The Canadian Journal of Statistics
Vol. xx, No. yy, 2015, Pages 1–24
La revue canadienne de statistique

Pseudo-Empirical Bayes Estimation of Small Area Means Based on the James-Stein Estimation in Linear Regression Models with Functional Measurement Error

Elaheh Torkashvand[1], Mohammad Jafari Jozani[1]* and Mahmoud Torabi[2]

[1] *University of Manitoba, Department of Statistics, Winnipeg, MB, CANADA, R3T 2N2*
[2] *University of Manitoba, Department of Community Health Sciences, Winnipeg, MB, CANADA, R3E 0W3*

Key words and phrases: Empirical Bayes; Jackknife method; James-Stein estimator; Mean squared prediction error; Small area estimation.

MSC 2010: Primary 62F10; secondary 62F30; tertiary 62C20

Abstract: Small area estimation plays an important role in making reliable inference for subpopulations (areas) for which relatively small samples or no samples are available. In model-based small area estimation studies, linear and generalized linear mixed models have been used extensively assuming that covariates are not subjected to measurement errors. Recently, there have been studies considering this problem under the functional measurement error for covariates using the maximum likelihood method and the method of moments. In this paper, we study the James-Stein estimator of the true covariate subject to the functional measurement error. To this end, we obtain a new pseudo-empirical Bayes (PEB) predictor of small area means based on the James-Stein estimator. Then, we show that the new PEB predictor is asymptotically optimal. The weighted and unweighted jackknife estimators of the mean squared prediction error of the new PEB predictor are also derived. Simulation studies are conducted to evaluate the performance of the proposed approach. We observe that the PEB predictor based on the James-Stein estimator performs better than those based on the maximum likelihood method and the method of moments. Finally, we apply the proposed methodology to a real dataset. *The Canadian Journal of Statistics* xx: 1–24; 2015 © 2015 Statistical Society of Canada

1. INTRODUCTION

Small area estimation is a statistical method to find estimates of means, totals or any other parameters associated with the quantity of interest in subpopulations. Depending on the number of sample units from subpopulations (areas), we can have either "direct" or "indirect" estimates of the parameters of interest. If there is enough sample units in the specific subpopulation to derive estimates of parameters, they are called "direct" estimates. However, due to the cost and operational consideration, it is not always possible to sample enough data to get direct estimates. In this case, "indirect" estimates of the parameters of interest can be obtained by borrowing information from other areas through a linking model which formulates the relationship between the auxiliary information and the mean of the response variable in areas (Rao, 2003).

Sometimes the auxiliary information of the model is subject to measurement error. In such situations, it is natural to study small area estimation problem under measurement errors. To this end, Ghosh and Sinha (2007), Datta et al. (2010) and Torabi (2011) studied the unit level regression model for small area estimation when the area level covariate is subject to the functional

* *Author to whom correspondence may be addressed.*
E-mail: M_Jafari_Jozani@umanitoba.ca

measurement error.

Consider a small area estimation problem, and let N_i be the known value of the population size of the ith area, y_{ij} and X_{ij} represent the value of the study variable and the covariate associated with the jth unit in the ith area, respectively. We consider the case where X_{ij}'s contain measurement errors as one might not be able to precisely measure the true value of the area-specific covariate. In particular, consider the following model

$$y_{ij} = b_0 + b_1 x_i + u_i + e_{ij}, \quad i = 1, \ldots, m, j = 1, \ldots, N_i, \tag{1}$$

with

$$X_{ij} = x_i + \eta_{ij}, \quad i = 1, \ldots, m, j = 1, \ldots, N_i, \tag{2}$$

where m, u_i's, η_{ij}'s, and e_{ij}'s are the number of areas, the area-level random effects, the measurement error and the random errors, respectively. Also, assume that e_{ij}'s, η_{ij}'s and u_i's are mutually independent. Furthermore, it is assumed that u_i's, η_{ij}'s, and e_{ij}'s are independently and identically distributed ($i.i.d.$) random variables with $e_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$, $\eta_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_\eta^2)$ and $u_i \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$, where σ_u^2, σ_e^2 and σ_η^2 are all unknown.

In this paper, the functional measurement error model is considered for the true area-specific covariate x_i's. That is, the true area-specific covariates are unknown fixed parameters (Carroll et al., 2010). If there is no measurement error in the covariate variables X_{ij}, (1) and (2) reduce to the unit level regression model (Battese et al., 1988).

In small area estimation, one interest is to estimate small area means, γ_i's, which are defined as follows

$$\gamma_i = \frac{\sum_{j=1}^{N_i} Y_{ij}}{N_i}, \quad i = 1, \ldots, m.$$

Assuming there is no selection bias, let $\{(y_{ij}, X_{ij}),\ i = 1, \ldots, m,\ j = 1, \ldots, n_i\}$, be the observed values of Y_{ij} and the covariates X_{ij} where n_i is the sample size from the ith area. The best (or Bayes) predictor of γ_i under the squared error loss function using the observed values of $\mathbf{y}_i^{(1)} = (y_{i1}, y_{i2}, \ldots, y_{in_i})$ is given by

$$\begin{aligned}
\widehat{\gamma}_i^B &= \widehat{\gamma}_i^B(x_i, \phi) \\
&= \mathrm{E}(\gamma_i | \mathbf{y}_i^{(1)}) \\
&= (1 - f_i B_i)\bar{y}_i + f_i B_i (b_0 + b_1 x_i),
\end{aligned} \tag{3}$$

where $\phi = (b_0, b_1, \sigma_u^2, \sigma_e^2)$, $f_i = 1 - n_i/N_i$ is the finite population correction factor, $\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, and $B_i = \sigma_e^2/(\sigma_e^2 + n_i\sigma_u^2)$ (Ghosh and Sinha, 2007).

To estimate the unknown x_i's, Ghosh and Sinha (2007), henceforth abbreviated GS, proposed a pseudo-Bayes (PB) predictor of γ_i by replacing x_i with its moment estimator, $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$. Datta et al. (2010) suggested a new PB predictor of γ_i by using the maximum likelihood (ML) estimates of x_i's, i.e.

$$\tilde{x}_i(\psi) = \bar{X}_i + \frac{b_1\sigma_\eta^2(\bar{y}_i - b_0 - b_1\bar{X}_i)}{\sigma_e^2 + n_i\sigma_u^2 + b_1^2\sigma_\eta^2}, \tag{4}$$

where $\psi = (\phi, \sigma_\eta^2)$. Ghosh and Sinha (2007) and Datta et al. (2010), henceforth abbreviated DRT, also obtained a pseudo-empirical Bayes (PEB) predictor of γ_i by using the method of moments estimators of ψ proposed by Ghosh and Sinha (2007).

In this paper, we propose to estimate the unknown x_i's by the James-Stein estimator as a natural competitor to the ML and the method of moments estimators. The James-Stein estimator can be used to estimate the mean vector of a multivariate normal distribution which is more efficient than the ML estimate in terms of the sum of the weighted mean squared error,

$$\mathrm{E}\left[\sum_{i=1}^{m} W_i(\widehat{\gamma}_i - \gamma_i)^2\right],$$

where $m \geq 3$ and W_i's are variance stabilizing constants (Efron and Morris, 1972). The James-Stein estimator can be obtained as an empirical Bayes estimator, and it is robust against misspecifying the prior distribution (Lehmann and Casella, 1998).

Whittemore (1989) previously proposed the James-Stein estimator for the linear regression model, $y_i = b_0 + b_1 x_i + e_i$, where $e_i \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$, $i = 1, \ldots, m$, with the measurement errors in the covariate. Following Efron and Morris (1972), one can temporarily use the prior distribution $N(\mu, \tau^2)$, on the unknown x_i's in the measurement model $X_i = x_i + \eta_i$ with $\eta_i \overset{i.i.d.}{\sim} N(0, \sigma_\eta^2)$ and known σ_η^2, to obtain the James-Stein estimator, as an empirical Bayes estimator, as follows

$$\widehat{x}_i = \widehat{B}\bar{X} + (1 - \widehat{B})X_i \qquad i = 1, \ldots, m,$$

where $\widehat{B} = (m-3)\sigma_\eta^2 / \sum_{i=1}^{m}(X_i - \bar{X})^2$ and $\bar{X} = m^{-1}\sum_{i=1}^{m} X_i$.

In this paper, we use a similar idea to construct the James-Stein estimator of the true covariate, x_i, subject to the functional measurement error (see Carroll et al., 2010, Sec. 9.1.3 and Carroll et al., 1999 for more details). To this end, in Section 2, we first obtain the James-Stein estimator of the true area-specific covariate, x_i, in the unit level regression model (1). In Section 3, a new PEB predictor of the small area mean, γ_i, is constructed and its optimality is investigated. In Section 4, the jackknife estimators of the mean squared prediction error (MSPE) of the proposed PEB predictor of the small area mean are derived. The performance of the proposed approach is evaluated through simulation studies in Section 5. The cross-sectional data from the New Zealand population is analyzed in Section 6. Some concluding remarks are given in Section 7. The Appendix is devoted to some of the proofs.

2. A NEW PSEUDO-BAYES PREDICTOR

Following models (1) and (2), let

$$\bar{y}_i = b_0 + b_1 x_i + u_i + \bar{e}_i, \tag{5}$$

and

$$\bar{X}_i = x_i + \bar{\eta}_i, \tag{6}$$

where $\bar{e}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} e_{ij}$ and $\bar{\eta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \eta_{ij}$. Define $\bar{Z}_i^*$ as a linear estimator of x_i based on $\bar{X}_i$ and $\bar{y}_i$ as follows

$$\begin{aligned}\bar{Z}_i^* &= \bar{X}_i + h_i(\bar{y}_i - b_0 - b_1\bar{X}_i)\\ &= x_i + h_i u_i + h_i\bar{e}_i + \bar{\eta}_i(1 - h_i b_1), \quad i = 1, \ldots, m.\end{aligned} \tag{7}$$

It is easy to show that $\bar{Z}_i^*$ is an unbiased estimator of x_i with

$$\bar{Z}_i^*|x_i \sim N\left(x_i, h_i^2(\sigma_u^2 + \frac{\sigma_e^2}{n_i}) + \frac{\sigma_\eta^2}{n_i}(1 - h_i b_1)^2\right).$$

We first obtain the best linear estimator of $\bar{Z}_i^*$ by minimizing $\text{var}_{h_i}(\bar{Z}_i^*|x_i)$ with respect to h_i. To this end, by taking the derivative of $\text{var}_{h_i}(\bar{Z}_i^*|x_i)$ with respect to h_i, we have

$$h_i\left(\sigma_u^2 + \frac{\sigma_e^2}{n_i} + b_1^2\frac{\sigma_\eta^2}{n_i}\right) = \frac{b_1\sigma_\eta^2}{n_i},$$

which results in the optimum value of h_i as follows

$$h_i^* = \frac{b_1\sigma_\eta^2}{n_i\sigma_u^2 + \sigma_e^2 + b_1^2\sigma_\eta^2}, \quad i = 1, \ldots, m.$$

One can easily see that $\bar{Z}_i^*$ using h_i^* leads to the ML estimator of x_i defined in (4).

To get the James-Stein estimator, the method described in Efron and Morris (1972, 1975) is used by temporarily assuming that $x_i \overset{i.i.d.}{\sim} N(\mu, \tau^2)$. Let $\sigma_{0i}^2 = \text{var}_{h_i^*}(\bar{Z}_i^*|x_i)$ and observe that

$$x_i|\bar{Z}_i^* \sim N\left(\frac{\sigma_{0i}^2}{\sigma_{0i}^2 + \tau^2}\mu + \frac{\tau^2}{\sigma_{0i}^2 + \tau^2}\bar{Z}_i{}^*, \frac{\sigma_{0i}^2\tau^2}{\sigma_{0i}^2 + \tau^2}\right), \quad i = 1, \ldots, m.$$

The Bayes estimator of x_i under the quadratic loss function is

$$\text{E}(x_i|\bar{Z}_i^*) = C_i\mu + (1 - C_i)\bar{Z}_i^*, \quad i = 1, \ldots, m, \tag{8}$$

where $C_i = \sigma_{0i}^2/(\sigma_{0i}^2 + \tau^2)$. Now, the James-Stein estimator of x_i is given by the empirical Bayes estimate of $\text{E}(x_i|\bar{Z}_i^*)$. To this end, we first need to estimate the unknown parameters τ^2 and μ. One can estimate τ^2 using the marginal distribution of $\bar{Z}_i{}^* \sim N(\mu, \sigma_{0i}^2 + \tau^2)$, $i = 1, \ldots, m$. Efron and Morris (1975) gave an estimate of τ^2 as the solution of

$$\tau^2 = \frac{\sum_{i=1}^m((\bar{Z}_i^* - \mu)^2 - \sigma_{0i}^2)\text{I}_i(\tau^2)}{\sum_{i=1}^m \text{I}_i(\tau^2)}, \tag{9}$$

where $\text{I}_i(\tau^2) = \left\{\text{var}(\bar{Z}_i^* - \mu)^2\right\}^{-1} = \frac{1}{2}(\sigma_{0i}^2 + \tau^2)^{-2}$ is the Fisher information of τ^2 contained in $(\bar{Z}_i^* - \mu)^2$.

In (9), $(\bar{Z}_i^* - \mu)^2$ is used to remove the effect of the non-centrality parameter. We also estimate μ using its ML estimator, given by

$$\bar{Z}^* = \frac{\sum_{i=1}^m \dfrac{\bar{Z}_i^*}{\sigma_{0i}^2 + \tau^2}}{\sum_{i=1}^m \dfrac{1}{\sigma_{0i}^2 + \tau^2}}.$$

Hence, the James-Stein estimator of x_i is obtained as follows

$$x_{i\text{JS}} = C_i\bar{Z}^* + (1 - C_i)\bar{Z}_i^*, \quad i = 1, \ldots, m. \tag{10}$$

The new PB predictor of γ_i, following (3), is then given by

$$\widehat{\gamma}_{i\text{JS}}^{\text{PB}} = (1 - f_i B_i)\bar{y}_i + f_i B_i (b_0 + b_1 x_{i\text{JS}}), \quad i = 1, \ldots, m. \tag{11}$$

Remark 1. *In the structural measurement error, we have the assumption that $x_i \overset{i.i.d}{\sim} N(\mu_x, \sigma_x^2)$, for $i = 1, \ldots, m$ (Carroll et al., 2010). One can easily observe that the James-Stein predictors of x_i's under the structural measurement error (assuming x_i's are random variables) are equal to the James-Stein estimators of x_i's under the functional measurement error (assuming x_i's are unknown fixed values). Note that, x_i's in the structural measurement error are predicted while they are being estimated in the context of the functional measurement error.*

We now derive the MSPE of $\widehat{\gamma}_{i\text{JS}}^{\text{PB}}$ in (11) as follows:

$$\begin{aligned}
\mathrm{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}} - \gamma_i)^2 &= \mathrm{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}} - \widehat{\gamma}_i^B)^2 + \mathrm{E}(\widehat{\gamma}_i^B - \gamma_i)^2 \\
&= (f_i B_i b_1)^2 \left\{ \left(C_i x_i (d_i - 1) + C_i \sum_{j \neq i} x_j d_j \right)^2 + \sigma_{0i}^2 \left(1 + C_i(d_i - 1)\right)^2 \right. \\
&\left. + C_i^2 \sum_{j \neq i} \sigma_{0j}^2 d_j^2 \right\} + f_i^2 \left(\sigma_e^2 \left(\frac{(1 - B_i)^2}{n_i} + \frac{1}{N_i - n_i} \right) + B_i^2 \sigma_u^2 \right) \qquad (12) \\
&\equiv g_{1i}(\delta), \quad i = 1, \ldots, m,
\end{aligned}$$

where

$$d_i = \frac{\dfrac{1}{(\sigma_{0i}^2 + \tau^2)}}{\sum_{j=1}^m \dfrac{1}{(\sigma_{0j}^2 + \tau^2)}},$$

and $\delta = (b_1, \sigma_u^2, \sigma_e^2, \sigma_\eta^2, (x_1, \ldots, x_m))^\top$.

In fact, MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PB}}$) in (12) is the measure of variability of $\widehat{\gamma}_{i\text{JS}}^{\text{PB}}$ where the model parameters are known. However, δ is unknown and we need to estimate δ to obtain a pseudo-empirical Bayes predictor of γ_i and to evaluate its optimality.

Remark 2. *One can easily observe that the* MSPE *obtained in (12) depends on $(x_1, \ldots, x_m)^\top$. As $m \to \infty$, the dimension of the parameter space goes to infinity. However, using (10), we note that $(x_1, \ldots, x_m)^\top$ can be obtained by estimating ψ and τ^2 no matter how large m is. Also, following the Bayesian view towards the functional measurement error, and by applying the information provided in the prior distribution $N(\mu, \tau^2)$ in (12), we have*

$$\begin{aligned}
\bar{\mathrm{E}}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}} - \gamma_i)^2 &= \mathrm{E}\{\mathrm{E}[(\widehat{\gamma}_{i\text{JS}}^{\text{PB}} - \gamma_i)^2 | x_1, \ldots, x_m]\} \\
&= (f_i B_i b_1)^2 \left(C_i^2 \tau^2 \left(1 + \sum_{j=1}^m d_j^2 - 2 d_i \right) + \sigma_{0i}^2 \left(1 + C_i(d_i - 1)\right)^2 \right. \\
&\left. + C_i^2 \sum_{j \neq i} \sigma_{0j}^2 d_j^2 \right) + f_i^2 \left(\sigma_e^2 \left(\frac{(1 - B_i)^2}{n_i} + \frac{1}{N_i - n_i} \right) + B_i^2 \sigma_u^2 \right),
\end{aligned}$$

where $\bar{\mathrm{E}}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}-\gamma_i)^2$ *is obtained using the total law of expectation with* $\bar{\mathrm{E}}$ *being used for double expectations.Thus, the unconditional* MSPE*, denoted by* $\bar{\mathrm{E}}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}-\gamma_i)^2$ *does not depend on the specific values of* x_i*'s.*

3. PSEUDO-EB PREDICTOR

In this section, we first obtain a pseudo EB (PEB) predictor, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, of γ_i by using the method of moments to estimate the model parameters. We then show that this PEB predictor is asymptotically optimal in the sense that $\frac{1}{m}\sum_{i=1}^m \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})^2 \to 0$ as $m\to\infty$. To this end, similar to Ghosh and Sinha (2007) and Datta et al. (2010), the method of moments estimator of $\psi=(b_0,b_1,\sigma_u^2,\sigma_e^2,\sigma_\eta^2)$, $\widehat{\psi}$, is used. Thus, we have the following consistent estimators (Ghosh and Sinha, 2007)

$$\widehat{\sigma}_\eta^2=\mathrm{MSW}_x, \qquad \widehat{\sigma}_e^2=\mathrm{MSW}_y,$$
$$\widehat{b}_1=\frac{\mathrm{MSB}_x}{\mathrm{MSB}_x-\mathrm{MSW}_x}\tilde{b}_1, \qquad \widehat{b}_0=\bar{y}-\widehat{b}_1\bar{X},$$

and

$$\widehat{\sigma}_u^2=\max\left\{0,\{\mathrm{MSB}_y-\mathrm{MSW}_y-\widehat{b}_1^2(\mathrm{MSB}_x-\mathrm{MSW}_x)\}\frac{m-1}{g_m}\right\},$$

where

$$\mathrm{MSB}_y=\frac{1}{m-1}\sum_{i=1}^m n_i(\bar{y}_i-\bar{y})^2, \quad \mathrm{MSW}_y=\frac{1}{n_T-m}\sum_{i=1}^m\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2,$$
$$\mathrm{MSB}_x=\frac{1}{m-1}\sum_{i=1}^m n_i(\bar{\mathbf{X}}_i-\bar{\mathbf{X}})^2, \quad \mathrm{MSW}_x=\frac{1}{n_T-m}\sum_{i=1}^m\sum_{j=1}^{n_i}(X_{ij}-\bar{\mathbf{X}}_i)^2,$$
$$n_T=\sum_{i=1}^m n_i, \quad \bar{\mathbf{X}}=\frac{1}{n_T}\sum_{i=1}^m n_i\bar{\mathbf{X}}_i,$$
$$\bar{y}=\frac{1}{n_T}\sum_{i=1}^m n_i\bar{y}_i, \quad \tilde{b}_1=\frac{\sum_{i=1}^m n_i\bar{y}_i(\bar{\mathbf{X}}_i-\bar{\mathbf{X}})}{(m-1)\mathrm{MSB}_x},$$

and $g_m=n_T-\sum_{i=1}^m n_i^2/n_T$. The PEB predictor of γ_i is then given by

$$\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}=(1-f_i\widehat{B}_i)\bar{y}_i+f_i\widehat{B}_i(\widehat{b}_0+\widehat{b}_1\widehat{x}_{i\mathrm{JS}}), \quad i=1,\ldots,m,$$

where $\widehat{B}_i$ and $\widehat{x}_{i\mathrm{JS}}$ are estimated values of B_i and $x_{i\mathrm{JS}}$ obtained by replacing consistent estimates of parameters. For asymptotic optimality of the proposed PEB predictor, we need the following assumptions from Ghosh and Sinha (2007)

$$(i)\ \max_{1\le i\le m} n_i \le k_0<\infty, \quad (ii)\ \frac{1}{m-1}\sum_{i=1}^m n_i(x_i-\bar{x})^2\to c\,(>0)\ \text{as}\ m\to\infty. \tag{13}$$

Theorem 3 establishes the asymptotic optimality of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$. In this paper, following Ghosh and Sinha (2007) and Datta et al. (2010), we refer to the optimality in the sense of Robbins (1956).

Theorem 1. *Under the assumptions (13), $\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$ is asymptotically optimal in the sense that*

$$\frac{1}{m}\sum_{i=1}^{m}\text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\widehat{\gamma}_{i\text{JS}}^{\text{PB}})^2 \to 0, \text{ as } m\to\infty, \tag{14}$$

where the expectation is with respect to random variables $\mathbf{X}, \mathbf{Y}$, and $x_{i\text{JS}}$, $(i=1,\ldots,m)$.

Proof. See the Appendix for the proof. ■

To get a measure of the variability of $\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$, we calculate MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$) which is represented as

$$\text{MSPE}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}) = \text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}}-\gamma_i)^2 + \text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\widehat{\gamma}_{i\text{JS}}^{\text{PB}})^2 + 2\text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}}-\gamma_i)(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\widehat{\gamma}_{i\text{JS}}^{\text{PB}}), \tag{15}$$

where $\text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}}-\gamma_i)^2 = g_{1i}(\delta)$. Note that, in (15), $\text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\widehat{\gamma}_{i\text{JS}}^{\text{PB}})^2 + 2\text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}}-\gamma_i)(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\widehat{\gamma}_{i\text{JS}}^{\text{PB}})$ which comes from the variability of the model parameters estimation is unknown. In Section 4, using the jackknife method, we present an estimate of MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$).

4. THE JACKKNIFE ESTIMATION OF MSPE($\widehat{\gamma}_{I\text{JS}}^{\text{PEB}}$)

To find an estimator of

$$\text{MSPE}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}) = \text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\gamma_i)^2,$$

we follow Jiang et al. (2002), Chen and Lahiri (2002) and Datta et al. (2010) using the weighted and unweighted jackknife methods. Let

$$M_{1i} := \text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}}-\gamma_i)^2 = g_{1i}(\delta),$$
$$M_{2i} := \text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\widehat{\gamma}_{i\text{JS}}^{\text{PB}})^2,$$

and

$$M_{3i} := \text{E}(\widehat{\gamma}_{i\text{JS}}^{\text{PB}}-\gamma_i)(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\widehat{\gamma}_{i\text{JS}}^{\text{PB}}).$$

We have

$$\text{MSPE}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}) = M_{1i} + M_{2i} + 2M_{3i}. \tag{16}$$

Due to the structure of M_{3i}, it is not easy to find the explicit form of M_{3i}. As it is shown in Theorem 3, $\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$ is asymptotically optimal for $\gamma_{i\text{JS}}^{\text{PB}}$ in the sense of Robbins (1956). Using Theorem 3 and the Cauchy-Schwarz inequality, similar to Ghosh and Sinha (2007), we have

$$\begin{aligned}\text{E}\left(\frac{1}{m}\sum_{i=1}^{m}\left|(\gamma_{i\text{JS}}^{\text{PB}}-\gamma_i)(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\gamma_{i\text{JS}}^{\text{PB}})\right|\right) &\le \frac{1}{m}\sum_{i=1}^{m}\left[\text{E}^{1/2}(\gamma_{i\text{JS}}^{\text{PB}}-\gamma_i)^2\text{E}^{1/2}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\gamma_{i\text{JS}}^{\text{PB}})^2\right]\\ &\le \max_{1\le i\le m}\text{E}^{1/2}(\gamma_{i\text{JS}}^{\text{PB}}-\gamma_i)^2\left[\frac{1}{m}\sum_{i=1}^{m}\text{E}^{1/2}(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}-\gamma_{i\text{JS}}^{\text{PB}})^2\right]\\ &\to 0, \quad \text{as } m\to\infty.\end{aligned}$$

Knowing the mean of some positive terms goes to zero in limit implies that each term goes to zero in limit. So, we have

$$\mathrm{E}^{1/2}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})^2 \to 0, \text{ as } m \to \infty,$$

for $i = 1, \dots, m$. Thus, as

$$M_{3i} \leq \mathrm{E}\left|(\gamma_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})\right| \leq max_{1\leq i\leq m}\mathrm{E}^{1/2}(\gamma_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2\mathrm{E}^{1/2}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}} - \gamma_{i\mathrm{JS}}^{\mathrm{PB}})^2,$$

we have $M_{3i} \to 0$, as $m \to \infty$, for $i = 1, \dots, m$. Therefore, we ignore the cross product term M_{3i}, similar to Ybarra and Lohr (2008) and Datta et al. (2010), and use the approximation

$$\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}) \approx M_{1i} + M_{2i}, \tag{17}$$

in the jackknife method. However, we address the magnitude of the ignored term, M_{3i}, in the simulation study (Section 5). In particular, we found that (17) is quite accurate for the proposed predictor $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$.

To find the jackknife estimator of MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$), denoted by $mspe_J$, let

$$\widehat{M}_{1iJ} = g_{1i}(\widehat{\delta}) - \sum_{l=1}^{m} w_l[g_{1i}(\widehat{\delta}_{-l}) - g_{1i}(\widehat{\delta})],$$

and

$$\widehat{M}_{2iJ} = \sum_{l=1}^{m} w_l(\widehat{\gamma}_{iJS,-l}^{\mathrm{PEB}} - \widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}})^2,$$

where $\widehat{\delta}_{-l}$ is obtained by removing observations of the lth area, $l = 1, \dots, m$, and finding the model parameters estimate by using the information of other areas. Note that $\widehat{\gamma}_{iJS,-l}^{\mathrm{PEB}}$ is also PEB predictor of γ_i without using the information of lth area. Finally, weight w_l, $l = 1, \dots, m$, can be either $1 - \bar{\mathbf{X}}_l^\top(\sum_{t=1}^m \bar{\mathbf{X}}_t\bar{\mathbf{X}}_t^\top)^{-1}\bar{\mathbf{X}}_l$, where $\bar{\mathbf{X}}_l = (1, \bar{X}_l)^\top$, or $(m-1)/m$ leading to the weighted (Chen and Lahiri, 2002) or unweighted (Jiang et al., 2002) jackknife estimators of MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$), respectively. Hence, the $mspe_J$ is defined in either case as

$$mspe_J = \widehat{M}_{1iJ} + \widehat{M}_{2iJ}, \tag{18}$$

noting that the $mspe_J$ is a nearly unbiased estimator of MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$) if and only if $M_{3i} \approx 0$. Similarly, we use $mspe_{wJ}$ to show the weighted jackknife estimator of MSPE($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$).

5. EMPIRICAL RESULTS

In this section, we first provide a numerical study to compare the performance of the MSPE of the proposed pseudo-Bayes predictor $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ with those obtained by GS and DRT. We then conduct a simulation study to evaluate the relative efficiency of the proposed PEB predictor, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, compared with the GS estimator, $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}}$, the DRT estimator, $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$ and the naive estimator (henceforth abbreviated NAI), $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$. The NAI estimator is obtained in the absence of the measurement error (Battese et al., 1988). The performance of the proposed jackknife estimator, $mspe_J$, is also evaluated and compared with its competitors based on the GS, DRT and NAI methods.

Following Ghosh and Sinha (2007) and Datta et al. (2010), we assume that the responses y_{ij} for the population units and the observed covariates which contain measurement errors are generated from the model given by (1) and (2) with $\sigma_\eta^2 = 25$, $\sigma_u^2 = 16$, $\sigma_e^2 = 100$, $b_1 = 2$,

$b_0 = 100$. The population has $N = 3950$ units spread across 20 areas of sizes N_i given by 50, 250, 50, 100, 200, 150, 50, 150, 100, 150, 100, 50, 300, 350, 400, 200, 250, 300, 350, and 400. The sample sizes (n_i) within areas are taken to be 1, 5, 1, 2, 4, 3, 1, 3, 2, 3, 2, 1, 6, 7, 8, 4, 5, 6, 7, and 8, respectively. We also generate x_i's ranging from a uniform distribution between 191 to 199 given by $\mathbf{x} = (197, 198, 197, 192, 192, 195, 192, 196, 194, 192, 191, 197, 191, 193, 199, 198, 194, 199, 191, 196)$, and treat them fixed through the simulation study. We conduct the simulation study for $R = 5000$ iterations and for each iteration, we generate small area population $\mathbf{Y}_i^{(r)} = (y_{i1}^{(r)}, y_{i2}^{(r)}, \ldots, y_{iN_i}^{(r)})$ and associated simple random samples $(y_{i1}^{(r)}, \ldots, y_{in_i}^{(r)})$ and $(X_{i1}^{(r)}, \ldots, X_{in_i}^{(r)})$, $i = 1, \ldots, m$, $r = 1, \ldots, R$, independently.

5.1. Efficiency of PB Estimators

We first evaluate the performance of the proposed PB estimator and compare it with the GS, NAI, and DRT estimators based on their MSPE's.

The MSPE of the GS predictor $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PB}}$ can be obtained from the equation (3.5) of GS as follows

$$\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PB}}) = f_i^2 \left[\sigma_e^2 \left\{ \frac{(1-B_i)^2}{n_i} + \frac{1}{N_i - n_i} \right\} + B_i^2 \sigma_u^2 + \frac{b_1^2 B_i^2 \sigma_\eta^2}{n_i} \right],$$

while the NAI estimator gives

$$\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PB}}) = f_i^2 \left[\sigma_e^2 \left\{ \frac{(1-B_i)^2}{n_i} + \frac{1}{N_i - n_i} \right\} + B_i^2 \sigma_u^2 \right].$$

Similarly, the $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}})$ can be obtained from the equation (2.5) of DRT given by

$$\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}) = \frac{f_i^2 \sigma_e^2 (1 - A_i)}{n_i} + \frac{1}{N_i} f_i \sigma_e^2,$$

where $A_i = \sigma_e^2 / (\sigma_e^2 + n_i \sigma_u^2 + b_1^2 \sigma_\eta^2)$.

Table 1 reports the values of $\mathrm{MSPE}(\widehat{\gamma}_i^{\mathrm{PB}})$ based on the James-Stein estimator given by (12), and the estimators based on the DRT and GS methods. The relative efficiency of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ over $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}$, defined by $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}})/\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})$, ranged from 93.90% to 193.15%, and the relative efficiency of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ over $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PB}}$ ranged from 102.43% to 312.33%. In particular, our results confirm that in small areas with a very small number of samples, the James-Stein estimator is more efficient than the corresponding estimators based on the GS and DRT methods. One can also notice that the NAI method results in the smallest values of $\mathrm{MSPE}(\widehat{\gamma}^{\mathrm{PB}})$, however, as we show in Section 5.2, the EMSPE and the relative bias associated with this method are the largest ones among all methods considered in this paper. This is because the NAI method does not account for the measurement errors in the auxiliary variables (see Section 5.2 for more details).

The James-Stein estimator dominates the ML estimator in terms of the sum of the weighted MSPE, that is,

$$\sum_{i=1}^{m} W_i \mathrm{E}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}} - \gamma_i)^2 \leq \sum_{i=1}^{m} W_i \mathrm{E}(\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}} - \gamma_i)^2,$$

with different weighting schemes as presented in Table 2. It means that although the $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})$ in some areas is slightly greater than the corresponding $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}})$ which is based on the ML estimates of x_i's, however, the sum of the weighted $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}})$ is smaller than the corresponding sum of the weighted $\mathrm{MSPE}(\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}})$.

TABLE 1: : Numerical values of the MSPE of $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PB}}$, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$ and $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}$ and the empirical MSPE of $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$ and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$

		MSPE				EMSPE			
Area	n_i	$\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PB}}$	$\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}$	$\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PB}}$	$\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PB}}$	$\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$	$\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$
1	1	86.58	27.72	53.54	15.21	204.50	30.79	60.29	42.63
2	5	14.86	12.46	12.74	8.93	49.44	13.92	15.27	14.02
3	1	86.58	27.72	53.54	15.21	185.36	30.16	56.97	41.33
4	2	40.18	21.71	28.30	12.62	89.90	24.65	32.89	26.31
5	4	18.79	14.04	15.41	9.86	52.38	15.40	18.07	15.73
6	3	25.65	14.43	19.76	11.04	54.59	16.51	22.22	17.82
7	1	86.58	30.47	53.54	15.21	196.91	33.20	60.72	43.54
8	3	25.65	14.97	19.76	11.04	55.05	16.88	22.10	17.98
9	2	40.18	17.56	28.30	12.62	85.08	20.35	32.04	23.73
10	3	25.65	17.00	19.76	11.04	74.50	19.13	23.32	19.86
11	2	40.18	25.59	28.30	12.62	120.66	26.55	31.87	26.73
12	1	86.58	27.72	53.54	15.21	182.17	30.77	59.18	41.13
13	6	12.33	11.19	10.93	8.17	55.20	12.48	13.32	12.71
14	7	10.58	8.97	9.60	7.53	31.04	10.20	11.20	10.34
15	8	9.29	9.07	8.59	6.98	50.56	10.10	10.70	10.37
16	4	18.79	14.74	15.41	9.86	64.02	16.29	17.82	16.29
17	5	14.86	10.93	12.74	8.93	37.31	12.60	14.60	12.92
18	6	12.33	11.64	10.93	8.17	57.54	12.94	13.59	13.06
19	7	10.58	9.86	9.60	7.53	50.77	10.77	11.52	11.05
20	8	9.29	8.05	8.59	6.98	24.68	9.16	9.84	9.26

As indicated in (10), the James-Stein estimator employs the information from the data in two ways. First, it combines the information in $\bar{X}_i$ and $\bar{y}_i$ ($i = 1, \ldots, m$), called $\bar{Z}_i^*$. Then, it uses the information of all areas to estimate each area covariate, $x_{i\mathrm{JS}}$ ($i = 1, \ldots, m$). This method improves the GS and the naive estimator in terms of combining the information of $\bar{X}_i$ and $\bar{y}_i$ ($i = 1, \ldots, m$). It also has an advantage over the DRT estimator as it utilizes the information of the whole dataset to estimate the specific area covariate.

5.2. Simulation Study

In this section, we conduct a simulation study to evaluate the performance of the new PEB predictor, $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$. To this end, we first compute $\gamma_i^{(r)}$'s based on the small area populations in the rth iteration. Then, we find $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}(r)}$, $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}(r)}$, $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}(r)}$, and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}(r)}$ from the sample units. In order to find the pseudo-empirical Bayes, model parameters are also estimated using the consistent estimators defined in Section 3. Table 3 shows the model parameters estimates and their corresponding biases and mean squared errors (MSE). The empirical MSPE of $\widehat{\gamma}_i^{\mathrm{PEB}}$ for different methods ($\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{i\mathrm{GS}}^{\mathrm{PEB}}$, $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$, and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$) is defined as

$$\mathrm{EMSPE}(\widehat{\gamma}_i^{\mathrm{PEB}}) = \frac{1}{R}\sum_{r=1}^{R}(\widehat{\gamma}_i^{\mathrm{PEB}(r)} - \gamma_i^{(r)})^2.$$

TABLE 2: : Numerical values of the weighted MSPE of $\widehat{\gamma}^{\text{PB}}_{\text{GS}}$, $\widehat{\gamma}^{\text{PB}}_{JS}$, $\widehat{\gamma}^{\text{PB}}_{\text{DRT}}$, and $\widehat{\gamma}^{\text{PB}}_{NAI}$ and the weighted EMSPE of $\widehat{\gamma}^{\text{PEB}}_{\text{GS}}$, $\widehat{\gamma}^{\text{PEB}}_{JS}$, $\widehat{\gamma}^{\text{PEB}}_{\text{DRT}}$, and $\widehat{\gamma}^{\text{PEB}}_{NAI}$ for different weight schemes

Weighted MSPE					Weighted EMSPE				
Weighting Schemes	$\widehat{\gamma}^{\text{PB}}_{\text{GS}}$	$\widehat{\gamma}^{\text{PB}}_{\text{JS}}$	$\widehat{\gamma}^{\text{PB}}_{\text{DRT}}$	$\widehat{\gamma}^{\text{PB}}_{\text{NAI}}$	Weighting Schemes	$\widehat{\gamma}^{\text{PEB}}_{\text{GS}}$	$\widehat{\gamma}^{\text{PEB}}_{\text{JS}}$	$\widehat{\gamma}^{\text{PEB}}_{\text{DRT}}$	$\widehat{\gamma}^{\text{PEB}}_{\text{NAI}}$
$W_i = \frac{1/(\sigma_u^2+\frac{\sigma_e^2}{N_i})}{\sum_i 1/(\sigma_u^2+\frac{\sigma_e^2}{N_i})}$	32.76	16.54	23.06	10.64	$W_i = \frac{1/(\widehat{\sigma}_u^2+\frac{\widehat{\sigma}_e^2}{N_i})}{\sum_i 1/(\widehat{\sigma}_u^2+\frac{\widehat{\sigma}_e^2}{N_i})}$	83.78	18.33	26.15	20.87
$W_i = \frac{N_i}{\sum_i N_i}$	19.77	12.92	15.38	9.10	$W_i = \frac{N_i}{\sum_i N_i}$	59.03	14.41	17.87	15.36
$W_i = \frac{1}{m}$	33.78	16.79	23.64	10.74	$W_i = \frac{1}{m}$	86.08	18.64	26.88	21.34
$W_i = \frac{1/\sigma_{0i}^2}{\sum_i 1/\sigma_{0i}^2}$	20.96	13.29	16.11	9.27	$W_i = \frac{1/\widehat{\sigma}_{0i}^2}{\sum_i 1/\widehat{\sigma}_{0i}^2}$	63.15	15.17	19.32	16.35

In addition, we decompose EMSPE($\widehat{\gamma}_i^{\text{PEB}}$) as

$$\text{EMSPE}(\widehat{\gamma}_i^{\text{PEB}}) = M_{1i} + M_{2i} + 2M_{3i},$$

where

$$M_{1i} = \frac{1}{R}\sum_{r=1}^{R}(\widehat{\gamma}_i^{\text{PB}(r)} - \gamma_i^{(r)})^2, M_{2i} = \frac{1}{R}\sum_{r=1}^{R}(\widehat{\gamma}_i^{\text{PEB}(r)} - \widehat{\gamma}_i^{\text{PB}(r)})^2,$$

and

$$M_{3i} = \frac{1}{R1}\sum_{r=1}^{R}(\widehat{\gamma}_i^{\text{PB}(r)} - \gamma_i^{(r)})(\widehat{\gamma}_i^{\text{PEB}(r)} - \widehat{\gamma}_i^{\text{PB}(r)}).$$

Table 1 gives the empirical MSPE of the PEB predictors. One can easily observe that the James-Stein predictor outperforms other methods in most areas. The relative efficiency of the PEB based on the James-Stein estimator over its counterpart based on the ML estimator ranges from 105.06% to 195.78%. Also, the relative efficiency of the PEB based on the James-Stein estimator over the PEB based on the method of moments ranges from 269.60% to 664.11%. Finally, the relative efficiency of the PEB based on the James-Stein estimator over the NAI estimator ranges from 100.00% to 138.44%. Table 2 gives the corresponding values of the weighted EMSPE of PEB predictors for different weight schemes.

TABLE 3: : The estimated values of the parameters and their corresponding biases and MSEs

	b_0	b_1	σ_e^2	σ_u^2	σ_η^2
Parameter Estimate	-356.65	4.34	100.21	13.83	24.92
Bias	564.92	2.90	14.81	14.01	3.70
MSE	290057746.75	7609.13	346.54	283.31	21.77

As we discussed in Section 4, the jackknife estimator of MSPE($\widehat{\gamma}_i^{\text{PEB}}$) is nearly unbiased if $M_{3i} \approx 0$ $(i = 1, \ldots, m)$. Table 4 gives the decomposition of the EMSPE for the James-Stein, DRT, GS and NAI predictors. The James-Stein approach gives relatively small values of M_{3i} compared to the GS and the NAI predictors, and it is similar to the DRT predictor.

We have also studied the performance of the weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_i^{\text{PEB}}$), denoted by $mspe_{wJ}$ and $mspe_J$, of the proposed PEB predictor, $\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$, and similar estimators for DRT, GS and NAI methods. The relative bias (RB) of MSPE estimators, $mspe$, is given by

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

where for example to calculate the RB of the weighted jackknife estimation of MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$) for area i we have

$$\text{E}(mspe) = \frac{1}{R}\sum_{r=1}^{R} mspe_{iw\text{J}}^{(r)}(\widehat{\gamma}_{iw\text{JS}}^{\text{PEB}}).$$

Table 5 presents the RB of the weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_i^{\text{PEB}}$) for GS, the James-Stein, DRT, and the naive predictors. The absolute value of the RB of the weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$) is less than 12% for all areas. On the other hand, the absolute value of the RB of the weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_{i\text{DRT}}^{\text{PEB}}$) is less than 15% for all areas while the NAI estimators result in the larger RBs. Interestingly, based on our simulation studies, weighted and unweighted jackknife methods perform very similar in terms of the RB.

TABLE 4: : The values of the components of the EMSPE of $\widehat{\gamma}_{\text{GS}}^{\text{PEB}}$, $\widehat{\gamma}_{\text{JS}}^{\text{PEB}}$, $\widehat{\gamma}_{\text{DRT}}^{\text{PEB}}$, and $\widehat{\gamma}_{\text{NAI}}^{\text{PEB}}$

Area	n_i	M_{1GS}	M_{2GS}	M_{3GS}	M_{1JS}	M_{2JS}	M_{3JS}	M_{1DRT}	M_{2DRT}	M_{3DRT}	M_{1NAI}	M_{2NAI}	M_{3NAI}
1	1	88.58	92.54	11.69	27.94	3.64	-0.39	54.76	6.42	-0.44	88.61	30.43	-38.20
2	5	15.16	32.67	0.81	12.67	2.30	-0.52	12.90	2.33	0.02	15.15	7.37	-4.25
3	1	87.42	79.37	9.29	28.15	3.25	-0.62	53.25	5.80	-1.04	87.47	29.05	-37.60
4	2	40.38	44.51	2.50	22.81	3.31	-0.73	28.89	4.04	-0.02	40.39	17.77	-15.92
5	4	18.65	30.72	1.50	13.88	2.30	-0.39	15.31	2.39	0.19	18.63	8.32	-5.61
6	3	26.16	22.61	2.91	14.59	1.90	0.01	19.86	2.10	0.13	26.13	10.08	-9.19
7	1	89.57	88.73	9.30	31.21	3.72	-0.87	54.78	6.45	-0.26	89.64	31.69	-38.89
8	3	25.58	24.01	2.73	14.97	2.07	-0.08	19.63	2.31	0.08	25.60	10.43	-9.02
9	2	39.91	36.02	4.58	17.87	2.56	-0.04	28.42	3.23	0.19	39.92	15.48	-15.83
10	3	26.24	40.16	4.05	17.41	2.67	-0.47	20.18	2.88	0.13	26.23	11.34	-8.86
11	2	39.40	76.02	2.62	25.24	3.84	-1.26	27.36	4.60	-0.04	39.47	18.86	-15.80
12	1	85.91	78.02	9.12	28.53	3.58	-0.68	53.07	6.63	-0.26	85.83	30.53	-37.61
13	6	12.30	42.94	-0.03	11.18	2.25	-0.48	10.86	2.28	0.09	12.30	6.36	-2.97
14	7	10.58	18.52	0.97	8.99	1.35	-0.07	9.58	1.38	0.12	10.58	3.98	-2.11
15	8	9.17	42.12	-0.37	9.00	1.97	-0.43	8.50	2.06	0.07	9.17	4.59	-1.70
16	4	18.42	42.53	1.53	14.74	2.51	-0.48	15.01	2.61	0.10	18.39	9.29	-5.70
17	5	14.55	19.12	1.82	11.08	1.53	-0.01	12.75	1.60	0.13	14.56	5.53	-3.59
18	6	12.50	45.59	-0.28	11.81	2.43	-0.65	11.02	2.51	0.03	12.48	6.64	-3.03
19	7	10.18	40.34	0.13	9.60	2.03	-0.43	9.17	2.07	0.14	10.18	5.08	-2.10
20	8	9.11	13.99	0.79	8.04	1.15	-0.02	8.53	1.17	0.07	9.10	3.14	-1.49

TABLE 5: : Percent relative bias of the jackknife estimators of MSPE and the bias of different PEB estimators of the γ_i's

		Relative Bias of $mspe(\widehat{\gamma}_i)$								Bias of $\widehat{\gamma}_i$			
Area	n_i	GS$_w$	GS$_{uw}$	JS$_w$	JS$_{uw}$	DRT$_w$	DRT$_{uw}$	NAI$_w$	NAI$_{uw}$	GS	JS	DRT	NAI
1	1	–683.51	-684.50	-7.18	-6.74	7.68	7.16	-12.18	-11.69	9.46	4.44	6.10	5.18
2	5	204.20	204.35	4.01	4.17	7.71	7.78	17.32	17.58	4.63	2.98	3.12	2.98
3	1	-713.26	-713.66	-6.17	-5.74	14.13	13.59	-9.90	-9.46	9.20	4.40	5.94	5.09
4	2	-579.42	-581.18	-5.63	-5.45	6.25	5.94	5.45	5.73	6.62	3.98	4.57	4.10
5	4	-60.77	-62.44	4.28	4.44	7.26	7.18	19.17	19.41	4.98	3.13	3.36	3.15
6	3	-1069.78	-1071.91	6.11	6.38	6.48	6.18	21.59	21.80	5.35	3.23	3.75	3.37
7	1	-478.40	-480.36	-11.48	-11.11	8.12	7.59	-13.37	-12.95	9.54	4.62	6.17	5.27
8	3	-1119.88	-1122.37	5.32	5.58	7.87	7.59	20.87	21.10	5.34	3.27	3.74	3.37
9	2	-772.00	-774.38	5.64	5.95	6.38	5.98	13.66	13.89	6.44	3.59	4.51	3.89
10	3	-308.56	-310.60	-0.81	-0.65	5.88	5.71	12.62	12.85	5.74	3.50	3.84	3.56
11	2	36.61	35.57	-6.33	-6.22	12.94	12.70	6.24	6.57	6.93	4.15	4.47	4.14
12	1	-525.68	-526.03	-6.82	-6.40	10.13	9.59	-9.12	-8.70	9.32	4.44	6.09	5.11
13	6	1572.66	1572.53	6.54	6.70	10.38	10.58	16.70	16.96	4.52	2.80	2.89	2.84
14	7	-207.99	-208.51	7.21	7.46	7.78	7.81	21.40	21.58	3.78	2.56	2.67	2.58
15	8	1846.38	1849.17	11.21	11.48	12.28	12.73	18.63	19.00	4.08	2.54	2.62	2.57
16	4	371.57	371.88	1.62	1.76	10.47	10.44	16.87	17.13	5.04	3.24	3.36	3.24
17	5	-669.14	-671.01	3.91	4.16	5.53	5.40	21.84	22.02	4.30	2.84	3.06	2.87
18	6	1121.81	1123.59	5.78	5.94	9.81	10.09	15.51	15.85	4.62	2.87	2.93	2.89
19	7	1008.16	1007.53	11.37	11.58	13.37	13.58	20.22	20.45	4.26	2.60	2.70	2.64
20	8	-371.06	-372.12	6.66	6.95	7.81	7.89	21.12	21.31	3.48	2.41	2.50	2.43

Table 5 also shows the empirical bias of the different PEB predictors. The PEB predictors based on the James-Stein estimate of the true area covariate is approximately equal to the bias of the NAI estimator. DRT and GS have larger biases in estimating the MSPE$(\widehat{\gamma}_i)$. Figure 1 illustrates the EMSPE and bias of the different PEB predictors.

6. APPLICATION

In this section, we apply the proposed approach using a cross sectional study in New Zealand to predict the diastolic blood pressure using the cholesterol level. The data frame contains 10529 observations on 58 variables including age, sex, height, weight, and education. In this study we focus on 222 Maori or Asian female observations. This dataset is available as `xs.nz` in the `VGAMdata package` in R.

To apply our proposed methodology, we first group the underlying population (i.e., the female) based on the age (16-32, 32-42, 42-52, and 52-88), Body Mass Index (BMI) (12.80-23.53, 23.53-25.86, 25.86-28.68, 28.68-88.43), ethnic (Maori or Asian), and smoking status (0 or 1). This results in 43 small areas with the number of samples in areas, denoted by n_i, $i = 1, \dots, 43$, ranges from 0 to 15, with 3 areas having no samples. We assume the sampling error of the covariate in each small area to be negligible in comparison with the measurement error and this seems to be a reasonable assumption due to our refine grouping. Hence, we consider the x_i's as

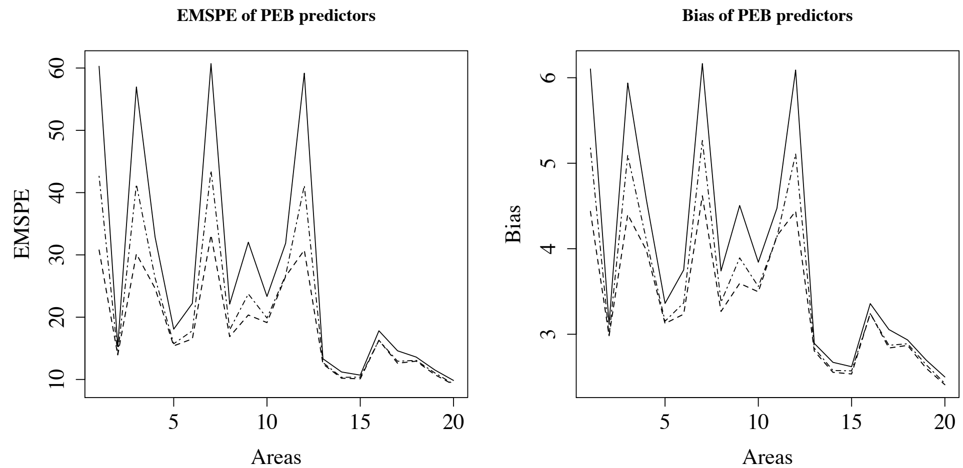

FIGURE 1: : EMSPE and bias of $\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}$ (dashed line), $\widehat{\gamma}_{i\mathrm{DRT}}^{\mathrm{PEB}}$ (solid line) and $\widehat{\gamma}_{i\mathrm{NAI}}^{\mathrm{PEB}}$ (dot line) for the different areas

the true mean of the cholesterol level in each area. The idea is to predict the average diastolic blood pressure in each area assuming that the cholesterol level of each individual is measured with error. To model this measurement error problem, the cholesterol levels of individuals in each area are assumed to be close to the average cholesterol level of the corresponding area with some error. It is worth noting that one can also use the data set to predict the diastolic blood pressure of each individual, however this is not of general interest in the context of the small area estimation and the problem will be studied in a separate work. Figure 2 shows the diastolic blood pressure versus the cholesterol level. We use equations (1) and (2) to model the data where X_{ij} is the observed value of the cholesterol level and Y_{ij} is the diastolic blood pressure for the j'th person in the i'th small area.

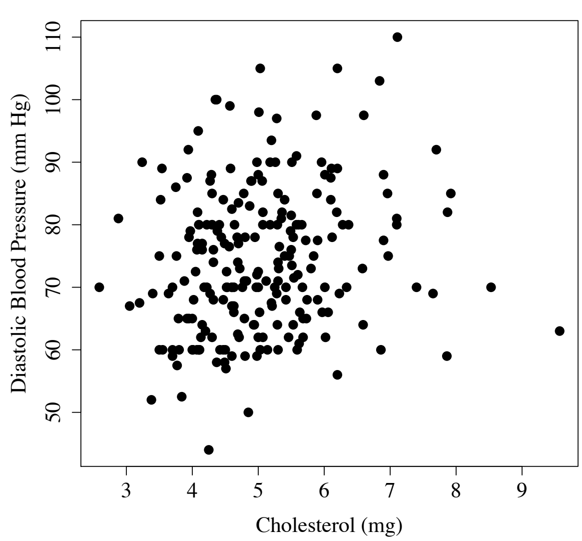

FIGURE 2: : Diastolic Blood Pressure versus Cholesterol

Using the method of moments, we get $\widehat{b}_0 = 24.62$, $\widehat{b}_1 = 9.86$, $\widehat{\sigma}_e^2 = 93.39$, $\widehat{\sigma}_u^2 = 26.07$, and $\widehat{\sigma}_\eta^2 = 0.97$. In this data set X_{ij}'s vary from 2 to 10 while Y_{ij}'s vary from 40 to 100. The estimated values of the diastolic blood pressure means, $\widehat{\gamma}_i$, using different approaches are given in Figure 3.

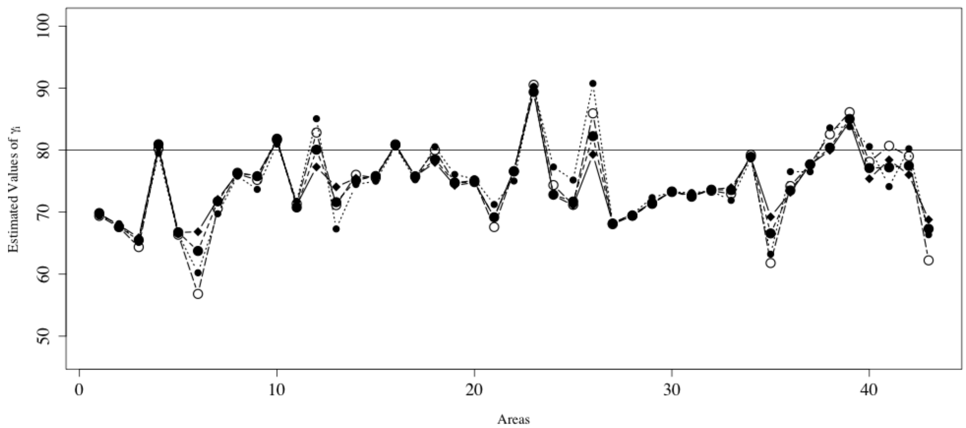

FIGURE 3: : ■▬▬, ●--, • . . . , and ○-. are corresponding to $\widehat{\gamma}^{\text{PEB}}_{i\text{JS}}$, $\widehat{\gamma}^{\text{PEB}}_{\text{NAIVE}}$, $\widehat{\gamma}^{\text{PEB}}_{\text{GS}}$, and $\widehat{\gamma}^{\text{PEB}}_{\text{DRT}}$

For the small areas with no sample units, we use $\widehat{\gamma}^{\text{PEB}}_{i\text{DRT}} = \widehat{b}_0$ as the DRT estimate of the true area covariate does not exist. The same will be used for $\widehat{\gamma}^{\text{PEB}}_{i\text{GS}}$, while for the James-Stein method, a natural estimator is $\widehat{\gamma}^{\text{PEB}}_{i\text{JS}} = \widehat{b}_0 + \widehat{b}_1 \widehat{x}_{i\text{JS}}$. We get $\widehat{\gamma}^{\text{PEB}}_{i\text{JS}} = 74.54$ for areas with no sample units. Finally, Figure 4 presents the weighted and unweighted jackknife estimates of MSPE$(\widehat{\gamma}^{\text{PEB}}_i)$ for different methods.

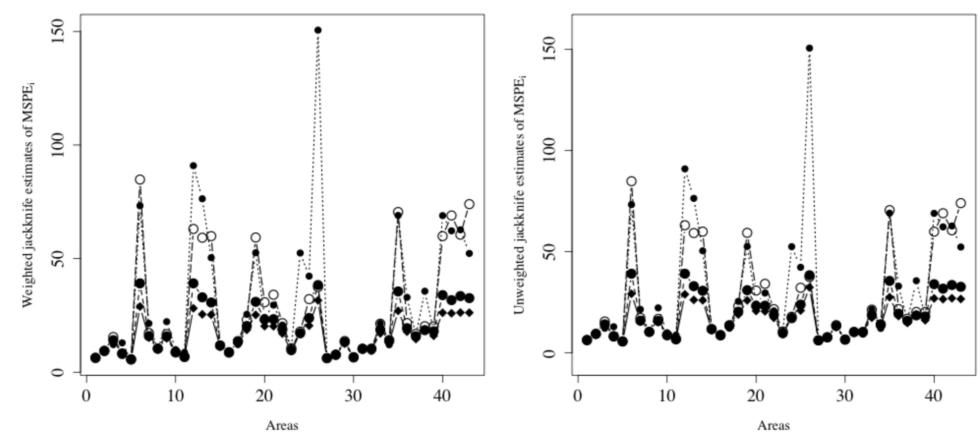

FIGURE 4: : ■▬▬, ●--, • . . . , and ○-. are corresponding to $\widehat{\gamma}^{\text{PEB}}_{i\text{JS}}$, $\widehat{\gamma}^{\text{PEB}}_{\text{NAIVE}}$, $\widehat{\gamma}^{\text{PEB}}_{\text{GS}}$, and $\widehat{\gamma}^{\text{PEB}}_{\text{DRT}}$

Based on our result, there are areas with the high diastolic blood pressure. These areas belong to areas with overweight women. The smoking status does not have a significant effect on the diastolic blood pressure, while Age is also an influential factor as older women with larger BMI suffer more from the high diastolic blood pressure.

7. CONCLUSION

In this paper, we obtained a new pseudo-empirical Bayes (PEB) predictors of small area means based on the James-Stein estimate of the true area-specific covariate for a unit level regression model with the functional measurement error. Our findings show that the James-Stein estimate of the true covariate performs better than the PEB of γ_i based on the ML estimate and also method of moment in terms of the weighted and unweighted mean squared prediction error. Specifically, if the range of the true area-specific covariate is small, the PEB based on the James-Stein performs very well which can be easily justified using the structure of the James-Stein estimator. We performed several simulation studies (not shown here) and observed that the PEB predictor based on the James-Stein estimate always dominates the corresponding PEB predictors (GS and DRT) in terms of MSPE($\widehat{\gamma}_i^{\text{PEB}}$) in small areas with one sample unit even if the range of the true area-specific covariate is large. In addition, the James-Stein method results in a predictor of the small area means for areas with no samples while at least one sampled unit is required to construct $\widehat{\gamma}_{i\text{GS}}^{\text{PEB}}$ and $\widehat{\gamma}_{i\text{DRT}}^{\text{PEB}}$, respectively.

We used the jackknife method to get weighted and unweighted jackknife estimators of MSPE($\widehat{\gamma}_{i\text{JS}}^{\text{PEB}}$). Our simulation results showed that these estimates perform well in terms of the relative bias of $mspe(\widehat{\gamma}_{i\text{JS}}^{\text{PEB}})$. Also, in our simulation studies (not shown here) we observed that the jackknife estimation of the MSPE has a large variance. Thus, we suggest to use this method with caution. This is specially important if one uses other measures of performance for the jackknife estimators such as the coefficient of variation. It is worth mentioning that, although the weighted and unweighted jackknife estimators of the MSPE have almost the same performance in terms of RB, we suggest using the weighted jackknife method as it results in a smaller variance. Further, using the result obtained from the simulation studies, the larger the sample size is in an area, the more similar performance different PEB's show. But, the total number of the areas does not have significant effect on the MSPE or EMSPE of the PEB predictors. Based on our simulation studies, we observed that the estimates of the model parameters using the method of moments (Ghosh and Sinha, 2007) do not perform well.

ACKNOWLEDGEMENTS

We would like to thank the editor and two anonymous referees for their constructive comments which improved the presentation of our results. Mohammad Jafari Jozani and Mahmoud Torabi gratefully acknowledge the research supports of the Natural Sciences and Engineering Research Council (NSERC) of Canada. Elaheh Torkashvand's research is supported by the University of Manitoba Graduate Fellowship (UMGF) and Manitoba Graduate Scholarship (MGS).

BIBLIOGRAPHY

BATTESE, G. E., HARTER, R. M., AND FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* 83:28–36.

CARROLL, R. J., ROEDER, K., AND WASSERMAN, L. (1999). Flexible parametric measurement error models. *Biometrics* 55:44–54.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A., AND CRAINICEANU, C. M. (2010). Measurement error in nonlinear models: a modern perspective. CRC press.

CHEN, S. AND LAHIRI, P. (2002). On mean squared prediction error estimation in small area estimation problems. *In: Proceeding of the Survey Research Methods Section. American Statistical Association, Washington* pp. 473–477.

CROWDER, M. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory* pp. 305–330.

DASGUPTA, A. (2008). Asymptotic theory of statistics and probability. Springer.

DATTA, G. S., RAO, J., AND TORABI, M. (2010). Pseudo-empirical Bayes estimation of small area means under a nested error linear regression model with functional measurement errors. *Journal of Statistical Planning and Inference* 140:2952–2962.

EFRON, B. AND MORRIS, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators part ii: The empirical Bayes case. *Journal of the American Statistical Association* 67:130–139.

EFRON, B. AND MORRIS, C. (1975). Data analysis using stein's estimator and its generalizations. *Journal of the American Statistical Association* 70:311–319.

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.

GHOSH, M., SINHA, K., AND KIM, D. (2006). Empirical and hierarchical Bayesian estimation in finite population sampling under structural measurement error models. *Scandinavian journal of statistics* 33:591–608.

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.

LEHMANN, E. L. AND CASELLA, G. (1998). Theory of point estimation, volume 31. Springer.

RAO, J. N. (2003). Small area estimation, volume 331. Wiley. com.

ROBBINS, H. 1956. An empirical Bayes approach to statistics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* pp. 157–163.

SMALL, C. G. AND YANG, Z. (1999). Multiple roots of estimating functions. *Canadian Journal of Statistics* 27:585–598.

TORABI, M. (2011). Small area estimation using survey weights with functional measurement error in the covariate. *Australian & New Zealand Journal of Statistics* 53:141–155.

WHITTEMORE, A. S. (1989). Errors-in-variables regression using stein estimates. *The American Statistician* 43:226–228.

YBARRA, L. M. AND LOHR, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika* 95:919–931.

APPENDIX

Proof of Theorem 1. Straightforward calculation shows that

$$
\begin{aligned}
&\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\left(\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PEB}}-\widehat{\gamma}_{i\mathrm{JS}}^{\mathrm{PB}}\right)^2\\
&=\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\Big(f_i(B_i-\widehat{B}_i)(\bar{y}_i-b_0-b_1x_{i\mathrm{JS}})-f_i\widehat{B}_i(b_0+b_1x_{i\mathrm{JS}})+f_i\widehat{B}_i(\widehat{b}_0+\widehat{b}_1\widehat{x}_{i\mathrm{JS}})\Big)^2\\
&=\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\Big(f_i(B_i-\widehat{B}_i)(\bar{y}_i-b_0-b_1x_i-b_1(x_{i\mathrm{JS}}-x_i)\\
&+f_i\widehat{B}_i((\widehat{b}_0-b_0)+(\widehat{b}_1-b_1)\widehat{x}_{i\mathrm{JS}}+b_1(\widehat{x}_{i\mathrm{JS}}-x_{i\mathrm{JS}}))\Big)^2\\
&\le\frac{5}{m}\sum_{i=1}^{m}\mathrm{E}\Big(f_i^2(B_i-\widehat{B}_i)^2(\bar{y}_i-b_0-b_1x_i)^2+f_i^2(B_i-\widehat{B}_i)^2b_1^2(x_{i\mathrm{JS}}-x_i)^2\\
&+f_i^2\widehat{B}_i^2(\widehat{b}_0-b_0)^2+f_i^2\widehat{B}_i^2(\widehat{b}_1-b_1)^2\widehat{x}_{i\mathrm{JS}}^2+f_i^2\widehat{B}_i^2b_1^2(\widehat{x}_{i\mathrm{JS}}-x_{i\mathrm{JS}})^2\Big)\\
&\le\frac{5}{m}\sum_{i=1}^{m}\mathrm{E}\Big((B_i-\widehat{B}_i)^2(\bar{y}_i-b_0-b_1x_i)^2+(B_i-\widehat{B}_i)^2b_1^2(x_{i\mathrm{JS}}-x_i)^2\\
&+(\widehat{b}_0-b_0)^2+(\widehat{b}_1-b_1)^2\widehat{x}_{i\mathrm{JS}}^2+b_1^2(\widehat{x}_{i\mathrm{JS}}-x_{i\mathrm{JS}})^2\Big).
\end{aligned}
\tag{1}
$$

The first inequality is due to the well-known partial sums moment inequality (e.g., DasGupta, 2008)

$$
\mathrm{E}|\sum_{i=1}^{n}X_i|^p\le n^{p-1}\sum_{i=1}^{n}\mathrm{E}|X_i|^p,\quad p>1,
\tag{2}
$$

while the second inequality follows from $f_i\le 1$ and $\widehat{B}_i\le 1$, $i=1,\ldots,m$. Under the assumption (13) and the consistency of $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_e^2$, we have

$$
\begin{aligned}
|B_i-\widehat{B}_i|&=\left|\frac{\sigma_e^2}{\sigma_e^2+n_i\sigma_u^2}-\frac{\widehat{\sigma}_e^2}{\widehat{\sigma}_e^2+n_i\widehat{\sigma}_u^2}\right|\\
&=\left|\frac{n_i(\widehat{\sigma}_u^2(\sigma_e-\widehat{\sigma}_e^2)+\widehat{\sigma}_e^2(\widehat{\sigma}_u^2-\sigma_u^2))}{(\sigma_e^2+n_i\sigma_u^2)(\widehat{\sigma}_e^2+n_i\widehat{\sigma}_u^2)}\right|\\
&\le\frac{k_0}{\sigma_e^2+\sigma_u^2}(|\widehat{\sigma}_e^2-\sigma_e^2|+|\widehat{\sigma}_u^2-\sigma_u^2|)\to 0,\ \text{as}\, m\to\infty.
\end{aligned}
\tag{3}
$$

Thus

$$
\max_{1\le i\le m}|B_i-\widehat{B}_i|^2\to 0,\ \text{as}\ m\to\infty.
\tag{4}
$$

In addition,

$$\max_{1\le i\le m}|B_i-\widehat{B}_i| \le \frac{k_0}{\sigma_e^2+\sigma_u^2}(\widehat{\sigma}_e^2+\sigma_e^2+\widehat{\sigma}_u^2+\sigma_u^2)$$
$$\le k_0+\frac{k_0}{\sigma_e^2+\sigma_u^2}(\text{MSB}_y+\text{MSW}_y). \quad (5)$$

Furthermore,

$$m^{-1}\sum_{i=1}^m \text{E}(\bar{y}_i-b_0-b_1x_i)^2 = m^{-1}\sum_{i=1}^m(\sigma_u^2+\frac{\sigma_e^2}{n_i})$$
$$\le \sigma_u^2+\sigma_e^2 = O(1). \quad (6)$$

Also,

$$\frac{1}{m}\sum_{i=1}^m|B_i-\widehat{B}_i|^2(\bar{y}_i-b_0-b_1x_i)^2 \le \max_{1\le i\le m}|B_i-\widehat{B}_i|^2\frac{1}{m}\sum_{i=1}^m(\bar{y}_i-b_0-b_1x_i)^2. \quad (7)$$

Now, from (4) and (6), we have

$$\frac{1}{m}\sum_{i=1}^m|B_i-\widehat{B}_i|^2(\bar{y}_i-b_0-b_1x_i)^2 \xrightarrow{p} 0 \quad \text{as } m\to\infty. \quad (8)$$

To use the Dominated Convergence Theorem (DCT), first note that by using (7), (2) and the Cauchy-Schwarz inequality, we obtain

$$\text{E}\left[\frac{1}{m}\sum_{i=1}^m|B_i-\widehat{B}_i|^2(\bar{y}_i-b_0-b_1x_i)^2\right]$$
$$\le \text{E}\left[\max_{1\le i\le m}|B_i-\widehat{B}_i|^2\frac{1}{m}\sum_{i=1}^m(\bar{y}_i-b_0-b_1x_i)^2\right]$$
$$\le \sqrt{\text{E}\left(\max_{1\le i\le m}|B_i-\widehat{B}_i|^2\right)^2\text{E}\left(\frac{1}{m}\sum_{i=1}^m(\bar{y}_i-b_0-b_1x_i)^2\right)^2}$$
$$\le \sqrt{\text{E}\left(\max_{1\le i\le m}|B_i-\widehat{B}_i|\right)^4\text{E}\left(\frac{1}{m}\sum_{i=1}^m(\bar{y}_i-b_0-b_1x_i)^4\right)}. \quad (9)$$

Also,

$$\text{E}\left(\frac{1}{m}\sum_{i=1}^m(\bar{y}_i-b_0-b_1x_i)^4\right) = 3\frac{1}{m}\sum_{i=1}^m(\sigma_u^2+\frac{\sigma_e^2}{n_i})^2$$
$$\le 3(\sigma_u^2+\sigma_e^2)^2 = O(1). \quad (10)$$

As (5) shows $\mathrm{E}\left(\max_{1\le i\le m}|B_i-\widehat{B}_i|\right)^4$ is integrable. Therefore

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}|B_i-\widehat{B}_i|^2(\bar{y}_i-b_0-b_1x_i)^2\right)\to 0,\quad \text{as } m\to\infty.$$

In the next step, we show

$$\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}\left[(B_i-\widehat{B}_i)^2(x_{i\mathrm{JS}}-x_i)^2\right]\to 0,\quad \text{as } m\to\infty. \tag{11}$$

By using the Cauchy-Schwarz inequality and (5), we show that

$$\frac{1}{m}\sum_{i=1}^{m}\mathrm{E}(x_{i\mathrm{JS}}-x_i)^4 \tag{12}$$

is finite. To this end, since $x_{i\mathrm{JS}}$, $1\le i\le m$, follows a normal distribution, we have

$$\mathrm{E}(x_{i\mathrm{JS}}^4)<\infty,\quad 1\le i\le m. \tag{13}$$

Also, (12) is finite from (2). Now using DCT, (11) follows easily.
To show $\mathrm{E}(\widehat{b}_0-b_0)^2\to 0$, as $m\to\infty$, we first write

$$\mathrm{E}(\widehat{b}_0-b_0)^2=\mathrm{E}\left[\mathrm{var}(\widehat{b}_0|\mathbf{X})+(\mathrm{E}(\widehat{b}_0|\mathbf{X})-b_0)^2\right]. \tag{14}$$

Note that

$$\begin{aligned}\mathrm{var}(\widehat{b}_0|\mathbf{X}) &= \mathrm{var}(\bar{y}-\widehat{b}_1\bar{\mathbf{X}}|\mathbf{X})\\ &= \mathrm{var}(\bar{y})+\mathrm{var}(\widehat{b}_1|\mathbf{X})\bar{\mathbf{X}}^2-2\bar{\mathbf{X}}\mathrm{cov}(\bar{y},\widehat{b}_1|\mathbf{X})\\ &= \frac{\sum_{i=1}^m n_i(\sigma_e^2+n_i\sigma_u^2)}{n_T^2}+\mathrm{var}(\widehat{b}_1|\mathbf{X})\bar{\mathbf{X}}^2-2\bar{\mathbf{X}}\frac{\mathrm{MSB}_x}{\mathrm{MSB}_x-\mathrm{MSW}_x}\frac{\sum_{i=1}^m n_i^2(\bar{\mathbf{X}}_i-\bar{\mathbf{X}})\mathrm{var}(\bar{y}_i)}{n_T(m-1)\mathrm{MSB}_x}\\ &\le \frac{\sigma_e^2+k_0\sigma_u^2}{n_T}+\mathrm{var}(\widehat{b}_1|\mathbf{X})\bar{\mathbf{X}}^2-2\frac{\bar{\mathbf{X}}\sum_{i=1}^m n_i(\bar{\mathbf{X}}_i-\bar{\mathbf{X}})(\sigma_e^2+n_i\sigma_u^2)}{(\mathrm{MSB}_x-\mathrm{MSW}_x)(m-1)n_T},\end{aligned} \tag{15}$$

where $\mathbf{X}$ is the vector of covariates with measurement error. Ghosh and Sinha (2007) showed that

$$\mathrm{var}(\widehat{b}_1|\mathbf{X})\le\frac{\sigma_e^2+k_0\sigma_u^2}{m-1}\frac{\mathrm{MSB}_x}{(\mathrm{MSB}_x-\mathrm{MSW}_x)^2} \tag{16}$$

By (15) and (16), we have $\mathrm{var}(\widehat{b}_0|\mathbf{X})=O(m^{-1})$. Noting that

$$\mathrm{E}\left(\frac{\sigma_e^2+k_0\sigma_u^2}{m-1}\frac{\mathrm{MSB}_x}{(\mathrm{MSB}_x-\mathrm{MSW}_x)^2}+\frac{\sigma_e^2+k_0\sigma_u^2}{n_T}-2\frac{\bar{\mathbf{X}}\sum_{i=1}^m n_i(\bar{\mathbf{X}}_i-\bar{\mathbf{X}})(\sigma_e^2+n_i\sigma_u^2)}{(\mathrm{MSB}_x-\mathrm{MSW}_x)(m-1)n_T}\right),$$

is finite, we have

$$\mathrm{E}(\mathrm{var}(\widehat{b}_0|\mathbf{X}))\to 0,\quad as\, m\to\infty. \tag{17}$$

Using orthogonal transformation introduced by Ghosh and Sinha (2007),

$$
\begin{aligned}
\mathrm{E}(\mathrm{E}(\widehat{b}_0|\mathbf{X}) - b_0)^2 &= \mathrm{E}(\mathrm{E}(\bar{y}) - \mathrm{E}(\widehat{b}_1|\mathbf{X})\bar{X} - b_0)^2 \\
&= \mathrm{E}\left(b_0 + b_1\bar{x} - b_1\bar{\mathbf{X}}\frac{\sum_{i=1}^m n_i\bar{\mathbf{X}}_i(x_i-\bar{x})}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)} - b_0\right)^2 \\
&= b_1^2\mathrm{E}\left(\bar{x} - \frac{\sum_{i=1}^m n_i\bar{\mathbf{X}}_i(x_i-\bar{x})}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)}\bar{\mathbf{X}}\right)^2 \\
&= b_1^2\mathrm{E}\left(\bar{x} - \frac{Z_1Z_2(\sum_{i=1}^m n_i(x_i-\bar{x})^2)^{\frac{1}{2}}}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)}\sqrt{n_T}\right)^2, \qquad (18)
\end{aligned}
$$

where $(Z_1, Z_2, \ldots, Z_m)^\top = \mathbf{C}(\sqrt{n_1}\bar{\mathbf{X}}_1, \ldots, \sqrt{n_m}\bar{\mathbf{X}}_m)^\top$, and $\mathbf{C}$ is an orthogonal matrix with first two rows given by

$$
\left(\sqrt{\frac{n_1}{n_T}}, \ldots, \sqrt{\frac{n_m}{n_T}}\right),
$$

and

$$
\left(\frac{\sqrt{n_1}(x_1-\bar{x})}{\sqrt{\sum n_i(x_i-\bar{x})^2}}, \ldots, \frac{\sqrt{n_m}(x_m-\bar{x})}{\sqrt{\sum n_i(x_i-\bar{x})^2}}\right).
$$

Moreover,

$$
\bar{x} - \frac{Z_1Z_2(\sum_{i=1}^m n_i(x_i-\bar{x})^2)^{\frac{1}{2}}}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)}\sqrt{n_T} \longrightarrow 0,
$$

as $m \to \infty$. Under (13), $\frac{Z_2}{\sqrt{m-1}} \to \sqrt{c}$, $\mathrm{MSB}_x - \mathrm{MSW}_x \to c$ and $\bar{\mathbf{X}} \to \bar{x}$. Similar to Ghosh and Sinha (2007), under the uniform integrability argument, $\mathrm{E}(\mathrm{E}(\widehat{b}_0|\mathbf{X}) - b_0)^2 \to 0$.
Next, we need to show that $\mathrm{E}((\widehat{b}_1 - b_1)^2\frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^2) \to 0$ as $m \to \infty$. Using the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
\mathrm{E}((\widehat{b}_1 - b_1)^2\frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^2) &\le \sqrt{\mathrm{E}(\widehat{b}_1 - b_1)^4\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^2\right)^2} \qquad (19) \\
&\le \sqrt{\mathrm{E}(\widehat{b}_1 - b_1)^4\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^4\right)}, \quad \text{by } C_r \text{ inequality.} \qquad (20)
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
(\widehat{b}_1 - b_1)^4 &= \left(\frac{\mathrm{MSB}_x}{\mathrm{MSB}_x - \mathrm{MSW}_x} \frac{\sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})}{\sum_{i=1}^m n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^2} - b_1 \right)^4 \\
&\leq \left(\frac{1}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)} \sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}}) - b_1 \right)^4 \\
&\leq 8 \left(\left(\frac{1}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)} \sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}}) \right)^4 + b_1^4 \right). \qquad (21)
\end{aligned}
$$

Again, by the Cauchy-Schwarz inequality for integration and (21) we get

$$
\begin{aligned}
&\mathrm{E} \left(\frac{1}{(m-1)(\mathrm{MSB}_x - \mathrm{MSW}_x)} \sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}}) \right)^4 \\
&\leq \sqrt{ \mathrm{E} \left(\frac{1}{\mathrm{MSB}_x - \mathrm{MSW}_x} \right)^8 \mathrm{E} \left(\frac{\sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})}{(m-1)} \right)^8 }. \qquad (22)
\end{aligned}
$$

By (2) and independency of $\bar{y}_i$ and $(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})$, we have

$$
\begin{aligned}
\frac{1}{(m-1)^8} \mathrm{E} \left(\sum_{i=1}^m n_i \bar{y}_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}}) \right)^8 &\leq \frac{m^7}{(m-1)^8} \sum_{i=1}^m \mathrm{E} \left(n_i^8 \bar{y}_i^8 (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^8 \right) \\
&\leq \frac{k_0^8 m^7}{(m-1)^8} \sum_{i=1}^m \mathrm{E} \left(\bar{y}_i^8 \right) \mathrm{E} \left(\bar{\mathbf{X}}_i - \bar{\mathbf{X}} \right)^8 < \infty.
\end{aligned}
$$

The last inequality is due to the fact that $\bar{y}_i$ and $\bar{X}_i - \bar{X}$ follow the normal distribution. Because $\frac{1}{(\mathrm{MSB}_x - \mathrm{MSW}_x)^8} \xrightarrow{p} \frac{1}{c^8}$, as $m \to \infty$, similar to Ghosh and Sinha (2007),

$$
\mathrm{E} \left(\frac{1}{(\mathrm{MSB}_x - \mathrm{MSW}_x)^8} \right) \to \frac{1}{c^8}
$$

as $m \to \infty$, under the uniform integrability argument. Therefore, (22) is finite, and DCT implies $\mathrm{E}(\widehat{b}_1 - b_1)^4 \to 0$. If $\mathrm{E}(\frac{1}{m} \sum_{i=1}^m \widehat{x}_{i\mathrm{JS}})^4 < \infty$, then $\mathrm{E}((\widehat{b}_1 - b_1)^2 \frac{1}{m} \sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^2) \to 0$. Note that

$$
\begin{aligned}
\widehat{x}_{i\mathrm{JS}}^4 &= (\widehat{C}_i \widehat{\bar{Z}}^* + (1 - \widehat{C}_i) \widehat{\bar{Z}}_i^*)^4 \\
&\leq 8(\widehat{C}_i^4 \widehat{\bar{Z}}^{*4} + (1 - \widehat{C}_i)^4 \widehat{\bar{Z}}_i^{*4}), \quad \text{by } C_r \text{ inequality}. \qquad (23)
\end{aligned}
$$

As $0 \leq C_i \leq 1$, $i = 1, \ldots, m$, one can easily show that

$$
\begin{aligned}
\widehat{\bar{Z}}_i^{*4} &\leq 8(\bar{X}_i^4 + (\bar{y}_i - (\bar{y} - \widehat{b}_1(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})))^4) \\
&\leq 8\bar{X}_i^4 + 216(\bar{y}_i^4 + \bar{y}^4 + \widehat{b}_1^4 (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^4), \quad \text{by } C_r \text{ inequality}. \qquad (24)
\end{aligned}
$$

Using the Cauchy-Schwarz inequality for the expectation and similar to (22), $\mathrm{E}(\widehat{b}_1^8) < \infty$. Thus, $\mathrm{E}(\widehat{\bar{Z}}_i^{*4}) < \infty$. In addition, $\mathrm{E}(\widehat{\bar{Z}}_i^{*p}) < \infty$, for $p = 1, 2, 3$. Also, we have

$$\widehat{\bar{Z}}^{*4} = \frac{(\sum_{i=1}^m \widehat{\bar{Z}}_i^* / (\widehat{\sigma}_{0i}^2 + \widehat{\tau}^2))^4}{(\sum_{i=1}^m 1/(\widehat{\sigma}_{0i}^2 + \widehat{\tau}^2))^4} = \frac{\sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \sum_{l=1}^m \widehat{\omega}_{ijkl} \widehat{\bar{Z}}_i^* \widehat{\bar{Z}}_j^* \widehat{\bar{Z}}_k^* \widehat{\bar{Z}}_l^*}{\sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \sum_{l=1}^m \widehat{\omega}_{ijkl}}, \tag{25}$$

where $\widehat{\omega}_{ijkl} = (\widehat{\sigma}_{0i}^2 + \widehat{\tau}^2)^{-1}(\widehat{\sigma}_{0j}^2 + \widehat{\tau}^2)^{-1}(\widehat{\sigma}_{0k}^2 + \widehat{\tau}^2)^{-1}(\widehat{\sigma}_{0l}^2 + \widehat{\tau}^2)^{-1}$. Now, using (25), we can easily observe that

$$\mathrm{E}(\widehat{\bar{Z}}^{*4}) = \mathrm{E}\left(\frac{\sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \sum_{l=1}^m \widehat{\omega}_{ijkl} \widehat{\bar{Z}}_i^* \widehat{\bar{Z}}_j^* \widehat{\bar{Z}}_k^* \widehat{\bar{Z}}_l^*}{\sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \sum_{l=1}^m \widehat{\omega}_{ijkl}}\right) = \mathrm{E}(\max_{1 \le i \le m} \widehat{\bar{Z}}_i^*)^4 \le \mathrm{E}(\max_{1 \le i \le m} \widehat{\bar{Z}}_i^{*4}).$$

From (24), we can simply find a bound for $\widehat{\bar{Z}}_i^{*4}$, $i = 1, \dots, m$. Therefore, noting that mean of finite numbers is finite and using (2) and (23) gives

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}\right)^4 < \infty. \tag{26}$$

To prove the last part of (1), first, we claim $\widehat{x}_{i\mathrm{JS}} \xrightarrow{p} x_{i\mathrm{JS}}$ as $m \to \infty$. Note that the method of moments estimate of variance components gives consistent estimates of the parameter. Thus, it remains to prove $\widehat{\tau}^2$ is a consistent estimator of τ^2.

As Small and Yang (1999) pointed out, based on Crowder (1986), the estimating equations always have one consistent root so, we need to show that (9) has a unique solution. Determining the exact number of roots is not possible due to complicated mathematical form of (9). To this end, we rely on numerical evaluations of (9) by plotting it for each run of the simulation. Numerical studies with different dataset show that there exist only one positive root for this function. Finally, using (2), we have

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^m (\widehat{x}_{i\mathrm{JS}} - x_{i\mathrm{JS}})^2\right) \le \mathrm{E}\left(\frac{2}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^4\right) + \mathrm{E}\left(\frac{2}{m}\sum_{i=1}^m x_{i\mathrm{JS}}^4\right).$$

As $x_{i\mathrm{JS}}$'s are normally distributed, $\mathrm{E}(\frac{2}{m}\sum_{i=1}^m x_{i\mathrm{JS}}^4) < \infty$. By (26), $\mathrm{E}(\frac{2}{m}\sum_{i=1}^m \widehat{x}_{i\mathrm{JS}}^4) < \infty$. Now, using DCT the result is obtained and this completes the proof.

■

Received 9 July 2013
Accepted 10 January 2015