Transformed Fay-Herriot Model with Measurement Error in Covariates

Sepideh Mosaferi[1] Rebecca C. Steorts[2]

[1]Department of Statistics
Iowa State University

[2]Department of Statistical Science and Computer Science
Duke University

August 06, 2020

Introduction

Statistical agencies are often asked to produce small area estimates for skewed variables. When domain sample sizes are too small to support direct estimators, effects of skewness or outliers of the response variable can be large. As such, it is important to appropriately account for the distribution of the response variable given available auxiliary information. Motivated by this issue and in order to stabilize the skewness and achieve normality in the response variable, we propose an area-level log-measurement error model on the response variable.

- We derive an empirical Bayes predictor of positive small area quantities subject to the covariates containing measurement error
- We propose a corresponding mean squared prediction error of EB predictor using a jackknife and a bootstrap method
- We illustrate the results through simulations

Motivation

Census of the Governments provides information on roads, tolls, airports, and other similar information at the local-government level as defined by the United States Census Bureau. The United States National Agricultural Statistics Service provides estimates regarding crop harvests (see Bellow and Lahiri (2011)). The United States Natural Resources Conservation Service provides estimates regarding roads at the county-level (e.g., Wang and Fuller (2003)), and the Australian Agricultural and Grazing Industries Survey provides estimates of the total expenditures of Australian farms (e.g., Chandra and Chambers (2011)). When domain sample sizes are too small to support direct estimators, the effect of skewness can be quite large, and it is critical to account for the distribution of the response variable given auxiliary information at hand.

Area-Level Logarithmic Model with Measurement Error

Consider m small areas and let Y_i $(i = 1, \ldots, m)$ denote the population characteristic of interest in area i, where often the information of interest is a population mean or proportion. A primary survey provides a direct estimator y_i of Y_i for some or all of the m small areas.

- Let $z_i = \theta_i + e_i$, where $z_i := \log y_i$, $\theta_i := \log Y_i$, and e_i is the sampling error distributed as $e_i \sim N(0, \psi_i)$.
- Let $\theta_i = \sum_{k=1}^{p} \beta_k \log X_{ik} + \nu_i$, where X_{ik} is the k-th covariate of the i-th small area, which is unknown but is observed by x_{ik}. The regression coefficient β_k is unknown and must be estimated, and ν_i is the random effect distributed as $\nu_i \sim N(0, \sigma_\nu^2)$, where σ_ν^2 is unknown.

Measurement Error Model for the Positively Skewed X_{ik}'s

Our proposed measurement error model is as follows:

$$w_{ik} := \log x_{ik} = \log X_{ik} + \eta_{ik} \qquad \text{for} \qquad k = 1, ..., p,$$

or in a vector form of

$$\boldsymbol{w}_i = \boldsymbol{W}_i + \boldsymbol{\eta}_i, \qquad \boldsymbol{\eta}_i \sim N_p(\mathbf{0}, \Sigma_i),$$

where $\boldsymbol{w}_i = (w_{i1}, ..., w_{ip})^\top$ and $\boldsymbol{W}_i = (W_{i1}, ..., W_{ip})^\top$ for $W_{ik} = \log X_{ik}$. Note that $\boldsymbol{W}_i$ is non-stochastic within the class of functional measurement error models.

Empirical Bayes Predictor

One can write

$$\begin{cases} z_i = \boldsymbol{W}_i^\top \boldsymbol{\beta} + \nu_i + \boldsymbol{\beta}^\top \boldsymbol{\eta}_i + e_i \\ \theta_i = \boldsymbol{W}_i^\top \boldsymbol{\beta} + \nu_i + \boldsymbol{\beta}^\top \boldsymbol{\eta}_i \end{cases}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$. Thus, for the pair (z_i, θ_i), we have the following joint normal distribution

$$\begin{pmatrix} z_i \\ \theta_i \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \boldsymbol{W}_i^\top \boldsymbol{\beta} \\ \boldsymbol{W}_i^\top \boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 + \psi_i & \boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 \\ \boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 & \boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 \end{pmatrix} \right]$$

We have the following conditional distribution $[\theta_i | z_i]$:

$$\theta_i | z_i \sim N \Big[\boldsymbol{W}_i^\top \boldsymbol{\beta} + \frac{\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2}{\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 + \psi_i} (z_i - \boldsymbol{W}_i^\top \boldsymbol{\beta}),$$

$$\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 - \frac{(\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2)^2}{\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 + \psi_i} \Big]$$

Empirical Bayes Predictor (continued)

Let

$$\theta_i | z_i \sim N\Big(\gamma_i z_i + (1-\gamma_i)\boldsymbol{W}_i^\top \boldsymbol{\beta}, \gamma_i \psi_i\Big),$$

where $\gamma_i = (\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2)/(\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 + \psi_i)$.

The parameter of interest is $Y_i := \exp(\theta_i)$ after transforming from the logarithmic scale back to the original scale. The corresponding Bayes predictor can be defined as $\hat{Y}_i := E(Y_i|z_i)$. The EB predictor of Y_i is

$$\hat{Y}_i^{\mathrm{EB}} = \exp\left\{\hat{\gamma}_i z_i + (1-\hat{\gamma}_i)\boldsymbol{w}_i^\top \hat{\boldsymbol{\beta}} + \frac{\hat{\gamma}_i \psi_i}{2}\right\}.$$

Estimation of Unknown Parameters

We discuss estimation of the unknown parameters $\boldsymbol{\beta}$ and σ_ν^2. Let

$$\sum_{i=1}^{m} \Big[D_i \Big(\boldsymbol{w}_i \boldsymbol{w}_i^\top - \Sigma_i \Big) \Big] \boldsymbol{\beta} = \sum_{i=1}^{m} D_i \boldsymbol{w}_i z_i \qquad (*)$$

Let $\boldsymbol{z} = (z_1, ..., z_m)^\top$ and $\boldsymbol{W}^\top = (\boldsymbol{W}_1, ..., \boldsymbol{W}_m)$. It follows that $\boldsymbol{z} \sim N_m(\boldsymbol{W}\boldsymbol{\beta}, D^{-1})$ where $D^{-1} = \text{diag}(D_1^{-1}, ..., D_m^{-1})$ and $D_i^{-1} = \boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + \sigma_\nu^2 + \psi_i$. The best linear unbiased estimator of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} = \Big(\boldsymbol{W}^\top D \boldsymbol{W} \Big)^{-1} \boldsymbol{W}^\top D \boldsymbol{z} = \Big(\sum_{i=1}^{m} D_i \boldsymbol{W}_i \boldsymbol{W}_i^\top \Big)^{-1} \sum_{i=1}^{m} D_i \boldsymbol{W}_i z_i$$

Note that $E(\boldsymbol{w}_i \boldsymbol{w}_i^\top) = \boldsymbol{W}_i \boldsymbol{W}_i^\top + \Sigma_i$ and $E(\boldsymbol{w}_i) = \boldsymbol{W}_i$.

Estimation of Unknown Parameters (continued)

We estimate $\boldsymbol{\beta}$ from

$$\sum_{i=1}^{m} \left[D_i \left(\boldsymbol{w}_i \boldsymbol{w}_i^\top - \Sigma_i \right) \right] \boldsymbol{\beta} = \sum_{i=1}^{m} D_i \boldsymbol{w}_i z_i$$

Note that D_i is not known as both $\boldsymbol{\beta}$ and σ_ν^2 are unknown. Take $E(z_i - \boldsymbol{w}_i^\top \boldsymbol{\beta})^2 = \sigma_\nu^2 + \psi_i$, then σ_ν^2 can be estimated from

$$m^{-1} \sum_{i=1}^{m} \left(z_i - \boldsymbol{w}_i^\top \boldsymbol{\beta} \right)^2 - m^{-1} \sum_{i=1}^{m} \psi_i \qquad (**)$$

If the above is less than zero, estimate σ_ν^2 as zero. One can estimate $\boldsymbol{\beta}$ and σ_ν^2 by iteratively solving the Eqs. $(*)$ and $(**)$.

Mean Squared Prediction Error of the EB Predictor

Let

$$
\begin{aligned}
M_{1i} &:= E[(\hat{Y}_i - Y_i)^2 | z_i] \\
&= \exp\left\{\psi_i \gamma_i\right\}\left[\exp\left\{\psi_i \gamma_i\right\} - 1\right] \exp\left\{2\left[\gamma_i z_i + (1-\gamma_i)\boldsymbol{W}_i^\top \boldsymbol{\beta}\right]\right\}
\end{aligned}
$$

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

Definition

The MSPE of the EB predictor $\hat{Y}_i^{\text{EB}}$ is

$$
\begin{aligned}
\text{MSPE}(\hat{Y}_i^{\text{EB}}) &= E[(\hat{Y}_i^{\text{EB}} - Y_i)^2 | z_i] \\
&\equiv E[(\hat{Y}_i - Y_i)^2 | z_i] + E[(\hat{Y}_i^{\text{EB}} - \hat{Y}_i)^2 | z_i] \\
&+ 2E[(\hat{Y}_i^{\text{EB}} - \hat{Y}_i)(\hat{Y}_i - Y_i) | z_i] \\
&= M_{1i} + M_{2i} + 2M_{3i}
\end{aligned}
$$

Asymptotic Optimality for Ignoring M_{3i}

Under assumptions $\min_{1\leq i\leq m} n_i \geq 1$, $\max_{1\leq i\leq m} n_i = K < \infty$, and $[p-1]^{-1}\sum_{k=1}^{p}(w_{ik}-\bar{w}_i)^2 \xrightarrow{p} c(>0)$ $\forall i = 1, \ldots, m$, we have

Lemma

$$E\Big(\frac{1}{m}\sum_{i=1}^{m}[(\hat{Y}_i - Y_i)(\hat{Y}_i^{EB} - \hat{Y}_i)|z_i]\Big) \leq E\Big(\frac{1}{m}\sum_{i=1}^{m}\Big|[(\hat{Y}_i - Y_i)(\hat{Y}_i^{EB} - \hat{Y}_i)|z_i]\Big|\Big)$$

$$\leq \frac{1}{m}\sum_{i=1}^{m}\Big[\Big(E[(\hat{Y}_i - Y_i)^2|z_i]\Big)^{1/2}\Big(E[(\hat{Y}_i^{EB} - \hat{Y}_i)^2|z_i]\Big)^{1/2}\Big]$$

$$\leq \max_{1\leq i\leq m}\Big(E[(\hat{Y}_i - Y_i)^2|z_i]\Big)^{1/2}\Big[\frac{1}{m}\sum_{i=1}^{m}\Big(E[(\hat{Y}_i^{EB} - \hat{Y}_i)^2|z_i]\Big)^{1/2}\Big] \xrightarrow{p} 0$$

Jackknife Estimator of MSPE

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

$$\hat{M}_{1i,J} = \hat{M}_{1i} - \frac{m-1}{m}\sum_{j=1}^{m}(\hat{M}_{1i} - \hat{M}_{1i(-j)})$$

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

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

Theorem

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

Parametric Bootstrap Estimator of MSPE

Consider the following bootstrap model:

$$z_i^\star | \boldsymbol{w}_i^\star, \nu_i^\star \overset{ind}{\sim} N(\boldsymbol{w}_i^{\star\top}\hat{\boldsymbol{\beta}} + \nu_i^\star, \psi_i)$$

$$\boldsymbol{w}_i^\star \overset{ind}{\sim} N_p(\boldsymbol{W}_i, \Sigma_i)$$

$$\nu_i^\star \overset{ind}{\sim} N(0, \hat{\sigma}_\nu^2)$$

Our proposed estimator of $\mathrm{MSPE}(\hat{Y}_i^{\mathrm{EB}})$ is

$$\mathrm{mspe}_B(\hat{Y}_i^{\mathrm{EB}}) = 2M_{1i}(\hat{\sigma}_\nu^2, \hat{\boldsymbol{\beta}}) - E_\star[M_{1i}(\hat{\sigma}_\nu^{\star 2}, \hat{\boldsymbol{\beta}}^\star)|z_i^\star] + E_\star[(\hat{Y}_i^{\mathrm{EB}\star} - \hat{Y}_i^{\mathrm{EB}})^2|z_i^\star]$$

Theorem

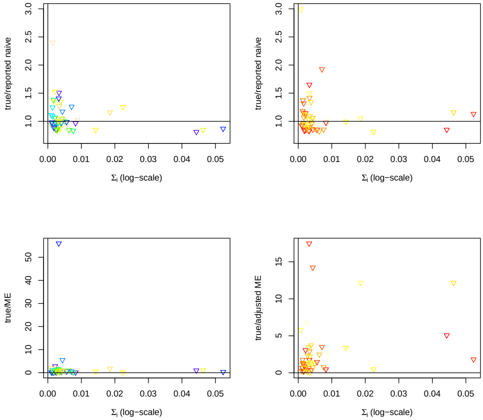
The bootstrap estimator of the MSPE has bias of order $O(m^{-1})$, such that

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

Comparison of Predictors from CoG

State	y_i	$\tilde{Y}_i$	$\hat{Y}_i^{EB}$	EMSE(y_i)	EMSE($\tilde{Y}_i$)	EMSE($\hat{Y}_i^{EB}$)
RI	191.641	202.928	204.907	6.523	5.390	5.103
AK	132.301	120.112	123.684	4.501	6.153	5.793
NV	420.299	422.992	431.912	8.077	8.170	8.449
MD	824.939	784.779	794.784	8.050	5.524	6.503
DE	64.273	72.674	73.130	3.003	5.113	5.182
...	...	...	...	...	...	...
KS	30.804	30.253	30.321	1.859	2.253	2.208
CA	238.284	248.558	248.800	6.991	6.244	6.223
TX	221.105	204.983	205.689	2.570	5.965	5.892
PA	81.653	83.551	83.668	2.888	3.628	3.666
IL	64.650	67.278	67.172	1.490	3.110	3.064

Comparison of Predictors from Model-Based Simulation



Up-Left) ratio of true to reported standard error for the naive model w/o covariate v.s. Σ_i in log-scale. Up-Right) ratio of true to reported standard error for the naive model w. covariate v.s. Σ_i in log-scale. Down-Left) ratio of true to predicted MSE for ME model w/o adjustment v.s. Σ_i in log-scale. Down-Right) ratio of true to predicted MSE for ME model w. adjustment v.s. Σ_i in log-scale.

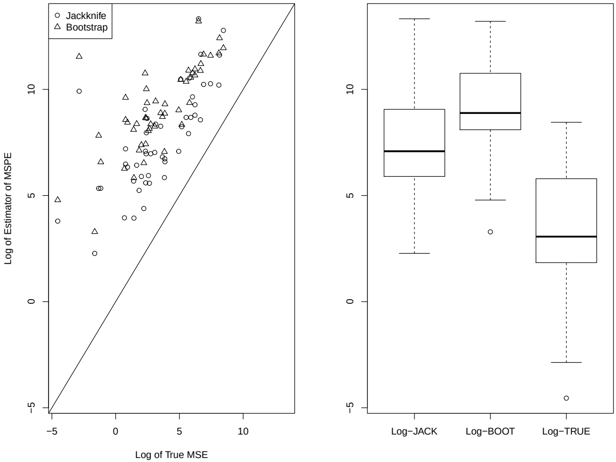
Main Investigation

The EB predictors cannot uniformly outperform the direct estimators due to propagated errors in the term $\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta}$, which is present in the term γ_i in the EB predictors through the simulations.

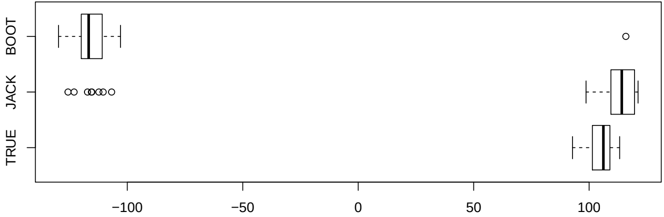
$$
\begin{aligned}
E[\hat{Y}_i^{\mathrm{EB}} - Y_i]^2 &= E\left[\exp\left\{\hat{\gamma}_i z_i + (1-\hat{\gamma}_i)\boldsymbol{w}_i^\top \hat{\boldsymbol{\beta}} + \frac{\hat{\gamma}_i \psi_i}{2}\right\} - Y_i\right]^2 \\
&= \exp\left(2\,\boldsymbol{W}_i^\top \boldsymbol{\beta}\right)\Big\{\exp(\gamma_i\psi_i)(\exp(\gamma_i\psi_i) - 1) \\
&\quad \times \exp\left(2(1-\gamma_i)^2 \boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + 2\gamma_i^2(\sigma_\nu^2 + \psi_i)\right)\Big\}
\end{aligned}
$$

Observe the $\lim_{\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} \to \infty} \exp\left(2(1-\gamma_i)^2 \boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta} + 2\gamma_i^2(\sigma_\nu^2 + \psi_i)\right) = \infty$ under the assumption that $\boldsymbol{\beta} \neq \mathbf{0}$.

Comparison of Jackknife and Bootstrap from CoG



Comparison of Jackknife and Bootstrap from Model-Based Simulation



Comparison of the Proposed Jackknife and Bootstrap Estimators from Model-Based Simulations

We consider the case of $m = 500$.

k	EMSE($\hat{Y}_i^{\text{EB}}$)	mspe$_J$($\hat{Y}_i^{\text{EB}}$)	mspe$_B$($\hat{Y}_i^{\text{EB}}$)	RB$_J$($\hat{Y}_i^{\text{EB}}$)	RB$_B$($\hat{Y}_i^{\text{EB}}$)
0	103.465	108.38	110.455*	4.908	6.991*
20	108.169	119.672	115.892	11.502	7.722
50	110.006	121.191	125.754*	11.184	15.747*
80	112.805	125.672*	129.761*	12.867*	16.955*
100	113.217	123.037*	124.597*	9.819*	11.379*

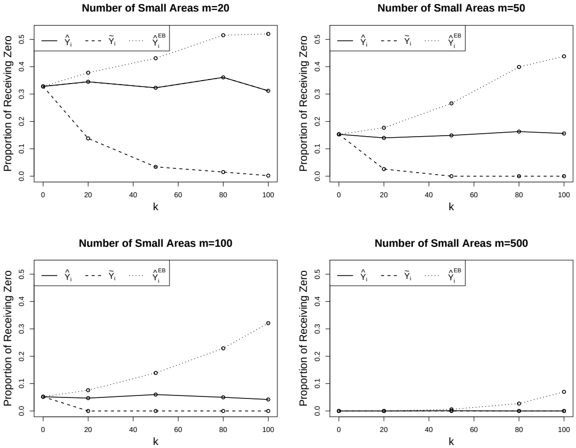
Note: * stands for a negative value.

MSPE Components Comparison

m	k	$\tilde{M}_1$	$\tilde{M}_2$	$\tilde{M}_3$	$\tilde{M}$	$\lvert\tilde{M}_3\rvert/(\tilde{M}_1+\tilde{M}_2)$
500	0	103.680	100.461	101.532	103.921	0.112
	20	108.237	101.446	104.833	108.302	0.033
	50	110.242	107.738	108.316	110.559	0.135
	80	112.656	107.559	110.101*	112.494	0.077
	100	113.091	107.899	110.358*	112.958	0.065

Note: * stands for a negative value.

Ratio of Zeros for σ_ν^2



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

Constrained Empirical Bayes Predictor

Often the weighted sum of the EB predictors are not equal to the weighted sum of the direct estimators. One way to resolve this issue is the benchmarking approach by consideration of the following benchmark constraint $\left[\sum_{i=1}^{m} w_i \hat{Y}_i = \sum_{i=1}^{m} w_i y_i \text{ where } \sum_{i=1}^{m} w_i = 1\right]$ $(\star)$.
A reasonable method for deriving estimators satisfying $(\star)$ is the constrained Bayes procedure, which minimizes the posterior risk subject to $(\star)$. A solution of the conditional optimality can be obtained by the method of Lagrange multipliers. Let $L(\boldsymbol{Y}, \hat{\boldsymbol{Y}})$ be a loss function for estimating $\boldsymbol{Y} = (Y_1, ..., Y_m)^\top$ by an estimator $\hat{\boldsymbol{Y}} = (\hat{Y}_1, ..., \hat{Y}_m)^\top$. The Lagrange function can be defined as

$$LM(\hat{\boldsymbol{Y}}, \lambda) = E\Big[L(\boldsymbol{Y}, \hat{\boldsymbol{Y}})|\boldsymbol{z}\Big] + \lambda\Big\{\sum_{i=1}^{m} w_i \hat{Y}_i - \sum_{i=1}^{m} w_i y_i\Big\},$$

where λ is the lagrange multiplier.

Constrained Empirical Bayes Predictor (Objective)

We propose an appropriate loss-function such that the benchmarked estimators of the positive parameters Y_i's are also positive. Therefore, we recommend a *weighted Kullback-Leibler loss*:

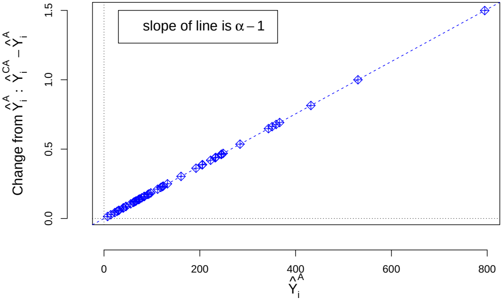
$$L_{KL}^{w}(\boldsymbol{Y}, \hat{\boldsymbol{Y}}) = \sum_{i=1}^{m} w_i \Big\{ \hat{Y}_i - Y_i - Y_i \log(\hat{Y}_i / Y_i) \Big\}.$$

Note that $L_{KL}^{w}(\boldsymbol{Y}, \hat{\boldsymbol{Y}}) \to \infty$ as $\hat{Y}_i \to 0$ or as $\hat{Y}_i \to \infty \, \forall \, i$. We have the following closed form solution

$$\hat{Y}_i^{CEB} = \alpha \, \hat{Y}_i^{EB} \quad \text{where} \quad \alpha = \Big[\sum_{j=1}^{m} w_j y_j \Big] \Big/ \Big[\sum_{j=1}^{m} w_j \, \hat{Y}_j^{EB} \Big]$$

Results of Comparisons

Empirical Measure of Uncertainties [in log-scale]	
$\text{EMSE}\left(\hat{Y}^{EB}\right) = 5.467$	$\text{EMSE}\left(\hat{Y}^{CEB}\right) = 5.465$
$\text{mspe}_{\text{JACK}}\left(\hat{Y}^{EB}\right) = 6.041$	$\text{mspe}_{\text{JACK}}\left(\hat{Y}^{CEB}\right) = 6.044$
$\text{mspe}_{\text{BOOT}}\left(\hat{Y}^{EB}\right) = 9.7096$	$\text{mspe}_{\text{BOOT}}\left(\hat{Y}^{CEB}\right) = 9.713$



Computational Complexities and Some Directions to Future Work

- The propagated error related to the term $\boldsymbol{\beta}^\top \Sigma_i \boldsymbol{\beta}$ can reduce the accuracy of EB predictors themselves and causes over-shrinking of $\hat{\sigma}_\nu^2$ towards zero. This leads to computational instabilities and one has to propose a bias-corrected EB predictor to rectify this.
- Bootstrap estimator is quite lengthy to be implemented and its accuracy is less than the Jackknife one in the non-linear transformed response variable case. This could be potentially fixed by proposing a bias-corrected $\hat{M}_{1i}$ in which its bias is smaller than reciprocal of the number of small areas and computationally easier. Additionally, $\hat{\boldsymbol{\beta}}, \hat{\sigma}_\nu^2$ should be calculated to as close to machine precision as possible; otherwise, Jackknife and Bootstrap may result in a decreased accuracy.

Thank You!

Questions: mosaferi@iastate.edu