\top \boldsymbol{w}_{ij} + \mathrm{E}\Big\{\boldsymbol{\beta}_2^\top \boldsymbol{x}_i + v_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\}.$$

One can also write

$$\mathrm{E}\Big\{v_i|\boldsymbol{x}_i, \boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2/n_i}\Big(\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \boldsymbol{x}_i\Big).$$

We then have

$$\mathrm{E}\Big\{y_{ij}|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \beta_0 + \boldsymbol{\beta}_1^\top \boldsymbol{w}_{ij} +$$

$$\mathrm{E}\Big[\Big\{\boldsymbol{\beta}_2^\top \boldsymbol{x}_i + \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2/n_i}(\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \boldsymbol{x}_i)\Big\}\Big|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big]. \tag{A. 1}$$

Next, we predict $\boldsymbol{x}_i$. The best predictor of $\boldsymbol{x}_i$ is given by $\mathrm{E}\{\boldsymbol{x}_i|\boldsymbol{y}_i^{(1)}, \mathbf{X}_i^{(1)}\}$, denoted by $\hat{\boldsymbol{x}}_i$. Due to normal distribution theory, $\hat{\boldsymbol{x}}_i$ is obtained by minimizing

$$Q(\boldsymbol{x}_i) = \{\boldsymbol{r}_i^{(1)} - \mathbf{1}_{n_i}\boldsymbol{\beta}_2^\top \boldsymbol{x}_i\}^\top (\sigma_e^2 I_{n_i} + \sigma_v^2 J_{n_i})^{-1}\{\boldsymbol{r}_i^{(1)} - \mathbf{1}_{n_i}\boldsymbol{\beta}_2^T \boldsymbol{x}_i\} +$$

$$(\boldsymbol{x}_i - \bar{\boldsymbol{X}}_i)^\top \Sigma_{i\eta}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{X}}_i) + (\boldsymbol{x}_i - \boldsymbol{\mu}_x)^\top \Sigma_x^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_x),$$

where $\boldsymbol{r}_i^{(1)} = \boldsymbol{y}_i^{(1)} - \beta_0 \mathbf{1}_{n_i} - \boldsymbol{w}_i^{(1)}\boldsymbol{\beta}_1$. We solve $\frac{\partial Q(\boldsymbol{x}_i)}{\partial \boldsymbol{x}_i} = \mathbf{0}$ to find $\hat{\boldsymbol{x}}_i$. After some algebra and simplification, we obtain

$$\hat{\boldsymbol{x}}_i = \bar{\boldsymbol{X}}_i - \Big(\mathbf{I}_{n_i} + \boldsymbol{\Sigma}_{i\eta}\boldsymbol{\Sigma}_x^{-1} + \frac{n_i}{\sigma_e^2 + n_i\sigma_v^2}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^\top\Big)^{-1}\Big\{\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\Sigma}_x^{-1}(\bar{\boldsymbol{X}}_i - \boldsymbol{\mu}_x) -$$

$$\frac{n_i \boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2}{\sigma_e^2 + n_i\sigma_v^2}(\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \bar{\boldsymbol{X}}_i)\Big\},$$

and then, after further simplification,

$$\boldsymbol{\beta}_2^\top \hat{\boldsymbol{x}}_i = \boldsymbol{\beta}_2^\top \bar{\boldsymbol{X}}_i - \Big(\frac{\sigma_e^2 + n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2 + n_i\boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2}\Big)\boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{i\eta}(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_{i\eta})^{-1}(\bar{\boldsymbol{X}}_i - \boldsymbol{\mu}_x) +$$

$$\Big(\frac{n_i\boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2}{\sigma_e^2 + n_i\sigma_v^2 + n_i\boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2}\Big)\Big(\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \bar{\boldsymbol{X}}_i\Big),$$

where $\mathbf{H}_i = \mathbf{I}_{n_i} + \boldsymbol{\Sigma}_{i\eta}\boldsymbol{\Sigma}_x^{-1}$. We can then write

$$\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \hat{\boldsymbol{x}}_i = (\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \bar{\boldsymbol{X}}_i)$$

$$+\Big(\frac{\sigma_e^2 + n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2 + n_i\boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2}\Big)\Big\{\boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{i\eta}(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_{i\eta})^{-1}(\bar{\boldsymbol{X}}_i - \boldsymbol{\mu}_x)\Big\}$$

$$-\Big(\frac{n_i\boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2}{\sigma_e^2 + n_i\sigma_v^2 + n_i\boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2}\Big)\Big(\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \bar{\boldsymbol{X}}_i\Big).$$

$$= \Big(\frac{\sigma_e^2 + n_i\sigma_v^2}{\sigma_e^2 + n_i\sigma_v^2 + n_i\boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2}\Big)\Big\{\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \boldsymbol{\mu}_x - \boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_x(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_{i\eta})^{-1}(\bar{\boldsymbol{X}}_i - \boldsymbol{\mu}_x)\Big\}.$$

Hence, for $j = n_i + 1, \ldots, N_i$, using (A. 1) we can write, after some simplification,

$$\mathrm{E}\Big\{y_{ij}|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \beta_0 + \boldsymbol{\beta}_1^\top \boldsymbol{w}_{ij}$$

$$+\Big\{\frac{\sigma_e^2}{\sigma_e^2 + n_i(\sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2)}\Big\}\Big[\boldsymbol{\beta}_2^\top\{\boldsymbol{\Sigma}_x(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_{i\eta})^{-1}\bar{\boldsymbol{X}}_i + \boldsymbol{\Sigma}_{i\eta}(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_{i\eta})^{-1}\boldsymbol{\mu}_x\}\Big]$$

$$+\Big\{\frac{n_i(\sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2)}{\sigma_e^2 + n_i(\sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{H}_i^{-1}\boldsymbol{\Sigma}_{i\eta}\boldsymbol{\beta}_2)}\Big\}(\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i). \tag{A. 2}$$

Hence, the best estimator of γ_i is

$$\hat{\gamma}_i^B = \mathrm{E}\Big\{\gamma_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \frac{n_i}{N_i}\bar{y}_i + \frac{1}{N_i}\sum_{j=n_i+1}^{N_i} E\Big\{y_{ij}|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\},$$

which leads to (3) by substituting (A. 2) in the above equation.

Moreover, to get $V\Big\{\gamma_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\}$, we can write

$$V\Big\{\gamma_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \Big(\frac{N_i - n_i}{N_i}\Big)^2 \Big[V\Big\{\boldsymbol{\beta}_2^\top \boldsymbol{x}_i + v_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} + \sigma_e^2/(N_i - n_i)\Big]. \tag{A. 3}$$

However, to get $V\Big\{\boldsymbol{\beta}_2^T \boldsymbol{x}_i + v_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\}$, we need to find the joint distribution of $(v_i, \boldsymbol{x}_i)$ conditional on the data. In particular, noting that

$$f(v_i, \boldsymbol{x}_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}) \propto \exp\Big[\frac{-1}{2}\Big\{\frac{n_i}{\sigma_e^2}(\bar{y}_i - \beta_0 - \boldsymbol{\beta}_1^\top \bar{\boldsymbol{w}}_i - \boldsymbol{\beta}_2^\top \boldsymbol{x}_i - v_i)^2 + \frac{v_i^2}{\sigma_v^2}$$
$$+(\bar{\boldsymbol{X}}_i - \boldsymbol{x}_i)^\top \boldsymbol{\Sigma}_{i\eta}^{-1}(\bar{\boldsymbol{X}}_i - \boldsymbol{x}_i) + (\boldsymbol{x}_i - \boldsymbol{\mu}_x)^\top \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_x)\Big\}\Big].$$

We get

$$-\frac{\partial^2 logf(v_i, \boldsymbol{x}_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi})}{\partial v_i^2} = (\sigma_e^2/n_i)^{-1} + \sigma_v^{-2},$$
$$-\frac{\partial^2 logf(v_i, \boldsymbol{x}_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi})}{\partial \boldsymbol{x}_i \partial \boldsymbol{x}_i^\top} = (\sigma_e^2/n_i)^{-1}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^\top + \boldsymbol{\Sigma}_{i\eta}^{-1} + \boldsymbol{\Sigma}_x^{-1},$$
$$-\frac{\partial^2 logf(v_i, \boldsymbol{x}_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi})}{\partial v_i \partial \boldsymbol{x}_i} = \boldsymbol{\beta}_2(\sigma_e^2/n_i)^{-1}.$$

Using the property of multivariate normal distribution, the variance-covariance matrix is the inverse of the Hessian matrix of the negative log-density, we obtain

$$V\Big\{\begin{pmatrix} v_i \\ \boldsymbol{x}_i \end{pmatrix}|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \begin{bmatrix} (\sigma_e^2/n_i)^{-1} + \sigma_v^{-2} & (\sigma_e^2/n_i)^{-1}\boldsymbol{\beta}_2^\top \\ \boldsymbol{\beta}_2(\sigma_e^2/n_i)^{-1} & (\sigma_e^2/n_i)^{-1}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^\top + \mathbf{M}_i \end{bmatrix}^{-1},$$

where $\mathbf{M}_i = \boldsymbol{\Sigma}_{i\eta}^{-1} + \boldsymbol{\Sigma}_x^{-1}$. After some algebra and simplification, we have

$$V\Big\{\begin{pmatrix} v_i \\ \boldsymbol{x}_i \end{pmatrix}|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \begin{bmatrix} \frac{\sigma_v^2(\sigma_e^2/n_i + \boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}\boldsymbol{\beta}_2)}{\sigma_v^2 + \sigma_e^2/n_i + \boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}\boldsymbol{\beta}_2} & \frac{-\sigma_v^2\boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}}{\sigma_e^2/n_i + \sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}\boldsymbol{\beta}_2} \\ \frac{-\mathbf{M}_i^{-1}\boldsymbol{\beta}_2\sigma_v^2}{\sigma_e^2/n_i + \sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}\boldsymbol{\beta}_2} & \mathbf{M}_i^{-1} - \frac{\mathbf{M}_i^{-1}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}}{\sigma_e^2/n_i + \sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}\boldsymbol{\beta}_2} \end{bmatrix}.$$

Hence, we get, after further simplification,

$$V\Big\{v_i + \boldsymbol{\beta}_2^\top \boldsymbol{x}_i|\boldsymbol{y}_i^{(1)}, \boldsymbol{X}_i^{(1)}, \boldsymbol{w}_i^{(1)}, \boldsymbol{\phi}\Big\} = \frac{(\sigma_e^2/n_i)(\sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}\boldsymbol{\beta}_2)}{\sigma_e^2/n_i + \sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}\boldsymbol{\beta}_2} \tag{A. 4}$$
$$= D_i(\sigma_v^2 + \boldsymbol{\beta}_2^\top \mathbf{M}_i^{-1}\boldsymbol{\beta}_2) = g_{1i}(\delta).$$

Now substituting (A. 4) in (A. 3), we get (4) which completes the proof of Theorem 2.1.

References

[1] S. Arima, G.S. Datta, B. Liseo, Objective Bayesian analysis of a measurement error small area model, Bayesian Anal. 7 (2012) 363–384.

[2] S. Arima, G.S. Datta, B. Liseo, Bayesian estimators for small area models when auxiliary information is measured with error, Scand. J. Stat. 42 (2015) 518–529.

[3] G.E. Battese, R.M. Harter, W.A. Fuller, An error-components model for prediction of county crop areas using survey and satellite data. J. Amer. Statist. Assoc. 83 (1988) 28–36.

[4] S. Chen, P. Lahiri, On mean squared prediction error estimation in small area estimation problems. In: Proceedings of the Survey Research Methods Section, Amer. Statist. Assoc., Washington, 2002, pp. 473–477.

[5] G.S. Datta, Model-based approach to small area estimation. In: Handbook of Statistics: Sample Surveys: Inference and Analysis, Vol. 29B, eds. D. Pfeffermann and C. R. Rao, Amsterdam, The Netherlands: North-Holland, 2009, pp. 251–288.
[6] G.S. Datta, J.N.K. Rao, M. Torabi, Pseudo-empirical Bayes estimation of small area means under a nested error linear regression model with functional measurement errors, J. Statist. Plann. Inference 140 (2010) 2952–2962.
[7] R.E. Fay, R.A. Herriot, Estimates of income for small places: an application of James-Stein procedures to census data, J. Amer. Statist. Assoc. 74 (1979) 269–277.
[8] M. Ghosh, K. Sinha, D. Kim, Empirical and hierarchical Bayesian estimation in finite population sampling under structural measurement error models, Scand. J. Stat. 33 (2006) 591–608.
[9] M. Ghosh, K. Sinha, Empirical Bayes estimation in finite population sampling under functional measurement error models, J. Statist. Plann. Inference 137 (2007) 2759–2773.
[10] J. Jiang, Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems, Chapman & Hall/CRC, 2017.
[11] J. Jiang, P. Lahiri, Mixed model prediction and small area estimation, Test 15 (2006) 1–96.
[12] J. Jiang, P. Lahiri, S.-M. Wan, A unified Jackknife theory for empirical best prediction with M-estimation, Ann. Statist. 30 (2002) 1782–1810.
[13] D. Pfeffermann, New important developments in small area estimation, Stat. Sci. 28 (2013) 40–68.
[14] N.G.N. Prasad, J.N.K. Rao, The estimation of mean squared error of small-area estimators, J. Amer. Statist. Assoc. 85 (1990) 163–171.
[15] J.N.K. Rao, I. Molina, Small Area Estimation, 2nd Edition, Wiley, Hoboken, New Jersey, 2015.
[16] M. Torabi, Small area estimation using survey weights with functional measurement error in the covariate, Aust. N. Z. J. Stat. 53 (2011) 141–155.
[17] M. Torabi, Small area estimation using survey weights under a nested error linear regression model with structural measurement error, J. Multivariate Anal. 109 (2012) 52–60.
[18] M. Torabi, G.S. Datta, J.N.K. Rao, Empirical Bayes estimation of small area means under a nested error linear regression model with measurement errors in the covariates, Scand. J. Stat. 36 (2009) 355–368.
[19] L.M.R. Ybarra, S.L. Lohr, Small area estimation when auxiliary information is measured with error, Biometrika 95 (2008) 919–931.